BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY
KARL PEARSON

VOLUME X

APRIL 1914 TO MAY 1915

CAMBRIDGE
AT THE UNIVERSITY PRESS

LONDON: FETTER LANE, E.C.
(C. F. CLAY, MANAGER)
AND H. K. LEWIS, GOWER STREET
WILLIAM WESLEY AND SON, 28, ESSEX STREET, STRAND
EDINBURGH: 100, PRINCES STREET
CHICAGO: UNIVERSITY OF CHICAGO PRESS
BOMBAY, CALCUTTA AND MADRAS: MACMILLAN AND CO., LIMITED
TORONTO: J. M. DENT AND SONS, LIMITED
TOKYO: THE MARUZEN-KABUSHIKI-KAISHA

Cambridge:
PRINTED BY JOHN CLAY, M.A.
AT THE UNIVERSITY PRESS.

CONTENTS OF VOL. X.

Memoirs.

Miscellanea.

Plates.

Corrigenda and Addenda.

The following corrections and additions to the memoir on the Partial Correlation Ratio, pp. 391—411, have been received from Mr L. Isserlis:

Page 396. Eqn (28) *for* $\frac{z}{\sigma_z}$ *read* $\frac{\bar{z}_{xy}}{\sigma_z}$; Eqn (30) *for* $r_{xy}=$ *read* $r_{xz}=$; Eqn (35) *for* $\frac{z}{\sigma_z}$ *read* $\frac{\bar{z}_{xy}}{\sigma_z}$.

Page 397. Eqn (37) *for* $r_{xy}=$ *read* $r_{xz}=$; Eqn (38) *for* cq_{xy} *read* cq_{xy^2}; Line following Eqn (45) *for* $d=-c\gamma_{xy}$ *read* $d=-cr_{xy}$; Eqn (46) *for* $\gamma_3' q_{xy}$ *read* $\gamma_3' q_{xy^2}$; Line following Eqn (46) *for* ${}_{xy}R_x^2$ *read* ${}_{xy}R_z^2$.

Page 402. Line 8 from top: after,

$$ {}_{xy}H_z^2 - {}_y\eta_z^2 = S_y \left\{ \left(a + \frac{c_y}{\sigma_y} \right)^2 \left(1 - {}_y\eta_x^2 \right) \frac{x_y}{N} \right\} $$

insert "provided the arrays of x for constant y are homoscedastic, or sufficiently so, for this to be an approximation."

Line 6 from bottom *for* $\gamma_3'\phi - r_3\theta$ *read* $\gamma_3'\phi - \gamma_3\theta$.

Expression at foot of page, factor $(r_{xy} q_{xy^2} - q_{x^2y})$ of first term in large curled bracket *should read* $(r_{xy} q_{x^2y} - q_{xy^2})$ and factor $(r_{xz} q_{x^2y} - r_{yz} q_{xy})$ of expression in numerator of last line of same bracket *should read* $(r_{xz} q_{x^2y} - r_{yz} q_{xy^2})$.

Page 403. In denominator of second term of expression in first line *for* $1 - {}_y\eta_x^2$ *read* $1 - {}_x\eta_y^2$.

Page 405. In denominator of left-hand side of Eqn (70) *for* $q_{x^2y^2} - r^2$ *read* $q_{x^2y^2} - r^2_{xy}$.

Page 406. Line below Eqn (71) and in Eqn (72) *for* $\frac{2r^2_{xy}}{1+r^2_{xy}}$ *read* $\frac{1+r^2_{xy}}{2r^2_{xy}}$.

Vol. X. Part I. April, 1914

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY
KARL PEARSON

CAMBRIDGE UNIVERSITY PRESS
C. F. CLAY, MANAGER
LONDON: FETTER LANE, E.C.
EDINBURGH: 100, PRINCES STREET
also
H. K. LEWIS, 136, GOWER STREET, LONDON, W.C.
WILLIAM WESLEY AND SON, 28, ESSEX STREET, LONDON, W.C.
CHICAGO: UNIVERSITY OF CHICAGO PRESS
BERLIN: A. ASHER AND CO.
LEIPSIC: BROCKHAUS
BOMBAY AND CALCUTTA: MACMILLAN AND CO., LIMITED
TORONTO: J. M. DENT AND SONS, LIMITED
TOKYO: THE MARUZEN-KABUSHIKI-KAISHA

Price Ten Shillings net.

[*Issued April 30,* 1914]

BIOMETRIC LABORATORY PUBLICATIONS

Drapers' Company Research Memoirs.

Biometric Series.

I. **Mathematical Contributions to the** Theory of Evolution.—XIII. On the Theory of Contingency and its Relation to Association and Normal Correlation. By KARL PEARSON, F.R.S. *Issued.* Price 4*s. net.*

II. **Mathematical Contributions to the** Theory of Evolution.—XIV. On the Theory of Skew Correlation and Non-linear Regression. By KARL PEARSON, F.R.S. *Issued.* Price 5*s. net.*

III. **Mathematical Contributions to the** Theory of Evolution.—XV. On the Mathematical Theory of Random Migration. By KARL PEARSON, F.R.S., with the assistance of JOHN BLAKEMAN, M.Sc. *Issued.* Price 5*s. net.*

IV. **Mathematical Contributions to the** Theory of Evolution.—XVI. On Further Methods of Measuring Correlation. By KARL PEARSON, F.R.S. *Issued.* Price 4*s. net.*

V. **Mathematical Contributions to the** Theory of Evolution.—XVII. On Homotyposis in the Animal Kingdom. By ERNEST WARREN, D.Sc., ALICE LEE, D.Sc., EDNA LEA-SMITH, MARION RADFORD, and KARL PEARSON, F.R.S. *Shortly.*

VI. **Albinism in Man.** By KARL PEARSON, E. NETTLESHIP, and C. H. USHER. Text, Part I, and Atlas, Part I. *Issued.* Price 35*s. net.*

VII. **Mathematical Contributions to the** Theory of Evolution.—XVIII. On a Novel Method of Regarding the Association of two Variates classed solely in Alternative Categories. By KARL PEARSON, F.R.S. *Issued.* Price 4*s. net.*

VIII. **Albinism in Man.** By KARL PEARSON, E. NETTLESHIP, and C. H. USHER. Text, Part II, and Atlas, Part II. *Issued.* Price 30*s. net.*

IX. **Albinism in Man.** By KARL PEARSON, E. NETTLESHIP, and C. H. USHER. Text, Part IV, and Atlas, Part IV. *Issued.* Price 21*s. net.*

Studies in National Deterioration. Price 3*s. net* each (IX excepted).

I. **On the Relation of Fertility in Man** to Social Status, and on the changes in this Relation that have taken place in the last 50 years. By DAVID HERON, M.A., D.Sc. *Issued.* Sold only with complete sets.

II. **A First Study of the Statistics of** Pulmonary Tuberculosis (Inheritance). By KARL PEARSON, F.R.S. *Issued.*

III. **A Second Study of the Statistics of** Pulmonary Tuberculosis. Marital Infection. By ERNEST G. POPE, revised by KARL PEARSON, F.R.S. With an Appendix on Assortative Mating by ETHEL M. ELDERTON. *Issued.*

V. **On the Inheritance of the Diathesis** of Phthisis and Insanity. A Statistical Study based upon the Family History of 1,500 Criminals. By CHARLES GORING, M.D., B.Sc. *Issued.*

VI. **A Third Study of the Statistics of** Pulmonary Tuberculosis. The Mortality of the Tuberculous and Sanatorium Treatment. By W. P. ELDERTON, F.I.A. and S. J. PERRY, A.I.A. *Issued.*

VII. **On the Intensity of Natural Selection** in Man. By E. C. SNOW, D.Sc. *Issued.*

VIII. **A Fourth Study of the Statistics of** Pulmonary Tuberculosis: the Mortality of the Tuberculous: Sanatorium and Tuberculin Treatment. By W. PALIN ELDERTON, F.I.A., and SIDNEY J. PERRY, A.I.A. *Issued.*

IX. **A Statistical Study of Oral Temperatures** in School Children with special reference to Parental, Environmental and Class Differences. By M. H. WILLIAMS, M.B., JULIA BELL, M.A., and KARL PEARSON, F.R.S. *Issued.* Price 6*s. net.*

Questions of the Day and of the Fray. Price 1*s. net* each (VII excepted).

I. **The Influence of Parental Alcoholism** on the Physique and Ability of the Offspring. A Reply to the Cambridge Economists. By KARL PEARSON, F.R.S.

II. **Mental Defect, Mal-Nutrition, and** the Teacher's Appreciation of Intelligence. A Reply to Criticisms of the Memoir on 'The Influence of Defective Physique and Unfavourable Home Environment on the Intelligence of School Children.' By DAVID HERON, D.Sc.

III. **An Attempt to correct some of the** Misstatements made by Sir VICTOR HORSLEY, F.R.S., F.R.C.S., and MARY D. STURGE, M.D., in their Criticisms of the Galton Laboratory Memoir: 'A First Study of the Influence of Parental Alcoholism,' &c. By KARL PEARSON, F.R.S.

IV. **The Fight against Tuberculosis and** the Death-rate from Phthisis. By KARL PEARSON, F.R.S.

V. **Social Problems: Their Treatment,** Past, Present and Future. By KARL PEARSON, F.R.S.

VI. **Eugenics and Public Health.** Lecture to the York Congress of the Royal Sanitary Institute. By KARL PEARSON, F.R.S.

VII. **Mendelism and the Problem of Mental** Defect. I. A Criticism of Recent American Work. By DAVID HERON, D.Sc. *Issued* (Double Number). *Price 2s. net.*

VIII. **Mendelism and the Problem of Mental** Defect. II. The Continuity of Mental Defect. By KARL PEARSON, F.R.S., and GUSTAV A. JAEDERHOLM.

THE ABOVE MAY BE OBTAINED FROM DULAU & CO., LTD., 37, SOHO SQUARE, LONDON, W.

VOLUME X APRIL, 1914 No. 1

BIOMETRIKA

CONGENITAL ANOMALIES IN A NATIVE AFRICAN RACE

BY HUGH STANNUS STANNUS, M.D. Lond., *Medical Officer, Nyasaland.*

(1) I HAVE thought it would be of interest to put on record some observations made by myself in Nyasaland during the past seven years, on the subject which appears as the title of this paper.

These observations relate to members of a native population of Bantu stock, belonging to several main tribes, namely, Mananja, Yao, Ngoni and Tumbuka, with a few references to the Nkonde in the north and the Nguru from the south-east.

My interest in the subject was aroused by the frequency with which some abnormalities were seen and I think the facts I bring forward will go to shew that this unusual incidence is real and not only the result of the ease with which observations may be made among a partially clothed community.

Statistics dealing with the subject, to be of value, must treat of large numbers, such have however only been possible in a few instances to be referred to later. I speak therefore largely from impressions in appraising the rarity or otherwise of any particular condition. It should be remembered in this direction that the cases now to be reported have been met with more or less casually, most of them while travelling on the path or in some village, few in the course of Native Hospital work and none in any Special Department.

Classification is a matter of some difficulty for many reasons and as the number of anomalies to be described is not very large it is perhaps more convenient to consider the various conditions according to the anatomical part affected.

One large section of congenital anomalies, Anomalies of Pigmentation, I have already dealt with (*Biometrika*, Vol. IX. pp. 333—365), and they will not be touched on in the present paper.

(2) Dealing with those deviations from the normal in which there is a change of a more or less general nature, I refer firstly to *Infantilism,* at the same time recognising that such a condition may not constitute a truly congenital anomaly.

To the class designated *Idiopathetic Infantilism* I should relegate a woman aged 22 years seen in 1911 at Zomba who presented the figure and development of a girl of 13. There was no breast development, no pubic or axillary hair and the rounded contours of the body and limbs usually associated with this age in a woman were wanting; menstruation had not commenced. In other respects she appeared normal and her mental development was but little if at all below the average.

(3) In W. Nyasa I encountered a very excellent example of the *Ateliotic Dwarf,* a perfect "little man," a man in miniature 1·25 metres in height. Another case which I think must be considered as one of simple dwarfism is here reproduced:—Samuti, aged 35, a Yao, 1·42 metres high. He is shewn together with a man of 1·85 metres. Samuti shews no other abnormality (Plate I, (1)).

No case of *Cretinism* or *Myxoedematous Dwarfism* has been seen. I may here mention that *Cachetic Infantilism* is well seen in some cases of spinal caries among Natives just as among Europeans.

A paper on "Congenital Humeral Micromely" in the *Nouvelle Iconographie de la Salpêtrière,* T. XXIV. pp. 463—471, Paris 1911, by Dr S. A. Kinnier Wilson and myself, contains references to two cases of *Achondroplasia* in Nyasaland. Since then I have heard of two other cases and seen a fifth:—Etimu, male, aged 25 years, a Yao, son of Masinjiri of Ndindi's near Chipoli, Dedza District. The subject stated that he had no children and that no member of the family was known to have been similarly affected. He is a perfect example of the condition as the photographs will attest, and further remarks are unnecessary (Plate I, (3) and (4)).

The following measurements were made and tracings of his hands are here depicted (Fig. 1):

(1)	Head: maximum length	20·1 cm.
(2)	„ breadth	15·8
(3)	circumference	60·0
(4)	Nose: length, base to root	3·6
(5)	breadth, across nostrils	4·5
(6)	Face: bizygomatic breadth	14·0
(7)	length, nasion to chin	11·3
(8)	„ to commissure of lips	6·7
(9)	Standing height	118·2
(10)	Span of arms	113·3
(11)	Arm: acromion to external condyle of humerus	20

(12)	Forearm: external humeral condyle to tip of ulnar tubercle	17 cm.
(13)	Forearm to tip of middle finger	32
(14)	Leg: top of iliac crest to head of fibula	25
(15)	„ „ „ to external malleolus	48
(16)	„ „ „ to sole of foot	54
(17)	Trunk: upper border of sternum to umbilicus	34
(18)	„ „ „ symphysis pubis	45

Left. Fig. 1. Etimu. Right.

(4) No case of actual *Gigantism* has been seen. Tallness or shortness often runs in families. The tallest man I have ever seen measured 1·92 metres. He was the father of an albinotic child and had internal strabismus but no signs of acromegaly (see Plate I, (2)). Another man who I have not seen but who was measured by Dr Davey at Kota Kota was 2·0 metres in height. No case of *Acromegaly* has been seen by myself.

(5) The following case in the want of development of the lower jaw and zygomatic arches might be considered as the converse to acromegaly (Fig. 2).

From the sketch the subject will at once be recognised as a type of Congenital Idiot, the above-mentioned features and ill-formed pinnae together with the rather bird-like appearance being characteristic.

Jaidi, male, aged 20 years, a Yao of Chumbosa, Bursali, is the second child of a family of three, the elder brother being dead and the younger sister normal. No family history was elicited.

Fig. 2. Jaidi.

The growth of the face is defective as before noted, the zygomatic arches are so little developed that there are practically no cheeks. The descending rami of the jaws converge very considerably so that the floor of the mouth is very narrow and the horizontal rami are so short that the symphysis is situated mid-way between the lower lip and the neck as they lie on one horizontal plane. The palate is high and narrow.

The following measurements were made:

Maximum occipito frontal	19·1 cm.
„ bi-parietal	13·8
Bizygomatic at junction of zygoma with temporal	12·3
Nose: length	4·7
breadth	3·8
Face: nasion to commissure of lips	7·7
„ „ symphysis of chin	11·2

Right external strabismus is present and vision defective.

Though mentally an imbecile with an impaired speech he is an excellent field labourer. He states that no woman would marry him but that he has had sexual intercourse and that he is capable of the act.

A few other cases of *Congenital Idiocy* have been seen and include an example of *Spastic Diplegia*, a *Mongol Idiot* aged 4 years in W. Nyasa district and two microcephalic idiots met with in adjacent villages in Chikala district, in neither of which were factors of etiological interest elicited.

(*a*) Aged 22, male, looked like a boy of 12 in physical development, the head was very small but no measurements were made; the palpebral fissures were markedly slanting downwards and inwards and an internal strabismus was present; the ears and palate were normal; the hands large and like those of a man.

(*b*) A male infant aged one year with so marked a degree of microcephaly as to approach in type anencephaly, the resemblance being the more marked as the protuberant eyes and lips were like those characteristically found in anencephalic monsters (Plate II, (7)).

(6) The following case is given at length (Plate II, (5) and (6)).

Masimosya, aged 19 years (1911), a Yao of Chipi's village Zomba, exhibits a marked want of development of sexual organs (male) associated with large breasts. The general form of the body is that of a woman; the attitude, voice, laugh, facial aspect and expression resemble those of a woman rather than of a man. The teeth are good, the body and limbs well developed and there is a fair deposit of subcutaneous fat. The breasts (see photo) are remarkable, being large, with large well-formed nipples and well-marked areolae, dark in colour. They have started to beome pendulous and resemble exactly those of a nulliparous woman of the same age. The abdomen is well formed and round the umbilicus there is a deposit of fat such as is commonly seen in women; the pelvis appears large. There is some hair in the axillae but none on the face or body. The pubes is rather prominent resembling the female mons veneris and there is some development of hair upon it. The penis is very small, only two inches in length and of infantile type, the glans is covered by a prepuce and there is no deformity. The scrotum is very small indeed and only contains one testicle, the left, which can be felt as a small body about the size of a bean, three-eighths of an inch long. The right testicle is not apparently present in the scrotum or inguinal canal. The scrotum shews no tendency to be divided nor is there anything in the arrangement of the skin to suggest labia. No rectal examination was made.

The subject is insane. He is fairly tractable and good-natured. He has delusions and hallucinations, it is reported, with various phases of the moon, when he is said to travel 15 miles to bathe in a certain stream, etc. He has tried to burn down some houses. I could get very little of his history. The mother and father are said to have been normal; the only other child, a girl, was insane and died in the Central Asylum. The subject once cohabited with a woman who was to have been his wife, but she ran away the next day and I was unable to find out from him if he had any sexual desire. Such is a case which would have been called one of *Partial Hermaphroditism* but in the absence of further data I shall not discuss it.

(7) *Obesity.* No cases of general obesity outside normal limits with possibly a congenital origin have been seen. *Steatopygy* does not occur.

(8) *Symmetrical Lipomatosis* is conveniently considered here though perhaps not strictly within the subject. Three old women have been seen all presenting the same abnormal feature, namely, the presence of symmetrical lipomata in both axillae, each about the size of a small orange. In a fourth case the affection was one-sided, the subject giving a history of the gradual descent of the tumour from the upper aspect of the shoulder into the arm-pit.

That these tumours were lipomata I can only support by clinical examination, they certainly were not of the nature of the pads seen in myxoedema and no signs of that disease were present. There is the possibility that they were accessory breasts but they did not present the characters found in undoubted cases of this condition. These tumours may have a similar pathogeny to the masses seen on either side of the back of the neck of men and specially described by Sir Jonathan Hutchinson; on account of their possible paleogenetic significance I have included notes on these cases here.

(9) *Lymphatism.* Post-mortem examination on a boy 10 years of age who died after receiving a blow on the head revealed a thymus gland of considerable bulk, 4 inches long. The blow had not severed the soft tissues over the skull and in the absence of any other evidence of injury or disease one might suspect the case to be one of lymphatism, an inherent disorder which had predisposed to death. In a second case, that of a woman aged 40 years who died after moderately severe burns, a body 4½ inches long of yellow colour and firm consistency was found lying on the anterior surface of the heart, the apex of this body being at a level with the 2nd costal cartilage.

(10) Coming now to Malformations, there is a well-defined *deformation of the skull* of which I have seen several examples, the main points of which are well shewn in the photographs. The extreme height of the cranium and marked dolicocephaly without bossing of the forehead, while the sides of the vault of the skull are flattened, are characteristic. The photographs depict a boy aged 7, son of Matikwiri, headman of Mlanje, whose two younger sisters are said to resemble him exactly in the deformity present (Plate IV, (12) and (13)).

The second case is a boy aged 15 years, the head measured 21·5 cm. long and 12·5 cm. broad (Plate III, (9)—(11)).

(11) *Congenital Ptosis* is not uncommon and is associated with the typical expression due to this disability. A slight degree of *Epicanthus* may be fairly often observed; more marked, it is sometimes seen associated with obliquity of the palpebral fissures giving a regular mongolian character to the face (Fig. 3).

Buphthalmos has been seen on two occasions in young adults with a history of its congenital nature but nothing else of note; tension normal and vision apparently good.

Microphthalmos was once seen associated with coloboma of the iris and choroid (see below).

Coloboma. This defect was met with in two brothers aged about 18 and 17 years, but neither parent nor, as far as I could ascertain, any other member of the family was similarly affected. Bwanali the elder presented a coloboma of the iris and choroid of the left eye; there was also a small opacity on the posterior surface of the lens, which however could not be traced more deeply but which suggested a remnant of an "arteria centralis."

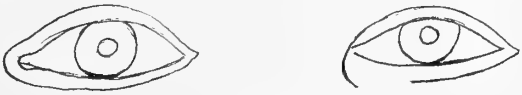
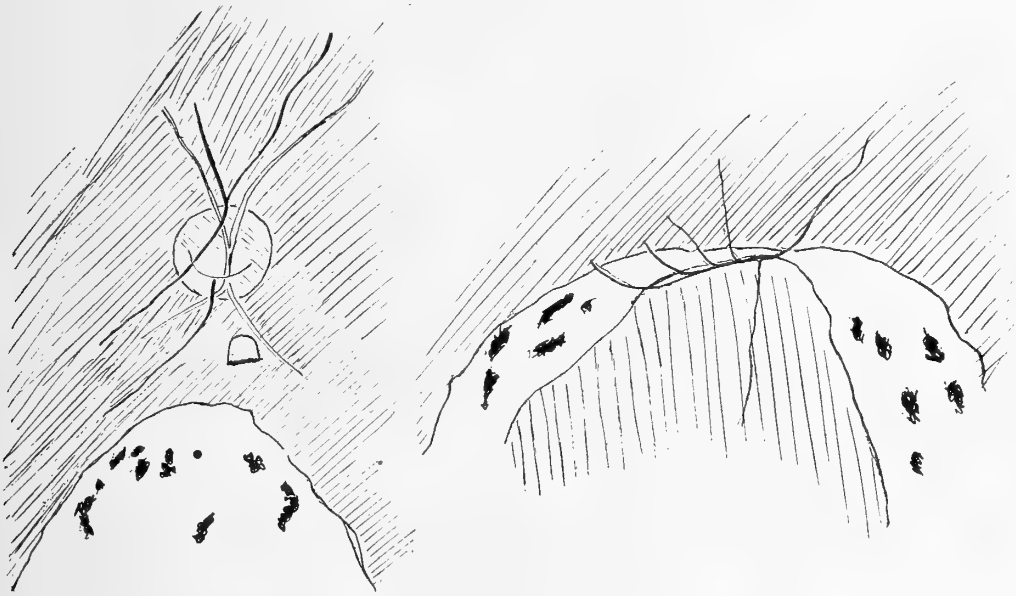


Fig. 3. Epicanthus.

The right cornea shewed some superficial opacities, the iris appeared normal, but examination of the fundus revealed a large white triangular area with the apex near the disc with here and there small masses of pigment. The middle portion of the white area was on a much deeper plane than the rest of the fundus, forming a posterior staphyloma, the whole composing a kind of posterior coloboma (Fig. 4).



Right eye. Left eye.

Fig. 4. Bwanali Coloboma.

This boy also had an accessory nipple.

The younger brother Pete presented on the right side a microphthalmic eye with coloboma of iris and choroid resembling the condition in his brother, with an opaque spot on the posterior surface of the lens. The eye is convergent and vision poor; he counts fingers at one yard. The left eye is normal.

Dermoid Cysts of the Face have been seen in the situations shewn in the sketch (Fig. 5). One of these was excised and found to contain the usual pultaceous mass mixed with hairs. These hairs examined microscopically were found to be spindle shaped, tapering at each end, brown diffuse and granular pigment was present in them.

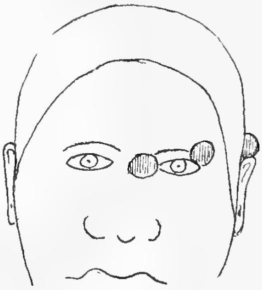


Fig. 5. Dermoid cysts of face.

A relic of the cleft between the median and upper external processes of the foetal face was on one occasion seen as a small pit at the lower extremity of and just external to an epicanthal fold.

(12) *Congenital Naevus.* Only two cases of naevus have been seen. One a woman presented a small naevus just to the left of the middle line on the forehead at the margin of the hairy scalp, 1 cm. in diameter. The second was a man with a similar growth 1 cm. in diameter on the lower lip just to the right of the middle line (Ching'waya of Zomba).

(13) *Ear.* The general conformation of the ear varies a good deal; some of the types are shewn in the sketches (Fig. 6) but all these must be considered as coming within the limits of normal variation. In one case a kind of *Accessory Lobule* was noted; the subject was an albino. A number of persons with *Accessory Auricles*

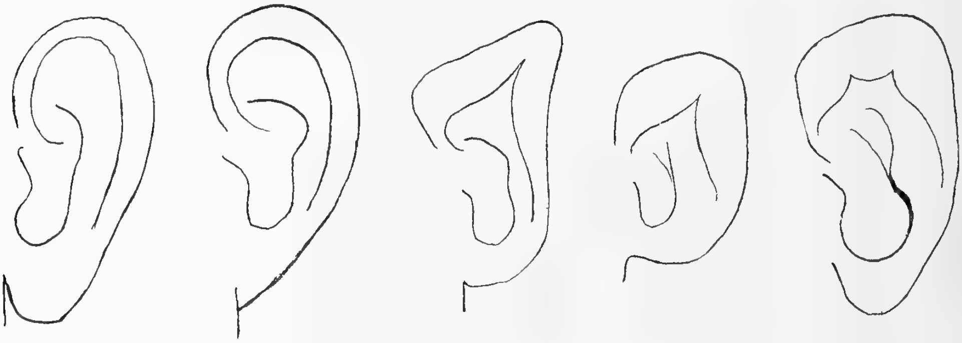


Fig. 6.

have been seen. These consist of little subcutaneous nodules of cartilage forming tubercles one to four in number situated just in front of the tragus, the affection being usually bilateral.

An abnormality seen affecting a woman in N. Nyasa consisted in the direct prolongation of the skin from the side of the head on to the outer surface of the pinna so that the upper margin of the ear was hidden, though easily felt beneath the skin.

Helical fistula. Under this name have been described the remains of the first branchial cleft found as little pits on the helix. The condition is certainly rare in England and persons exhibiting the anomaly are sometimes shewn as interesting cases at medical societies. That heredity plays a part in its incidence is well known as illustrated by a case shewn by Dr Prichard at the Royal Society of Medicine, an infant with symmetrical helical fistulae, whose mother, four siblings, maternal grandmother and two great-aunts all exhibited the same defect. Having noted this same anomaly in quite a number of natives I became interested to ascertain the actual incidence. The statistics given below embody the results of my observations covering nearly 6500 individuals of all tribes. The populations of whole villages were taken so that consecutive unselected persons were dealt with.

Tribe		Number examined	Right	Left	Both sides
N. Angoni	Males	416	7	6	3
	Females	612	13	8	5
Achewa	Males ...	100	—	4	1
	Females...	136	—	3	—
Atonga	Males ...	1941	34	22	12
	Females...	2576	69	53	23
Wankonde*		455	4	8	5
Awemba		48	1	—	1
Anyanja		65	1	1	—
Ahenga		142	3	5	—
Totals		6491	132	110	50

Thus among 6491 individuals of all ages and both sexes a total of 292 were found to have helical fistula (4·5 $^{\circ}/_{\circ}$). It was more commonly unilateral, affecting the right side a little more often than the left, giving percentages of 2·08 and 1·69 respectively and for bilateral cases 0·77 $^{\circ}/_{\circ}$. Taking each sex we see that the proportions between the three numbers are almost the same.

2457 males	...	41	32	16
3324 females	...	82	64	28

---

* These figures were kindly supplied by Dr Davey.

The actual incidence in the two sexes is however greater among females than males in the proportion of 5·2 °/o to 3·6 °/o. An abnormality occurring so frequently as 45 per mille might almost be considered to be a variation within the limits of the normal. The fact remains, however, that it is the persistence of a foetal character and abnormal, if the whole of mankind be taken into consideration.

Dealing more in detail with this defect, there is some variation in the exact site of the fistula; the sketches (Fig. 7) serve to illustrate the extremes of position in three directions.

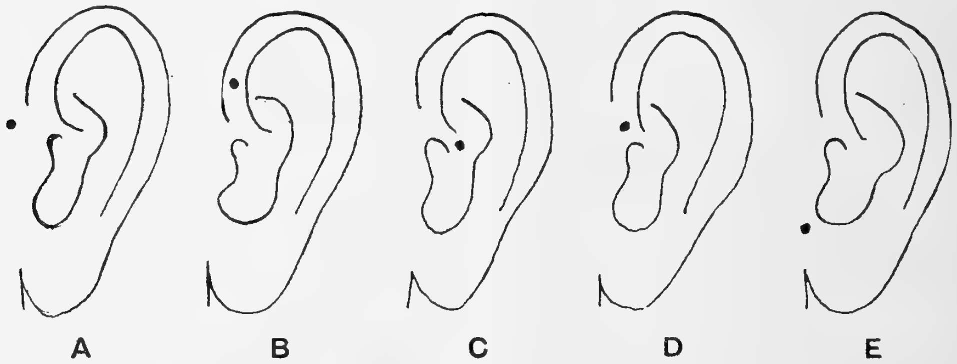


Fig. 7.

Three cases presented two pits on the same side, one each in positions *A* and *B*. In these three cases the affection was bilateral and symmetrical. The common position at which the pit is found is in *D*. In another case not included in the series a pit was observed resembling those above mentioned but situated at the junction of the tragus and lobule as in *E*.

These helical fistulae, which I have described as little pits, consist of a small opening on the skin 1 or 2 mm. in diameter leading into a blind sac 1 or 2 mm. deep; often this sac opens out into a little ampulla which can be seen and felt under the skin. The ampulla and canal are generally filled with a little plug of sebaceous matter.

In three cases the skin in this situation looked like scar tissue and presented a honey-combed appearance, there being several openings into the ampulla giving the impression that an abscess had formed at some past date in the ampulla with consequent loss of tissue.

The fistula is so common and so unremarkable that most tribes have no name for it and one cannot elicit long pedigrees to shew its incidence in families. Cases of heredity were common enough but the type was not necessarily the same in members of the same family; thus a mother with Left Fistula had a child with Right and Left, or again, three brothers were seen two with the Left side affected, the third with Right Fistula.

No malformations in connection with other branchial clefts have been seen.

(14) *Lips, Mouth and Palate.* Most natives shew a well-marked tubercle in the median line on the "red" margin of the upper lip; in a few however this is replaced by a distinct groove which involves the red margin of the lip or only the subjacent fold of mucous membrane (see sketch Fig. 8 and photo, Plate V, (16)).

Fig. 8.

These cases resemble one of a Hindu (recorded in the *Lancet*, Oct. 2, 1909, by Thurston), who besides having the median hare-lip was the subject of polydactylism. In one of my cases there was a considerable gap between the upper central incisors but no further abnormalities were present.

In a single case notching of the upper lip was found to the left of the middle line with a mark running up to the nostril which looked like a scar. There was no question of any operation having been performed, though the condition resembled exactly an artificial repair of a lateral hare-lip (Fig. 9). A similar case has been shewn at the W. Lond. Med. Chir. Society in which there was, besides, a deformity of the nose and a family history of hare-lip. I have only seen one case of ordinary *Hare-Lip*, a Blantyre boy aged 10 years (1909), the affection being left-sided and unassociated with any cleft of the palate (Plate V, (17)). Among 30,000 natives examined in the northern districts of this country no case was seen.

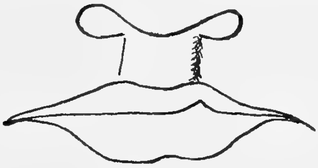


Fig. 9.

No case of typical *Cleft Palate* has come to my notice; on the other hand I have seen three cases which owing to their non-association with defects in the upper lip are of great interest. All three cases, one a boy aged 10 years (1906), the other two adult males, presented complete *Absence of the Premaxilla* and attached teeth. In the boy there was also a *Median Perforation in the hard Palate.* Congenital perforations of the palate apart from clefts are apparently rare in Europe. Dundas Grant (*Roy. Soc. Med.* April 1910) has recorded the case of a girl aged 16 years with a perforation above and to the right of the base of the uvula with no history of trauma or syphilis. Prof. Karl Pearson has drawn my attention to a skull which was brought by Du Chaillu from Fernand Vas in the Congo (see *Biometrika,* Vol. VIII, Plate XXVI); this shews congenital absence of the premaxilla, but the two maxillae have not approximated in the mid-line in front as in my own cases, and we do not know the condition of the soft parts, but it is interesting to see this anomaly from another part of Africa.

(15) *Teeth.* Native children are said to be born sometimes with teeth; it is possible that this is not very rare as there is a common superstition regarding them. I have seen one case with this history, to be mentioned later, as having deformities of the lower extremities. A gap of as much as $\frac{1}{3}$ of an inch between the lower central incisors has been noticed a number of times, the other teeth all being regular and touching one another. A similar condition may be seen also affecting the upper pair of incisors, one that I am not conversant with among Europeans. Among 1500 natives examined for statistical purposes in regard to caries the following numerical abnormalities were noted:

(*a*) Complete reduplication of the set of teeth in an adult, the second set lying on the palatal side of what appeared to be the normal set. I have every reason to believe that this was a case of true reduplication, that is to say, the result of growth from doubled enamel organs and not of retention of the deciduous teeth.

(*b*) Reduplication of upper incisors.

(*c*) Reduplication of right lower bicuspid.

(*d*) Reduplication of both bicuspids in the lower jaw on each side and in the upper jaw on the right side in a woman aged 24 years.

A single case of a *Bifid Extremity to the Tongue* was seen in an albino child.

(16) *Polymazia and Polythelia.* 14 cases of these anomalies have been met with casually, so that I imagine this anomaly by excess is comparatively not uncommon. Short notes of these cases are given below for purposes of comparison:

(*a*) Male adult, accessory nipple springing from the skin at the right sternal edge opposite the 3rd intercostal space, it was large and well formed like a woman's but there was nothing resembling an accessory mamma beneath it.

(*b*) Male aged 45. Insane and suffering from spinal caries. There was a rudimentary accessory nipple in Scarpa's triangle on the right side $1\frac{1}{2}''$ below Poupart's ligament.

(*c*) Adult female, an accessory nipple on the right breast, small but well formed and lying above the one proper to the breast; both are patent and milk can be drawn through both.

(*d*) Adult male, the accessory nipple is situated in a line with the left nipple below it and half-way between it and the costal margin.

(*e*) and (*f*) Two women each had two nipples to the right breast.

(*g*) A young woman was found to have two nipples on the left breast (Plate V, (18)).

(*h*) Male with congenital coloboma iridis mentioned above has an accessory nipple just above and to the inner side of the right nipple.

(*i*) Young male adult has just at the outer edge of the areola of the left breast a very small accessory nipple, and beyond this and above it over the third intercostal space another flat nipple with areola and hairs.

(*j*) Male, presents a rudimentary nipple in the left groin just below the middle of Poupart's ligament.

(*k*) Young adult male shews a small accessory nipple just below and internal to the right nipple; his brother, father, and grandfather are all possessed of the same identical anomaly. The subject has no children, no nephews or nieces.

(*l*) Female in hospital with syphilis has a small accessory nipple springing from the skin of the chest wall just internal to the point of the left pendant breast.

(*m*) A woman with well-formed accessory breast in the right axilla. It is breast-shaped and pendant though there is no nipple. The woman volunteered the fact that it was a breast and said it swelled with pregnancy. The right breast was twice as big as the left.

(*n*) An old woman with symmetrical masses in each axilla resembling rather the symmetrical lipomata mentioned elsewhere: see p. 6. She states that they appeared at puberty and thinks them to be breasts but denies that they enlarged with pregnancy.

In the Japanese this condition has been shewn to be not unrare, and among them tuberculosis has been found to be more frequent than among the normal population. I can only support the idea with one case (No. *b*).

(17) *Meningocoele and Spina Bifida.* No typical case has been noted. A man was seen with a little dipple of the skin over the lower part of the sacrum in the median line having a little fold of skin on either side forming two small vertical lips.

(18) *Penis, Testicle; Hernia.*

Epispadias, hypospadias and extroversion of the bladder have never been seen.

I have seen a boy aged 18 years with a short penis enclosed in a fold of skin from the upper surface of the scrotum (Fig. 10). The boy had other deformities which are described later. When examining a number of recruits I was surprised to find in a large proportion the right testicle hanging lower than the left, the reverse of what is known to occur among Europeans. On examination of 400 consecutive men, adults, between the ages of 30 and 40 years, I found in 166 or 41·5 °/₀ the right testicle lower than the left. In the remainder or 58·5 °/₀ the right testicle was on a level with the left, or rather higher in the scrotum. I also got the impression that, associated with right lower testicles, the testicles and penis were large. In another series of 280 men, the left was lower than or on the same level as the right in 185; the right lower in 88. There were two cases of left cryptorchidism, one of right cryptorchidism; one each left and right hydrocoeles and two right inguinal bubonocoeles.

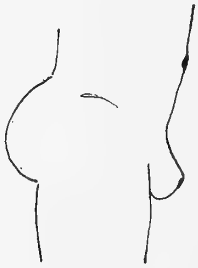


Fig. 10. Boy aged 18.

I have come across a number of cases of undescended testis among other natives, in some associated with a swelling in the inguinal canal, in others there was complete cryptorchidism. Inguinal hernia is not infrequent in adult males but I can give no figures relating to a large number of persons. In a single man it was associated with umbilical hernia. I have never seen a femoral hernia. Umbilical hernia is common enough especially in children. The following figures though small in number give some idea of the frequent incidence of the condition. They refer to all the children in a single village and may therefore be said to be unselected in any way.

Age	–	+	+ +
0— 1 year	18	13	6
1— 2 years	44	27	12
3—10 „	102	44	10
Totals	164	84	28

=276,

giving roughly 60 °/₀ $\underbrace{30\,°/_0 \quad 10\,°/_0}_{40\,°/_0}$

That the hernia diminishes in size even to disappearance after childhood, as indicated by these figures, is certainly true as the same incidence is undoubtedly not found among adults. The protrusion is sometimes very marked and takes the shape of the finger of a glove, some several inches long and curving downwards (Plate IV, (14)). Writing recently E. M. Corner in doubting the commonness of congenital sacs in hernia in general, as insisted on by some writers, has shewn in a series extending to between two and three thousand observations that herniae in children are often multiple and associated particularly with a ventral hernia, a diastema, which though very rare at birth is common in young children and of the nature of a true hernia. He believes that this ventral protrusion, which is certainly not congenital, is caused by increased abdominal pressure due to gaseous distension of the bowels the result of fermentative processes, and that other herniae are due to the same cause. Among native children abdominal distension is almost the rule, "pot-bellied" is an expression always used in speaking of them. This distension is due largely, I believe, to fermentative processes, and also a second factor, absent in European children, namely enlarged spleen. Of 50 children under the age of 5 years taken from among those with umbilical hernia, 43 or 86 °/₀ were found to have the ventral protrusion as described by Corner. 18 of these had enlarged spleens and 20 shewed a considerable abdominal distension. In none was any other hernia found. In these cases we see *par excellence* the effect of intra-abdominal pressure, in producing first ventral hernia and secondly umbilical hernia. The weakness of the umbilical scar is due, I have little doubt, to the method of treating the cord at birth. The custom prevailing among many is to bind the whole cord and placenta on to the child's abdomen till it separates; with others the greater part of the cord is so treated after severing the placenta; in any case there must be considerable tension, I think, at the umbilicus and sepsis is more likely to occur. Cursham Corner has said that the size of the bulging is proportional to the length of cord left proximal to the ligature, and the same principle adapted to natives who use no ligature may be true, and thus account for the very "long" umbilical hernias.

I am therefore inclined to agree with Corner that the umbilical and the ventral herniae of children are due largely to intra-abdominal pressure, but though my numbers are small, the absence of any other hernia among my cases must be taken to mean that for their production there is another factor to be taken into account, and that is, I believe, in Corner's cases some congenital structural anomaly, namely a congenital sac, and, conversely, I think congenital sacs are uncommon among natives of this country.

(19) *Malformations of the Extremities.* Various forms of *Congenital Talipes* are met with which call for no special comment.

A peculiar condition characterised by symmetrical shortening of the humeri has been observed and forms the subject of a paper by Dr S. A. Kinnier Wilson and myself referred to above; certain deformities of the hands and feet are also therein dealt with.

Since this paper was written I have seen three other cases of *Congenital Humeral Micromely,* one of which I mention here as there is a family history of the defect, a point of some interest and one which I had not elicited in previous cases.

Gobedi, male, aged 22 years, a Yao employed as a machila carrier in Zomba, exhibits the deformity in typical form well represented in the photograph (Plate V, (19)).

The head of each humerus appears to be poorly developed and though movement at the shoulder joint is free, a certain amount of fine crepitus is elicited, such as was found in several of the other cases.

The point of interest however is the fact that the maternal aunt is stated to have had the same congenital anomaly.

The subject has no brothers or sisters and his own two young children are stated to be normals, his mother and father and more remote relations are not known to be affected.

Besides these the following cases deserve mention.

A boy was seen, 18 years of age, with a peculiar deformation of the hands, stated to be congenital; the fingers and thumb shewed considerable thickening about the 1st interphalangeal joints with marked ulnar deflection; the bridge of the nose was depressed, the lips very thick, and epicanthus present. There was also the penile deformity above mentioned, except for which I should have doubted the statement in regard to the congenital nature of the hand deformity (Fig. 11).

Fig. 11. Boy aged 18.

I saw at Bandawe a female infant aged 1½ years presenting multiple deformities. The astragalus of the left foot was apparently implanted in a cup-shaped depression on the lower end of a very much shortened thigh. The femur of this leg was short but around it there was an abnormal amount of muscle as if the usual amount of muscle for a normal had been cramped up into the shortened limb; the foot could be freely moved by the child. The left foot had only a hallux and two toes with a partial cleft between the hallux and the adjacent toe, but I think four metatarsal bones. The right thigh was also somewhat shortened but the bones of the leg apparently both present, the knee-joint could not be distinctly made out and was flail. Right talipes equinovarus present, also right internal strabismus. No history of similar deformity in family, a brother a year older was born with two upper incisors. Father and mother normal. The father has two other wives with six and ten children respectively, all normal. Such gross congenital deformities are from time to time recorded in Europe, thus Lockart Mummery described a case of congenital absence of the femur in a male child, etc., in the *Brit. Med. Jour.* for November 5, 1910.

In a male 35 years of age I found *Congenital Absence of the Right Fibula,* the tibia being bowed forward with 8 inches shortening of the limb, the foot on the same side had only three metatarsal bones and three digits including the hallux. A woman was seen with congenital shortening of one leg to the extent of four inches.

A single case of unilateral *Congenital Dislocation of the Hip* has been met with.

(20) *Split Hand and Split Foot Deformities.* The photograph, Plate IV, (15), serves to shew moderately well the deformities met with in a male child aged 5 years (1905): in the absence of a skiagram it is impossible to go into the detail of the bony conditions present. There was no admitted history of similar or other deformity in the family.

A second case, Ndala of Njalusi's Mangoche, shewed a similar deformity of the left hand but in a less degree; he was otherwise normal and stated that no other members of his family were similarly affected (Plate VI, (20)).

These cases are interesting to compare with those collected and classified by Lewis and Embleton in *Biometrika,* Vol. VI, 1908.

(21) *Shortening of the Fourth Metatarsal Bone.* When first I entered the country my attention was attracted by a number of natives who presented a shortening of the fourth toe.

Since then Captain Hughes has noticed the condition in Egypt. The description he gives is as follows (*Lancet,* July 16, 1910):—"The fourth toe is markedly retracted usually behind the level of the fifth toe. The phalanges are not apparently abnormally short, and the metatarsal bone can be felt unfractured but with the head very much farther back than usual. Commonly the digit is pushed upwards by the pressure inwards of the fifth toe. The condition is sometimes unilateral sometimes bilateral." He adds that in one case the second metatarsal and, in another, the third metatarsal were also shortened. In a single case he saw a similar condition in the hand, shortening of the second and fifth metacarpals.

The above description corresponds exactly with the condition seen in this country. I have also seen other toes than the fourth affected, and I shew a photograph of a man's feet with involvement of the metatarsal of the hallux; in another case the fifth was affected; in another case, a woman, the common variety was associated with shortening of the third metatarsal of the left foot (see Plate VI, (22), (24) and Fig. 12).

Fig. 12.

Fig. 13.

(22) *Syndactyly* of various degrees has been observed; sketches of two examples are given in Fig. 13.

(23) *Polydactyly* is not at all uncommon. I have casually come across some dozen cases in five years.

In the majority the supernumerary digit consists of a miniature phalanx attached to the skin of the hand or foot at the level of the head of the fifth metacarpal or -tarsal bone. Such digits are often removed in childhood, leaving a small cartilaginous nodule at the seat of removal. Most commonly it is a symmetrical affection of both hands and feet; in other cases hands or feet alone (Plate VI, (23)), or one extremity only, present the deformity. In some the accessory digit is well formed and an accessory metatarsal or metacarpal bone more or less complete is present. In one case it was the hallux which was reduplicated, the two digits being partially fused. In another, reported to me, the supernumerary digit in each hand was situated on the radial side of the first finger with probably an accessory metacarpal bone in connection with it. The feet had extra digits beyond the fifth toes.

(24) The following case is of some interest:

Chibisa, male, an Angoni of Kawenga's, aged 30 years. The deformities involve all segments of the right upper and lower limbs and to a minor extent the left limbs (Plate VI, (21) and Figs. 15—18). On the right side there is shortening of the humerus and forearm (10 cm. difference between the two sides), but elongation of all the segments of the middle finger and its metacarpal bone; the middle finger itself measures 10½ cm. The metacarpal bones and phalanges of the other fingers are, I think, absolutely a little shortened. The left arm and hand are normal, except that this hand as also the right hand shew a little nodule at the base of the little finger where a supernumerary digit was removed.

Fig. 15.

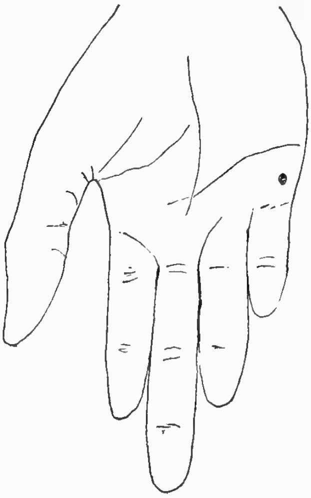


Fig. 16.

Fig. 17.

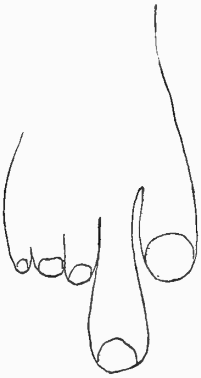


Fig. 18.

The right foot presents a similar condition to the right hand, elongation of phalanges and metatarsal affecting the second toe, the toe itself being 7 cm. long. The tibia is somewhat bowed outwards. The left foot presents shortening of the metatarsal bone of the hallux.

The photograph and sketches illustrate some of these points. Other measurements were as follows:

Height 166·5 cm.; span of arms 164·5 cm.;

Maximum fronto-occipital 18·0; maximum biparietal 13·8;

Nose length and width 4·4.

(25) *Congenital Anomalies of the Kidney.* Post-mortem examination on a native prisoner who died of pellagra revealed the presence of a double kidney on the left side and none on the right.

From the sketches (Fig. 19) it will be seen that the upper part was the one proper to the side while the lower half was the abnormal portion.

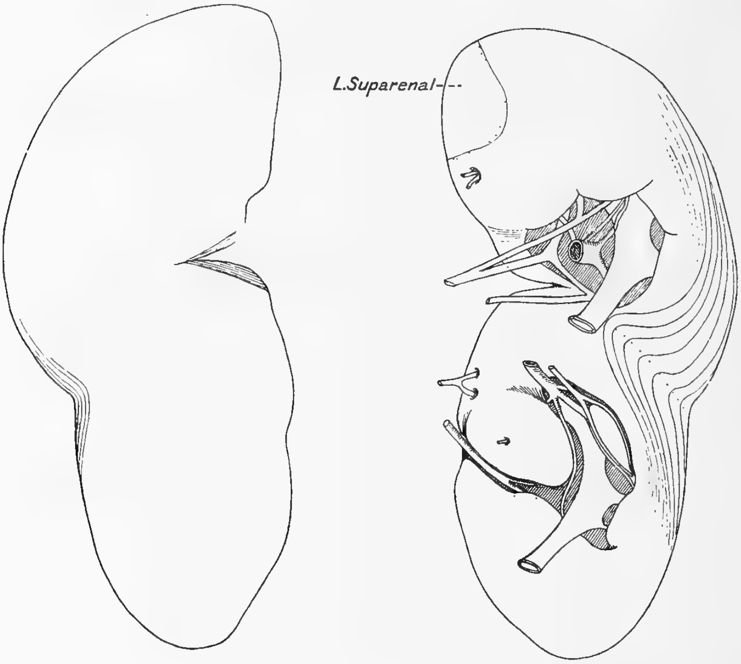


Fig. 19. Churinigu. Kidney of Left side double.

The two parts were really very distinct, partly separated by a groove and cleft.

The lower viscus had been felt during life as a tumour in the abdomen of unknown nature as it lay along the left side of the vertebral column. The kidney was unfortunately removed before dissection of vessels, etc. was made, but the sketch shews the arrangement of these at the hilum of the kidney.

The two ureters united below the lower pole of the double organ, the distal ureter being nearly twice the normal size.

The bladder was normal; there was no right ureter. The suprarenal body of the right side was in its normal position and appeared normal. No other abnormalities were remarked.

(26) Some suggestive observations have been made by Dr Ewald Stier, published in the *Deutsche Zeitschrift für Nervenheilkunde* (Band XLIV, Heft 1–2, S. 21), from which the generalisation is made that in all anomalies of overgrowth the right side of the body is much more frequently involved than the left, whereas in anomalies of undergrowth the left is more commonly the site of the condition than the right, this distribution being the result of a preponderance of persons with a leading or superior left cerebral hemisphere, as with left-handed persons the converse was found to be true. In other words, the plus anomalies occur on the right side in right-handed people and the minus anomalies on the left side, the left hemisphere being the superior hemisphere, the converse being true.

I have therefore tabulated my observations, and though small in number they tend to confirm the idea assuming that the African native is right-handed. This remains unproved and a less marked superiority of the left hemisphere may account for non-conformity of my few cases to Stier's rule.

	Right	Bilateral	Left
Plus Anomalies:			
Reduplication of teeth ...	2	3	0
Polymazia	1	1	0
Polythelia	10	0	5
Polydactyly	1	4	2
Minus Anomalies:			
Hare-lip	0	0	2
Cryptorchidism	2	0	2
Absence of Fibula	1	0	0
" " and Tibia...	0	0	1
Split hand, foot	0	1	1
Shortened metacarpal, -tarsal	0	1	3
Syndactyly	1	0	0
Coloboma iris	1	1	0
Plus and Minus together:			
Chibisa	1 (+)	—	1 (−)

In considering these cases it should be remembered that the majority of my observations have been made casually among natives met in the bush, in villages, etc., others in the course of routine work among troops, prisoners, etc., the few were the result of special investigation.

(27) *Concluding Remarks.* The notes of cases which I have thus collected together form rather a medley of facts but I think certain deductions may be made from them.

It would appear that

(1) The slighter the anomaly the greater the frequency with which it may be observed.

(2) The more marked degrees of deformity are only seen in children and those in places where European influence is felt.

(3) Cases of heredity are only seen among the lesser anomalies.

(4) The least obvious congenital anomaly is a helical fistula, and this is found in 4·6 °/₀ of the population and is frequently inherited.

The difference in the observed incidence between the minor anomalies and those of more marked proportions may be real or only apparent. I think the latter supposition is true for reasons which can be deduced from the facts given above.

It is the custom among all the tribes of this country to destroy all deformed children at birth. Any minimal deformity such as a helical fistula is of course unrecognised, an accessory nipple is probably hardly noticeable, accessory digits which can be removed by a nick with a knife are matters of no import, while a foot with six well-formed toes would hardly be considered worthy of note. These abnormalities are therefore comparatively common, but hare-lip, cleft-palate, deformities common enough in Europe, are among the rarest in this country; a child with a hare-lip would be seen to resemble a hare and would be immediately destroyed. Children with the greater deformities would certainly be destroyed. In recent years under European influence native customs fall into abeyance and so we see my single case of hare-lip in a boy aged 10 at Blantyre, a township of 25 years standing, a child with gross deformities of the lower extremities born practically on a mission station; or, to quote another example, an albino reported by myself was the fifth albino child born, the first four having been killed at birth by order of a chief, who in later years came under the influence of an up-country mission station, for which the living albino has to thank his survival. The gross abnormality of absence of premaxilla would pass unnoticed as the deformity is slight. History relates that in the case of the child with lobster claw deformity of hands and feet, it was only saved from a summary death by the efforts of the mother.

I think with the evidence as it stands one may with fairness say that congenital anomalies are common among the natives of this country. Secondly, I think one may also deduce from the facts stated that abnormalities of all kinds are at least not uncommon. In the few cases in which I have adduced statistics there can be no doubt, in other cases it is rather a matter of one's impression.

I have shewn that certain congenital anomalies among natives of Nyasaland are common and have attempted to argue that probably many of them are common.

(28) Very few statistics are available for comparison, but I should like to refer to some by writers in Egypt. Prof. Madden cites in a letter to the *Lancet* a case of cleft-palate which he operated on as the first in 11 years during surgical work at the Kasr-el-ainy Hospital, and assigns as the cause of the lack of such cases the "truly awful struggle for existence" which would eliminate infants so handicapped. The *Lancet* remarked (*Lancet*, July 3, 1909), in an annotation upon this letter, that Prof. Elliot Smith considers it to be impossible to endeavour to explain this rarity of congenital defects in Egypt, unless the time-honoured scapegoat of our too modern civilisation be invoked to account for their frequency in other countries. Statistics of the Kasr-el-ainy Hospital compiled by Dr Day are quoted in 1907; among 2630 total surgical admissions the only congenital deformities were 5 hare-lips, 2 talipes, 2 imperforate anus, 1 extroversion of bladder; in 1908, 2702 admissions, 3 hare-lips, 2 imperforate anus, 1 hypospadias, 1 undescended testicle, 1 meningo-encephalocoele. Capt. G. W. G. Hughes, R.A.M.C., in a paper to the *Lancet*, July 16, 1910, referring to this annotation, remarks "Readers will be interested to hear that our too modern civilisation is innocent of this slur," and goes on to shew that many congenital defects are by no means uncommon. Dealing with males between the ages of 14 and 21 years he gives the following figures:

Hare-lip in 0·041 %.
Cleft-palate 0·016 %.
Polydactylism 0·058 % and 0·04 %, in two series.
Shortened metatarsal 0·37 % and 0·23 %.
Other deformities of fingers and toes 0·22 %.
Talipes 0·16 %.

Among the thousands of ancient Egyptian bodies which Prof. G. Elliot Smith has unearthed and examined, a single case of cleft-palate was met with, a female of 20 years of age with a skull of negroid type, of between the 4th and 6th century B.C.; only one case of talipes (T. equinovarus) was recorded.

It is obvious that in Egypt surgical treatment is not sought in cases of cleft-palate and rarely for other congenital defects but many of them are common enough.

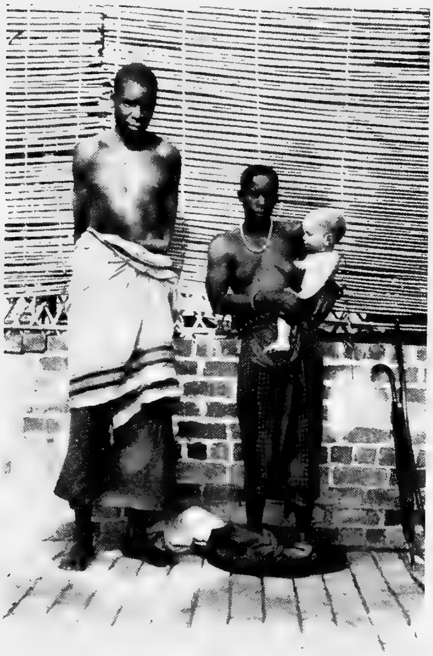
May the rarity of defects among the ancient peoples of Egypt be due to the same cause that acts in Nyasaland to-day? Were the children affected with deformities killed at birth and "thrown onto the dust-heap" where their remains were soon lost trace of? Of chief interest to me are the figures published by Captain Hughes. He shews that a shortening of the 4th metatarsal bone occurs in percentages rising to 0·37 of males examined. This defect is peculiarly common in this country. Again, polydactylism occurs in 0·05 % and other deformities of fingers and toes in 0·22 % of Egyptians, both deformities very frequently met with by myself in Nyasaland.

He however does not mention polythelia and polymazia nor helical fistula. I should be very interested to learn if this last insignificant anomaly was looked for. There is no doubt that one of them, helical fistula, occurs with a frequency in Nyasaland which cannot be rivalled by any other among peoples of any race. I think one may also say with certainty that the incidence of others (shortened 4th metatarsal and polydactylism) in this country is far in excess of that among Europeans, though probably much about the same as in Egypt.

Upon what hypothesis can these facts be explained? Is there a single cause or are there many at work? These are questions which I shall not attempt to enter into, but by simply recording my observations I shall hope to stimulate others to do the same, for only by accumulating facts can it be hoped that such problems will ever be solved.

(1)

Samuti, an Ateliotic Dwarf.



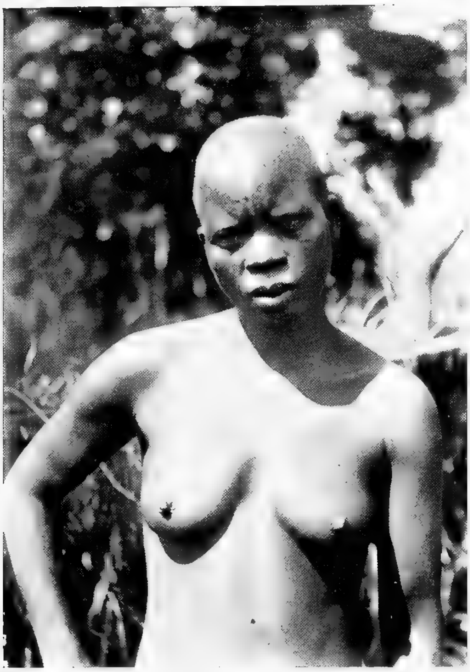
(2)

Subgiant, Height 1·92 metres, with Wife and Albinotic Child.

(3)

(4)

Etimu, aged 25, an Achondroplasic Dwarf.



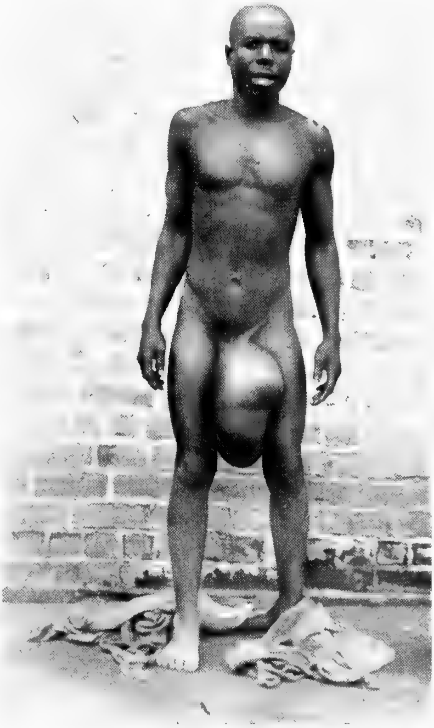
(5)

(6)

Masimosya, aged 19, Gynaecomastos, with other features which were formerly described as those of Partial Hermaphroditism.

(7)

Microcephalic Infant.



(8)

[Case of Hydrocele testis included by an oversight of Dr Stannus in the photographs, and engraved in consequence. Discovered too late to rearrange plates.]

(9)

(10)

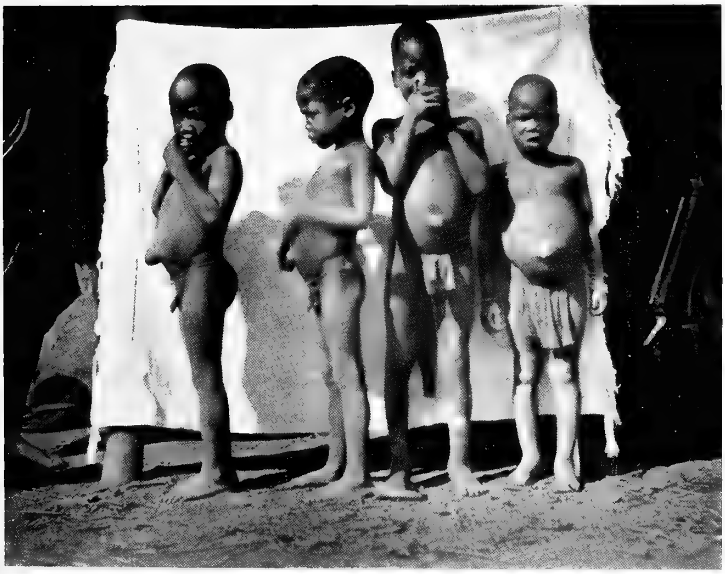
(11)

Boy, aged 15 years, showing Scaphocephaly.

(12)

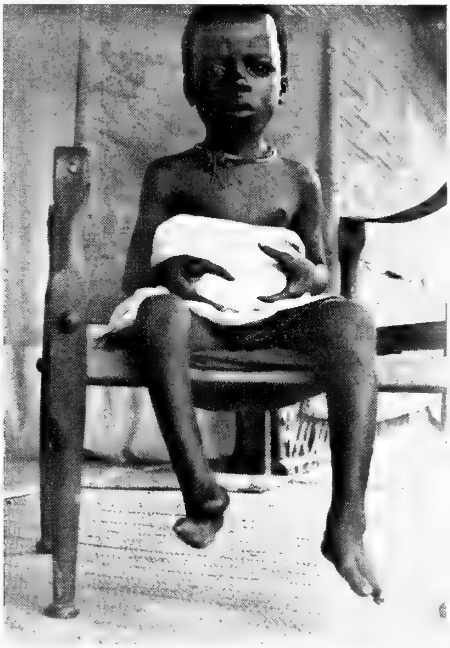
(13)

Son of Matikwiri, aged 7, a case of Scaphocephaly.



(14)

Cases of Umbilical Hernia.



(15)

Split Hand and Foot in a child, aged 5.

(16)

Case showing faint medium depression of upper lip.

(17)

Blantyre boy, aged 10, with Hare-lip.

(18)

Young woman with two nipples on left breast.

(19)

Gobedi, aged 22, with congenital Humeral Micromely.

(20) Ndala, Split Hand, left only.

(21) Chibisa, aged 30, elongation of all segments of middle finger, and its metacarpal bone.

(22) Shortening of the fourth metatarsal bone.

(23) Case of Polydactyly.

(24) Shortening of the left great toe.

TABLES OF POISSON'S EXPONENTIAL BINOMIAL LIMIT.

By H. E. SOPER, M.A.

In his treatise, *Recherches sur la Probabilité des Jugements*, Paris, 1837, Poisson[*] shows that the series of frequencies

$$p^n + np^{n-1}q + \frac{n(n-1)}{2!}p^{n-2}q^2 + \ldots + \frac{n!}{r!(n-r)!}p^{n-r}q^r + \ldots,$$

given by the expanded terms of the binomial

$$(p+q)^n,$$

becomes in the limit, when q is diminished, and n increased, indefinitely, but so that nq remains finite and equal to m, the exponential series

$$e^{-m}\left(1 + m + \frac{m^2}{2!} + \ldots + \frac{m^r}{r!} + \ldots\right);$$

and he points out that the terms of this series will give the proportional frequencies of the occurrences

$$0, \quad 1, \quad 2, \ldots r, \ldots$$

times, in any sample, of an event, every occurrence of which is equally likely in the sample and independent of the other occurrences, and which is of such frequency that m events occur in the sample on an average.

The series is arrived at by "Student[†]," when considering the theoretical frequencies in sample drops of a liquid of minute corpuscles supposed distributed at random throughout the mass of the liquid.

The event may also occur in time, each occurrence being supposed to take place with equal probability in any finite period taken as the sample, and to act independently of the occurrences of all the other events. A physical example, which appears by the closeness of the observed to the theoretical frequencies to

---

[*] pp. 205 *et seq.*

[†] *Biometrika*, Vol. v. p. 351, "On the Error of counting with a Haemacytometer."

satisfy these conditions, is the number of α-particles discharged per $\frac{1}{8}$-minute or $\frac{1}{4}$-minute interval from a film of polonium[*].

In vital statistics the sample may be an individual or house or community and the event an accident or disease and so on. But it must be borne in mind that for such series as the above to be applicable the occurrence of one event in the sample must not preclude or influence in any way the occurrence of a second.

The probability of x occurrences, m being the mean number, in a sample, is

$$e^{-m}\,m^x/x\,!$$

and in the tables which follow this is evaluated for $m = 0\cdot1,\ 0\cdot2 \ldots$ to $15\cdot0$ and for $x = 0, 1, 2 \ldots$ up to such an integer as gives a figure in the sixth place of decimals, the number of places tabulated.

The terms of the series were calculated, each by a fractional operation upon the preceding, beginning with the modal term and going both forward and back. Thus if $m = 7\cdot6$ the term $e^{-7\cdot6} \times (7\cdot6)^7/7\,!$ was first calculated by tables of logarithms, and the succeeding terms were then obtained seriatim by the operations

$$\frac{7\cdot6}{8},\ \frac{7\cdot6}{9},\ \frac{7\cdot6}{10},\ \text{etc.,}$$

and the preceding ones by the operations

$$\frac{7}{7\cdot6},\ \frac{6}{7\cdot6},\ \frac{5}{7\cdot6},\ \text{etc.,}$$

done with a mechanical calculator, first a multiplication and then a division.

Seven places of decimals were thus calculated and the series is checked by the total, which differs from unity by the remainder (a figure in the eighth or later place of decimals in all the present cases) and the algebraical sum of the errors of seventh figure approximations.

Poisson's exponential series has been previously calculated to four places of decimals by L. von Bortkewitsch[†] for values of m from 0·1 to 10·0.

The present tables give the probability of each number of times of occurrence of the event. For the sums of these values, that is, the probability of occurrence of the event, a given number of times or greater, or a given number of times or less, reference must be made to a second paper in this issue of *Biometrika*[‡], where such probabilities are calculated for integral values of m from 1 to 30.

---

[*] See Rutherford and Geiger: "The Probability Variations in the Distribution of α-Particles," *Philosophical Magazine*, Vol. xx. p. 700, 1910. See also E. C. Snow, "Note on the Probability Variations, &c.," Vol. xxii. p. 198, 1911, who finds the variance of experiment from theory to be such as would occur once in six experiments and once in three experiments respectively of the limited time taken, were theory exact. In a note to the first paper H. Bateman gives a proof of the exponential series of probabilities arrived at from considerations of this problem.

[†] *Das Gesetz der kleinen Zahlen*, 1898. A comparison of the table printed therein with the present table shows agreement except as to the fourth figure; the nearest fourth figure is not given, in rather many instances, in the tables of Bortkewitsch.

[‡] Lucy Whitaker, B.Sc. "On the Poisson Law of Small Numbers," Vol. x. p. 37 *et seq.*

TABLE of $e^{-m}m^x/x!$: General Term of Poisson's Exponential Expansion ("Law of Small Numbers").

x	*m*										*x*
	0·1	*0·2*	*0·3*	*0·4*	*0·5*	*0·6*	*0·7*	*0·8*	*0·9*	*1·0*	
0	·904837	·818731	·740818	·670320	·606531	·548812	·496585	·449329	·406570	·367879	*0*
1	·090484	·163746	·222245	·268128	·303265	·329287	·347610	·359463	·365913	·367879	*1*
2	·004524	·016375	·033337	·053626	·075816	·098786	·121663	·143785	·164661	·183940	*2*
3	·000151	·001092	·003334	·007150	·012636	·019757	·028388	·038343	·049398	·061313	*3*
4	·000004	·000055	·000250	·000715	·001580	·002964	·004968	·007669	·011115	·015328	*4*
5	—	·000002	·000015	·000057	·000158	·000356	·000696	·001227	·002001	·003066	*5*
6	—	—	·000001	·000004	·000013	·000036	·000081	·000164	·000300	·000511	*6*
7	—	—	—	—	·000001	·000003	·000008	·000019	·000039	·000073	*7*
8	—	—	—	—	—	—	·000001	·000002	·000004	·000009	*8*
9	—	—	—	—	—	—	—	—	—	·000001	*9*
x	*1·1*	*1·2*	*1·3*	*1·4*	*1·5*	*1·6*	*1·7*	*1·8*	*1·9*	*2·0*	*x*
0	·332871	·301194	·272532	·246597	·223130	.201897	·182684	·165299	·149569	·135335	*0*
1	·366158	·361433	·354291	·345236	·334695	.323034	·310562	·297538	·284180	·270671	*1*
2	·201387	·216860	·230289	·241665	·251021	.258428	·263978	·267784	·269971	·270671	*2*
3	·073842	·086744	·099792	·112777	·125510	·137828	·149587	·160671	·170982	·180447	*3*
4	·020307	·026023	·032432	·039472	·047067	·055131	·063575	·072302	·081216	·090224	*4*
5	·004467	·006246	·008432	·011052	·014120	·017642	·021615	·026029	·030862	·036089	*5*
6	·000819	·001249	·001827	·002579	·003530	·004705	·006124	·007809	·009773	·012030	*6*
7	·000129	·000214	·000339	·000516	·000756	·001075	·001487	·002008	·002653	·003437	*7*
8	·000018	·000032	·000055	·000090	·000142	·000215	·000316	·000452	·000630	·000859	*8*
9	·000002	·000004	·000008	·000014	·000024	·000038	·000060	·000090	·000133	·000191	*9*
10	—	·000001	·000001	·000002	·000004	·000006	·000010	·000016	·000025	·000038	*10*
11	—	—	—	—	—	·000001	·000002	·000003	·000004	·000007	*11*
12	—	—	—	—	—	—	—	—	·000001	·000001	*12*
x	*2·1*	*2·2*	*2·3*	*2·4*	*2·5*	*2·6*	*2·7*	*2·8*	*2·9*	*3·0*	*x*
0	·122456	·110803	·100259	·090718	·082085	·074274	·067206	·060810	·055023	·049787	*0*
1	·257159	·243767	·230595	·217723	·205212	·193111	·181455	·170268	·159567	·149361	*1*
2	·270016	·268144	·265185	·261268	·256516	·251045	·244964	·238375	·231373	·224042	*2*
3	·189012	·196639	·203308	·209014	·213763	·217572	·220468	·222484	·223660	·224042	*3*
4	·099231	·108151	·116902	·125409	·133602	·141422	·148816	·155739	·162154	·168031	*4*
5	·041677	·047587	·053775	·060196	·066801	·073539	·080360	·087214	·094049	·100819	*5*
6	·014587	·017448	·020614	·024078	·027834	·031867	·036162	·040700	·045457	·050409	*6*
7	·004376	·005484	·006773	·008255	·009941	·011836	·013948	·016280	·018832	·021604	*7*
8	·001149	·001508	·001947	·002477	·003106	·003847	·004708	·005698	·006827	·008102	*8*
9	·000268	·000369	·000498	·000660	·000863	·001111	·001412	·001773	·002200	·002701	*9*
10	·000056	·000081	·000114	·000158	·000216	·000289	·000381	·000496	·000638	·000810	*10*
11	·000011	·000016	·000024	·000035	·000049	·000068	·000094	·000126	·000168	·000221	*11*
12	·000002	·000003	·000005	·000007	·000010	·000015	·000021	·000029	·000041	·000055	*12*
13	—	·000001	·000001	·000001	·000002	·000003	·000004	·000006	·000009	·000013	*13*
14	—	—	—	—	—	·000001	·000001	·000001	·000002	·000003	*14*
15	—	—	—	—	—	—	—	—	—	·000001	*15*

TABLE—(*continued*).

x	*m*										*x*
	3·1	*3·2*	*3·3*	*3·4*	*3·5*	*3·6*	*3·7*	*3·8*	*3·9*	*4·0*	
0	·045049	·040762	·036883	·033373	·030197	·027324	·024724	·022371	·020242	·018316	*0*
1	·139653	·130439	·121714	·113469	·105691	·098365	·091477	·085009	·078943	·073263	*1*
2	·216461	·208702	·200829	·192898	·184959	·177058	·169233	·161517	·153940	·146525	*2*
3	·223677	·222616	·220912	·218617	·215785	·212469	·208720	·204588	·200122	·195367	*3*
4	·173350	·178093	·182252	·185825	·188812	·191222	·193066	·194359	·195119	·195367	*4*
5	·107477	·113979	·120286	·126361	·132169	·137680	·142869	·147713	·152193	·156293	*5*
6	·055530	·060789	·066158	·071604	·077098	·082608	·088102	·093551	·098925	·104196	*6*
7	·024592	·027789	·031189	·034779	·038549	·042484	·046568	·050785	·055115	·059540	*7*
8	·009529	·011116	·012865	·014781	·016865	·019118	·021538	·024123	·026869	·029770	*8*
9	·003282	·003952	·004717	·005584	·006559	·007647	·008854	·010185	·011643	·013231	*9*
10	·001018	·001265	·001557	·001899	·002296	·002753	·003276	·003870	·004541	·005292	*10*
11	·000287	·000368	·000467	·000587	·000730	·000901	·001102	·001337	·001610	·001925	*11*
12	·000074	·000098	·000128	·000166	·000213	·000270	·000340	·000423	·000523	·000642	*12*
13	·000018	·000024	·000033	·000043	·000057	·000075	·000097	·000124	·000157	·000197	*13*
14	·000004	·000006	·000008	·000011	·000014	·000019	·000026	·000034	·000044	·000056	*14*
15	·000001	·000001	·000002	·000002	·000003	·000005	·000006	·000009	·000011	·000015	*15*
16	—	—	—	·000001	·000001	·000001	·000001	·000002	·000003	·000004	*16*
17	—	—	—	—	—	—	—	—	·000001	·000001	*17*
x	*4·1*	*4·2*	*4·3*	*4·4*	*4·5*	*4·6*	*4·7*	*4·8*	*4·9*	*5·0*	*x*
0	·016573	·014996	·013569	·012277	·011109	·010052	·009095	·008230	·007447	·006738	*0*
1	·067948	·062981	·058345	·054020	·049990	·046238	·042748	·039503	·036488	·033690	*1*
2	·139293	·132261	·125441	·118845	·112479	·106348	·100457	·094807	·089396	·084224	*2*
3	·190368	·185165	·179799	·174305	·168718	·163068	·157383	·151691	·146014	·140374	*3*
4	·195127	·194424	·193284	·191736	·189808	·187528	·184925	·182029	·178867	·175467	*4*
5	·160004	·163316	·166224	·168728	·170827	·172525	·173830	·174748	·175290	·175467	*5*
6	·109336	·114321	·119127	·123734	·128120	·132270	·136167	·139798	·143153	·146223	*6*
7	·064040	·068593	·073178	·077775	·082363	·086920	·091426	·095862	·100207	·104445	*7*
8	·032820	·036011	·039333	·042776	·046329	·049979	·053713	·057517	·061377	·065278	*8*
9	·014951	·016805	·018793	·020913	·023165	·025545	·028050	·030676	·033416	·036266	*9*
10	·006130	·007058	·008081	·009202	·010424	·011751	·013184	·014724	·016374	·018133	*10*
11	·002285	·002695	·003159	·003681	·004264	·004914	·005633	·006425	·007294	·008242	*11*
12	·000781	·000943	·001132	·001350	·001599	·001884	·002206	·002570	·002978	·003434	*12*
13	·000246	·000305	·000374	·000457	·000554	·000667	·000798	·000949	·001123	·001321	*13*
14	·000072	·000091	·000115	·000144	·000178	·000219	·000268	·000325	·000393	·000472	*14*
15	·000020	·000026	·000033	·000042	·000053	·000067	·000084	·000104	·000128	·000157	*15*
16	·000005	·000007	·000009	·000012	·000015	·000019	·000025	·000031	·000039	·000049	*16*
17	·000001	·000902	·000002	·000003	·000004	·000005	·000007	·000009	·000011	·000014	*17*
18	—	—	·000001	·000001	·000001	·000001	·000002	·000002	·000003	·000004	*18*
19	—	—	—	—	—	—	—	·000001	·000001	·000001	*19*
x	*5·1*	*5·2*	*5·3*	*5·4*	*5·5*	*5·6*	*5·7*	*5·8*	*5·9*	*6·0*	*x*
0	·006097	·005517	·004992	·004517	·004087	·003698	·003346	·003028	·002739	·002479	*0*
1	·031093	·028686	·026455	·024390	·022477	·020708	·019072	·017560	·016163	·014873	*1*
2	·079288	·074584	·070107	·065852	·061812	·057982	·054355	·050923	·047680	·044618	*2*
3	·134790	·129279	·123856	·118533	·113323	·108234	·103275	·098452	·093771	·089235	*3*

TABLE—(*continued*).

x	*m*										*x*
	5·1	*5·2*	*5·3*	*5·4*	*5·5*	*5·6*	*5·7*	*5·8*	*5·9*	*6·0*	
4	·171857	·168063	·164109	·160020	·155819	·151528	·147167	·142755	·138312	·133853	*4*
5	·175294	·174785	·173955	·172821	·171401	·169711	·167770	·165596	·163208	·160623	*5*
6	·149000	·151480	·153660	·155539	·157117	·158397	·159382	·160076	·160488	·160623	*6*
7	·108557	·112528	·116343	·119987	·123449	·126717	·129782	·132635	·135268	·137677	*7*
8	·069205	·073143	·077077	·080991	·084871	·088702	·092470	·096160	·099760	·103258	*8*
9	·039216	·042261	·045390	·048595	·051866	·055192	·058564	·061970	·065398	·068838	*9*
10	·020000	·021976	·024057	·026241	·028526	·030908	·033382	·035943	·038585	·041303	*10*
11	·009273	·010388	·011591	·012882	·014263	·015735	·017298	·018952	·020696	·022529	*11*
12	·003941	·004502	·005119	·005797	·006537	·007343	·008216	·009160	·010175	·011264	*12*
13	·001546	·001801	·002087	·002408	·002766	·003163	·003603	·004087	·004618	·005199	*13*
14	·000563	·000669	·000790	·000929	·001087	·001265	·001467	·001693	·001946	·002228	*14*
15	·000191	·000232	·000279	·000334	·000398	·000472	·000557	·000655	·000766	·000891	*15*
16	·000061	·000075	·000092	·000113	·000137	·000165	·000199	·000237	·000282	·000334	*16*
17	·000018	·000023	·000029	·000036	·000044	·000054	·000067	·000081	·000098	·000118	*17*
18	·000005	·000007	·000008	·000011	·000014	·000017	·000021	·000026	·000032	·000039	*18*
19	·000001	·000002	·000002	·000003	·000004	·000005	·000006	·000008	·000010	·000012	*19*
20	—	—	·000001	·000001	·000001	·000001	·000002	·000002	·000003	·000004	*20*
21	—	—	—	—	—	—	—	·000001	·000001	·000001	*21*
x	*6·1*	*6·2*	*6·3*	*6·4*	*6·5*	*6·6*	*6·7*	*6·8*	*6·9*	*7·0*	*x*
0	·002243	·002029	·001836	·001662	·001503	·001360	·001231	·001114	·001008	·000912	*0*
1	·013682	·012582	·011569	·010634	·009772	·008978	·008247	·007574	·006954	·006383	*1*
2	·041729	·039006	·036441	·034029	·031760	·029629	·027628	·025751	·023990	·022341	*2*
3	·084848	·080612	·076527	·072595	·068814	·065183	·061702	·058368	·055178	·052129	*3*
4	·129393	·124948	·120530	·116151	·111822	·107553	·103351	·099225	·095182	·091226	*4*
5	·157860	·154936	·151868	·148674	·145369	·141969	·138490	·134946	·131351	·127717	*5*
6	·160491	·160100	·159461	·158585	·157483	·156166	·154648	·152939	·151053	·149003	*6*
7	·139856	·141803	·143515	·144992	·146234	·147243	·148020	·148569	·148895	·149003	*7*
8	·106640	·109897	·113018	·115994	·118815	·121475	·123967	·126284	·128422	·130377	*8*
9	·072278	·075707	·079113	·082484	·085811	·089082	·092286	·095415	·098457	·101405	*9*
10	·044090	·046938	·049841	·052790	·055777	·058794	·061832	·064882	·067935	·070983	*10*
11	·024450	·026456	·028545	·030714	·032959	·035276	·037661	·040109	·042614	·045171	*11*
12	·012429	·013669	·014986	·016381	·017853	·019402	·021028	·022728	·024503	·026350	*12*
13	·005832	·006519	·007263	·008064	·008926	·009850	·010837	·011889	·013005	·014188	*13*
14	·002541	·002887	·003268	·003687	·004144	·004644	·005186	·005774	·006410	·007094	*14*
15	·001033	·001193	·001373	·001573	·001796	·002043	·002317	·002618	·002949	·003311	*15*
16	·000394	·000462	·000540	·000629	·000730	·000843	·000970	·001113	·001272	·001448	*16*
17	·000141	·000169	·000200	·000237	·000279	·000327	·000382	·000445	·000516	·000596	*17*
18	·000048	·000058	·000070	·000084	·000101	·000120	·000142	·000168	·000198	·000232	*18*
19	·000015	·000019	·000023	·000028	·000034	·000042	·000050	·000060	·000072	·000085	*19*
20	·000005	·000006	·000007	·000009	·000011	·000014	·000017	·000020	·000025	·000030	*20*
21	·000001	·000002	·000002	·000003	·000003	·000004	·000005	·000007	·000008	·000010	*21*
22	—	—	·000001	·000001	·000001	·000001	·000002	·000002	·000003	·000003	*22*
23	—	—	—	—	—	—	—	·000001	·000001	·000001	*23*

TABLE—(*continued*).

x	*m*										*x*
	7·1	*7·2*	*7·3*	*7·4*	*7·5*	*7·6*	*7·7*	*7·8*	*7·9*	*8·0*	
0	·000825	·000747	·000676	·000611	·000553	·000500	·000453	·000410	·000371	·000335	*0*
1	·005858	·005375	·004931	·004523	·004148	·003803	·003487	·003196	·002929	·002684	*1*
2	·020797	·019352	·018000	·016736	·015555	·014453	·013424	·012464	·011569	·010735	*2*
3	·049219	·046444	·043799	·041282	·038889	·036614	·034455	·032407	·030465	·028626	*3*
4	·087364	·083598	·079934	·076372	·072916	·069567	·066326	·063193	·060169	·057252	*4*
5	·124057	·120382	·116703	·113031	·109375	·105742	·102142	·098581	·095067	·091604	*5*
6	·146800	·144458	·141989	·139405	·136718	·133940	·131082	·128156	·125171	·122138	*6*
7	·148897	·148586	·148074	·147371	·146484	·145421	·144191	·142802	·141264	·139587	*7*
8	·132146	·133727	·135118	·136318	·137329	·138150	·138783	·139232	·139499	·139587	*8*
9	·104249	·106982	·109596	·112084	·114440	·116660	·118737	·120668	·122449	·124077	*9*
10	·074017	·077027	·080005	·082942	·085830	·088661	·091427	·094121	·096735	·099262	*10*
11	·047774	·050418	·053094	·055797	·058521	·061257	·063999	·066740	·069473	·072190	*11*
12	·028267	·030251	·032299	·034408	·036575	·038796	·041066	·043381	·045736	·048127	*12*
13	·015438	·016754	·018137	·019586	·021101	·022681	·024324	·026029	·027794	·029616	*13*
14	·007829	·008616	·009457	·010353	·011304	·012312	·013378	·014502	·015684	·016924	*14*
15	·003706	·004136	·004603	·005107	·005652	·006238	·006867	·007541	·008260	·009026	*15*
16	·001644	·001861	·002100	·002362	·002649	·002963	·003305	·003676	·004078	·004513	*16*
17	·000687	·000788	·000902	·001028	·001169	·001325	·001497	·001687	·001895	·002124	*17*
18	·000271	·000315	·000366	·000423	·000487	·000559	·000640	·000731	·000832	·000944	*18*
19	·000101	·000119	·000141	·000165	·000192	·000224	·000259	·000300	·000346	·000397	*19*
20	·000036	·000043	·000051	·000061	·000072	·000085	·000100	·000117	·000137	·000159	*20*
21	·000012	·000015	·000018	·000021	·000026	·000031	·000037	·000043	·000051	·000061	*21*
22	·000004	·000005	·000006	·000007	·000009	·000011	·000013	·000015	·000018	·000022	*22*
23	·000001	·000002	·000002	·000002	·000003	·000004	·000004	·000005	·000006	·000008	*23*
24	—	—	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000003	*24*
25	—	—	—	—	—	—	—	·000001	·000001	·000001	*25*

x	*8·1*	*8·2*	*8·3*	*8·4*	*8·5*	*8·6*	*8·7*	*8·8*	*8·9*	*9·0*	*x*
0	·000304	·000275	·000249	·000225	·000203	·000184	·000167	·000151	·000136	·000123	*0*
1	·002459	·002252	·002063	·001889	·001729	·001583	·001449	·001326	·001214	·001111	*1*
2	·009958	·009234	·008560	·007933	·007350	·006808	·006304	·005836	·005402	·004998	*2*
3	·026885	·025239	·023683	·022213	·020826	·019517	·018283	·017120	·016025	·014994	*3*
4	·054443	·051740	·049142	·046648	·044255	·041961	·039765	·037664	·035656	·033737	*4*
5	·088198	·084854	·081576	·078368	·075233	·072174	·069192	·066289	·063467	·060727	*5*
6	·119067	·115967	·112847	·109716	·106581	·103449	·100328	·097224	·094143	·091090	*6*
7	·137778	·135848	·133805	·131659	·129419	·127094	·124693	·122224	·119696	·117116	*7*
8	·139500	·139244	·138823	·138242	·137508	·136626	·135604	·134446	·133161	·131756	*8*
9	·125550	·126866	·128025	·129026	·129869	·130554	·131084	·131459	·131682	·131756	*9*
10	·101696	·104031	·106261	·108382	·110388	·112277	·114043	·115684	·117197	·118580	*10*
11	·074885	·077550	·080179	·082764	·085300	·087780	·090197	·092547	·094823	·097020	*11*
12	·050547	·052993	·055457	·057935	·060421	·062909	·065393	·067868	·070327	·072765	*12*
13	·031495	·033426	·035407	·037435	·039506	·041617	·043763	·045941	·048147	·050376	*13*
14	·018222	·019578	·020991	·022461	·023986	·025565	·027196	·028877	·030608	·032384	*14*
15	·009840	·010703	·011615	·012578	·013592	·014657	·015773	·016941	·018161	·019431	*15*
16	·004981	·005485	·006025	·006604	·007221	·007878	·008577	·009318	·010102	·010930	*16*
17	·002373	·002646	·002942	·003263	·003610	·003985	·004389	·004823	·005289	·005786	*17*
18	·001068	·001205	·001356	·001523	·001705	·001904	·002121	·002358	·002615	·002893	*18*
19	·000455	·000520	·000593	·000673	·000763	·000862	·000971	·001092	·001225	·001370	*19*
20	·000184	·000213	·000246	·000283	·000324	·000371	·000423	·000481	·000545	·000617	*20*

TABLE—(*continued*).

x	m										x
	8·1	*8·2*	*8·3*	*8·4*	*8·5*	*8·6*	*8·7*	*8·8*	*8·9*	*9·0*	
21	·000071	·000083	·000097	·000113	·000131	·000152	·000175	·000201	·000231	·000264	*21*
22	·000026	·000031	·000037	·000043	·000051	·000059	·000069	·000081	·000093	·000108	*22*
23	·000009	·000011	·000013	·000016	·000019	·000022	·000026	·000031	·000036	·000042	*23*
24	·000003	·000004	·000005	·000006	·000007	·000008	·000009	·000011	·000013	·000016	*24*
25	·000001	·000001	·000002	·000002	·000002	·000003	·000003	·000004	·000005	·000006	*25*
26	—	—	—	·000001	·000001	·000001	·000001	·000001	·000002	·000002	*26*
27	—	—	—	—	—	—	—	—	·000001	·000001	*27*
x	*9·1*	*9·2*	*9·3*	*9·4*	*9·5*	*9·6*	*9·7*	*9·8*	*9·9*	*10·0*	*x*
0	·000112	·000101	·000091	·000083	·000075	·000068	·000061	·000055	·000050	·000045	*0*
1	·001016	·000930	·000850	·000778	·000711	·000650	·000594	·000543	·000497	·000454	*1*
2	·004624	·004276	·003954	·003655	·003378	·003121	·002883	·002663	·002459	·002270	*2*
3	·014025	·013113	·012256	·011452	·010696	·009987	·009322	·008698	·008114	·007567	*3*
4	·031906	·030160	·028496	·026911	·025403	·023969	·022606	·021311	·020082	·018917	*4*
5	·058069	·055494	·053002	·050593	·048266	·046020	·043855	·041770	·039763	·037833	*5*
6	·088072	·085091	·082154	·079262	·076421	·073632	·070899	·068224	·065609	·063055	*6*
7	·114493	·111834	·109147	·106438	·103714	·100981	·098246	·095514	·092790	·090079	*7*
8	·130236	·128609	·126883	·125065	·123160	·121178	·119123	·117004	·114827	·112599	*8*
9	·131683	·131467	·131113	·130623	·130003	·129256	·128388	·127405	·126310	·125110	*9*
10	·119832	·120950	·121935	·122786	·123502	·124086	·124537	·124857	·125047	·125110	*10*
11	·099133	·101158	·103090	·104926	·106661	·108293	·109819	·111236	·112542	·113736	*11*
12	·075176	·077555	·079895	·082192	·084440	·086634	·088770	·090843	·092847	·094780	*12*
13	·052623	·054885	·057156	·059431	·061706	·063976	·066236	·068481	·070707	·072908	*13*
14	·034205	·036067	·037968	·039904	·041872	·043869	·045892	·047937	·050000	·052077	*14*
15	·020751	·022121	·023540	·025006	·026519	·028076	·029677	·031319	·033000	·034718	*15*
16	·011802	·012720	·013683	·014691	·015746	·016846	·017992	·019183	·020419	·021699	*16*
17	·006318	·006884	·007485	·008123	·008799	·009513	·010266	·011058	·011891	·012764	*17*
18	·003194	·003518	·003867	·004242	·004644	·005074	·005532	·006021	·006540	·007091	*18*
19	·001530	·001704	·001893	·002099	·002322	·002563	·002824	·003105	·003408	·003732	*19*
20	·000696	·000784	·000880	·000986	·001103	·001230	·001370	·001522	·001687	·001866	*20*
21	·000302	·000343	·000390	·000442	·000499	·000563	·000633	·000710	·000795	·000889	*21*
22	·000125	·000144	·000165	·000189	·000215	·000245	·000279	·000316	·000358	·000404	*22*
23	·000049	·000057	·000067	·000077	·000089	·000102	·000118	·000135	·000154	·000176	*23*
24	·000019	·000022	·000026	·000030	·000035	·000041	·000048	·000055	·000064	·000073	*24*
25	·000007	·000008	·000010	·000011	·000013	·000016	·000018	·000022	·000025	·000029	*25*
26	·000002	·000003	·000003	·000004	·000005	·000006	·000007	·000008	·000010	·000011	*26*
27	·000001	·000001	·000001	·000001	·000002	·000002	·000002	·000003	·000004	·000004	*27*
28	—	—	—	—	·000001	·000001	·000001	·000001	·000001	·000001	*28*
29	—	—	—	—	—	—	—	—	—	·000001	*29*
x	*10·1*	*10·2*	*10·3*	*10·4*	*10·5*	*10·6*	*10·7*	*10·8*	*10·9*	*11·0*	*x*
0	·000041	·000037	·000034	·000030	·000028	·000 25	·000023	·000020	·000018	·000017	*0*
1	·000415	·000379	·000346	·000317	·000289	·000264	·000241	·000220	·000201	·000184	*1*
2	·002095	·001934	·001784	·001646	·001518	·001400	·001291	·001190	·001097	·001010	*2*
3	·007054	·006574	·006125	·005705	·005313	·004946	·004603	·004283	·003984	·003705	*3*

TABLE—(*continued*).

x	*m*										*x*
	10·1	*10·2*	*10·3*	*10·4*	*10·5*	*10·6*	*10·7*	*10·8*	*10·9*	*11·0*	
4	·017811	·016764	·015773	·014834	·013946	·013107	·012313	·011564	·010856	·010189	*4*
5	·035979	·034199	·032492	·030855	·029287	·027786	·026350	·024978	·023667	·022415	*5*
6	·060565	·058139	·055777	·053482	·051252	·049089	·046991	·044960	·042995	·041095	*6*
7	·087387	·084716	·082072	·079458	·076878	·074334	·071830	·069367	·066949	·064577	*7*
8	·110326	·108013	·105668	·103296	·100902	·098493	·096072	·093646	·091218	·088794	*8*
9	·123810	·122415	·120931	·119364	·117720	·116003	·114219	·112375	·110475	·108526	*9*
10	·125048	·124863	·124559	·124139	·123606	·122963	·122215	·121365	·120418	·119378	*10*
11	·114817	·115782	·116633	·117368	·117987	·118492	·118882	·119159	·119323	·119378	*11*
12	·096637	·098415	·100110	·101719	·103239	·104667	·106003	·107243	·108386	·109430	*12*
13	·075080	·077218	·079318	·081375	·083385	·085344	·087248	·089094	·090877	·092595	*13*
14	·054165	·056259	·058355	·060450	·062539	·064618	·066683	·068730	·070754	·072753	*14*
15	·036471	·038256	·040071	·041912	·043777	·045663	·047567	·049485	·051415	·053352	*15*
16	·023022	·024388	·025795	·027243	·028729	·030252	·031810	·033403	·035026	·036680	*16*
17	·013678	·014633	·015629	·016666	·017744	·018863	·020022	·021220	·022458	·023734	*17*
18	·007675	·008292	·008943	·009629	·010351	·011108	·011902	·012732	·013600	·014504	*18*
19	·004080	·004451	·004848	·005271	·005720	·006197	·006703	·007237	·007802	·008397	*19*
20	·002060	·002270	·002497	·002741	·003003	·003285	·003586	·003908	·004252	·004618	*20*
21	·000991	·001103	·001225	·001357	·001502	·001658	·001827	·002010	·002207	·002419	*21*
22	·000455	·000511	·000573	·000642	·000717	·000799	·000889	·000987	·001093	·001210	*22*
23	·000200	·000227	·000257	·000290	·000327	·000368	·000413	·000463	·000518	·000578	*23*
24	·000084	·000096	·000110	·000126	·000143	·000163	·000184	·000208	·000235	·000265	*24*
25	·000034	·000039	·000045	·000052	·000060	·000069	·000079	·000090	·000103	·000117	*25*
26	·000013	·000015	·000018	·000021	·000024	·000028	·000032	·000037	·000043	·000049	*26*
27	·000005	·000006	·000007	·000008	·000009	·000011	·000013	·000015	·000017	·000020	*27*
28	·000002	·000002	·000003	·000003	·000004	·000004	·000005	·000006	·000007	·000008	*28*
29	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000002	·000003	·000003	*29*
30	—	—	—	—	—	·000001	·000001	·000001	·000001	·000001	*30*
x	*11·1*	*11·2*	*11·3*	*11·4*	*11·5*	*11·6*	*11·7*	*11·8*	*11·9*	*12·0*	*x*
0	·000015	·000014	·000012	·000011	·000010	·000009	·000008	·000008	·000007	·000006	*0*
1	·000168	·000153	·000140	·000128	·000116	·000106	·000097	·000089	·000081	·000074	*1*
2	·000931	·000858	·000790	·000727	·000670	·000617	·000568	·000522	·000481	·000442	*2*
3	·003445	·003202	·002976	·002764	·002568	·002385	·002214	·002055	·001907	·001770	*3*
4	·009559	·008965	·008406	·007879	·007382	·006915	·006476	·006062	·005674	·005309	*4*
5	·021221	·020082	·018997	·017963	·016979	·016043	·015153	·014307	·013504	·012741	*5*
6	·039259	·037487	·035778	·034130	·032544	·031017	·029549	·028137	·026782	·025481	*6*
7	·062253	·059979	·057755	·055584	·053465	·051400	·049388	·047432	·045530	·043682	*7*
8	·086376	·083970	·081579	·079206	·076856	·074529	·072231	·069962	·067725	·065523	*8*
9	·106531	·104496	·102427	·100328	·098204	·096060	·093900	·091728	·089548	·087364	*9*
10	·118249	·117036	·115743	·114374	·112935	·111430	·109863	·108239	·106562	·104837	*10*
11	·119324	·119164	·118899	·118533	·118068	·117508	·116854	·116110	·115281	·114368	*11*
12	·110375	·111220	·111964	·112607	·113149	·113591	·113933	·114175	·114320	·114363	*12*
13	·094243	·095820	·097322	·098747	·100093	·101358	·102539	·103636	·104647	·105570	*13*
14	·074721	·076656	·078553	·080409	·082219	·083982	·085694	·087350	·088950	·090489	*14*
15	·055294	·057236	·059177	·061110	·063035	·064946	·066841	·068716	·070567	·072391	*15*
16	·038360	·040065	·041793	·043541	·045306	·047086	·048877	·050678	·052484	·054293	*16*
17	·025047	·026396	·027780	·029198	·030648	·032129	·033639	·035176	·036739	·038325	*17*
18	·015446	·016424	·017440	·018492	·019581	·020706	·021865	·023060	·024288	·025550	*18*
19	·009023	·009682	·010372	·011095	·011852	·012641	·013465	·014322	·015212	·016137	*19*

TABLE—(*continued*).

x	m										x
	11·1	*11·2*	*11·3*	*11·4*	*11·5*	*11·6*	*11·7*	*11·8*	*11·9*	*12·0*	
20	·005008	·005422	·005860	·006324	·006815	·007332	·007877	·008450	·009051	·009682	*20*
21	·002647	·002892	·003153	·003433	·003732	·004050	·004388	·004748	·005129	·005533	*21*
22	·001336	·001472	·001620	·001779	·001951	·002136	·002334	·002547	·002774	·003018	*22*
23	·000645	·000717	·000796	·000882	·000975	·001077	·001187	·001307	·001435	·001575	*23*
24	·000298	·000335	·000375	·000419	·000467	·000521	·000579	·000642	·000712	·000787	*24*
25	·000132	·000150	·000169	·000191	·000215	·000242	·000271	·000303	·000339	·000378	*25*
26	·000057	·000065	·000074	·000084	·000095	·000108	·000122	·000138	·000155	·000174	*26*
27	·000023	·000027	·000031	·000035	·000041	·000046	·000053	·000060	·000068	·000078	*27*
28	·000009	·000011	·000012	·000014	·000017	·000019	·000022	·000025	·000029	·000033	*28*
29	·000004	·000004	·000005	·000006	·000007	·000008	·000009	·000010	·000012	·000014	*29*
30	·000001	·000002	·000002	·000002	·000003	·000003	·000003	·000004	·000005	·000005	*30*
31	—	·000001	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000002	*31*
32	—	—	—	—	—	—	—	·000001	·000001	·000001	*32*

x	*12·1*	*12·2*	*12·3*	*12·4*	*12·5*	*12·6*	*12·7*	*12·8*	*12·9*	*13·0*	x
0	·000006	·000005	·000005	·000004	·000004	·000003	·000003	·000003	·000002	·000002	*0*
1	·000067	·000061	·000056	·000051	·000047	·000042	·000039	·000035	·000032	·000029	*1*
2	·000407	·000374	·000344	·000317	·000291	·000268	·000246	·000226	·000208	·000191	*2*
3	·001641	·001522	·001412	·001309	·001213	·001124	·001042	·000965	·000894	·000828	*3*
4	·004966	·004643	·004341	·004057	·003791	·003541	·003307	·003088	·002882	·002690	*4*
5	·012017	·011330	·010679	·010062	·009477	·008924	·008400	·007905	·007436	·006994	*5*
6	·024233	·023037	·021892	·020794	·019744	·018740	·017781	·016864	·015988	·015153	*6*
7	·041889	·040151	·038467	·036836	·035258	·033733	·032259	·030837	·029464	·028141	*7*
8	·063358	·061230	·059142	·057095	·055091	·053129	·051212	·049339	·047511	·045730	*8*
9	·085181	·083000	·080828	·078665	·076515	·074381	·072266	·070171	·068100	·066054	*9*
10	·103069	·101261	·099418	·097544	·095644	·093720	·091777	·089819	·087849	·085870	*10*
11	·113376	·112308	·111168	·109959	·108686	·107352	·105961	·104516	·103023	·101483	*11*
12	·114321	·114180	·113947	·113624	·113215	·112720	·112142	·111484	·110749	·109940	*12*
13	·106406	·107153	·107811	·108380	·108860	·109251	·109554	·109769	·109897	·109940	*13*
14	·091965	·093376	·094720	·095994	·097197	·098326	·099381	·100360	·101263	·102087	*14*
15	·074185	·075946	·077670	·079355	·080997	·082594	·084143	·085641	·087086	·088475	*15*
16	·056103	·057909	·059709	·061500	·063279	·065043	·066788	·068513	·070213	·071886	*16*
17	·039932	·041558	·043201	·044859	·046529	·048208	·049895	·051586	·053279	·054972	*17*
18	·026843	·028167	·029521	·030903	·032312	·033746	·035204	·036683	·038183	·039702	*18*
19	·017095	·018086	·019111	·020168	·021258	·022379	·023531	·024713	·025925	·027164	*19*
20	·010342	·011033	·011753	·012504	·013286	·014099	·014942	·015816	·016721	·017657	*20*
21	·005959	·006409	·006884	·007383	·007908	·008459	·009036	·009640	·010272	·010930	*21*
22	·003278	·003554	·003849	·004162	·004493	·004845	·005216	·005609	·006023	·006459	*22*
23	·001724	·001885	·002058	·002244	·002442	·002654	·002880	·003122	·003378	·003651	*23*
24	·000869	·000958	·001055	·001159	·001272	·001393	·001524	·001665	·001816	·001977	*24*
25	·000421	·000468	·000519	·000575	·000636	·000702	·000774	·000852	·000937	·001028	*25*
26	·000196	·000219	·000246	·000274	·000306	·000340	·000378	·000420	·000465	·000514	*26*
27	·000088	·000099	·000112	·000126	·000142	·000159	·000178	·000199	·000222	·000248	*27*
28	·000038	·000043	·000049	·000056	·000063	·000071	·000081	·000091	·000102	·000115	*28*
29	·000016	·000018	·000021	·000024	·000027	·000031	·000035	·000040	·000046	·000052	*29*
30	·000006	·000007	·000009	·000010	·000011	·000013	·000015	·000017	·000020	·000022	*30*
31	·000002	·000003	·000003	·000004	·000005	·000005	·000006	·000007	·000008	·000009	*31*
32	·000001	·000001	·000001	·000002	·000002	·000002	·000002	·000003	·000003	·000004	*32*
33	—	—	—	·000001	·000001	·000001	·000001	·000001	·000001	·000002	*33*
34	—	—	—	—	—	—	—	—	—	·000001	*34*

TABLE—(*continued*).

x	*m*										*x*
	13·1	*13·2*	*13·3*	*13·4*	*13·5*	*13·6*	*13·7*	*13·8*	*13·9*	*14·0*	
0	·000002	·000002	·000002	·000002	·000001	·000001	·000001	·000001	·000001	·000001	*0*
1	·000027	·000024	·000022	·000020	·000019	·000017	·000015	·000014	·000013	·000012	*1*
2	·000175	·000161	·000148	·000136	·000125	·000115	·000105	·000097	·000089	·000081	*2*
3	·000766	·000709	·000657	·000608	·000562	·000520	·000481	·000445	·000411	·000380	*3*
4	·002510	·002341	·002183	·002035	·001897	·001768	·001648	·001535	·001429	·001331	*4*
5	·006575	·006180	·005807	·005455	·005123	·004810	·004514	·004236	·003974	·003727	*5*
6	·014356	·013596	·012872	·012183	·011526	·010902	·010308	·009743	·009206	·008696	*6*
7	·026867	·025639	·024458	·023322	·022230	·021181	·020173	·019207	·018280	·017392	*7*
8	·043994	·042304	·040661	·039064	·037512	·036007	·034547	·033132	·031762	·030435	*8*
9	·064036	·062046	·060088	·058161	·056269	·054410	·052588	·050802	·049054	·047344	*9*
10	·083887	·081901	·079916	·077936	·075963	·073998	·072046	·070107	·068185	·066282	*10*
11	·099901	·098281	·096626	·094940	·093227	·091489	·089730	·087953	·086162	·084359	*11*
12	·109059	·108109	·107094	·106017	·104880	·103687	·102441	·101146	·099804	·098418	*12*
13	·109898	·109773	·109566	·109279	·108914	·108473	·107957	·107370	·106713	·105989	*13*
14	·102833	·103500	·104087	·104595	·105024	·105373	·105644	·105836	·105951	·105989	*14*
15	·089807	·091080	·092291	·093439	·094522	·095539	·096488	·097369	·098181	·098923	*15*
16	·073530	·075141	·076717	·078255	·079753	·081208	·082618	·083981	·085295	·086558	*16*
17	·056661	·058345	·060019	·061683	·063333	·064966	·066580	·068173	·069741	·071283	*17*
18	·041237	·042786	·044348	·045920	·047500	·049086	·050675	·052266	·053856	·055442	*18*
19	·028432	·029725	·031043	·032385	·033750	·035135	·036539	·037962	·039400	·040852	*19*
20	·018623	·019619	·020644	·021698	·022781	·023892	·025030	·026193	·027383	·028597	*20*
21	·011617	·012332	·013074	·013846	·014645	·015473	·016329	·017213	·018125	·019064	*21*
22	·006917	·007399	·007904	·008433	·008987	·009565	·010168	·010797	·011452	·012132	*22*
23	·003940	·004246	·004571	·004913	·005275	·005656	·006057	·006478	·006921	·007385	*23*
24	·002151	·002336	·002533	·002743	·002967	·003205	·003457	·003725	·004008	·004308	*24*
25	·001127	·001233	·001348	·001470	·001602	·001744	·001895	·002056	·002229	·002412	*25*
26	·000568	·000626	·000689	·000758	·000832	·000912	·000998	·001091	·001191	·001299	*26*
27	·000275	·000306	·000340	·000376	·000416	·000459	·000507	·000558	·000613	·000674	*27*
28	·000129	·000144	·000161	·000180	·000201	·000223	·000248	·000275	·000305	·000337	*28*
29	·000058	·000066	·000074	·000083	·000093	·000105	·000117	·000131	·000146	·000163	*29*
30	·000025	·000029	·000033	·000037	·000042	·000047	·000053	·000060	·000068	·000076	*30*
31	·000011	·000012	·000014	·000016	·000018	·000021	·000024	·000027	·000030	·000034	*31*
32	·000004	·000005	·000006	·000007	·000008	·000009	·000010	·000012	·000013	·000015	*32*
33	·000002	·000002	·000002	·000003	·000003	·000004	·000004	·000005	·000006	·000006	*33*
34	·000001	·000001	·000001	·000001	·000001	·000001	·000002	·000002	·000002	·000003	*34*
35	—	—	—	—	—	·000001	·000001	·000001	·000001	·000001	*35*

x	*14·1*	*14·2*	*14·3*	*14·4*	*14·5*	*14·6*	*14·7*	*14·8*	*14·9*	*15·0*	*x*
0	·000001	·000001	·000001	·000001	·000001	—	—	—	—	—	*0*
1	·000011	·000010	·000009	·000008	·000007	·000007	·000006	·000006	·000005	·000005	*1*
2	·000075	·000069	·000063	·000058	·000053	·000049	·000045	·000041	·000038	·000034	*2*
3	·000352	·000325	·000300	·000277	·000256	·000237	·000219	·000202	·000186	·000172	*3*
4	·001239	·001153	·001073	·000999	·000929	·000864	·000803	·000747	·000694	·000645	*4*
5	·003494	·003275	·003070	·002876	·002694	·002523	·002362	·002211	·002069	·001936	*5*
6	·008212	·007752	·007316	·006902	·006510	·006139	·005787	·005454	·005138	·004839	*6*
7	·016541	·015726	·014946	·014199	·013486	·012804	·012152	·011530	·010937	·010370	*7*
8	·029153	·027913	·026715	·025559	·024443	·023367	·022330	·021331	·020370	·019444	*8*
9	·045673	·044040	·042447	·040894	·039380	·037907	·036472	·035078	·033723	·032407	*9*
10	·064399	·062537	·060700	·058887	·057101	·055343	·053614	·051915	·050247	·048611	*10*
11	·082547	·080730	·078910	·077089	·075270	·073456	·071648	·069850	·068062	·066287	*11*

TABLE—(*continued*).

x	*m*										*x*
	14·1	*14·2*	*14·3*	*14·4*	*14·5*	*14·6*	*14·7*	*14·8*	*14·9*	*15·0*	
12	·096993	·095530	·094034	·092507	·090951	·089371	·087769	·086148	·084510	·082859	*12*
13	·105200	·104349	·103437	·102469	·101446	·100371	·099247	·098076	·096862	·095607	*13*
14	·105951	·105839	·105654	·105396	·105069	·104672	·104209	·103681	·103089	·102436	*14*
15	·099594	·100195	·100723	·101181	·101567	·101881	·102125	·102298	·102402	·102436	*15*
16	·087768	·088923	·090021	·091063	·092045	·092967	·093827	·094626	·095361	·096034	*16*
17	·072795	·074277	·075724	·077135	·078509	·079842	·081133	·082380	·083581	·084736	*17*
18	·057023	·058596	·060158	·061708	·063243	·064761	·066259	·067735	·069187	·070613	*18*
19	·042317	·043793	·045277	·046768	·048264	·049763	·051263	·052762	·054257	·055747	*19*
20	·029834	·031093	·032373	·033673	·034992	·036327	·037678	·039044	·040422	·041810	*20*
21	·020031	·021025	·022045	·023090	·024161	·025256	·026375	·027517	·028680	·029865	*21*
22	·012838	·013570	·014329	·015114	·015924	·016761	·017623	·018511	·019424	·020362	*22*
23	·007870	·008378	·008909	·009462	·010039	·010640	·011264	·011911	·012584	·013280	*23*
24	·004624	·004957	·005308	·005677	·006065	·006472	·006899	·007345	·007812	·008300	*24*
25	·002608	·002816	·003036	·003270	·003518	·003780	·004057	·004348	·004656	·004980	*25*
26	·001414	·001538	·001670	·001811	·001962	·002123	·002294	·002475	·002668	·002873	*26*
27	·000739	·000809	·000884	·000966	·001054	·001148	·001249	·001357	·001473	·001596	*27*
28	·000372	·000410	·000452	·000497	·000546	·000598	·000656	·000717	·000784	·000855	*28*
29	·000181	·000201	·000223	·000247	·000273	·000301	·000332	·000366	·000403	·000442	*29*
30	·000085	·000095	·000106	·000118	·000132	·000147	·000163	·000181	·000200	·000221	*30*
31	·000039	·000044	·000049	·000055	·000062	·000069	·000077	·000086	·000096	·000107	*31*
32	·000017	·000019	·000022	·000025	·000028	·000032	·000035	·000040	·000045	·000050	*32*
33	·000007	·000008	·000009	·000011	·000012	·000014	·000016	·000018	·000020	·000023	*33*
34	·000003	·000003	·000004	·000005	·000005	·000006	·000007	·000008	·000009	·000010	*34*
35	·000001	·000001	·000002	·000002	·000002	·000002	·000003	·000003	·000004	·000004	*35*
36	—	·000001	·000001	·000001	·000001	·000001	·000001	·000001	·000002	·000002	*36*
37	—	—	—	—	—	—	—	·000001	·000001	·000001	*37*

ON THE POISSON LAW OF SMALL NUMBERS.

BY LUCY WHITAKER, B.SC.

PART I. THEORY AND APPLICATION TO CELL-FREQUENCIES.

(1) *Introductory.*

Let p denote the probability of the happening of a certain event A, and $q = 1 - p$, the probability of its failure in one trial. Then it is well known that the distribution of the frequencies of occurrence $n, n-1, n-2 \dots$ times in a series N of n trials is given by the terms of the point binomial

$$N(p+q)^n \quad \dots\dots\dots\dots\text{(i)}.$$

The fitting of point-binomials plotted on an elementary base c to observed frequency distributions has been discussed by Pearson*, and he has indicated that, if c be unknown, the problem can be solved in terms of the three moment coefficients μ_2, μ_3, μ_4 required to find c, p and n. In actual practice but few cases of frequency can be found which are describable in terms of a point-binomial, and of these few a considerable section have n negative, p greater than unity and q negative; thus defying at present interpretation, however well they may serve as an analytical expression of the frequency.

The hypothesis made in deducing the binomial $(p+q)^n$ as a description of frequency is clearly that each trial shall be absolutely independent of those which precede it. In this respect it may be said that binomial frequencies belong to the teetotum class of chances, and not to those of card-drawings, when each drawing is unreplaced. In the latter case the "contributory cause groups are not independent," and our series corresponds to the hypergeometrical rather than to the binomial type of progression†.

Using the customary notation $\beta_1 = \mu_3^2/\mu_2^3$, $\beta_2 = \mu_4/\mu_2^2$, the binomial is determined from:

$$\left.\begin{aligned} n &= 2/\{3 - \beta_2 + \beta_1\}, \quad c = \sigma\sqrt{6 - 2\beta_2 + 3\beta_1} \\ pq &= \tfrac{1}{2}(3 - \beta_2 + \beta_1)/(6 - 2\beta_2 + 3\beta_1) \end{aligned}\right\} \dots\dots\dots\text{(ii)}.$$

---

* "Skew Variation in Homogeneous Material," *Phil. Trans.* Vol. 186, A, p. 347, 1895.

† *Phil. Trans.* Vol. 186, A, p. 381, 1895.

In order that n should be positive, it is needful that

$$3 - \beta_2 + \beta_1, \quad = \tfrac{1}{2}(6 - 2\beta_2 + 2\beta_1),$$

should be positive. If this is satisfied clearly c will be real because β_1 is always positive. Further then

$$pq = p(1-p) = \frac{1}{4} \times \frac{6 - 2\beta_2 + 2\beta_1}{6 - 2\beta_2 + 3\beta_1}$$

is always less than a quarter and p and q will therefore be real. If the reader will turn to Rhind's diagram, *Biometrika*, Vol. VII. p. 131, he will see that the line $3 - \beta_2 + \beta_1 = 0$ cuts off all curves of Types III, IV, V and VI, and includes a portion only of Type I, with a part of its U and J varieties. The binomial description of frequency, therefore, is not—considering our experience of frequency distributions—likely to be of very universal application.

(2) *Further Limitations.*

Now let us still further limit our binomial by supposing:

(i) that the unit of grouping of the observed frequencies corresponds to the actual binomial base unit c and (ii) that the first of the observed frequencies corresponds to the term Np^n of the binomial*.

In this case the mean m of the observed frequency measured from the first term of the frequency will be equal to the nq of the binomial and the standard deviation of the observed distribution will be equal to $\sqrt{npq}$. We have thus:

$$p = \sigma^2/m, \quad q = 1 - \sigma^2/m, \quad n = m^2/(m - \sigma^2) \text{..................(iii)}$$

and n and q will both be negative, if m be less than σ^2. The condition for a positive binomial is therefore that σ be less than $\sqrt{m}$.

(3) *Probable errors of the constants of a Binomial Frequency.*

It is desirable to find the probable errors of p and n as determined by these formulae. We have:

$$\mu_1' = nq, \qquad \mu_2 = npq,$$
$$\delta\mu_1' = q\delta n + n\delta q, \quad \delta\mu_2 = pq\delta n + nq\delta p + np\delta q,$$

assuming deviations may be represented by differentials.

Hence, since $dp = -dq$:

$$\delta\mu_2 - (p-q)\,\delta\mu_1' = q^2\delta n \quad \text{and} \quad p\delta\mu_1' - \delta\mu_2 = nq\delta q.$$

Square each of these results, sum for all samples and divide by the number of samples, and we have:

$$\sigma^2_{\mu_2} + (p-q)^2\sigma^2_{\mu_1'} - 2(p-q)\,\sigma_{\mu_2}\sigma_{\mu_1'}r_{\mu_2\mu_1'} = q^4\sigma_n^{\,2}$$
$$\sigma^2_{\mu_2} + p^2\sigma^2_{\mu_1'} - 2p\sigma_{\mu_2}\sigma_{\mu_1'}r_{\mu_2\mu_1'} = n^2q^2\sigma_q^{\,2}.$$

---

* The exact nature of these limitations must be fully appreciated. The best fitting binomial to a given frequency distribution will usually be far from one in which the first term of the binomial corresponds to the first observed frequency. The modes of the binomial and the observed frequency will closely correspond, but the "tails" of the binomial may be quite insignificant and correspond to no observed frequencies.

Now σ_{μ_2} is the standard deviation of variations in μ_2 and therefore

$$\sigma^2_{\mu_2} = (\mu_4 - \mu_2^2)/N.$$

Similarly $\sigma_{\mu_1'}$ is the standard deviation of variations in the mean and therefore $\sigma^2_{\mu_1'} = \mu_2/N$. Lastly the product $\sigma_{\mu_2}\sigma_{\mu_1'}r_{\mu_2\mu_1'}$ measures the correlation between deviations in μ_2 and μ_1' and is known to be μ_3/N[*].

Thus we have:

$$q^4\sigma_n^2 = \frac{1}{N}\{\mu_4 - \mu_2^2 + (p-q)^2\mu_2 - 2(p-q)\mu_3\},$$

$$n^2q^2\sigma_q^2 = \frac{1}{N}\{\mu_4 - \mu_2^2 + p^2\mu_2 - 2p\mu_3\}.$$

But[†]

$$\left.\begin{array}{l}\mu_4 = npq\,\{1 + 3(n-2)\,pq\}, \\ \mu_3 = npq\,(p-q), \quad \mu_2 = npq\end{array}\right\} \qquad \text{(iv)}.$$

Whence after some purely algebraical reductions we deduce:

$$\sigma_n = \frac{n\,p}{\sqrt{N}\,q}\sqrt{2\left(1-\frac{1}{n}\right)} = \frac{\sigma^2}{\sqrt{N}\,q^2}\sqrt{2\left(1-\frac{1}{n}\right)} \qquad \text{(v)},$$

$$\sigma_p = \sigma_q = \frac{p}{\sqrt{N}}\sqrt{2 + \frac{1-3p}{np}} \qquad \text{(vi)}.$$

Formulae (v) and (vi) are very important; they enable us to obtain the probable errors for n and p when a binomial limited in the present manner is fitted to a frequency distribution[‡].

We see at once, that as n grows large and q grows small

$$\sigma_p = \sigma_q \text{ approaches the limit } \sqrt{2/N},$$

or the probable error, ·67449 $\sqrt{2/N}$, of p and q is finite. But σ^2 being finite σ_n becomes infinitely great, or the probable error of n indefinitely large. Thus when the n of the binomial is very large, q being very small, the probable error of its determination is so great that its actual value is not capable of being found accurately. Again, suppose N embraced 200 observations, the probable error of q would be of the order ·07; if N corresponded to only eighteen observations, then the probable error of q would be of the order ·22. It is clearly wholly impossible

---

[*] *Biometrika*, Vol. II. "On the Probable errors of Frequency Constants," see p. 275 (iv), p. 276 (vii), and p. 279 (xii).

[†] *Phil. Trans.* Vol. 186, A, p. 347, 1895.

[‡] There is no difficulty in obtaining the probable errors of n and p from the more general values in (ii). In this case

$$\sigma_n = \tfrac{1}{2}n^2\sqrt{\sigma^2_{\beta_1} + \sigma^2_{\beta_2} - 2\sigma_{\beta_1}\sigma_{\beta_2}r_{\beta_1\beta_2}},$$

$$\sigma_p = \sigma_q = \frac{pq}{\beta_1}\sqrt{\sigma^2_{\beta_1} + \tfrac{1}{4}n^2\beta_1^2\sigma^2_{\beta_2} - n\beta_1\sigma_{\beta_1}\sigma_{\beta_2}r_{\beta_1\beta_2}}.$$

The values of σ_{β_1}, σ_{β_2} and $r_{\beta_1\beta_2}$ for different values of β_1 and β_2 have been tabled by Rhind, *Biometrika*, Vol. VII. pp. 136—141.

from series of observations even of the order 200, much less of order 18, to assert that q is or is not really a "small quantity." Thus the observed value of q corresponding to a population of extremely small q might easily show $q = {\cdot}15$ to ${\cdot}50$!.

(4) *Poisson—Law of Small Numbers.*

A last limitation of the point-binomial is made by supposing the mean $m = nq$ to remain finite, but q to be indefinitely small. We write:

$$N(p+q)^n = N(1-q+q)^n = N(1-q)^{\frac{m}{q}}\left(1 + \frac{q}{1-q}\right)^{\frac{m}{q}}$$

$$= N(1-q)^{\frac{m}{q}}(1+q)^{\frac{m}{q}} \text{ nearly}$$

$$= Ne^{-m}\left(1 + m + \frac{m^2}{2\,!} + \frac{m^3}{3\,!} + \dots\right).$$

Here the successive terms give the frequency of occurrence of 0, 1, 2, 3... successes on the basis of each success not being prejudiced by what has previously occurred. This is the Law of Small Numbers. It was first published by Poisson in 1837*. It was adopted later by Bortkewitsch, who published a small treatise expanding by illustrations Poisson's work†. The same series was deduced later by "Student" in ignorance of both Poisson and Bortkewitsch's papers, when dealing with the counts made with a haemacytometer‡.

The mean is at m from the first group, the other moments as "Student" has shewn § are:

$$\mu_2 = m, \quad \mu_3 = m, \quad \mu_4 = 3m^2 + m.$$

Hence $\qquad \beta_1 = 1/m, \quad \beta_2 - 3 = 1/m.$

When the mean value is large, β_1, β_2 and the higher β's approach the values given by the Gaussian curve.

Clearly the Poisson-Exponential formula contains only the single constant $m = \mu_1'$ and its probable error is therefore ${\cdot}67449\sigma/\sqrt{N} = {\cdot}67449\sqrt{\frac{m}{N}}$. This will, if N be reasonably large and m not too big, be a small or at any rate a finite quantity (i.e. not like σ_n for q very small). Hence it might be supposed, although erroneously, that the Poisson-Exponential formula was capable of great accuracy in addition to its great simplicity. But this is to neglect the fundamental assumptions on which it is based, namely:

(i) that the data actually correspond to a binomial,

(ii) that in that binomial q is small and n large.

Clearly (i) shows us that, if we can find the binomial, it will actually be closer to the observed frequency than Poisson's merely approximate formula.

---

* *Recherches sur la Probabilité des Jugements.* Paris, 1837, pp. 205 *et seq.*

† *Das Gesetz der kleinen Zahlen*, Leipzig, 1898.

‡ "On the Error of Counting with a Haemacytometer," *Biometrika*, Vol. v, pp. 351—5, 1907.

§ They may be deduced at once from (iv).

Secondly (ii) can only be justified as an assumption by actually ascertaining the form of the binomial from the data and testing whether n is large and q small and *positive*. It appears absurd to base our formula on an approximation to a binomial of a particular kind when, on testing in the actual problem, such a binomial does not describe the results. As a merely empirical formula, the Poisson-Exponential of course can be tested by the usual processes for measuring goodness of fit, but no such test nor any discussion of the probable errors of their results have been provided by Bortkewitsch himself nor by Mortara, who has followed recently his lines in a work to be considered later. As a matter of fact in the cases dealt with by Bortkewitsch, by Mortara and by "Student," n will be found almost as frequently small and negative as large and positive, and q takes a great variety of values large and negative and large and positive, as well as small and positive. Thus the initial assumptions made from which the "law of small numbers" is deduced are by no means justified on the material to which it has so far been applied.

(5) *Application of the Law of Small Numbers to determine the Probable Errors of Small Frequencies.* Given a distribution of frequency for a population $\overline{N}$ let $\overline{n}_{st}$ be the frequency in the cell of the sth row and tth column of a contingency table (or if we drop t, $\overline{n}_s$ would stand for the frequency of any class). Then if we take a random sample of N individuals from this population, the chance that an individual is taken out of the $\overline{n}_{st}$ cell is $\overline{n}_{st}/\overline{N}$, and that it is not is $1-\frac{\overline{n}_{st}}{\overline{N}}$. Therefore if the original population be so large that the withdrawal of an individual does not affect the next draw, the frequency of individuals in M random samples of N will be given by the terms of the binomial:

$$M\left\{\left(1-\frac{\overline{n}_{st}}{\overline{N}}\right)+\frac{\overline{n}_{st}}{\overline{N}}\right\}^N.$$

Now, if $\overline{n}_{st}/\overline{N}$ be very small, and N large this will approximate to the Poisson series:

$$Me^{-m}\left(1+m+\frac{m^2}{2\,!}+\frac{m^3}{3\,!}+\dots\right),$$

where $m=\frac{\overline{n}_{st}}{\overline{N}}\times N$. But $\overline{n}_{st}/\overline{N}$ will approximately be the mean proportion of the whole in the st cell of the sample itself $=n_{st}/N$, or $m=n_{st}$. Thus if in any cell of a contingency table, or in any sub-class of a frequency whatsoever, we have a frequency n_{st} small as compared to the population N, then in sampling, this small frequency will have a distribution approximating to the Poisson Law, and tending as n_{st} becomes larger to approach the Gaussian distribution*. It would appear,

---

* Such approach is usually *assumed* when we speak of

$$\cdot 67449\sqrt{n_{st}\left(1-\frac{n_{st}}{N}\right)}$$

as the probable error of the frequency n_{st}. But such a "probable error" has really no meaning if n_{st} be very small and the exponential law be applied.

therefore, that the Poisson Law of Small Numbers should be applied in order to deal with the errors of random sampling in any *small* frequency, and an appeal should not be made—as is usually the case—to Sheppard's Tables on the assumption that the frequency is Gaussian.

The following Table I illustrates the results obtained (*a*) from the Binomial, (*b*) from the Poisson-Exponential and (*c*) from the normal curve on the two hypotheses that (i) the frequency is 10 in the 1000 and (*b*) is 30 in the 1000. But here a word must be said as to which Gaussian is to be compared with the Binomial or the Poisson-Exponential. The usual method of fitting a Gaussian is to give it the same mean and standard-deviation as the material to which we are fitting it. For example, we should compare the Poisson exponential with a Gaussian at mean m and with standard-deviation $\sqrt{m}$, or the point binomial with mean nq

TABLE I.

Comparison of Binomial, Poisson-Exponential and Gaussian for cell-frequency variations in samples for case of 10 *and* 30 *in a total population of* 1000

Percentage Frequency

Number in Cell	*10* in *1000*				*30* in *1000*		
	Binomial	Poisson-Exponential	Gaussian		Binomial	Poisson-Exponential	Gaussian
0	·00004	·00005	·00132	*19*	·00848	·00894	·01100
1	·00044	·00045	·00327	*20*	·01287	·01341	·01553
2	·00020	·00227	·00735	*21*	·01857	·01916	·02118
3	·00739	·00757	·01491	*22*	·02556	·02613	·02792
4	·01861	·01892	·02736	*23*	·03362	·03408	·03544
5	·03745	·03783	·04539	*24*	·04233	·04260	·04373
6	·06274	·06306	·06806	*25*	·05110	·05112	·05198
7	·08999	·09080	·09224	*26*	·05927	·05898	·05970
8	·11282	·11260	·11300	*27*	·06613	·06553	·06625
9	·12561	·12511	·12514	*28*	·07107	·07021	·07104
				29	·07367	·07263	·07360
10	·12574	·12511	·12526	*30*	·07375	·07263	·07367
11	·11431	·11374	·11334				
12	·09516	·09478	·09271	*31*	·07137	·07029	·07126
13	·07305	·07291	·06854	*32*	·06684	·06590	·06659
14	·05202	·05208	·04580	*33*	·06064	·05991	·06013
15	·03454	·03472	·02767	*34*	·05334	·05286	·05246
16	·02148	·02170	·01511	*35*	·04553	·04531	·04423
17	·01256	·01276	·00746	*36*	·03775	·03776	·03602
18	·00693	·00709	·00333	*37*	·03042	·03061	·02835
19	·00362	·00373	·00134	*38*	·02384	·02417	·02156
20	·00179	·00187	·00049	*39*	·01819	·01859	·01584
21	·00085	·00089	·00016	*40*	·01351	·01394	·01125
22	·00038	·00040	·00005	*41*	·00979	·01020	·00771
23	·00016	·00018	·00001	*42*	·00691	·00729	·00511

and standard-deviation $\sqrt{npq}$. These will, however, not be identical standard deviations as p is not truly unity. In ordinary practice, in testing for example the 30 in 1000 frequency, we should put the centre of our Gaussian at our 30 group, and use a standard deviation $= \sqrt{30\ (1-30/1000)} = \sqrt{30 \times \cdot 97} = 5\cdot 39444$ to enter the table of the probability integral. This is, of course, the Gaussian we obtain by the method of least squares, but to assume that it is "the best" is to argue in a circle, because we then take least squares as a test of what is best[*]. It is not the Gaussian which is directly reached by proceeding either to a limit of the Binomial or to the Exponential, for example, by applying Stirling's Theorem. It will be seen by examining Table II that the Gaussian curve develops out of the exponential by a mode at the point midway between the two equal terms, rather than by a mode at the mean, which coincides with the centre of the second of them. If we apply Stirling's Theorem to the term[†]

$$N \frac{\lfloor n}{\lfloor n-r \lfloor r} p^{n-r} q^r$$

of the binomial $N(p+q)^n$ it becomes

$$u_r = \frac{N}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{1}{2}\{r-nq+\frac{1}{2}(p-q)\}^2/(npq)},$$

i.e. the *ordinate* of a Gaussian curve of Standard Deviation $\sqrt{npq}$ and mean at $nq - \frac{1}{2}(p-q)$. These give for the Poisson-Exponential the Gaussian with standard-deviation $\sqrt{m}$ and mean $m - \frac{1}{2}$. The above type of curve which gives frequencies by coordinates and not by areas has been termed by Sheppard a 'spurious curve of frequency'; at the same time it is the method by which Laplace and Poisson first reached the normal curve, and the real point at issue is whether we shall get better approximations to the discontinuous frequencies of the binomials by using Gaussian ordinates than by using the areas of a Gaussian curve. At the same time it has been shewn[‡] that if a Gaussian curve gives a series of frequencies by its areas, then if its standard-deviation be σ^2, a spurious Gaussian frequency curve with standard deviation given by $\sigma_0^2 = \sigma^2 + \frac{1}{12}h^2$, h being the sub-range, will closely give the frequencies by its ordinates. It seems probable therefore that the Gaussian curve with mean at $nq - \frac{1}{2}(p-q)$ and standard deviation $\sqrt{npq - \frac{1}{12}}$ will more closely represent the binomial for cell frequency variation by its areas,

---

[*] There is a further flaw in this treatment—the Gaussian is continuous, the Binomial and the Poisson-Exponential are not. If t_r be the rth term of either of the latter series, we ought really to make

$$S_0^\infty \left[\left\{ t_r - \int_{r-\overline{m}}^{r+1-\overline{m}} \frac{N}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{x}{\sigma^2}} dx \right\}^2 \Big/ t_r \right] = u,$$

a minimum by the conditions $du/d\overline{m} = du/d\sigma = 0$. No complete solution of this problem has hitherto been determined.

[†] The final form for u_r may be obtained by neglecting the terms in $\frac{1}{n^2}$ in the formula given by Pearson, *Phil. Trans.* Vol. 186, A, p. 347, footnote.

[‡] *Biometrika*, Vol. III. p. 311.

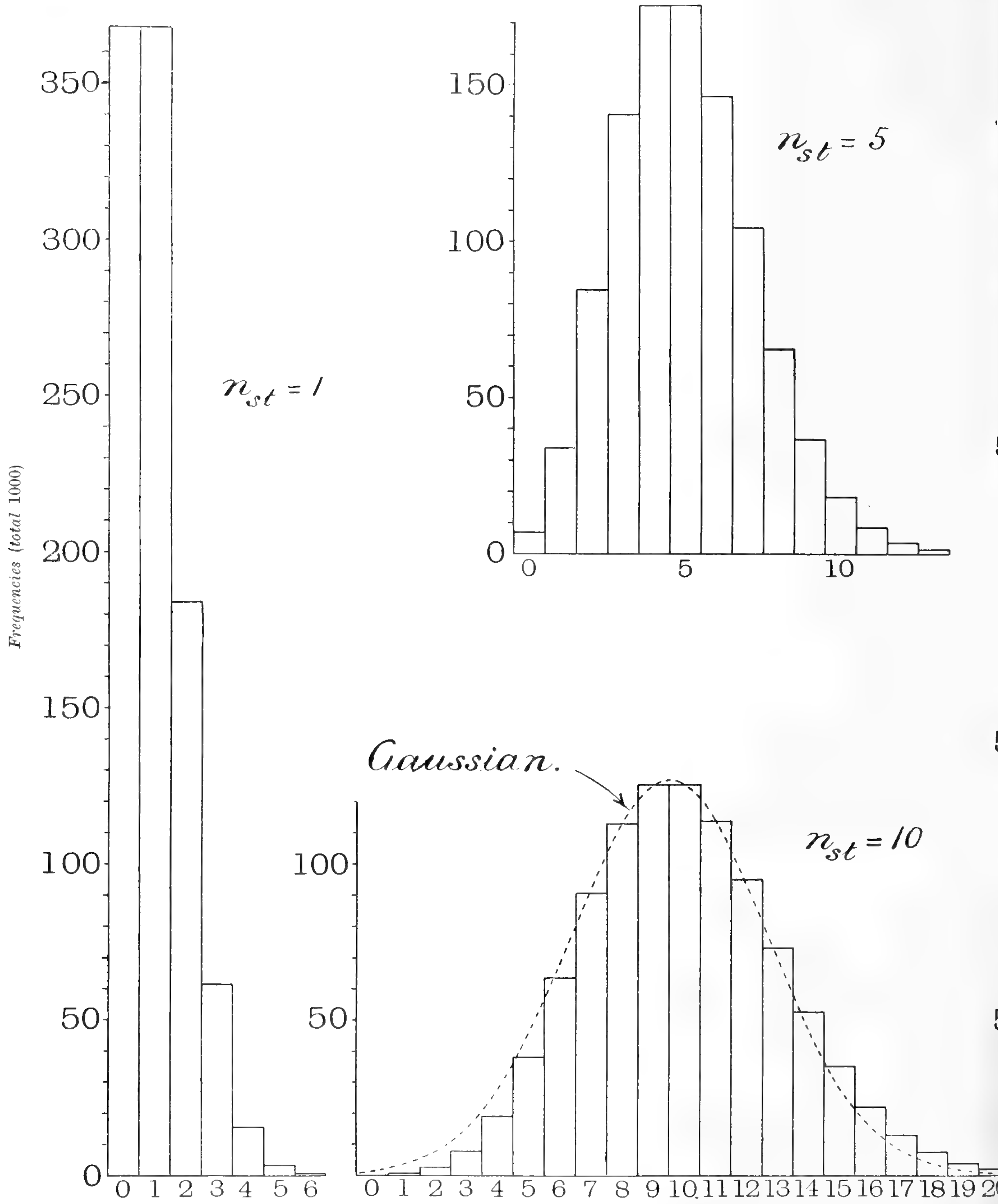
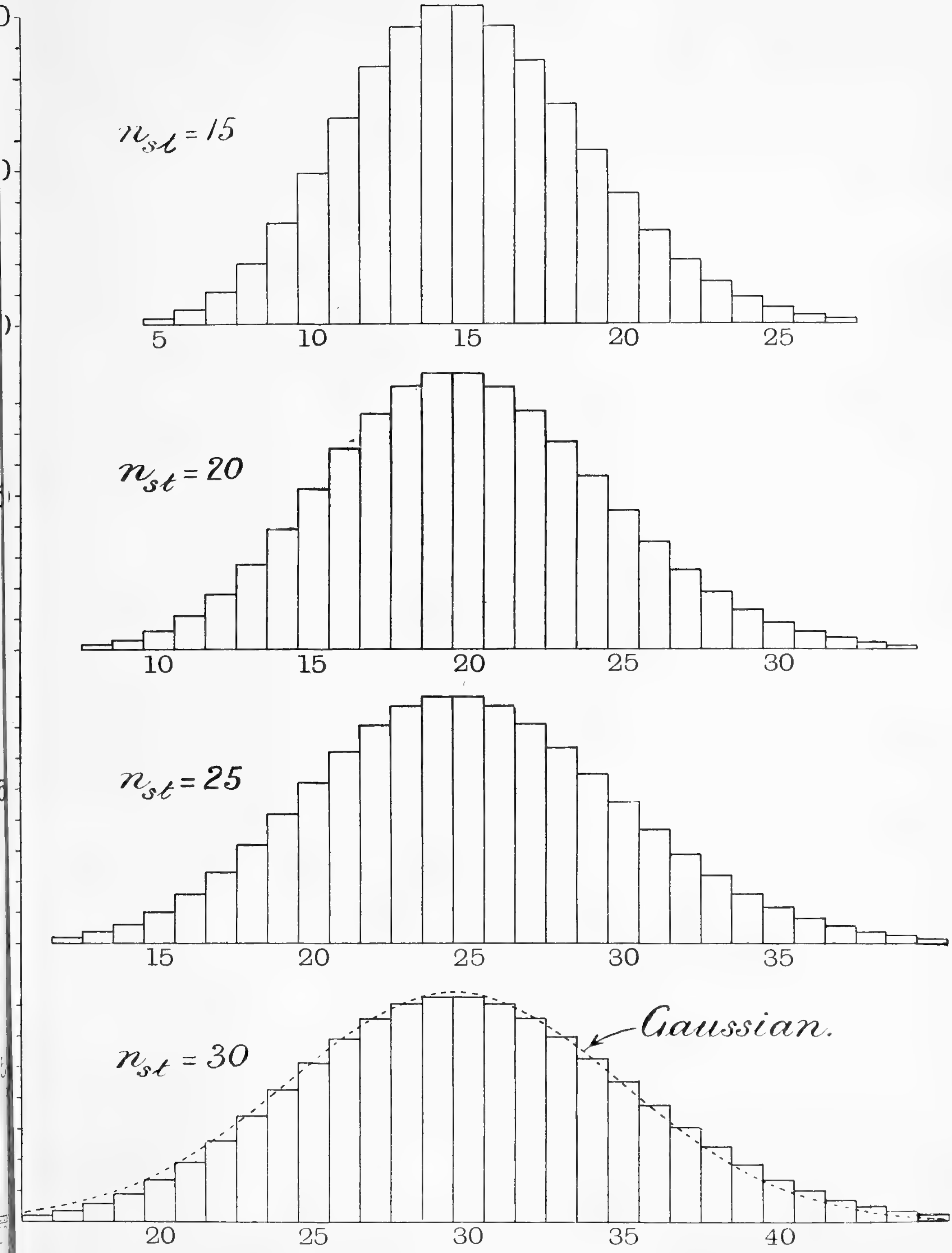
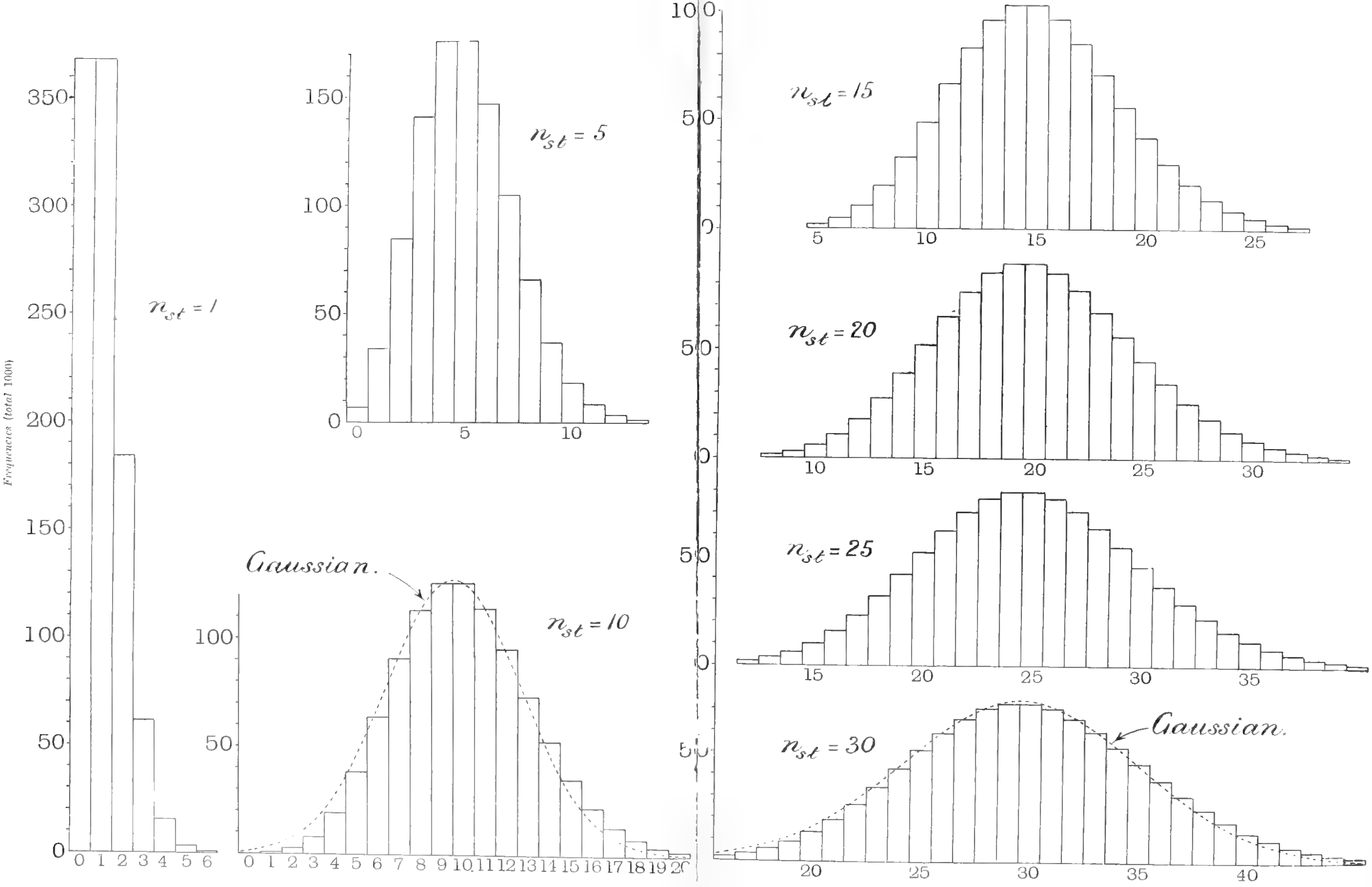

Frequencies (total 1000)
350
300
250
200
150
100
50
0
0 1 2 3 4 5 6
n_{st} = 1
150
100
50
0
0 5 10
n_{st} = 5
Gaussian.
100
50
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
n_{st} = 10

Plate VII



$n_{st} = 15$
5
10
15
20
25
$n_{st} = 20$
10
15
20
25
30
$n_{st} = 25$
15
20
25
30
35
$n_{st} = 30$
Gaussian.
20
25
30
35
40


Frequencies (total 1000)
$n_{st} = 1$
$n_{st} = 5$
$n_{st} = 10$
$n_{st} = 15$
$n_{st} = 20$
$n_{st} = 25$
$n_{st} = 30$
Gaussian.
Gaussian.
350
300
250
200
150
100
50
0
0 1 2 3 4 5 6
0 5 10
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5 10 15 20 25
10 15 20 25 30
15 20 25 30 35
20 25 30 35 40

than if we apply the ordinary process of mean nq, standard deviation $\sqrt{npq}$, and Sheppard's table for areas to the frequencies. It will be noted that this amounts to using Sheppard's correction on the crude second-moment and slightly shifting the central ordinate towards the side of greater frequency. This is the Gaussian curve used in Table I.

The object of the present section of our work is to indicate how far it is legitimate to use the Poisson-Exponential up to cell frequencies of the order 30 in a population of about 1000* and how far we then reach a state of affairs, which for practical purposes may be described by ordinary tables of the Gaussian. It will be seen from Table I that the Poisson-Exponential even for $n_{st} = 10$ and 30 is not extremely divergent from the Binomial.

In Plate VII the transition of the exponential histograms of frequency towards the Gaussian form is indicated for cell-frequency = 1, 5, 10, 15, 20, 25 and 30; in the cases of 10 and 30 the corresponding Gaussian curves are drawn.

It will be seen that with due caution the Poisson-Exponential may be reasonably used up to frequencies of about 30 in the 1000, and that after that it would be fairly satisfactory to use the areas of the Gaussian curve as provided in the usual tables.

(6) In order to table the results of the Poisson-Exponential for easy use, it seemed desirable to turn them into percentages of excess and defect. For example take the distribution for a frequency 5. It is:

		Per cent. of Cases in which:			
0	·006,737,945	a defect of 5		occurs:	0·674
1	·033,689,725	„	4 or more	„ :	4·043
2	·084,224,310	„	3 or more	„ :	12·465
3	·140,373,850	„	2 or more	„ :	26·503
4	·175,467,310	„	1 or more	„ :	44·049
5	·175,467,310	the true value		„ :	17·547
6	·146,222,755	an excess of 1 or more		„ :	38·404
7	·104,444,825	„	2 or more	„ :	23·782
8	·065,278,015	„	3 or more	„ :	13·337
9	·036,265,564	„	4 or more	„ :	6·809
10	·018,132,782	„	5 or more	„ :	3·183
11	·008,242,173	„	6 or more	„ :	1·370
12	·003,434,238	„	7 or more	„ :	0·545
13	·001,320,860	„	8 or more	„ :	0·202

---

* Of course in the Poisson-Exponential itself the total frequency plays no part; it is only useful in testing the validity of the approximation.

Thus we see that if the true value of the frequency be 5 for the average sample, it will only lie outside the range 1 to 10 in ·674 + 1·370 = 2·044 cases per cent., or the odds are 49 to 1 that the value found will be from 1 to 10.

On the other hand it will lie outside the range 2 to 8 in 4·043+6·809=10·852 °/₀ of cases, or once in about 9 trials the frequency will lie outside this range. Or, again, once in about every four trials (25·8 °/₀) the result will fall outside the range 3 to 7.

On the other hand if we write $\sigma = \sqrt{5(1 - \cdot 005)} = 2\cdot 23047$, we have $-4\cdot 5$ and $+5\cdot 5$ as the deviations from a mean 5 of all beyond 0·5 and above 10·5, giving $x/\sigma = -2\cdot 0175$ and $+2\cdot 4658$ respectively. These cut off tail areas of ·02181 and ·00684, respectively. Thus in 2·865—not 2·044—per cent. of cases we should assert that the frequency would lie outside the range 1 to 10, or the odds that it would lie inside this range are now only about 34 to 1, not 49 to 1. Calculated from the Gaussian the frequencies outside ranges 2 to 8 and 3 to 7 correspond to 10·1 °/₀ and 26·2 °/₀ of the trials instead of 10·9 °/₀ and 25·8 °/₀. If we take for the standard-deviation of our Gaussian $\sqrt{npq - \frac{1}{12}} = 2\cdot 21171$, we find that the odds in the first case are still only 35 to 1, but the percentages in the other two cases are 11·3 and 25·8.

It will be clear that near the centre of the curve—especially when we equalise the excess and defect of the Gaussian by taking equal ranges on both sides—it does not give bad percentages of frequency, but that it does not lend itself to the accurate determination of the range for reasonable working odds such as 50 to 1.

It will be noted that the total area in excess and defect of 2 and more = 23·782 + 26·503 = 50·285, or corresponds very nearly to the "probable error." Actually the Gaussians with standard deviations of 2·23047 and 2·21171 give probable errors of 1·504 and 1·492 respectively, so that the Gaussian with 1·5 as the probable error is very nearly accurate.

Table II gives the Poisson-Exponential; it will enable the reader to appreciate the range of probable variation in small frequencies. Thus we realise that in 37 °/₀ of cases in which the true frequency is 1, the cell will be found empty; in 13·5 per cent. of cases it will be empty when the actual frequency is 2, and in 5 °/₀ of cases when the frequency is 3 and in 1·8 °/₀ when the frequency is 4. These results indicate how rash it is to assume that a sample 4-fold table with one zero quadrant signifies perfect dependence or association in the attributes of the material sampled. The second line below gives the percentages of cases that 0 would appear in a cell when the actual number to be expected is that in the first line calculated from Table II on the usual theory of *a priori* probabilities:

Actual ...	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9 & over*
Percentage ...	63·21	23·25	8·55	3·15	1·16	0·43	0·16	0·06	0·02	0·01

TABLE II.

Table of Poisson-Exponential for Cell Frequencies 1 *to* 30.

Cell Frequencies

	x	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Per cent. occurrence of values differing by *x* or more in defect from Actual.	*22*										
	21										
	20										
	19										
	18										
	17										
	16										
	15										
	14										
	13										
	12										
	11										
	10										·005
	9									·012	·050
	8								·034	·123	·277
	7							·091	·302	·623	1·033
	6						·248	·730	1·375	2·123	2·925
	5					·674	1·735	2·964	4·238	5·496	6·708
	4				1·832	4·043	6·197	8·177	9·963	11·569	13·014
	3			4·979	9·158	12·465	15·120	17·299	19·124	20·678	22·022
	2		13·534	19·915	23·810	26·503	28·506	30·071	31·337	32·390	33·282
	1	36·788	40·601	42·319	43·347	44·049	44·568	44·971	45·296	45·565	45·793
	Actual	36·788	27·067	22·404	19·537	17·547	16·062	14·900	13·959	13·176	12·511
Per cent. occurrence of values differing by *x* or more in excess from Actual.	*1*	26·424	32·332	35·277	37·116	38·404	39·370	40·129	40·745	41·259	41·696
	2	8·030	14·288	18·474	21·487	23·782	25·602	27·091	28·338	29·401	30·323
	3	1·899	5·265	8·392	11·067	13·337	15·276	16·950	18·411	19·699	20·845
	4	·366	1·656	3·351	5·113	6·809	8·392	9·852	11·192	12·422	13·554
	5	·059	·453	1·191	2·136	3·183	4·262	5·335	6·380	7·385	8·346
	6	·008	·110	·380	·813	1·370	2·009	2·700	3·418	4·146	4·875
	7	·001	·024	·110	·284	·545	·883	1·281	1·726	2·203	2·705
	8	·000	·005	·029	·092	·202	·363	·572	·823	1·110	1·428
	9	—	·001	·007	·027	·070	·140	·241	·372	·532	·719
	10	—	·000	·002	·008	·023	·051	·096	·159	·242	·346
	11	—	—	·000	·002	·007	·018	·036	·065	·105	·160
	12	—	—	—	·001	·002	·006	·013	·025	·044	·071
	13	—	—	—	·000	·001	·002	·005	·009	·017	·030
	14	—	—	—	—	·000	·001	·002	·003	·007	·013
	15	—	—	—	—	—	·000	·001	·001	·002	·006
	16	—	—	—	—	—	—	·000	·000	·001	·002
	17	—	—	—	—	—	—	—	—	·000	·001
	18	—	—	—	—	—	—	—	—	—	·001
	19	—	—	—	—	—	—	—	—	—	·000
	20	—	—	—	—	—	—	—	—	—	—
	21	—	—	—	—	—	—	—	—	—	—
	22	—	—	—	—	—	—	—	—	—	—
	23	—	—	—	—	—	—	—	—	—	—
	24	—	—	—	—	—	—	—	—	—	—
	25	—	—	—	—	—	—	—	—	—	—
	26	—	—	—	—	—	—	—	—	—	—
	27	—	—	—	—	—	—	—	—	—	—
	28	—	—	—	—	—	—	—	—	—	—

TABLE II—(*continued*).

Cell Frequencies

	x	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
Per cent. occurrence of values differing by *x* or more in defect from Actual.	*22*										
	21										
	20										—
	19									—	—
	18								—	—	—
	17							—	—	·000	·000
	16						—	·000	·000	·001	·002
	15					·000	·000	·001	·002	·004	·007
	14				·000	·001	·002	·004	·008	·015	·026
	13			·000	·001	·004	·009	·018	·032	·052	·078
	12		·001	·003	·009	·021	·040	·067	·104	·151	·209
	11	·002	·008	·022	·047	·086	·138	·206	·289	·387	·500
	10	·020	·052	·105	·181	·279	·401	·543	·706	·886	1·081
	9	·121	·229	·374	·553	·763	1·000	1·260	1·538	1·832	2·139
	8	·492	·760	1·073	1·423	1·800	2·199	2·612	3·037	3·467	3·901
	7	1·510	2·034	2·589	3·162	3·745	4·330	4·912	5·489	6·056	6·613
	6	3·752	4·582	5·403	6·206	6·985	7·740	8·467	9·167	9·840	10·486
	5	7·861	8·950	9·976	10·940	11·846	12·699	13·502	14·260	14·975	15·651
	4	14·319	15·503	16·581	17·568	18·475	19·312	20·087	20·808	21·479	22·107
	3	23·198	24·239	25·168	26·004	26·761	27·451	28·084	28·665	29·203	29·703
	2	34·051	34·723	35·317	35·846	36·322	36·753	37·146	37·505	37·836	38·142
	1	45·989	46·150	46·311	46·445	46·565	46·674	46·774	46·865	46·948	47·026
	Actual	11·938	11·437	10·994	10·599	10·244	9·922	9·629	9·360	9·112	8·884
Per cent. occurrence of values differing by *x* or more in excess from Actual.	*1*	42·073	42·404	42·695	42·956	43·191	43·404	43·597	43·776	43·939	44·091
	2	31·130	31·846	32·486	33·064	33·588	34·066	34·503	34·909	35·283	35·630
	3	21·871	22·798	23·639	24·408	25·114	25·765	26·367	26·928	27·451	27·939
	4	14·596	15·559	16·450	17·280	18·053	18·776	19·451	20·088	20·686	21·251
	5	9·261	10·129	10·953	11·736	12·478	13·184	13·852	14·491	15·099	15·677
	6	5·593	6·297	6·983	7·650	8·297	8·923	9·526	10·111	10·675	11·219
	7	3·219	3·742	4·266	4·791	5·311	5·825	6·329	6·826	7·313	7·789
	8	1·769	2·128	2·501	2·884	3·275	3·669	4·064	4·461	4·856	5·248
	9	·929	1·160	1·407	1·671	1·947	2·232	2·523	2·824	3·127	3·433
	10	·467	·607	·762	·933	1·117	1·312	1·516	1·732	1·954	2·182
	11	·225	·305	·396	·502	·619	·746	·882	1·030	1·185	1·348
	12	·104	·148	·201	·261	·331	·411	·497	·595	·699	·809
	13	·047	·069	·097	·131	·172	·219	·272	·333	·400	·473
	14	·020	·031	·046	·063	·086	·114	·144	·182	·223	·269
	15	·008	·014	·021	·030	·042	·057	·074	·096	·121	·149
	16	·003	·006	·009	·013	·020	·028	·036	·050	·064	·081
	17	·001	·002	·004	·006	·009	·014	·017	·025	·033	·042
	18	·001	·000	·002	·002	·004	·006	·008	·012	·017	·022
	19	·000	·000	·001	·001	·002	·003	·003	·006	·008	·011
	20	—	—	·001	·000	·001	·002	·002	·003	·004	·005
	21	—	—	·000	·000	·000	·001	·001	·002	·002	·003
	22	—	—	—	—	—	·000	·000	·001	·001	·001
	23	—	—	—	—	—	—	—	·000	·000	·001
	24	—	—	—	—	—	—	—	—	—	·000
	25	—	—	—	—	—	—	—	—	—	—
	26	—	—	—	—	—	—	—	—	—	—
	27	—	—	—	—	—	—	—	—	—	—
	28	—	—	—	—	—	—	—	—	—	—

TABLE II—(*continued*).

Cell Frequencies

	x	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*
Per cent. occurrence of values differing by *x* or more in defect from Actual.	*22*	—	—	—	—	—	—	—	—	—	·000
	21	—	—	—	—	—	—	·000	·000	·000	·001
	20	—	—	—	—	·000	·000	·001	·001	·001	·002
	19	—	—	·000	·000	·001	·001	·002	·003	·004	·006
	18	·000	·000	·001	·001	·002	·004	·006	·009	·012	·017
	17	·001	·002	·003	·005	·008	·011	·016	·023	·031	·041
	16	·003	·006	·010	·015	·022	·031	·043	·056	·073	·092
	15	·012	·020	·030	·043	·059	·078	·102	·129	·160	·195
	14	·039	·058	·081	·109	·142	·180	·224	·273	·328	·387
	13	·111	·150	·198	·252	·314	·384	·460	·543	·632	·727
	12	·277	·355	·443	·540	·647	·762	·884	1·014	1·151	1·293
	11	·625	·763	·912	1·072	1·240	1·417	1·601	1·791	1·987	2·187
	10	1·290	1·512	1·743	1·983	2·229	2·482	2·739	3·000	3·263	3·528
	9	2·455	2·778	3·107	3·440	3·775	4·111	4·446	4·781	5·114	5·444
	8	4·336	4·769	5·200	5·626	6·048	6·463	6·872	7·274	7·669	8·057
	7	7·157	7·689	8·208	8·713	9·204	9·682	10·147	10·599	11·038	11·465
	6	11·107	11·704	12·277	12·827	13·358	13·867	14·357	14·830	15·285	15·724
	5	16·292	16·900	17·477	18·025	18·549	19·048	19·525	19·981	20·417	20·836
	4	22·696	23·250	23·771	24·263	24·730	25·172	25·591	25·990	26·371	26·734
	3	30·168	30·603	31·010	31·391	31·753	32·094	32·416	32·721	33·011	33·287
	2	38·426	38·691	38·938	39·168	39·387	39·593	39·786	39·970	40·143	40·308
	1	47·097	47·164	47·226	47·283	47·340	47·392	47·440	47·486	47·530	47·572
	Actual	8·671	8·473	8·288	8·115	7·952	7·799	7·654	7·517	7·387	7·264
Per cent. occurrence of values differing by *x* or more in excess from Actual.	*1*	44·232	44·363	44·485	44·603	44·708	44·810	44·906	44·997	45·083	45·165
	2	35·955	36·258	36·542	36·812	37·062	37·299	37·525	37·739	37·942	38·135
	3	28·397	28·828	29·235	29·620	29·982	30·326	30·653	30·965	31·262	31·546
	4	21·785	22·290	22·770	23·227	23·660	24·074	24·469	24·847	25·208	25·555
	5	16·230	16·758	17·264	17·748	18·211	18·655	19·083	19·493	19·888	20·269
	6	11·744	12·251	12·740	13·213	13·669	14·110	14·538	14·951	15·351	15·738
	7	8·254	8·709	9·153	9·585	10·007	10·418	10·819	11·210	11·591	11·962
	8	5·637	6·022	6·402	6·777	7·146	7·509	7·866	8·218	8·562	8·901
	9	3·742	4·052	4·362	4·670	4·978	5·284	5·588	5·890	6·188	6·484
	10	2·415	2·654	2·895	3·138	3·385	3·632	3·880	4·129	4·377	4·625
	11	1·517	1·692	1·873	2·057	2·246	2·438	2·633	2·831	3·030	3·230
	12	·927	1·051	1·181	1·315	1·456	1·599	1·747	1·899	2·053	2·210
	13	·552	·637	·727	·821	·921	1·025	1·134	1·247	1·362	1·481
	14	·320	·376	·437	·500	·570	·643	·720	·801	·885	·973
	15	·181	·217	·256	·298	·345	·394	·448	·504	·564	·626
	16	·100	·122	·147	·173	·204	·237	·272	·311	·352	·395
	17	·054	·067	·082	·098	·118	·139	·162	·188	·215	·245
	18	·028	·036	·045	·055	·067	·080	·095	·111	·129	·149
	19	·015	·019	·024	·030	·037	·045	·054	·065	·076	·089
	20	·007	·010	·013	·016	·020	·025	·031	·037	·044	·052
	21	·004	·005	·007	·009	·011	·014	·017	·021	·025	·030
	22	·002	·002	·004	·005	·006	·007	·009	·011	·014	·017
	23	·001	·001	·002	·003	·003	·004	·005	·006	·007	·010
	24	·001	·001	·001	·002	·002	·002	·003	·003	·004	·006
	25	·000	·000	·001	·001	·001	·001	·001	·002	·002	·003
	26	—	—	·000	·000	·000	·001	·001	·001	·001	·002
	27	—	—	—	—	—	·000	·000	·000	·000	·001
	28	—	—	—	—	—	—	—	—	—	·000

PART II. CRITICISMS OF PREVIOUS APPLICATIONS OF POISSON'S LAW OF SMALL NUMBERS.

(7) We now turn to the illustrations which various authors have given of the Law of Small Numbers.

"Student's" Cases. We take first the series given by "Student" in his memoir on counting with a Haemacytometer*. They are of special importance because the series at first appear of fairly adequate size, namely consisting of 400 individuals, and further we should anticipate that the Law of Small Numbers would hold in his cases. He obtains better fits with the binomial than with the exponential but, as he remarks, he has one more constant at his disposal. On the other hand, if the exponential be a true approximation, the binomial ought to come out with a large n and a small but positive q. "Student" finds for his four series:

$$\begin{array}{rl} \text{I.} & 400 \times (1\cdot1893 - \cdot1893)^{-3\cdot6054}. \\ \text{II.} & 400 \times (\cdot97051 + \cdot02949)^{46\cdot2084}. \\ \text{III.} & 400 \times (1\cdot0889 - \cdot0889)^{-20\cdot2473}. \\ \text{IV.} & 400 \times (\cdot9525 \ \ + \cdot0475)^{98\cdot5263}. \end{array}$$

II. and IV. may, perhaps, be held fairly to satisfy the conditions, although it is not certain if 46 is to be considered a large n or ·05 a very small q.

I. and III. fail to satisfy the conditions at all, unless the probable errors of q and n are such that q might really be a small positive quantity and n really large and positive. The following are the values for the four series of n and q and their probable errors:

$$\begin{array}{rll} \text{I.} & q = -\cdot1893 \pm \cdot0647, & n = -\ \ 3\cdot6054 \pm \ \ 1\cdot2209. \\ \text{II.} & q = +\cdot0295 \pm \cdot0457, & n = \ \ \ 46\cdot2084 \pm 71\cdot7373. \\ \text{III.} & q = -\cdot0889 \pm \cdot0534, & n = -20\cdot2473 \pm 12\cdot1165. \\ \text{IV.} & q = +\cdot0475 \pm \cdot0452, & n = \ \ \ 98\cdot5263 \pm 93\cdot7494. \end{array}$$

Now while these results are very satisfactory for II. and IV., they are not wholly conclusive for I. and III. We can approach the matter from another standpoint; the probable error of q for $p = 1$ is

$$\cdot67449 \frac{1}{\sqrt{N}} \sqrt{2} = \cdot67449 \times \cdot0707$$

in "Student's" cases. Thus the deviation of q from q a very small quantity is for I. 2·68 times the S. D., and for III. 1·26 times the S. D. Since q may be either positive or negative, we may reasonably apply the probability tables and the odds against deviations occurring as great as these are in one trial about 250 to 1 and 9 to 1 respectively. Hence in four trials we should still have large odds against their combined appearance.

---

* *Biometrika*, Vol. v. p. 356.

We have said that the results for II. and IV. are fairly satisfactory, *i.e.* we mean that they are consistent with q being small and positive and n being large; but of course they are also consistent with q being negative and n being small and negative.

It will be obvious from these results for "Student's" data that it is extremely difficult to test the legitimacy of the hypothesis on which the "Law of Small Numbers" is based. In none of the cases dealt with by Bortkewitsch, much less in those dealt with by Mortara, are the populations (N) anything like as extensive as those considered by "Student." But populations of even 400 give, as we see, too large values of the probable errors of q and n for us to be certain of our conclusions.

(8) *Bortkewitsch's Cases.* Taking Bortkewitsch next, he deals with the following cases:

I. Suicides of Children in Prussia for 25 years: (*a*) Boys, (*b*) Girls, 25 cases.

II. Suicides of Women in eight German States for 14 years: 112 cases or 8 subseries of 14.

III. Accidental Deaths in 11 Trade Societies in 9 years: 99 cases, or 11 subseries of 9.

IV. Deaths from the Kick of a Horse in 14 Prussian Army Corps for 20 years: 280, or, as Bortkewitsch, 200 cases.

It will be noted at once that Bortkewitsch's populations (N) are far too small for any effective determination of the legitimacy of his application of Poisson's formula to his data.

We take his cases in order:

I. (*a*) *Suicides of Boys.*

TABLE III.

Number of Suicides ...	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7* and over
Number of Years ...	4	8	5	3	4	0	1	0

The binomial is:

$$25\,[1{\cdot}2033 - {\cdot}2033]^{-9{\cdot}6425}.$$

$$\text{Mean } 1{\cdot}9600 \text{ and } \mu_2 = 3{\cdot}2584.$$

We have $q = -{\cdot}2033 \pm {\cdot}2421, \quad n = -9{\cdot}6425 \pm 10{\cdot}9416.$

If q were really zero its probable error would be $\pm{\cdot}1908$. Clearly 25 cases are wholly inadequate to test the legitimacy of applying the Poisson-Exponential to the frequency*. But to what extent is the reader made conscious by Bortkewitsch that his cases fail entirely to demonstrate the legitimacy of applying his hypotheses?

---

* The χ^2 for the binomial is 2·379 and for the exponential 2·836, showing a somewhat better result for the binomial.

I. (*b*) *Suicides of Girls.*

TABLE IV.

Number of Suicides ...	*0*	*1*	*2*	*3*
Number of Years ...	15	9	1	0

The binomial is:

$$25\,[\cdot 7418 + \cdot 2582]^{1\cdot 7041}.$$

$$\text{Mean} = \cdot 4400 \text{ and } \mu_2 = \cdot 3264.$$

We find $$q = \cdot 2582 \pm \cdot 1012, \quad n = 1\cdot 7041 \pm \cdot 7850.$$

As in the case of the boys' suicides, if q were practically zero its probable error would be ± ·1908, and there is nothing in this result again to justify us in asserting that q is indefinitely small and n indefinitely large.

Actually we have:

TABLE V.

Number of Suicides per Year.

	0	*1*	*2*	*3*
Actual	15	9	1	0
Bortkewitsch ...	16·1	7·1	1·8	—
Binomial (*a*) ...	15·0	8·9	1·1	—
Binomial (*b*) ...	15·2	8·7	1·1	—

(*a*) is the binomial considered above, (*b*) is the binomial obtained by taking n a whole number = 2, and $q = \text{mean}/2 = \cdot 22$, *i.e.* $25\,(\cdot 78 + \cdot 22)^2$.

It is clear that either (*a*) or (*b*) gives better results than the Poisson-Exponential. Applying the test of goodness to fit, we have

$$\chi^2 = \cdot 007 \text{ for the binomial } (a),$$

$$\chi^2 = \cdot 610 \text{ for Bortkewitsch's solution.}$$

Both give $P > \cdot 60$ but the first is much better than the second.

If both boys and girls are taken together, we find the binomial

$$25\,(\cdot 9333 + \cdot 0667)^{36}.$$

This is the nearest approach to a small q and big n we have so far found—*i.e.* the nearest approach so far to an exponential, but it is reached by a process, *i.e.* that of adding together two series of entirely different means and variabilities in a manner which cannot be justified, for Bortkewitsch's hypothesis depends essentially on the *homogeneity* of his material. Even here the fit of the point binomial is slightly better than that of the exponential.

II. *Suicides of Women in Eight German States.* Bortkewitsch gives the following table:

TABLE VI.

State	Number of Suicides of Women per Year											Totals
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
(*a*) Schaumburg-Lippe ...	4	4	2	4	—	—	—	—	—	—	—	14
(*b*) Waldeck	1	4	3	4	1	1	—	—	—	—	—	14
(*c*) Lübeck	1	3	2	4	3	1	—	—	—	—	—	14
(*d*) Reuss ä. L.	1	3	3	3	2	1	1	—	—	—	—	14
(*e*) Lippe	2	3	1	2	3	1	2	—	—	—	—	14
(*f*) Schwarzburg-Rudolstadt ...	—	1	—	2	—	5	3	2	1	—	—	14
(*g*) Mecklenburg-Strelitz ...	—	1	2	1	4	—	1	—	2	2	1	14
(*h*) Schwarzburg-Sonderhausen	—	—	4	—	2	2	1	—	—	3	2	14
Totals	9	19	17	20	15	11	8	2	3	5	3	112

The resulting binomials are:

$$(a)\quad 14\,(\ \cdot 9714 + \cdot 0286)^{50\cdot 0024},$$
$$(b)\quad 14\,(\ \cdot 8571 + \cdot 1429)^{15\cdot 4996},$$
$$(c)\quad 14\,(\ \cdot 5819 + \cdot 4181)^{6\cdot 1503},$$
$$(d)\quad 14\,(1\cdot 0058 - \cdot 0058)^{-456\cdot 2944},$$
$$(e)\quad 14\,(1\cdot 3929 - \cdot 3929)^{-7\cdot 2727},$$
$$(f)\quad 14\,(\ \cdot 6071 + \cdot 3929)^{13\cdot 0909},$$
$$(g)\quad 14\,(1\cdot 5792 - \cdot 5792)^{-9\cdot 1267},$$
$$(h)\quad 14\,(1\cdot 6609 - \cdot 6609)^{-8\cdot 5376}.$$

Thus it will be seen that of the eight binomials only four have a positive q, and of these only *one* can be said to have a very small q, and even in this case the n is not indefinitely large. Of the four negative binomials three have quite substantial q's, and the fourth with its small negative q corresponds most closely to the Poisson-Exponential. The probable error of q for $q = 0$ is $\pm \cdot 2549$. The number, 14, of cases taken is therefore wholly inadequate to test whether the Poisson-Exponential may be applied to these data. The mean value of q is negative and $= - \cdot 0820 \pm \cdot 0901$, and the standard deviation of $q = \cdot 3928 \pm \cdot 0637$, which are within the limits of random sampling of $q = 0$ with a standard deviation of ·3779. We shall return to a different manner of considering the point later. At present we wish only to indicate that the hypothesis is that q is a very small positive quantity and that data which give q a standard deviation of ·3928, or in the next example of ·4714 are really inadequate to test such a hypothesis; for in the resulting binomials q may easily lie anywhere between $+ \cdot 8$ and $- \cdot 8$, and it is not possible to demonstrate that its real value is practically an exceeding small positive quantity.

III. *Accidental Deaths in* 11 *Trade Societies.* Bortkewitsch provides data from which the following table is deduced:

TABLE VII.

Index Number of Society	Accidental Deaths															Totals
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	
13	—	—	—	—	1	1	1	1	3	1	—	—	—	—	1	9
14	—	2	3	2	1	1	—	—	—	—	—	—	—	—	—	9
12	2	1	3	—	1	1	—	1	—	—	—	—	—	—	—	9
20	—	—	1	3	2	2	—	—	—	—	1	—	—	—	—	9
23	—	—	—	1	2	1	2	—	1	1	—	1	—	—	—	9
27	—	4	3	1	1	—	—	—	—	—	—	—	—	—	—	9
29	—	—	—	2	3	—	—	1	2	—	—	—	1	—	—	9
41	—	1	—	1	1	2	1	2	1	—	—	—	—	—	—	9
40	2	1	2	1	—	1	1	1	—	—	—	—	—	—	—	9
42	1	—	—	1	1	4	1	—	1	—	—	—	—	—	—	9
55	—	—	2	1	1	3	1	1	—	—	—	—	—	—	—	9
Totals ...	5	9	14	13	14	16	7	7	8	2	1	1	1	—	1	99

The resulting binomials are:

$$(13)\quad 9\,(\cdot 4914+\cdot 5086)^{15\cdot 5108},$$

$$(14)\quad 9\,(\cdot 6184+\cdot 3816)^{6\cdot 6962},$$

$$(12)\quad 9\,(1\cdot 9227-\cdot 9227)^{-2\cdot 7696},$$

$$(20)\quad 9\,(1\cdot 1282-\cdot 1282)^{-33\cdot 8000},$$

$$(23)\quad 9\,(\cdot 9921+\cdot 0079)^{784\cdot 0502},$$

$$(27)\quad 9\,(\cdot 5229+\cdot 4771)^{3\cdot 9589},$$

$$(29)\quad 9\,(1\cdot 4130-\cdot 4130)^{-14\cdot 2589},$$

$$(41)\quad 9\,(\cdot 8454+\cdot 1546)^{33\cdot 0626},$$

$$(40)\quad 9\,(2\cdot 0342-1\cdot 0342)^{-2\cdot 7934},$$

$$(42)\quad 9\,(\cdot 9322+\cdot 0678)^{67\cdot 2397},$$

$$(55)\quad 9\,(\cdot 6154+\cdot 3846)^{11\cdot 2667}.$$

Of these eleven binomials seven have a positive q; only one of these (23) actually corresponds to a really small q and large n, although a second, (42), approximates to this condition. In the five other cases the q's are quite substantial; in (13) the q is larger than p. Of the four negative q's none can be said to be so small and the n so large as to suggest that they really correspond to the Poisson-Exponential. The probable error of q for $q=0$ is, however, $\pm\cdot 3180$, and thus for such small series, no test whatever can be really reached of the legitimacy of applying the Poisson-Exponential to such data. We may note, indeed, that seven of the eleven values of q exceed the probable error and two of these are more than three times the probable error. We should only expect *two* negative values of q as great or greater than $\cdot 9227$ in 80 trials, whereas two have occurred in 9 trials, so that the odds are considerably against such an experience. The mean value of q is $-{\cdot}0469 \pm {\cdot}0959$ and the standard deviation of q is $\cdot5127 \pm \cdot0678$, both results compatible with q indefinitely small and a standard deviation $= \cdot4714$. The main problem, however, of the legitimacy of applying the Poisson-Exponential to such series cannot be answered by data involving only total frequencies of 9 to 14 cases in the individual series.

Bortkewitsch examines the matter from another standpoint. He clubs the results given for each application of the Poisson-Exponential together and examines the observed totals against the sums of the calculated totals. Thus calculating the 11 Poisson-Exponential series* and adding them together Bortkewitsch finds for observed and calculated deaths:

TABLE VIII.

Accidental Deaths in 11 *Trade-Societies.*

Number of Deaths	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13* & over	Totals
Observed Frequencies	5	9	14	13	14	16	7	7	8	2	1	1	1	1	99
Sums of 11 Exponentials	3·7	9·6	13·9	15·2	14·3	12·3	9·8	7·3	5·8	3·3	2·0	1·2	0·7	0·6	99
Single Binomial ...	3·8	9·5	13·9	15·6	14·8	12·4	9·6	6·9	4·8	3·1	2·0	1·2	0·7	0·7	99

If we attempt to fit a *single* binomial to the observed line of totals, we obtain:

$$m = 4\cdot3636, \quad \sigma^2 = 7\cdot5849$$

leading to the negative binomial:

$$99\,(1\cdot7382 - \cdot7382)^{-5\cdot9111}.$$

Here: $q = -\cdot7382 \pm \cdot1829$†, $n = -5\cdot9111 \pm \cdot1391$,

or the constants are significantly substantial with regard to their probable errors. The resulting frequencies are given in the last line of the table above. The reader

---

* The values of the means and standard deviations for the eleven societies are:

	m	σ		m	σ		m	σ
13	7·889	1·969	*23*	6·222	2·485	*40*	2·889	2·424
14	2·556	1·343	*27*	1·889	0·994	*42*	4·556	2·061
12	2·556	2·217	*29*	5·889	2·885	*55*	4·333	1·633
20	4·333	2·211	*41*	5·111	2·079			

All these means are less than 10, which is the limit reached by Bortkewitsch's Tables for the Poisson-Exponential. Bortkewitsch says he has taken the societies for which "the statistics indicated the smallest numbers of such accidents." This is not very clear. It is certain that a society with a mean number of accidents $= 100$, if it consisted of 200,000 members, would be more suitable for application of the exponential, than one with a mean of 8 if it only contained 10,000 members. Both Bortkewitsch and Mortara confine their results to means less than 10, and seem to indicate that "smallness" has been determined by the absolute frequencies, but clearly it is relative frequency with which we have to deal. The use of such a term as *Das Gesetz der kleinen Zahlen* for the Poisson-Exponential seems open to serious objection, if it be associated with "m" an absolutely small number, and not with smallness of "q."

† For $q = 0$, the probable error would be $\pm\cdot0959$ and accordingly q is very divergent from the Poisson-Exponential value of zero.

will be surprised to see how closely the single negative binomial determined by *two* constants gives the same result as the sum of the eleven Poisson-Exponentials determined by *eleven* constants, no one of which is really of any significance for its own exponential*. If we apply the condition for "goodness of fit," $\chi^2 = 5{\cdot}83$ for the single binomial and $\chi^2 = 5{\cdot}88$ for the sum of the eleven Poisson exponentials, leading to $P = {\cdot}950$ and $P = {\cdot}951$ respectively, or the fit with a single negative binomial is slightly better than that with eleven exponentials. The two constants are significant, the eleven constants have no real significance for their individual series, as is demonstrated by the fact that the binomials for these series do not approximate to the Poisson-Exponential type.

We may now consider the previous case of suicides of women from the same standpoint†. The following are the data as given by Bortkewitsch:

TABLE IX.

Suicides of Women in Eight German States.

Number of Suicides	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10* & over	Totals
Observed Frequencies	9	19	17	20	15	11	8	2	3	5	3	112
Sum of 8 Exponentials	8·0	16·9	20·3	18·7	15·1	11·4	8·3	5·6	3·6	2·1	2·0	112
Single Binomial ...	12·6	18·4	18·8	16·4	13·2	9·9	7·2	5·1	3·5	2·4	4·5	112

For the single binomial we have:

$$m = 3{\cdot}4732, \quad \sigma^2 = 8{\cdot}2312,$$

leading to:

$$112\,(2{\cdot}3699 - 1{\cdot}3699)^{-2{\cdot}5354},$$

where

$$q = -1{\cdot}3699 \pm {\cdot}1490, \quad n = -2{\cdot}5354 \pm {\cdot}3076.$$

If q were very small its probable error would be $\pm {\cdot}0901$. The values of q and n are quite significant, q is large and negative and n is small and negative. The resulting frequencies are given in the last line of the table as "Single Binomial." Turning now to the test of "goodness of fit," we have for the sum of the 8 exponentials $\chi^2 = 7{\cdot}957$, and for the single binomial $\chi^2 = 7{\cdot}740$, leading to $P = {\cdot}633$

---

* If the reader will turn to the first footnote on p. 53 he will note that for *nine* cases, the standard deviations of the means $(\sigma/\sqrt{9})$ are roughly about ·7 or errors of ±1 to ±1·5 may easily occur in the means. Hence with the possible exception of (13) and (27) the m's have not significant differences, and are not typical of the individual societies.

† The values of the means and standard deviations are:

	m	σ		m	σ
Schaumburg-Lippe	1·429	1·178	Lippe	2·857	1·995
Waldeck ...	2·214	1·378	Schwarzburg-Rudolstadt ...	5·143	1·767
Lübeck	2·571	1·223	Mecklenburg-Strelitz ...	5·286	2·889
Reuss ä. L. ...	2·643	1·631	Schwarzburg-Sonderhausen	5·642	3·061

The standard deviation of the mean is here $\sigma/\sqrt{14}$, or, say, ·5. Thus errors of 1 might easily occur in the values of m. There are probably significant differences between the first five and the last three states, but not between the first five among themselves or the last three among themselves. Thus the Poisson-Exponentials, if correct in theory, are not significant for the individual states.

and ·654 respectively. Thus again the single binomial with only two constants give a fit slightly better, than the sum of eight exponentials with eight constants.

Bortkewitsch looking at the observed frequencies and the sum of 8 or 11 exponentials—without using any satisfactory test for "goodness of fit"—assumes that the coincidence is so good as to justify his hypothesis. But a better fit can be obtained with two instead of 8 or 11 constants by simply using a negative binomial. We must note here that Bortkewitsch is using the final coincidence merely as justification of the Poisson-Exponential; the total frequency is not describable in terms of the 8 or 11 constants as it is in terms of the two, for these eight constants are not really significant for his individual eleven trade societies or for the suicides in the individual eight states. If he wants to describe the total, he has no constants by which he can do it. If, on the other hand, he wishes to describe what has occurred in the individual societies or states, we have seen that their binomials differ very widely from Poisson-Exponentials. If, lastly, no stress be laid on the individual cases as having too large probable errors, but only on the general coincidence with total frequencies, then the same coincidence would justify us in using a single binomial with two constants only*. It appears to us that to properly test the Poisson-Exponential, we need not 9 or 14 instances in the individual case, but several hundred instances,—more, indeed, than "Student" has taken—and that no proof of the "Law of Small Numbers" can be obtained on data such as those of Bortkewitsch or Mortara.

IV. *Deaths from the Kick of a Horse in Prussian Army Corps, omitting four Corps with Bortkewitsch.*

Here the results are:

TABLE X.

Number of Deaths ...	*0*	*1*	*2*	*3*	*4*	Totals
Number of Corps ...	109	65	22	3	1	200

Whence:

$$m = \cdot 61, \quad \mu_2 = \cdot 6079$$

and the binomial is:

$$200\,(\cdot 996{,}557 + \cdot 003{,}443)^{177 \cdot 17107}.$$

This is the first of Bortkewitsch's illustrations for which his hypothesis that q is small and n large is really justified by his data. For:

$$q = \cdot 0034 \pm \cdot 0670,$$
$$n = 177\cdot 1711 \pm 3449\cdot 103.$$

The probable error of q for q really zero is $\pm \cdot 0674$.

---

* Of course immensely better general total fits are obtained by using the sums of the actual 8 or 11 binomials than by the Poisson-Exponential sum or the single binomial, but the results in that case involve 16 or 22 non-significant constants.

The actual results as given by the binomial and the Poisson-Exponential are:

TABLE XI.

Number of Deaths ...	*0*	*1*	*2*	*3*	*4* and over
Observed	109	65	22	3	1
Binomial	108·6	66·4	20·2	4·1	0·7
Exponential	108·7	66·3	20·2	4·1	0·7

Actually if we work to two decimal places in the frequencies we have $\chi^2 = \cdot 61$ for both binomial and exponential, or the goodness of fit is practically identical.

In this case it seemed worth discussing the binomial fit more at length. Taking the moment coefficients about the mean we have:

(i) Mean $= nq = \cdot 6100$.

(ii) $\mu_2 = npq = \cdot 6079$.

(iii) $\mu_3 = npq\,(p - q) = \cdot 590{,}562$.

(iv) $\mu_4 = npq\,(1 + 3npq - 6pq) = 1\cdot 643{,}373$.

We have already discussed the binomial from (i) and (ii), giving χ^2 for goodness of fit $= \cdot 6096$. Using (ii) and (iii) we have for the binomial

$$200\,(\cdot 985{,}739 + \cdot 014{,}261)^{43\cdot 24351},$$

giving

$$\chi^2 = \cdot 665.$$

Using (iii) and (iv) we have:

$$200\,(\cdot 979{,}524 + \cdot 020{,}057)^{30\cdot 3090},$$

giving

$$\chi^2 = \cdot 707.$$

Putting:

$$\beta_2 = \mu_4/\mu_2^2 \text{ and } \beta_1 = \mu_3^2/\mu_2^3,$$

we have:

$$\beta_2 - 3 = (1 - 6pq)/npq, \quad \beta_1 = (1 - 4pq)/npq,$$

and working from β_1 and β_2 we find:

$$200\,(\cdot 969{,}150 + \cdot 030{,}850)^{18\cdot 9645},$$

and in this case

$$\chi^2 = 1\cdot 1286.$$

This of course does not give a bad fit, but it is clear that working from the *lowest moment coefficients*, as we might anticipate, gives the best results.

But if q be the chance of death from the kick of a horse, and n the number of men in an army corps, then the binomial should be

$$200\,(p + q)^n.$$

Now it is obvious that none of the binomials give, by their value of n any approach to the real number of men in an army corps. If we start with the number of men n in an army corps as 50,000*, we have $nq = \cdot 61$ and $q = \cdot 000{,}0122$, thus reaching the binomial

$$200\,(\cdot 999{,}9878 + \cdot 000{,}0122)^{50{,}000},$$

giving as compared against Bortkewitsch:

	Binomial	*Bortkewitsch*
0	108·6876	108·6703
1	66·3002	66·2889
2	20·2213	20·2181
3	4·1115	4·1110
4 and over	·7035	·7034
and $\chi^2 =$	·608,298	·608,318

or, the slight advantage to the binomial exists but is of no significance.

Now it seems to us that in this case the use of the exponential is justified for the total frequencies, but as far as describing those frequencies is concerned, it gives no better result than the binomial. But as in the other five of Bortkewitsch's cases the Exponential is not justified by the individual series themselves†.

It is perfectly true that the exponential has a definite theory behind it, and is interpretable in terms of that theory, i.e. we must suppose the probability of an occurrence very small and the chance of its repetition absolutely identical. But is the second of these conditions ever likely to be demonstrable *a priori*, or must

---

* This supposes that every man in the army corps is equally liable to death from the kick of a horse; of course a very arbitrary assumption.

† To illustrate the idleness of the application of the Poisson-Exponential even to these data for the Prussian Army Corps, we give here the binomials for the whole of the 14 corps.

Index Number of Corps	*Binomial*
G	$20\,(\cdot 95 + \cdot 05)^{16\cdot 0000}$
I	$20\,(1\cdot 325 - \cdot 325)^{-2\cdot 4615}$
II	$20\,(1\cdot 5667 - \cdot 5667)^{-1\cdot 0588}$
III	$20\,(\cdot 9 + \cdot 1)^{6\cdot 0000}$
IV	$20\,(\cdot 6 + \cdot 4)^{1\cdot 0000}$
V	$20\,(\cdot 6318 + \cdot 3682)^{1\cdot 4938}$
VI	$20\,(1\cdot 0912 - \cdot 0912)^{-9\cdot 3202}$
VII	$20\,(\cdot 9 + \cdot 1)^{6\cdot 0000}$
VIII	$20\,(\cdot 65 + \cdot 35)^{1\cdot 0000}$
IX	$20\,(\cdot 8115 + \cdot 1885)^{3\cdot 4483}$
X	$20\,(1\cdot 05 - \cdot 05)^{-15\cdot 0000}$
XI	$20\,(1\cdot 11 - \cdot 11)^{-11\cdot 3636}$
XIV	$20\,(1\cdot 05 - \cdot 05)^{-24\cdot 0000}$
XV	$20\,(1\cdot 1 - \cdot 1)^{-4\cdot 0000}$

One seeks in vain through these binomials for any approach to q very small and positive and n very large and positive. In no case does n approach the number of men in an army corps, say 50,000, or q equal the chance of a death from the kick of a horse, say, ·0000122! It seems impossible by clubbing such equations together to give any satisfactory proof that the Poisson-Exponential really does apply to individual cases. In the 20 years involved, there were doubtless great changes in both the training and the personnel of each army corps, and the results obtained may be just as much due to such causes as to the errors of small samples.

not we *a posteriori* demonstrate it from the data themselves? Child suicide may be influenced by example, by environmental conditions in different districts, possibly even by meteorological conditions in different years. Again, even in different army corps the conditions may be far from uniform, the spirit of the corps, the teaching with regard to the handling of horses, the experience of past life according to whether the corps is raised in town or rural districts may all tell. Even Bortkewitsch before he gets his best fit removes four corps or 80 observations from his data. We do not criticise this removal, but even unremoved he says the fit of theory with experience leaves "wie man sieht, nichts zu wünschen übrig" (p. 25). But the binomial is before removal:

$$280\,(1\cdot085{,}714 - \cdot085{,}714)^{-8\cdot166{,}694}$$

in which q is not very small and is negative, and n is not very large and is not positive. It is true that the probable error of q for q insignificant is in this case $\pm\cdot0570$, but this only shows that the data were insufficient in quantity to determine whether the exponential could be applied or not.

(9) *Mortara's Cases.*

Mortara* in an interesting paper has realised the possibility of repetitions not being independent and has discussed a constant Q', by which he proposes to test such influence. This quantity Q' should be unity, if the Bortkewitschian hypothesis can be applied. He then takes 16 or 17 districts with records of 10 years, and calculates the mean number of deaths from some special cause per year, say, for each district for those years. If this mean number exceeds 10, he casts out that district, presumably on the ground either (i) that such a number is no longer small, or (ii) that it differentiates the district from those with lower numbers. Thus Bologna with 10·9 deaths by murder is excluded and Bergamo with 8·4 is included, although $Q' = 1$ for both. Bologna with 7·1 deaths from smallpox is included, but Pavia with 12·3 is excluded although the Q' of the former is 2·5 and that of the latter 1·7. What method should be employed in dealing with the frequency of the excluded districts which may amount to 50 °/₀ of all districts is not discussed. Having thus reduced his available districts, Mortara proceeds to apply the exponential to each *individual* district; he adds up the results for each district and compares his totals with the observed totals. It will thus be observed that he fits his exponential to *ten* observations, and then adds together five or more districts to get his totals. We can equally well apply this process by fitting a binomial to each 10 observations and then adding up such results. But it is quite clear that on the basis of *ten* observations, it is, owing to the large probable errors, wholly impossible to assert, whether a binomial of the kind required by the Bortkewitsch-Mortara hypothesis,—i.e. one of very small positive q and very large positive n—really is justified. We can illustrate this at once from Mortara's Tables (see his pp. 42 and 45) for deaths from Chronic Alcoholism. The

---

* "Sulle variazioni di frequenza di alcuni fenomeni demografici rari," *Annali di Statistica*, Serie v. Vol. IV. pp. 5—81. Roma, 1912.

observed numbers, and those deduced from the binomials are given in the accompanying table. At the foot are the observed totals, Mortara's exponential totals and the binomial totals.

TABLE XII. *Deaths from Chronic Alcoholism.*

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14* & over	
Calabria	1	3	4	—	2	—	—	—	—	—	—	—	—	—	—	Observed
	1·49	2·84	2·70	1·71	·81	·31	·10	·03	·01	—	—	—	—	—	—	Mortara
	1·18	2·85	3·06	1·91	·76	·20	·03	—	—	—	—	—	—	—	—	Binomial
Foggia	1	2	4	—	2	1	—	—	—	—	—	—	—	—	—	O.
	1·00	2·30	2·65	2·03	1·17	·54	·21	·07	·02	·01	—	—	—	—	—	M.
	·96	2·29	2·70	2·08	1·18	·53	·19	·06	·01	—	—	—	—	—	—	B.
Siracusa	2	1	3	—	2	2	—	—	—	—	—	—	—	—	—	O.
	·82	2·05	2·56	2·14	1·34	·67	·28	·10	·03	·01	—	—	—	—	—	M.
	1·12	2·16	2·33	1·85	1·21	·69	·35	·17	·07	·03	·01	—	—	—	—	B.
Potenza	2	—	2	2	1	1	1	1	—	—	—	—	—	—	—	O.
	·41	1·30	2·09	2·23	1·78	1·14	·61	·28	·11	·04	·01	—	—	—	—	M.
	·78	1·61	1·95	1·80	1·41	·98	·63	·38	·21	·12	·06	·03	·01	·01	—	B.
Catanzaro	1	1	3	1	—	—	1	1	1	—	—	1	—	—	—	O.
	·15	·63	1·32	1·85	1·95	1·63	1·14	·69	·36	·17	·07	·03	·01	—	—	M.
	·88	1·35	1·46	1·36	1·17	·95	·75	·57	·43	·31	·23	·16	·12	·08	·17	B.
Salerno	1	1	1	—	2	—	1	1	2	—	—	1	—	—	—	O.
	·06	·31	·79	1·35	1·72	1·75	1·49	1·09	·69	·39	·20	·09	·04	·02	·01	M.
	·40	·86	1·18	1·31	1·27	1·14	·95	·77	·59	·45	·33	·24	·17	·12	·22	B.
Cosenza	2	—	1	—	1	—	3	1	—	1	—	—	1	—	—	O.
	·06	·29	·75	1·29	1·68	1·75	1·51	1·12	·73	·42	·22	·10	·05	·02	·01	M.
	·43	·88	1·17	1·27	1·23	1·10	·93	·75	·59	·44	·33	·24	·17	·13	·33	B.
Bologna	—	—	3	1	1	1	1	—	—	—	1	—	—	—	2	O.
	·01	·06	·21	·49	·88	1·24	1·47	1·49	1·32	1·04	·74	·48	·28	·15	·14	M.
	·40	·76	·97	1·06	1·05	·98	·87	·76	·64	·53	·43	·35	·28	·21	·71	B.
Totals	10	8	21	4	11	5	7	4	3	1	1	2	1	—	2	O.
	4·00	9·78	13·07	13·09	11·33	9·03	6·81	4·87	3·27	2·08	1·24	·70	·38	·19	·16	M.
	6·15	12·75	14·82	12·64	9·28	6·57	4·70	3·46	2·54	1·88	1·38	1·02	·75	·55	1·43	B.

The following are the binomials for the 8 districts out of 16 which Mortara has selected.

Reggio Calabria	$10\,({\cdot}7842 + {\cdot}2158)^{+8{\cdot}8049}$
Foggia	$10\,({\cdot}9609 + {\cdot}0391)^{58{\cdot}7769}$
Siracusa	$10\,(1{\cdot}3000 - {\cdot}3000)^{-8{\cdot}3333}$
Potenza	$10\,(1{\cdot}5500 - {\cdot}5500)^{-5{\cdot}8182}$
Catanzaro	$10\,(2{\cdot}7524 - 1{\cdot}7524)^{-2{\cdot}3967}$
Salerno	$10\,(2{\cdot}3510 - 1{\cdot}3510)^{-3{\cdot}7750}$
Cosenza	$10\,(2{\cdot}5308 - 1{\cdot}5308)^{-3{\cdot}3970}$
Bologna	$10\,(3{\cdot}3161 - 2{\cdot}3161)^{-2{\cdot}6769}$

Examining these we see that there are only *two* in which q and n are positive and only *one* in q is small and positive and n moderately large. The probable error of q for 10 observations on the assumption that n is very large and q very small is ± ·3016 and is quite inconsistent with the last four districts being samples from exponentially distributed frequencies. The other four districts may or may not belong to such frequencies—the data are wholly inadequate to determine whether they do or not. Reggio Calabria and Foggia have the lowest Q's, i.e. 0·9 and 1·0. But that six districts out of an *already selected* eight give *negative* q and a seventh a relative large q and small n suggests the inapplicability of the hypothesis adopted. If we seek for "goodness of fit" of the totals, we find:

	Binomial	*Exponential*
$\chi^2 =$	25·12	47·92
$P =$	·0336	·0000

Thus the odds against the binomial system are 28 to 1, but the odds against the exponential are enormous. It does not seem possible to justify the treatment of such data by the use of the Poisson-Exponential.

Let us turn to a second of Mortara's illustrations, that of deaths from small-pox. He rejects first six out of the 17 districts, the remaining ten are given in Table XIII. The districts give the following binomials:

Venezia	$10(\cdot 9500 + \cdot 0500)^{16}$
Bologna	$10(\cdot 9889 + \cdot 0111)^{81}$
Treviso	$10(2\cdot 2000 - 1\cdot 2000)^{-\cdot 8333}$
Pavia	$10(1\cdot 8000 - \cdot 8000)^{-1\cdot 5000}$
Cagliari	$10(4\cdot 5190 - 3\cdot 5190)^{-\cdot 5968}$
Padova	$10(3\cdot 6833 - 2\cdot 6833)^{-\cdot 8944}$
Verona	$10(5\cdot 6000 - 4\cdot 6000)^{-\cdot 5217}$
Brescia	$10(9\cdot 9727 - 8\cdot 9727)^{-\cdot 3678}$
Bergamo	$10(2\cdot 3821 - 1\cdot 3821)^{-2\cdot 8219}$
Catanzaro	$10(15\cdot 6128 - 14\cdot 6128)^{-\cdot 2669}$
Vicenza	$10(3\cdot 4854 - 2\cdot 4854)^{-1\cdot 6497}$

Out of the eleven cases only two give q small and positive; not *a single one* gives for q anything like the chance of a death from small-pox in the district, nor for n anything like the population of the district. There is an increasing divergence from the positive binomial as Mortara's Q' increases in value. We see that in nine cases, however, a negative binomial not the exponential is required to describe the frequencies. The probable error of q, for insignificant q is as before ± ·3016, and therefore it is improbable that q is zero in at least 9 out of these 11 districts.

Examining the totals we find

	Binomial	*Exponential*
$\chi^2 =$	9·64	570·79
$P =$	·67	·000,000

TABLE XIII.

Deaths from Small-pox (1900—1909).

	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12* or more	
Venezia	4	5	—	1	—	—	—	—	—	—	—	—	—	Observed
	4·49	3·60	1·44	·38	·08	·01	—	—	—	—	—	—	—	Mortara
	4·40	3·71	1·46	·36	·06	·01	—	—	—	—	—	—	—	Binomial
Bologna	4	4	1	1	—	—	—	—	—	—	—	—	—	O.
	4·07	3·66	1·65	·49	·11	·02	—	—	—	—	—	—	—	M.
	4·04	3·68	1·65	·49	·11	·02	·01	—	—	—	—	—	—	B.
Treviso	5	3	1	—	—	1	—	—	—	—	—	—	—	O.
	3·68	3·68	1·84	·61	·15	·03	·01	—	—	—	—	—	—	M.
	5·18	2·36	1·18	·61	·32	·17	·09	·05	·03	·01	—	—	—	B.
Pavia	4	3	2	—	—	1	—	—	—	—	—	—	—	O.
	3·01	3·62	2·17	·87	·26	·06	·01	—	—	—	—	—	—	M.
	4·14	2·76	1·53	·79	·40	·19	·09	·05	·02	·01	·01	—	—	B.
Cagliari	5	1	1	1	—	1	—	—	—	—	1	—	—	O.
	1·23	2·57	2·70	1·89	·99	·42	·15	·04	·01	—	—	—	—	M.
	4·07	1·89	1·17	·79	·55	·39	·28	·21	·15	·11	·08	·06	·25	B.
Padova	3	3	—	2	—	1	—	—	—	—	1	—	—	O.
	·91	2·18	2·61	2·09	1·25	·60	·24	·08	·03	·01	—	—	—	M.
	3·12	2·03	1·40	·98	·70	·50	·36	·26	·19	·13	·10	·07	·16	B.
Verona	4	3	—	1	—	—	1	—	—	—	—	—	1	O.
	·91	2·18	2·61	2·09	1·25	·60	·24	·08	·03	·01	—	—	—	M.
	4·07	1·74	1·09	·75	·54	·40	·31	·23	·18	·14	·11	·09	·35	B.
Brescia	2	3	2	2	—	—	—	—	—	—	—	—	1*	O.
	·37	1·22	2·01	2·21	1·82	1·20	·66	·31	·13	·05	·02	—	—	M.
	4·29	1·42	·87	·62	·47	·37	·30	·24	·20	·17	·14	·12	·79	B.
Bergamo	2	—	2	2	—	1	—	1	1	1	—	—	—	O.
	·20	·79	1·54	2·00	1·95	1·52	·99	·55	·27	·12	·04	·02	·01	M.
	·86	1·41	1·57	1·46	1·23	·98	·74	·54	·38	·27	·18	·12	·24	B.
Catanzaro	3	3	1	1	1	—	—	—	—	—	—	—	1*	O.
	·20	·79	1·54	2·00	1·95	1·52	·99	·55	·27	·12	·04	·02	·01	M.
	4·80	1·20	·71	·50	·38	·31	·25	·21	·18	·16	·14	·12	1·04	B.
Vicenza	3	—	1	1	1	1	—	1	1	—	—	—	1	O.
	·17	·68	1·39	1·91	1·95	1·60	1·09	·64	·33	·15	·06	·02	·01	M.
	1·28	1·50	1·42	1·23	1·02	·82	·65	·51	·39	·30	·23	·17	·48	B.
Totals	39	28	11	12	2	6	1	2	2	1	2	—	4	O.
	19·24	24·97	21·50	16·54	11·76	7·58	4·38	2·25	1·07	·46	·16	·06	·03	M.
	40·25	23·70	14·05	8·58	5·78	4·16	3·08	2·30	1·72	1·20	·99	·75	3·31	B.

* 1 at '12 or more' in cases of Brescia and Catanzaro was found to signify 1 at 20 in the case of Brescia, and 1 at 27 in case of Catanzaro, if the means were to agree with those given by Mortara.

In other words the binomials give a reasonable total fit, the exponentials a practically impossible one.

But there is another question to be asked in such series as those of Mortara: What justification is there in cutting off at 10 cases, say of murder? A province may have a million inhabitants and, perhaps, 40 murders occur in a year*. Hence the binomial is for ten year returns

$$10 \times \left(\frac{24{,}999}{25{,}000} + \frac{1}{25{,}000}\right)^{1{,}000{,}000}$$

but this is as close as anything can be desired to the exponential series. It may be reasonable to apply a separate series to districts giving 4·2 and 36·6 murders per annum respectively, but it is difficult to see why the latter district should be altogether excluded from treatment. If the theory of the binomial be applicable *at all*, then it applies practically as well to districts with 40 murders as to districts with 4; for, we need no indefinitely small q to get a closely exponential series. If we take the case of deaths by murder, Mortara has retained only 6 out of 16 provinces, yet his criterion Q' (see his Table, p. 51) is not more divergent from unity for the rejected provinces than for those retained; the binomials are indeed

Reggio Treviso	$10(\cdot 7000 + \cdot 3000)^{9\cdot 3333}$
Venezia	$10(\cdot 5619 + \cdot 4381)^{9\cdot 5869}$
Vicenza	$10(\cdot 9571 + \cdot 0429)^{114\cdot 2191}$
Padova	$10(\cdot 4774 + \cdot 5226)^{11\cdot 8638}$
Pavia	$10(1\cdot 8162 - \cdot 8162)^{-9\cdot 0664}$
Bergamo	$10(\cdot 8857 + \cdot 1143)^{73\cdot 4908}$

only one of which gives q small and positive and n large.

The mean Q' for the retained provinces is ·967 with a range from ·7 to 1·4 and for the rejected 1·03 with a range from ·8 to 1·4. Even if—which is not the case—the probability of an individual being murdered were too great for the exponential, it ought to follow the binomial, but this, as a rule, it does not do, unless we give some wholly new interpretations to q and n; the actual values render the theory of the binomial as stated inapplicable.

(10) *Mortara's Criterion.*

As a matter of fact the only test of whether an exponential will legitimately fit a given series or not is to determine the binomial $(p+q)^n$ and ascertain whether p is slightly less than unity. But:

$$\begin{aligned} p &= npq/nq \\ &= \frac{(\text{Standard Deviation})^2}{\text{Mean}}. \end{aligned}$$

---

* We assume that each individual is equally likely to be murdered. But if there be a graduated probability for murder throughout the community, what right have we to apply Poisson's series at all? The essential basis of the application—equal chance of each individual—is wanting.

Now if m_s be the number of deaths, say, occurring in any year and there be l years under consideration, then:

$$(\text{Standard Deviation})^2 = \frac{S_1^2 (m_s - nq)^2}{l},$$

or, if we use the form preferred by Bortkewitsch*

$$= \frac{S_1^2 (m_s - nq)^2}{l - 1}.$$

Hence: $$p = \frac{S_1^2 (m_s - nq)}{(l-1)\, nq}.$$

This in other notation is Mortara's Q'^2, the only criterion he actually uses provided by his equation (17 *ter*), p. 18. Thus his Q', which he says must not differ much from 1, is only $\sqrt{p}$, and it would be better to use p—which has a direct physical meaning—than Mortara's $Q' = \sqrt{p}$. Clearly Mortara's somewhat elaborate process of deducing Q', does not amount to more than saying: Fit a point binomial and test if p is slightly less than unity. We contend that it is best straight off to fit the binomial.

It is true that Mortara does not reach his Q'^2, our p, by the simple process of asking whether the binomial is one with a positive probability less than unity. He endeavours to obtain it by considering whether there is "lumpiness" in the observations. But it seems to us clearer and briefer to ask: Are the contributory cause-groups independent as in teetotum spinning? If so, the data will fit a true binomial and p will of necessity be a positive quantity less than unity. If they are not of this character then p must of necessity be greater than unity. It is of interest to see how Mortara's test of dependence of contributory cause groups leads to a criterion, but he actually only gets his Q'^2, i.e. our binomial p after a series of hypotheses which much limit, and that in no very obvious manner,

---

* The use of $\sqrt{l}$ or $\sqrt{l-1}$ in the value of the standard deviation when l is small has been several times discussed. It may be dealt with as follows: The probable errors of a mean as deduced by the two processes are

$$E = \cdot 67449 \,.\, \sigma/\sqrt{l},$$

and $$E' = \cdot 67449 \,.\, \sigma/\sqrt{l-1},$$

now $$E' = \cdot 67449\, \sigma/\sqrt{l}\left(1 + \frac{1}{2l} + \dots\right)$$

$$= \cdot 67449 \frac{1}{\sqrt{l}}\left(\sigma + \frac{1}{\sqrt{2l}}\frac{\sigma}{\sqrt{2l}} + \dots\right).$$

Now the probable error of σ is $\cdot 67449 \dfrac{\sigma}{\sqrt{2l}}$, and $\dfrac{1}{\sqrt{2l}}$ is less and often much less than ·67449. Hence if we only know σ from the observations themselves, and this is the usual case, we have:

$$E' = \cdot 67449 \frac{1}{\sqrt{l}}\, \sigma',$$

where σ' differ from σ by a quantity usually far less than the probable error of σ. In other words the refinement of using E' for E is idle having regard to the accuracy of our observations; and the form used by Bortkewitsch and Mortara with $\sqrt{l-1}$ for $\sqrt{l}$ is of no importance.

the nature of those contributory causes groups. Of course if their dependence were of the nature of successive draws from a pack, then the result would be a hypergeometrical series and Q'^2 would have no physical meaning for the series at all.

(11) We will deal with one further illustration out of many considered by Mortara which are of like character. In the case of Marriages of Uncle and Niece (see Table XIV, p. 65), where the distribution of Q's is the most favourable for his theory, the binomials are

Reggio Marche	$10(\cdot 7000 + \cdot 3000)^{1\cdot 0}$
Umbria	$10(\cdot 9000 + \cdot 1000)^{5\cdot 0}$
Basilicata	$10(1\cdot 4000 - \cdot 4000)^{-1\cdot 5}$
Sardegna	$10(\cdot 44545 + \cdot 55455)^{1\cdot 9836}$
Emilia	$10(\cdot 9818 + \cdot 0182)^{120\cdot 1000}$
Abruzzi	$10(\cdot 8429 + \cdot 1571)^{17\cdot 8182}$
Lazio	$10(1\cdot 2548 - \cdot 2548)^{-12\cdot 1646}$
Puglie	$10(1\cdot 5111 - \cdot 5111)^{-7\cdot 0435}$
Veneto	$10(1\cdot 3444 - \cdot 3444)^{-1\cdot 30645}$
Toscana	$10(2\cdot 2667 - 1\cdot 2667)^{-4\cdot 26315}$
Calabria	$10(1\cdot 3584 - \cdot 3584)^{-24\cdot 8305}$

of which only one (Emilia) approaches the conditions for an exponential distribution. If we test the totals at the foot of Table XIV, we find the result much to the advantage of the binomial, for which $P = \cdot 902$ as against $\cdot 714$ for the exponential.

(12) On Mortara's own showing nearly all the Q's of his numerous series are greater than unity, and very few of the binomials are positive. If we consider the distribution of Q's, given in his work omitting Table 13 (Deaths from Malaria) we find a range from ·5 to 3·6 with a mean Q' at

$$1\cdot 2565 \pm \cdot 0847,$$

while for the distribution of all the p's in the binomials we have determined, we find a range from ·4 to 15·6 with a mean p at $2\cdot 5655 \pm \cdot 3817$.

These results are sufficient to show that there is no real distribution of p round the value unity but the binomials have a distinct tendency to be *negative.*

(13) But the whole theory of Poisson's exponential law in the hands of Bortkewitsch and Mortara appears essentially vague. The binomial is built up on the assumption of the repetition n times of a number of independent events, of which the chance of occurrence is identical and equal to q. The population is n and the chance of occurrence q in the case of each individual. The mean frequency of occurrence is nq. But if q be very small we have seen that the series is

$$e^{-m}\left(1 + m + \frac{m^2}{2\,!} + \frac{m^3}{3\,!} + \dots\right),$$

TABLE XIV.

Marriages of Uncle and Niece (1900—1909).

		0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16 & over*
Marche	O.	7	3	—	—													
	M.	7·41	2·22	·33	·04													
	B.	7	3	—	—													
Umbria	O.	6	3	1	—	—												
	M	6·06	3·03	·76	·13	·02												
	B.	5·90	3·28	·73	·08	—												
Basilicata	O.	6	3	—	1	—												
	M	5·49	3·29	·99	·20	·03												
	B.	6·04	2·59	·92	·31	·10	·03	·01										
Sardegna	O.	2	5	3	—	—	—	—										
	M.	3·33	3·66	2·01	·74	·20	·05	·01										
	B.	2·01	4·96	3·03	—	—	—	—										
Emilia	O.	1	3	2	2	1	1	—										
	M.	1·11	2·44	2·68	1·97	1·08	·48	·18	·05	·01								
	B.	1·09	2·43	2·70	1·98	1·08	·47	·17	·05	·01								
Abruzzi	O.	—	3	1	3	2	—	1	—	—								
	M.	·61	1·70	2·38	2·23	1·56	·87	·41	·16	·06	·02							
	B.	·48	1·58	2·48	2·43	1·68	·87	·34	·11	·03	—							
Lazio	O.	1	1	2	3	—	2	—	1	—	—							
	M.	·45	1·40	2·17	2·24	1·73	1·07	·55	·25	·10	·03	·01						
	B.	·63	1·56	2·09	2·00	1·54	1·01	·59	·31	·14	·06	·03	·01	·01				
Puglie	O.	—	3	1	2	—	1	1	2	—	—	—	—	—				
	M.	·27	·98	1·77	2·13	1·91	1·38	·83	·42	·19	·08	·03	·01	—				
	B.	·55	1·30	1·77	1·80	1·53	1·14	·77	·49	·29	·16	·10	·05	·03	·01	·01		
Veneto	O.	1	—	1	1	3	1	—	1	2	—	—	—	—	—	—		
	M.	·11	·50	1·13	1·69	1·90	1·71	1·28	·82	·46	·23	·10	·04	·02	·01	—		
	B.	·21	·70	1·26	1·62	1·67	1·46	1·13	·79	·51	·30	·17	·09	·05	·02	·01	·01	
Toscana	O.	—	—	1	2	2	2	1	1	—	—	—	—	—	—	—	1	
	M.	·04	·24	·66	1·19	1·60	1·73	1·56	1·20	·81	·49	·26	·13	·06	·02	·01	—	
	B.	·31	·73	1·07	1·25	1·27	1·17	1·01	·83	·65	·50	·37	·27	·19	·13	·09	·06	·10
Calabria	O.	—	—	—	—	—	2	2	—	1	1	1	1	—	1	—	—	1
	M.	—	·01	·05	·16	·36	·64	·94	1·20	1·33	1·32	1·17	·95	·70	·48	·31	·18	·20
	B.	·00	·03	·11	·26	·48	·73	·96	1·16	1·21	1·17	1·04	·87	·69	·51	·37	·25	·16
Totals	O.	24	24	12	14	8	9	5	5	3	1	1	1	—	1	—	1	1
	M.	24·88	19·47	14·93	12·72	10·39	7·93	5·76	4·10	2·96	2·17	1·57	1·13	·78	·51	·32	·18	·20
	B.	24·22	22·16	16·46	11·73	9·35	6·88	4·98	3·74	2·84	2·19	1·71	1·29	·97	·67	·48	·32	·26

from which n has disappeared, and in this exponential we have seen that Bortkewitsch and Mortara suppose m small, i.e. 10 or under. We have seen that there is no reason why m should be absolutely small, and that the name given by Bortkewitsch to the Poisson-Exponential—i.e. the "Law of Small Numbers"—is misleading. But supposing the mean occurrence m to be small, it by no means follows that q need be small and n finite. For if $q = \cdot 2$ and $n = 4$, m would be "small"—and the sort of small number with which our authors deal, but the mere fact that the mean frequency of occurrence was 2 would not justify our using the Poisson-Exponential for

$$(\cdot 8 + \cdot 2)^4.$$

The fact is that when our authors speak of the deaths in a Prussian Army corps from the kick of a horse, or the suicides of schoolgirls, or the deaths from chronic alcoholism as being "small," they really mean small as compared with the number of persons exposed to risk. They had probably in mind all the men in the army corps, all school-girls or all individuals liable to death in the towns considered. But are all men in the army corps,—or only the cavalry, the artillery, etc.,—equally liable to death from the kick of a horse? Is every school-girl equally liable to commit suicide or only a very few morbid and unhealthy minded girls? Is every individual equally liable to die of chronic alcoholism, or only perhaps the 10 or 12 confirmed and aged drunkards in a town? The moment we realise these doubts, what is the population n to be considered? It is not m being small, but the smallness of m/n that leads us to believe that the binomial may have passed into an exponential. But if only six school-girls per year in a community are in the least likely to commit suicide, what is the justification for the "law of small numbers," if the average number of suicides be ·65? Further, if we pass to even a large community in which the tendency to commit suicide is graded—a very probable state of affairs—m might be small and n large, and yet since q is not constant, the binomial and its exponential limit would not be applicable; and this non-applicability would not depend on "lumpiness"—i.e. contagion or example in occurrence. Thus the probability might be:

$$(p_1 + q_1)(p_2 + q_2)(p_3 + q_3) \dots (p_n + q_n)$$

with all the p's independent (as in spinning differently divided teetotums) and not correlated (as they would be in drawing successive non-returned cards from a pack). It would seem therefore that *a priori* we should not expect the conditions for the exponential to be fulfilled in most of the cases selected by Bortkewitsch and Mortara, although with perfect mixing we might expect it in the cases cited by "Student."

(14) In order to test this point on adequate numbers, the ages at death of all persons dying over 70 years of age were extracted for a period of three complete years from the notices of death in the *Times* newspaper for the years 1910—1912: see Table XV. These announcements of death are those of individuals in a fairly limited class, which may be considered stable in numbers for these three years.

TABLE XV.

Deaths per day of the Aged from the Times *newspaper.*

	Number of Deaths per diem	*70* Years and Over			*80* Years and Over			*85* Years and Over			*90* Years and Over		
		Observed	Binomial	Exponential	Observed	Binomial	Exponential	Observed	Binomial	Exponential	Observed	Binomial	Exponential
Men.	*0*	33	32·44	25·94	222	218·27	198·26	484	484·57	480·83	831	829·01	828·25
	1	110	107·50	97·11	339	332·78	338·99	391	391·78	396·16	225	230·70	232·00
	2	170	184·30	181·78	262	271·70	289·81	164	162·04	163·20	38	32·85	32·49
	3	246	217·68	226·84	151	157·71	165·18	45	45·69	44·82	2	3·44	3·26
	4	187	199·09	212·30	79	72·93	70·61	11	9·87	9·23			
	5	142	150·23	158 95	32	28·56	24·15	1	2·05	1·76			
	6	84	97·34	99·18	6	9·84	6·88						
	7	69	55·66	53·04	4	3·06	1·68						
	8	31	28·65	24·82	1	1·15	·44						
	9	19	13·47	10·36									
	10	4	5·86	3·88									
	11	1	3·78	1·80									
Women.	*0*	46	52·92	32·41	162	152·81	126·78	364	363·61	336·25	633	632·31	635·69
	1	140	139·03	114·11	267	274·37	273·47	376	375·20	397·30	350	350·88	346·27
	2	207	200·29	200·89	271	269·90	294·92	218	217·35	234·72	94	94·04	94·31
	3	221	209·30	235·78	185	192·49	212·04	89	93·12	92·45	17	16·20	15·31
	4	169	177·31	207·54	111	111·23	114·34	33	32·87	27·31	2	2·57	4·42
	5	119	129·17	146·15	61	55·25	49·33	13	10·11	6·45			
	6	87	83·89	85·77	27	24·45	17·73	2	2·81	1·27			
	7	44	49·73	43·14	8	9·88	5·46	1	·93	·25			
	8	35	27·38	18·99	3	3·70	1·47						
	9	18	14·17	7·43	1	1·92	·46						
	10	4	6·96	2·62									
	11	4	3·27	·84									
	12	1	1·48	·25									
	13	1	1·10	·08									

Table XVI shows that the announcements of deaths over 70 years of age only amount to 3·74 per day for males and 3·52 for females. These are certainly "small numbers," but "small" with regard to what? Are we to consider n as the number of the population which embraces, (i) all the individuals of the limited classes of the same range of ages as the defunct, (ii) all the individuals announced as dead on the same day, (iii) all the individuals of whatever ages of the class which announces deaths in the *Times*? Or, should we refer to all the individuals in the community of that range of ages, or the whole community at large, i.e. the chance that in a population of so many millions an individual over 70 or 80 as the case may be will die and have their death announced in the *Times* newspaper? Well, it really does not matter, because if for any one or all of these populations the binomial $(p+q)^n$ applied, we should get if q were small and n large, the Poisson series

$$e^{-m}\left(1+m+\frac{m^2}{2\,!}+\frac{m^3}{3\,!}+\dots\right),$$

and this quite regardless of the size of n. If therefore we did find a series in which q was very small and n large, we might not be able to say to which, if any of the above populations n applied. On the other hand the mere fact that m is small is no justification for the use of the "law of small numbers" as is sometimes implied. If it be argued that the small number of people who die over 80 and have their names recorded in the *Times* are drawn from a *small* population, we reply so it may be argued are the school children who commit suicide, the uncles who feel any inclination to marry their nieces, or the men liable to die of chronic alcoholism; and we can in the case of the announcement of deaths test the values of q and n on fairly adequate numbers. As a matter of fact we do not know, in attempting to apply the Poisson formula, what is the population from which we are drawing our individuals, and the justification of the Poisson formula lies only in showing that there actually does exist a binomial for which q is small and n large. We might imagine that as we got to the higher ages practically every person of that age would die, or that in our notation q would be 1 nearly and p be a very small quantity; thus an approach might be made to the Poisson-Exponential. But the approach to the Poisson-Exponential arises not through q approaching unity but from q becoming very small. Nor again in the lower age groups do we find ourselves left with a *positive* binomial.

In all cases except women over 90 years of age, we find that a negative binomial best fits the observations. Even in the case of the announcements of deaths of women over 90 years, we find that the approach of the binomial to the Poisson exponential depends on

$$\left(1+\frac{1}{53{\cdot}3333}\right)^{53{\cdot}3333}$$

being measured with sufficient approximation by $e=2{\cdot}71828$. But

$$(1{\cdot}01875)^{53{\cdot}333}=2{\cdot}69323,$$

and is therefore not a very close approximation, a result shown when we use a binomial by the substantial improvement in the measure P of "goodness of fit." Even in this case we are not prepared to say what is the population for which the $q = ·01875$ in the case of these announcements of deaths of women over 90 years of age. It can scarcely be that there are only 29 women over 90 years of age living in the country, whose deaths are likely to be announced in the *Times* when they occur. Further the probable error of q is such that actually this case might equally well be a random sample from material following a negative binomial. Analysing our material we see that our first two cases of males and the first three of females are such that they could not possibly be random samples from positive binomials, the probable errors of q are too small. Next, seven cases out of the eight do give actually negative binomials and the eighth might, having regard to its probable errors, well be a negative binomial. Thus although our daily occurrences are certainly in Bortkewitsch and Mortara's sense "small numbers," they give no support to the use of a Poisson-Exponential.

TABLE XVI.

Constants for Deaths of Aged.

Men.

Age over	p	q	Probable Error of q	n	Probable Error of n	m	Binomial P	Exponential P
70 years ...	1·12965	− ·12965	± ·03314	− 28·8747	± 7·3734	3·7436	·1355	·0045
80 years ...	1·12152	− ·12152	± ·03349	− 14·0703	± 3·8704	1·7099	·9358	·1129
85 years ...	1·01903	− ·01903	± ·02902	− 43·2996	± 67·5797	·8239	·9737	·9715
90 years ...	1·00654	− ·00654	± ·02934	− 42·8498	± 192·3069	·2801	·6741	·6672

Women.

Age over	p	q	Probable Error of q	n	Probable Error of n	m	Binomial P	Exponential P
70 years ...	1·34012	− ·34012	± ·04161	− 10·3522	± 1·2307	3·5210	·8084	·0000
80 years ...	1·20770	− ·20770	± ·03294	− 10·4400	± 1·8309	2·1569	·9686	·0018
85 years ...	1·14507	− ·14507	± ·03077	− 8·1447	± 1·9627	1·1816	·9860	·1062
90 years ...	·98125	+ ·01875	± ·02779	+ 29·0573	± 43·0634	·5447	·9848	·8116

If it be said that these "small numbers" differ in character from those used by our authors, the reply must be: we know in none of these cases the real population from which deaths are to be considered as drawn. The chances of death are certainly graduated with age, but the chances of suicide are graduated with temperament, and the same is true of alcoholism, or again the chance of death by accident is graduated with occupation. At any rate until those who support the use of the "law of small numbers" demonstrate its application on material, where the probable errors are sufficiently small for us to measure the true value of q and n, no advance can be made. Nor until we have clear ideas of the population n in which the chance is q, is it possible to assert that it may be used for the suicides of school children, and the marriage of uncle and niece, and must not be used for the deaths of aged people, which certainly occur in "smaller" numbers.

In the illustrations of deaths we have taken, certainly the Poisson-Exponential is not the rule, although the distributions appear to approach it, as towards a limit, when the number of deaths approach zero. But our data which show the rule of the negative binomial appear to show it in no more marked manner than much of the data selected by Mortara himself indicate the negative binomial, although owing to the sparsity of his material his results are far more erratic and unreliable. Nor is Bortkewitsch much behind Mortara in the evidence he produces for a negative binomial being as reasonable a description—possibly owing to inherent lumpiness—as a positive binomial of these "small number" frequencies.

(15) *Conclusions.*

(*a*) The Poisson-Exponential gives a fairly reasonable method of dealing with the probable deviations of small sub-frequencies in the case of random sampling. When the average value of a sub-frequency is not more than 3 °/₀ of a population, then Poisson's formula suffices in most practical cases to determine the range of error likely to be made. Tables are given to assist its use.

(*b*) The application of the Poisson-Exponential to various data by Bortkewitsch and Mortara has hardly been justified by those writers, for they have not tested whether the probability q is small and positive and the power n large and positive in the cases considered by them. When this is actually done, it is found that their hypotheses, having regard to the probable errors of q and n, are largely unjustified in the case of their illustrations. Even in such cases where it is justified, a binomial gives a better result as measured by the test for goodness of fit.

(*c*) Negative binomials repeatedly occur and give just as good fits, where they occur, as positive binomials. In the illustrations taken by Mortara, the frequency 10 used is so small that it is not possible to assert that either positive or negative binomials are demanded by the data. Still the average p of his results is very significantly in excess of unity.

(*d*) Mortara like Bortkewitsch cuts out of his data straight off all districts with, on the average, more than 10 cases in the year. But the q obtained from 20, 40, or even 100 cases in a population of 100,000 is a small q in the sense that the resulting binomial is adequately expressed by a Poisson-Exponential. There appears to be no valid reason for such a procedure, except the experience that many such cases actually give negative binomials[*]. It seems to us theoretically unjustifiable to apply the exponential to 8 cases say in a district of 100,000, and not apply it to 12 cases in a district of 200,000. Actually p may be 1·4 in the first case and only 0·9 in the second.

(*e*) We consider that the reasonable method in every case is not to start with the Poisson-Exponential, which screens the truth or falsity of the *a priori* hypotheses, but to fit a binomial regardless of the magnitude of p. The fact that quite as good fits are obtained with negative as with positive binomials suggests that a new interpretation of these cases of "negative probability" is requisite. Several cases of the interrelation of "contributory cause groups" which provide a series represented by a negative binomial $(p-q)^{-n}$ have been recognised[†]. A general interpretation based on a very simple conception seems needed for these demographic cases in which the law of small numbers appears far more often to correspond to a negative than to a positive binomial.

This paper was worked out in the Biometric Laboratory, and I have to thank Professor Karl Pearson for his aid at various stages.

---

[*] Can we cite in addition perhaps, the fact that existing tables of $m^x e^{-m}/x!$ do not extend beyond $m=10$?

[†] Pearson, *Biometrika*, Vol. IV. p. 208.

THE RELATIONSHIP BETWEEN THE WEIGHT OF THE SEED PLANTED AND THE CHARACTERISTICS OF THE PLANT PRODUCED. II.

By J. ARTHUR HARRIS, Ph.D., Carnegie Institution of Washington, U.S.A.

I. Introductory Remarks.

1. In *Biometrika*, Vol. IX. pp. 11—21, March 1913, were published constants showing the relationship between the weight of the seed planted and the number of pods on the plants produced in twenty experimentally grown series of *Phaseolus vulgaris.* From the economic view point, number of pods is the most important character which could have been chosen, total weight of seed matured only excepted. But to the student of morphogenesis, or of the physiology of seed production, other characters are of equal interest, while the comparison of the correlations for various features must yield results of significance.

The purpose of the present communication is the presentation of the constants measuring the influence of the weight of the seed planted upon the number of ovules formed and the number of seeds developing in the pods of the matured plant.

These various relationships have now been worked out for a relatively large bulk of material. Altogether there are 29 individual series belonging to 5 varieties, involving 17,953 plants, from which 119,192 determinations of the number of ovules and seeds per pod have been made. The reply to the possible suggestion that the expenditure of effort in the collection and analysis of such masses of data is quite unjustifiable is twofold. First, a major portion of the labour involved was necessary for investigations not touched upon here. Secondly, there are many problems of morphogenesis and physiology which can only be solved by the amassing of large series of accurately determined biometric constants which when sufficiently numerous may themselves be the materials for statistical analysis. The data here contained are recorded in partial fulfilment of such requirements for certain definite morphological and physiological problems.

The present paper is limited strictly to matters of fact; general discussions are reserved until further data—much of which is already available in a raw state—are reduced.

II. MATERIALS.

The first paper may be consulted for details not entered here. The data analysed are drawn in part from the series already considered for the relationship between weight planted and number of pods produced. In addition to the White Flageolet, Navy and Ne Plus Ultra varieties already treated, several lots of Burpee's Stringless and two of Golden Wax are available.

III. ANALYSIS OF DATA.

2. Data for Number of Ovules and Seeds per Pod.

Tables III—VI, similar to those of the preceding paper, give in a condensed form the data for the correlations discussed. Table I* gives the correlations

TABLE I. *Correlation and Partial Correlation Coefficients.*

Series	Number of Plants	Correlation, Weight and Pods r_{wp}	Number of Pods Examined	Correlation, Weight and Ovules r_{wo}	Partial Correlation, ${}_pr_{wo}$	Correlation, Weight and Seeds r_{ws}	Partial Correlation, ${}_pr_{ws}$
LL	1141	−·008 ± ·020	8043	·026 ± ·008	·027 ± ·008	−·013 ± ·008	−·013 ± ·008
LG	182	·066 ± ·050	806	·153 ± ·023	·140 ± ·023	−·100 ± ·024	−·103 ± ·024
GG	750	−·368 ± ·021	6310	·018 ± ·008	·029 ± ·008	·004 ± ·008	·016 ± ·008
GGH	583	·208 ± ·027	5251	·045 ± ·010	·019 ± ·009	·024 ± ·010	−·004 ± ·009
*GGH*2	499	·176 ± ·029	3502	·093 ± ·011	·083 ± ·011	·063 ± ·011	·049 ± ·011
GGHH	396	·193 ± ·033	2656	−·022 ± ·013	−·042 ± ·013	−·029 ± ·013	−·048 ± ·013
GGD	514	·159 ± ·039	1438	·107 ± ·018	·089 ± ·018	·071 ± ·018	·068 ± ·018
*GGD*2	449	·215 ± ·030	1227	·044 ± ·019	·018 ± ·019	·079 ± ·019	·062 ± ·019
GGDD	342	·137 ± ·036	807	·101 ± ·023	·092 ± ·024	·089 ± ·024	·076 ± ·024
HH	1484	·177 ± ·017	14029	·010 ± ·006	−·039 ± ·006	·007 ± ·006	−·054 ± ·006
HHH	1271	·145 ± ·019	11230	−·000 ± ·006	−·030 ± ·006	·016 ± ·006	−·014 ± ·006
HD	1416	·129 ± ·018	5581	−·044 ± ·009	−·067 ± ·009	−·049 ± ·009	−·052 ± ·009
HDD	1204	·121 ± ·019	5449	−·029 ± ·009	−·065 ± ·009	−·010 ± ·009	−·030 ± ·009
DD	513	·282 ± ·027	1827	·098 ± ·016	·009 ± ·016	·050 ± ·016	·008 ± ·016
DDD	459	·215 ± ·030	2018	·044 ± ·015	·000 ± ·015	·046 ± ·015	·006 ± ·015
DH	670	·258 ± ·024	5955	·075 ± ·009	−·005 ± ·009	·076 ± ·009	−·013 ± ·009
DHH	565	·152 ± ·028	5019	·045 ± ·010	·008 ± ·010	·011 ± ·010	−·025 ± ·010
USC	530	·150 ± ·029	2569	·059 ± ·013	·032 ± ·013	·031 ± ·013	·024 ± ·013
USS	680	·155 ± ·025	6605	·023 ± ·008	−·000 ± ·008	·041 ± ·008	·024 ± ·008
USH	361	·129 ± ·035	3406	·032 ± ·012	·001 ± ·012	·037 ± ·012	·020 ± ·012
USHH	224	·143 ± ·044	1743	·112 ± ·016	·098 ± ·016	·011 ± ·016	−·004 ± ·016
USD	312	·195 ± ·037	802	·127 ± ·023	·098 ± ·024	·071 ± ·024	·067 ± ·024
USDD	237	·241 ± ·041	851	·238 ± ·022	·175 ± ·023	·131 ± ·023	·090 ± ·023
FSC	586	·147 ± ·027	2876	·047 ± ·013	·017 ± ·013	·089 ± ·012	·073 ± ·013
FSS	868	·098 ± ·023	7809	·021 ± ·008	·001 ± ·008	·026 ± ·008	·004 ± ·008
FSH	475	·100 ± ·031	4541	·049 ± ·010	·018 ± ·010	−·045 ± ·010	−·073 ± ·010
FSHH	427	·121 ± ·032	3837	·015 ± ·011	−·013 ± ·011	·040 ± ·011	·017 ± ·011
FSD	428	·130 ± ·032	1449	·060 ± ·018	−·027 ± ·018	−·019 ± ·018	−·036 ± ·018
FSDD	387	·144 ± ·034	1556	·037 ± ·017	·013 ± ·017	·047 ± ·017	·024 ± ·017

---

* The weight of the seed planted was weighted with the number of pods counted. Thus $\bar{w}$ and σ_w differ slightly from those of Table II of the first paper. Sheppard's correction was used for seed weight, but not for the integral variates ovules per pod or seeds per pod.

between weight of seed planted and ovules per pod, r_{wo}, and between weight planted and number of seeds matured per pod, r_{ws}. The partial correlation coefficients,

$$_pr_{wo} = \frac{r_{wo} - r_{wp}\, r_{po}}{\sqrt{1 - r^2_{wp}}\,\sqrt{1 - r^2_{po}}}, \qquad _pr_{ws} = \frac{r_{ws} - r_{wp}\, r_{ps}}{\sqrt{1 - r^2_{wp}}\,\sqrt{1 - r^2_{ps}}},$$

showing the correlation for weight (w) and ovules (o) and weight and seeds (s) for constant numbers of pods (p) per plant are also given. These require in addition to the correlations here given r_{wp}, r_{po} and r_{ps}, the correlations between the number of pods per plant and the number of ovules and seeds in these pods. Values of r_{wp} are available from the preceding paper (*Biometrika*, Vol. IX. p. 21, Table VII) and from a supplementary table giving nine additional constants*. For the reader's convenience these are reprinted in this table. The values of r_{po} and r_{ps} will be published in connection with another problem.

The probable errors have all been calculated on the basis of the number of pods examined as N. There is considerable question whether the actual number of seeds planted should not have been used instead; the degree of trustworthiness of a constant is perhaps not greater than is indicated by the lowest number of actual measurements (irrespective of the number of associated measures taken). The point is not of the greatest practical importance for the present case, since the number of series is so large that conclusions can be drawn from the run of the constants as a whole and too much weight need not be given to individual series.

A glance at the table shows that the correlations are low throughout. The suggestion naturally arises that some of the extremely low values may be due to non-linear regression. The regression straight line equations and the results of Blakeman's test† are given in Table II. Here r, η and the straight line equation for the regression of ovules and seeds per pod on weight planted (in working units) are determined by the conventional formulae. The final two columns give the values of

$$\left\{\tfrac{1}{2}\sqrt{\zeta}\cdot\frac{1}{\sqrt{1 + (1 - r^2)^2 - (1 - \eta^2)^2}}\right\}\Big/\chi_1$$

when $\zeta = \eta^2 - r^2$ and $\chi_1 = \cdot 67449/\sqrt{N}$.

All the straight lines are shown in Diagram 1. The empirical means are indicated in all of the cases where it can be done without confusion. The slope is very slight and the agreement of observed and predicted means not very close, especially near the ends of the range, *where the number of observations is small.* There is, however, no clear indication that a curve of a higher order would describe the results better than a straight line. This irregularity is precisely what is to be expected in cases of low correlation.

---

* Harris, J. Arthur, "An Illustration of the Influence of Substratum Heterogeneity upon Experimental Results." *Science*, N. S. Vol. XXXVIII. pp. 345—346, 1913.

† Blakeman, J., *Biometrika*, Vol. IV. pp. 332—350, 1905.

TABLE II.

Tests for Linearity of Regression.

Series	Correlation, r, and Probable Error	Correlation-Ratio, η, and Probable Error	Regression Straight Line Equation	Blakeman's Criterion, Test A	Blakeman's Criterion, Test B
For Ovules:					
USS	·0232 ± ·0083	·0657 ± ·0083	5·4230 + ·0074 w	3·720	1·688
DHH	·0445 ± ·0095	·0788 ± ·0095	4·9385 + ·0257 w	3·431	1·151
USDD	·2381 ± ·0218	·2978 ± ·0211	3·6886 + ·1001 w	11·096	2·102
*GGD*2	·0442 ± ·0192	·1276 ± ·0189	4·7224 + ·0137 w	3·152	1·907
FSS	·0209 ± ·0076	·0403 ± ·0076	5·5606 + ·0153 w	2·263	·754
HH	·0098 ± ·0057	·0661 ± ·0057	5·3600 + ·0056 w	5·159	1·678
For Seeds:					
USS	·0407 ± ·0083	·0946 ± ·0082	3·5870 + ·0206 w	5·182	2·351
DHH	·0111 ± ·0095	·0541 ± ·0095	4·1521 + ·0106 w	5·573	1·869
USDD	·1313 ± ·0227	·1932 ± ·0223	2·1840 + ·0940 w	8·712	1·650
*GGD*2	·0794 ± ·0191	·1760 ± ·0187	2·4735 + ·0346 w	4·181	2·529
FSS	·0261 ± ·0076	·0499 ± ·0076	3·0712 + ·0269 w	2·793	·931
HH	·0068 ± ·0057	·0953 ± ·0057	4·2119 + ·0058 w	8·421	2·739

Blakeman's criterion has been applied in two ways, *A* and *B*. In the first the actual number of pods examined has been taken as N. In test *B* the number of seeds planted (not the weighted number) has been used in obtaining χ_1. If the first test be accepted as the proper one, it follows that regression cannot safely be regarded as linear. But there are two important points to be taken into account. The correlation ratio η depends upon the squares of the differences in means, hence it has always a positive value, which may be very substantial because of the errors of sampling when the number of individuals per array is small. Thus when r approaches zero η is limited by $\bar{\eta}$, the mean values of η for zero correlation*. Hence a test for linearity based on a comparison of η with a very low value of r may be misleading. Again, as pointed out above, the significance of both r and η should perhaps be tested on the basis of the lowest number of measurements. If this be done, as it is in test *B*, there is found very little evidence for non-linear regression. Certainly, one cannot possibly assert that the low values of r, which is seen throughout these experiments, is due to the number of ovules (seeds) per pod at first becoming larger and then decreasing after a maximum is reached as one passes from the lowest to the highest grade of seed weight.

The results of Table I are also shown graphically in Diagram 2. Here the relationships for weight of seed planted and number of pods on the plant developing are also indicated as a basis of comparison. The values of both r_{wo} and r_{ws} are in general conspicuously lower than the low values of r_{wp}. But very few of them drop below the zero bar; one is forced to the conclusion that there is a distinct though very slight correlation between weight and ovules and between weight and seeds.

---

* See K. Pearson, *Biometrika*, Vol. viii. pp. 254—256, 1911.

Consider in somewhat greater detail the signs and magnitudes of these correlations*.

Of the 26 values of r_{wo} only 4 are negative. The mean value of the 22 positive coefficients is $+ \cdot 0673$; the mean of the 4 negative is $- \cdot 0236$; the mean of all (regarding signs) is $+ \cdot 0533$.

For the relationship between weights of seed planted and number of seed matured per pod, r_{ws}, 21 constants are positive and 5 are negative. The mean of the positive coefficients is $+ \cdot 0502$; the mean of the negative values is $- \cdot 0303$; for all 26 correlations the mean (regarding signs) is $+ \cdot 0348$.

Thus both correlations are (as is clear from the diagrams) unquestionably positive but very low.

Apparently the relationship for weight and ovules is slightly closer than that for weight and seeds per pod, but the difference is too slight to justify any final conclusion.

Consider now the question whether the observed correlations r_{wo}, r_{ws} are to be regarded as direct biological relationships between the two variables w and o or w and s, or whether they are to be looked upon as merely necessary resultants of other interdependences. At present, the only other demonstrated correlation which might tend to bring about sensible values of r_{wo} and r_{ws} is that between number of pods per plant and number of ovules formed and number of seeds developing per pod. Since number of pods per plant is known to be correlated with weight of seed planted, while both number of ovules and number of seeds per pod are correlated with number of pods per plant, some correlation must be expected between weight planted and number of ovules and seeds per pod. If now the observed values of r_{wo} and r_{ws} which are always small, are merely the necessary resultant of the relationships r_{wp}, r_{po}, r_{ps}, one would expect the partial correlation coefficients, ${}_pr_{wo}$, ${}_pr_{ws}$, to be sensibly zero. If these partial correlations are not sensibly zero, it can only mean that there is a direct (causal) relationship other than the one just considered between number of ovules (or seeds) and the weight of the seed planted.

The partial correlations and the correlations are shown side by side in Diagrams 3 and 4. The lowering of the degree of interdependence between both weight and ovules and weight and seeds by the correction for number of pods per plant is clearly marked. In a number of cases in which the correlation coefficient is positive the partial correlation coefficient is negative.

Thus only 4 of the 26 values of r_{wo} are negative, while 9 of the partial correlation coefficients have the minus sign. In only 5 cases is r_{ws} negative, but in 11 of the series, the sign of ${}_pr_{ws}$ is negative. The mean values of the partial correlations are very close indeed to zero. Thus ${}_p\bar{r}_{wo} = \cdot 0186$ as compared with $\bar{r}_{wo} = \cdot 0533$; ${}_p\bar{r}_{ws} = \cdot 0099$ as against $\bar{r}_{ws} = \cdot 0348$.

---

* I have already shown (*Science*, N. S. Vol. XXXVIII. pp. 345—346, 1913) that the *LL*, *LG* and *GG* series are open to question because of the lack of certain precautions in the cultures; while they are included in the table of fundamental constants to avoid any possible criticism of selection of series they will be left out of account in the following discussions.

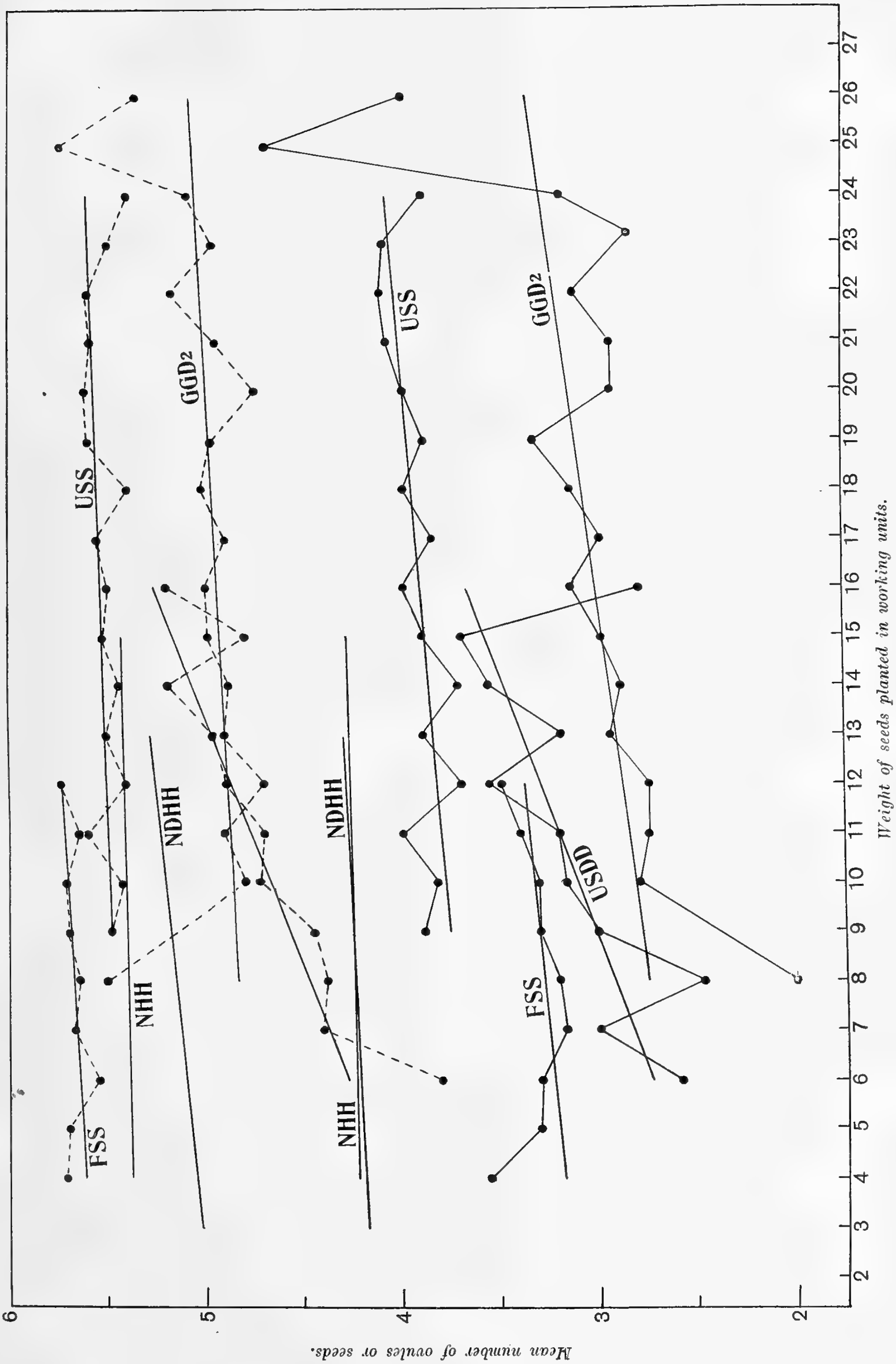


Diagram 1. Regression of ovules per pod and of seeds per pod on weight of seed planted. The six upper lines are for ovules, the six lower for seeds.

Diagram 2. Comparison of the known correlations for weight of seeds planted and characteristics of individuals developing. Solid dots and firm lines $= r_{wp}$; circles and firm lines $= r_{wo}$; circles and broken lines $= r_{ws}$.

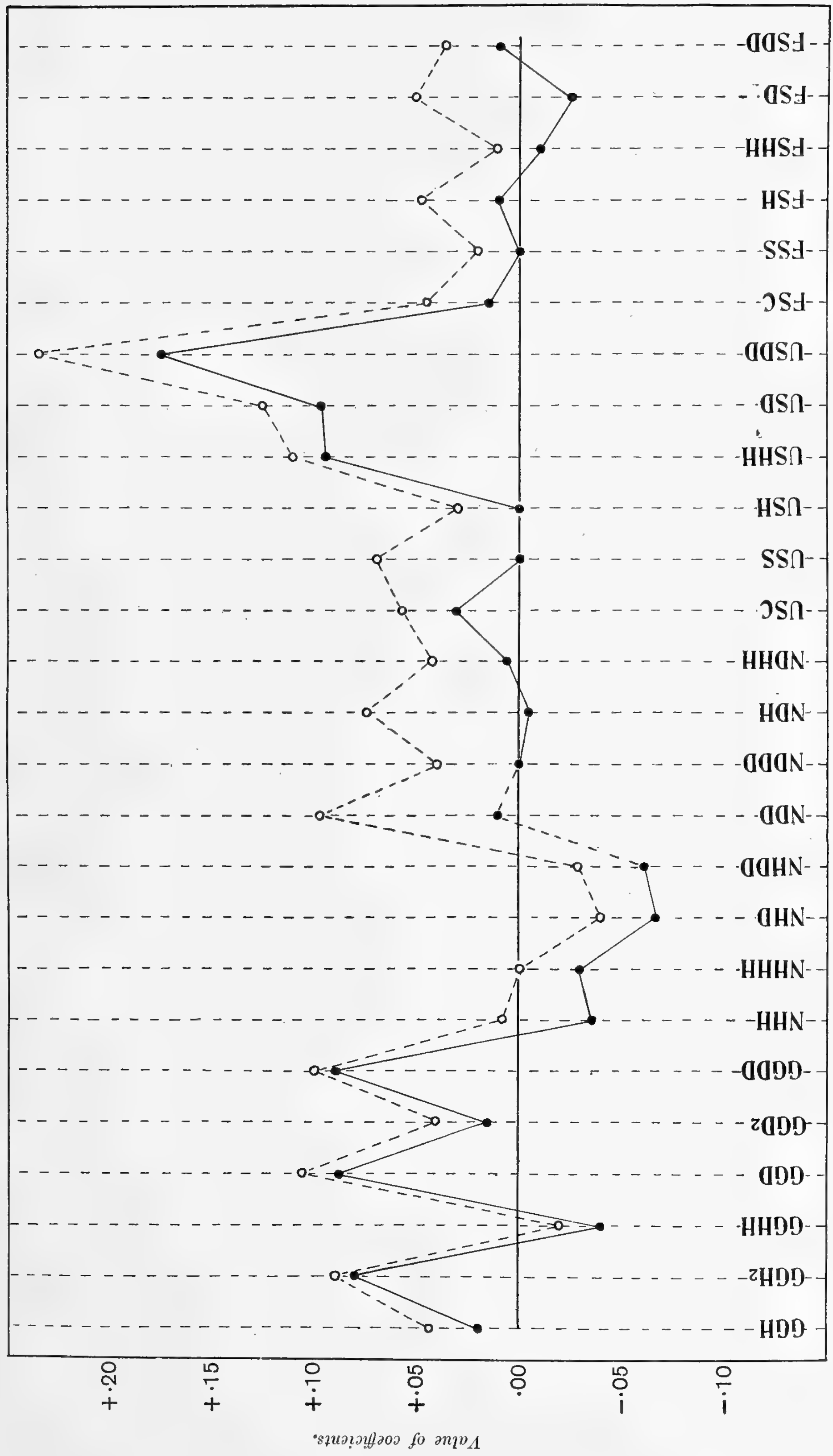


DIAGRAM 3. Comparison of correlation of weight of seed planted and ovules per pod with the partial correlation for constant number of pods per plant. Broken lines $=r_{wo}$; solid lines $={}_{p}r'_{wo}$.

DIAGRAM 4. Comparison of correlation of weight of seed planted and number of seeds per pod with the partial correlation for constant number of pods per plant. Broken lines $= r_{ws}$; solid lines $= {}_p r_{ws}$.

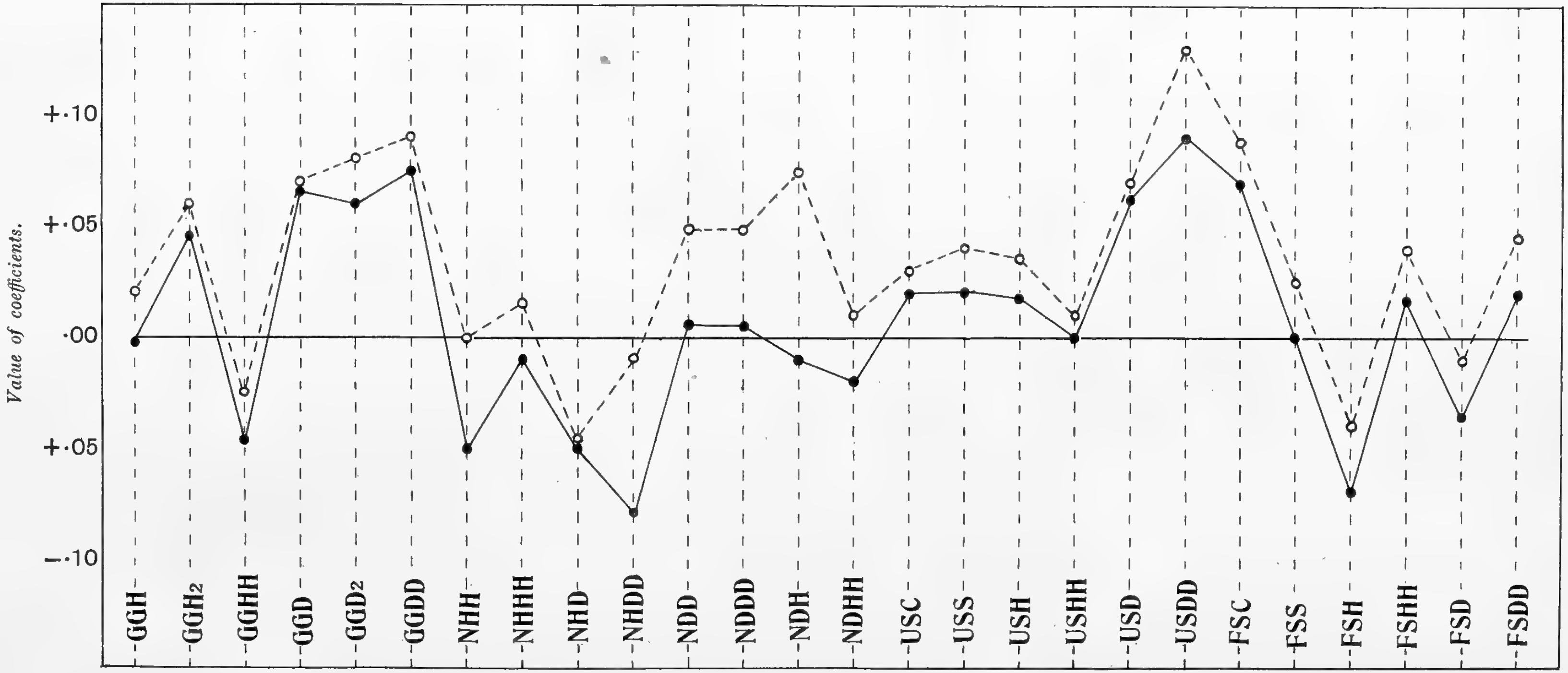


TABLE III.

Weight of Seed Planted	Series *GG*			Series *GGH*			Series *GGH2*			Series *GGHH*			Series *GGD*			Series *GGD2*			Series *GGDD*		
	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds
·050—·075 (3)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·075—·100 (4)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·100—·125 (5)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·125—·150 (6)	—	—	—	6	32	23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·150—·175 (7)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	8	6
·175—·200 (8)	—	—	—	7	41	32	12	69	49	—	—	—	—	—	—	2	11	4	1	6	5
·200—·225 (9)	—	—	—	25	124	100	44	206	143	—	—	—	4	18	17	—	—	—	5	21	13
·225—·250 (10)	7	41	21	47	261	192	13	62	56	3	17	7	8	34	18	5	24	14	7	32	16
·250—·275 (11)	12	53	33	74	425	302	51	278	192	—	—	—	18	92	44	19	94	53	12	49	28
·275—·300 (12)	8	37	20	185	1073	801	52	271	217	—	—	—	32	147	92	24	113	54	36	172	107
·300—·325 (13)	28	171	65	221	1242	950	109	578	453	79	414	343	84	417	206	60	296	177	75	370	233
·325—·350 (14)	122	750	341	327	1793	1330	229	1192	920	105	522	390	87	418	236	75	367	218	112	546	350
·350—·375 (15)	154	908	429	352	2023	1519	367	1898	1494	172	898	733	105	493	270	93	462	288	154	719	431
·375—·400 (16)	177	983	516	561	3124	2354	404	2037	1562	264	1385	1091	161	762	452	108	541	341	127	622	393
·400—·425 (17)	464	2780	1324	721	4059	3167	526	2792	2224	357	1881	1489	188	900	504	156	765	468	106	517	340
·425—·450 (18)	418	2554	1308	462	2655	2026	397	2146	1618	493	2558	2079	146	694	410	170	865	543	64	311	212
·450—·475 (19)	374	2249	1158	518	2999	2282	403	2157	1661	391	1992	1598	180	864	528	189	932	634	57	271	167
·475—·500 (20)	515	3147	1554	643	3686	2762	362	1940	1515	358	1886	1425	128	648	356	109	522	322	25	119	75
·500—·525 (21)	781	4679	2369	431	2435	1862	251	1306	1020	218	1130	878	118	596	352	95	472	281	11	56	33
·525—·550 (22)	724	4217	2178	301	1709	1263	145	815	598	126	640	514	107	538	314	56	291	176	2	13	12
·550—·575 (23)	683	4200	2029	204	1188	925	58	305	246	44	231	170	24	116	65	24	119	69	11	62	44
·575—·600 (24)	483	2931	1451	82	465	344	37	190	155	28	135	113	24	129	69	30	154	96	—	—	—
·600—·625 (25)	406	2458	1142	45	273	203	19	105	82	—	—	—	14	75	35	4	23	19	—	—	—
·625—·650 (26)	256	1500	735	30	167	101	11	70	57	18	91	74	8	39	23	8	43	32	—	—	—
·650—·675 (27)	255	1512	731	9	54	43	5	27	22	—	—	—	2	12	9	—	—	—	—	—	—
·675—·700 (28)	93	512	241	—	—	—	7	42	32	—	—	—	—	—	—	—	—	—	—	—	—
·700—·725 (29)	159	942	532	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·725—·750 (30)	141	798	392	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·750—·775 (31)	23	126	77	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·775—·800 (32)	20	116	63	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·800—·825 (33)	7	44	21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

TABLE IV.

Weight of Seed Planted	Series *USC*			Series *USS*			Series *USH*			Series *USHH*			Series *USD*			Series *USDD*			Series *LL*			Series *LG*		
	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds
·075—·100 (4)	5	27	15	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·100—·125 (5)	2	8	7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·125—·150 (6)	3	16	12	—	—	—	—	—	—	—	—	—	—	—	—	7	27	18	14	64	36	—	—	—
·150—·175 (7)	—	—	—	—	—	—	10	51	39	—	—	—	—	—	—	17	84	51	15	82	37	5	21	10
·175—·200 (8)	7	34	26	—	—	—	10	54	38	—	—	—	—	—	—	33	145	81	90	464	251	1	3	1
·200—·225 (9)	18	93	75	73	401	284	10	50	43	—	—	—	11	54	29	91	407	273	122	623	313	5	21	11
·225—·250 (10)	25	136	81	176	957	670	93	476	370	28	134	104	7	33	16	160	761	507	286	1433	791	7	31	15
·250—·275 (11)	126	703	511	557	3125	2268	251	1262	1009	167	855	658	32	147	79	216	1018	697	264	1306	673	26	121	61
·275—·300 (12)	293	1627	1215	1136	6215	4208	396	2038	1550	178	931	677	113	520	286	149	747	529	460	2342	1277	67	306	159
·300—·325 (13)	444	2488	1840	1207	6689	4717	668	3448	2673	319	1647	1204	135	642	357	93	462	299	620	3128	1641	87	405	228
·325—·350 (14)	535	2960	2202	1120	6137	4190	521	2697	2072	396	2079	1578	161	762	434	55	286	197	917	4565	2388	66	312	161
·350—·375 (15)	348	1955	1447	771	4288	3031	581	2974	2296	290	1548	1110	142	680	362	25	121	80	923	4662	2394	134	657	274
·375—·400 (16)	294	1664	1201	421	2324	1686	267	1386	1064	215	1124	812	47	230	118	5	26	14	1311	6729	3501	125	603	263
·400—·425 (17)	134	759	587	335	1866	1291	144	744	624	65	351	262	36	172	96	—	—	—	887	4455	2335	77	357	161
·425—·450 (18)	104	584	435	182	989	730	120	617	511	44	243	186	28	136	76	—	—	—	864	4460	2320	92	436	185
·450—·475 (19)	94	539	389	229	1289	912	119	607	479	31	169	123	31	151	89	—	—	—	564	2886	1463	62	310	138
·475—·500 (20)	78	426	308	160	905	655	76	399	339	10	51	31	27	130	72	—	—	—	534	2713	1353	13	64	30
·500—·525 (21)	39	225	165	128	716	524	110	575	446	—	—	—	8	38	22	—	—	—	134	700	391	25	127	60
·525—·550 (22)	10	54	44	60	338	249	10	45	20	—	—	—	10	53	28	—	—	—	15	77	38	8	41	13
·550—·575 (23)	10	54	42	30	165	122	10	55	40	—	—	—	7	43	28	—	—	—	23	102	58	6	28	6
·575—·600 (24)	—	—	—	20	108	78	10	60	44	—	—	—	7	35	19	—	—	—	3	14	5	—	—	—

TABLE V.

Weight of Seed Planted	Series *HH*			Series *HHH*			Series *HD*			Series *HDD*			Series *DD*			Series *DDD*			Series *DH*			Series *DHH*		
	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds
·050—·075 (3)	—	—	—	—	—	—	2	8	4	4	15	9	—	—	—	—	—	—	8	40	37	2	8	8
·075—·100 (4)	5	21	14	—	—	—	—	—	—	9	43	36	6	27	17	4	17	10	51	258	209	—	—	—
·100—·125 (5)	23	107	92	—	—	—	10	42	27	54	254	193	65	269	189	17	80	58	212	1086	876	10	51	47
·125—·150 (6)	222	1154	869	76	419	316	57	279	182	349	1766	1319	151	631	477	47	224	159	642	3213	2474	108	535	431
·150—·175 (7)	880	4784	3771	792	4333	3421	291	1386	909	995	5129	3811	480	2047	1507	195	910	718	1690	8648	7095	345	1761	1427
·175—·200 (8)	2601	14018	10873	2276	12335	9808	1007	4781	3227	1378	6875	5174	562	2468	1792	524	2452	1845	1907	9784	7960	1414	7342	6063
·200—·225 (9)	3777	20563	16219	3349	17852	14108	1552	7437	4952	1309	6433	4887	335	1501	1113	614	2907	2188	985	5153	4264	1472	7590	6253
·225—·250 (10)	3157	17128	13705	2751	14617	11730	1225	5725	3790	763	3762	2789	169	739	539	420	1980	1506	352	1832	1484	1165	6008	4978
·250—·275 (11)	2081	11228	8998	1373	7344	5945	856	3943	2584	394	1955	1437	55	251	176	140	670	527	88	471	405	395	2084	1689
·275—·300 (12)	904	4871	3852	466	2603	2068	412	1947	1265	146	747	581	4	15	13	39	194	159	20	109	86	81	422	334
·300—·325 (13)	249	1374	1043	78	411	338	117	537	353	29	145	98	—	—	—	18	83	59	—	—	—	27	143	107
·325—·350 (14)	83	431	274	40	224	162	39	198	119	19	114	96	—	—	—	—	—	—	—	—	—	—	—	—
·350—·375 (15)	47	254	154	20	126	103	13	55	39	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
·375—·400 (16)	—	—	—	9	48	40	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

TABLE VI.

Weight of Seed Planted	Series *FSC*			Series *FSS*			Series *FSH*			Series *FSHH*			Series *FSD*			Series *FSDD*		
	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds	Pods	Total Ovules	Total Seeds
·050—·075 (3)	5	35	26	—	—	—	—	—	—	—	—	—	4	20	8	3	15	12
·075—·100 (4)	5	27	19	59	339	210	13	66	47	10	59	48	9	45	24	4	22	18
·100—·125 (5)	70	420	295	184	1054	614	84	466	386	23	117	85	30	138	107	40	204	155
·125—·150 (6)	198	1193	806	636	3546	2117	401	2175	1742	112	619	490	61	293	225	286	1430	1108
·150—·175 (7)	475	2886	2039	1490	8487	4764	751	4176	3393	365	2006	1636	248	1200	851	676	3361	2528
·175—·200 (8)	823	4990	3466	2188	12382	7100	1093	6102	4834	1490	8247	6683	494	2493	1766	350	1779	1387
·200—·225 (9)	764	4679	3325	1762	10083	5863	1269	7036	5581	1186	6500	5289	361	1790	1205	155	784	620
·225—·250 (10)	354	2185	1562	1077	6162	3605	624	3498	2650	540	2996	2382	149	747	521	41	210	166
·250—·275 (11)	152	937	660	314	1783	1084	246	1380	1067	91	493	395	56	264	189	1	5	4
·275—·300 (12)	30	180	130	99	569	349	50	281	195	20	124	91	37	194	121	—	—	—
·300—·325 (13)	—	—	—	—	—	—	10	59	43	—	—	—	—	—	—	—	—	—

IV. Recapitulation.

The facts presented in this paper and in the preceding studies justify the following conclusions.

1. In *Phaseolus vulgaris* there is a sensible relationship between the weight of the seed planted and the number of pods on the plant developing from it. The correlation is always low, averaging only about ·166, but under proper experimental conditions the coefficients have always been found to be positive. When experiments are not made with all necessary precautions substratum heterogeneity may completely obscure the influence of seed weight, reducing the correlation to practically zero or even bringing about a substantial negative correlation.

2. There is also a significant positive correlation between the weight of the seed planted and the number of ovules and the number of seeds in the pods produced by the plant developing from it. These correlations are so low that on relatively small samples negative values may be found. They average only about one-fifth to one-third the magnitude of the correlation for weight planted and pods per plant.

The relationship for weight and ovules is numerically higher than that for weight and seed, but on the basis of the number of series now available the difference cannot be asserted to be significant.

3. Morphogenetically and physiologically, the observed correlations between weight and ovules and weight and seeds are to be regarded as the resultant of two other correlations, namely, that between the weight of the seed planted and the number of pods per plant and that between the number of pods on the plant and the characteristics of these pods. This conclusion is based on the fact that the partial correlation coefficient for weight of seed planted and number of ovules or seeds per pod for constant number of pods per plant is practically zero.

Cold Spring Harbor, N.Y.
August 20, 1913.

ON THE PROBABILITY THAT TWO INDEPENDENT DISTRIBUTIONS OF FREQUENCY ARE REALLY SAMPLES OF THE SAME POPULATION, WITH SPECIAL REFERENCE TO RECENT WORK ON THE IDENTITY OF TRYPANOSOME STRAINS

By KARL PEARSON, F.R.S.

(1) In *Biometrika*, Vol. VIII. p. 250, I discussed fully the mathematical process requisite for measuring the probability that two independent distributions of frequency are really samples of the same population. As far as I am aware this is the only complete theory of the subject which has been published. I believe it to be scientifically adequate, and it has already been applied to a large number of problems*.

Before that paper was published, it had been usual to compare any constants of two frequency distributions together, and by a due consideration of their difference relative to the combination of their probable errors to determine the probability of the identity of those constants. This could be repeated for any number of corresponding constants, and if theoretical curves of frequency had been fitted, their divergence or correspondence measured by the divergence or correspondence of their complete series of constants. The method above referred to, however, as based on the general theory of sampling, calls for no hypothesis as to the general theory of frequency. It takes the observed distributions and measures the probability that both are samples from a large population. The population may be homogeneous or heterogeneous; provided the samples are truly random samples we obtain a measure of the probability of their common origin.

In the course of a long statistical experience I have learnt that it is wholly impossible to reach any safe conclusions as to the identity or non-identity of populations by any process of mere graphical comparison of frequency distributions.

---

* In actual practice the χ^2 test of "goodness of fit" should always be made with not too fine grouping at the terminals, especially when any group in the tails appears to be contributing largely to the total of χ^2. This point was recognised *ab initio* (*Phil. Mag.* Vol. L. p. 164), and has recently been re-emphasised by Edgworth, *Journal R. Statistical Society*, Vol. LXXVII. p. 198.

The distributions in appearance are wholly dependent on the choice of scales and the eye alone cannot possibly make any measure of the degree of accordance, which will have scientific value.

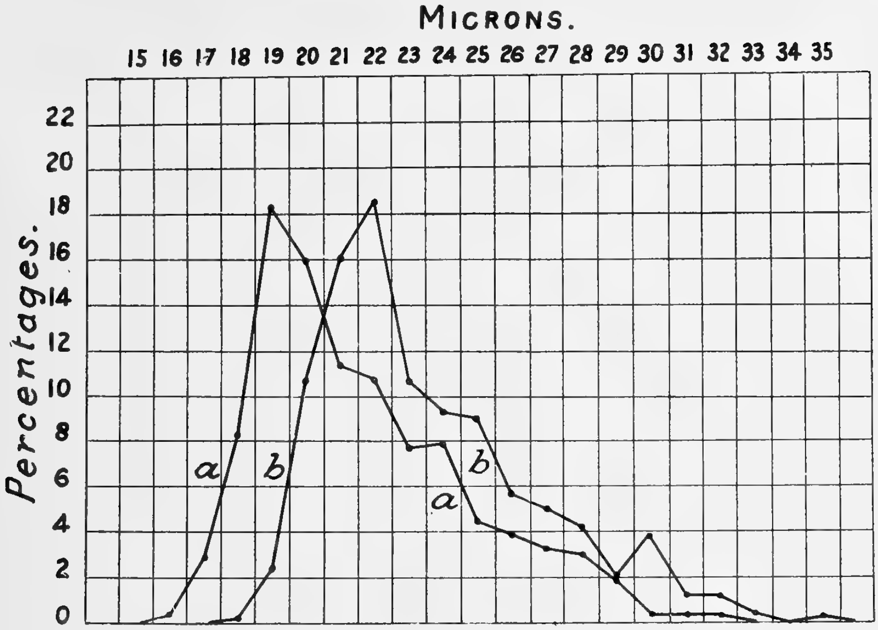
In the accompanying Diagram I. for example, we have the frequency distributions permille of two strains of trypanosomes, (*aa*) from a Donkey and (*bb*) from a Hartebeeste. These we are told are identical. Below (*cc*) and (*dd*) are given the frequency distributions of head-breadths for two races, Egyptian and English women, separated by 7000 years interval. These strains we know to be different, but the eye that judges (*aa*) and (*bb*) to be the "same"* might well suppose (*cc*) and (*dd*) to be also the same. Actually when we come to the quantitative measure of divergence, the probability that (*aa*) and (*bb*) are samples of the same thing is $P <$ ·000,000,1, while the probability that (*cc*) and (*dd*) are the same is $P =$ ·001. In other words it is 10,000 times as likely that Egyptians of 6000 B.C. and the English of 1680 A.D. are the same strain as that the trypanosomes from the Hartebeeste and those from the Mzimba Donkey are of the same strain. Both may indeed be of the "same strain" if a sufficiently wide meaning be given to the term. But is such a racial resemblance as we find between the Prehistoric Egyptian woman and the English woman *diluted* 10,000 *times* what we understand in ordinary language by the "same strain"? All the mathematician can understand by "sameness of strain" is the identity which corresponds to random samples of the same population. If the identity has been modified by a long evolutionary process, by markedly differential environment or treatment, is it not better to have some measure of a scientific nature of the extent of the difference or of the sameness? The eye can never provide any judgment of value on such a point. Especially is this the case if the graphs represent percentages, as the degree of divergence is of course a function of the *number employed to determine the percentages.* A deviation of frequency by percentages based upon samples of 200 might look to the eye absolutely like the deviation of frequency due to samples of 2000, but the scientific measure of the probability of sameness would be widely modified.

That the reader should have evidence how excellent is the test, I have taken the cranial lengths (Flower's measurement) of 67 female skulls dug up in Liverpool Street and compared them with the like lengths of 142 female skulls dug up in Church Lane, Whitechapel. It is possible that both these sets of crania formed part of the contents of plague pits, or there may be an interval in date of a century between them†. Diagram I *bis* shows the data arranged as percentage frequency curves. The χ^2 for 17 groups proceeding by 2 mm. ranges = 19·38, giving $P =$ ·250, or once in four trials, if the material drawn from were the same, we should obtain pairs of samples more divergent than the pair recorded. In other words we can be confident that the Liverpool Street and Whitechapel crania represent persons

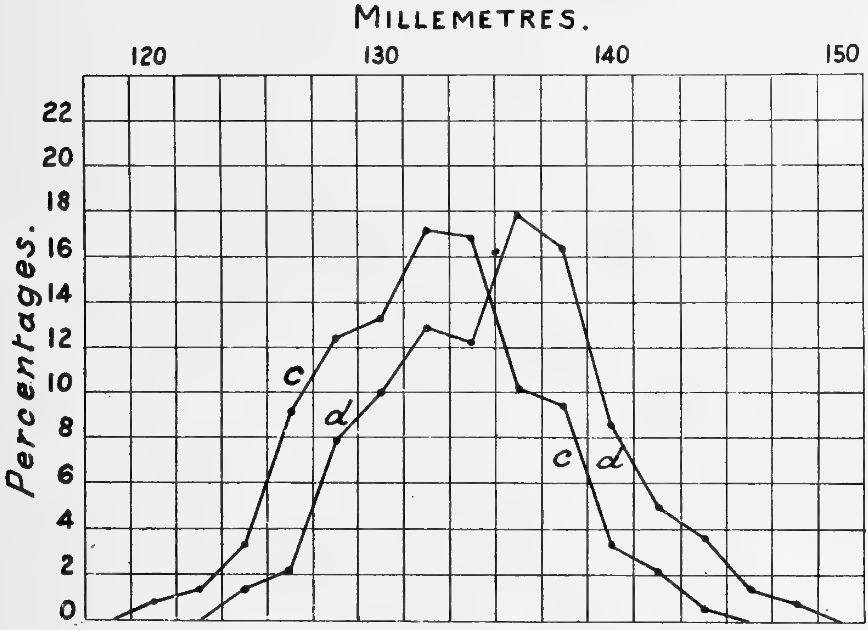
---

* An attempt to define the word "sameness" as used by writers on trypanosome strains would doubtless serve a useful purpose, and emphasise the fact that we can only define "sameness" by appeal to the theory of sampling, or by the adoption of some quantitative measure of the grade of likeness.

† See *Biometrika*, Vol. III. p. 191 and Vol. V. p. 86.

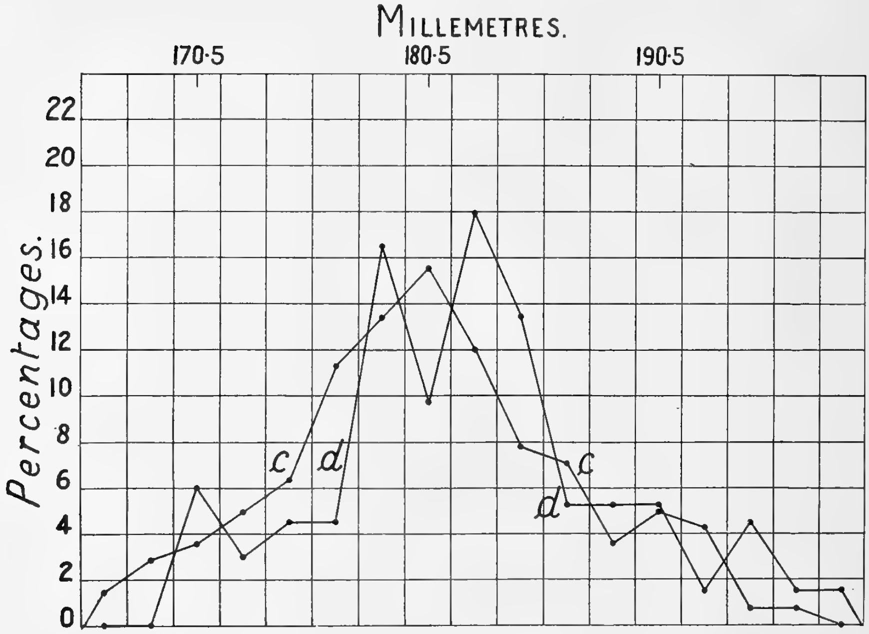


aa. Mzimba Strain of Trypanosome.
bb. Strain from Hartebeeste.



cc. Cranial Breadths of Egyptian Women 6000 B.C.
dd. Cranial Breadths of English Women 1680 A.D.

DIAGRAM I.



cc. Whitechapel ♀'s.
dd. Moorfields ♀'s.

DIAGRAM I *bis.*

of the same strain. That $P = \cdot 25$ and not, say, ·85 may be merely a result of random sampling, or it may arise from some difference of period or social class.

(2) In a long series of papers recently published in the *Proceedings of the Royal Society, Section* B, conclusions are reached as to the identity of various strains of trypanosomes. These conclusions are largely based on a comparison of graphs of the frequency percentages obtained by measurement of hundreds of trypanosomes.

To some extent mean values are given for the different strains, but no arguments whatever can be based on them, for in no case has the probable error of the difference been calculated. Even if it had been calculated, this constant alone would not have sufficed to determine the sameness or difference of the strains. Further, the percentages of various forms in the strains are sometimes given; but again no attempt has been made to determine whether the differences of these percentages are or are not significant. It seems sufficient here to consider the far more valid test of the sameness or diversity of the frequency-distributions as a whole.

I shall divide my investigation into four parts:

(i) The probability of identity of the strains on the evidence presented in the reports of the Commission of the Royal Society, Nyasaland, 1912.

(ii) The probability that the host or the animal in which the trypanosome is cultivated makes essential differences in the distributions of frequency.

(iii) The probability that the strains are alike after allowance has been made for the host.

(iv) The nature of the heterogeneity which is statistically demonstrable in the bulk of trypanosome measurements.

I should like before considering the material to indicate one or two very important points. I am not concerned here with the truth or error of the conclusions drawn by Sir David Bruce and his collaborators. I am only concerned with the nature of the process by which they have drawn their inferences. That process consists in a measurement of the individual trypanosomes and an appeal to the statistics of these measurements—in short to what I should term biometric reasoning. There may well be other means of discussing the resemblances of the different strains of trypanosome,—either by microscopic examinations of divergencies in the life history of the different strains or by differentiation in their action on different hosts, or otherwise. But in the present case the appeal to statistics of measurement has been made. Drs Stephens and Fantham in their paper on *T. rhodesiense* (*R. S. Proc.* Vol. 85, B. p. 227) actually term their work a "biometric study," and the later papers of Sir David Bruce and others are no less "biometric." Now if an appeal be made to statistics, then by a statistical method alone can the answer be given. Further, that method must be the analysis of the modern fully equipped and highly trained statistician. Such a statistician, and he alone, can assert or deny *on the basis of statistics* the probability of any of these strains of trypanosomes being samples of the same population; he alone is in a position to judge the value of the evidence provided by the frequency distributions. If he finds substantial "divergence" where Sir David Bruce and his collaborators assert "sameness," then either statistical theory is wrong, or Sir David Bruce understands by "sameness" something quite different from the "sameness" of the statistician, and something which cannot be judged by the methods of statistics, to which accordingly no appeal should have been made, or only an appeal after a long series of control experiments. The "sameness" postulated by Sir David Bruce is something quite incompatible with the "sameness" found by the statistician when he investigates two samples of 100 crania of the same race or two samples of 1000 blood corpuscles of two series of frogs of the same race. It is what the statistician calls marked divergence and not sameness. If it be asserted that the extreme divergence actually existing between the strains of trypanosomes statistically discussed is due to difference of individual host and not to difference of strain, it will be clear that the divergence and not the sameness ought to have come out of the statistical investigation, and then control investigations ought to have been made to explain that divergence by environmental or other differences. But this is *à priori* to assume the identity of the strains and *à posteriori* to seek an explanation of marked divergence deduced statistically, whereas in the actual papers this great divergence is assumed to be statistical sameness and this sameness used as an argument for identity of strains. The statistician coming to the data critically does not of course assert dogmatically that any two strains are not of identical race. What he does assert is that no argument for the sameness of the strains can be based on the statistics provided; for these actually show wide divergence, and he asks if the strains are *à priori* assumed to be "same," for a full *à posteriori* examination of the sources of the divergence.

The scope of the present paper is not the complete investigation of all the data of the Royal Society Commission, nor an endeavour to obtain from the published data the full conclusions which may be legitimately drawn from them. Its purpose is to illustrate the statistical methods which ought to be applied to such material and to indicate the essential necessity of control experiments on strains known to be the same or accepted as different. A point should be noted here, namely, that I have only found two cases where the strains on the basis of the statistical evidence are said to be different. The first is in the case of *Trypanosoma evansi* and *Trypanosoma brucei.* Sir David Bruce* gives (1911) the frequency distribution of lengths of 820 individuals of *T. evansi* and compares it by means of a graph of percentages with *T. brucei.* The percentages of the latter appear to be deduced from the lengths for two series of 160 trypanosomes and 200 trypanosomes cultivated in a variety of animals (Uganda, 1909, and Zululand, 1894) and published in the preceding year†, but no reference is given in the paper to the original of the percentages in the graph, nor is any demonstration given in the paper of 1910 of the statistical sameness of the Uganda and Zululand strains—there is merely said to be "marked resemblance‡" where the trained statistician finds marked divergence§. Stephens and Fantham‖ use the curve of 1911 to assert that there is a "general resemblance between the curves representing the measurements of these trypanosomes (*T. gambiense, T. rhodesiense, T. brucei*)" and consider that this "general resemblance" shows that "the method is a trustworthy one." It is not clear what "the method" referred to really signifies. The statistical comparison of means and maximum and minimum lengths without statement of probable errors, and the mere graphical examination of frequency curves are wholly inadequate to determine sameness or

---

* *R. S. Proc.* Vol. 84, B, p. 186, 1911.

† *R. S. Proc.* Vol. 83, B, pp. 5 and 11, 1910.

‡ *R. S. Proc.* Vol. 83, B, p. 12.

§ The two distributions are as follows:

	13	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	Totals
Uganda, 1909	—	—	1	2	4	6	10	26	14	14	12	9	12	6	6	12	10	7	1	2	3	3	—	160
Zululand, 1894	4	3	11	11	20	32	17	4	4	3	5	3	7	7	10	13	13	8	10	8	3	1	3	200

These give $\chi^2 = 101 \cdot 18$, leading to $P < \cdot 000,001$, or not once in a million trials would two so divergent distributions be obtained by sampling the same population.

‖ *R. S. Proc.* Vol. 85, B, p. 233, 1912.

divergence. Only a month later than Stephen and Fantham's paper appeared another paper by Sir David Bruce and others* comparing human trypanosomes from Nyasaland with *T. brucei* and *T. rhodesiense* and *T. gambiense.* But the curve for *T. brucei* is wholly different from that of a year earlier. Instead of a minimum at 24 microns there is now a maximum at 24 microns, and the "general resemblance" of *T. brucei* to *T. evansi* is much *increased.* We are now told that *T. rhodesiense* (Stephens and Fantham) is "a distinct species, nearly related to *T. brucei* and *T. gambiense,*" and the conclusion drawn that "the human trypanosome disease of North-east Rhodesia and Nyasaland is not the disease known as Sleeping Sickness in Uganda and the West Coast of Africa†." But the divergence between the frequency distributions of *T. brucei* and the human trypanosome of Nyasaland when accurately measured is of exactly the same order as that which suffices to demonstrate the identity of the human Nyasaland trypanosome and *T. rhodesiense.* Thus the two cases in which divergence is asserted, i.e. (i) *T. brucei* and *T. evansi,* (ii) *T. brucei* and *T. rhodesiense,* seem to be differentiated largely on the base of unanalysed statistical evidence of a nature precisely like that which in other cases is interpreted to mean close "general resemblance" or "sameness." We do not feel that we are in the possession of independent evidence of differentiation which would enable us to test how far statistical divergency corresponds to recognised morphological differences of strain,—a fundamental requisite if we are to interpret as "sameness" a statistical divergence of an extremely high order.

In concluding these introductory remarks we must refer to the types of trypanosome in Nyasaland recognised by Sir David Bruce and his colleagues as distinct on other grounds than numerical measurements. They are:

(i) *T. brucei* vel *rhodesiense.* This is said to be the cause of the human trypanosome disease of Nyasaland. The modal length appears to be 24 to 25 microns‡. According to Bruce and colleagues *T. gambiense* appears to have a mode of 20 microns, but there is evidence for a submode at 26.

(ii) *T. pecorum.* This is said to be the cause of trypanosome diseases of domestic animals in both Uganda and Nyasaland. The modal length varies from 13 to 14§. There is no statistical evidence of bimodality.

(iii) *T. simiae.* This attacks monkey, goat and warthog. Oxen, dogs, white rats, etc., are said to be immune. The length distribution appears to be very homogeneous and with a single mode at 18 microns‖.

---

* *R. S. Proc.* Vol. 85, B, p. 431, 1912.

† *R. S. Proc.* Vol. 85, B, p. 433, 1912. In 1913, however, we find that "there is some reason for the belief that *T. rhodesiense* and *T. brucei* are one and the same species," see Sir David Bruce and others, *R. S. Proc.* Vol. 86, B, p. 407.

‡ *R. S. Proc.* Vol. 84, B, p. 331. Stephens and Fantham's measurements on *T. rhodesiense* suggest modes at 20 and 26. *Ibid.* Vol. 85, B, p. 231. The double mode—roughly 18 to 20 and 28 to 29—appears in the Zululand (1894) and Uganda (1909) strains of *T. brucei. Ibid.* Vol. 83, B, p. 12.

§ *R. S. Proc.* Vol. 82, B, p. 468, and Vol. 87, B, p. 14.

‖ *R. S. Proc.* Vol. 85, B, p. 477, and Vol. 87, B, p. 48.

(iv) *T. caprae.* This is found in waterbuck, ox, goat and sheep. The distribution of length is apparently homogeneous and the mode at 25 microns*.

I leave out of account several forms of trypanosome referred to by Sir David Bruce and colleagues, e.g. *T. vivax, T. uniforme, T. ingens,* etc., of which no large series of measurements were at my disposal.

With the exception of *T. simiae,* which occurs in the warthog, the above trypanosomes appear to be found generally in the wild game and all of them are found in the *Glossina morsitans.* Sir David Bruce and his colleagues suppose the differentiation into these classes to precede the consideration of individual strain, but the exact *modus differentiationis* is not clear from the memoirs.

(3) *Method of Investigation.* The actual formula employed in the present investigation is very simple and can be applied by anyone able to do ordinary arithmetic. If N and N' be the sizes of two samples and the corresponding frequencies:

$$f_2,\ f_2,\ f_3,\ \dots f_p,\ \dots f_s,$$
$$f_1',\ f_2',\ f_3',\ \dots f_p',\ \dots f_s',$$

where f_p, f_p' are the frequencies falling in the p^{th} category, then if

$$\chi^2 = S_1^s \left\{ \frac{NN' \left(\frac{f_p}{N} - \frac{f_p'}{N'}\right)^2}{f_p + f_p'} \right\}$$

be calculated, the probability P that the observed or a greater divergence between the two series would arise from sampling the same population is obtained by determining P from χ^2 by my method of testing "goodness of fit." This method was first published in the *Phil. Mag.*, Vol. 50, p. 157, 1900. The shortest method of actually determining P is by aid of Palin Elderton's tables for P with argument χ^2 issued in *Biometrika*, Vol. I. p. 155, 1902. This is the process used in the measurements of sameness and divergence provided below.

(4) *On the Probability of the identity of the Strains discussed by Sir David Bruce and others.*

(*a*) I take first the question of the "sameness" of the Wild-game strains of trypanosomes as isolated from five antelopes—reedbuck, waterbuck, oribi, and two hartebeeste. Sir David Bruce and others discuss these strains in a paper† of February, 1912, and conclude, apparently from the statistical data, that "the five Wild-game strains resemble each other closely and all belong to the same species."

Now these Wild-game strains have a distinct advantage for they are all obtained from the trypanosomes ultimately taken from the rat as host; they were passed from the infected antelope through healthy goat, monkey or dog, which

---

* *R. S. Proc.* Vol. 86, B, p. 278.

† *R. S. Proc.* Vol. 86, B, p. 407, 1913. In the Table p. 405 for 2500 trypanosomes under the heading 31 microns read a frequency of 33 not 53.

became infected, to the rat. The frequencies of lengths of the trypanosomes in microns were as follows:

From Rat	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	Totals
Hartebeeste (1) ...	—	—	—	1	12	53	80	92	53	46	45	28	25	21	10	19	6	6	2	—	1	500
Hartebeeste (2) ...	1	1	11	30	47	79	51	51	44	37	36	35	28	24	11	3	8	1	—	2	—	500
Oribi	—	1	10	22	77	109	90	57	28	19	23	15	21	14	6	5	2	—	1	—	—	500
Waterbuck ...	—	1	2	8	26	59	74	58	58	44	27	34	33	26	17	19	9	3	2	—	—	500
Reedbuck ...	—	5	30	57	90	81	53	27	17	16	18	23	18	25	18	9	8	2	2	1	—	500
Mzimba (Donkey) Strain	—	2	14	41	91	79	56	53	38	39	22	19	16	15	9	2	2	2	—	—	—	500

I questioned first whether the strains found in the two Hartebeeste were the same; they give

$$\chi^2 = 108{\cdot}69, \text{ and therefore } P < {\cdot}000{,}000{,}1.$$

In other words not once in 10,000,000 trials would two such divergent samples arise if the Hartebeeste strains were samples of the same population. I now compare the Waterbuck and the Oribi; these provide $\chi^2 = 109{\cdot}25$ and $P < {\cdot}000{,}000{,}1$, and again the *extraordinary* divergence, not the sameness, is the statistical feature. The reader may rest assured that equally incompatible results arise when we compare the other antelopes. Statistically we are compelled to assert either that the trypanosome strains in these different antelopes were different species, or that, not only the infected species of antelope, but the individual antelope of the same species (as in the case of the two Hartebeeste) immensely modifies the strain of trypanosome. In short not the "sameness" of the strains, but their great statistical divergence is the fact which impresses itself on the biometrician. No biometrician could possibly accept the view of Sir David Bruce and his colleagues that*:

"It is evident from these tables and charts that the various strains of this trypanosome, as they occur in wild game are remarkably alike. This is what might be expected. Here the trypanosome is at home; it is leading a natural life. It may be supposed to be saved from variation by constantly passing and repassing between the antelope and the tsetse fly."

Our authors, it will be noted, directly appeal for "likeness" of strains to the tables and charts.

With these immense measures of statistical differentiation, we ask: what would be the values of χ^2 and P, if examples of *differentiated* strains of trypanosomes could be found? If differences of host or treatment can produce these wide divergences, how without a preliminary study of the same strain in different hosts and under different treatments can we be certain whether these large divergences mean the same strain differently treated, or different species of trypanosomes?

---

* *R. S. Proc.* Vol. 86, B, p. 406.

(*b*) The next comparison I make is between the Mzimba (Donkey) Strain taken through rats and the above wild-game strains. I have added the data for the Mzimba Strain to the last table (p. 93): it is given by Sir David Bruce and others in a paper on the Mzimba Strain*. I compare the Reedbuck and the Mzimba (Donkey) strains first. We find:

$$\chi^2 = 53{\cdot}37, \quad P = {\cdot}000{,}05.$$

Thus only once in 20,000 trials would a divergence as great as this arise, if the two strains were samples from the same population.

The results of comparing the Mzimba strain with Waterbuck and Hartebeeste (1) give respectively

$$\chi^2 = 114{\cdot}23, \quad P = < {\cdot}000{,}000{,}1,$$

and

$$\chi^2 = 171{\cdot}00, \quad P = < {\cdot}000{,}000{,}1.$$

These give for practical purposes *impossibility* of a common source, thus still further demonstrating that the marked feature of the wild-game and Mzimba strains is *divergence*, not sameness.

Sir David Bruce and his colleagues write† : "The trypanosome of the Mzimba strain is the same species as that occurring in the wild-game inhabiting the Proclaimed Area, Nyasaland." In an earlier paper a diagram‡ is given of the frequency distribution of 3600 trypanosomes of Human strain taken from the rat alone. These are drawn from four native cases of sleeping sickness in Nyasaland and from one European case from Portuguese East Africa. As the individual cases for the rats alone are not given, they have had to be read off the percentage diagram, but the frequencies must be very nearly correct. This Human strain may be compared with the *T. rhodesiense*, the *T. brucei*, the Mzimba (Donkey) strain and a strain obtained from a native woman suffering from "Kaodzera," the so-called sleeping sickness of Nyasaland. The frequencies of these five strains are given in the following table. I first compare the trypanosomes of Nyasaland given as (*b*) above with *T. brucei* and *T. rhodesiense*, for this is the comparison made by the authors themselves§.

Taking the trypanosomes of Nyasaland (*b*) and the *T. brucei* as figured in percentage curves by Bruce and others, we have

$$\chi^2 = 72{\cdot}17, \quad P < {\cdot}000{,}000{,}1,$$

or it is impossible to ascribe any degree of sameness to these two strains. We now compare the Nyasaland strain (*b*) with *T. rhodesiense*, and find

$$\chi^2 = 69{\cdot}95, \quad P = {\cdot}000{,}01\,;$$

---

* *R. S. Proc.* Vol. 87, B, p. 31, 1913.

† *R. S. Proc.* Vol. 87, B, p. 34.

‡ *R. S. Proc.* Vol. 86, B, p. 301.

§ *R. S. Proc.* Vol. 85, B, pp. 431 and 433.

thus once in 100,000 trials two such divergent samples might be drawn. Although there is less divergence than in the case of *T. brucei* and Nyasaland (*b*), it is idle to speak of such a degree of divergence as sameness.

Length in Microns.

	12	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*
Mzimba (Donkey) (*a*) ...	—	—	—	—	2	14	41	91	79	56	53	38	39	22	19	16	15	9
Human, Native Woman (*b*)	—	—	1	4	19	42	63	81	75	91	65	66	93	91	107	110	104	87
Human, mixed (*c*) ...	—	—	—	1	4	46	111	159	219	288	312	365	359	314	314	231	218	198
T. brucei	—	5	8	14	17	40	63	55	66	63	75	87	93	80	82	72	50	38
T. rhodesiense	1	3	10	19	29	35	67	54	92	51	74	56	68	59	85	61	72	50

Length in Microns.—(continued).

	30	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	Totals	Remarks
Mzimba (Donkey) (*a*) ...	2	2	2	—	—	—	—	—	—	—	500	*R. S. Proc.* Vol. 87, B, p. 31. Rats only.
Human, Native Woman (*b*)	49	27	23	13	7	1	1	—	—	—	1220	*R. S. Proc.* Vol. 85, B, p. 427. Various hosts.
Human, mixed (*c*) ...	132	125	90	59	30	13	8	2	2	—	3600	*R. S. Proc.* Vol. 86, B, p. 301. Read from diagram. Rats.
T. brucei	27	26	18	11	4	4	—	—	2	—	1000	*R. S. Proc.* Vol. 84, B, p. 331. Read from diagram
T. rhodesiense	52	28	13	13	5	1	1	—	—	1	1000	*R. S. Proc.* Vol. 85, B, p. 227. Various hosts.

To further establish our point let us compare the Human strain (*c*) for 3600 trypanosomes with the *T. rhodesiense.* Here $\chi^2 = 325{\cdot}47$ leading to $P < {\cdot}000{,}000{,}01$. In other words the great degree of divergence for the case of the Nyasaland native woman is exceeded at least a thousand times, when we take the big example of four natives and one European.

Sir David Bruce and his colleagues write of these strains:

"(1) The trypanosome of the human trypanosome disease of Nyasaland is *T. rhodesiense* (Stephens and Fantham)." In other words the $P = {\cdot}000{,}01$ is interpreted as sameness.

"(2) This is a distinct species, nearly related to *T. brucei* and *T. gambiense*, but more closely resembling the former than the latter." In other words they at this date distinguished between *T. brucei* and *T. rhodesiense**, and as a result of this distinction proposed to call the human trypanosome disease of North-east Rhodesia and Nyasaland by the name "Kaodzera" as not being identical with the sleeping sickness of Uganda and the West Coast of Africa. If we, however, compare *T. brucei* and *T. rhodesiense* we find $\chi^2 = 46{\cdot}83$ and $P = {\cdot}019$. In other

---

* *R. S. Proc.* Vol. 85, B, p. 433, 1912.

words once in about 50 trials we might expect to get two samples from the same population as divergent or more divergent than the distributions found for *T. brucei* and *T. rhodesiense.* We have in fact in the cases of these two trypanosomes reached our first instance of comparative sameness, and the statistics should have shown Sir David Bruce and his colleagues that *T. brucei* and *T. rhodesiense* were relatively the same, and though both differed from the human trypanosome of Nyasaland widely, the approach to *T. rhodesiense* was only slightly closer.

The accordance—speaking in a relative sense—of *T. rhodesiense* and *T. brucei* was asserted by Stephens and Fantham in March, 1912*. In May, 1912, Bruce and others, speaking of the *T. rhodesiense*, term it a distinct species; in February, 1913, they say—although without publishing further frequency distributions—that "There is some reason for the belief that *T. rhodesiense* and *T. brucei* (Plimmer and Bradford) are one and the same species,"† and in a further paper of the same month, "Evidence is accumulating than *T. rhodesiense* and *T. brucei* (Plimmer and Bradford) are identical‡." In May, 1913 (*R. S. Proc.* Vol. 87, B, p. 34), we are told that the Mzimba strain is identical with the wild-game strain and that "it has already been concluded that this species is *T. brucei* vel *T. rhodesiense.*" As far as the statistics of the subject go the only really weighty evidence for the identity is that of 1912, on which, without statistical analysis, the distinction between the two species was asserted.

(*c*) We will next consider the possible identification of *T. gambiense* with *T. rhodesiense* and with *T. brucei.*

The second identification is suggested by Sir D. Bruce and others in the words§:

"Whether these slight differences are fundamental or only accidental it is impossible at present to say, but enough has been written to show that *Trypanosoma gambiense* and *Trypanosoma brucei* approach each other very closely in shape and size."

The following table|| provides the data for *T. gambiense* to be compared with the distribution of *T. rhodesiense* ranging from 12 to 39 in the last table.

Microns.

12	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	Totals
—	1	—	9	21	56	79	114	122	110	85	85	61	47	49	47	44	31	20	11	4	4	—	—	—	—	—	—	1000

The trypanosomes are from a variety of hosts.

For the 28 classes we have, $\chi^2 = 140{\cdot}27$ and $P < {\cdot}000,000,1$. The chief point therefore is the complete divergence, not the resemblance of the two series.

---

* *R. S. Proc.* Vol. 85, p. 233, 1912.

† *R. S. Proc.* Vol. 86, B, p. 407.

‡ *R. S. Proc.* Vol. 86, B, p. 302.

§ *R. S. Proc.* Vol. 84, B, p. 332.

|| *R. S. Proc.* Vol. 84, B, p. 330.

Stephens and Fantham, who term their work a "biometric study," speak of "the general resemblance between the curves representing the measurements of these three trypanosomes (*T. gambiense, T. rhodesiense, T. brucei*)." They continue: "We do not consider, however, that identity of measurement would necessarily imply identity of species. We still believe that the difference in internal morphology, namely the presence of the posterior nucleus, is sufficient to separate *T. rhodesiense* both from *T. gambiense* and *T. brucei*[*]." As a matter of fact the "biometric study" of the data does not indicate identity in the measurements, but confirms the result of internal morphology by proclaiming wide differentiation[†].

(*d*) We can now compare *T. brucei* and *T. gambiense.* Of these Sir David Bruce writes: "Whether these slight differences are fundamental or only accidental it is impossible at present to say, but enough has been written to show that *Trypanosoma gambiense* and *Trypanosoma brucei* approach each other very closely in size and shape[‡]." The biometric commentary on this is that for length of the two series $\chi^2 = 126{\cdot}52$, giving $P < {\cdot}000{,}000{,}1$ and that as far as *size* is concerned the samples differ immeasurably, i.e. far beyond the limits of the calculated tables of P.

We should thus conclude, merely from the statistical evidence, for close sameness in *T. brucei* and *T. rhodesiense* but for marked divergence of both from *T. gambiense.*

---

[*] *R. S. Proc.* Vol. 85, B, p. 233.

[†] In a later section of this memoir I show that Stephens and Fantham have been markedly biased in their judgment of even and odd units of measurement (p. 129 below), and that the recognition of this makes a wide difference in the goodness of fit of my resolution into components to their data for *T. rhodesiense.* It seems desirable therefore to inquire whether this bias affects the test of "sameness" of *T. rhodesiense* with *T. gambiense*, *T. brucei*, and the Human strains (*b*) and (*c*), see the Tables pp. 95—6. The data were accordingly classified into groups of two microns, starting with 12 and 13, 14 and 15, etc., so as to get rid of the even bias as far as possible, and we find:

Strains compared	Old Unit Ranges			New Two Unit Ranges		
	n	χ^2	P	n	χ^2	P
T. rhodesiense and *T. gambiense*	28	140·27	< ·000,000,1	14	118·73	< ·000,000,1
T. rhodesiense and *T. brucei* ...	28	46·83	·019	14	25·76	·018
T. rhodesiense and Human strain (*b*)	28	69·95	·000,01	14	45·92	·000,06
T. rhodesiense and Human strain (*c*)	28	325·47	< ·000,000,01	14	253·37	< ·000,000,01

The bias towards even numbers of Stephens and Fantham has thus not substantially influenced our results, which still show the relative likeness of *T. rhodesiense* and *T. brucei*, and the marked divergence of the former from *T. gambiense* and the human strains.

[‡] *R. S. Proc.* Vol. 84, B, p. 332.

(*e*) It seemed well worth while to investigate how far the two Nyasaland strains of Human Trypanosomes given in the table on p. 95 agree or differ. The first (*b*) of these strains from a native woman of Nyasaland may be compared with (*c*) a compound strain from four natives and a European. We find

$$\chi^2 = 172{\cdot}36$$

giving $P < {\cdot}000{,}000{,}1.$

In other words, the two Nyasaland strains from human beings are indefinitely differentiated. I now compare the Mzimba (Donkey) strain* (*a*) with human strains (*b*) and (*c*), we find:

for (*a*) and (*b*) $\chi^2 = 223{\cdot}16$ giving P certainly $< {\cdot}000{,}000{,}01$;
for (*a*) and (*c*) $\chi^2 = 348{\cdot}55$,, ,, $< {\cdot}000{,}000{,}01$.

Thus the trypanosome strain found in the donkey appears to be absolutely incomparable with that found in man in Nyasaland, just as the strain found in the donkey differed from that found in wild-game.

(*f*) We may now turn to a memoir† by Sir David Bruce and others comparing the Mvera cattle strain, the wild-game strain, and the wild *Glossina morsitans* strain. They give on p. 13 of that paper the graphs for 500 specimens of *T. pecorum*, the wild-game strain, and of the wild *Glossina morsitans* strain taken from a variety of hosts. The following are the frequencies:

Microns.

Strain	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	Totals
Mvera Cattle Strain ...	1	1	15	64	101	136	114	59	8	1	—	500
Wild-Game Strain	—	—	2	34	85	172	119	63	22	3	—	500
Wild *G. morsitans* Strain ...	1	4	16	42	129	147	103	42	15	1	—	500

We compare first Mvera cattle strain with the wild-game strain and find for our 10 categories

$$\chi^2 = 34{\cdot}554, \quad P = {\cdot}000{,}243.$$

This is a relatively low degree of divergence considering that P has been running into 1 in 10,000,000! But it means that if these two strains were samples of one and the same population, we should only expect two such divergent samples to occur 1 in 4000 trials.

---

* This Mzimba strain of trypanosome is discussed in a paper headed: 'Morphology of the various strains of Trypanosome causing Disease in Man in Nyasaland.—The Mzimba Strain' (*R. S. Proc.* Vol. 87, B, p. 26); it is said to be of the Nagana type and is identified by Sir David Bruce and colleagues with *T. brucei* vel *rhodesiense*, the source of the human trypanosome disease.

† *R. S. Proc.* Vol. 87, B, p. 4.

Next we find for Mvera cattle strain and the wild *Glossina morsitans* strain,

$$\chi^2 = 40{\cdot}508, \quad \text{or } P = {\cdot}000{,}008,$$

or only once in 125,000 trials would a pair of samples so divergent arise when testing the same material.

Lastly, testing the resemblances of wild-game strain and wild *G. morsitans* strain, we find

$$\chi^2 = 35{\cdot}41, \quad \text{or } P = {\cdot}000{,}2,$$

not such a gigantic divergency as we have found in many cases, but a difference so great that it only occurs once in 5000 trials requires explanation as divergency and cannot be used as an argument for "sameness."

It will thus be quite clear that as far as the measurements of length go, there is wide divergence to be accounted for between the trypanosomes found in the cattle, the wild-game and the tsetse fly, and that statistically this *divergence* is the remarkable feature. Yet the conclusion of Sir David Bruce and his colleagues, arguing very largely from the frequency distributions, is that "The Mvera cattle strain, the wild-game strain and the wild *G. morsitans* strain belong to the same species of trypanosome, *T. pecorum**."

It will be seen that actual statistical analysis does not in any way confirm the bulk of the conclusions reached by Sir David Bruce and his collaborators. The strains may or may not be ultimately of like origin, but what is quite clear from the analysis is that, if we are to rely on the measurements, then it is the divergence, not the sameness of these strains, which should have been emphasised. No stronger evidence could be deduced of the danger of appeal to statistics when the statistics are not handled by the trained statistician. The mere appeal to the resemblance of frequency curves given in the form of percentages, often based on widely different totals, is an only too common error of medical investigations; it is by no means confined to the Scientific Commission of the Royal Society, Nyasaland. But it has recently become so marked a feature of Series B of the *Proceedings of the Royal Society*, that a vigorous protest is really needful. Thus in the very last part issued (Vol. 87, B, p. 89) occurs a paper on "The Trypanosomes causing Dourine." In this paper there may be microscopic evidence to differentiate the strains *A*, *B* and *C* dealt with; on that I cannot express an

---

* A further conclusion is also reached (*Ibid.* p. 26) "*T. pecorum*, Nyasaland, is identical with the species found and described in Uganda." Unfortunately the species found in Uganda is dealt with in a paper (*R. S. Proc.* Vol. 82, B, p. 468) which provides no frequency distributions, and does not tell us the total number on which the mean length—13·3 microns—is based. The mean value of the *T. pecorum*, Nyasaland is 13·954 (*R. S. Proc.* Vol. 87, B, p. 3) and the standard deviation is 1·393 in microns, thus the probable error of the mean is $\cdot 67449 \times \cdot 0623$. Assuming the Uganda trypanosome to be the same strain and to have the same variability as the *T. pecorum*, Nyasaland, the difference of the means $= \cdot 654$, with a probable error of $\cdot 67449 \times \sqrt{2} \times \cdot 0623 = \cdot 67449 \times \cdot 088$, thus the deviation of the means is 7·73 times its standard deviation. A deviation so great would only occur about once in 4×10^{14} trials, i.e., would be practically impossible if the two strains were identical. Here again it is excessive divergence not sameness which the statistics indicate.

opinion. But on pp. 92—3 percentage frequency curves are drawn for the three strains, and the following remark is made:

> "A survey of the curves obtained by plotting out in percentages the various lengths of trypanosomes encountered in each of the three strains is of interest. It will be observed that in the case of rats the curves of each of the strains correspond fairly closely."

Now what do the authors mean by "fairly closely"? In their conclusions they identify *B* and *C* and differentiate *A*. Unfortunately they have not given their actual frequencies, and I have had to endeavour to reconstruct them from the percentage curves. There results for the rat-data:

Microns.

	16	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	Totals
Berlin Strain *A* ...	1	1	10	9	12	17	17	22	28	48	47	57	55	42	39	37	28	13	8	6	3	500
Frankfurt Strain *B* ...	—	—	1	3	5	1	4	10	20	22*	18*	25	24	35*	23	15	18	15	8	3	—	250
East Prussian Strain *C*	—	1	4	3	6	12	15	22	24	27	28	37	31	16	10	7	5	2	—	—	—	250

We obtain the following results:

Strains *A* and *B*: $\chi^2 = 31{\cdot}11$, $P = {\cdot}0627$,

Strains *A* and *C*: $\chi^2 = 43{\cdot}37$, $P = {\cdot}0034$,

Strains *B* and *C*: $\chi^2 = 72{\cdot}72$, $P = < {\cdot}000{,}001$.

Thus to judge from rats only, *B* and *C* are far more divergent from each other than either is from *A*; in other words the strain *A* is intermediate between *B* and *C* and closer to *B*, from which it is not immensely divergent; two such samples as *A* and *B* might, as far as the length distributions go, be drawn from common material once in 16 trials.

Now of course no one suggests that a conclusion drawn from this rat-material is to replace one drawn from guinea-pig material, but the statistician cannot agree that for rats "the strains correspond very closely"; and he finds it illogical to place the evidence of the rat-data on one side and proceed to draw conclusions from the ocular inspection of the guinea-pig curves, without noticing that the conclusion is markedly opposed to the proper deduction from rat-data. Indeed while the guinea-pig-data† give a relatively high degree of relationship between *B* and *C* ($P = {\cdot}0157$) it is not as high as the rats give between *A* and *B* ($P = {\cdot}0627$); and while the

---

* The values given by the percentage graphs in these cases are respectively 21, 17 and 34, and the total appears to be 247 and not 250 as stated. Either 247 were used or the graph is in error. The three individuals were introduced in a way calculated not to increase divergence.

† The frequency distributions for the guinea-pigs have had to be reconstructed from the percentage curves, the necessary data not being published by the authors.

relationships of A and B ($P < \cdot000{,}000{,}1$) and A and C ($P < \cdot000{,}001$) are very low, the origin of the second hump in the guinea-pig distribution for A requires much more analysis and the certainty by control experiments, that it always repeats itself, and is not the result of hitting a "pocket."

It seems to me that any statistical analysis by modern methods of the trypanosome data compels us to confess that either statistical methods must be discarded entirely in these trypanosome investigations, or they must be pushed to their logical conclusion, and used as the fundamental instrument of research which can guide our enquiries by inference and suggestion when, and when only, it is handled by the trained craftsman. Thus far the use made of statistical methods seems merely to have confused the issues, and brave would be the man who would venture to say after reading this section of our present paper that any two strains discussed by the commission are definitely "same" or certainly differentiated.

(5) *On the Probability that the Animal in which the Trypanosome is cultivated makes essential Differences in the Distributions of Frequency.*

But the very method which casts apparent discredit on the results at present reached seems able to lead us to definite conclusions provided we start with it as the fundamental mode of investigation. Really very little inspection seems to indicate that not only the host but the period of infection materially influences the frequency distribution. These points have not been wholly disregarded by the investigators in this field, but they have had no quantitative measure by which they could appreciate the relative influence of the various environmental factors. Nor indeed could the method be fully applied without experimental observations on trypanosomes of the same strain subjected to differential treatment. Knowing in such cases the quantitative divergence produced, we should be in a position to infer whether two strains from different sources were separate species or merely modified by differential environment. Until we have such quantitative measure no hypothesis of sameness or difference can flow from statistical treatment; nobody as yet knows how much to attribute to environment, how much to attribute to individuality of strain.

In endeavouring to throw light on this matter we are, however, checked at the very start by the absence of effective material. In some cases the period of infectivity is not given; in others we are not always able to break up the total frequency by reference to the host, or to a single host. And even when we merely classify by one type of animal as host, we may have reduced our material to such small numbers that samples may be "same," which on larger numbers would show the marked divergence due to the emphasis of smaller differences*. Some suggestive points can, however, be effectively dealt with and they are treated in the following paragraphs.

---

* It may not be possible to differentiate Bavarian from Würtemberger on samples of 50 crania, although quite possible on samples of 400.

(*a*) I ask what difference is made when a strain is passed through various animals (goat, monkey, dog, rat) or through a single animal alone. Taking the wild-game strain discussed by Sir David Bruce and others*, we have:

Microns.

	10	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	Totals
Wild-Game Strain (from various animals)	—	2	34	85	172	119	63	22	3	500
Wild-Game Strain (from a single rat 510)	1	9	53	137	163	117	14	6	—	500

Here we find $\chi^2 = 65{\cdot}37$ and $P < {\cdot}000{,}000{,}1$. In other words the distribution of lengths of the trypanosomes of the wild-game strain obtained from various animals differs so enormously from that obtained from a single rat that the two cannot be looked upon as samples of the same population. The moment this result is realised we appreciate that (i) it is impossible to compare two strains developed in a variety of animals unless we have previously tested on the same strain the equal valency of these animals, (ii) a series of animals of even the same species may quite possibly give widely divergent results from those obtained for a single animal. Thus passing from a variety of animals in wild-game strain to a variety in wild *G. morsitans* strain makes less difference ($P = {\cdot}000{,}008$)—although great enough—than passing from a variety of hosts to a single rat in the wild-game strain. This rule is not universal, but it illustrates the absolutely essential need for testing the effect of change of host before questioning the identity or non-identity of two strains.

(*b*) I now turn to the Mvera cattle strain, and ask what differentiation is produced by the dog and goat as hosts. The data are very sparse and unless we get a high degree of resemblance may be worth little. They run†:

---

* *R. S. Proc.* Vol. 87, B, pp. 6 and 8.

† *R. S. Proc.* Vol. 87, B, p. 3. I tested the relative interchangeability of goat and sheep in the case of *T. caprae*. The data are as follows: (*R. S. Proc.* Vol. 86, B, p. 280)

Microns.

T. caprae	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	Totals
Goat ...	—	—	3	7	11	35	43	50	33	28	27	17	5	1	—	260
Sheep ...	—	—	—	1	10	12	29	39	31	28	20	5	3	1	1	180

leading to $\chi^2 = 18{\cdot}088$ and $P = {\cdot}1133$ or the resemblance is considerable although not so great as we find between goat and dog for the Mvera cattle strain.

Microns.

	9	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	Totals
Mvera Cattle, Goat ...	1	1	3	14	22	26	19	13	1	100
„ „ Dog ...	—	—	3	11	27	30	21	8	—	100

We have $\chi^2 = 5{\cdot}396$ leading to $P = {\cdot}714$, or in 71 pairs of samples out of 100 from a homogeneous population, we should get more divergent results. It follows therefore that, as far as these small series of this strain go, goat and dog are interchangeable as hosts.

Let us go a stage further and ask whether ox is interchangeable with goat and dog. The following is the frequency distribution for the trypanosomes through the ox:

Microns.

	9	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	Totals
Mvera Cattle, Ox ...	—	—	7	18	33	44	49	21	7	1	180

Compared with the goat strain, this gives

$$\chi^2 = 9{\cdot}559 \text{ and } P = {\cdot}3888,$$

and compared with the dog strain

$$\chi^2 = 9{\cdot}461 \text{ and } P = {\cdot}3973.$$

Thus in about two out of five trials from a same population we should get pairs of samples differing more than the dog and goat strains do from the ox strain. We conclude that while for practical purposes dog, goat and ox strains in the Mvera cattle trypanosomes are interchangeable, yet the dog and goat strain are nearly twice as much alike as the ox strain is to either. Lastly—although it is rather a rash proceeding—I compare rat with goat and dog. It is rash because only 40 trypanosomes through the rat were measured, and this is wholly inadequate for real determination. The frequencies for the lengths are:

	9	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	Totals
Mvera Cattle, Rat	—	—	—	—	2	16	12	10	—	40
„ „ Dog and Goat	1	1	6	25	49	56	40	21	1	200

We find $\chi^2 = 21{\cdot}329$ and $P = {\cdot}0064$. The small series of rat trypanosomes probably accounts for no smaller value of P, but the odds of 155 to 1 are sufficient to show that rat series must not be mixed with series from the goat, dog or ox. This confirms the view obtained for the wild-game strain, that a strain taken through the rat as host is incomparable with strains from other animals.

(*c*) The totals considered for one species of host in (*a*) and (*b*) are rather small. Larger numbers are forthcoming for the so-called Mzimba strain of trypanosomes taken from a donkey at Mzimba. The frequencies are here*:

Microns.

	16	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	Totals
Mzimba Strain, Dog	3	8	17	56	69	67	47	27	22	10	12	7	4	4	4	2	1	360
„ „ Rat	2	14	41	91	79	56	53	38	39	22	19	16	15	9	2	2	2	500

We find $\chi^2 = 25{\cdot}499$ and $P = {\cdot}0619$. Thus only about once in 16 trials should we get such a degree of divergence as the two samples present, drawing them from the same population. This is very far from such a divergence as we have noted in the rat and dog for the Mvera cattle strain, or in the case of rat against other animals in the wild-game strain, which was extremely large. The only explanations that occur to me here are:

(i) In the case of the wild-game strain and the Mvera cattle strain a single rat seems to have provided all the trypanosomes, while in the case of the Mzimba strain two rats were used; this might lessen the influence of individuality.

(ii) In the case of the Mvera cattle strain and the wild-game strain the trypanosomes were ultimately taken from a great number of individuals. In the Mvera cattle case we are told that 32 °/ₒ of the herd were affected, and we have some details of 16 head of cattle and 5 donkeys naturally infected†. In the wild-game case, the wild game affected were very numerous, covering cases of eland, reedbuck, waterbuck, bushbuck, oribi, koodoo, hartebeeste, buffalo and hyaena. Now can we start with the hypothesis that all the individual cattle and all the individual wild game were each bitten by a fly carrying the same strain of trypanosome? Have we any more right to suppose *à priori* that *one* wild-game strain of trypanosome and *one* cattle strain of trypanosome exist, and ask whether these two are identical, than to ask whether the strains carried by hyaena and hartebeeste are the same? We have already (p. 93) seen that the strains from two hartebeeste are extremely divergent. What right have we *à priori* to classify all wild-game trypanosomes together and call them a wild-game strain? And if two antelopes, whether of the same or of different species, give widely different results, why are the trypanosomes of oxen of the same herd or donkeys and oxen from the same neighbourhood to be classed *à priori* as of one species?

---

* *R. S. Proc.* Vol. 87, B, p. 31.

† *R. S. Proc.* Vol. 87, B, p. 15.

If we turn to the Mvera cattle, we find there were *four* sources of trypanosomes for the ox, *two* for the goat, and the *same two* for the dog—these two sources being two of the four cattle sources. There was only one source for the rat, but I have not discovered how far it was identical with one of those for ox or goat*. In the Mzimba donkey strain there was *one* source for dog and rat. In the wild-game strain there were, I make out, *eight* sources of trypanosomes for the goat, *four* for the dog, and only *one* for the rat†.

Thus the individuality, which might be supposed to influence the result, because we are treating of trypanosomes in this case from a single rat, in the Mvera cattle case from a single rat, and in the Mzimba data from only two rats, may really arise from the fact that the rat strains in each case are derived from a single source, while the dog, goat and ox strains show a multiplicity of sources. The troublesome point is that the experimental part of the work has not been designed to answer what seem to me fundamental questions. We cannot directly inquire what difference the host makes because different hosts have rarely been treated with the strain from a unique source. We can say that dog and goat are interchangeable for the Mvera cattle strain, because both drew trypanosomes from the same two sources; but we cannot determine whether the difference in the ox is due to difference of the host, or to the introduction of two more sources. Similarly the divergence between the trypanosomes from rat and from other animals for the wild-game strain may be due to using one rat and therefore one source, and not the many sources of the other animals, or it may really be due to the differentiation of the host. In the same way the difference between the two hartebeeste may be due to individuality in the same species, or to infection from different strains.

(*d*) To some slight extent we may appreciate the effect of individuality by comparing the two rats 512 and 513 in the case of the single source, the Mzimba strain‡.

The frequencies are as follows:

Microns.

Mzimba Strain	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	Totals
Rat 512 ...	—	5	17	37	36	26	25	18	22	14	9	11	9	7	1	1	2	240
Rat 513 ...	2	9	24	54	43	30	28	20	17	8	10	5	6	2	1	1	—	260

The numbers are not as large as we should like; but they give

$$\chi^2 = 17{\cdot}89, \quad P = {\cdot}3306.$$

---

* *R. S. Proc.* Vol. 87, B, pp. 2 and 15.

† *R. S. Proc.* Vol. 87, B, pp. 6 and 8 compared with 5. Rat from p. 8.

‡ *R. S. Proc.* Vol. 87, B, pp. 29 and 31.

Clearly then two samples as divergent as those found would occur on the average once in three trials. It follows that two individual rats are really interchangeable and we note that the extent to which ox is interchangeable with dog or goat for the cattle strain is very much the degree in which two rats are interchangeable. To judge from this single instance, individuality within the same species of host is not very important, and when we find two hartbeeste differing as those considered on p. 93, it seems much more likely, with the information we have at present got, that the hartebeeste were infected with different strains of trypanosome than that their individuality produced the enormous divergence noted. Again the sensible divergence between Mzimba strain in dog and rat on p. 104 is probably due to difference of host, but the enormous difference in the wild-game strain between a single *rat* and dog and goat on p. 103 is probably due to differences in the strains of trypanosomes in the various types of wild game dealt with. We may consider whether the dog and goat data for the wild-game strain differ sensibly. We have*

Microns.

	11	*12*	*13*	*14*	*15*	*16*	*17*	*18*	Totals
Wild-Game Strain, Goat ...	1	16	37	73	38	26	8	1	200
,, ,, ,, Dog ...	—	12	31	57	50	24	6	—	180

Here $\chi^2 = 6{\cdot}04$ and $P = {\cdot}5378$. Thus in more than half the trials we should obtain from homogeneous material pairs of samples more divergent than those for dog and goat. This confirms the view formerly expressed that as far as trypanosomes are concerned dog and goat are interchangeable. We cannot yet say that they are not interchangeable with the rat, as the mixture of strains in dog and goat and the uniqueness of strain in the rat may account for the marked divergence of the latter. Sir David Bruce and his colleagues do not appear to have noticed the wide divergence of the distribution of the rat from the dog and goat either as indicating the heterogeneity of the wild-game and the cattle strains of trypanosomes, or as suggesting such wide differentiation of strain by the host, that rat-material cannot be mixed with that from dog and goat. They do, however, remark of the wild-game strain: "In this the rat is not a suitable animal, since many strains of *T. pecorum* have no effect on it†." This suggests that *T. pecorum* is not homogeneous and that the rat exercises a selective influence on its strains. The suggested rejection of the rat data seems, however, to be based upon the inconvenience of its non-infectivity, and not on what might turn out to be of great importance a selective influence on wild-game or cattle strains. It is not possible to test this selective power in the present instance, as we do not actually know how heterogeneous either the cattle or wild-game material used really was.

* *R. S. Proc.* Vol. 87, B, p. 7.

† *R. S. Proc.* Vol. 87, B, p. 7.

(*e*) If we turn to the *T. pecorum* strain as actually found in the tsetse fly, we see that Sir David Bruce and his colleagues deal with these trypanosomes passed through a variety of animals, of which only goat and dog supply sufficient numbers for any even approximately accurate treatment. The data are as follows*:

Microns.

	9	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	Totals
Wild *G. morsitans* strain: Goat	1	3	12	21	55	60	32	12	4	—	200
,, ,, ,, Dog	—	—	3	14	34	41	40	19	9	—	160
Wild *G. morsitans* strain: Rat	—	1	—	3	22	28	19	6	1	—	80

For goat and dog we find $\chi^2 = 19{\cdot}518$, which give $P = {\cdot}0125$. The resemblance is therefore far less than we have found for goat and dog in other strains, only once in 80 trials from homogeneous material would two samples of such divergent character arise. Before we comment on this it seems desirable to compare the very inadequate rat data.

For rat and goat we have

$$\chi^2 = 12{\cdot}201, \quad P = {\cdot}1434.$$

For rat and dog we have

$$\chi^2 = 11{\cdot}370, \quad P = {\cdot}1245.$$

Accordingly we see that for this material the rat strain (i) lies between the dog and goat strains, and (ii) is definitely interchangeable with dog and with goat, while the dog and goat are much more divergent. Now the sparsity here of all the data must prevent any dogmatism; all we can reach is suggestion for further investigation. But the following points should be noted†. The trypanosomes through the goats were obtained from *six* different goats, infected directly from the wild fly; the trypanosomes from the dogs were obtained from only *four* different sources, namely from a monkey directly infected by the wild fly, from a dog directly infected, and from two goats (89 and 125), the former only of which is identical with one of the former six goat sources. Lastly, the rats were infected from one dog alone, upon which the tsetse flies had directly fed. This dog is not identical with one of the dog sources. Now unless we assume that all the strains of the trypanosome found in the tsetse fly are identical—which is certainly not in accordance with the differences found in the strains of wild game from the "fly-country"—it is by no means certain that the trypanosomes obtained from wild *G. morsitans*, through goat, dog and rat as above noted came from anything like the same sources. Further, the closer resemblance between rat and dog strains

---

* *R. S. Proc.* Vol. 87, B, p. 11.

† *R. S. Proc.* Vol. 87, B, pp. 10, 11, and 19 to 22.

may simply be the result of the rat strain having been developed in the dog as host. The divergence between the dog and goat strain may again be solely due to the greater variety of sources in the goat. The data from the wild *G. morsitans* experiments seem to indicate that the observed divergences between the strain from rat and the strain from goat or dog may not be due to difference of host, but to difference of source from which the material was drawn, and to difference of treatment of the individual stock of trypanosomes, *e.g.* the number of hosts, etc., through which it has passed.

It seems absolutely certain that at the present time most light would be thrown on the conditions for asserting sameness or diversity of strains, by well devised experiments on strains from single sources passed through different species of hosts in different manners, in order to determine the exact measure of divergence produced by host and by treatment, and ultimately to devise a standard treatment for all strains which we desire to compare.

The exact nature not only of host, but of standard treatment is most vital. We can demonstrate the influence of treatment at once by considering the "percentages of posterior nuclear forms among short and stumpy forms" recorded by Sir David Bruce and his colleagues for the wild-game strain*. All the trypanosomes were from rats, and although the date of infection of the rat is, I think, not stated, the dates of first extraction will be after much the same interval, and we can therefore classify by date from first extraction. We find the following table:

Wild-Game Strains.

Percentage of Posterior-Nuclear Forms among Short and Stumpy Forms.

From first Extraction	21 °/。 and under	22 °/。 and over	Totals
6 days and under ...	18	6	24
7 days and over ...	6	18	24
Totals ...	24	24	48

Using Sheppard's formula for the four-fold table, we have for tetrachoric r

$$r = \cdot 707\,;$$

or, the correlation between this character of the trypanosome and the time after infection of extraction is very considerable. It will be obvious that in a standardised treatment this time of extraction will play a most important part. But it again is not independent of the species of trypanosome, for if we take the wild *Glossina morsitans* strains†, we find:

---

* *R. S. Proc.* Vol. 86, pp. 396—404, Tables III, VI, IX, XII and XV.

† *R. S. Proc.* Vol. 86, B, pp. 410—418, Tables III, VI, IX, XII and XV. I have added one percentage by random selection from the complete table by lot in order to give 60 cases, and save labour in fractionising.

Percentage of Posterior-Nuclear Forms among Short and Stumpy Forms.

From first Extraction	7 °/。 and under	8 °/。 and over	Totals
6 days and under ...	12	18	30
7 days and over ...	18	12	30
Totals ...	30	30	60

leading to $r = -\cdot 309$.

In other words using tsetse fly strains and not wild-game strains, but the same host, we find that now the correlation is *negative* or the longer the infection the smaller the percentage. Actually the five *G. morsitans* strains show remarkably irregular results compared with the results for the wild-game strains; the extractions were spread over much the same period, 13 to 14 days on the average, but were somewhat more numerous for the *G. morsitans*. Thus even the same method of extraction may give widely varying results according to the nature of the strain producing the infection, although the host be the same.

To the statistician who examines the frequency distributions provided by Sir David Bruce and his colleagues for both wild-game strains and *Glossina morsitans* strains, there can hardly remain a doubt about the heterogeneity of the material in each case. We have already demonstrated this statistically for the wild-game strains. These strains not only differ by immense differences *inter se*, but *intra se* they are clearly heterogeneous. Whether this heterogeneity is due to the mixture of separate strains, to dimorphism within the strain, or to the combination of material drawn from the rat at various stages of infection, it is not possible on the material at present available to determine finally. The same remarks apply with even greater certitude to the wild *G. morsitans* strains than to the wild-game strains. But we shall return to this point in the last section of this paper. We have already noted that Sir David Bruce and his colleagues identify—against the weight of the statistical evidence—the Mvera cattle strain, the wild-game strain and the wild *G. morsitans* strain as belonging to the same species *T. pecorum**. They had previously identified other strains in wild game, *G. morsitans* and human beings† with *T. rhodesiense* which they elsewhere describe as vel *brucei*‡. This is again, I hold, against the weight of statistical evidence. But it is not clear from the memoirs themselves what is the exact process by which an individual fly, an individual human being, or the blood from a specimen of wild game is credited with carrying a *homogeneous* strain. The sizes are so different in the cases of *T. pecorum* and *T. simiae* that there may be no difficulty in distinction, but the range is so great and to the statistician the material seems so heterogeneous in the case of *T. brucei* vel *rhodesiense* that, perhaps, a fuller description by the authors

---

* *R. S. Proc.* Vol. 87, B, p. 26.

† *R. S. Proc.* Vol. 86, B, p. 42.

‡ *R. S. Proc.* Vol. 86, B, p. 426.

of the process of differentiation would aid him. This is of especial importance if it should turn out, as I suspect, that the trypanosomes classed as *T. brucei* are either dimorphic, or belong to two different species.

In another paper* we find the trypanosomes from *G. morsitans*, on the basis of their infective powers on monkey, goat and dog, resolved into *T. brucei* vel *rhodesiense, T. pecorum, T. simiae* and *T. caprae.* But it is clear that the differentiation was not done solely by infectivity, or there would have been no means of distinguishing *T. brucei* and *T. pecorum* which attack all three—monkey, dog and goat. The question arises, whether *T. pecorum, T. simiae* and *T. caprae* being readily identified by microscopic examination or size, the remainder was classed as *T. brucei,* in which case the question of the heterogeneity of this group, which appears to attack all animals, is rather supported than otherwise by this paper.

Frequencies of the Various Strains for Length.

Length in Microns.

Strain	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*
T. pecorum ...	2	6	42	193	452	618	453	178	51	5	—	—	—	—	—	—	—	—
T. simiae	—	—	—	—	—	7	28	76	93	126	92	47	22	6	2	1	—	—
T. caprae	—	—	—	—	—	—	—	—	—	1	—	3	8	23	49	79	95	80
(i) *T. rhodesiense*	—	—	—	1	3	10	19	29	35	67	54	92	51	74	56	68	59	85
(ii) *T. brucei* ...	—	—	—	—	5	8	14	17	40	63	55	66	63	75	87	93	80	82
(iii) *T. gambiense* ...	—	—	—	—	1	—	9	21	56	79	114	122	110	85	85	61	47	49
(iv) Mzimba Strain	—	—	—	—	—	—	—	8	27	79	175	180	139	109	72	66	36	32
(v) *G. morsitans* ...	—	—	—	—	—	—	7	31	148	230	326	252	237	184	143	115	130	110
(vi) Wild Game ...	—	—	—	—	—	—	1	8	53	118	252	381	348	285	200	162	149	135
(vii) Human Strain	—	—	—	—	—	1	10	41	154	325	494	528	577	512	525	511	464	425
(viii) Chituluka ...	—	—	—	—	—	—	1	8	48	81	78	71	44	46	56	53	98	120

Length in Microns—(continued).

Strain	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	Totals	Source
T. pecorum ...	—	—	—	—	—	—	—	—	—	—	—	—	—	2000	*R. S. Proc.* 87, B, p. 13
T. simiae	—	—	—	—	—	—	—	—	—	—	—	—	—	500	*Ibid.* 85, B, p. 477
T. caprae	68	57	24	9	2	2	—	—	—	—	—	—	—	500	*Ibid.* 86, B, p. 278
(i) *T. rhodesiense*	61	72	50	52	28	13	13	5	1	1	—	—	1	1000	*Ibid.* 85, B, p. 227
(ii) *T. brucei* ...	72	50	38	27	26	18	11	4	4	—	—	2	—	1000	*Ibid.* 84, B, p. 331
(iii) *T. gambiense*	47	44	31	20	11	4	4	—	—	—	—	—	—	1000	*Ibid.* 84, B, p. 330
(iv) Mzimba Strain	24	22	16	7	4	4	—	—	—	—	—	—	—	1000	*Ibid.* 87, B, p. 31
(v) *G. morsitans* ...	127	133	113	96	54	44	11	7	2	—	—	—	—	2500	*Ibid.* 86, B, p. 419
(vi) Wild Game ...	125	110	62	55	33	12	7	3	1	—	—	—	—	2500	*Ibid.* 86, B, p. 405
(vii) Human Strain	372	347	307	198	167	123	77	36	12	11	2	1	—	6220	*Ibid.* 86, B, p. 330
(viii) Chituluka ...	111	128	138	99	117	91	63	27	11	9	1	1	—	1500	*Ibid.* 86, B, p. 291

---

* *R. S. Proc.* Vol. 86, B, p. 422.

At any rate the exact method of differentiation adopted would be of interest to the statistician. The result of the paper is that the four species of trypanosomes occur in quite comparable permilles of tsetse flies caught in the sleeping sickness area of Nyasaland, and there is no evidence to show that they or other strains also may not occur side by side in the same fly or in the same specimen of wild game. Further, these compound strains would then appear in different proportions in the host. Some such hypothesis seems very needful to account for the extreme heterogeneity of the wild game, wild *G. morsitans*, and human strains as recorded by Sir David Bruce and his colleagues. The following table gives a comparison of what appear to be homogeneous strains—*T. pecorum, T. simiae* and *T. caprae*—with what appear *statistically* to be heterogeneous strains, i.e. *T. brucei, T. rhodesiense, T. gambiense,* the Mzimba strain, the wild-game and wild *G. morsitans* strains of human type, and the human strains themselves. The table above, gives the means, standard deviations and coefficients of variation of these strains. It will be seen that the first three are of a very different character to the last five. The variation of the latter is about double that of the admittedly pure strains, and throughout the whole course of our further work this possibility of heterogeneity, and the differential selection of the components by the host must be borne carefully in mind. Great divergences do not discourage the use of biometric methods, and we get occasionally identities of strains which are quite beyond the limits of chance coincidence and which point to definite possibilities if only host, environment, and treatment are once effectively standardised. I propose to try to throw some light on these points in the remaining sections of this paper.

Means, Standard Deviations and Coefficients of Variation of eleven Strains of Trypanosomes.

Series	Mean	Standard Deviation	Coefficient of Variation
T. pecorum	13·992 ± ·019	1·2816 ± ·014	9·16 ± ·099
T. simiae	17·870 ± ·050	1·6558 ± ·035	9·27 ± ·199
T. caprae	25·508 ± ·063	2·1011 ± ·045	8·58 ± ·184
(i) *T. rhodesiense* ...	23·577 ± ·100	4·6764 ± ·071	19·83 ± ·311
(ii) *T. brucei* ...	23·529 ± ·094	4·3938 ± ·066	18·67 ± ·291
(iii) *T. gambiense* ...	22·113 ± ·081	3·7867 ± ·057	17·12 ± ·266
(iv) Mzimba Strain ...	21·413 ± ·063	2·9586 ± ·045	13·82 ± ·212
(v) *G. morsitans* ...	22·695 ± ·058	4·3002 ± ·041	18·95 ± ·187
(vi) Wild Game ...	22·622 ± ·047	3·4541 ± ·033	15·27 ± ·174
(vii) Human Strain ...	23·796 ± ·035	4·1262 ± ·025	17·34 ± ·108
(viii) Chituluka ...	26·172 ± ·084	4·8414 ± .060	18·50 ± ·235

(6) *On the Probability that Strains are alike after allowance for the Host.*

(*a*) Luckily in certain cases the treatment has been more or less alike. Thus in the wild *Glossina morsitans* strain, the tsetse flies brought to the Laboratory from the "fly-country" were in one strain (I) fed on a monkey and in the case of four other strains (II to IV) fed on dogs. From these animals thus infected others were inoculated, but in each case only the trypanosomes from a single rat were used for purposes of measurement and comparison. The following table gives the frequency distributions of the five strains, and chiefly on the basis of these distributions, Sir David Bruce and his colleagues conclude that:

"The five wild *Glossina morsitans* strains resemble each other closely, and all belong to the same species of trypanosome." (p. 421.)

*Wild G. morsitans Strains**.

Microns.

	15	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	Totals
Strain I	—	3	11	25	56	62	75	53	31	28	44	16	23	26	22	12	6	4	1	2	—	500
,, II	—	—	1	4	20	43	67	44	48	34	42	35	33	43	28	23	19	13	2	1	—	500
,, III	—	19	72	84	85	44	24	40	33	27	18	21	12	11	3	4	3	—	—	—	—	500
,, IV	1	5	37	60	71	54	34	25	17	13	15	19	34	30	31	30	8	14	1	1	—	500
,, V	6	4	27	57	94	49	37	22	14	13	11	19	25	23	29	27	18	13	7	3	2	500
Sum	7	31	148	230	326	252	237	184	143	115	130	110	127	133	113	96	54	44	11	7	2	2500

Investigating the statistical measure of resemblance in the usual way we have the following series of results:

Strains I and II: $\chi^2 = 81{\cdot}88$, $P < {\cdot}000,000,1$,
Strains I and III: $\chi^2 = 139{\cdot}71$, $P < {\cdot}000,000,01$,
Strains I and IV: $\chi^2 = 100{\cdot}15$, $P < {\cdot}000,000,1$,
Strains I and V: $\chi^2 = 115{\cdot}77$, $P < {\cdot}000,000,1$,
Strains II and III: $\chi^2 = 328{\cdot}12$, $P < {\cdot}000,000,01$,
Strains II and IV: $\chi^2 = 184{\cdot}88$, $P < {\cdot}000,000,01$,
Strains II and V: $\chi^2 = 208{\cdot}79$, $P < {\cdot}000,000,01$,
Strains III and IV: $\chi^2 = 122{\cdot}79$, $P < {\cdot}000,000,1$,
Strains III and V: $\chi^2 = 147{\cdot}20$, $P < {\cdot}000,000,1$,
Strains IV and V: $\chi^2 = 23{\cdot}90$, $P = {\cdot}2470$.

Statistically therefore there is not the faintest resemblance whatever between any pair of these strains except the IV and V. These strains are for practical purposes interchangeable. In one out of every four trials two pairs of samples of 500 from the *same* trypanosome population would give results more divergent than those observed. But what is the source of this resemblance? Why are these two strains alike and all the others widely divergent? There is nothing whatever in the paper to account for this agreement, and it is the more remarkable because Strains IV and V are to the statistician the most compound looking of all the strains. But some uniformity of origin or treatment has caused the two components to appear in like proportions, and at the back of this resemblance there is some vital point, if we could follow it up. Were the two dogs bitten by the same

---

* *R. S. Proc.* Vol. 86, B, p. 409 *et seq.*

fly, or Rats 658 and 660 really inoculated from the same dog? Clearly there is a point here which ought to be cleared up, for otherwise the statistician could only conclude that the wild *G. morsitans* strains are widely divergent, and that their compound nature suggests that the tsetse fly carries various types of trypanosomes and these in varying proportions.

(*b*) I now turn to the five human strains dealt with by Sir David Bruce and his colleagues. Let us first consider the human strains compounded from various animals. The following table gives the length distributions:

Human Strains. A: Compounds from Various Animals.*

Microns.

	14	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*
Strain I, Mkanyanga ...	1	4	19	42	63	81	75	91	65	66	93	91	107
„ II, E	—	2	2	12	55	108	159	210	188	215	177	138	83
„ III, Chituluka ...	—	1	8	48	81	78	71	44	46	56	53	98	120
„ IV, Chipochola ...	—	2	4	32	68	110	101	109	106	95	95	74	64
„ V, Chibibi ...	—	1	8	20	58	117	122	123	107	93	93	63	51
Sum	1	10	41	154	325	494	528	577	512	525	511	464	425

Human Strains. A: Compounds from Various Animals—(continued).

Microns.

	27	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	Totals
Strain I, Mkanyanga ...	110	104	87	49	27	23	13	7	1	1	—	—	1220
„ II, E	60	34	26	18	8	3	—	1	—	1	—	—	1500
„ III, Chituluka ...	111	128	138	99	117	91	63	27	11	9	1	1	1500
„ IV, Chipochola ...	50	38	26	16	5	3	1	1	—	—	—	—	1000
„ V, Chibibi ...	41	43	30	16	10	3	—	—	—	—	1	—	1000
Sum	372	347	307	198	167	123	77	36	12	11	2	1	6220

We may compare the strains precisely as in the case of the wild *G. morsitans* strains. We find:

Strains I and II: $\chi^2 = 408{\cdot}50$, $P = < {\cdot}000,000,01$,
Strains I and III: $\chi^2 = 204{\cdot}99$, $P = < {\cdot}000,000,01$,
Strains I and IV: $\chi^2 = 180{\cdot}63$, $P = < {\cdot}000,000,01$,
Strains I and V: $\chi^2 = 205{\cdot}40$, $P = < {\cdot}000,000,01$,
Strains II and III: $\chi^2 = 923{\cdot}62$, $P = < {\cdot}000,000,001$,
Strains II and IV: $\chi^2 = 79{\cdot}01$, $P = < {\cdot}000,000,5$,
Strains II and V: $\chi^2 = 77{\cdot}66$, $P = < {\cdot}000,000,5$,
Strains III and IV: $\chi^2 = 531{\cdot}32$, $P = < {\cdot}000,000,01$,
Strains III and V: $\chi^2 = 563{\cdot}82$, $P = < {\cdot}000,000,01$,
Strains IV and V: $\chi^2 = 16{\cdot}81$, $P = {\cdot}7733$.

---

* *R. S. Proc.* Vol. 86, B, pp. 287, 291, 295, and 297. For Strain I see *R. S. Proc.* Vol. 85, B, p. 423.

Again we have the remarkable result that all the human strains are statistically divergent beyond any possible comparison, except those of Chipochola and Chibibi which show a high degree of correspondence. Now is this result the outcome of treatment? We note the following diversity of hosts:

	Strain I.		Strain II.		Strain III.		Strain IV.		Strain V.	
	Gross	Percentage	Gross	Percentage	Gross	Percentage	Gross	Percentage	Gross	Percentage
Men ...	60	4·9	—	0·0	—	0·0	—	0·0	—	0·0
Monkey ...	100	8·2	160	10·7	160	10·7	160	16·0	160	16·0
Goat ...	20	1·6	60	4·0	80	5·3	80	8·0	80	8·0
Sheep ...	60	4·9	20	1·3	—	0·0	—	0·0	—	0·0
Dog ...	260	21·3	260	17·3	260	17·3	260	26·0	260	26·0
Guinea Pig	120	9·8	—	0·0	—	0·0	—	0·0	—	0·0
Rat ...	600	49·2	1000	66·7	1000	66·7	500	50·0	500	50·0
Totals ...	1220	—	1500	—	1500	—	1000	—	1000	—

Now it will be clear at once that the percentages of trypanosomes drawn from various types of host are identical only in the case of Strains IV and V, which we have found in close accordance. But there is not great divergence in source between Strains II and III although Strain I shows fairly wide differences. We find, however, that II and III are statistically very unlike, the next closest resemblances, although very slight, being between II and IV and V. It would not seem therefore that the degree of similarity is *wholly* determined by similarity of hosts. I have accordingly reinvestigated the five human strains by taking rats only. But, of course, even then it is of vital importance to be certain that the process of transfer from man to rat was the same in all five cases, and of this no evidence is provided.

Human Strains. B: From Rat only*.

Microns.

	15	16	17	18	19	20	21	22	23	24	25	26	27
Strain I, Mkanyanga ...	—	1	1	21	40	52	49	30	31	36	33	48	52
„ II, E, Rat 728 ...	—	—	2	4	15	30	57	72	85	72	59	44	26
„ II, E, Rat 726... ...	—	—	2	24	30	42	60	61	87	73	55	27	20
„ III, Chituluka, Rat 952	1	3	21	27	23	15	10	15	19	21	34	44	36
„ III, Chituluka, Rat 953	—	1	17	26	20	19	15	14	26	18	33	40	34
„ IV, Chipochola, Rat 1337	—	—	4	6	16	29	53	61	59	69	56	51	36
„ V, Chibibi, Rat 1660 ...	—	—	—	4	17	29	46	63	69	73	52	40	31
Sum	1	5	47	112	161	216	290	316	376	362	322	294	235

---

* *R. S. Proc.* Vol. 86, B, pp. 288, 289, 292, 293, 295, and 298. For Strain I see *R. S. Proc.* Vol. 85, B, p. 423.

Human Strains. B: From Rat only—(continued).

Microns.

	28	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	Totals
Strain I, Mkanyanga ...	57	57	35	21	18	11	6	1	—	—	—	600
„ II, E, Rat 728 ...	12	15	6	1	—	—	—	—	—	—	—	500
„ II, E, Rat 726	8	4	4	2	1	—	—	—	—	—	—	500
„ III, Chituluka, Rat 952	41	48	28	43	27	23	10	4	5	1	1	500
„ III, Chituluka, Rat 953	37	47	37	33	41	23	12	4	3	—	—	500
„ IV, Chipochola, Rat 1337	31	14	13	2	—	—	—	—	—	—	—	500
„ V, Chibibi, Rat 1660 ...	33	25	11	6	1	—	—	—	—	—	—	500
Sum	219	210	134	108	88	57	28	9	8	1	1	3600

This table with its two pairs of rats inoculated from the same strains is peculiarly instructive. We can compare II, Rat 726, with II, Rat 728.

We find: $\chi^2 = 36{\cdot}195$, giving $P = {\cdot}0048$.

This is far from the high degree of divergence we have found between the compound human strains, but it is not satisfactory as a measure of the agreement of the same strain in two hosts of the same species.

Applying the same test to the two Rats 952 and 953 of Strain III we have:

$$\chi^2 = 14{\cdot}715, \text{ giving } P = {\cdot}9038.$$

This is, of course, quite satisfactory. We should not hesitate to assert identity of strains and of treatment in the case of the trypanosomes from these two rats. The statistician will feel fairly confident that there is a factor of divergence between the trypanosomes of the two rats in Strain II, which does not occur in the two rats of Strain III. He will be almost certain that the strain was not conveyed through the same steps or at the same stage of the disease to the rats in Strain II. Unfortunately dates and processes are not discussed. Sir David Bruce and his colleagues say that it is remarkable how much alike these distributions for Rats 726 and 728 are, and again for the distributions for Rats 952 and 953 that they also closely resemble each other. "It is curious and striking that the same strain of trypanosome growing in two different animals should show this remarkable similarity*." The interesting point is that the statistician would agree with the remarkable similarity in the latter case, but the *divergence* not the remarkable resemblance in the first case would force him to seek for some explanation in treatment. It will, I think, be clear from these illustrations that a strain of trypanosomes, even if obviously compound, can be taken from a single source and after inoculation into two different individuals of the same species be identified as *same*; but to insure this result on every repetition the greatest caution will have to be exercised as to identity of process and treatment.

---

* *R. S. Proc.* Vol. 86, B, pp. 289 and 293.

There are still further results of importance to be ascertained, however, from our table of human strains. Let us compare Strains IV and V, which we found resembled each other closely even for compounded hosts. We now reach

$$\chi^2 = 14{\cdot}035 \text{ and } P = {\cdot}5229.$$

Or, the probability that these two strains are identical has been *reduced* by selecting out the rat data only. But the result is still so high that no one would hesitate to assert that Chipochola and Chibibi were suffering from a disease due to the same strain of trypanosome. The correspondence is so close that we have combined Strains III and V for all other comparisons. In the case of Strain III, we have added together the results for Rats 952 and 953. Such addition is less reasonable for Rats 726 and 728, but without doing this, it is impossible to decide which rat is to represent the E strain. I have then made the following comparisons:

Strains IV and V with III: $\chi^2 = 525{\cdot}67$, $P < {\cdot}000,000,01$.

There is accordingly no similarity at all between the Chituluka strain and that common to Chipochola and Chibibi.

Strains IV and V with II: $\chi^2 = 64{\cdot}70$, $P < {\cdot}000,001$.

Thus the strain from the European E from Portuguese East Africa diverges from the Nyasaland strain widely, but not as widely as that of Chituluka does from those of Chipochola and Chibibi.

Strain I with Strain III: $\chi^2 = 126{\cdot}13$, $P < {\cdot}000,000,1$,
Strain I with IV and V: $\chi^2 = 217{\cdot}82$, $P < {\cdot}000,000,01$.

Thus the trypanosomes from Mkanyanga are widely divergent from those of the three other Nyasaland cases. Nor are they any closer to the European E:

Strain I with Strain II: $\chi^2 = 331{\cdot}37$, $P < {\cdot}000,000,01$.

Thus with the exception of the Chipochola and Chibibi strains, the trypanosome distributions from human sources differ widely. Nor is this to be wondered at, if the human beings owe their trypanosomes to *Glossina morsitans*, for in that case we should expect the human strains to be as diverse as we have found those from the tsetse fly itself. It would remain to explain the close similarity of the Chipochola and Chibibi cases. It would be interesting to know the history of these cases with regard to locality and to the possibility of a unique source of infection.

(*c*) In the case last dealt with, namely that of Chipochola and Chibibi, we have the remarkable feature that the strains although significantly identical, whether treated in the rat alone or in compounded distributions from various hosts, resemble each other somewhat less closely in the single host series. This is not generally the rule. Some of the big divergencies we have already noticed become far less appreciable, nay, even become resemblances when we confine our attention to one species of host. The chief misfortune which then too often arises is the paucity of the total numbers that we have at our disposal. I will consider, however, from this aspect the relations of the three strains wild *G. morsitans*, wild game, and Mvera cattle.

I compare first the lengths of 200 trypanosomes from wild *G. morsitans* and wild-game strains. These yield for the host, goat*:

Microns.

From Goat	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	Totals
Wild *G. morsitans* Strain	1	3	12	21	55	60	32	12	4	—	—	200
Wild-Game Strain ...	—	—	1	16	37	73	38	26	8	1	—	200

giving: $\chi^2 = 26{\cdot}782$ and $P = {\cdot}0015$.

To further test this, I take the same two strains in the dog as host†:

Microns.

From Dog	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	Totals
Wild *G. morsitans* Strain	—	—	3	14	34	41	40	19	9	—	160
Wild-Game Strain ...	—	—	—	12	31	57	50	24	6	—	180

Here $\chi^2 = 7{\cdot}045$ and $P = {\cdot}3171$.

The value we had previously found for a mixture of all strains was $P = {\cdot}0002$. Thus the two strains may be considered as identical when we deal with the trypanosomes from the dog, as showing considerable divergence when we take the goat, and as showing marked divergence when we take a great variety of hosts. The weight of evidence in favour of a standardised treatment thus becomes very great.

Let us look at precisely the same material for the wild-game strain and for the Mvera cattle strain, first for the goat and then for the dog as host‡. The grave difficulty is the paucity of measurements thus differentiated.

Microns.

From Goat	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	Totals
Wild-Game Strain ...	—	—	1	16	37	73	38	26	8	1	200
Mvera Cattle Strain ...	1	1	3	14	22	26	19	13	1	—	100

This gives $\chi^2 = 14{\cdot}670$, leading to $P = {\cdot}1013$.

---

* *R. S. Proc.* Vol. 87, B, pp. 6 and 11.

† *R. S. Proc.* Vol. 87, B, pp. 6 and 11.

‡ *R. S. Proc.* Vol. 87, B, pp. 3 and 5.

Microns.

From Dog	*11*	*12*	*13*	*14*	*15*	*16*	*17*	Totals
Wild-Game Strain ...	—	12	31	57	50	24	6	180
Mvera Cattle Strain ...	3	11	27	30	21	8	—	100

This leads to $\chi^2 = 15{\cdot}992$, $P = {\cdot}0138$.

Previously (p. 98) on the total series of different hosts we had found $P = {\cdot}000{,}243$. Thus by referring our material to individual hosts, we have reduced the degree of divergency between the wild-game and Mvera cattle strains, but it would be still hazardous to state that these strains are identical.

Lastly, we turn to the Mvera cattle strain and the wild *G. morsitans* strain dealing with dog and goat as hosts separately*:

Microns.

From Goat	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	Totals
Wild *G. morsitans* Strain ...	1	3	12	21	55	60	32	12	4	200
Mvera Cattle Strain ...	1	1	3	14	22	26	19	13	1	100

This gives $\chi^2 = 7{\cdot}968$, $P = {\cdot}4368$.

And again:

Microns.

From Dog	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	Totals
Wild *G. morsitans* Strain ...	—	—	3	14	34	41	40	19	9	160
Mvera Cattle Strain ...	—	—	3	11	27	30	21	8	—	100

resulting in $\chi^2 = 11{\cdot}120$, $P = {\cdot}0852$.

The Mvera cattle strain and the *Glossina morsitans* strain had for all hosts a divergence measured by $P = {\cdot}000{,}008$. Thus the great bulk of this divergence is due to multiplicity of hosts†.

To sum up the results obtained for *T. pecorum* in Mvera cattle, wild *G. morsitans* and wild-game strains, the identification of these strains was quite illegitimate on the basis of the compound host frequencies. It is reasonable on the basis of

---

* *R. S. Proc.* Vol. 87, B, pp. 3, 10—11.

† It is worthy of note that in comparisons with the cattle strain the goat appears to give closer results than the dog, but the dog appears the better in the comparison of the *G. morsitans* and wild-game strains.

trypanosomes taken from a single species of host. But how far the resemblance in these cases is produced by a selective influence of the host and not necessarily by an identity of all the members of the strain before transference to the host is not demonstrated.

On the other hand while divergence due to host will account for the divergences which are so notable in *T. pecorum*, it will not account for the divergences in the human strains; these are startlingly conspicuous even if we confine our attention to a single species of host. Precisely the same remarks apply to the trypanosomes similar to those causing disease in human beings found in wild game and in the tsetse fly itself. There must be another source for these divergences.

(7) *Discussion of the Heterogeneity which is statistically demonstrable in the bulk of the Trypanosome Measurements.*

The reader who has attentively followed the course of the argument in the previous sections will be prepared for the next step in this memoir, the attempt to account for the large divergences between strains of trypanosomes in individuals of the same species by the heterogeneity of those strains. My suggestion is that the strain in one fly differs from that in another because the components do not appear in the same proportion, the strain in one specimen of wild game from that in another, or in one man from that in another because they have been bitten by a fly containing the components in unlike proportions. The host does make some difference, either by nutrition or selection of trypanosomes, but it is a minor difference. Thus consider what we may probably hold to be pure strains and observe the average differences in length found by Sir David Bruce and his colleagues:

Microns.

*T. simiae**		*T. caprae*†		*T. pecorum*			
				Mvera Cattle ‡		*Wild G. morsitans* §	
Goat	17·3	Waterbuck	26·8	Donkey	13·5	Goat	13·5
Monkey	18·1	Ox	25·7	Ox	14·2	Monkey	13·6
—		Goat	25·3	Goat	13·8	Dog	14·2
—		Sheep	25·6	Dog	13·8	Guinea Pig	14·6
—		—		Rat	14·8	Rat	14·0
Max. Difference	0·8	Max. Difference	1·5	Max. Difference	1·3	Max. Difference	1·1

We may thus anticipate that in a pure strain the change of host would hardly make a difference of more than 2 microns in the average length. We must

---

* *R. S. Proc.* Vol. 85, B, p. 479.

† *R. S. Proc.* Vol. 86, B, p. 279.

‡ *R. S. Proc.* Vol. 87, B, p. 3.

§ *R. S. Proc.* Vol. 87, B, p. 10.

accordingly be prepared for some such change as this in the shifting of the mean when the host is varied.

We have next to inquire what type of curve accurately describes the strains which we are fairly certain are homogeneous.

If the reader will turn back to p. 110 he will note at once a marked difference between the distributions for *T. caprae, T. pecorum* and *T. simiae* when compared with those entitled Mzimba strain, human strain, wild-game strain, *T. brucei, T. rhodesiense, T. gambiense* and the wild *G. morsitans* strain. The coefficients of variation of the former group are all under 9·5 (mean = 9·00), the coefficients of variation of the latter group are all over 13·5 (mean = 17·29). We recognise therefore a totally different order of variability. Even in absolute variation as measured by the standard deviations we find the first group with its mean S. D. = 1·68 and the second with its mean 3·96. An examination of the graphs scattered through the trypanosome papers to which we have referred will, we think, convince the statistician that we have to deal with heterogeneous and not skew homogeneous material*. It becomes of course important to ascertain whether in the pure strains a Gaussian curve will suffice to describe the frequency closely enough for statistical purposes, for, if it does, the analysis into at any rate two Gaussian components of the heterogeneous strains becomes relatively direct, if laborious. I will consider the *T. pecorum, T. simiae,* and *T. caprae* strains from this standpoint.

(*a*) *T. pecorum* (see p. 110).

Mean = 13·992 microns. S.D. = 1·2816 microns.

Microns	Observed Values	Calculated Values
9 and under	2	0·46
10	6	5·98
11	42	45·41
12	193	192·52
13	452	456·70
14	618	607·12
15	453	452·49
16	178	188·98
17	51	44·16
18 and over	5	6·20

$\chi^2 = 7·630$
$P = ·572$

Hence in 57 out of 100 trials from material following the Gaussian distribution a more divergent sample than that observed would actually be obtained. We can therefore conclude that a simple Gaussian frequency adequately describes the distribution in size of *T. pecorum*. This is illustrated in Diagram II.

---

* Note especially the bimodal graphs in *R. S. Proc.* Vol. 83, B, pp. 5 and 11, for both the Uganda and Zululand strains of *T. brucei*, in Vol. 86, B, pp. 291—293, for human strains, in Vol. 86, B, pp. 395, 397 for wild-game strains and pp. 409, 411, 417 and 419 for *G. morsitans* strains.

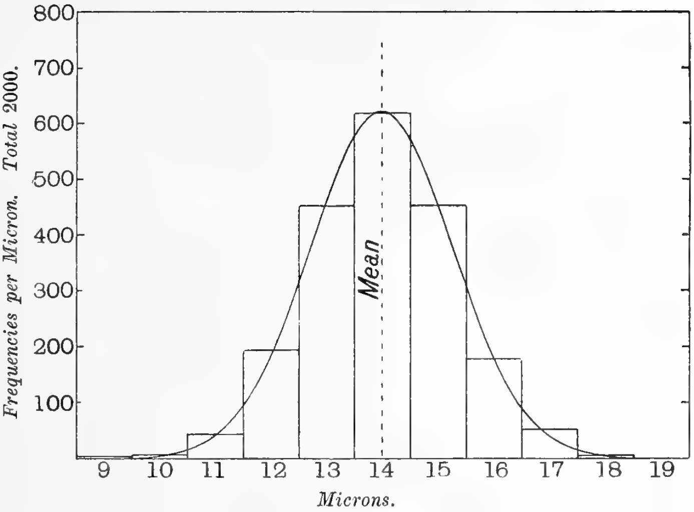


DIAGRAM II. Gaussian fitted to *T. pecorum* Frequency.

(*b*) *T. simiae* (see p. 110).

Mean = 17·870 microns. S.D. = 1·6558 microns.

Microns	Observed Values	Calculated Values
14 and under	7	10·46
15	28	27·63
16	76	63·92
17	93	103·78
18	126	118·32
19	92	94·66
20	47	53·18
21	22	20·96
22	6	5·80
23 and over	3	1·29

$\chi^2 = 8{\cdot}149$

$P = {\cdot}520$

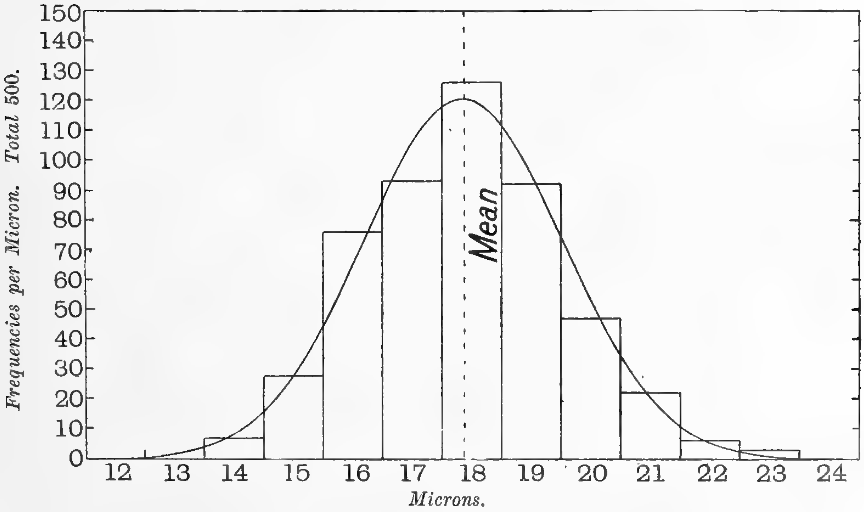


DIAGRAM III. Gaussian fitted to *T. simiae* Frequency.

We conclude that the Gaussian adequately describes the distribution of *T. simiae.* In more than half the trials we should get a worse sample. See for graphical fit, Diagram III.

(*c*) *T. caprae* (see p. 110).

Mean = 25·508 microns. S.D. = 2·1011 microns.

Microns	Observed Values	Calculated Values
20 and under	4	4·28
21	8	9·82
22	23	23·95
23	49	46·74
24	79	73·05
25	95	91·38
26	80	91·54
27	68	73·45
28	57	47·16
29	24	24·26
30	9	9·98
31 and over	4	4·38

$\chi^2 = 5{\cdot}175$

$P = {\cdot}921$

This is a still more excellent fit; if the Gaussian represented the population, in 92 °/ₒ of samples we should get a more divergent sample than that observed. The curve is given in Diagram IV.

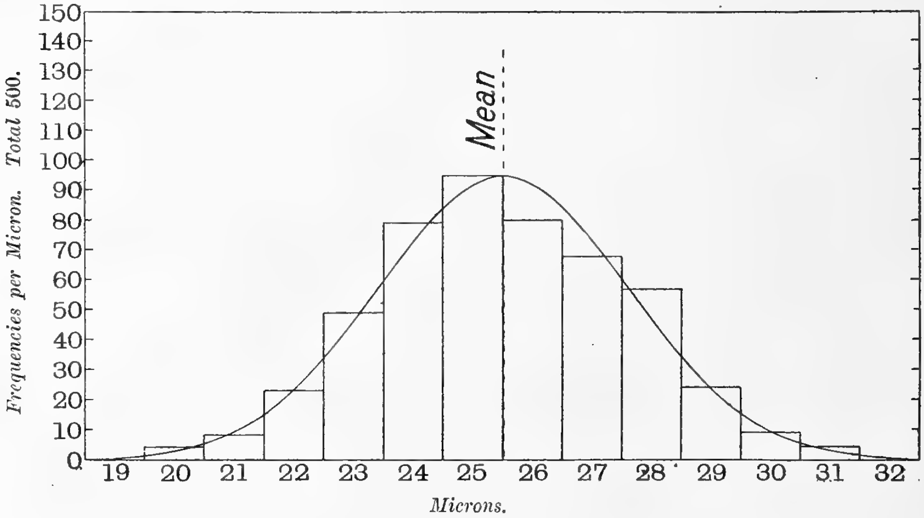


Diagram IV. Gaussian fitted to *T. caprae* Frequency.

It will be clear from the above three illustrations of what we may term homogeneous trypanosome strains that the Gaussian curve of frequency suffices to describe adequately such material. It is equally clear that no Gaussian can possibly describe such skew distributions as we get in the wild-game strain or wild tsetse fly strain of the trypanosome species identified by Sir David Bruce and colleagues as *T. rhodesiense**. It is equally impossible in the case of the human strains figured in the paper of February 1913†. I illustrate this on the frequency distribution for 6220 trypanosomes of human strains‡.

	Observed	Calculated		Observed	Calculated
14 and under	1	75·45	*26*	425	520·55
15	10	62·51	*27*	372	444·42
16	41	101·51	*28*	347	357·96
17	154	155·50	*29*	307	271·88
18	325	224·73	*30*	198	194·81
19	494	306·27	*31*	167	131·68
20	528	393·73	*32*	123	83·91
21	577	477·39	*33*	77	50·44
22	512	545·93	*34*	36	28·61
33	525	588·91	*35*	12	15·30
24	511	599·17	*36* and over	14	14·18
25	464	575·04			

Here $\chi^2 = 501$ and $P < \cdot000,000,001$. In other words description by a Gaussian is absolutely impossible. The histogram of observations and the curve are shewn on Diagram V.

Now the suggestion that flowed at once from these results was the compound nature of *all* the material classed under the headings:

(i) *T. rhodesiense.*
(ii) *T. brucei.*
(iii) *T. gambiense.*
(iv) Mzimba Strain.
(v) Wild *G. morsitans* Strain.
(vi) Wild-Game Strain.
(vii) Human Strain.

With the experience of the Gaussian fitting the homogeneous strains, the direct step was to investigate whether the above material could be analysed into two Gaussian components and to determine how nearly these components were in agreement. The method of carrying out this analysis was provided in the first of my series of *Contributions to the Mathematical Theory of Evolution*§. There was nothing to prevent the process being applied to every individual frequency given by the trypanosome workers, except the very laborious arithmetic. The method was applied to the above seven cases, and also (viii) for the purposes of illustration to a single human case, that of Chituluka, a native of Nyasaland, who

---

* See *R. S. Proc.* Vol. 86, B, pp. 407 and 419.

† See *R. S. Proc.* Vol. 86, B, pp. 285 *et seq.*

‡ See *R. S. Proc.* Vol. 86, B, p. 300.

§ *Phil. Trans.* Vol. 185, A, pp. 71—110, 1894.

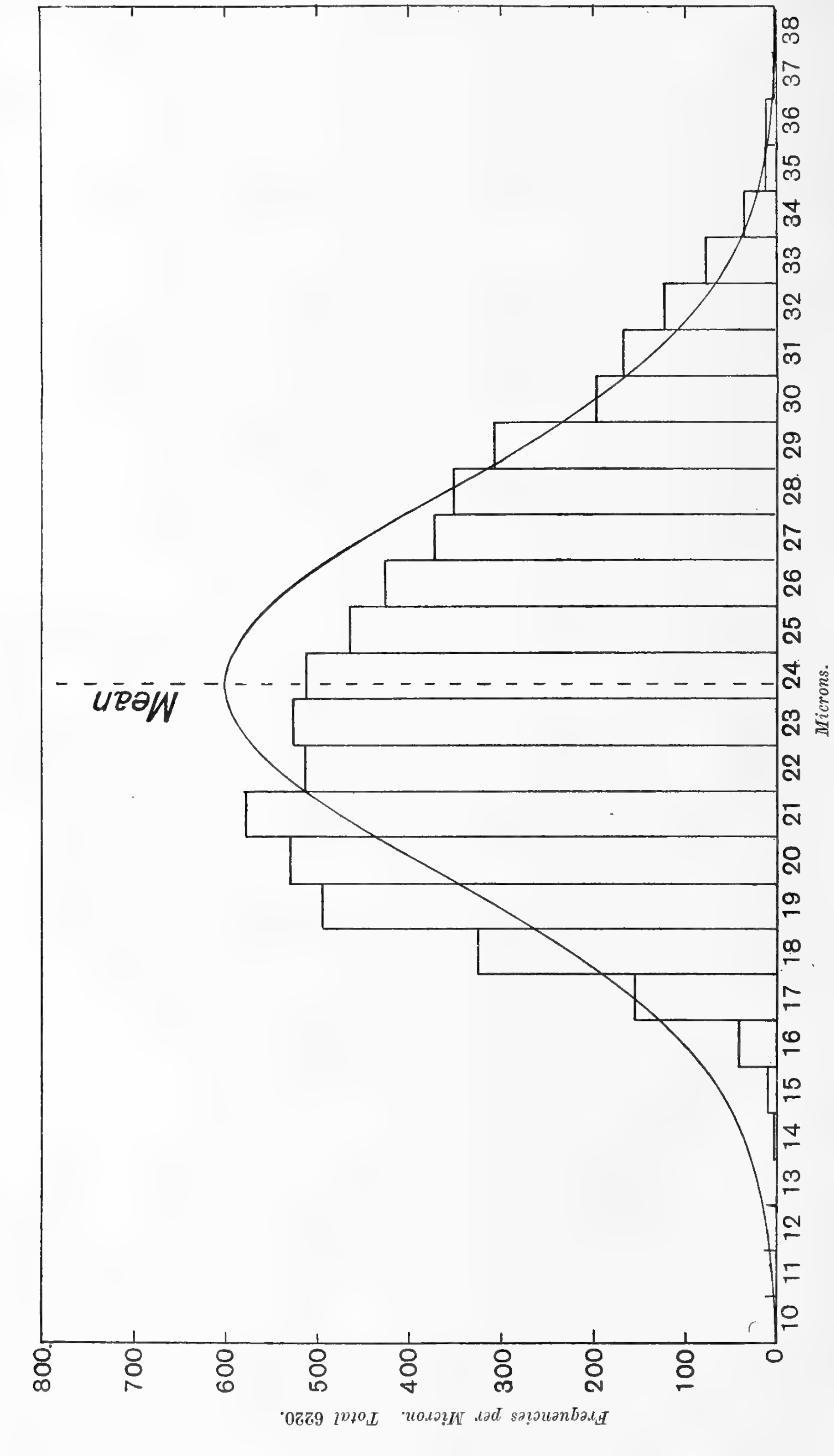


DIAGRAM V. Failure of Gaussian to fit Frequency Distribution of Human Trypanosomes.

died of sleeping sickness*. With the single exception of *T. brucei* every one of these distributions broke up into two components, and into two components with strikingly close means. I propose to call these two components *T. minus* and *T. majus.* I do not assert that they are distinct species; they may be dimorphic groups of one and the same trypanosome species. But the recognition of their existence seems to bring some order at least into the chaos we have already noted as existing in the trypanosome measurements. Two human strains or two wild-game strains differ from each other with such wide divergence in their frequencies because these two groups *T. minus* and *T. majus* are mixed in the individual in different proportions.

Strain	Means		Standard Deviations		Coefficients of Variation		Size of Populations	
	T. minus	*T. majus*	*T. minus*	*T. majus*	*T. minus*	*T. majus*	*T. minus*	*T. majus*
T. rhodesiense ...	18·7418	26·1122	2·3184	3·4397	12·370	13·173	343·95 34·4 %	656·05 65·6 %
T. brucei	[19·8244	26·1122	2·6439	3·4134	13·337	13·072	410·83 41·08 %	589·17 58·92 %]
T. gambiense ...	19·8926	26·2463	2·0566	2·6260	10·339	10·005	650·5 65·1 %	349·5 34·9 %
Mzimba Strain ...	19·8966	24·0508	1·3961	3·1028	7·017	12·901	634·96 63·5 %	365·04 36·5 %
G. morsitans Strain	19·6475	27·1966	1·7503	2·7013	8·908	9·932	1490·7 59·6 %	1009·3 40·4 %
Wild-Game Strain	20·4418	25·8263	1·6332	2·8799	7·990	11·151	1487·7 59·5 %	1012·3 40·5 %
Human Strain ...	20·3687	26·2930	1·9444	3·4470	9·536	13·110	2621·4 42·15 %	3598·0 57·85 %
Chituluka	19·8410	28·7875	1·9785	2·8823	9·972	10·012	438·53 29·24 %	1061·47 70·76 %
Means	19·8315	25·9542	1·8498	3·0328	9·360	11·712	—	—
T. simiae	17·870	—	1·6558	—	9·270	—	500 100 %	—
T. caprae	—	25·508	—	2·1011	—	8·580	—	500 100 %

The table below gives the chief biometric characters of *T. minus* and *T. majus* as found from the seven resolutions. The mean values of the constants for *T. minus* and for *T. majus* are placed at the foot; in calculating these mean values, Chituluka's data have been excluded as already included in the human strain, and also those for *T. brucei* not directly resolved.

At the foot of the table I have placed the constants for *T. simiae* and *T. caprae,* the nearest pure strains to *T. minus* and *T. majus* respectively. I do not in the

---

* *R. S. Proc.* Vol. 86, B, p. 291.

least suggest there is any identity, but comparison may bring home to the trypanosome worker the average sizes of the two components*. The differences of the variabilities are, however, much larger, and the influence of host on variability as well as on mean ought to be studied.

It will be seen at once that the divergence in the individual means of *T. minus* from the general mean is very slight, at most a micron, and well within the limits which arise, as we have seen, from difference of host. It is a most remarkable fact that from six independent reductions the mean size of *T. minus* should come out so nearly 19·8 microns. In *T. majus* the correspondence is not so good; the average of about 26 microns falls to 24 in the Mzimba strain and rises to 28·8 in the case of Chituluka†. Still it does not appear to me that these changes of mean of the *T. majus* component are absolutely beyond the variation due to differences of host and treatment. Another more serious matter is the comparatively wide range found for the variabilities; but even here it is impossible to assert that such differences will not occur with difference of host. For example the Mvera cattle strain, a fair sample of the simple *T. pecorum*, gives:

Host	Mean	Standard Deviation	Coefficients of Variation
Goat ...	13·80	1·462	10·592
Rat ...	14·75	·839	5·689
Dog ...	13·79	1·087	7·885

Here while the means are within one micron, the differences in variability are of the same order as those found in *T. majus* from different hosts.

Again, taking a pure homogeneous strain as *T. caprae* with goat and sheep as host, which are scarcely so differentiated as man and antelope, we find:

Host	Mean	Standard Deviation	Coefficients of Variation
Goat ...	25·31	2·187	8·642
Sheep ...	25·60	1·923	7·512

Lastly, taking *T. simiae* for goat and monkey we have:

Host	Mean	Standard Deviation	Coefficients of Variation
Monkey ...	17·26	1·403	8·127
Goat ...	18·11	1·687	9·315

---

* The maximum average length of *T. caprae* is 26·8 in the waterbuck and of *T. simiae* 18·1.

† It should be noted that with the *whole* of the human data the mean is 26·33 and that Chituluka's mean is very exceptional.

I think we may conclude that, allowing for the errors of random sampling and the errors arising from the resolving process, the deviations observed in the variability of our two components do not invalidate the hypotheses:

(i) That the widely divergent results obtained from different strains are due to the existence in the same individual of two types of trypanosome with very varying percentages from individual to individual.

(ii) That one of these types has a mean length of about 19·8 microns and a variability of about 1·8 microns, the other a mean of about 26·0 microns and a variability of about 3·0 microns. The means may vary 1 or 2 microns with the nature of the host and the variability 0·5 to 1 micron.

The large type predominates in the Nyasaland human strains*, on the average in about the ratio of 3 to 2, but the smaller type predominates in the *G. morsitans* and wild-game strains in about the same ratio; while in the trypanosomes classed as *T. rhodesiense,* and *T. gambiense* as well as in the strain from the Mzimba donkey the preponderance is still of the smaller type and the ratio approaches 13 to 7. Whether these ratios are peculiar to the host or due to the infecting fly, it is not at present possible to determine. But the hypothesis of the existence of these two types,—whether as a dimorphism of *T. rhodesiense* or as independent species seems to bring some order into the apparent chaos of recent trypanosome measurements.

The following paragraphs give the calculated constants of the reductions, and the numbers of the diagrams showing the nature of the compound frequencies:

(i) *T. rhodesiense.*

$$\text{Mean} = 23{\cdot}577,$$

$$\mu_2 = 21{\cdot}86874, \qquad \mu_4 = 1079{\cdot}10255,$$

$$\mu_3 = +\,4{\cdot}01986. \qquad \mu_5 = +\,1105{\cdot}74834.$$

Reducing nonic:

$$24q^9 - 298{\cdot}7232q^7 - {\cdot}5817q^6 + 1114{\cdot}7684q^5 + 34{\cdot}7620q^4$$
$$- 1179{\cdot}2495q^3 + 12{\cdot}9808q^2 + {\cdot}0891q + {\cdot}0001 = 0$$

where† $p_2 = -10q$.

The root is $p_2 = -12{\cdot}2578$. This leads to the two components in the Table p. 125. The histogram of the observations and the two component Gaussian curves with their compound are given in Diagram VI.

The resolution is not a very good one; for 24 groups $\chi^2 = 37{\cdot}48$, and $P = {\cdot}05$, or once in 20 trials only we should get a worse result. But an examination of either the graph or the original frequency shows at once the cause of this divergence. *In their measurements Drs Stephens and Fantham have had a strange bias in favour*

---

* The European from Portuguese East Africa had predominance of *T. minus.* See *R. S. Proc.* Vol. 86, B, p. 288.

† Notation of the memoir *Phil. Trans.* Vol. 185, A, p. 84, Eqn. (29).

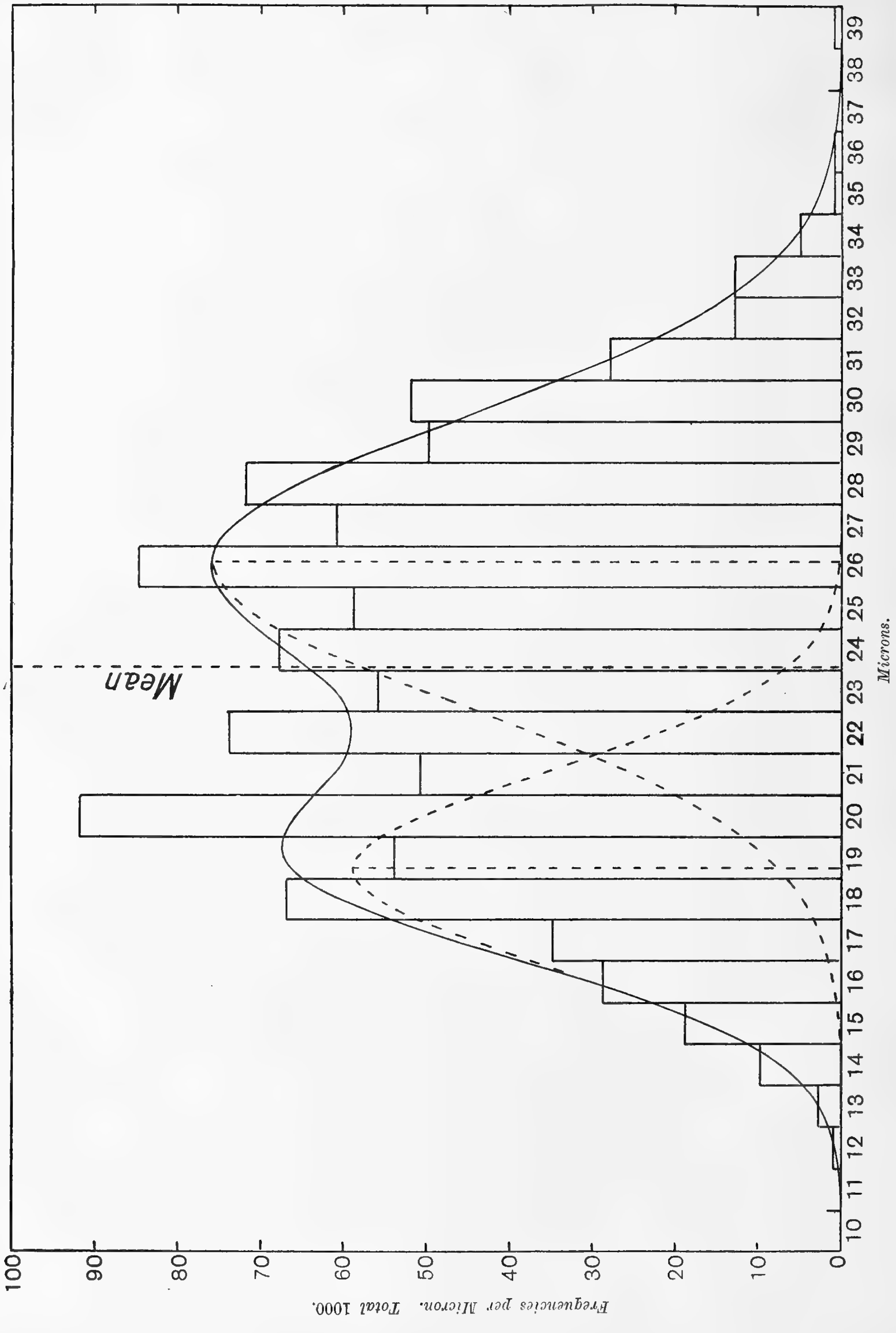


DIAGRAM VI. Resolution of the Frequency of *T. rhodesiense* into *T. minus* and *T. majus.*

of even numbers. No curve whatever could fit the data satisfactorily under the circumstances! Either they used a scale graduated to 2 microns only, and had a prejudice in favour of the scale markings, or else their even numbers were in some way more conspicuous than their odd. Whatever the source of this peculiarity may be, there can be no doubt of the bias*.

The only way to obtain a reasonable measure of the goodness of fit in Stephens and Fantham's results for *T. rhodesiense* is to group from 10 to 12, 12 to 14 and so on in comparing the observed and calculated frequencies. If this be done we find $\chi^2 = 5{\cdot}03$ for 13 groups and $P = {\cdot}957$, a splendid fit. The frequencies are as follows:

	10–14	*14–16*	*16–18*	*18–20*	*20–22*	*22–24*	*24–26*	*26–28*	*28–30*	*30–32*	*32–34*	*34–36*	*36–38*	Totals
Observed	9	38·5	83·0	133·5	134·0	127·0	135·5	139·5	112·0	60·5	22·0	4·0	1·5	1000
Calculated	7·17	34·67	92·99	132·91	124·79	124·28	146·35	145·56	106·55	56·22	21·36	5·84	1·17	999·85

(ii) *T. brucei.* The data for this trypanosome were taken from Sir David Bruce and colleagues' diagram†. I have not come across the original publication with the measurements involved in this diagram. Describing this species in July 1910‡, the authors speak of its well-marked dimorphism. This is very obvious in the graphs for length given for the Uganda 1909 and Zululand 1894 strains, but the numbers given are far too slender (160 and 200 respectively) to justify any attempt at analytical resolution. Graphically we may take it that roughly the following are the means of the components:

	T. minus.	*T. majus.*
Uganda 1909	20 microns	28 microns
Zululand 1894	18 microns	29 microns.

These are not very widely divergent from the values

19·8 microns 26·0 microns

we have found from the seven resolutions.

In May 1911 § the two curves for Uganda and Zululand appear to be added together to give a *T. brucei* curve of length distribution. This is again markedly bimodal with one component mean at 18·75 microns and the other at 27·5 microns, both approximative. Thus far *T. brucei* appears quite well to fit in with our other material. But in September 1911 appears the diagram of *T. brucei* said to be

---

* Bias of this or of a similar character is not uncommon—even in the pages of this Journal. I remember once pointing out to a Scotch anthropometer his prejudice in favour of whole centimetres. He looked at his results, recognised the bias, and then gravely told me that it was not due to any personal bias, but that the Creator must have designed Scotsmen on the metric scale!

† *R. S. Proc.* Vol. 84, B, p. 331.

‡ *R. S. Proc.* Vol. 83, B, p. 2.

§ *R. S. Proc.* Vol. 84, B, p. 186.

based on 1000 individuals. Here there is a mode about 24·0, with possibly a submode at 19 microns, but the evidence for dimorphism has largely disappeared. It is very desirable that we should know the details of this curve, i.e. the nature of the hosts and so forth, for it apparently replaces the earlier data and remains the standard *T. brucei* distribution. It certainly shows nothing of the definite heterogeneity (or dimorphism) of the previous Uganda material.

Its constants are as follows:

$$\text{Mean} \quad 23{\cdot}5290,$$

$$\mu_2 = 19{\cdot}30583, \qquad \mu_4 = 996{\cdot}87764,$$

$$\mu_3 = 10{\cdot}54837, \qquad \mu_5 = 2146{\cdot}37930.$$

$$24q^9 - 101{\cdot}8618q^7 - 4{\cdot}0057q^6 + 140{\cdot}6937q^5 + 62{\cdot}0835q^4$$
$$- 29{\cdot}3940q^3 + 11{\cdot}2371q^2 + 1{\cdot}4413q + {\cdot}0331 = 0.$$

No suitable root of this equation exists and accordingly it would appear that this distribution is not rigidly reducible to Gaussian components. This result is so remarkable in view of the obviously bi-modal character of the earlier *T. brucei* distribution, and the resolution into two components of all the other seven distributions, said to be allied to *T. brucei*, that I determined to consider the matter further by fitting Gaussians to the 'tails' of the *T. brucei* distribution*. I chose as the right-hand 'tail' the frequency from 28 to 38 inclusive, and as the left-hand 'tail' the frequency from 13 to 18 microns inclusive. The two resulting components were:

T. minus.	*T. majus.*
$\overline{m}_1 = 20{\cdot}0817\ (19{\cdot}83)$,	$\overline{m}_2 = 26{\cdot}4359\ (25{\cdot}95)$,
$\sigma_1 = 2{\cdot}8685\ (1{\cdot}85)$,	$\sigma_2 = 3{\cdot}6399\ (3{\cdot}03)$,
$n_1 = 628{\cdot}16$,	$n_2 = 467{\cdot}52$.

The totals populations for each component are clearly not very good and their combination exceeds by 9·6 °/₀ the total observed population; but the means are not widely divergent from the average values resulting from our six resolutions, as the numbers given in brackets testify. Accordingly I determined to select the means of the components at values near the mean values of six reductions, and after one or two slight betterments, determine the sizes of the populations and their standard deviations so as to give the mean, and second and third moments of the observed population. These provided:

T. minus.	*T. majus.*
$\overline{m}_1 = 19{\cdot}8244$,	$\overline{m}_2 = 26{\cdot}1122$,
$\sigma_1 = 2{\cdot}6439$,	$\sigma_2 = 3{\cdot}4134$,
$n_1 = 410{\cdot}83$,	$n_2 = 589{\cdot}17$.

---

* *Biometrika*, Vol. II. p. 1 and Vol. VI. p. 65.

The following table gives the observed and calculated values:

Microns	Observed	Calculated	Microns	Observed	Calculated
13	5	3·44	*26*	82	72·74
14	8	5·80	*27*	72	67·98
15	14	12·25	*28*	50	59·71
16	17	22·79	*29*	38	48·10
17	40	37·05	*30*	27	36·04
18	63	52·87	*31*	26	24·79
19	55	66·68	*32*	18	15·67
20	66	75·44	*33*	11	9·09
21	63	78·43	*34*	4	4·84
22	75	77·49	*35*	4	2·37
23	87	75·61	*36*	—	1·75
24	93	74·71	*37*	—	
25	80	74·36	*38*	2	

From these results we find $\chi^2 = 29{\cdot}92$ and $P = {\cdot}22$. Thus more often than once in five trials we should get a worse divergence than the observed, if the sample were taken from the calculated population. Some endeavour was made to better the fit by small variations from the above solution, discussed by least squares, but no improvement was effected. The two components are represented in Diagram VII (p. 132).

(iii) *T. gambiense.*

$$\text{Mean} = 22{\cdot}1130,$$

$$\mu_2 = 14{\cdot}3389, \qquad \mu_4 = 531{\cdot}3585,$$

$$\mu_3 = 29{\cdot}1104, \qquad \mu_5 = 2429{\cdot}0948.$$

Reducing nonic:

$$24q^9 - 71{\cdot}7810q^7 - 30{\cdot}5070q^6 - 300{\cdot}0260q^5 + 869{\cdot}6372q^4$$
$$- 278{\cdot}8475q^3 - 270{\cdot}9547q^2 + 58{\cdot}9108q + 14{\cdot}6050 = 0.$$

This leads to $p_2 = -10q = -9{\cdot}1777$, and the components given in the Table p. 125. The two Gaussians and their compound are given in Diagram VIII (p. 133). We find $\chi^2 = 11{\cdot}96$, giving for $n' = 18$, $P = {\cdot}80$ a splendid fit.

(iv) Mzimba Strain (from Donkey).

$$\text{Mean} = 21{\cdot}4130,$$

$$\mu_2 = 8{\cdot}7531, \qquad \mu_4 = 293{\cdot}5629,$$

$$\mu_3 = 26{\cdot}6602, \qquad \mu_5 = 1926{\cdot}7045.$$

The reducing nonic:

$$24q^9 + 53{\cdot}5186q^7 - 25{\cdot}5876q^6 - 41{\cdot}5706q^5 - 171{\cdot}2637q^4$$
$$+ 227{\cdot}1211q^3 - 37{\cdot}3371q^2 - 30{\cdot}8995q + 8{\cdot}6177 = 0.$$

The required root is $p_2 = -10q = -4{\cdot}0000$, which leads to the two components given in the Table on p. 125. The two components and their compound curve are

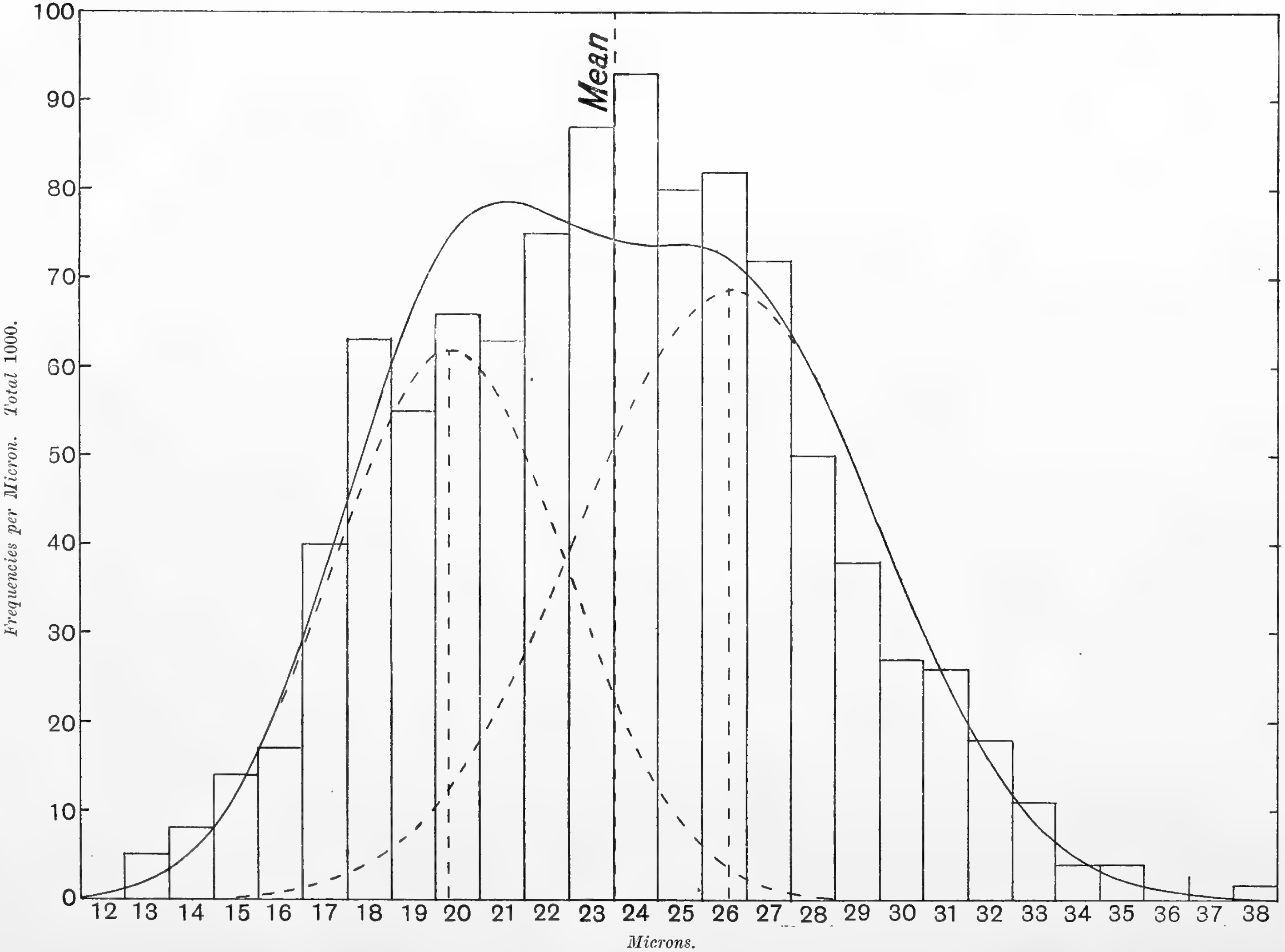


Diagram VII. Resolution of the Frequency of *T. brucei* into *T. minus* and *T. majus*.

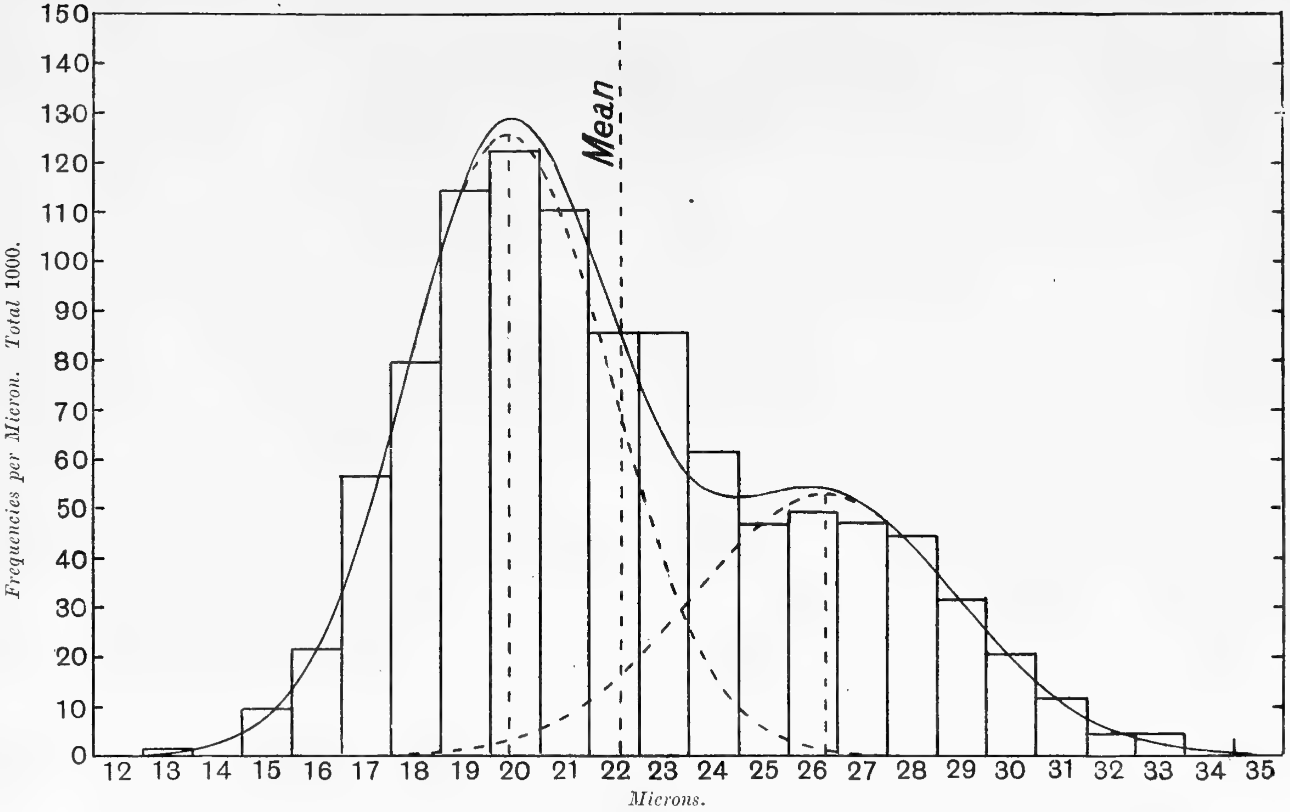


DIAGRAM VIII. Resolution of the Frequency of *T. gambiense* into *T. minus* and *T. majus*.

figured on Diagram IX on this page. We have $\chi^2 = 19{\cdot}28$, giving for $n' = 17$, $P = {\cdot}26$ a fairly reasonable fit.

(v) Wild *G. morsitans* Strain.

$$\text{Mean} = 22{\cdot}6952,$$

$$\mu_2 = 18{\cdot}4918, \qquad \mu_4 = 758{\cdot}4420,$$

$$\mu_3 = 43{\cdot}0246, \qquad \mu_5 = 3954{\cdot}8788.$$

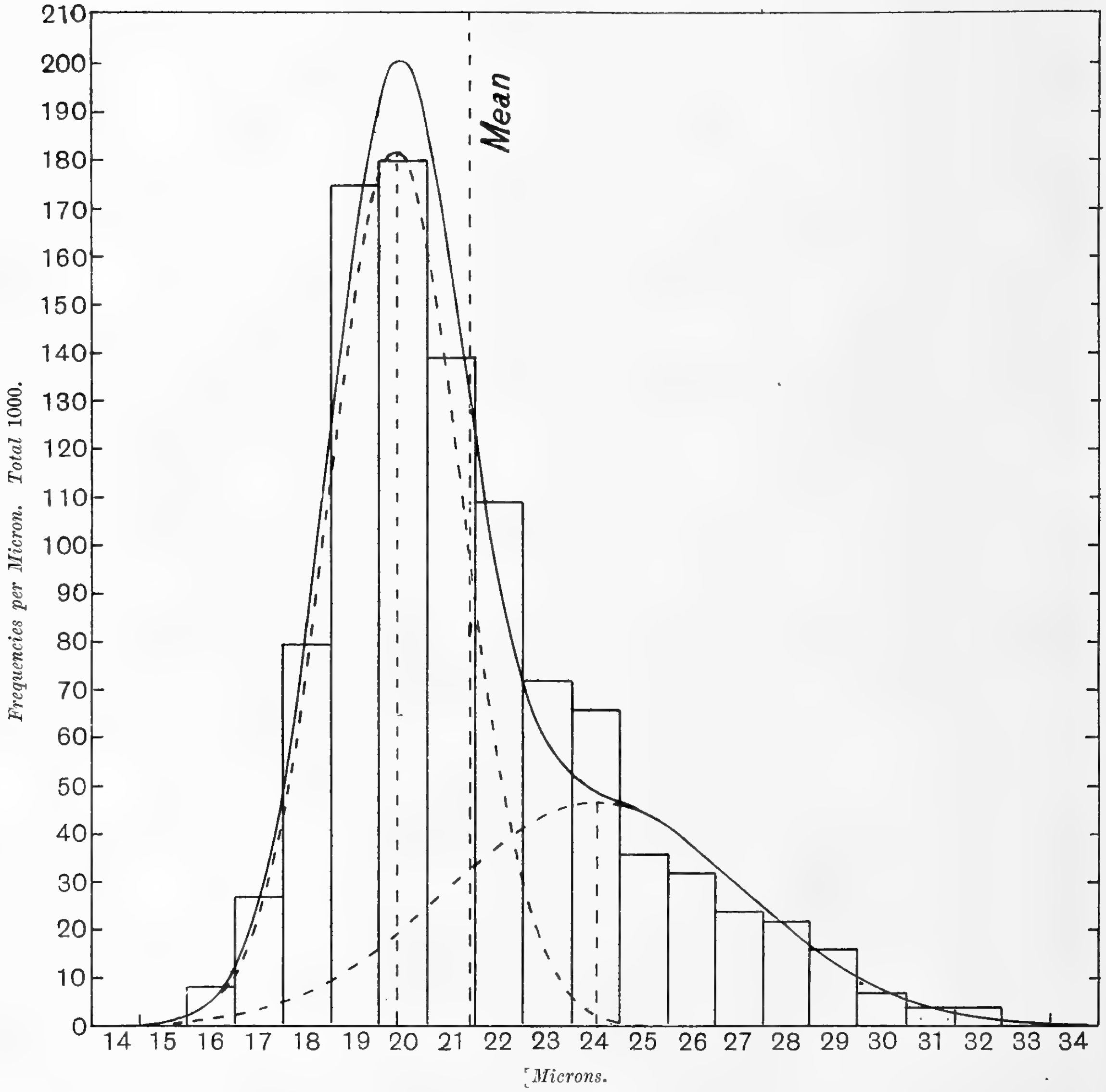


DIAGRAM IX. Resolution of the Frequency of the Mzimba Strain into *T. minus* and *T. majus*.

Reducing nonic:

$$24q^9 - 224{\cdot}6115q^7 - 66{\cdot}6402q^6 - 595{\cdot}9589q^5 + 5079{\cdot}3305q^4$$
$$- 4500{\cdot}7030q^3 - 1460{\cdot}5459q^2 + 879{\cdot}6116q + 152{\cdot}2340 = 0.$$

The required root is $p_2 = -10q = -4{\cdot}75085$, which leads to the components given in Table on p. 125. These components with their compound curve are drawn in Diagram X (p. 136). Here $\chi^2 = 92{\cdot}75$ which for 20 groups gives $P < {\cdot}000{,}000{,}1$. Thus although the *G. morsitans* strain breaks up into two components the combined curve is not a probable description of the frequency. One would like to test another sample of this strain, at present it tells against the validity of our reduction.

(vi) *Wild-Game Strain.*

$$\text{Mean} = 22{\cdot}6220,$$
$$\mu_2 = 11{\cdot}9310, \qquad \mu_4 = 404{\cdot}4932,$$
$$\mu_3 = 29{\cdot}0514, \qquad \mu_5 = 2247{\cdot}6657.$$

Reducing nonic:

$$24q^9 - 18{\cdot}9446q^7 - 30{\cdot}3834q^6 - 250{\cdot}2869q^5 + 351{\cdot}7475q^4$$
$$+ 118{\cdot}6154q^3 - 212{\cdot}3972q^2 + 15{\cdot}4222q + 14{\cdot}4283 = 0.$$

The root required is $p_2 = -10q = -6{\cdot}9859$. There result the two components provided in the Table p. 125. The two components and their compound are figured on Diagram XI. (p. 137). We find $\chi^2 = 12{\cdot}61$ giving for $n' = 19$, $P = {\cdot}81$, an excellent fit.

(vii) *Human Strain.*

$$\text{Mean} = 23{\cdot}7963,$$
$$\mu_2 = 17{\cdot}0252, \qquad \mu_4 = 713{\cdot}1660,$$
$$\mu_3 = 27{\cdot}1389, \qquad \mu_5 = 3034{\cdot}1222.$$

Reducing nonic:

$$24q^9 - 131{\cdot}3796q^7 - 26{\cdot}5147q^6 - 89{\cdot}8059q^5 + 964{\cdot}4176q^4$$
$$- 674{\cdot}2755q^3 - 114{\cdot}7894q^2 + 81{\cdot}4492q + 9{\cdot}5887 = 0.$$

The root is given by $p_2 = -10q = -8{\cdot}5576$, which leads to the components given in the Table on p. 125. The two curves and their compound are figured in Diagram XII (p. 138). Although the two components merely from the graphical point of view do not give a bad fit, the number of trypanosomes involved is so large that the deviations are not reconcileable with random sampling from two such components. We find $\chi^2 = 79{\cdot}67$, giving $P < {\cdot}000{,}001$.

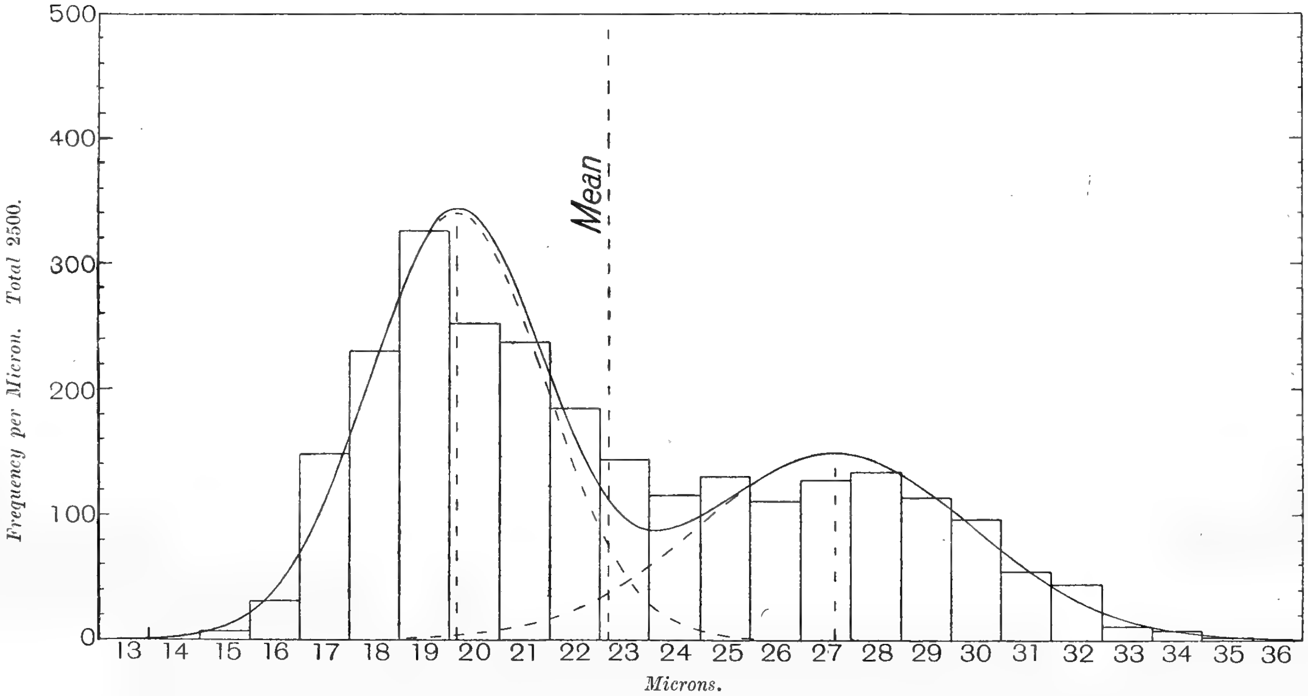


Diagram X. Resolution of the Frequency of the Wild *G. morsitans* Strain into *T. minus* and *T. majus.*

In order to determine how far heterogeneity of treatment or material might be responsible we took further frequencies. In the first place we dealt with the 3600 measurements for trypanosomes through the rat only. The frequencies are:

15	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	Totals
1	5	47	112	161	216	290	316	376	362	322	294	235	219	210	134	108	88	57	28	9	8	1	1	3600

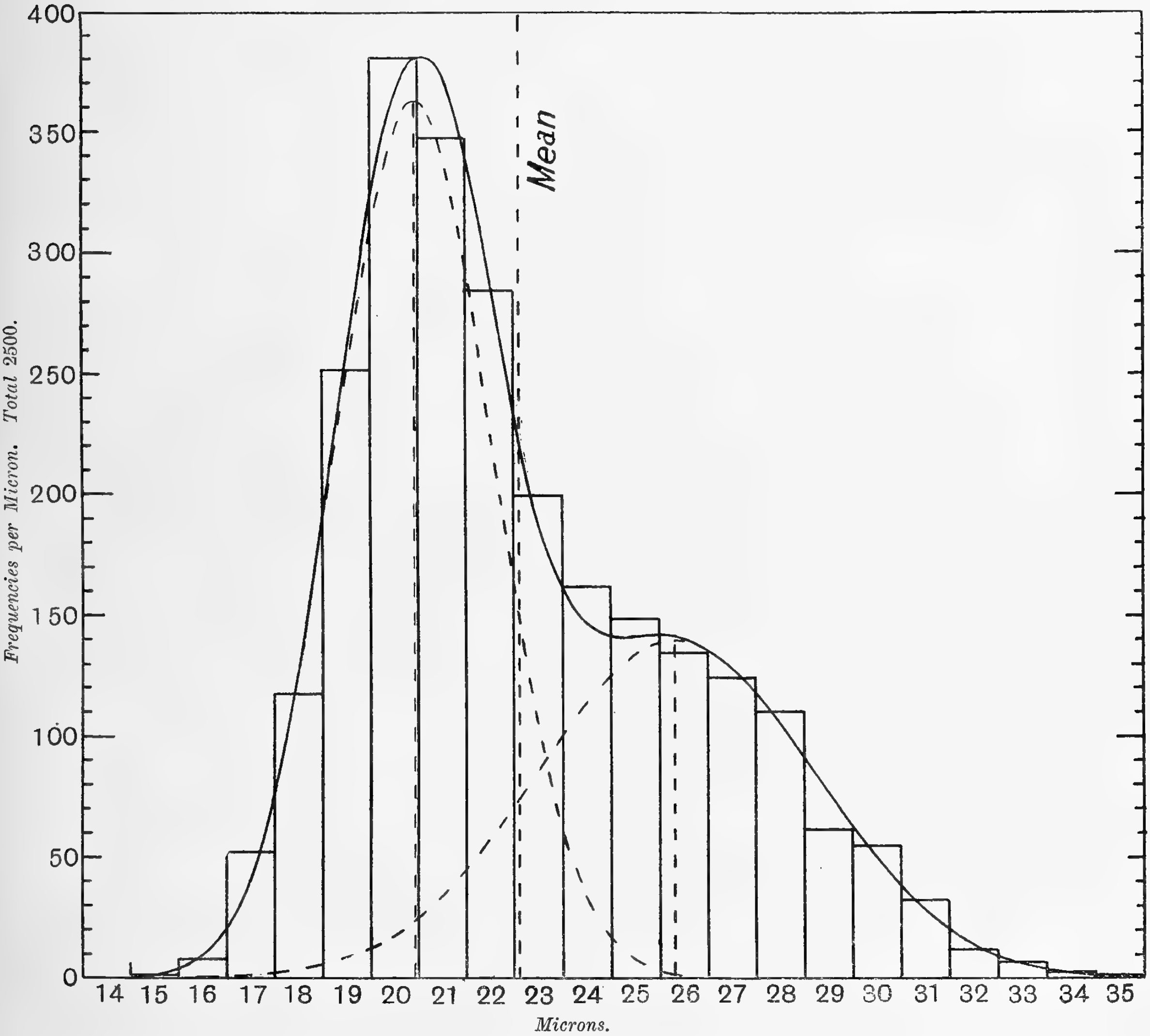


DIAGRAM XI. Resolution of the Frequency of the Wild-Game Strain into *T. minus* and *T. majus*.

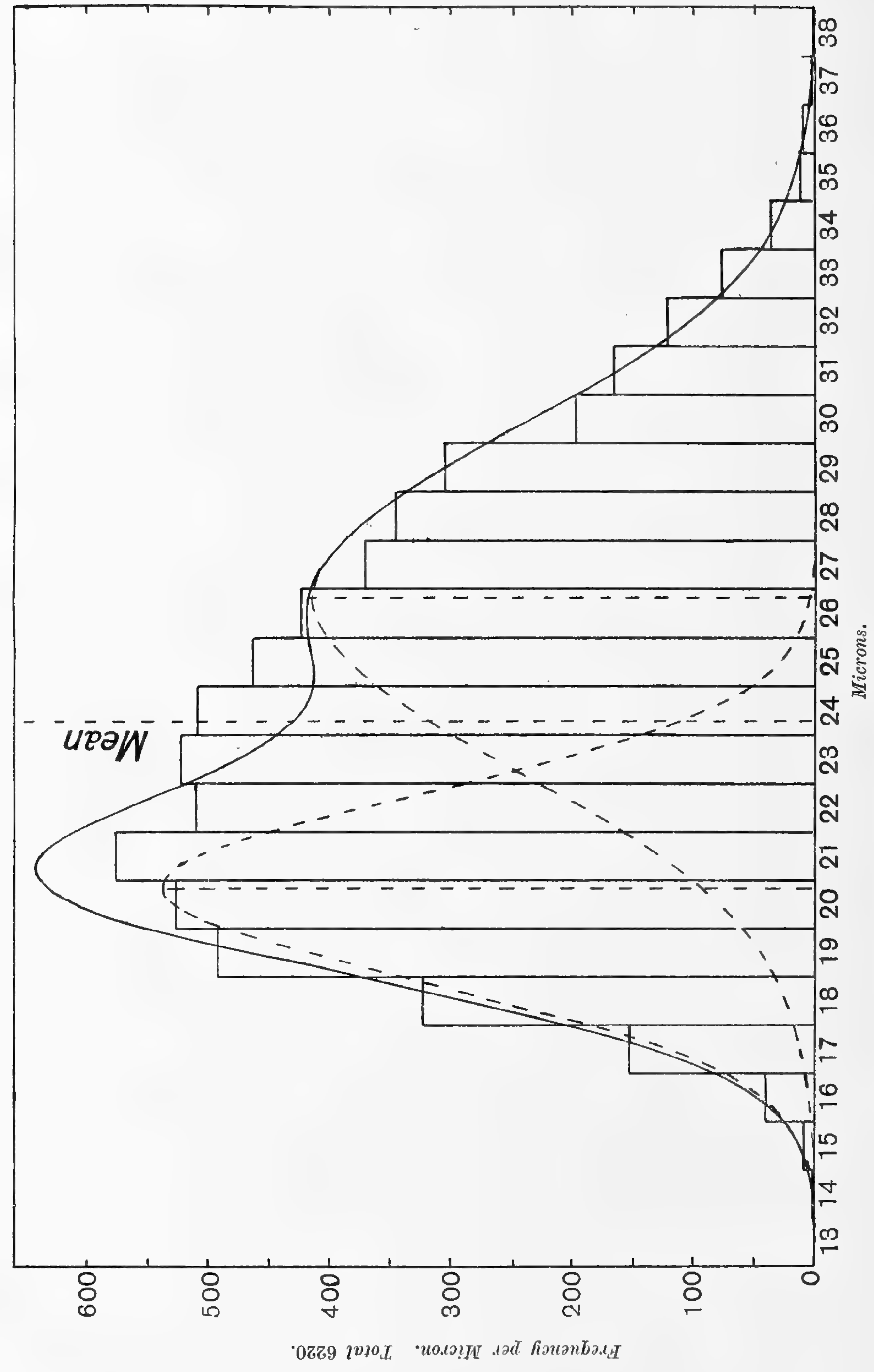


Diagram XII. Resolution of the Frequency of the Human Strain into *T. minus* and *T. majus*.

These give:

$$\text{Mean}: 24\cdot6175,$$

$$\mu_2 = 15\cdot25897, \qquad \mu_4 = 602\cdot23008,$$

$$\mu_3 = 19\cdot21542, \qquad \mu_5 = 2023\cdot21556,$$

leading to the reducing nonic:

$$24q^9 - 80\cdot8739q^7 - 13\cdot2924q^5 - 42\cdot3159q^5 + 306\cdot5227q^4$$
$$- 166\cdot4257q^3 - 24\cdot8654q^2 + 12\cdot6008q + 1\cdot2081 = 0$$

which gives $$p_2 = -10q = -7\cdot0031.$$

This provides the two components:

T. minus.	*T. majus.*
$m_1 = 21\cdot6772,$	$m_2 = 26\cdot9993,$
$\sigma_1 = 2\cdot2404,$	$\sigma_2 = 3\cdot2981,$
$n_1 = 1611\cdot18,$	$n_2 = 1988\cdot82.$

The components and their compound are figured in Diagram XIII, p. 140, and we find for $n = 21$, $\chi^2 = 52\cdot68$ and $P = \cdot00016$. There has thus been much improvement of goodness of fit, although the result is still unsatisfactory.

It is impossible, however, to look through the graphs given by Sir David Bruce and others for the human strains* without being convinced of their fundamentally bimodal character, although there appears to be much evidence of its being disguised by heterogeneity of host and treatment.

(viii) Diagram XIV (p. 141) gives the resolution for the human strain from Chituluka†. The constants

$$\text{Mean} = 26\cdot172,$$

$$\mu_2 = 23\cdot52260, \qquad \mu_4 = 1179\cdot30786,$$

$$\mu_3 = -37\cdot13226, \qquad \mu_5 = -3248\cdot43805,$$

leading to the reducing nonic:

$$24q^9 - 393\cdot8678q^7 - 49\cdot6370q^6 + 520\cdot2910q^5 + 8226\cdot9435q^4$$
$$- 12493\cdot5620q^3 - 101\cdot1017q^2 + 855\cdot7520q + 63\cdot2383 = 0.$$

The value of the root is $p_2 = -10q = -16\cdot2295$ and this leads to the components given in the Table p. 125, and illustrated in the diagram. The graph while giving broadly some of the features of the case is by no means a satisfactory fit; for $n = 21$ groups, $\chi^2 = 86$ and P is $< \cdot000,000,1$. The diagram suggests that we are probably dealing with a mixture of three components with means about 18·5, 25·5 and 31·0, but at present we have no satisfactory method of performing multiple resolutions of this character.

---

* *R. S. Proc.* Vol. 86, B, pp. 285—302.

† *R. S. Proc.* Vol. 86, B, p. 291.

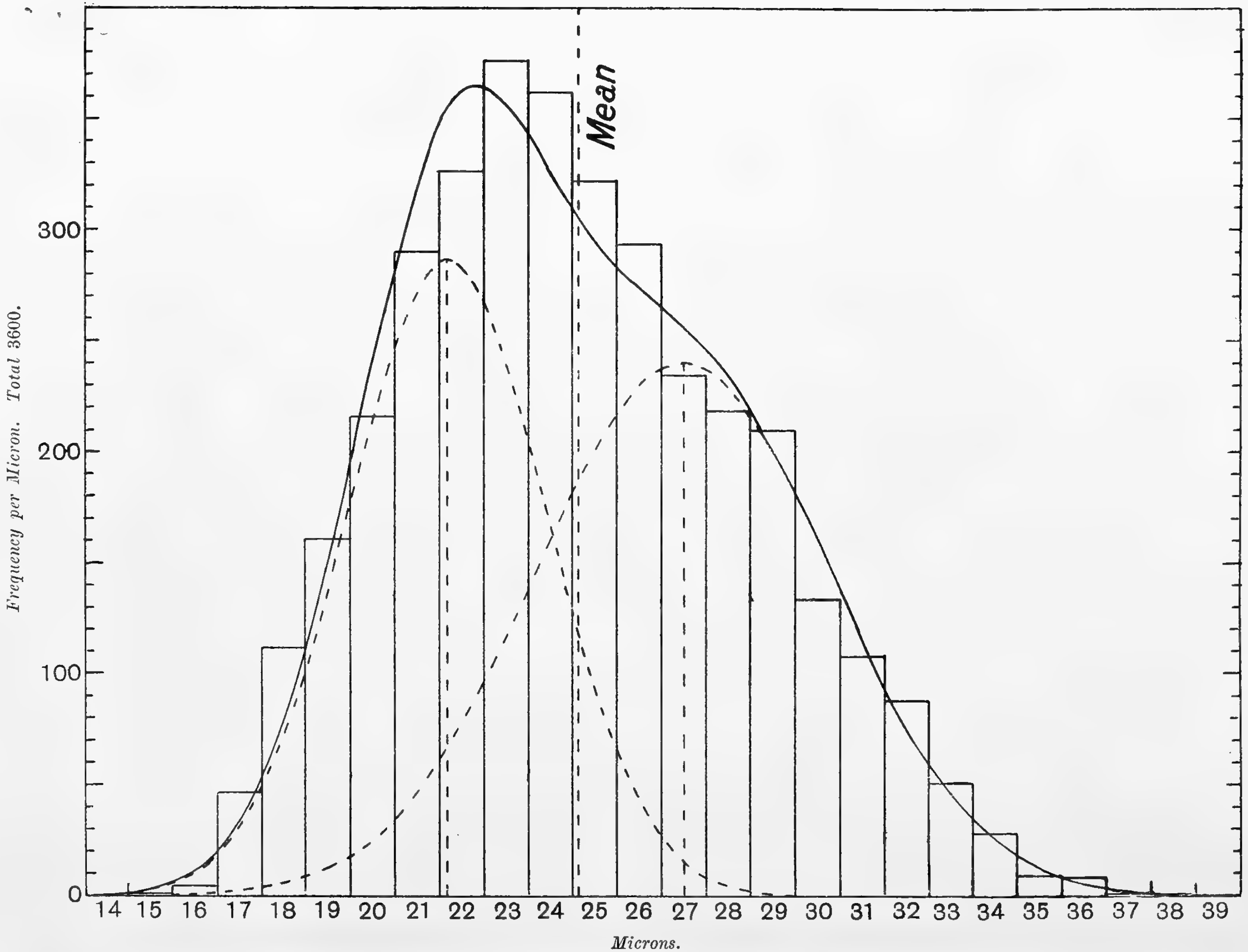


Diagram XIII. Resolution of the Frequency of Trypanosomes from Human Strain, through Rats only into *T. minus* and *T. majus*.

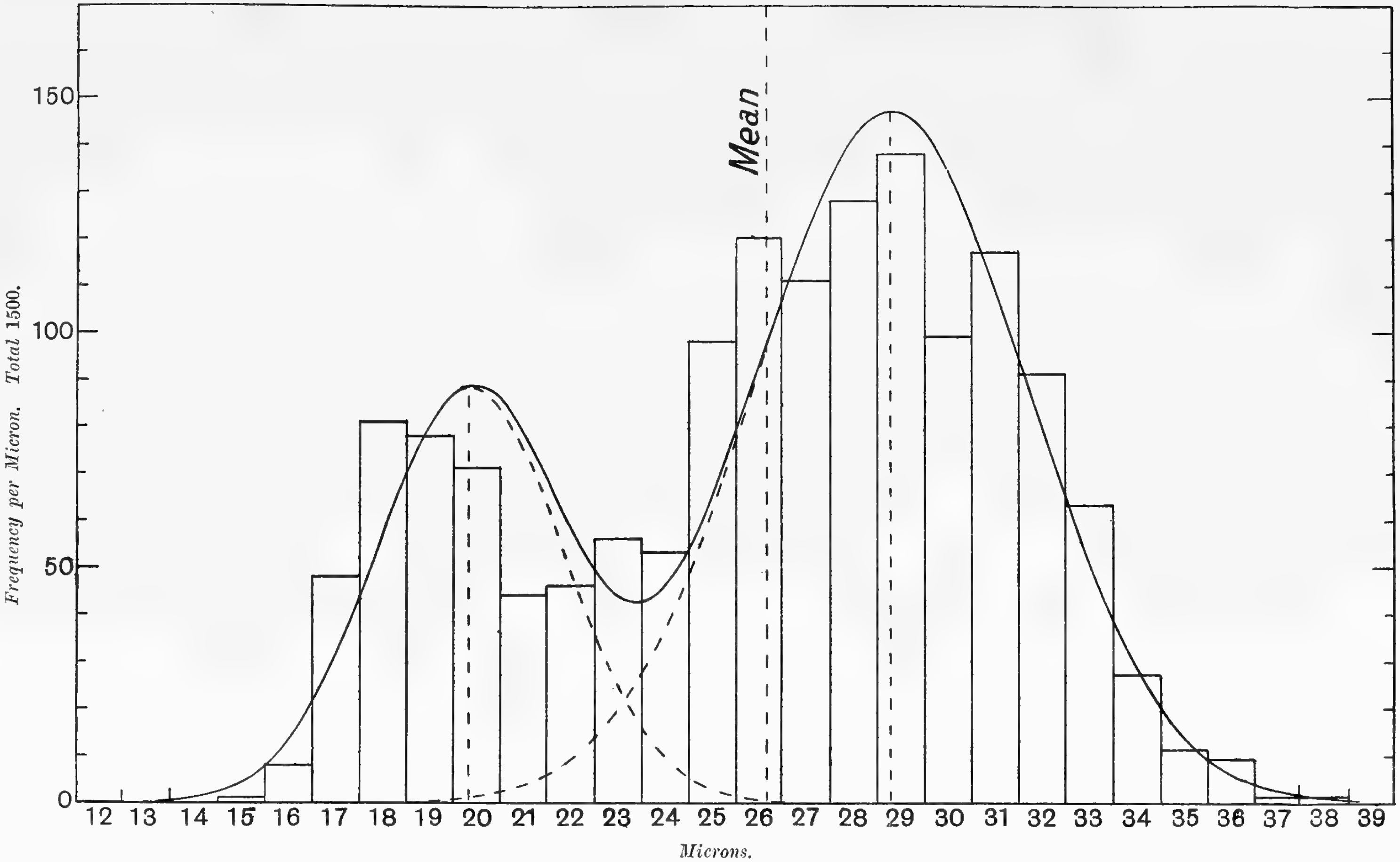


DIAGRAM XIV. Resolution of the Frequency of Trypanosomes from the Native Chituluka into *T. minus* and *T. majus*.

It will be seen that the following strains, *T. rhodesiense, T. brucei, T. gambiense,* the *Mzimba,* and wild game, give either reasonable or excellent results as combined frequencies of *T. minus* and *T. majus.* On the other hand the *G. morsitans* and the human strains break up into reasonable pairs of components, but the goodness of fit test is not fulfilled. In the case of the human strain, we better matters somewhat by taking the strain through the rat only, but the fit is still bad. If we confine our attention to a single human being, the case of Chituluka, we still do not get a satisfactory fit, although few statisticians could look at the four diagrams published by Sir David Bruce and others for Chituluka*, and not recognise the character of the material as being at least bimodal. The same applies to the Mkanyanga data of an earlier paper†, it is distinctly bimodal. But besides this bimodal character there are certain other features in the human data, and to a lesser extent in the *G. morsitans,* which appear to some extent to disguise the bimodal features. I am not prepared to assert definitely that this is the appearance of a third component. It is of course easy to improve the fit of the distribution by the introduction of such a third component, but the remarkable excellence of a bimodal resolution for *T. rhodesiense, T. gambiense,* and the wild-game strain makes me hesitate at present to adopt such an expedient.

Owing to the courtesy of Sir David Bruce (who heard from Sir John Rose Bradford that I was much puzzled over the differentiation of strains) I have been able to examine a series of drawings of the various strains of trypanosomes. There is no other morphological differentiation which impresses itself *a priori* on the layman and statistician, and which might serve as a new measure of the possibility of differentiation into *T. minus* and *T. majus.* But it occurs to me that an index of breadth to length of the nucleus might just possibly serve as a differential character of even more importance than the length. It is only a suggestion and considerable caution would have to be used in selecting only nuclei not near the dividing stage. But it would be of striking interest to see how far the resulting frequency distributions for the nuclear indices were or were not bimodal. I think a classification according to nuclear index might possibly—to judge from the drawings—cut across the forms "intermediate" in length. But this is only a suggestion which may appear idle to the student of the subject‡. Some difficulty might also arise from the doubt as to whether the index was really greater than 100, or the nucleus as a whole had set itself athwart the "length" of the trypanosome. This difficulty would certainly have to be considered in the "stumpy" *T. brucei* and *T. gambiense*

---

* *R. S. Proc.* Vol. 86, B, pp. 291 to 293.

† *R. S. Proc.* Vol. 85, B, p. 428.

‡ Several students of the subject with whom I discussed the matter stated that they considered the nucleus so mobile and so impermanent in form, that a "nuclear index" would prove of little value. I think much objection could *a priori* be raised to the use of the trypanosome "length" on the same grounds. The problem is rather, whether in dealing with large numbers we do reach an average type. It would only be possible *a posteriori* to justify the use of a nuclear index, i.e. if it were found to differ sensibly from one pure strain to a second, and if it confirmed in such cases as *T. rhodesiense* resolutions based on length frequencies.

forms, but I am inclined to think that the index really passes through the value 100. Undoubtedly this range of index, or possible athwartness of the nucleus is not conspicuous in the simple strains like *T. pecorum*, *T. simiae* and *T. caprae*.

Conclusions. (i) If appeal be made to statistical measurements, judgment between identity and diversity of strain must be formed by means of accepted statistical processes and not by mere comparison of graphs.

(ii) Statistical processes show that the conclusions already formed as to the identity of trypanosome strains from mere inspection of the graphs cannot be confirmed.

(iii) There must be some standardised process of treatment both in regard to host, and to method of and stage of infectivity at extraction.

(iv) Even making allowance for differences due to host and treatment, we find remarkable divergences in the very strains asserted to be identical.

(v) It would appear that some order would be brought into the chaos, if we could consider the strains described as *T. brucei*, *T. rhodesiense*, *T. gambiense*, the wild-game, the Mzimba, and very probably the tsetse fly and the human strains as really consisting of two components, which for the time I have termed *T. minus* and *T. majus*. It is highly desirable that additional measurements should be made (? a nuclear index ascertained) to determine whether these lead also to similar components.

I do not assume that this is a final solution of the problem, nor do I assert that *T. minus* and *T. majus* represent necessarily, although probably, distinct strains; they may be dimorphic forms of one and the same strain occurring in different proportions. But, I believe, that the suggestion of their existence may help to explain some anomalies of the present chaos. I ought also to state quite frankly that this paper is not written in a merely critical spirit. I believe that the trypanosome workers have undertaken in their elaborate systems of measurements most laborious and most valuable work, but, I think, the time has now come when without trained statistical aid, but little further progress will be made in a very important and urgent matter.

The very large amount of arithmetical work in this paper would never have got carried through had I not had the ever ready assistance of my colleague Miss Julia Bell; to Mr H. E. Soper also I owe help in the arithmetical work, but I have to thank him in particular for the careful preparation of the diagrams, and the planimetric determination of their frequencies by aid of which the χ^2 for all but two of the compound curves was found. In the case of *T. brucei* and *T. rhodesiense* actual calculation of the areas of the normal curves was used.

ON HOMOTYPOSIS AND ALLIED CHARACTERS IN EGGS OF THE COMMON TERN

By WILLIAM ROWAN, K. M. PARKER, B.Sc., and JULIA BELL, M.A.

(1) *Origin of the material and method of measurement.*

The settlement of Common Terns, which provided material for the present work, is one of old establishment on Blakeney Point, Norfolk. This is a shingle spit of some 8 miles in length on the north coast of Norfolk, about 12 miles west of Cromer. The colony is situated on the very end of the point, with water on three sides. Here the spit is a combination of dunes, salt marsh and shingle, and for the most part the nests are found on the open shingle on the seaward side of the dunes. Nests are plentiful in the embryo dunes in some years, though this year (1913) none were found there. The colony was more scattered than usual and covered the greater part of a mile of sea front. To avoid missing any clutches, Miss K. M. Parker, B.Sc., and Mr William Rowan divided the nesting area into suitable well marked plots and worked these one after another. Each of these again were worked in strips, till a patch was completed, when the workers moved on to a remote one, to give the birds a chance of settling down again. After measurement each egg was numbered with indelible ink, so that any one egg was never measured twice. In all 203 clutches were handled.

(2) *Reduction of the material.*

The principal part of the work of tabling and reduction was carried out by Julia Bell[*]. The characters dealt with were:

(i)	Length of Egg	L
(ii)	Breadth of Egg, maximum value	B
(iii)	Lateral Girth at section with maximum breadths	G_b
(iv)	Longitudinal Girth	G_l
(v)	Length-Breadth Index	B/L
(vi)	Mottling, as determined from a scale of typical eggs	M
(vii)	Ground Colour, as determined from a tint scale	C

---

[*] The authors have to thank Miss B. M. Cave for certain tables and their correlation coefficients. The Editor is responsible for the actual wording of this paper.

The Length of egg L may be considered as the easiest character to determine and needs no further comment.

The Breadth of egg B should be closely related to the Lateral Girth G_b, and in most cases the relationship $G_b = \pi B$ is very closely satisfied. If we sum and take the means we have

$$\pi = \text{Mean Lateral Girth}/\text{Mean Breadth}.$$

This gives in the present material:

$$\pi = 3{\cdot}224 \text{ as against } 3{\cdot}142,$$

which marks an error of about 2·6 %, rather larger than we might anticipate, and possibly due to the inclusion of a certain number of slightly damaged eggs, and the measurement of the eggs in the field and not in the laboratory. The relation between G_b and B is a useful test of accuracy and should be determined with a slide rule before the egg is finally replaced in the nest, or lost sight of.

The Longitudinal Girth G_l is somewhat more difficult to measure, and a rough test of its accuracy not so easy to determine as in the case of G_b. We have, however, developed a formula for determining G_l in terms of B and L, and on testing it we find that as a rule the differences are below 1·5 mm. Such a formula may be useful as emphasising the need for remeasurements, when the observed and calculated girths have values much in excess of 1·5 mm. We are not prepared to say, however, that the coefficients in this formula can be extended beyond the case of the Common Tern.

While the Length-Breadth Index is valuable as giving a measure of the ellipticity of the egg, it is not of much influence on the apparent *oval* shape, unless we suppose some theoretical geometrical construction for the egg. If we suppose the blunt end of the egg to be approximately spherical, the hemisphere ending with the maximum breadth, then the egg might be considered as divided into two portions, the upper or hemispherical with radius $\frac{1}{2}B$ and the lower with length from the base of the hemisphere (or 'equator') to the lower pole $= L - \frac{1}{2}B$. The ratio of these two segments of the length depends only on the index B/L. Thus it is conceivable that this index has actually as much association with ovality as with ellipticity, although without some geometric theory of egg-shape, we are not able to make any dogmatic assertion as to the value of B/L. It seems, however, a character of considerable interest as being free of absolute size and also some measure of shape. If $I = B/L$ and O be the ratio of $\frac{1}{2}B$ to $L - \frac{1}{2}B$, i.e. $O = \dfrac{B/L}{2 - B/L} = I/(2 - I)$, we may consider O a measure of the ovality, and we have correlated O for eggs of the same clutch as well as I. Of course, since O is a function of I, there will be relatively little difference in the results.

The mottling is a far more difficult matter for determination. The points which may be considered are:

(i) Size and shape of individual splodges.

(ii) Portion of the egg over which these splodges are distributed.

(iii) Area of mottled surface as compared with whole area of the egg.

The fieldworkers selected 9 typical mottlings (see Plate IX) and named these *a, b, c, d, e, f, g, h, i*; they then compared each recorded egg with these and selected the letter which marked the egg on the scale most resembling the egg to be recorded. There is little doubt that in this manner they divided the whole series of eggs into differentiated classes. But it may be doubted whether the judgment made depended on one only of the above three characteristics. Hence when we came to arrange the eggs *a, b, c, d,* ... *h, i* on a scale of mottling, we found that the order would not be the same when we classified in turn by each of the three characteristics. We endeavoured to place the eggs in order by extent of mottling, i.e. by (iii), but we think that the relatively low value of the homotyposis which has resulted is possibly due to size and shape of the mottlings, (i), having had as much influence on the classification as the extent of area mottling. Even position on the egg, (ii), can influence judgment considerably. We believe that in future work on eggs, it would be desirable to classify the mottling of each by using the three characteristics independently. Even then an ocular appreciation, as this must be, may fail to give a very close measure of the nature of the mottling and thus weaken any homotypic correlation.

The Ground Colour of these eggs varies through all shades of brown to brownish greens and blue-greens. The fieldworkers attempted to give the value or depth of ground-colour pigmentation without regard to the brown or green shade of colouring. The scale of values is given at the foot of Plate VIII.

A point seemed worth consideration: assuming the pigments to be deposited on the egg in its passage through the oviduct, it was conceivable that greater pressure might indicate greater intensity of pigmentation. We accordingly selected the *broader* egg in each clutch and investigated for every pair of eggs from the same clutch whether the broader or narrower egg had the larger mass of mottling and greater density of ground colour. We reached the following results:

The broader egg in every possible clutch-pair has

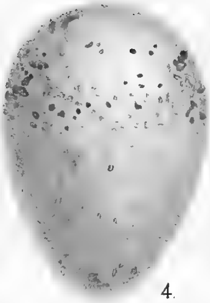
Greater mottling	in	26	cases	More dense ground colour		in	25	cases
The same	,,	37	,,	The same	,,	,,	39	,,
Less	,,	40	,,	Less dense	,,	,,	37	,,

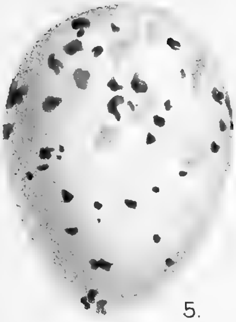
Perhaps not very much stress is to be laid on these results, but they suggest that the total amount of pigment deposited is *less* the broader the egg, i.e. for the same bird a relatively smaller egg will be more pigmented. A solution of this

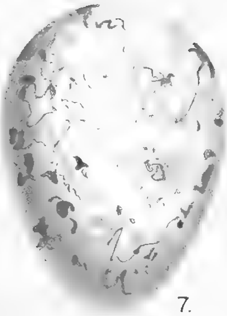

2.

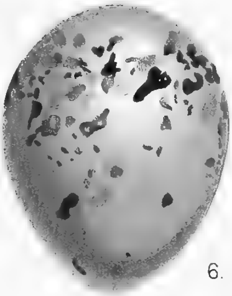

1.

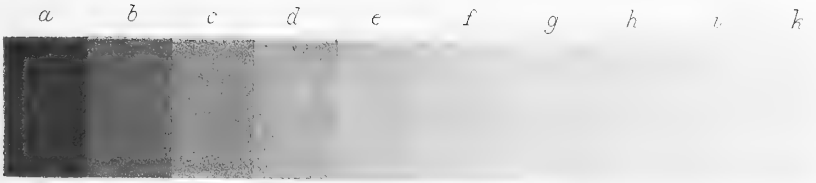

3.


4.


5.


7.

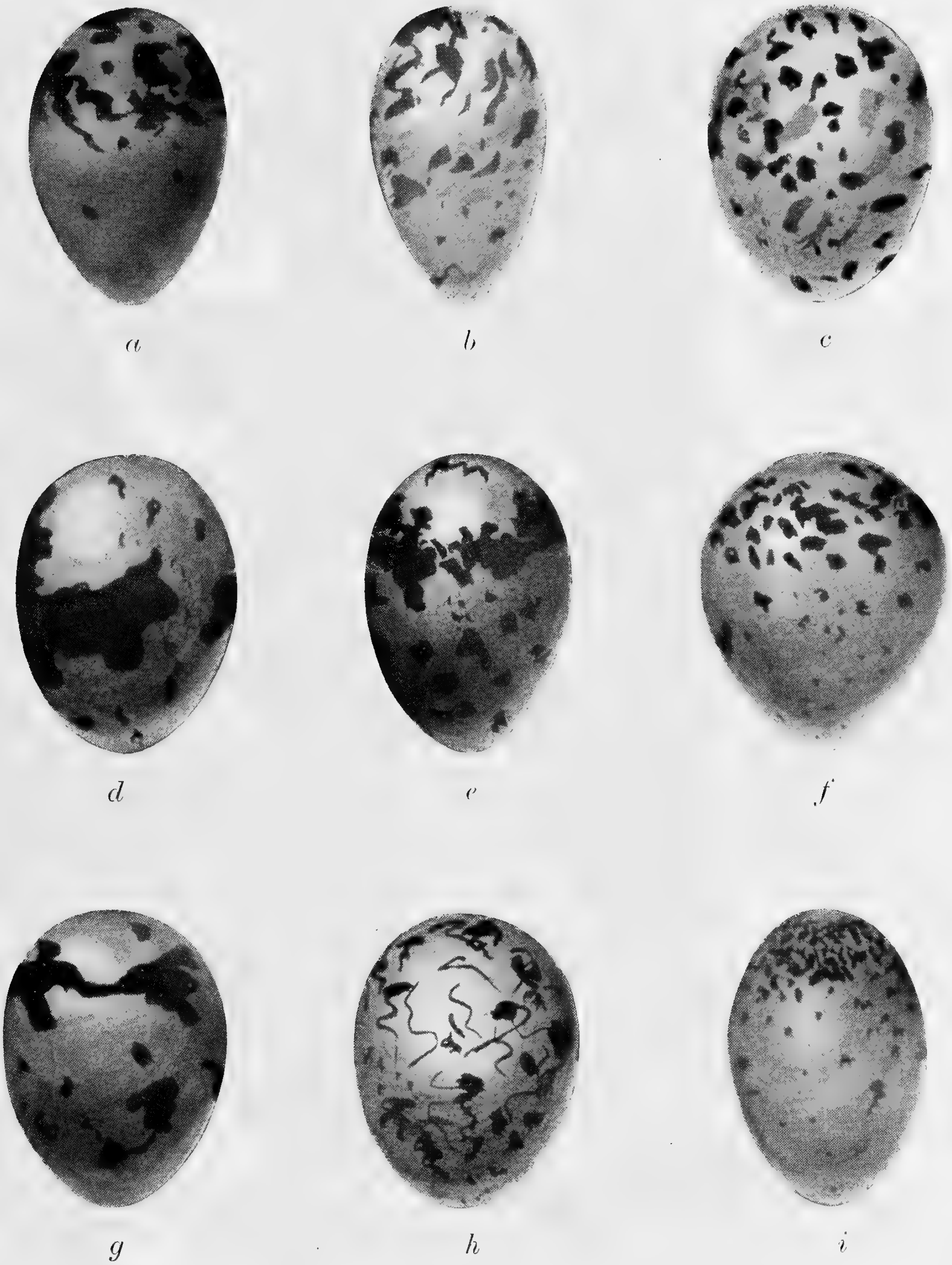

6.



Colour Value Scale

Cambridge University Press

SAMPLE EGGS, COMMON TERN, NATURAL SIZE



TYPES OF MOTTLING OF EGGS OF COMMON TERN.

rather unexpected result may, perhaps, be found in the suggestion that the total amount of pigment is the *same* in both eggs, but the mottling and ground colour will appear denser on the smaller surface of the smaller egg. The point deserves consideration on the basis of larger numbers and possibly better defined measures of pigmentation.

(3) *Means and Variability.*

Table I gives the means, standard deviations and coefficients of variation of the several characters studied. It will be seen that the tern's egg has for quantitative characters relatively small variation. The values of the coefficients of variation are less than many of those which we find for the human skull (3 to 8), but greater than those we know for the wing of the wasp. It is very doubtful whether the coefficients of variation of the indices should be included in such considerations, for the object of the use of these coefficients is to get rid of absolute lengths, and this is already done in the case of indices[†]. It is noteworthy that the length of the egg is only slightly more variable than the breadth and the breadth-girth is actually more variable than the length-girth.

TABLE I.

Means and Variabilities (Absolute Measurements in Centimetres).

Character	Mean	Standard Deviation	Coefficient of Variation
Length L	4·14 ± ·007	·180 ± ·005	4·34 ± ·12
Breadth B	2·98 ± ·004	·099 ± ·010	3·33 ± ·09
Girth G_l	11·39 ± ·015	·376 ± ·010	3·30 ± ·09
Girth G_b	9·59 ± ·014	·347 ± ·010	3·62 ± ·10
Index B/L	72·04 ± ·136	3·449 ± ·096	[4·79 ± ·13]
Index of Ovality, O* ...	56·35 ± ·171	4·334 ± ·121	[7·69 ± ·22]

(4) *Correlations.*

If we turn to the correlation of characters in the same egg, we note that while the ordinary product-moment correlation r has been calculated for all measurable pairs of characters, this is not possible for the ground colour or the mottling. Where mottling has been used with a quantitative character there η has been calculated and both corrections used. Where mottling has been considered in conjunction with ground colour, there we have adopted mean square contingency correcting for both number of cells and for class-index correlations.

---

* $O = (B/L)/\{2 - (B/L)\}$.

[†] For example, if we take $1/O$ for our index of ovality its mean = 176·32, the standard deviation = 11·24 and the coefficient of variation = 6·38. Is O or $1/O$ the more variable? It does not seem that the coefficient of variation can help us in such a problem.

Certain facts are at once obvious from this Table, others are obscured. In the first place length and breadth of the egg of the Common Tern have a relatively small relationship, while the relationship between the two girths is between two and three times as great. This probably flows from the consideration that the correlation of G_l and G_B arises from B being a factor in both and only secondarily from the correlation between L and B. The correlation of the longitudinal girth with egg length is 60% higher than that of longitudinal girth with egg breadth; both these correlations are more substantial than that of the longitudinal girth, G_l, with the egg index, B/L. The egg index correlated with length is large and negative, and with breadth considerable and positive, precisely the results we should anticipate would appear if the correlation were largely *spurious**.

TABLE II.

Correlations of Characters in the same Egg.

Characters	Symbols	Correlation	Remarks
Length and Breadth	L , B	$+\cdot 2220 \pm \cdot 0374$	—
Longitudinal and Equatorial Girths	G_l , G_b	$+\cdot 5297 \pm \cdot 0284$	—
Length and Longitudinal Girth ...	L , G_l	$+\cdot 8804 \pm \cdot 0088$	—
Breadth and Longitudinal Girth ...	B , G_l	$+\cdot 5216 \pm \cdot 0286$	—
Index and Longitudinal Girth ...	B/L, G_l	$-\cdot 3832 \pm \cdot 0336$	—
Index and Length	B/L, L	$-\cdot 7284 \pm \cdot 0185$	—
Index and Breadth	B/L, B	$+\cdot 5033 \pm \cdot 0294$	—
Mottling and Ground Colour ...	M, C	$+\cdot 2260$ (corrected C_2)	More mottling, deeper ground colour
Mottling and Index	M, B/L	$-\cdot 1550$ (corrected η)	Less mottling, higher index
Mottling and Breadth	M, B	$-\cdot 1803$ (corrected η)	Less mottling, greater breadth
Ground Colour and Index	C , B/L	$\cdot 0000$ (corrected η)	No relationship
Ground Colour and Breadth ...	C , B	$-\cdot 1506$ (corrected η)	Fainter ground colour, greater breadth

In order to ascertain how far it was possible to predict the longitudinal girth from length and breadth, double (for L and B) and triple (for L, B and B/L) regression formulae were worked out. The following equations resulted:

(i) $$G_l - \bar{G}_l = 1\cdot 2701\,(B - \bar{B}) + 1\cdot 6415\,(L - \bar{L}),$$

or, $$G_l = 1\cdot 2701\,B + 1\cdot 6415\,L + \cdot 8224,$$

and (ii) $$G_l - \bar{G}_l = -\,17\cdot 2930\,(B - \bar{B}) + 14\cdot 6374\,(L - \bar{L}) + \cdot 7636\,(I - \bar{I}),$$

or, $$G_l = -\,17\cdot 2930\,B + 14\cdot 6374\,L + \cdot 7636\,B/L - 52\cdot 7239.$$

The first seventeen eggs were taken as a random set to test these results upon with the following values:

---

* As a matter of fact the correlation of index and length for a constant breadth is $-\cdot 997$ and of index and breadth for a constant length is $+\cdot 996$ instead of unity. These values indicate how closely the linearity of regression holds in these quantitative measurements.

TABLE III.

Observed and Calculated Longitudinal Girths.

Egg Number	Observed Girth G_l	Calculated Girth		Difference	
		(ii)	(i)	Δ_2	Δ_1
1	11·40	11·14	11·20	+ ·26	+ ·20
2	11·65	11·83	11·74	− ·18	− ·09
3	12·10	12·24	12·07	− ·14	+ ·03
4	10·80	11·46	10·84	− ·66	− ·04
5	11·70	11·23	11·31	+ ·47	+ ·39
6	11·20	11·27	11·34	− ·07	− ·14
7	12·15	13·19	12·31	− 1·04	− ·16
8 (i)	11·20	11·19	11·27	+ ·01	− ·07
8 (ii)	11·30	11·09	11·27	+ ·21	+ ·03
9 (i)	11·50	11·44	11·61	+ ·06	− ·11
9 (ii)	11·40	11·36	11·45	+ ·04	− ·05
10	11·50	11·52	11·61	− ·02	− ·11
11	11·80	11·55	11·72	+ ·25	+ ·08
12	11·90	11·62	11·74	+ ·28	+ ·16
13 (i)	11·10	11·01	10·94	+ ·09	+ ·16
13 (ii)	10·80	10·75	10·78	+ ·05	+ ·02
13 (iii)	11·70	11·45	11·55	+ ·25	+ ·15
—	—	—	Root mean square Δ	·354	·146

To judge by this small sample we obtain only increased inaccuracy by taking the more complicated formula. We shall only make an error of about $1\frac{1}{2}$ mm. if we calculate the longitudinal girth from

$$G_l = 1{\cdot}2701\, B + 1{\cdot}6415\, L + {\cdot}8224,$$

and for the egg of the Common Tern at least this is a convenient formula for verifying measurements in the field.

The remaining correlations indicate sensible correlations, but these correlations might well be substantially higher had a better scale of mottling been adopted *ab initio.* In the first place we see that the mottling and the ground colour are sensibly correlated, and the deeper the ground colour the more intense is the mottling*.

We have already seen (p. 146) that for eggs of the same clutch the broader has less intensity of ground colour and more meagre mottling. This is true for the eggs of the Common Tern in general, although it is probable that a better classification of mottling would bring out more marked correlations. The

---

* This might probably be asserted interracially as well as intraracially, compare for example the swallow with the skylark, the lapwing with the ringed plover, etc.

following are the orders (*a*) of mottling chosen, (*b*) of breadth classes, (*c*) of index classes:

(*a*) Assumed Order of Mottling	(*b*) Order of Breadth		(*c*) Order of Index	
Class	B	Class	B/L	Class
$g+e+d$	a	3·00	a	72·64
a	c	2·99	$f+i$	72·54
b	$g+e+d$	2·97	c	72·30
c	$f+i$	2·96	$g+e+d$	72·27
h	h	2·96	h	71·95
$f+i$	b	2·95	b	70·54
—	Mean	2·98	Mean	72·30

The relationship is small, but exists. It seems reasonable to suppose that the order of mottling classes as given by B or B/L, where there is only one displacement, may be a better one than that we have selected. But if in the mottling order b and c were interchanged, it would agree with the B classification, in so far that the three classes of least and of most mottling in the two classifications would be the same.

We now turn to the ground colour. We see that the ground colour is fainter, when the egg has greater breadth, but that there is no relation of the index to the intensity of ground colour. The results of p. 147 are thus confirmed by the general correlation of ground colour and breadth. Although there is no high-correlation, we may assert that it is probable that the intensity of pigment does not depend on the pressure during transit of the oviduct, but rather on a constant amount of pigment being distributed over a larger surface.

(5) *Homotyposis in Eggs of the same Clutch.*

The homotyposis, or degree of resemblance in character between eggs of the same clutch may be studied on the present material. The chief direct and cross homotypic correlations are given in Table IV.

Pearson has shewn* that the degree of resemblance of undifferentiated 'like organs' might be expected to be equal to that of pairs of brethren, i.e. about ·50, and proved that this is so for many homotypes in the vegetable kingdom, a result which has been since confirmed by much as yet unpublished material from the animal kingdom, including a number of series of birds' eggs. Thus the mean value of the homotyposis for eggs of the Common Tern could hardly be improved upon. Only the colour characters show irregularity, especially the mottling, a

---

* "On Homotyposis in the Vegetable Kingdom," *Phil. Trans.* Vol. 197, A, pp. 285—379, 1900.

feature we have already indicated as difficult to measure. It will be seen that the correlation of the ground colour of an egg with the mottling of a second (·3989) has come out greater than the organic correlation between mottling and ground colour in the same egg (·2260).

TABLE IV.

Homotypic Correlations.

	Symbols	Characters	Correlation
Direct	L, L	Lengths of Eggs in same clutch	·4643 ± ·0346
	B, B	Breadths of Eggs in same clutch	·5176 ± ·0326
	G_l, G_l	Longitudinal Girths of Eggs in same clutch	·5076 ± ·0327
	G_b, G_b	Equatorial Girths of Eggs in same clutch	·4621 ± ·0350
		Mean value	·4879
	M, M	Mottling of Eggs in same clutch	·3500
	C, C	Ground colour of Eggs in same clutch	·5709
		Mean of six characters	·4788
Cross	L, B	Length of one Egg with Breadth of a second	·0922 ± ·0441
	C, M	Ground colour of one Egg with Mottling of a second ...	·3989 ± ·0379
	L, G_l	Length of one Egg with Longitudinal Girth of a second ...	·4229 ± ·0362
	B, G_l	Breadth of one Egg with Longitudinal Girth of a second ...	·2530 ± ·0416
	G_l, G_b	Longitudinal Girth of one Egg with Equatorial Girth of a second	·2603 ± ·0413
Index	B/L, B/L	Indices of two Eggs of same clutch	·5537 ± ·0308
	O, O	Indices of ovality of two Eggs of same clutch	·5527 ± ·0309
	I/O, I/O	Inverse of indices of ovality	·5361 ± ·0317
		Mean of three Index Correlations ...	·5475
		Mean of nine Homotypic Correlations	·5017

We feel that the classification by mottling is at present too uncertain, and that until the result cited has been confirmed with larger numbers and more definite categories, it would be idle to consider whether, while a given bird has usually highly or lowly pigmented eggs both as to ground colour and mottling, yet when in the individual egg there is an excess of mottling pigment, there may be some tendency to a relatively less increase of ground colour. Thus the correlation in the individual egg might possibly be less than the correlation between eggs of the same clutch. Such considerations must be postponed until the fact itself is adequately demonstrated.

Another relation suggested by Pearson[*] is that the cross homotypic correlation of the characters x and y should on the average equal $\frac{1}{2}$ (correlation of x and x + correlation of y and y) × (the organic correlation of x and y). It is clearly impossible from what has just been said to apply this to the cross homotyposis of ground colour and mottling. We can apply it to the five cases in which quantitative measurements have been made. Table V gives the requisite data, the last two columns giving respectively the calculated and observed cross correlations.

TABLE V.

Cross Homotypic Correlations.

Characters		Direct Correlations		Organic Correlation (1) and (2)	Cross Correlation	
(1)	(2)	(1) and (1)	(2) and (2)		Calculated	Observed
L	B	·4643	·5176	·2220	·1090	·0922
L	G_l	·4643	·5076	·8804	·4278	·4229
G_l	G_b	·5076	·4621	·5297	·2568	·2603
B	G_l	·5176	·5076	·5216	·2674	·2530
G_l	B/L	·5076	·5537	– ·3832	– ·2033	– ·2007

When we compare the calculated and observed cross correlations, we see a striking agreement, or the theory that cross homotyposis is the product of direct homotyposis and the organic correlation of the characters under investigation holds very closely for the egg of the Common Tern.

The general results obtained are in good accord with those reached by previous observers, and the authors hope to investigate one or two doubtful points on fuller material this year.

---

[*] *Phil. Trans.* Vol. 197, A, p. 290.

APPENDIX OF CORRELATION TABLE

TABLE A. *Length and Breadth of Egg.*

Breadth.

Length.	2·55—2·59	2·60—2·64	2·65—2·69	2·70—2·74	2·75—2·79	2·80—2·84	2·85—2·89	2·90—2·94	2·95—2·99	3·00—3·04	3·05—3·09	3·10—3·14	3·15—3·19	3·20—3·24	Totals
3·55—3·59	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
3·60—3·64	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
3·65—3·69	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
3·70—3·74	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
3·75—3·79	—	—	—	—	—	1	1	—	—	2	—	—	—	1	5
3·80—3·84	—	—	—	—	—	1	—	—	1	2	—	—	1	—	5
3·85—3·89	1	—	—	—	—	3	1	3	5	2	—	—	—	—	15
3·90—3·94	—	1	—	—	—	1	—	3	3	1	—	1	—	—	10
3·95—3·99	—	—	—	1	—	2	3	4	6	4	4	1	—	—	25
4·00—4·04	—	—	—	—	—	2	5	6	3	5	5	—	—	—	26
4·05—4·09	—	—	—	1	1	—	2	5	5	8	3	2	—	—	27
4·10—4·14	—	—	—	—	1	1	—	13	12	8	5	2	3	—	45
4·15—4·19	—	—	—	—	—	—	3	4	9	3	2	3	—	—	24
4·20—4·24	—	—	—	—	—	1	1	5	5	8	6	2	2	1	31
4·25—4·29	—	—	—	—	—	—	2	4	4	4	5	4	—	—	23
4·30—4·34	—	—	—	—	1	—	2	6	2	4	1	2	2	—	20
4·35—4·39	—	—	—	—	—	—	1	3	4	3	2	—	1	—	14
4·40—4·44	—	—	—	—	1	—	2	1	—	1	2	—	—	—	7
4·45—4·49	—	—	—	—	—	1	—	1	1	—	—	2	—	—	5
4·50—4·54	—	—	—	—	—	—	1	1	—	2	—	1	1	—	6
4·55—4·59	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
4·60—4·64	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1
4·65—4·69	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
4·70—4·74	—	—	—	—	—	—	—	1	1	—	—	—	—	—	2
Totals	1	2	1	2	4	14	24	61	61	57	35	20	10	2	294

TABLE B. *Girth L, and Girth B.* Girth *B.*

Girth *L.*	8·20—8·39	8·40—8·59	8·60—8·79	8·80—8·99	9·00—9·19	9·20—9·39	9·40—9·59	9·60—9·79	9·80—9·99	10·00—10·19	10·20—10·39	10·40—10·59	10·60—10·79	Totals
10·00—10·09	—	1	—	—	—	—	—	—	—	—	—	—	—	1
10·10—10·19	—	—	—	—	—	—	—	—	—	—	—	—	—	0
10·20—10·29	1	1	—	—	—	—	—	—	—	—	—	—	—	2
10·30—10·39	—	—	—	—	—	—	—	—	—	—	—	—	—	0
10·40—10·49	1	—	—	—	2	—	—	—	—	—	—	—	—	3
10.50 –10·59	—	—	1	—	—	—	1	—	—	—	—	—	—	2
10·60—10·69	—	—	—	—	—	1	—	—	—	—	—	—	—	1
10·70—10·79	—	—	—	1	1	—	2	—		—	—	—	—	4
10·80—10·89	—	—	—	—	5	3	1	2	2	—	—	—	—	13
10·90—10·99	—	—	—	1	2	4	7	3	1	—	—	—	—	18
11·00—11·09	—	—	—	1	—	4	9	2	—	—	—	—	—	16
11·10—11·19	—	—	—	—	2	5	9	5	2	—	—	1	—	24
11·20—11·29	—	—	—	—	2	5	10	8	5	—	1	—	—	31
11·30—11·39	—	—	—	1	—	3	8	7	2	—	—	—	1	22
11·40—11·49	—	—	—	—	—	7	13	12	7	—	—	—	1	40
11·50—11·59	—	—	—	—	1	3	8	11	3	2	—	—	—	28
11·60—11·69	—	—	—	1	—	3	5	5	5	5	1	—	2	27
11·70—11·79	—	—		—	1	3	2	6	4	1	—	—	—	17
11·80—11·89	—	—	—	—	—	1	3	2	4	5	2	—	—	17
11·90—11·99	—	—	—	—	1	2	4	—	4	1	—	—	—	12
12·00—12·09	—	—	—	—	—	—	—	—	3	2	1	—	—	6
12·10– 12·19	—	—	—	—	—	—	2	1	—	1	2	—	—	6
12·20—12·29	—	—	—	—	—	—	—	—	—	1	—	—	—	1
Totals	2	2	1	5	17	44	84	64	42	18	7	1	4	291

TABLE C.

Length of Egg and Girth L.

Length.

Girth L.	3·55—3·59	3·60—3·64	3·65—3·69	3·70—3·74	3·75—3·79	3·80—3·84	3·85—3·89	3·90—3·94	3·95—3·99	4·00—4·04	4·05—4·09	4·10—4·14	4·15—4·19	4·20—4·24	4·25—4·29	4·30—4·34	4·35—4·39	4·40—4·44	4·45—4·49	4·50—4·54	4·55—4·59	4·60—4·64	4·65—4·69	4·70—4·74	Totals
10·00—10·09	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
10·10—10·19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
10·20—10·29	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
10·30—10·39	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
10·40—10·49	—	—	—	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
10·50—10·59	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
10·60—10·69	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
10·70—10·79	—	—	—	—	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
10·80—10·89	—	—	—	—	2	2	6	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	13
10·90—10·99	—	—	—	—	—	2	4	4	4	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	18
11·00—11·09	—	—	—	—	—	—	1	1	8	5	1	1	—	—	—	—	—	—	—	—	—	—	—	—	17
11·10—11·19	—	—	—	—	1	—	—	1	5	8	5	3	1	—	—	—	—	—	—	—	—	—	—	—	24
11·20—11·29	—	—	—	—	—	1	—	1	4	6	7	10	3	—	—	—	—	—	—	—	—	—	—	—	32
11·30—11·39	—	—	—	—	—	—	—	—	1	2	6	8	2	1	1	1	—	—	—	—	—	—	—	—	22
11·40—11·49	—	—	—	—	—	—	—	—	—	2	5	10	12	6	2	3	—	—	—	—	—	—	—	—	40
11·50—11·59	—	—	—	—	—	—	—	—	—	—	2	5	2	9	5	5	1	—	—	—	—	—	—	—	29
11·60—11·69	—	—	—	—	—	—	—	—	—	—	—	4	4	5	6	3	4	3	—	—	—	—	—	—	29
11·70—11·79	—	—	—	—	—	—	—	—	—	—	—	2	—	5	3	4	2	—	1	—	—	—	—	—	17
11·80—11·89	—	—	—	—	—	—	—	—	—	—	—	1	—	5	3	2	5	1	—	—	—	—	—	—	17
11·90—11·99	—	—	—	—	—	—	—	—	—	1	—	1	—	—	3	2	1	—	2	2	—	—	—	—	12
12·00—12·09	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	3	1	1	—	—	—	—	6
12·10—12·19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	2	—	1	—	1	6
12·20—12·29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
12.30—12·39	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
12·40—12·49	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	1	0	1	1	5	5	15	10	25	27	27	46	24	31	23	21	14	7	5	6	0	1	0	2	297

TABLE D.

Breadth of Egg and Girth L.

Breadth.

Girth *L.*	*2·55—2·59*	*2·60—2·64*	*2·65—2·69*	*2·70—2·74*	*2·75—2·79*	*2·80—2·84*	*2·85—2·89*	*2·90—2·94*	*2·95—2·99*	*3·00—3·04*	*3·05—3·09*	*3·10—3·14*	*3·15—3·19*	*3·20—3·24*	Totals
10·00—10·09	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
10·10—10·19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
10·20—10·29	1	—	1	—	—	—	—	—	—	—	—	—	—	—	2
10·30—10·39	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
10·40—10·49	—	1	—	—	—	2	—	—	—	—	—	—	—	—	3
10·50—10·59	—	—	—	1	—	—	—	—	1	—	—	—	—	—	2
10·60—10·69	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
10·70—10·79	—	—	—	—	—	1	1	—	2	—	—	—	—	—	4
10·80—10·89	—	—	—	—	—	5	1	3	1	3	—	—	—	—	13
10·90—10·99	—	—	—	1	—	1	4	4	3	5	—	—	—	—	18
11·00—11·09	—	—	—	—	1	1	2	4	7	—	1	—	—	—	16
11·10—11·19	—	—	—	—	1	—	2	10	4	4	2	—	—	1	24
11·20—11·29	—	—	—	—	—	1	1	7	10	6	4	2	1	—	32
11·30—11·39	—	—	—	—	1	—	2	6	5	5	2	1	—	—	22
11·40—11·49	—	—	—	—	—	1	3	8	13	8	6	1	—	—	40
11·50—11·59	—	—	—	—	—	—	2	6	5	9	4	1	1	—	28
11·60—11·69	—	—	—	—	1	—	1	4	4	5	7	4	2	—	28
11·70—11·79	—	—	—	—	—	2	—	3	1	8	1	2	—	—	17
11·80—11·89	—	—	—	—	—	—	1	2	2	1	3	5	2	1	17
11·90—11·99	—	—	—	—	—	—	3	2	2	—	3	1	1	—	12
12·00—12·09	—	—	—	—	—	—	—	—	—	2	2	1	1	—	6
12·10—12·19	—	—	—	—	—	—	—	2	—	1	—	2	1	—	6
12·20—12·29	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
12·30—12·39	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
12·40—12·49	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
Totals	1	2	1	2	4	14	24	61	61	57	35	20	10	2	294

TABLE E. *Girth L and Index* 100 *Breadth/Length.*

Girth *L.*

Index.	*10·00—10·09*	*10·10—10·19*	*10·20—10·29*	*10·30—10·39*	*10·40—10·49*	*10·50—10·59*	*10·60—10·69*	*10·70—10·79*	*10·80—10·89*	*10·90—10·99*	*11·00—11·09*	*11·10—11·19*	*11·20—11·29*	*11·30—11·39*	*11·40—11·49*	*11·50—11·59*	*11·60—11·69*	*11·70—11·79*	*11·80—11·89*	*11·90—11·99*	*12·00—12·09*	*12·10—12·19*	*12·20—12·29*	*12·30—12·39*	*12·40—12·49*	Totals
62·0—63·9	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	2	—	—	1	5
64·0—65·9	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	—	1	4	—	—	—	—	—	8
66·0—67·9	—	—	1	—	1	—	—	—	—	1	1	—	—	1	2	3	4	3	1	2	1	1	—	—	—	22
68·0—69·9	—	—	—	—	—	1	—	—	—	—	1	1	2	3	9	5	5	3	3	—	3	1	—	—	—	37
70·0—71·9	—	—	—	—	—	—	—	1	2	2	2	7	9	8	12	8	6	6	3	3	1	1	1	—	—	72
72·0—73·9	1	—	1	—	—	—	—	1	4	5	3	7	7	2	9	6	5	4	5	2	—	1	—	—	—	63
74·0—75·9	—	—	—	—	2	—	—	1	4	3	8	6	8	4	8	3	4	—	3	1	1	—	—	—	—	56
76·0—77·9	—	—	—	—	—	1	1	1	1	5	1	2	2	3	—	2	2	—	1	—	—	—	—	—	—	22
78·0—79·9	—	—	—	—	—	—	—	—	—	1	—	—	3	—	—	—	—	—	—	—	—	—	—	—	—	4
80·0—81·9	—	—	—	—	—	—	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
82·0—83·9	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
84·0—85·9	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	0	2	0	3	2	1	4	13	18	16	24	32	22	40	28	28	17	17	12	6	6	1	0	1	294

TABLE F. *Length of Egg and Index* 100 *Breadth/Length.*

Length.

Index.	*3·55—3·59*	*3·60—3·64*	*3·65—3·69*	*3·70—3·74*	*3·75—3·79*	*3·80—3·84*	*3·85—3·89*	*3·90—3·94*	*3·95—3·99*	*4·00—4·04*	*4·05—4·09*	*4·10—4·14*	*4·15—4·19*	*4·20—4·24*	*4·25—4·29*	*4·30—4·34*	*4·35—4·39*	*4·40—4·44*	*4·45—4·49*	*4·50—4·54*	*4·55—4·59*	*4·60—4·64*	*4·65—4·69*	*4·70—4·74*	Totals
62·0—63·9	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	—	2	5
64·0—65·9	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	2	2	2	—	—	—	—	8
66·0—67·9	—	—	—	—	—	—	1	1	—	—	1	1	—	1	2	7	5	1	—	2	—	—	—	—	22
68·0—69·9	—	—	—	—	—	—	—	—	1	1	2	1	5	5	7	6	5	2	1	1	—	—	—	—	37
70·0—71·9	—	—	—	—	—	—	—	1	3	4	6	19	11	12	9	2	2	1	1	1	—	—	—	—	72
72·0—73·9	1	—	1	—	—	1	3	—	6	9	5	14	5	9	5	3	1	—	—	—	—	—	—	—	63
74·0—75·9	—	—	—	1	1	—	5	6	10	8	12	6	3	3	—	1	—	—	—	—	—	—	—	—	56
76·0—77·9	—	—	—	—	1	1	6	1	3	4	1	4	—	1	—	—	—	—	—	—	—	—	—	—	22
78·0—79·9	—	—	—	—	—	1	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
80·0—81·9	—	—	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
82·0—83·9	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
84·0—85·9	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	0	1	1	5	5	15	10	25	26	27	45	24	31	23	20	14	7	5	6	0	1	0	2	294

TABLE G.

Breadth of Egg and Index 100 *Breadth/Length.*

Breadth.

Index.	*2·55—2·59*	*2·60—2·64*	*2·65—2·69*	*2·70—2·74*	*2·75—2·79*	*2·80—2·84*	*2·85—2·89*	*2·90—2·94*	*2·95—2·99*	*3·00—3·04*	*3·05—3·09*	*3·10—3·14*	*3·15—3·19*	*3·20—3·24*	Totals
62·0—63·9	—	—	—	—	1	1	—	2	1	—	—	—	—	—	5
64·0—65·9	—	—	—	—	1	—	4	2	1	—	—	—	—	—	8
66·0—67·9	1	1	—	1	1	1	4	9	2	2	—	—	—	—	22
68·0—69·9	—	—	—	1	1	2	5	11	7	7	1	2	—	—	37
70·0—71·9	—	—	—	—	—	4	4	22	20	12	8	1	1	—	72
72·0—73·9	—	1	1	—	—	4	5	8	14	12	9	7	2	—	63
74·0—75·9	—	—	—	—	—	2	1	7	11	17	9	6	3	—	56
76·0—77·9	—	—	—	—	—	—	1	—	5	3	7	2	3	1	22
78·0—79·9	—	—	—	—	—	—	—	—	—	1	1	2	—	—	4
80·0—81·9	—	—	—	—	—	—	—	—	—	3	—	—	—	—	3
82·0—83·9	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
84·0—85·9	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	1	2	1	2	4	14	24	61	61	57	35	20	10	2	294

TABLE H.

Ground Colour and Mottling.

Ground Colour.

Mottling.	$a+b$	c	d	e	f	$g-k$	Totals
$g+e+d$	12 (+1·21)	15 (+3·96)	4 (−5·28)	16 (+ ·70)	11 (−4·55)	15 (+3·96)	73
a	4 (+·16)	5 (+1·07)	4 (+ ·69)	4 (−1·45)	5 (− ·54)	4 (+ ·07)	26
b	3 (−4·24)	9 (+1·59)	3 (−3·23)	14 (+3·73)	13 (+2·56)	7 (− ·41)	49
c	12 (−2·92)	11 (−4·27)	18 (+5·16)	18 (−3·17)	29 (+7·48)	13 (−2·27)	101
h	5 (+3·08)	— (−1·97)	2 (+ ·35)	3 (+ ·28)	— (−2·77)	3 (+1·03)	13
$f+i$	7 (+2·71)	4 (− ·39)	6 (+2·31)	6 (− ·08)	4 (−2·18)	2 (−2·39)	29
Totals	43	44	37	61	62	44	291

TABLE J.

Index 100 Breadth/Length and Mottling.

Index.

Mottling.	62·0—63·9	64·0—65·9	66·0—67·9	68·0—69·9	70·0—71·9	72·0—73·9	74·0—75·9	76·0—77·9	78·0—79·9	80·0—81·9	82·0—83·9	84·0—85·9	Totals
g+*d*+*e*	2	1	2	9	16	21	15	7	—	—	—	—	73
a	—	—	1	4	9	3	6	1	1	—	1	—	26
b	2	1	9	6	15	10	3	3	—	—	—	—	49
c	1	4	8	12	20	22	22	9	2	1	—	1	102
h	—	1	—	2	3	1	5	—	—	—	—	—	12
f+*i*	—	1	2	4	8	4	5	2	1	2	—	—	29
Totals	5	8	22	37	71	61	56	22	4	3	1	1	291

TABLE K.

Breadth of Egg and Mottling.

Breadth of Egg.

Mottling.	2·55—2·59	2·60—2·64	2·65—2·69	2·70—2·74	2·75—2·79	2·80—2·84	2·85—2·89	2·90—2·94	2·95—2·99	3·00—3·04	3·05—3·09	3·10—3·14	3·15—3·19	3·20—3·24	Totals
g+*d*+*e*	—	—	—	—	2	6	4	17	21	10	5	6	2	—	73
a	—	—	—	—	—	1	1	6	4	6	3	3	2	—	26
b	—	1	1	1	—	1	9	13	3	10	7	1	1	1	49
c	1	1	—	—	2	3	8	14	22	19	17	9	5	1	102
h	—	—	—	1	—	1	—	1	4	4	—	1	—	—	12
f+*i*	—	—	—	—	—	2	2	8	7	7	3	—	—	—	29
Totals	1	2	1	2	4	14	24	59	61	56	35	20	10	2	291

TABLE L.

Ground Colour and Index 100 B/L.

Ground Colour.

Index.	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*k*	Totals
62·0—62·9	—	—	1	—	2	—	1	—	—	—	4
63·0—63·9	—		—	—	1	—	—	—	—	—	1
64·0—64·9	—	—	—	—	3	1	—	1	—	—	5
65·0—65·9	—	—	—	1	—	—	1	1	—	—	3
66·0—66·9	—	2	2	3	3	2	—	2	—	—	14
67·0—67·9	1	3	2	1	—	1	—	—	—	—	8
68·0—68·9	1	2	3	2	7	4	1	4	—	—	24
69·0—69·9	2	1	3	2	—	4	—	2	—	—	14
70·0—70·9	1	6	5	4	7	10	—	2	—	—	35
71·0—71·9	2	1	5	5	9	6	4	1	4	—	37
72·0—72·9	—	5	4	7	8	11	2	—	—	—	37
73·0—73·9	1	2	6	—	5	6	2	—	—	1	23
74·0—74·9	2	4	2	4	7	5	1	—	1	—	26
75·0—75·9	1	1	5	5	3	5	3	5	—	—	28
76·0—76·9	1	1	1	1	4	2	3	—	—	—	13
77·0—77·9	—	1	4	—	1	2	1	—	—	—	9
78·0—78·9	—	—	—	1	—	1	—	—	—	—	2
79·0—79·9	—	—	—	—	1	1	—	—	—	—	2
80·0—80·9	—	3	—	—	—	—	—	—	—	—	3
81·0—81·9	—	—	—	—	—	—	—	—	—	—	0
82·0—82·9	—	—	1	—	—	—	—	—	—	—	1
83·0—83·9	—	—	—	—	—	—	—	—	—	—	0
84·0—84·9	—	—	—	—	—	—	—	—	—	—	0
85·0—85·9	—		—	1	-	—	—	—	—	—	1
Totals	12	32	44	37	61	61	19	18	5	1	290

TABLE M.

Ground Colour and Breadth.

Ground Colour.

Breadth.	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*k*	Totals
2·55—2·59	—	—	—	1	—		—	—	—	—	1
2·60—2·64	—	—		—	1	1	—	—	—	—	2
2·65—2·69	—	—	—		1	—	—	—	—	—	1
2·70—2·74	1	—	—	—	1	—	—	—	—	—	2
2·75—2·79	—	—	—	—	1	2	—	1	—	—	4
2·80—2·84	1	1	1	3	2	4	—	2	—	—	14
2·85—2·89	—	3	1	3	8	7	—	2	2	—	26
2·90—2·94	4	8	13	6	12	7	3	6	—	—	59
2·95—2·99	3	3	8	8	11	15	5	5	2	—	60
3·00—3·04	—	7	11	10	10	10	6	1	1	—	56
3·05—3·09	3	4	6	2	10	8	2	—	—	—	35
3·10—3·14	—	4	1	2	3	6	3	—	—	1	20
3·15—3·19	—	2	2	1	1	2	1	1	—	—	10
3·20—3·24	—	—	1	1	—	—	—	—	—	—	2
Totals	12	32	44	37	61	62	20	18	5	1	292

TABLE N.

Breadth of Egg in Pairs of same Clutch.

	2·55—2·59	2·60—2·64	2·65—2·74	2·75—2·79	2·80—2·84	2·85—2·89	2·90—2·94	2·95—2·99	3·00—3·04	3·05—3·09	3·10—3·14	3·15—3·19	Totals
2·55—2·59	—	1	—	—	—	—	—	—	—	—	—	—	1
2·60—2·64	1	—	—	—	—	—	—	—	—	—	—	—	1
2·65—2·74	—	—	—	—	—	—	—	—	—	—	—	—	0
2·75—2·79	—	—	—	—	—	—	1	1	—	—	—	—	2
2·80—2·84	—	—	—	—	2	2	4	2	1	—	1	—	12
2·85—2·89	—	—	—	—	2	4	4	5	1	—	—	—	16
2·90—2·94	—	—	—	1	4	4	10	14	8	5	1	—	47
2·95—2·99	—	—	—	1	2	5	14	16	12	5	—	—	55
3·00—3·04	—	—	—	—	1	1	8	12	12	8	8	2	52
3·05—3·09	—	—	—	—	—	—	5	5	8	4	3	—	25
3·10—3·14	—	—	—	—	1	—	1	—	8	3	4	—	17
3·15—3·19	—	—	—	—	—	—	—	—	2	—	—	—	2
Totals	1	1	0	2	12	16	47	55	52	25	17	2	230

TABLE O.

Length of Egg in Pairs of same Clutch.

	3·75—3·79	3·80—3·84	3·85—3·89	3·90—3·94	3·95—3·99	4·00—4·04	4·05—4·09	4·10—4·14	4·15—4·19	4·20—4·24	4·25—4·29	4·30—4·34	4·35—4·39	4·40—4·44	4·45—4·49	4·50—4·54	Totals
3·75—3·79	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
3·80—3·84	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	2
3·85—3·89	—	—	4	2	2	—	1	1	1	1	—	—	—	—	—	—	12
3·90—3·94	—	—	2	2	3	—	—	—	—	1	—	1	—	—	—	—	9
3·95—3·99	1	—	2	3	6	6	2	2	—	2	—	—	—	—	—	—	24
4·00—4·04	—	—	—	—	6	8	3	2	3	1	4	1	—	—	—	—	28
4·05—4·09	—	—	1	—	2	3	4	4	4	1	1	1	3	—	—	—	24
4·10—4·14	—	2	1	—	2	2	4	8	6	2	3	4	1	—	1	—	36
4·15—4·19	—	—	1	—	—	3	4	6	2	5	1	2	—	—	—	—	24
4·20—4·24	—	—	1	1	2	1	1	2	5	2	2	2	—	1	—	3	23
4·25—4·29	—	—	—	—	—	4	1	3	1	2	—	—	1	—	—	—	12
4·30—4·34	—	—	—	1	—	1	1	4	2	2	—	4	5	—	—	—	20
4·35—4·39	—	—	—	—	—	—	3	1	—	—	1	5	4	—	—	—	14
4·40—4·44	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1
4·45—4·49	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
4·50—4·54	—	—	—	—	—	—	—	—	—	3	—	—	—	—	—	—	3
Totals	1	2	12	9	24	28	24	36	24	23	12	20	14	1	1	3	234

TABLE P.

Girth B in Pairs of same Clutch.

	8·20—8·39	8·40—8·79	8·80—8·99	9·00—9·19	9·20—9·39	9·40—9·59	9·60—9·79	9·80—9·99	10·00—10·19	10·20—10·39	10·40—10·59	10·60—10·79	Totals
8·20— 8·39	2	—	—	—	—	—	—	—	—	—	—	—	2
8·40— 8·79	—	—	—	—	—	—	—	—	—	—	—	—	0
8·80— 8·99	—	—	—	—	1	1	—	—	—	—	—	—	2
9·00— 9·19	—	—	—	4	3	6	1	2	—	—	—	—	16
9·20— 9·39	—	—	1	3	10	18	4	—	—	—	—	1	37
9·40— 9·59	—	—	1	6	18	20	14	8	—	—	—	1	68
9·60— 9·79	—	—	—	1	4	14	24	5	7	1	—	1	57
9·80— 9·99	—	—	—	2	—	8	5	10	4	2	—	—	31
10·00—10·19	—	—	—	—	—	—	7	4	—	—	—	—	11
10·20—10·39	—	—	—	—	—	—	1	2	—	—	—	—	3
10·40—10·59	—	—	—	—	—	—	—	—	—	—	—	—	0
10·60—10·79	—	—	—	—	1	1	1	—	—	—	—	—	3
Totals ...	2	0	2	16	37	68	57	31	11	3	0	3	230

TABLE Q.

Girth L in Pairs of same Clutch.

	10·20—10·29	10·30—10·39	10·40—10·49	10·50—10·59	10·60—10·69	10·70—10·79	10·80—10·89	10·90—10·99	11·00—11·09	11·10—11·19	11·20—11·29	11·30—11·39	11·40—11·49	11·50—11·59	11·60—11·69	11·70—11·79	11·80—11·89	11·90—11·99	12·00—12·09	12·10—12·19	Totals
10·20—10·29	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
10·30—10·39	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
10·40—10·49	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
10·50—10·59	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
10·60—10·69	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
10·70—10·79	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	2
10·80—10·89	—	—	—	—	—	2	—	—	3	1	2	—	1	—	—	2	—	—	—	—	11
10·90—10·99	—	—	—	1	—	—	—	4	2	—	1	2	5	—	—	—	—	—	—	—	15
11·00—11·09	—	—	—	—	1	—	3	2	4	3	—	2	1	—	2	—	—	—	—	—	18
11·10—11·19	—	—	—	—	—	—	1	—	3	6	4	2	—	—	1	2	1	1	—	—	21
11·20—11·29	—	—	—	—	—	—	2	1	—	4	4	6	3	1	—	1	1	2	—	—	25
11·30—11·39	—	—	—	—	—	—	—	2	2	2	6	—	5	—	2	3	—	1	—	—	23
11·40—11·49	—	—	—	—	—	—	1	5	1	—	3	5	6	8	3	—	—	—	—	—	32
11·50—11·59	—	—	—	—	—	—	—	—	—	—	1	—	8	2	2	—	4	2	1	1	21
11·60—11·69	—	—	—	—	—	—	—	—	2	1	—	2	3	2	6	3	1	—	—	—	20
11·70—11·79	—	—	—	—	—	—	2	—	—	2	1	3	—	—	3	2	2	2	—	1	18
11·80—11·89	—	—	—	—	—	—	—	—	—	1	1	—	—	4	1	2	—	2	—	—	11
11·90—11·99	—	—	—	—	—	—	—	—	—	1	2	1	—	2	—	2	2	—	—	—	10
12·00—12·09	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1
12·10—12·19	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	2
Totals	1	0	1	1	1	2	11	15	18	21	25	23	32	21	20	18	11	10	1	2	234

TABLE R.

Mottling in Pairs of Eggs of same Clutch.

	$g+e+d$	a	b	c	h	$f+i$	Totals
$g+e+d$	26 (+5·54)	4 (− ·81)	12 (+2·37)	19 (−5·97)	4 (− ·21)	3 (− ·91)	68
a	4 (− ·81)	4 (+2·87)	1 (−1·27)	7 (+1·12)	— (− ·99)	— (− ·92)	16
b	12 (+2·37)	1 (−1·27)	8 (+3·47)	9 (−2·75)	1 (− ·98)	1 (− ·84)	32
c	19 (−5·97)	7 (+1·12)	9 (−2·75)	40 (+9·52)	5 (− ·14)	3 (−1·77)	83
h	4 (− ·21)	— (− ·99)	1 (− ·98)	5 (− ·14)	2 (+1·13)	2 (+1·19)	14
$f+i$	3 (− ·91)	— (− ·92)	1 (− ·84)	3 (−1·77)	2 (+1·19)	4 (+3·25)	13
Totals	68	16	32	83	14	13	226

TABLE S.

Ground Colour of one Egg with Mottling of the other Egg for Pairs of same Clutch.

Ground Colour.

Mottling.	$a+b$	c	d	e	f	$g-k$	Totals
$g+e+d$	8 (−3·67)	11 (+1·43)	13 (+3·73)	8 (−5·16)	16 (+ ·75)	11 (+2·92)	67
a	1 (−1·96)	1 (−1·43)	— (−2·35)	6 (+2·66)	4 (+ ·13)	5 (+2·95)	17
b	2 (−3·40)	5 (+ ·57)	4 (− ·29)	3 (−3·09)	12 (+4·94)	5 (+1·26)	31
c	17 (+2·72)	10 (−1·71)	10 (−1·35)	24 (+7·89)	15 (−3·67)	6 (−3·88)	82
h	8 (+5·56)	1 (−1·00)	1 (− ·94)	— (−2·75)	4 (+ ·81)	— (−1·69)	14
$f+i$	3 (+ ·74)	4 (+2·14)	3 (+1·20)	3 (+ ·45)	— (−2·96)	— (−1·57)	13
Totals	39	32	31	44	51	27	224

In Tables R—T, the contingency of each cell is given in brackets.

TABLE T.

Ground Colour in Pairs of Eggs of same Clutch.

	a+*b*	*c*	*d*	*e*	*f*	*g*−*k*	Totals
a+*b*	22 (+15·15)	7 (+1·55)	5 (− ·45)	1 (− 7·08)	3 (−5·96)	1 (−3·22)	39
c	7 (+ 1·55)	6 (+1·67)	6 (+1·67)	4 (− 2·42)	4 (−3·12)	4 (+ ·65)	31
d	5 (− ·45)	6 (+1·67)	8 (+3·67)	7 (+ ·58)	4 (−3·12)	1 (−2·35)	31
e	1 (− 7·08)	4 (−2·42)	7 (+ ·58)	20 (+10·47)	12 (+1·43)	2 (−2·97)	46
f	3 (− 5·96)	4 (−3·12)	4 (−3·12)	12 (+ 1·43)	20 (+8·28)	8 (+2·49)	51
g−*k*	1 (−3·22)	4 (+ ·65)	1 (−2·35)	2 (− 2·97)	8 (+2·49)	8 (+5·41)	24
Totals	39	31	31	46	51	24	222

TABLE U.

Length of one Egg with Breadth of the other Egg for Pairs of same Clutch.

Length of one Egg of Pair.

Breadth of Second Egg of Pair.	*3·75—3·79*	*3·80—3·84*	*3·85—3·89*	*3·90—3·94*	*3·95—3·99*	*4·00—4·04*	*4·05—4·09*	*4·10—4·14*	*4·15—4·19*	*4·20—4·24*	*4·25—4·29*	*4·30—4·34*	*4·35—4·39*	*4·40—4·44*	*4·45—4·49*	*4·50—4·54*	Totals
2·55—2·59	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
2·60—2·64	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
2·65—2·74	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
2·75—2·79	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	2
2·80—2·84	—	—	1	1	3	—	1	2	2	1	—	—	—	—	—	1	12
2·85—2·89	—	—	—	1	3	3	—	3	3	1	—	2	—	—	—	—	16
2·90—2·94	—	1	4	1	7	3	6	7	2	5	2	5	3	1	—	—	47
2·95—2·99	1	—	4	2	2	8	9	6	6	4	3	5	4	—	—	1	55
3·00—3·04	—	1	2	2	2	5	2	13	5	4	5	4	6	—	—	1	52
3·05—3·09	—	—	—	—	4	6	2	4	2	3	2	—	1	—	1	—	25
3·10—3·14	—	—	—	1	2	3	3	1	1	4	—	2	—	—	—	—	17
3·15—3·19	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	2
Totals	1	2	12	9	24	28	24	36	22	23	12	18	14	1	1	3	230

TABLE V.

Length and Girth L. in Pairs of same Clutch.

Length of First Egg.

Length Girth of Second Egg.	3·70—3·79	3·80—3·89	3·90—3·99	4·00—4·09	4·10—4·19	4·20—4·29	4·30—4·39	4·40—4·49	4·50—4·59	Totals
10·20—10·39	—	—	1	—	—	—	—	—	—	1
10·40—10·59	—	2	—	—	—	—	—	—	—	2
10·60—10·79	—	1	2	—	—	—	—	—	—	3
10·80—10·99	—	3	8	6	7	1	1	—	—	26
11·00—11·19	1	2	11	13	6	5	1	—	—	39
11·20—11·39	—	2	5	13	19	5	4	—	—	48
11·40—11·59	—	2	3	6	21	10	8	2	1	53
11·60—11·79	—	2	3	8	4	4	15	—	2	38
11·80—11·99	—	—	—	5	2	10	4	—	—	21
12·00—12·19	—	—	—	1	1	1	—	—	—	3
Totals ...	1	14	33	52	60	36	33	2	3	234

TABLE W.

Breadth and Girth L. in Pairs of same Clutch.

Breadth of First Egg.

Length Girth of Second Egg.	2·55—2·59	2·60—2·64	2·65—2·69	2·70—2·74	2·75—2·79	2·80—2·84	2·85—2·89	2·90—2·94	2·95—2·99	3·00—3·04	3·05—3·09	3·10—3·14	3·15—3·19	Totals
10·20—10·39	—	1	—	—	—	—	—	—	—	—	—	—	—	1
10·40—10·59	1	—	—	—	—	—	—	—	1	—	—	—	—	2
10·60—10·79	—	—	—	—	—	1	—	1	1	—	—	—	—	3
10·80—10·99	—	—	—	—	—	1	3	5	8	6	2	1	—	26
11·00—11·19	—	—	—	—	—	4	3	12	8	7	5	—	—	39
11·20—11·39	—	—	—	—	—	2	3	10	13	7	7	5	1	48
11·40—11·59	—	—	—	—	2	—	6	7	12	11	5	5	1	49
11·60—11·79	—	—	—	—	—	3	1	7	6	14	3	4	—	38
11·80—11·99	—	—	—	—	—	1	—	5	6	5	2	2	—	21
12·00—12·19	—	—	—	—	—	—	—	—	—	2	1	—	—	3
Totals ...	1	1	0	0	2	12	16	47	55	52	25	17	2	230

TABLE X.

Girth L. and Girth B. in Pairs of same Clutch.

Breadth Girth of First Egg.

Length Girth of Second Egg.	*8·20—8·39*	*8·40—8·59*	*8·60—8·79*	*8·80—8·99*	*9·00—9·19*	*9·20—9·39*	*9·40—9·59*	*9·60—9·79*	*9·80—9·99*	*10·00—10·19*	*10·20—10·39*	*10·40—10·59*	*10·60—10·79*	Totals
10·20—10·39	1	—	—	—	—	—	—	—	—	—	—	—	—	1
10·40—10·59	1	—	—	—	—	—	—	1	—	—	—	—	—	2
10·60—10·79	—	—	—	—	1	1	1	—	—	—	—	—	—	3
10·80—10·99	—	—	—	—	2	6	12	2	3	—	—	—	1	26
11·00—11·19	—	—	—	—	4	10	14	7	2	1	—	—	1	39
11·20—11·39	—	—	—	—	3	9	11	12	8	4	—	—	1	48
11·40—11·59	—	—	—	2	3	4	14	17	6	3	2	—	—	51
11·60—11·79	—	—	—	—	3	5	7	15	5	2	1	—	—	38
11·80—11·99	—	—	—	—	—	2	9	3	6	1	—	—	—	21
12·00—12·19	—	—	—	—	—	—	—	2	1	—	—	—	—	3
Totals ...	2	0	0	2	16	37	68	59	31	11	3	0	3	232

TABLE Y.

Index 100 *Breadth/Length in Pairs of same Clutch.*

	62	*63*	*64*	*65*	*66*	*67*	*68*	*69*	*70*	*71*	*72*	*73*	*74*	*75*	*76*	*77*	*78*	*79*	*80*	Totals
62	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
63	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
64	—	—	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	2
65	—	—	—	—	—	·5	—	—	1	—	—	—	—	—	—	—	—	—	—	1·5
66	—	—	—	—	—	1	1	—	—	—	—	—	·5	—	—	—	—	—	—	2·5
67	—	—	1	·5	1	1	·5	—	—	2	1	1	·5	—	—	—	—	—	—	8·5
68	—	—	—	—	1	·5	5	1·5	·5	1	1	—	—	—	—	—	—	—	—	10·5
69	1	—	—	—	—	—	1·5	3	1·5	1·5	1·5	—	—	1	·5	—	—	—	—	11·5
70	—	—	—	1	—	—	·5	1·5	—	3	2·5	·5	—	2	·5	—	—	—	—	11·5
71	—	—	1	—	—	2	1	1·5	3	7	7·5	5	3	1·5	—	—	—	—	—	32·5
72	—	—	—	—	—	1	1	1·5	2·5	7·5	14	5	6	6·5	3	1	—	·5	—	49·5
73	—	—	—	—	—	1	—	—	·5	5	5	3	2·5	1·5	2·5	2	—	·5	—	23·5
74	—	—	—	—	·5	·5	—	—	—	3	6	2·5	10	1	2	·5	—	—	—	26
75	—	—	—	—	—	—	—	1	2	1·5	6·5	1·5	1	3	3	2·5	—	—	—	22
76	—	—	—	—	—	—	—	·5	·5	—	3	2·5	2	3	1	2	—	—	—	14·5
77	—	—	—	—	—	—	—	—	—	—	1	2	·5	2·5	2	2	—	—	—	10
78	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	·25	·25	0·5
79	—	—	—	—	—	—	—	—	—	—	·5	·5	—	—	—	—	·25	·5	·25	2
80	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	·25	·25	—	0·5
Totals	1	0	2	1·5	2·5	8·5	10·5	11·5	11·5	32·5	49·5	23·5	26	22	14·5	10	·5	2	·5	230

TABLE Z.

Index $\frac{100\,B/L}{2-B/L}$ *in Pairs of same Clutch.*

	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	Totals
45	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
46	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
47	—	—	—	—	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	2
48	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
49	—	—	—	—	—	·5	·75	·25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1·5
50	—	—	1	—	·5	2	·25	·25	—	1	·5	·25	·75	·5	1	—	—	—	—	—	—	—	8
51	—	—	—	—	·75	·25	1	3	—	—	1·5	·5	·5	—	—	—	—	—	—	—	—	—	7·5
52	1	—	—	—	·25	·25	3	1	2·5	·5	1	1·75	·75	—	—	1	—	—	—	—	—	—	13
53	—	—	—	—	—	—	—	2·5	1	·5	—	·5	—	—	·25	·25	1	—	—	—	—	—	6
54	—	—	—	1	—	1	—	·5	·5	—	3	2·5	·5	·5	·25	1·25	—	—	—	—	—	—	11
55	—	—	1	—	—	·5	1·5	1	—	3	6	5	5·5	2·5	1·5	1·5	—	—	—	—	—	—	29
56	—	—	—	—	—	·25	·5	1·75	·5	2·5	5	8	5	3	4·5	4·5	1·5	—	·5	—	—	—	37·5
57	—	—	—	—	—	·75	·5	·75	—	·5	5·5	5	3	1·5	2·25	1·75	2	—	1·5	—	—	1	26
58	—	—	—	—	—	·5	—	—	—	·5	2·5	3	1·5	2	4·5	·5	3	·5	1	—	—	—	19·5
59	—	—	—	—	—	1	—	—	·25	·25	1·5	4·5	2·25	4·5	3·5	·75	1	1·5	—	—	—	—	21
60	—	—	—	—	—	—	—	1	·25	1·25	1·5	4·5	1·75	·5	·75	1	·5	2	—	—	—	—	15
61	—	—	—	—	—	—	—	—	1	—	—	1·5	2	3	1	·5	4	·5	1·5	—	—	—	15
62	—	—	—	—	—	—	—	—	—	—	—	—	—	·5	1·5	2	·5	3	·5	—	—	—	8
63	—	—	—	—	—	—	—	—	—	—	—	·5	1·5	1	—	—	1·5	·5	—	—	—	—	5
64	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	·5	0·5
65	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	·5	0·5
66	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	·5	·5	—	2
Totals	1	0	2	1	1·5	8	7·5	13	6	11	29	37·5	26	19·5	21	15	15	8	5	·5	·5	2	230

TABLE AA.

2 L/B − 1 in Pairs of same Clutch.

	1·52—1·54	*1·55—1·57*	*1·58—1·60*	*1·61—1·63*	*1·64—1·66*	*1·67—1·69*	*1·70—1·72*	*1·73—1·75*	*1·76—1·78*	*1·79—1·81*	*1·82—1·84*	*1·85—1·87*	*1·88—1·90*	*1·91—1·93*	*1·94—1·96*	*1·97—1·99*	*2·00—2·02*	*2·03—2·05*	*2·06—2·08*	*2·09—2·11*	*2·12—2·14*	*2·15—2·17*	*2·18—2·20*	*2·21—2·23*	Totals
1·52—1·54	—	1	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
1·55—1·57	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
1·58—1·60	—	—	—	1	2	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
1·61—1·63	—	—	1	4	1	3	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
1·64—1·66	—	—	2	1	2	—	1	3	1	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	13
1·67—1·69	—	—	—	3	—	2	4	2	4	4	1	2	—	1	—	—	1	—	—	—	—	—	—	—	24
1·70—1·72	—	—	—	1	1	4	4	1	2	5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	19
1·73—1·75	—	—	2	—	3	2	1	2	5	3	2	—	—	—	—	1	—	—	—	—	—	—	—	—	21
1·76—1·78	1	—	1	1	1	4	2	5	6	5	3	1	—	1	1	1	—	—	—	—	—	—	—	—	33
1·79—1·81	—	—	—	—	2	4	5	3	5	4	6	1	—	2	—	1	1	—	—	—	—	—	—	—	34
1·82—1·84	—	—	—	—	—	1	1	2	3	6	6	—	—	—	1	—	—	—	—	—	1	—	—	—	21
1·85—1·87	—	—	—	—	—	2	—	—	1	1	—	—	1	1	—	—	—	—	—	1	—	—	—	—	7
1·88—1·90	—	—	—	—	1	—	—	—	—	—	—	1	—	2	—	—	—	—	—	—	—	—	—	—	4
1·91—1·93	—	—	—	—	—	1	—	—	1	2	—	1	2	—	3	—	—	—	—	—	—	—	—	1	11
1·94—1·96	—	—	—	—	—	—	—	—	1	—	1	—	—	3	2	—	1	—	—	—	—	—	—	—	8
1·97—1·99	—	—	—	—	—	—	—	1	1	1	—	—	—	—	—	—	1	1	—	—	1	—	—	—	6
2·00—2·02	—	—	—	—	—	1	—	—	—	1	—	—	—	—	1	1	—	—	—	—	—	—	—	—	4
2·03—2·05	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
2·06—2·08	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
2·09—2·11	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
2·12—2·14	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	2
2·15—2·17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
2·18—2·20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
2·21—2·23	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
Totals	2	1	6	11	13	24	19	21	33	34	21	7	4	11	8	6	4	1	0	1	2	0	0	1	230

TABLE BB.

Girth L and Index 100 B/L *in Pairs of same Clutch.*

Length—Girth of First Egg.

Index of Second Egg.	*10·20—10·39*	*10·40—10·59*	*10·60—10·79*	*10·80—10·99*	*11·00—11·19*	*11·20—11·39*	*11·40—11·59*	*11·60—11·79*	*11·80—11·99*	*12·00—12·19*	Totals
62·0—62·9	—	—	—	—	—	—	1	—	—	—	1
63·0—63·9	—	—	—	—	—	—	—	—	—	—	0
64·0—64·9	—	—	—	—	—	—	2	1	—	—	3
65·0—65·9	—	—	—	—	—	—	1	—	—	—	1
66·0—66·9	1	1	—	1	—	—	2	—	2	—	7
67·0—67·9	—	—	—	—	—	3	1	2	1	—	7
68·0—68·9	—	—	—	—	1	1	3	9	1	—	15
69·0—69·9	—	—	—	—	1	1	2	5	—	—	9
70·0—70·9	—	—	1	1	5	6	5	2	2	—	22
71·0—71·9	—	—	—	3	10	13	8	2	3	1	40
72·0—72·9	—	—	—	6	3	6	7	6	7	—	35
73·0—73·9	—	—	—	3	8	3	8	3	3	—	28
74·0—74·9	—	—	1	5	4	7	6	4	2	2	31
75·0—75·9	—	—	1	2	4	2	3	3	—	—	15
76·0—76·9	—	1	—	4	2	2	1	1	—	—	11
77·0—77·9	—	—	—	1	1	2	—	—	—	—	4
78·0—78·9	—	—	—	—	—	1	—	—	—	—	1
79·0—79·9	—	—	—	—	—	1	1	—	—	—	2
Totals	1	2	3	26	39	48	51	38	21	3	232

MISCELLANEA.

I. The Statistical Study of Dietaries, a reply to Professor Karl Pearson.

By Professor D. NOËL PATON, F.R.S.

Professor Pearson's criticism of Miss Lindsay's Study of the Diets of the Labouring Classes in the City of Glasgow (*Biometrika*, Vol. IX. Oct. 1913) is a good example of the danger of one who does not understand the problems involved and who is ignorant of the work already done upon a subject attempting to discredit the results of an investigation by the application of mathematics according to his own fancy and in, what seems to me, a totally illegitimate manner.

Not appreciating the questions which were under investigation, he starts his criticism by demanding that our studies should afford a solution of problems other than those we had before us, and, because he does not find the solution of these problems, he proceeds to abuse the work.

Apparently in his opinion the object of the studies should have been to determine what effect the diets which the families were taking at the time of the study had upon the physique of the various individuals. He states that, if adequate anthropometric observations had been secured in such a study, it would have been at once possible to co-relate these with the diets. It is unnecessary to point out, as was pointed out in the Report, that the physique is determined by the whole previous condition of life and by the influence of heredity, and that it is absurd to attempt to relate it *solely* to the diet (Report, pp. 3 and 4).

The objects of the studies are quite clearly stated on p. 4 of the Report: "Do the working classes of this city get such a diet as will enable them to develop into strong, healthy, energetic men, and, as men, will enable them to do a strenuous day's work; or are the conditions of the labouring classes such that a suitable diet is not obtainable? Further, if a suitable diet is obtainable, and is obtained, is it procured, or can it be procured, at a cost low enough to leave a margin sufficient to cover the other necessary expenses of the family life, with something over for those pleasures and amenities without which the very continuance of life is of doubtful value?"

It was accepted as proved by previous work that for the labouring classes: "If a family diet......gives a yield of energy of less than 3500 Calories per man per day it is insufficient for active work, and if less than 3000 it is quite inadequate for the proper maintenance of growth and normal activity."

The first question investigated was: "Did the families examined receive this supply of energy?" As regards the poorest classes this was answered in the negative. The validity of this conclusion has not been challenged by Professor Pearson.

The second question considered was whether the diets contained a sufficient supply of protein. Previous work indicates that this is probably something above 110 grms. per man per diem. It was shown that in families with regular incomes of over 20*s.* a week the average protein intake was above 110 grms., and that in families with regular incomes and in those with irregular incomes of under 20*s.* a week the average protein intake was under 110 grms. This conclusion has not been refuted.

Accepting our premises, the final conclusion was (p. 27) "that while the labouring classes with a regular income of over 20*s.* a week generally manage to secure a diet approaching the proper standard for active life, those with a smaller income and those with an irregular income entirely fail to get a supply of food sufficient for the proper development and growth of the body and for the maintenance of the capacity for active work."

The main points proposed for the study were thus elucidated.

The part of the Report to which Professor Pearson specially directs his criticism is not the main problem, but that dealt with on pp. 30 and 31—The Physique of Children in Relationship to Diet, a subject taken up at the suggestion of Dr Chalmers. Professor Pearson, having declared the data totally insufficient, proceeds to apply his statistical methods not to refute Miss Lindsay's conclusion, but to demolish other conclusions upon the relationship of physique to income which were never deduced by us.

The very guarded conclusion in the Report was: "These show very markedly the relationship between the physique and the food. *When the weight is much below the average for that age almost without exception the diet is inadequate.*"

Weights alone were considered. Thirty-six children, boys and girls, were dealt with. As the relationship of weight to income was not under consideration, they were classified not according to the income but according to the energy value of the family diet. Hence Professor Pearson's remarks upon this point are quite beside the mark.

I give below, in a re-arranged form, the Table from Appendix IV. The individuals are placed in two groups according to the energy value of their diets, with, opposite each child, the average weight for the age, taken from the Report of the Anthropometric Committee published in the Transactions of the British Association for the Advancement of Science, 1883, and with the difference between the weight of the child and the average weight. The differences between groups 1 and 2 are sufficiently marked and warrant the conclusion as stated above.

That is, of the children in families the diets of which yielded more than 3000 Calories per man per day:

10 were above the standard or not more than 5 lbs. below it,
8 were more than 5 lbs. below it,

while of the children in families in which the diet yielded less than 3000 Calories

3 were above the standard or not more than 5 lbs. below it,
15 were more than 5 lbs. below it.

It must be remembered that the 'standard' is for the children of all classes and not for those of the poorer classes.

The fact that the average age of the children in the second group was about 1¾ years greater than that of the children in the first group does not account for the marked difference.

The last question which Miss Lindsay had to consider was, how the necessary supply of energy and of protein might be supplied without increased expenditure, and she was right in stating that these can be more cheaply purchased in vegetable than in animal foods. She

TABLE A.

Family Diets above 3000 *Calories per Man per Day.*

Number	Calories	Age in Years	Sex	Weight in lbs.	Standard Weight in lbs.	Difference
2	4003	7	♀	39	46·7	− 7·7
2	4003	10	♂	63	67·5	− 4·5
2	4003	8	♂	50	54·9	− 4·9
2	4003	5	♂	35	39·9	− 4·9
36	4091	3·25	♂	35	34·0	− 1·0
4	3882	8	♀	45	52·2	− 7·2
32	3822	6·25	♀	39	42·4	— 3·4
4	3882	6	♀	39	42·4	− 3·4
4	3882	10	♂	56	67·5	−11·5
39	3422	10·5	♀	55	65	−10·0
50	3471	6·25	♀	37	42·4	− 5·4
50	3215	6	♀	47	42·4	+ 4·6
31	3116	6	♀	43	42·4	+ 0·6
18	3248	5·5	♀	43	41·0	+ 2·0
54	3282	5	♀	33	39·6	− 6·6
58	3030	6	♂	38	44·4	− 6·4
*30**	3136	5·5	♂	21	41	−20·0
49	3341	5·5	♂	42	41	+ 1

* Family with rickets.

TABLE B.

Family Diets below 3000 *Calories per Man per Day.*

Number	Calories in Diet	Age in years	Sex	Weight in lbs.	Standard Weight in lbs.	Difference
14	2690	13	♀	76	87	−11·0
14	2690	12	♀	60	76·4	−16·4
14	2690	10	♀	45·5	62·0	−17·5
15	2936	10	♀	56	62·0	− 6·0
17	2931	9·75	♀	44	62·0	−18·0
55	2686	5·75	♀	42	42·4	− 0·4
14	2690	9	♂	45	60·4	−15·4
41	2723	6·75	♂	53	49·7	+ 3·3
14	2690	6	♂	36	44·4	− 8·4
57	2974	5	♂	37	39·9	− 2·9
3	2891	5	♂	37	39·9	− 2·9
42	2772	5·5	♀	34	41·0	− 7·0
24	2412	11	♀	39	68·1	−29·1
21	2329	9	♀	37·5	55·5	−18·0
24	2412	6	♀	28	42·4	−14·4
21	2329	11	♂	60	72·0	−12·0
10	2435	8	♂	43	54·9	−11·9
59	1978	5	♂	26	39·9	−13·9

undoubtedly starts with the well-known conclusion that a Calorie in the food absorbed in a mixed diet from whatever source, protein, fat or carbohydrate is of equal dynamic value. Previous work amply justifies this.

She was not foolish enough to attempt to draw any conclusion from her investigations as to the relative value of animal and vegetable food in the diets on the physical development of the individuals.

Professor Pearson seems entirely unable to grasp the fundamental fact that the physical development of the individual depends largely upon his past conditions of life. To co-relate it with the special constituents of the food which he habitually eats will require not only an enormous series of studies, but a full investigation of the character of the various food stuffs and of the mode of cooking.

These points I tried to explain to him when I wrote to him in summer. He did not write to me as, in his criticism, he says he did. Miss Lindsay forwarded to me a letter from him to her, and I wrote a reply to Professor Pearson which he did not acknowledge.

In conclusion I would say that before he expects his criticism of a physiological problem to be taken seriously, he had better make some attempt to understand the nature of the problem. Certainly it is not my intention to waste time in replying further to his criticism unless in the future it is more pertinent than is his present contribution.

II. The Statistical Study of Dietaries. A Rejoinder.

By KARL PEARSON, F.R.S.

I PUBLISH Professor Noel Paton's reply because it is very typical of the type of difficulty which we meet with at present, when we assert that what is really statistical work must be undertaken only by the adequately trained statistician and that when it is not, then the investigation cannot be considered as falling into the field of science.

Professor Paton states that the following question given on p. 4 of the Report formulated its object: "Do the working classes of this city get such a diet as will enable them to develop into strong, healthy, energetic men, and as men, will enable them to do a strenuous day's work; or are the conditions of the labouring classes such that a suitable diet is not obtainable?"...

Now Professor Paton either assumes that the sample taken of the diet of the individual family was their customary diet, or he does not. If he does, then the question: Was the diet such as would enable the working classes "to *develop* into strong, healthy, energetic men"? has meaning. If he does not, not only is it idle, but the section dealing with the physique of the children on the basis of a sample diet taken as a rule *for a week* (occasionally for a fortnight), is beside the point.

But anyhow, I ask how he can possibly ascertain how the working classes will "develop into strong, healthy, energetic men," if he does not take an adequate anthropometric survey of the families subjected to the dietaries recorded? He says that it is accepted and proved that "If a family diet...gives a yield of energy of less than 3500 calories per man per day it is insufficient for active work; and if less than 3000, it is quite inadequate for the proper maintenance of growth and normal activity." He further assumes with Miss Lindsay that calories from animal and vegetable foods have equal "dynamic value." I assert that neither of these conclusions, which he accepts, are based on adequate research and they are in fact refuted by Miss Lindsay's own material. For, if it can be shown that animal and vegetable calories have different results on the physical development of the children, it is clear that the first statement as to how many calories are needful for the proper maintenance of growth has no significance until a statement is made with regard to the source of the calories. Professor Paton cites no evidence for his statements; from what I have read on the subject of calories, I feel convinced that most of the data on the matter would not stand for five minutes any adequate statistical analysis. The *Report*, Professor Paton tells us, shows "very markedly the relationship between the physique and the food." Yet in a previous paragraph he says "that the physique is determined by the whole previous condition of life and by the influence of heredity, and that it is absurd to attempt to relate it *solely* to the diet."

Now the only way to ascertain whether there was a *marked* relationship between the food and the physique of the children was to correlate the two *for a constant age* and investigate whether the correlations were such, *having regard to their probable errors*, that they could be considered significant. I did this with the result that the total calories in the food and the girls' weight for constant age was not definitely significant with regard to the probable error, while in the case of the boys the probable error was so large that it was impossible to say whether the relationship was really considerable or not. In fact no *marked* relationship could be deduced from Miss Lindsay's data, they were too inadequate. If Professor Paton's statement as to the influence of heredity is to be trusted, then even my correction for age was inadequate, and the data ought to be corrected also for physique of parent! If so, why was the parent not measured?

Professor Paton places before the readers of *Biometrika* two tables on which this "marked" relationship is asserted by him to rest. One of the cases in his Table A, No. 32, is erroneously placed in this table; the details show that the number of calories was 2949 and not 3822*; it should be in Table B. These tables contain 16 boys' weights and 20 girls' weights. Professor Paton takes the British Association measurements, which are, of course, wholly inadequate as a test of Glasgow children, and making no real correction for age† considers whether the children in the two tables were or were not above the quite arbitrary limit of 5 lbs. below standard. He gives us no measure at all of the significance of the result, which is based on the vagaries of sampling 16 boys of ages from 3 to 11, and 20 girls from 5 to 13; and he supposes in some way that this treatment can possibly refute the correlation coefficient, ${}_ar_{wC_F}$, of weight and food calories for constant age with its probable error! I can, however, throw more light on the matter. Owing to the great courtesy of Dr Chalmers, Medical Officer of Health for Glasgow, I have been able to more than treble the number of weights of the boys and girls subjected to the dietaries. The results for total calories in food, C_F, now are‡:

Girls, 69	Boys, 55
${}_ar_{wC_F} = +\cdot 21 \pm \cdot 08$,	${}_ar_{wC_F} = +\cdot 05 \pm \cdot 09$.

Thus the relation for boys is now quite insignificant, and for girls may well be insignificant also. At any rate although both correlations are positive, there is no "marked" relationship between the physique and the dietary. Of course, it may be said that these weights (w) have been taken at some interval after the dietaries were recorded, but unless we assume the dietary to be a rough measure of the permanent feeding of the family, whose physique has been gradually built up for years before the dietaries were recorded, the observations must be discarded as of no value at all for testing physique, or as Professor Paton phrases it "development."

---

* In the Appendix V of Rickety Families, it is given again; this time as 2329 calories.

† The deviation at each age would have to be measured in terms of the standard-deviation of weight at that age; naturally the deviations are larger for older children.

‡ I have to thank Miss B. M. Cave for the present series of correlations.

But the most interesting point ascertained from the new material is the confirmation of the result that the higher the proportion of animal to vegetable calories the greater the weight. In *Biometrika*, Vol. IX, p. 533, we had for 16 boys and 20 girls:

Boys: $_{a}r_{w,\ c_V/c_A} = -\cdot 23 \pm \cdot 16,$

Girls: $_{a}r_{w,\ c_V/c_A} = -\cdot 12 \pm \cdot 15.$

We now have for 55 boys and 69 girls:

Boys: $_{a}r_{w,\ c_V/c_A} = -\cdot 30 \pm \cdot 08,$

Girls: $_{a}r_{w,\ c_V/c_A} = -\cdot 24 \pm \cdot 08.$

These results seem to indicate that Miss Lindsay and Professor Paton, who supports her view, are in error when they consider a calory the same whether it be from animal or vegetable food. On the other hand, our larger numbers now indicate that:

(i) For a constant age the *expenditure* on vegetable or on animal food has no sensible relation to weight.

(ii) For a constant age the number of calories in vegetable food has no sensible relation to weight.

(iii) For a constant age the number of calories in animal food has a positive correlation with weight for both girls and boys, being definitely significant in the first case ($+\cdot 32 \pm \cdot 07$) and not so in the second ($+\cdot 08 \pm \cdot 09$).

(iv) For a constant age the correlations of weight with ratio of expenditure on vegetable and animal foods are for both boys and girls quite insignificant as compared with their probable errors.

I am extremely obliged to Dr Chalmers for doing his best to supply additional material. As far as it goes, it tends to show that calories are of far more importance than expenditures, but that calories from animal food are more closely related to physique than are calories from vegetable food*. The new material supports my criticisms that the failure to distinguish between animal and vegetable calories stultified the advice given by Miss Lindsay, i.e. to spend money on oatmeal rather than on eggs. It also indicates that no safe conclusions with regard to dietaries can be drawn until a reasonable anthropometric survey accompanies the record of dietaries, and the whole is reduced with adequate statistical knowledge.

One point I can allow Professor Paton. It was an oversight on my part, when I said that I had written to both Miss Lindsay and to himself; the letters in which Miss Lindsay and he stated that to follow up the families now would be impossible were both replies to one and the same letter of mine addressed to Miss Lindsay. The additional facts I desired were in their opinion unascertainable, and further correspondence did not seem to me likely to be of any service in achieving the end I had in view, namely to render of real service to science a piece of recording work from which in my opinion then and in my opinion still, very misleading conclusions had been drawn, and which conclusions in their turn had been exaggerated in the press résumés of the paper. I do not think any such work as that done on dietaries by Miss Lindsay and Professor Noel Paton will be of real value until (i) these dietaries are accompanied by a thorough anthropometric survey of the whole families of the dieted and (ii) the equality of animal and vegetable food calories ceases to be considered as a dogmatic truth.

---

* Of course the results show that on such data as are available, the food has relatively little relation to the weight, there is no "marked" relationship.

III. Note on the essential Conditions that a Population breeding at random should be in a Stable State.

By K. PEARSON, F.R.S.

Let us deal with bi-parental inheritance in the first place. Let x be a character in the father, mean $\bar{x}$, standard deviation σ_1; let y be the same character in the mother, $\bar{y}$ its mean, and σ_2 its standard deviation. Let z be the character in offspring of one sex, σ_3 be the standard deviation of all offspring of this sex and $\bar{z}$ the mean. Let μ_2', μ_3', μ_4'; $\mu_2'', \mu_3'', \mu_4''$; and $\mu_2''', \mu_3''', \mu_4'''$, be the moment coefficients about the means respectively of father, mother and offspring frequency distributions. Let $\bar{z}_{xy}$ be the mean of the offspring of those parents, who have characters x and y, and let the array of frequency of such offspring be given by $f_3(u)\,du$ about $\bar{z}_{xy}$, i.e. the character of any offspring in this array is $\bar{z}_{xy}+u$, where u is independent of the parental characters x and y, but $\bar{z}_{xy}$ is a function of x and y the parental characters. Some writers have suggested that the offspring character should be taken as a blend of the parental characters, i.e.

$$z=\tfrac{1}{2}(x+y),$$

understanding by blend the *mean* of the parental characters. This appears to be very unsatisfactory for:

(*a*) It supposes the parental characters to fix absolutely the offspring characters which is far from a result of experience.

(*b*) It supposes the mother to reproduce the female size of character in the male and the female offspring alike, whereas she contributes to each the sex character of her own stock, i.e. if she is a tall woman, she would contribute absolutely more to a son than to a daughter. The late Sir Francis Galton got over this difficulty by "reducing female measures to their male equivalents." This he did by altering absolute measurements in the ratio of male to female mean measurements. Thus he would take for the mean of his array of offspring

$$\bar{z}_{xy}=\tfrac{1}{2}\left(x+\frac{\bar{x}}{\bar{y}}y\right)$$

if he were dealing with male offspring. A more reasonable hypothesis is to assume that

$$z=\tfrac{1}{2}\sigma_3\left(\frac{x}{\sigma_1}+\frac{y}{\sigma_2}\right)+u \quad\text{..........(i).}$$

This will practically agree with Sir Francis's form, if the coefficients of variation in the two sexes are the same, i.e. $\sigma_1/\bar{x}=\sigma_2/\bar{y}$.

If we measure u from the mean of the array of offspring we have

$$\bar{z}=\tfrac{1}{2}\sigma_3\left(\frac{\bar{x}}{\sigma_1}+\frac{\bar{y}}{\sigma_2}\right) \quad\text{..........(ii).}$$

We shall now suppose the offspring to follow the law (i), or

$$z-\bar{z}=\tfrac{1}{2}\sigma_3\left(\frac{x-\bar{x}}{\sigma_1}+\frac{y-\bar{y}}{\sigma_2}\right)+u \quad\text{..........(iii),}$$

where x and y are uncorrelated (mating at random), and u represents other influences than the parental, and is therefore uncorrelated with x and y*. The frequency distributions of x and y

---

* This assumes the homoscedasticity of the arrays of offspring due to pairs of fathers and mothers with characters x and y.

may be taken as given by $f_1(x-\bar{x})$ and $f_2(y-\bar{y})$. Let $N_1 \times N_2$ be the total number of possible matings

$$=\iint f_1(x-\bar{x})\,f_2(y-\bar{y})\,dx\,dy$$

and the total number of offspring N_3 in any array

$$=\int f_3(u)\,du.$$

I now propose to give the expression for the nth moment coefficient about the mean, i.e. μ_n''', of the population of offspring of a given sex. We have

$$N_1 \times N_2 \times N_3 \times \mu_n''' = \iiint \left\{\tfrac{1}{2}\sigma_3{}^n \left(\frac{x-\bar{x}}{\sigma_1}+\frac{y-\bar{y}}{\sigma_2}\right)+u\right\}^n f_1(x-\bar{x})\,f_2(y-\bar{y}) \times f_3(u)\,dx\,dy\,du,$$

the integration being extended over the whole of the frequency distributions of father, mother and offspring. Thus

$$\mu_n''' = \frac{\sigma_3{}^n}{N_1 N_2 N_3}\iiint \sum_{s=0}^{s=n}\left[\frac{\underline{|n}}{\underline{|n-s}\,\underline{|s}} \sum_{t=0}^{t=n-s}\left\{\frac{\underline{|n-s}}{\underline{|n-s-t}\,\underline{|t}}\,\frac{\{\frac{1}{2}(x-\bar{x})\}^{n-s-t}\{\frac{1}{2}(y-\bar{y})\}^t}{\sigma_1{}^{n-s-t}\,\sigma_2{}^t}\right\}\frac{u^s}{\sigma_3{}^s}\right]$$
$$\times f_1(x-\bar{x})\,f_2(y-\bar{y})\,f_3(u)\,dx\,dy\,du.$$

Now x, y and u being independent we have

$$\frac{1}{N_1}\int (x-\bar{x})^{n-s-t} f_1(x-\bar{x})\,dx = \mu'_{n-s-t}$$

$$\frac{1}{N_2}\int (y-\bar{y})^t f_2(y-\bar{y})\,dy = \mu_t''$$

$$\frac{1}{N_3}\int u^s f_3(u)\,du = \mu_s{}^{\text{iv}}.$$

Thus
$$\mu_n''' = \sigma_3{}^n \sum_{s=0}^{s=n}\left[\frac{\underline{|n}}{2^{n-3}\,\underline{|n-s}\,\underline{|s}} \sum_{t=0}^{t=n-s}\left\{\frac{\underline{|n-s}}{\underline{|n-s-t}\,\underline{|t}}\,\frac{\mu'_{n-s-t}}{\sigma_1{}^{n-s-t}}\,\frac{\mu_t''}{\sigma_2{}^t}\,\frac{\mu_s{}^{\text{iv}}}{\sigma_3{}^s}\right\}\right] \text{..............(iv).}$$

Thus we reach, remembering that $\mu_1' = \mu_1'' = \mu_1''' = 0$,

$$\mu_2''' = \frac{1}{4}\sigma_3{}^2\left(\frac{\mu_2'}{\sigma_1{}^2}+\frac{\mu_2''}{\sigma_2{}^2}\right)+\mu_2{}^{\text{iv}} \text{..........................(v),}$$

$$\mu_3''' = \frac{1}{8}\sigma_3{}^3\left(\frac{\mu_3'}{\sigma_1{}^3}+\frac{\mu_3''}{\sigma_2{}^3}\right)+\mu_3{}^{\text{iv}} \text{..........................(vi),}$$

$$\mu_4''' = \frac{1}{16}\sigma_3{}^4\left(\frac{\mu_4'}{\sigma_1{}^4}+6\,\frac{\mu_2'}{\sigma_1{}^2}\,\frac{\mu_2''}{\sigma_2{}^2}+\frac{\mu_4''}{\sigma_2{}^4}\right)+\frac{3}{2}\sigma_3{}^2\left(\frac{\mu_2'}{\sigma_1{}^2}+\frac{\mu_2''}{\sigma_2{}^2}\right)\mu_2{}^{\text{iv}}+\mu_4{}^{\text{iv}} \text{...............(vii).}$$

But $\mu_2' = \sigma_1{}^2$, $\mu_2'' = \sigma_2{}^2$, and $\mu_2''' = \sigma_3{}^2$. Hence we must have

$$\mu_2{}^{\text{iv}} = \tfrac{1}{2}\sigma_3{}^2 \text{..........................(viii).}$$

If as usual we take $\beta_1 = \mu_3{}^2/\mu_2{}^3$ and $\beta_2 = \mu_4/\mu_2{}^2$ we find from (vi) and (vii), writing $s^2 = \mu_2{}^{\text{iv}}$

$$\sqrt{\beta_1{}^{\text{iv}}} = \left(\frac{\sigma_3}{s}\right)^3\left\{\sqrt{\beta_1'''}-\frac{1}{8}\left(\sqrt{\beta_1'}+\sqrt{\beta_2'}\right)\right\} \text{..........................(ix),}$$

$$\beta_2{}^{\text{iv}} = \frac{\sigma_3{}^4}{s^4}\left\{\beta_2'''-\frac{1}{16}(\beta_2'+\beta_2''+6)-3\,\frac{\sigma_2{}^2}{s^2}\right\} \text{..........................(x).}$$

Whence by the use of (viii)

$$\sqrt{\beta_1{}^{\text{iv}}} = 2\sqrt{2}\left\{\sqrt{\beta_1'''}-\frac{1}{8}\left(\sqrt{\beta_1'}+\sqrt{\beta_1''}\right)\right\} \text{..........................(xi),}$$

$$\beta_2{}^{\text{iv}} = 4\left\{\beta_2'''-\frac{1}{16}(\beta_2'+\beta_2''+6)\right\}-6 \text{..........................(xii).}$$

Hence in order that the offspring population should be stable, it is needful that in the array of offspring for given parents :

(*a*) $$s = \frac{1}{\sqrt{2}}\sigma_3.$$

(*b*) $$\sqrt{\beta_1^{\text{iv}}} = 2\sqrt{2}\left\{\sqrt{\beta_1'''} - \frac{1}{8}\left(\sqrt{\beta_1'} + \sqrt{\beta_1''}\right)\right\} = 2\sqrt{2}\sqrt{\beta_1'''}\left(1 - \frac{1}{4}\right) = \frac{3\sqrt{2}}{2}\sqrt{\beta_1'''},$$

if $\beta_1''' = \beta_1' = \beta_1''$, i.e. the skewness be the same for fathers, mothers and offspring.

$$\beta_2^{\text{iv}} = \frac{1}{2}(7\beta_2''' - 15),$$

if $$\beta_2''' = \beta_2' = \beta_2''.$$

Thus, we have for the array of offspring of given parents

$$\left.\begin{aligned} s &= \frac{1}{\sqrt{2}}\sigma_3 \\ \beta_1^{\text{iv}} &= \frac{9}{2}\beta_1''' \\ \beta_2^{\text{iv}} - 3 &= \frac{7}{2}(\beta_1''' - 3) \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(xiii)}.$$

Accordingly the variability of the array is less than that of the population of offspring ; and the array (unless $\beta_1''' = 0$, $\beta_1''' = 3$) is more skew and has greater kurtosis than the general population.

If r_{12}, r_{23}, r_{31} be the three correlations of father, mother and offspring we know that the mean standard-deviation of the offspring of arrays having the same parents is

$$s' = \sigma_3\sqrt{\frac{1 - r_{13}^2 - r_{23}^2 - r_{12}^2 + 2r_{12}r_{13}r_{31}}{1 - r_{12}^2}},$$

and this equals if there be no assortative mating

$$(r_{12} = 0), \quad \sigma_3\sqrt{1 - r_{13}^2 - r_{23}^2}.$$

If we could assume this equal to s we must have, since

$$s = \frac{1}{\sqrt{2}}\sigma_3,$$

$$\frac{1}{\sqrt{2}} = \sqrt{1 - r_{13}^2 - r_{23}^2},$$

leading to $$r_{13}^2 + r_{23}^2 = \tfrac{1}{2},$$

or if the two parental correlations are equal to

$$r_{13} = r_{23} = \cdot 5.$$

In other words, if the parental influences were equal and there were no assortative mating and the character in the array of offspring had the mean value

$$\frac{1}{2}\sigma_3\left(\frac{x}{\sigma_1} + \frac{y}{\sigma_2}\right),$$

then the population could only be stable if

$$r_{13} = r_{23} = 0 \cdot 5.$$

But this apparently noteworthy result only begs the question. By the general theory of correlation the mean of the array of offspring is

$$\bar{z} + \sigma_3\left(\frac{r_{13} - r_{12}r_{23}}{1 - r_{12}^2}\,\frac{x - \bar{x}}{\sigma_1} + \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2}\,\frac{y - \bar{y}}{\sigma_2}\right),$$

$$= \bar{z} + \sigma_3\left(r_{13}\frac{x - \bar{x}}{\sigma_1} + r_{23}\frac{y - \bar{y}}{\sigma_2}\right)$$

if there be no assortative mating,

$$= \sigma_3\left(r_{13}\frac{x}{\sigma_1} + r_{23}\frac{y}{\sigma_2}\right) + \bar{z} - \sigma_3\left(\frac{r_{13}\bar{x}}{\sigma_1} + \frac{r_{23}\bar{y}}{\sigma_2}\right).$$

Hence if we *assume* the mean of array of offspring to be given by

$$\sigma_3 \frac{1}{2}\left(\frac{x}{\sigma_1}+\frac{y}{\sigma_2}\right)$$

(i) the second portion of the expression must be zero, i.e. mean of whole population of offspring must coincide with mean of array of offspring where parents have the mean values and (ii) we must have $r_{13}=r_{23}=\frac{1}{2}$. In other words the form of our assumption involves both the equal influence of the parents and the value of the parental correlation.

From the standpoint of heredity no such assumption is legitimate. Neither in Mendelian theory nor in biometric formula, nor again in actual observation is it permissible to suppose that the mean of the array of offspring is determined solely by the parents. Still less is it possible to suppose the actual character of the offspring to be the mean of that of the parents (i.e. put $u=0$). If it were we should have $z=\frac{1}{2}(x+y)$, whence flow

$$\left.\begin{aligned} \mu_2{}''' &= \frac{1}{4}(\mu_2{}'+\mu_2{}'') \\ \mu_3{}''' &= \frac{1}{8}(\mu_3{}'+\mu_3{}'') \\ \mu_4{}^{\text{iv}} &= \frac{1}{16}(\mu_4{}'+6\mu_2{}'\mu_2{}''+\mu_4{}'') \end{aligned}\right\}\quad\text{..................................(xiv).}$$

But these equations assume that $\mu_2{}^{\text{iv}}$, $\mu_3{}^{\text{iv}}$ and $\mu_4{}^{\text{iv}}$ are all zero—an absurdity in itself and contrary to all experience, whether biometric or Mendelian. For non-assortative mating and equal potency of parents, they lead to parental correlations of the order ·7 and to an impossibility of stability in any population*.

In fact any such relations as (xiv) are inconceivable on the basis of both biometric as well as Mendelian theory and observation. Parental correlations have never been observed anywhere near such a value as 0·7. Equations (xiii) are, however, suggestive; they show that if the parental distribution be symmetrical and mesokurtic, the array of offspring will remain so after selection; but if the parental distribution does not possess these characters, then any selection of individual parents will emphasize the asymmetry and the kurtosis in the resulting array of offspring; or continued selection of this type will lead to greater and greater divergence from the normal or Gaussian frequency distribution.

---

* If we assume that the mean of the array of offspring of parents of characters x and y is given by $lx+my$, it is only another way of asserting that the regression is linear and that

$$l=\frac{r_{12}-r_{13}r_{23}}{1-r_{23}{}^2}\frac{\sigma_3}{\sigma_1},\quad m=\frac{r_{13}-r_{12}r_{23}}{1-r_{23}{}^2}\frac{\sigma_3}{\sigma_2}.$$

If we make $l=m$, or give equal weight to the parents, it is only rational to suppose that $\sigma_1=\sigma_2$ and $r_{12}=r_{13}$, which lead us to

$$l=m=\frac{r_{12}}{1+r_{23}}\frac{\sigma_3}{\sigma_1}.$$

Hence the mean of the array is $$\frac{r_{12}}{1+r_{23}}\frac{\sigma_3}{\sigma_1}(x+y),$$

and whether we make x constant and y constant or $x+y$ constant leads to precisely the same variability in the array, i.e.

$$s=\sigma_3\sqrt{\frac{1-r_{12}{}^2-r_{13}{}^2-r_{23}{}^2+2r_{12}r_{13}r_{23}}{1-r_{23}{}^2}}=\sigma_3\sqrt{1-\frac{2r_{12}{}^2}{1+r_{23}}}.$$

If assortative mating be zero, this equals

$$\sigma_3\sqrt{1-2r_{12}{}^2}$$

and, if to reach the results for $\mu_2{}'''$ given above we put this zero, we must have

$$r_{12}=\sqrt{\cdot 50}=0\cdot 7 \text{ nearly.}$$

IV. The Elimination of Spurious Correlation due to position in Time or Space.

By "STUDENT."

IN the *Journal of the Royal Statistical Society* for 1905[*], p. 696, appeared a paper by R. H. Hooker giving a method of determining the correlation of variations from the "instantaneous mean" by correlating corresponding differences between successive values. This method was invented to deal with the many statistics which give the successive annual values of vital or commercial variables; these values are generally subject to large secular variations, sometimes periodic, sometimes uniform, sometimes accelerated, which would lead to altogether misleading values were the correlation to be taken between the figures as they stand.

Since Mr Hooker published his paper, the method has been in constant use among those who have to deal statistically with economic or social problems, and helps to show whether, for example, there really *is* a close connection between the female cancer death rate and the quantity of imported apples consumed per head!

Prof. Pearson, however, has pointed out to me that the method is only valid when the connection between the variables and time is linear, and the following note is an effort to extend Mr Hooker's method so as to make it applicable in a rather more general way.

If x_1, x_2, x_3, etc., y_1, y_2, y_3, etc., be corresponding values of the variables x and y, then if x_1, x_2, x_3, etc., y_1, y_2, y_3, etc. are randomly distributed in time and space, it is easy to show that the correlation between the corresponding nth differences is the same as that between x and y.

Let ${}_nD_x$ be the nth difference.

For $\qquad {}_1D_x = x_1 - x_2, \quad \therefore {}_1D_x^2 = x_1^2 - 2x_1 x_2 + x_2^2.$

Summing for all values and dividing by N and remembering that since x_1 and x_2 are mutually random $S(x_1, x_2) = 0$, we get[†]

$$\sigma^2_{{}_1D_x} = 2\sigma_x^2.$$

Again, $\qquad {}_1D_y = y_1 - y_2, \quad \therefore {}_1D_x\,{}_1D_y = x_1 y_1 - x_2 y_1 - x_1 y_2 + x_2 y_2.$

Summing for all values and dividing by N, and remembering that x_1 and y_2 and x_2 and y_1 are mutually random

$$r_{{}_1D_x\,{}_1D_y}\,\sigma_{{}_1D_x} \cdot \sigma_{{}_1D_y} = 2r_{xy}\,\sigma_x\,\sigma_y,$$

$$\therefore r_{{}_1D_x\,{}_1D_y} = r_{xy}.$$

Proceeding successively $\qquad r_{{}_nD_x\,{}_nD_y} = r_{{}_{n-1}D_x\,{}_{n-1}D_y} = \dots = r_{xy} \quad \dots\dots\dots\dots\dots\dots(1).$

Now suppose x_1, x_2, x_3, etc. are not random in space or time; the problems arising from correlation due to successive positions in space are exactly similar to those due to successive occurrence in time, but as they are to some extent complicated by the second dimension, it is perhaps simpler to consider correlation due to time.

Suppose then $\qquad x_1 = X_1 + bt_1 + ct_1^2 + dt_1^3 + \text{etc.}, \quad x_2 = X_2 + bt_2 + ct_2^2 + dt_2^3 + \text{etc.}$

where X_1, X_2, etc. are independent of time and t_1, t_2, t_3 are successive values of time, so that $t_n - t_{n-1} = T$, and suppose $y_1 = Y_1 + b't_1 + c't_1^2 +$ etc. as before.

---

[*] The method had been used by Miss Cave in *Proc. Roy. Soc.* Vol. LXXIV. pp. 407 *et seq.* that is in 1904, but being used incidentally in the course of a paper it attracted less attention than Hooker's paper which was devoted to describing the method. The papers were no doubt quite independent.

[†] The assumption made is that n is sufficiently large to justify the relations

$$S_1^{n-1}(x)/(n-1) = S_2^n(x)/(n-1) = S_1^n(x)/n \text{ and } S_1^{n-1}(x^2)/(n-1) = S_2^n(x^2)/(n-1) = S_1^n(x^2)/n,$$

being taken to hold.

Then
$$ {}_1D_x = {}_1D_X - bT - cT(t_1+t_2) - dT(t_1^2+t_1t_2+t_2^2) - \text{etc.} $$
$$ {}_1D_x = {}_1D_X - \{bT + cT^2 + dT^3 + \text{etc.}\} - t_1\{2cT + 3dT^2 + 4eT^3 + \text{etc.}\} $$
$$ - t_1^2\{3dT + 6eT^2 + \text{etc.}\} - \text{etc.} $$

In this series the coefficients of t_1, t_2, etc. are all constants and the highest power of t_1 is one lower than before, so that by repeating the process again and again we can eliminate t from the variable on the right-hand side, provided of course that the series ends at some power of t.

When this has been done, we get
$$ {}_nD_x = {}_nD_X + \text{a constant}, $$
$$ {}_nD_y = {}_nD_Y + \text{a constant}, $$
so
$$ r_{{}_nD_x\,{}_nD_y} = r_{{}_nD_X\,{}_nD_Y} = r_{XY}, $$
and of course $r_{{}_{n+1}D_x\,{}_{n+1}D_y} = r_{{}_nD_x\,{}_nD_y}$, for ${}_nD_x$ and ${}_nD_y$ are now random variables independent of time.

Hence if we wish to eliminate variability due to position in time or space and to determine whether there is any correlation between the residual variations, all that has to be done is to correlate the 1st, 2nd, 3rd...nth differences between successive values of our variable with the 1st, 2nd, 3rd...nth differences between successive values of the other variable. When the correlation between the two nth differences is equal to that between the two $(n+1)$th differences, this value gives the correlation required.

This process is tedious in the extreme, but that it may sometimes be necessary is illustrated by the following examples: the figures from which the first two are taken were very kindly supplied to me by Mr E. G. Peake, who had been using them in preparing his paper "The Application of the Statistical Method to the Bankers' Problem" in *The Bankers' Magazine* (July—August, 1912). The material for the next is taken from a paper in *The Journal of Agricultural Science* by Hall and Mercer, on the error of field trials, and are the yields of wheat and straw on 500 $\frac{1}{500}$ acre plots into which an acre of wheat was divided at harvest. The remainder are from the three Registrar-Generals' returns.

	I	II	III	IV	V	VI
Correlation between ...	Sauerbeck's Index numbers.	Marriage Rate	Yield of Grain	Tuberculosis Death Rate.		
and	Bankers' Clearing House returns per head	Wages	Yield of Straw	Infantile Mortality		
				Ireland	England	Scotland
Raw figures ...	− ·33	− ·52	+ ·753	+ ·63	+ ·35	+ ·02
First difference ...	+ ·51	+ ·67	+ ·590	+ ·75	+ ·69	+ ·51
Second difference ...	+ ·30	+ ·58	+ ·539	+ ·74	+ ·74	+ ·65
Third difference ...	+ ·07	+ ·52	+ ·530	—	—	—
Fourth difference ...	+ ·11	+ ·55	+ ·524	—	—	—
Fifth difference ...	+ ·05	+ ·58	—	—	—	—
Sixth difference ...	—	+ ·55	—	—	—	—
Number of cases	41 years	57 years	500 plots	42 years		

The difference between I and II is very marked, and would seem to indicate that the causal connection between index numbers and Bankers' clearing house rates is not altogether of the same kind as that between marriage rate and wages, though all four variables are commonly taken as indications of the short period trade wave. I had hoped to investigate this subject more thoroughly before publishing this note, but lack of time has made this impossible.

V. On certain Errors with regard to Multiple Correlation occasionally made by those who have not adequately studied this Subject.

By KARL PEARSON, F.R.S.

(1) It is well-known* that if we endeavour to predict the value of a variate x_0 from n correlated variates $x_1, x_2, \dots x_n$, by determining a *linear* function of $x_1, x_2, \dots x_n$ which has the maximum correlation R_n with x_0, then the value of R_n^2 is given by

$$R_n^2 = 1 - \Delta/\Delta_{00},$$

where Δ is the determinant

$$\Delta = \begin{vmatrix} 1, & r_{01}, & r_{02}, & \dots & r_{0n} \\ r_{10}, & 1, & r_{12}, & \dots & r_{1n} \\ r_{20}, & r_{21}, & 1, & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{n0}, & r_{n1}, & r_{n2}, & \dots & 1 \end{vmatrix}$$

and Δ_{pq} is the minor corresponding to the constituent of the pth column and qth row.

The system I propose to consider is that in which all correlations like r_{0p} are equal, whatever p be, to a constant ρ, and all correlations r_{pq}, where p and q may take any values from 1 to n, are the same and equal to ϵ. We now have for the value of Δ the expression

$$\begin{vmatrix} 1, & \rho, & \rho, & \dots & \rho \\ \rho, & 1, & \epsilon, & \dots & \epsilon \\ \rho, & \epsilon, & 1, & \dots & \epsilon \\ \rho, & \epsilon, & \epsilon, & \dots & 1 \end{vmatrix}.$$

To evaluate this determinant add all the rows but the first together, giving

$$n\rho, \quad 1+(n-1)\,\epsilon, \quad 1+(n-1)\,\epsilon, \quad \dots \quad 1+(n-1)\,\epsilon,$$

multiply the result by $\rho/(1+(n-1)\,\epsilon)$ and subtract from the first row. We have

$$\begin{vmatrix} 1-\dfrac{n\rho^2}{1+(n-1)\,\epsilon}, & 0, & 0, & \dots & 0 \\ \rho, & 1, & \epsilon, & \dots & \epsilon \\ \rho, & \epsilon, & 1, & \dots & \epsilon \\ \dots & \dots & \dots & \dots & \dots \\ \rho, & \epsilon, & \epsilon, & \dots & 1 \end{vmatrix} = \left\{1-\frac{n\rho^2}{1+(n-1)\,\epsilon}\right\} \times \Delta_{00}.$$

Hence

$$R_n^2 = 1 - \left(1 - \frac{n\rho^2}{1+(n-1)\,\epsilon}\right) = \frac{n\rho^2}{1+(n-1)\,\epsilon} \quad \dots\dots\dots\dots (\text{i}),$$

or

$$R_n\dagger = \rho\sqrt{\frac{n}{1+(n-1)\,\epsilon}} \quad \dots\dots\dots\dots (\text{ii}).$$

Hence proceeding to the limit we have

$$R_\infty = \rho/\sqrt{\epsilon} \quad \dots\dots\dots\dots (\text{iii}).$$

---

* *Biometrika*, Vol. VIII. p. 439.

† The sign of R_n must be determined from other considerations.

Thus if n variates are equally correlated (ϵ) among themselves, and equally correlated (ρ) with another variable, we shall not indefinitely increase the accuracy with which the last variable will be predicted from the others by increasing indefinitely the number of the variates n.

Illustration. The coefficient of multiple correlation is required as we increase the number of brothers from whom a prediction of a character in a given brother is made. The fraternal correlation $= \cdot 5$.

Number of Brothers	R_n
1	·5000
2	·5774
3	·6124
4	·6325
5	·6455
6	·6547
10	·6742
∞	·7071

Compare against these results *two* parents only in a population where there is no assortative mating and the parental correlation $= \cdot 5$. Here $\epsilon = 0$, $\rho = \cdot 5$ and $n = 2$, $\therefore R = \frac{1}{2}\sqrt{2} = \cdot 7071$, or two parents will give more information than 10 brothers and sisters, and as much in fact as an indefinite number. Suppose the parents tend to select their like, i.e. suppose there is assortative mating in the population, say, $\epsilon = \cdot 15$, then with the same intensity of parental correlation

$$R = \cdot 6594,$$

or, two parents will give us more information than six brothers and sisters.

Now this illustration brings out the real nature of the effect of increasing the number of variables from which we predict. Such increase has very little value, if those variables are fairly highly correlated with each other. To be effective they must be highly correlated with the variate we wish to predict and correlated very slightly with each other.

Even in this case there is a limit to the degree of correlation reached when the number of variates is indefinitely increased, namely $\rho/\sqrt{\epsilon}$, and it is clear that if ρ be small and ϵ fairly large, no very great increase of correlation is obtained if we use an indefinitely great number of variates. For example if $\rho = \cdot 05$ and $\epsilon = \cdot 5$, we find $R_\infty = \cdot 0707$ only. Even if ρ were ·10, we should only raise R to ·1414, could we predict from an indefinitely large number of such correlated variates*. Indeed as long as ϵ is not less than ρ we gain singularly little by combining large numbers of variates. For example if ρ were ·4, and $\epsilon = \cdot 4$ ten such variates would only raise the correlation to ·5898, and an indefinitely large number to ·6325, which is less than double the single correlation. Yet there are apparently many persons who believe that by taking a number of low correlations, a high relationship can be reached!

Actually there is a limit to what relations can possibly exist between a variate x_0 and a series of equally correlated variables $x_1 \ldots x_n$. Since R must be less than unity, we have

$$\rho \sqrt{\frac{n}{1+(n-1)\epsilon}} < 1,$$

or

$$\epsilon > \frac{n\rho^2 - 1}{n-1}.$$

Thus if $n = 10$ and $\rho = \cdot 5$, ϵ must be $> \cdot 1667$. Or, it would be impossible for 10 variates to have a correlation ·5 with another variable, and a zero correlation with each other.

---

[*] Even if ρ were ·10 and ϵ as low as ·10 we should not raise R for endless variates of this order of correlation above ·3163, while from compounding ten such variates we should only obtain a correlation about double that of a single variate, i.e. $R = \cdot 2294$.

If we suppose a number of variates n to be uncorrelated with each other, but correlated $r_{01}, r_{02}, \dots r_{0n}$ with another variable x_0, then we have from the determinant as given below

$$\Delta = \left| \begin{array}{llll} 1\,, & r_{01}, & r_{02}, & \dots\ r_{0n} \\ r_{10}, & 1\,, & 0\,, & \dots\ 0 \\ r_{20}, & 0\,, & 1\,, & \dots\ 0 \\ \dots & \dots & \dots & \dots \\ r_{n0}, & 0\,, & 0\,, & \dots\ 1 \end{array} \right| = (1 - r_{10}{}^2 - r_{20}{}^2 - \dots r_{0n}{}^2)\Delta_{00}.$$

$$\therefore\ R^2 = r_{10}{}^2 + r_{20}{}^2 + \dots + r_{0n}{}^2,$$

or

$$R = \sqrt{n}\sqrt{\frac{r_{10}{}^2 + r_{20}{}^2 + \dots + r_{0n}{}^2}{n}}.$$

Therefore, if n variables, uncorrelated among themselves, be correlated with an additional variable, it is necessary that the root mean square of their correlations should be less than $\frac{1}{\sqrt{n}}$. We see therefore that it must either be impossible to find a large number of variables uncorrelated among themselves, which are correlated with an additional variable, or else their correlations with this variable must be extremely low. The last result shows us the fallacy of supposing that correlations are simply added together for a combined effect; clearly when the variates are uncorrelated among themselves, we add by the *sum of the squares.* For example, if $r_{01} = r_{02} = \dots = r_{0n} = {\cdot}03$ *one hundred* such variables would only raise R to ·30. On the other hand if the variates are highly correlated together, say $e = {\cdot}81$, an indefinitely great number of such variables would only raise the multiple correlation to ·0333, if the individual correlation were ·0300.

We are now in a position to apply our results to the problem of the relative intensity of heredity and environment. This problem has been singularly misunderstood especially by the popular exponents of Eugenics. Some illustrations of this may be given here. Major Leonard Darwin writes as follows in the Journal of the Eugenics Education Society: "It is impossible to compare heredity as a whole with environment as a whole as far as their effects are concerned; for no living being can exist for a moment without either of them[*]. Moreover, in order to compare two things so as to be able to use the words more or less in connection with such a comparison, we must have a *common unit of measurement* applicable to them both. But what is the unit by which both heredity and environment may be measured? I myself have no idea. May we not be discussing questions as illogical as enquiring what portion of the area of a rectangle is due to its width and what to its length? *Is it ever wise to use words in scientific literature without endeavouring to attach a definite meaning to them*[†] *?*

It is hard to conceive a paragraph of the same length more full of evidence of complete ignorance of the methods used in modern science for comparing correlated variates! Yet it goes out as the opinion of the President of a Society which is endeavouring to spread the scientific doctrines of Eugenics among the people! Major Darwin begins by stating that it is needful to have a common unit of measurement in order to compare two variates. To begin with we are not comparing *two* things, but we are comparing the influence of two things on

[*] There would in our sense be no heredity if the average child born to noteworthy parents was equal to the average child of the whole community. Yet it is perfectly easy to understand how living beings could exist under such a law of reproduction. Major Darwin seems to be confusing two things, the fact that a man is born true to his species, and the fact that he resembles his immediate ancestry. It is the latter fact only which concerns us when we compare heredity and environment, i.e. how variation of immediate ancestry affects the individual's physical or mental characters. But without such heredity individuals might quite well exist.

[†] *The Eugenics Review*, Vol. v. p. 152. The italics are mine.

a third, i.e. the intensity of a certain environmental influence and the intensity of a certain somatic character in the parent, say, on the intensity of the somatic character in the offspring. Yet Major Darwin tells us we cannot do this because we cannot measure these things in the *same* unit!—How suavely yet forcibly Sir Francis Galton himself would have ridiculed such ignorance in high places as is passed by the Editor of the Eugenics Journal!—We can hear him now telling us how the intensity of each character could be measured by its grade, and how the problem turned on whether the same change in grade in the environment and in the parental somatic character produced greater or less change in the grade of the filial somatic character. When we inquire whether inter-racially stature is more closely related to cephalic index or to eye colour, are we to be met by the statement that these characters cannot be compared because they cannot be measured in a 'common unit,' and then be told that it is not "wise to use words in scientific literature without endeavouring to attach a definite meaning to them?" Every trained statistician knows that each character is measured in the unit of its own variability—in what he terms its standard deviation*, and that this standard deviation provides him with a measure of the frequency of each value of the variate in question. It seems to me that the only correct sentence in this paragraph, is the author's statement that he himself has no idea what unit is 'common' to heredity and environment.

But our author continues:

"Take any quality, and we find that the human beings composing any community differ more or less considerably as regards that quality. Now we can measure the correlation between the differences shown in this quality and the differences of environment to which the members of the community in question had previously been exposed†. This is one correlation. Then we can also measure the correlation coefficient between, say, father and son, as regards the quality in question. Here is a second correlation; and if we are told that the relative influence of environment and heredity is measured by the ratio between these two correlation coefficients, we certainly do thus get a clear conception of what is meant‡."

But has the writer really obtained a clear conception of what such coefficients of correlation mean, when in the next paragraph he continues:

"Imagine an ideal republic, in some respects similar to that designed by Plato, where not only were all the children removed from their parents, but where they were all treated exactly alike. In these circumstances none of the differences between the adults could have anything to do with the differences of environments, and all must be due to some differences in inherent factors. In fact the environment correlation coefficient would be nil, whilst the hereditary correlation coefficient might be high§."

Could any better evidence be adduced that the President of the Eugenics Education Society did not know what a coefficient of correlation meant at that date? The coefficient of correlation for the environment might be anything from -1 to $+1$; the only obvious fact would be that you could not find its value, except in the form $0/0$, from an environment which precluded any measure of variation. How again Sir Francis would have smiled at the notion that the coefficient of correlation for a constant environment must be nil. Why should we follow such

---

* Of course he may or does need other constants to help in the description of the frequency.

† *loc. cit.* p. 153.

‡ This seems to contradict the writer's previous assertion that two things are incomparable, if they have not a '*common unit*'!

§ I wrote at once to Major Darwin pointing out the error of such a statement and he withdrew it in the next number. But the harm done by an article of this kind cannot be reversed by correcting a single misstatement.

advice as that given by the President of the Society to avoid as far as possible "such phrases as the relative influence of heredity and environment," when on his own showing he does not in the least appreciate the methods by which this relative influence is measured?

Then Major Darwin continues: "Surely what we want to know is how we can do most good—whether by attending to reforms intended to affect human surroundings, or to reforms intended to influence mankind through the agency of heredity. But does this ratio [that of the environmental and hereditary correlation] give us any sure indication of the relative amount of attention which should be paid to these two methods of procedure?" Our only reply can be that these correlations certainly do, and that as long as the President of the Eugenics Education Society fails to grasp their meaning, he is doing grave harm to the science of eugenics.

We measure the change in the character of an individual which would be produced by a change of a like or an allied character in a parent, such change being one of which we have experience; we measure the change which would be produced in the character of the individual by changes in the environment such as we have experience of, i.e. when we move the individual from a badly ventilated to a well ventilated house, from a back to back to a through house, from a low wage to a high wage, and so forth, and we find the resulting changes are of a wholly different order in these cases to what happens when we change the physical characters, the health or habits which define the parents. It is on the basis of this that we assert that the relative strength of heredity is far greater than the strength of environment. To this reasoning, apart from such arguments as the above or those to be immediately dealt with, reply is only made by talk as to the impossibility of an individual surviving if you deprived him of his normal environment! It would be just as reasonable to assert that everything must be due to heredity, because a race of supermen would breed supermen! What the scientific eugenist has endeavoured to measure are the influences of such range of differences in environment as occur in everyday experience and are therefore producible from the political, economic and social standpoints, not the absence of all environment at all. But while this is recognised by some of the popular eugenic writers, they have approached the problem from another standpoint which indicates equally how little they grasp modern statistical theory. We admit, they say, that the environmental correlations may be of the order ·03 or ·05 and the inheritance correlations of the order ·50. But this is the correlation of *one* character in environment. You ought to take ten or twenty, and then you will have multiplied up environment to be more effective than heredity, for $\cdot 03 \times 20 = \cdot 60$. In the first place we may suggest that it would be just as reasonable, if the argument were a valid one to multiply up the favourable hereditary characters, to take weight, height, muscular activity, health, intelligence, caution, and many other desirable factors, and these not only in one parent but in brothers, sisters, aunts, uncles and grandparents and treat the cross-correlation of these with the character under discussion. But although every improvement in stock would reflect itself in improvement in offspring, correlations cannot be *added* together—any more than forces by simple arithmetical addition. You do not combine two hereditary correlations any more than two environmental correlations by mere addition. You must proceed by the combinatory process indicated at the commencement of this paper, which is one of course familiar to every trained statistician.

Yet here is a statement which the Editor of the *Eugenics Review* admits to its pages without contradiction*:

> The point that we wish to make is this. In the face of so much ignorance concerning, not only heredity itself, but also its complement, the influence of environment, how can any one be justified in making sweeping generalisations with reference to these subjects?
>
> Such generalisations, however, are made. It is said that we have a definite proof that inheritance is of far greater strength than environment. This argument takes the following shape. The correlations between parent and offspring for a number of features have been calculated, and the mean is found to

---

* Vol. v. p. 219, in an article by A. M. Carr-Saunders.

be somewhere about ·5. Correlations between individuals and various aspects of their environment have also been worked out—as, for instance, mental ability and conditions of clothing, or between myopia and the age of learning to read[*]—and the mean value is found to be about ·03. It is then said that the mean "nature value" is at least five to ten times as great as the mean "nurture value," and upon this is founded the generalisation that "nature" is of far greater importance than "nurture"[†]. It may be questioned, however, whether such a comparison does not involve a serious mistake. For if we consider the two mean values that are compared, we find that, whereas the "mean nature value" is the mean value of a number of observations, *all of which provide a full measure of the strength of heredity*, the "mean nurture value" is the mean value of a number of observations, each of which measures only the strength of *some one isolated aspect of environment*. It would appear then that the full strength of inheritance has been compared, not with the full strength of environment, but with the average of a number of small isolated aspects of the latter. As a matter of fact it is quite beyond our power at present to sum up the full effect of environment upon the individual and compare it with the full effect of heredity. We are, therefore, justified in saying that we neither know in particular cases how far the environment can produce any effect, nor can we make any definite statement as to the comparative strength of "nature" and "nurture."

Now this is the doctrine passed by the Editors of the *Eugenics Review*, the journal of a society, which has assumed the mantle of Francis Galton[‡], and it is passed, because the editorial committee of that society does not grasp the meaning of multiple correlation! The passages in italics have been so printed to draw our readers' attention to them. In the first place, of course, a single correlation coefficient does *not* provide a full measure of the strength of heredity. In the table cited the coefficients are those for *one* parent or for *one* brother or sister. Each relative—and those for independent stocks are either non-correlated or inter-correlated very slightly—provides such a coefficient, and further each character in such relatives may be correlated with the character under discussion in the subject in question. In the next place the environment factors do not consist of "some one *isolated* aspect of environment." All these factors or aspects are closely interlinked, and this was a fact well-known to the workers in the Galton Laboratory. The real interpretation of such a difference as ·50 and ·03 in the average values of *single* coefficients can only be appreciated by those who are conversant with the theory of multiple correlation, and it is quite clear that those who profess to guide the public in this very difficult problem—which is essentially a scientific problem—lack any adequate knowledge of the sole instrument by which any conclusion can be drawn.

The writer appears to be wholly ignorant of the nature of multiple correlation in the first place, and in the second entirely to overlook the very high correlations which exist between environmental factors. Bad wages, bad habits, bad housing, uncleanliness, insanitary surroundings, crowded rooms, danger of infection, etc., etc. are all closely associated together, and while the order of correlation between environmental and physical characters is low, that between individual environmental factors is in our experience very high. Thus the problem of multiple correlation illustrates closely the theory developed in the first part of this note; we have to deal with a low ρ and a high ϵ.

For example, if we take the environmental factors to have an average inter-correlation of ·70, then an infinity of such factors for a mean environmental and individual correlation of ·03 would

---

[*] As the writer phrases this correlation, it is very liable to be misinterpreted. What the Galton Laboratory did was to show that myopia was very markedly inherited, and that the theory that it was largely due to school environment was incorrect, because children who began to read late, i.e. went *late* to school, were not less myopic than those who went early.

[†] Karl Pearson, *Nature and Nurture*, Eugenics Laboratory, Lectures VI. p. 25.

[‡] If there was one point on which Francis Galton felt strongly and wrote it was on this point of the relatively great intensity of "nature" as compared with "nurture." I do not stand alone in recognising it as an essential part of his teaching: "I am inclined to agree with Francis Galton," writes Charles Darwin, "in believing that education and environment produce only a small effect on the mind of anyone, and that most of our qualities are innate."

only raise the correlation to ·0359 against a *single* parental correlation of ·5000; if the correlation was ·05 instead of ·03, we should have the total possible environmental multiple correlation ·0598 as against ·5000. Even if we raise the average environmental correlation to ·1 and the inter-environmental factor correlation be reduced to ·5, the multiple correlation of an infinity of factors is only ·1414 as against the *single* factor of heredity ·5000. Even if we could pick out *one hundred* environmental factors which had no inter-correlations—which experience shows is wholly impossible—and each of these independent factors was correlated to the extent of ·05 with the mental or physical characters of an individual they would only just reach the hereditary influence of a *single* character in a *single* parent.

Now let us suppose an absolutely idle case, namely that the environmental factors had the same correlation as a parent, i.e. ·5, with the character of the individual, and only a correlation of ·6 with each other, then if we could use an indefinitely great number of such factors the multiple correlation would only be $\cdot 5/\sqrt{\cdot 6} = \cdot 6455$, while the correlation with two parents, with no assortative mating, would be ·7071. Even with assortative mating, it suffices to take only the four grandparents into account to show that heredity acts in excess of an environmental scheme even so preposterous as is suggested above. If we take the parental correlations ·50, the grandparental ·25, and those of assortative mating ·15, we have for the determinant:

$$\Delta = \begin{vmatrix} 1, & \cdot 50, & \cdot 50, & \cdot 25, & \cdot 25, & \cdot 25, & \cdot 25 \\ \cdot 50, & 1, & \cdot 15, & \cdot 50, & \cdot 50, & 0, & 0 \\ \cdot 50, & \cdot 15, & 1, & 0, & 0, & \cdot 50, & \cdot 50 \\ \cdot 25, & \cdot 50, & 0, & 1, & \cdot 15, & 0, & 0 \\ \cdot 25, & \cdot 50, & 0, & \cdot 15, & 1, & 0, & 0 \\ \cdot 25, & 0, & \cdot 50, & 0, & 0, & 1, & \cdot 15 \\ \cdot 25, & 0, & \cdot 50, & 0, & 0, & \cdot 15, & 1 \end{vmatrix}.$$

Add together the second and third rows multiplied by ·3951, and the fourth, fifth, sixth and seventh multiplied by ·0456 and subtract the result from the first. The first row then becomes

$$\begin{vmatrix} \cdot 5593, & 0, & 0, & 0, & 0, & 0, & 0 \end{vmatrix}$$

the others of course remaining the same.

Hence $$\Delta = \cdot 5593 \times \Delta_{00},$$

and $$R^2 = 1 - \Delta/\Delta_{00} = 1 - \cdot 5590 = \cdot 4407.$$

Therefore $$R = \cdot 6639.$$

Or together grandparents and parents would influence a man's character more than an *infinity* of environmental factors of the same grade of correlation, because the latter factors are far more highly correlated together than several of our relatives.

Actually of course we are dealing with average values; the average value of environmental correlation with individual character being in our experience of the order ·03 to ·05 and the inter-environmental factor correlations of the order ·5 to ·7. But these averages enable us to appreciate the total effect.

The doctrine taught by the writers in the *Eugenics Review*, that we know nothing of the relative intensity of environment and heredity and that it is unwise "to use words in scientific literature without endeavouring to attach a definite meaning to them" only demonstrate how far the Editors of that Journal are removed from any appreciation themselves of modern statistical methods. How far the doctrine is removed from the very strong views held on this point by Francis Galton, only those who have studied his writings and know how strongly he felt personally on the subject are in the least competent to appreciate.

VI. Formulae for the Determination of the Capacity of the Negro Skull from External Measurements.

By L. ISSERLIS, B.A.

§ 1. Formulae for the determination of the capacity of the human skull from external measurements, were obtained by Lee and Pearson*. The material they employed consisted of various series of measurements of Bavarian, Aino and Naqada skulls. Measurements of Ancient and modern Egyptian and other non-European skulls were employed, chiefly for purposes of comparison. The formulae, some of which will be quoted later, were intended primarily for the prediction of the capacity of European skulls, from external measurements. Doubt has been thrown on several occasions on the applicability of these formulae to the Negro skull, one of the reasons alleged being the supposed difference in thickness of the bone of European and Negro crania.

The publication† of the late Dr R. Crewdson Benington's researches on the negro skull has made it possible to obtain similar formulae for negro skulls, and to test how far these can be applied to the prediction of the capacity of European skulls and conversely to test the applicability of Lee and Pearson's Equations to the negro skull.

§ 2. The material is fully described in Dr Benington's Study. The crania dealt with in the present paper are Benington's series A, B, C.

A. *Congo Crania* in the Royal College of Surgeons. These crania provide 46 males and and 21 females, as owing to various defects no capacity is available for numbers 25, 38, 48, 54 among the males and numbers 69, 72, 75, 79, 82, 85 among the females.

B. *Crania from the Gaboon, Group I*, brought by Du Chaillu from Fernand Vaz in 1864. Of the 50 male and 44 female crania in the series, 2 males (numbers 3 and ?) and 1 female (number 2) are defective, leaving 48 male and 43 female crania available.

C. *Crania from the Gaboon, Group II*, brought by Du Chaillu from Fernand Vaz in 1880. Two of the 18 males (numbers $12a$ and 20) and two of the 19 females (numbers 8 and 18) are defective.

Altogether 110 male and 81 female crania have been dealt with. The correlation has been calculated of the capacity (C) and the product of the breadth, length and total height (B, L and H), for each group and for the aggregates of 110 male, and of 81 female crania.

Correlation coefficients have also been calculated for the capacity and breadth, capacity and length, and capacity and total height, but for the aggregates of the three groups only. Regression formulae are given in all cases. It is to be observed that Dr Crewdson Benington's measurements of capacity were taken with mustard seed, packing and measuring glass and that the error of measurement or rather his average difference as compared with other workers in the Biometric Laboratory was under 10 cm^3.

In comparing the regression formulae obtained here, with those given by Lee and Pearson for European and other skulls it must be remembered that in all their formulae except (12) and (13) of p. 247 they employed the auricular height and not the total height. In the present paper as in Dr Benington's study H denotes the total height. Lee and Pearson denote this by H' and use H for the auricular height.

It was not possible here to use the auricular height as it was not available for the whole of the Gaboon series B and C.

---

* *Phil. Trans.* Vol. 196, Series A, pp. 225—264.

† *Biometrika*, Vol. VIII. Nos. 3 and 4, Dec. 1911.

Taking first the male skulls, the mean value of the capacity and the product *BLH*, their standard deviations and the correlations are given in the following table.

TABLE I.

	Mean capacity in cm.^{3}	Mean value of BLH in cm.^{3}	σ_C in cm.^{3}	σ_{BLH} in cm.^{3}	$r_{C,\,BLH}$
46 Congo Skulls ...	1344	3303	126·22	282·99	·872
48 Gaboon (1864) ...	1379	3295	108·30	230·30	·822
16 Gaboon (1880) ...	1447	3463	109·60	266·42	·808
110 Negro skulls ...	1375	3323	120·74	265·20	·842

The corresponding regression lines are

for the 46 Congo ...	$C = {\cdot}0003889BLH + 59 \pm \frac{41}{\sqrt{n}}$ ....(1),
48 Gaboon (1864)	$C = {\cdot}0003865BLH + 105{\cdot}5 \pm \frac{62}{\sqrt{n}}$ ....(2),
16 Gaboon (1880)	$C = {\cdot}0003323BLH + 297 \pm \frac{65}{\sqrt{n}}$ ....(3),
110 male negro skulls	$C = {\cdot}0003849BLH + 96 \pm \frac{65}{\sqrt{n}}$ ....(4).

Lee and Pearson's corresponding equation for males is

$$C = {\cdot}000266LBH' + 524{\cdot}6^{*} \text{ ....} (P).$$

This is not a regression line, but is obtained by method of least squares from the results for various races in their table 20.

The formulae 1—4 can be used to predict the capacity of an individual skull from external measurements. The probable errors of the mean were calculated by the formula $0{\cdot}67449\sigma_C \frac{\sqrt{1-r^2}}{\sqrt{n}}$ where n is the number of skulls in the group to which the formula is applied. If we substitute in (1)—(4) the mean values of B, L, H for the Bavarian male skulls used by Lee and Pearson, viz.:

$$B = 150{\cdot}5,$$
$$L = 180{\cdot}6,$$
$$H = 133{\cdot}8,$$

we obtain, from (1), $C = 1474 \pm \frac{41}{\sqrt{n}}$

" (2), $C = 1471 \pm \frac{62}{\sqrt{n}}$

" (3), $C = 1506 \pm \frac{65}{\sqrt{n}}$

" (4), $C = 1496 \pm \frac{65}{\sqrt{n}}$.

---

* *Loc. cit.* Equation (12). H' = total height.

The measured capacities of these German skulls have a mean value of 1503 c.c. a result which is in very close agreement with (4) the formula based on 110 skulls. 1503 is the mean capacity of 100 skulls so that $\frac{65}{\sqrt{n}}=6{\cdot}5$. Thus the difference between the actual mean capacity of German skulls and the mean capacity estimated by the negro formula is less than 10 cm^3. although the mean capacity of German male skulls exceeds that of negro males by

$$1503-1375=128 \text{ cm}^3.$$

If the above values of B, L, H are substituted in Lee and Pearson's formula (P) on p. 4 we obtain $C=1492$.

On the other hand if we substitute the mean values of the dimensions of the 110 male negro skulls, $B=137$, $L=178$, $H=135$ in formula P we obtain $C=1400$ as compared with the measured mean of 1375.

This is not as good a reconstruction as our formula (4) or as the formulae of Lee and Pearson employing auricular height, and is probably due to the fact that P is obtained by the method of least squares from 11 means only.

§ 4. An approximation to the influence of the thickness of the bone of the skull on predictions of capacity from external measurements can be obtained by differentiating the equation

$$C=kBLH+\text{const.}$$

and putting

$$dB=dL=dH=t.$$

We obtain

$$dC=k(BL+LH+HB)\,t,$$

or if we observe that in the equations the constant is comparatively small

$$\frac{dC}{C}=\left(\frac{1}{B}+\frac{1}{L}+\frac{1}{H}\right)t$$

with $B=150{\cdot}5$, $L=180{\cdot}6$, $H=133{\cdot}8$

$$\frac{dC}{1500}=t\,({\cdot}02) \text{ approximately.}$$

Thus a difference of 10 cm^3. in capacity corresponds to a difference of $\frac{1}{3}$mm. in thickness which is about 5 $^\circ/_\circ$ of the thickness (say 6 mm.) of the human skull.

We may fairly conclude then, that there is no appreciable difference in the thickness of the negro skull as compared with the European.

§ 4. The female crania yield very similar results. The following is the table for the female skulls.

TABLE II.

	Mean capacity in cm.^3	Mean BLH in cm.^3	σ_C in cm.^3	σ_{BLH} in cm.^3	$r_{C,BLH}$
21 Congo Skulls ...	1206	2858	107·7	268·2	·9077
43 Gaboon (1864) ...	1232	2924	126·7	270·95	·8814
17 Gaboon (1880) ...	1240	2964	97·31	265·8	·8560
81 Negro skulls ...	1227	2956	117	255·72	·7668

The corresponding regression lines are

21 Congo skulls	$C = \cdot 0003645\,BLH + 164 \pm \frac{45}{\sqrt{n}}$ ....(5),
43 Gaboon (1864)	$C = \cdot 0004122\,BLH + 27 \pm \frac{60}{\sqrt{n}}$ ....(6),
17 Gaboon (1880)	$C = \cdot 0003134\,BLH + 311 \pm \frac{50}{\sqrt{n}}$ ....(7),
81 Negro skulls	$C = \cdot 0003508\,BLH + 204 \pm \frac{75}{\sqrt{n}}$ ....(8).

The corresponding Lee and Pearson formula obtained by the method of least squares is

$$C = \cdot 000156\,LBH' + 812 \quad \ldots\ldots\ldots\ldots (Q).$$

The mean values of B, L, H' for the Bavarian female skulls discussed by Lee and Pearson are

$$B = 144{\cdot}11,$$
$$L = 173{\cdot}59,$$
$$H' = 128{\cdot}07.$$

With these values, we deduce from 5—8 the following values for C.

$$(5) \quad C = 1331 \pm \frac{45}{\sqrt{n}}$$
$$(6) \quad C = 1347 \pm \frac{60}{\sqrt{n}}$$
$$(7) \quad C = 1315 \pm \frac{50}{\sqrt{n}}$$
$$(8) \quad C = 1327 \pm \frac{75}{\sqrt{n}}$$

The mean of the measured values of the capacities of these skulls is 1337 and formula (8) based on 81 negro skulls gives a result in very close agreement.

If the above values of B, L, H' are substituted in Lee and Pearson's formula Q we obtain $C = 1284$ a result which differs from the true value much more seriously than the prediction by the negro regression formula.

Again, if we insert the mean values

$$B = 130{\cdot}75,$$
$$L = 171{\cdot}33,$$
$$H = 129{\cdot}81,$$

of the 81 female negro crania in the formula Q we get $C = 1266$ as against the mean of the measured values which is $C = 1227$, demonstrating again the fact that the formulae P, Q based on 11 means are not as good as the regression formulae.

§ 5. We add tables of the correlation between capacity and breadth, capacity and length, and capacity and total height for the 110 male and the 81 female skulls, and for comparison reprint the corresponding value for German (Bavarian) skulls.

TABLE III.

Correlation Males.

	Negro	German
Capacity and Breadth ...	·4977	·6720
Capacity and Height ...	·6080 (total height)	·2431 (auricular height)
Capacity and Length ...	·7433	·5152

TABLE IV.

Females.

	Negro	German
Capacity and Breadth ...	·7578	·7068
Capacity and Height ...	·5450 (total height)	·4512 (auricular height)
Capacity and Length ...	·6699	·6873

The corresponding regression lines are given in the tables below :

TABLE V.

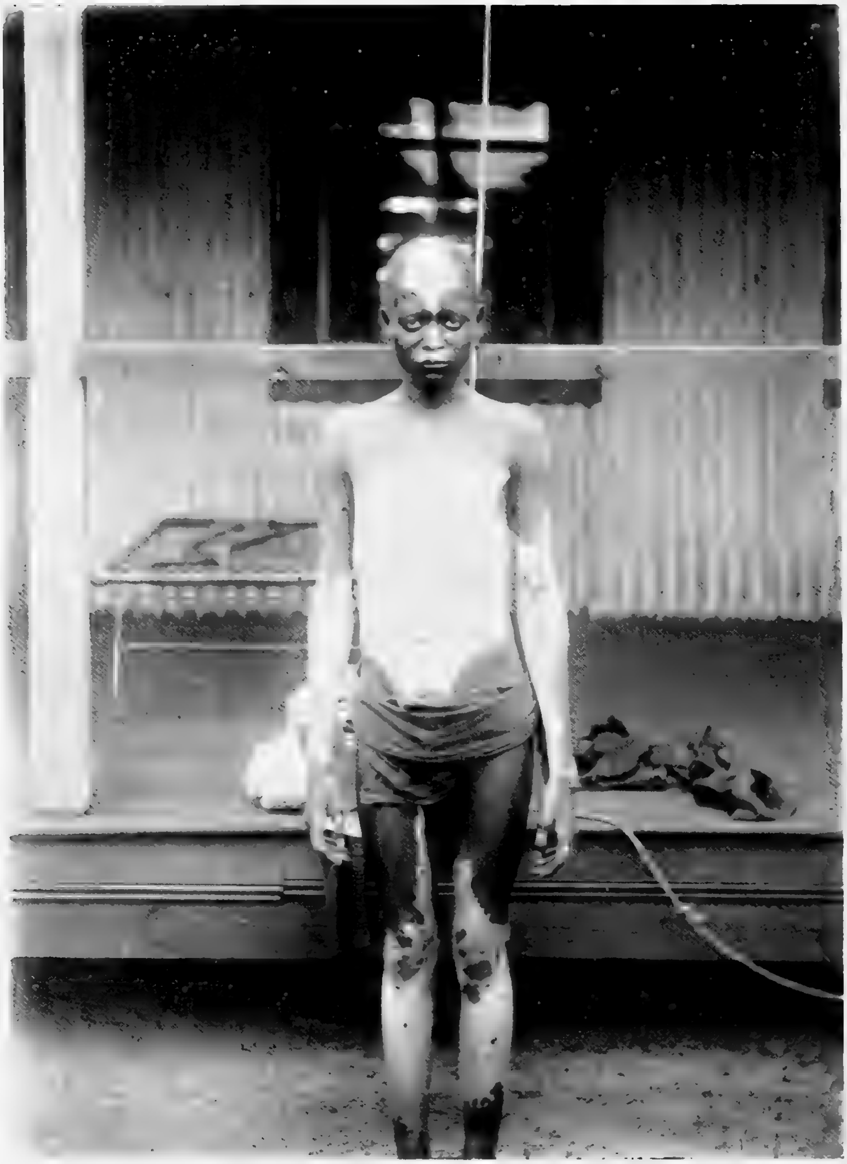
Males.

Negro	German
(9) $C=12{\cdot}6356B-356{\cdot}1\pm\frac{105}{\sqrt{n}}$	$C=13{\cdot}432B-517{\cdot}34$
(10) $C=12{\cdot}8301L-1087\pm\frac{81}{\sqrt{n}}$	$C=9{\cdot}892L-282{\cdot}55$
(11) $C=15{\cdot}3265H'-694\pm\frac{96}{\sqrt{n}}$ (H'=total height)	$C=5{\cdot}264H+868{\cdot}05$ (auricular height)

TABLE VI.

Females.

Negro	German
(12) $C=17{\cdot}872B-1114\pm\frac{76}{\sqrt{n}}$	$C=15{\cdot}716B-927{\cdot}66$
(13) $C=12{\cdot}46\,L-908\pm\frac{87}{\sqrt{n}}$	$C=12{\cdot}055L-755{\cdot}53$
(14) $C=10{\cdot}871H'-184\pm\frac{98}{\sqrt{n}}$ (H'=total height)	$C=10{\cdot}993H+82{\cdot}13$ (auricular height)



Dr Maynard's Piebald Negro

No great degree of accuracy can be expected in reconstructing the capacity of a skull from a single measurement, but the remarkable difference of formula (11) for negro skulls from the corresponding German formula is of course due to their referring to different measurements of the height. If we insert $H = 133{\cdot}8$ in (11), which is the mean total height of the Bavarian skulls we get $C = 1356{\cdot}7 \pm \frac{96}{\sqrt{n}}$ instead of the measured mean $C = 1503$ cm^3.

Similarly equation (9) gives $C = 1555{\cdot}6 \pm \frac{105}{\sqrt{n}}$ instead of 1503 when we insert the German mean $B = 150{\cdot}5$.

Thus 9—14 are of little use for our purpose.

VII. Note on a Negro Piebald. (C. D. MAYNARD.)

THE remarkably interesting photograph of a negro piebald on Plate X has been forwarded to the Editor by Dr C. D. Maynard. The native comes from the district round Chai Chai. Dr Maynard writes from Ressano Garcia, and states that the hospital attendant took the photograph. The extraordinary interest of the case arises from the fact that the thighs and feet are of normal negro pigmentation, but in the other patches we have varying degrees of pigmentation of the skin down to albinotic white. Unfortunately there is no dorsal view, but the back is stated to be also affected with albinotic areas. The boy reported that he was in the same condition when born, and that the nature and areas of the pigmentation had not altered.

VIII. Note on Infantile Mortality and Employment of Women, from the *Report on Condition of Woman and Child Wage-earners in the United States,* Volume XIII. Infant Mortality and its Relation to the Employment of Mothers.

BY ETHEL M. ELDERTON.

THE author of this *Report* emphasizes the difficulty of determining the effect of women's employment and points out that

> "It would be possible to draw positive conclusions as to the relative importance of this particular factor only by point-to-point comparison of the infant mortality for a period of years in two large communities, or two classes of large communities, in which all the material conditions were substantially common, with the single important exception that in one a considerable proportion of the married female population of child-bearing age were at work outside of their homes and in the other community with which the comparison was made none of the women were so employed.
>
> To admit of entirely sound conclusions, it would be necessary that the populations—and especially the women—of both communities should be of like ages, races, and physical health, that their living conditions should be practically identical, and that, in a general way, the child-bearing women should be of about the same grade of intelligence.......In default of some such comparison on a broad scale of the mortality of the infants of working and non-working women of similar ages, races, intelligence, and living conditions, no one can determine accurately how many of the deaths of working women's infants are due to the mother's work and how many to the other conditions of their lives and environment." (p. 18).

The author illustrates the point by taking the six New England States and giving the infant deathrate, percentage of women of 16 years and over who are breadwinners, percentage of foreign-born to the population and percentage of population living in towns of 4000 and more inhabitants, and showing that, though the states with the highest infant mortality have also the largest number of women employed, they have also the largest percentage of foreign-born and of those living in urban surroundings, and that it is therefore impossible without further investigation to assign the infant deathrate to any of these three factors.

A further investigation has been undertaken into the 32 Massachusetts cities and the deathrate under a year is given, the percentage of foreign-born, the births per 1000 of the population*, the percentage of women gainfully employed and the percentage illiterate, and a comparison is made between the ten cities with the highest and the ten cities with the lowest infant deathrate and percentage of women employed and the other factors enumerated. The conclusion is reached that "These comparisons indicate, superficially at least, that a more direct relation exists between infant mortality and the birthrate, the percentage of foreign-born, and the percentage of female illiteracy than between infant mortality and the employment of women." (p. 38).

There can be no doubt that a direct study of the infant mortality in relation to women's employment can only properly be made, when we confine our attention to women, employed and unemployed, who are actually mothers and live in the same town, and when we correct for age†, and if possible home conditions. Still if we take a series of different towns the right method must be to correct by the method of partial correlation for such divergent factors as we are able to ascertain and allow for in the series of towns investigated. I have endeavoured to apply modern statistical methods to the data of this Report, taking as measures of the environmental conditions in the towns: D the general deathrate, i = percentage of illiteracy, f = percentage of foreign-born population, e = percentage of females employed 10 years of age and upwards (*note, not percentage of employed mothers*, so we may be largely measuring effect of child labour on future motherhood), and d = deaths under one year per 1000 births. Then we have for correlations:

$$r_{de} = {\cdot}68, \quad r_{di} = {\cdot}70, \quad r_{df} = {\cdot}74.$$

Hence numbers of foreign-born and of illiterate appear to be slightly more influential on infantile mortality than employment of women. These values are certainly high and the first is the sort of crude value which is used as an argument against the employment of women. Proceeding to partial correlations we have

$$_{i}r_{de} = {\cdot}36, \quad _{f}r_{de} = {\cdot}43, \quad _{f}r_{ei} = {\cdot}48,$$

$$_{e}r_{di} = {\cdot}42, \quad _{e}r_{df} = {\cdot}57, \quad _{f}r_{di} = {\cdot}31.$$

We next corrected for two factors and found:

$$_{if}r_{de} = {\cdot}34, \quad _{ef}r_{di} = {\cdot}12, \quad _{ie}r_{df} = {\cdot}43.$$

Thus we see that illiteracy has least influence on the infantile deathrate and the presence of foreign-born most.

But even the presence of foreign-born and of illiterates is not a very complete measure of environmental effects liable to influence the infantile mortality in different towns as apart from employment of women. Many women employed means industrial conditions and possibly generally bad environment. I have taken as a measure of this the general deathrate D and find

$$r_{Dd} = {\cdot}71, \quad r_{De} = {\cdot}47, \quad r_{Di} = {\cdot}60, \quad r_{Df} = {\cdot}49.$$

Whence I find:

$$_{D}r_{de} = {\cdot}57, \quad _{D}r_{di} = {\cdot}62, \quad _{D}r_{df} = {\cdot}75,$$

$$_{D}r_{fe} = {\cdot}49, \quad _{D}r_{ei} = {\cdot}61, \quad _{D}r_{if} = {\cdot}68,$$

showing very substantial relations after correction for a general measure of poor environment.

---

[*] The author is not very confident of the full accuracy of the complete registration of births.

[†] Young women are often employed up to the birth of their first one or two children, but the deathrate of these elder-born is heavier than the deathrate of those who immediately follow.

Next proceeding to allow for two factors we find

$$ {}_{fD}r_{di} = \cdot 23, \quad {}_{fD}r_{ei} = \cdot 44, \quad {}_{fD}r_{de} = \cdot 35, $$

the latter result shows that general deathrate and illiteracy are about equally influential on the relation of employment of women to infantile mortality. Finally I corrected for all three factors and found:

$$ {}_{ifD}r_{de} = \cdot 28 $$

or 60 °/ₒ of the crude correlation $r_{de} = \cdot 68$ is due to women being most employed in towns where the general deathrate is high, where illiterates are frequent and the population is largely foreign-born. How much further the relationship would be reduced, could we equalise other features of these Massachusetts cities, it is not possible to predict. The examination of the individuals in one city appears to me to be the only satisfactory method of disentangling the numerous factors which influence infant mortality. We commend, however, the study of the first part of this *Report*, as it deals very clearly with the difficulties which arise, and will counteract the tendency, which is prevalent, to assert causation whenever association is observed. The author lays stress on avoiding such logical confusions.

Part II of the *Report* deals with infant mortality and its relation to the employment of mothers in Fall River, Massachusetts. In 1908 the attempt was made to visit the homes of each of the mothers of the 859 infants who died during the year and to ascertain details concerning her occupation, etc. In 279 cases the family could not be traced. In 266 cases prior to the birth of the child the mother was at work outside the home while in 314 cases the mother's work was limited to household duties or other work carried on entirely at home. Thus only the cases of *deaths* are dealt with and the causes of death are compared in the two groups of cases (1) when the mother was at work outside the home prior to the birth of the child and (2) when the mother's work was carried on entirely in the home.

I hold that this method will never prove as satisfactory as that employed in districts in England; in England certain districts are chosen and every baby within that area is visited and the deathrate per number born in one group can be compared with another and the circumstances surrounding those babies who survive and those who die in the first year of life in a given district can be analysed.

I do not think that the fact that a rather higher percentage of all deaths from gastritis etc. in Fall River occur when the mother works away from home and a rather higher percentage from congenital debility at birth when the mother does not work away from home will help us much in discovering the influence of the employment of the mother on infant mortality, nor do I think it will throw much light on the question of stillbirths with which the Report also deals. It is found that there are no more stillbirths proportional to all deaths when the mother is industrially employed, but it seems to me that this tells us nothing about the number of stillbirths proportional to all births. The real question is whether mothers employed away from home in factory or workshop, whose other circumstances are the same, lose more children in the first year of life or have more children stillborn than the mothers who are only employed in their homes and I do not think a comparison of causes of death will lead us much further, and I think it may lead to difficulties.

When dealing with the mother's work after childbirth in relation to the causes of infant mortality it is pointed out that the smaller percentage of deaths from congenital disease among the children of mothers who returned to work after childbirth was owing to the fact that most of the children dying from this group of causes died in the early weeks of life before the mother returned to work. For this same reason the number of deaths from gastritis etc. of children whose mothers returned to work is exaggerated, for we are missing out a whole series of illnesses which have ceased to add to the child deathrate by the time the mother returns to work and we must increase in this way the percentage of deaths of any disease of the later months of a child's first year of life.

It seems to me that a comparison of deaths in this way will really give very little information; an excess of deaths from one disease means a defect in some other disease; it is shown that when the baby is nursed exclusively by the mother 26·0 per cent. of the deaths were from diarrhoea, gastritis, etc.; when partly nursed the percentage was 52·3 and when artificial food was exclusively employed the percentage of deaths from diarrhoea etc. was 42·9; the baby certainly dies less from gastritis when it is breast fed but it dies in greater numbers from other causes. Here again there is a difficulty; deaths from congenital diseases fall on the first weeks of life when breast feeding is the rule, while deaths from gastritis etc. fall on the later months of child life when "partial breast feeding" has become more common and I do not think it is possible to draw any conclusions from a comparison of deaths from one disease to deaths from all diseases as to the importance of artificial feeding in relation to deaths from gastritis.

Interesting information is given as to the reasons for artificial feeding; the numbers are not large enough to justify any definite conclusions, but this is such an important part of any inquiry into the influence of artifical feeding on the infant deathrate that one welcomes its inclusion in a report of this kind.

We have been requested by Professor F. M. Urban to insert the accompanying announcement.

ANNOUNCEMENT.

A prize of One Hundred Dollars ($100.00) is offered for the best paper on the Availability of Pearson's Formulae for Psychophysics.

The rules for the solution of this problem have been formulated in general terms by William Brown. It is now required (1) to make their formulation specific, and (2) to show how they work out in actual practice. This means that the writer must show the steps to be taken, in the treatment of a complete set of data (Vollreihe), for the attainment in every case of a definite result. The calculations should be arranged with a view to practical application, i.e. so that the amount of computation is reduced to a minimum. If the labour of computation can be reduced by new tables, this fact should be pointed out.

The paper must contain samples of numerical calculation, but it is not necessary that the writer have experimental data of his own. In default of new data, those of F. M. Urban's experiments on lifted weights (all seven observers) or those of H. Keller's acoumetrical experiments (all results of one observer in both time-orders) are to be used.

Papers in competition for this Prize will be received, not later than December 31st, 1914, by Professor E. B. Titchener, Cornell Heights, Ithaca, N.Y., U.S.A. Such papers are to be marked only with a motto, and are to be accompanied by a sealed envelope, marked with the same motto, and containing the name and address of the writer. The Prize will be awarded by a committee consisting of Professors William Brown, E. B. Titchener and F. M. Urban.

The committee will make known the name of the successful competitor on July 1, 1915. The unsuccessful papers, with the corresponding envelopes, will be destroyed (unless called for by their authors) six months after the publication of the award.

Corrigendum. Dr Derry has most kindly pointed out a slip on p. 307, Vol. VIII; the value of 100 $(B-H)/L$ for Congo female crania is $+1·9$ and not $-1·9$, which brings these crania nearer to their proper place, and the remarks on this point p. 308 should accordingly be cancelled.

Journal of Anatomy and Physiology

CONDUCTED BY

SIR WILLIAM TURNER, K.C.B. ARTHUR THOMSON, University of Oxford
ALEX. MACALISTER, University of Cambridge ARTHUR KEITH, Royal College of Surgeons
ARTHUR ROBINSON, University of Edinburgh

VOL. XLVIII

ANNUAL SUBSCRIPTION 21/- POST FREE

CONTENTS OF PART III.—APRIL 1914

A. Francis Dixon. Note on Two Cases of Well-marked Suprasternal Bones. Professor Peter Thompson. The Development of the Lobus Quadratus of the Liver, with Special Reference to an Unusual Anomaly of this Lobe in the Adult. Professor F. G. Parsons. The Characters of the English Thigh-Bone. Frederic Wood Jones, D.Sc. The Lower Ends of the Wolffian Ducts in a Female Pig Embryo. D. Davidson Black, B.A., M.B. (Tor.). Two Cases of Cardiac Malformation—more especially of the Infundibular Region. Ralph Thompson, Ch.M., F.R.C.S. Figures relative to Congenital Abnormalities of the Upper Urinary Tract, and some Points in the Surgical Anatomy of the Kidneys, Ureter, and Bladder. Harold Rischbieth, M.A., M.D., B.C. (Cantab.), F.R.C.S. (England). Anomaly of the Inferior Vena Cava: Duplication of the Post-Renal Segment. Bernard Coen, M.B. A Communication as to the Causation of large Vascular Grooves found on the Inner Aspect of the Os Parietale. Rupert M. Downes, M.D., M.S. (Melb.). The Interrelationship of some Trunk Measurements and their Relation to Stature. Professor J. T. Wilson, M.B., F.R.S. Observations upon Young Human Embryos. Review: Biedl. Innere Sekretion: ihre physiologische Grundlagen und ihre Bedeutung für die Pathologie.

LONDON: CHARLES GRIFFIN AND COMPANY, Ltd., Exeter Street, Strand

---

JOURNAL OF THE ROYAL ANTHROPOLOGICAL INSTITUTE

Vol. XLIII. July—December, 1913

Contents:—

Sollas, W. J., M.A., Sc.D., LL.D., F.R.S. Paviland Cave: an Aurignacian Station in Wales (*The Huxley Memorial Lecture for* 1913). (With Plates XXI—XXIV.) Johnston, Sir H. H., G.C.M.G., K.C.B., D.Sc. A Survey of the Ethnography of Africa; and the Former Racial and Tribal Migrations in that Continent. Evans, Ivor H. N. Folk Stories of the Tempassuk and Tuaran Districts, British North Borneo. Dundas, Hon. Charles. History of Kitui. Parsons, F. G. On some Bronze Age and Jutish Bones from Broadstairs, with Type Contours of all the Bronze Age Skulls in the Royal College of Surgeons' Museum. Seligmann, C. G., M.D. Some Aspects of the Hamitic Problem in the Anglo-Egyptian Sudan. (With Plates XXV—XXXVIII.) Hilton-Simpson, M. W., F.R.G.S. Some Arab and Shawia Remedies and Notes on the Trepanning of the Skull in Algeria. (With Plate XXXIX.) Miscellanea. Sanderson, Meredith G. Native Games of Central Africa. Index, Title Page and List of Fellows.

WITH NINETEEN PLATES AND MANY ILLUSTRATIONS IN THE TEXT.

PRICE 15s. NET

LONDON: THE ROYAL ANTHROPOLOGICAL INSTITUTE, 50, Great Russell Street, W.C.
or through any Bookseller

MAN

A MONTHLY RECORD OF ANTHROPOLOGICAL SCIENCE

Published under the direction of the Royal Anthropological Institute of Great Britain and Ireland.

Each number of **MAN** consists of at least 16 Imp. 8vo. pages, with illustrations in the text together with one full-page plate; and includes Original Articles, Notes, and Correspondence; Reviews and Summaries; Reports of Meetings; and Descriptive Notices of the Acquisitions of Museums and Private Collections.

Price, 1s. Monthly or 10s. per Annum prepaid.

TO BE OBTAINED FROM THE

ROYAL ANTHROPOLOGICAL INSTITUTE, 50, Great Russell Street, W.C.

AND THROUGH ALL BOOKSELLERS

CONTENTS

(All Rights reserved)

The publication of a paper in *Biometrika* marks that in the Editor's opinion it contains either in method or material something of interest to biometricians. But the Editor desires it to be distinctly understood that such publication does not mark assent to the arguments used or to the conclusions drawn in the paper.

Biometrika appears about four times a year. A volume containing about 500 pages, with plates and tables, is issued annually.

Papers for publication and books and offprints for notice should be sent to Professor Karl Pearson, University College, London. It is very desirable that a copy of all measurements made, not necessarily for publication, should accompany each manuscript. In all cases the papers themselves should contain not only the calculated constants, but the distributions from which they have been deduced. Diagrams and drawings should be sent in a state suitable for direct photographic reproduction, and if on decimal paper it should be *blue* ruled, and the lettering only pencilled.

Papers will be accepted in German, French or Italian. In the first case the manuscript should be in Roman not German characters.

Contributors receive 25 copies of their papers free. Fifty additional copies may be had on payment of 7/- per sheet of eight pages, or part of a sheet of eight pages, with an extra charge for Plates; these should be ordered when the final proof is returned.

The subscription price, payable in advance, is 30*s.* *net* per volume (post free); single numbers 10*s.* *net*. Volumes I, II, III, IV, V, VI, VII, VIII and IX (1902—13) complete, 30*s.* *net* per volume. Bound in Buckram 34/6 *net* per volume. Index to Volumes I to V, 2*s.* *net*. Subscriptions may be sent to C. F. Clay, Cambridge University Press, Fetter Lane, London, E.C., either direct or through any bookseller, and communications respecting advertisements should also be addressed to C. F. Clay.

Till further notice, new subscribers to *Biometrika* may obtain Vols. I—IX together for £10 *net*—or bound in Buckram for £12 *net*.

The Cambridge University Press has appointed the University of Chicago Press Agents for the sale of *Biometrika* in the United States of America, and has authorised them to charge the following prices:—$7.50 net per volume; single parts $2.50 net each.

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS.

Vol. X. Parts II and III. November, 1914

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY
KARL PEARSON

CAMBRIDGE UNIVERSITY PRESS
C. F. CLAY, MANAGER
LONDON: FETTER LANE, E.C.
EDINBURGH: 100, PRINCES STREET
also
H. K. LEWIS, 136, GOWER STREET, LONDON, W.C.
WILLIAM WESLEY AND SON, 28, ESSEX STREET, LONDON, W.C.
CHICAGO: UNIVERSITY OF CHICAGO PRESS
BOMBAY AND CALCUTTA: MACMILLAN AND CO., LIMITED
TORONTO: J. M. DENT AND SONS, LIMITED
TOKYO: THE MARUZEN-KABUSHIKI-KAISHA

Price Twenty Shillings net.

[*Issued December* 3, 1914]

BIOMETRIC LABORATORY PUBLICATIONS

Drapers' Company Research Memoirs.

Biometric Series.

I. **Mathematical Contributions to the** Theory of Evolution.—XIII. On the Theory of Contingency and its Relation to Association and Normal Correlation. By KARL PEARSON, F.R.S. *Price* 4*s. net.*

II. **Mathematical Contributions to the** Theory of Evolution.—XIV. On the Theory of Skew Correlation and Non-linear Regression. By KARL PEARSON, F.R.S. *Price* 5*s. net.*

III. **Mathematical Contributions to the** Theory of Evolution.—XV. On the Mathematical Theory of Random Migration. By KARL PEARSON, F.R.S., with the assistance of JOHN BLAKEMAN, M.Sc. *Price* 5*s. net.*

IV. **Mathematical Contributions to the** Theory of Evolution.—XVI. On Further Methods of Measuring Correlation. By KARL PEARSON, F.R.S. *Price* 4*s. net.*

V. **Mathematical Contributions to the** Theory of Evolution.—XVII. On Homotyposis in the Animal Kingdom. By ERNEST WARREN, D.Sc., ALICE LEE, D.Sc., EDNA LEA-SMITH, MARION RADFORD, and KARL PEARSON, F.R.S. [*Shortly.*

VI. **Albinism in Man.** By KARL PEARSON, E. NETTLESHIP, and C. H. USHER. Text, Part I, and Atlas, Part I. *Price* 35*s. net.*

VII. **Mathematical Contributions to the** Theory of Evolution.—XVIII. On a Novel Method of Regarding the Association of two Variates classed solely in Alternative Categories. By KARL PEARSON, F.R.S. *Price* 4*s. net.*

VIII. **Albinism in Man.** By KARL PEARSON, E. NETTLESHIP, and C. H. USHER. Text, Part II, and Atlas, Part II. *Price* 30*s. net.*

IX. **Albinism in Man.** By KARL PEARSON, E. NETTLESHIP, and C. H. USHER. Text, Part IV, and Atlas, Part IV. *Price* 21*s. net.*

Studies in National Deterioration.

I. **On the Relation of Fertility in Man** to Social Status, and on the changes in this Relation that have taken place in the last 50 years. By DAVID HERON, M.A., D.Sc. *Price* 3*s.* Sold only with complete sets.

II. **A First Study of the Statistics of** Pulmonary Tuberculosis (Inheritance). By KARL PEARSON, F.R.S. *Price* 3*s. net.*

III. **A Second Study of the Statistics of** Pulmonary Tuberculosis. Marital Infection. By ERNEST G. POPE, revised by KARL PEARSON, F.R.S. With an Appendix on Assortative Mating by ETHEL M. ELDERTON. *Price* 3*s. net.*

IV. **The Health of the School-Child in relation** to its Mental Characters. By KARL PEARSON, F.R.S. [*Shortly.*

V. **On the Inheritance of the Diathesis** of Phthisis and Insanity. A Statistical Study based upon the Family History of 1,500 Criminals. By CHARLES GORING, M.D., B.Sc. *Price* 3*s. net.*

VI. **A Third Study of the Statistics of** Pulmonary Tuberculosis. The Mortality of the Tuberculous and Sanatorium Treatment. By W. P. ELDERTON, F.I.A., and S. J. PERRY, A.I.A. *Price* 3*s. net.*

VII. **On the Intensity of Natural Selection** in Man. (On the Relation of Darwinism to the Infantile Death-rate.) By E. C. SNOW, D.Sc. *Price* 3*s. net.*

VIII. **A Fourth Study of the Statistics of** Pulmonary Tuberculosis: the Mortality of the Tuberculous: Sanatorium and Tuberculin Treatment. By W. PALIN ELDERTON, F.I.A., and SIDNEY J. PERRY, A.I.A. *Price* 3*s. net.*

IX. **A Statistical Study of Oral Temperatures** in School Children with special reference to Parental, Environmental and Class Differences. By M. H. WILLIAMS, M.B., JULIA BELL, M.A., and KARL PEARSON, F.R.S. *Price* 6*s. net.*

Questions of the Day and of the Fray.

I. **The Influence of Parental Alcoholism** on the Physique and Ability of the Offspring. A Reply to the Cambridge Economists. By KARL PEARSON, F.R.S. *Price* 1*s. net.*

II. **Mental Defect, Mal-Nutrition, and** the Teacher's Appreciation of Intelligence. A Reply to Criticisms of the Memoir on 'The Influence of Defective Physique and Unfavourable Home Environment on the Intelligence of School Children.' By DAVID HERON, D.Sc. *Price* 1*s. net.*

III. **An Attempt to correct some of the** Misstatements made by Sir VICTOR HORSLEY, F.R.S., F.R.C.S., and MARY D. STURGE, M.D., in their Criticisms of the Memoir: 'A First Study of the Influence of Parental Alcoholism,' &c. By KARL PEARSON, F.R.S. *Price* 1*s. net.*

IV. **The Fight against Tuberculosis and** the Death-rate from Phthisis. By KARL PEARSON, F.R.S. *Price* 1*s. net.*

V. **Social Problems: Their Treatment,** Past, Present and Future. By KARL PEARSON, F.R.S. *Price* 1*s. net.*

VI. **Eugenics and Public Health.** Lecture to the York Congress of the Royal Sanitary Institute. By KARL PEARSON, F.R.S. *Price* 1*s. net.*

VII. **Mendelism and the Problem of Mental** Defect. I. A Criticism of Recent American Work. By DAVID HERON, D.Sc. (Double Number.) *Price* 2*s. net.*

VIII. **Mendelism and the Problem of Mental** Defect. II. The Continuity of Mental Defect. By KARL PEARSON, F.R.S., and GUSTAV A. JAEDERHOLM. *Price* 1*s. net.*

IX. **Mendelism and the Problem of Mental Defect.** III. On the Graduated Character of Mental Defect, and on the need for standardizing Judgments as to the Grade of Social Inefficiency which shall involve Segregation. By KARL PEARSON, F.R.S. (Double Number.) *Price* 2*s. net.*

THE ABOVE MAY BE OBTAINED FROM DULAU & CO., LTD., 37, SOHO SQUARE, LONDON, W.

VOLUME X NOVEMBER, 1914 NOS. 2 AND 3

A PIEBALD FAMILY.

BY E. A. COCKAYNE, M.D., M.R.C.P.

In spite of the great interest, which they have always excited, well authenticated examples of piebalds in the dark races have been found to be rare. In the white races they are much less conspicuous, in part owing to the presence of clothing and in part owing to the lack of contrast between the pigmented and unpigmented skin, but the likelihood of their coming under the notice of a skilled observer is much greater. The scarcity of records shows that piebalds in the white races also must be very uncommon. Last year I met with a case in a baby, and found that the child belonged to a family, many of whose members showed a similar defect of pigmentation. The family, belonging to a farming stock, originally came from the neighbourhood of Bury St Edmunds in Suffolk, and the anomaly is known to have descended directly through six generations. The oldest member, with whom I have talked, is fairly certain that it was present in one generation at least before this.

Of the first two generations in the pedigree (see Plate XI), I could obtain no definite information except the statement as to the existence of the piebaldism in I. 1 and II. 1, but of the third, III. 2 is said to have had a frontal blaze of white hair and white skin on the neck and forearms, which was very conspicuous owing to its marked contrast with the neighbouring weather-stained normal skin.

III. 4 appears to have been the only member of the family who showed a marked dislike to the condition and always wore a wig to hide the frontal blaze. III. 2, whose family name was C——*, had fifteen children. The first, IV. 2, a male, with dark hair, married twice, and had eight normal children, five by the first wife and three by the second. The second child, IV. 5, was a piebald, with a large frontal blaze, white skin on the front of the neck and arms, and blue eyes. He transmitted the condition to all his three children. V. 3, the eldest boy, aged 22 and unmarried, possesses dark hair, with a V-shaped frontal blaze of white or cream coloured hair, the apex of the V commencing near the coronal suture and spreading out to a width of 3½ inches, as it reaches the forehead. The eyebrows and some of the eyelashes are white. The next boy, V. 4, is aged 18. He has light hair and a very large blaze of unpigmented hair, which covers the whole of the top of the head. His eyebrows and eyelashes are white, and the eyes are blue. Both boys have white patches on the front of the neck and on the arms (see Plate XII).

---

* Names preserved in the confidential register of the Galton Laboratory.

Next in the fourth generation were twins, IV. 6 and 7, both piebalds. They were evidently not uniovular, because one had dark hair, and one light, and the white blazes were dissimilar in extent, but it is uncertain which had the larger. Both died at an early age.

The next, IV. 8, a girl, was normal with dark hair and eyes and remained unmarried. Next came a woman, IV. 10, who was a piebald with a large frontal blaze, white eyebrows and eyelashes, and white skin on the front of the neck and forearms. *The right eye was blue, and the left brown* (see Plate XIII (B)). Her child, aged 13, is quite normal with light hair and dark eyes.

The next child, IV. 12, Mrs W——, has a large frontal blaze and dark brown irides. There is a large irregular patch of white skin extending from just below the chin to the heads of the clavicles, and round it the skin appears to be more deeply pigmented than the rest of the skin of the neck. There are a few small islands of pigmented skin near the edge of the unpigmented area. The skin of the anterior aspect of the forearms is unpigmented from the elbows to the wrists, and here also, there are some small islands of pigmented skin in marked contrast to the unpigmented area, in which they lie (see Plate XIV).

The first two children of this individual were daughters, V. 8 and V. 9, both piebald, the third a normal son, V. 10, and then three more piebald daughters, V. 11—13. The first of the daughters, Mrs G——, V. 8 (see Plate XIV), is very fair with a very large frontal blaze covering the whole of the top of the head, and her eyebrows and eyelashes are white. Her normal hair has pale creamy diffused pigment and, according to the individual hair, some to a decided number of granules*. The hair of the blaze has no diffused pigment and no granules. The irides are light brown, but the outer segments on both sides are paler and greenish in colour. The skin of the forehead and base of the nose is very pale in colour. She has a large white patch on the skin of the front of the neck, beginning just below the chin and widening out so as to embrace that over the inner ends of both clavicles. As in her mother there appears to be some concentration of pigment round this white area, and there are small isolated areas of pigmented skin near its edge. She has unpigmented skin on the anterior aspect of both forearms.

Of her two children the first, VI. 1, a boy aged 3, is normal, the second, VI. 2, a boy aged $1\frac{1}{2}$, is a piebald (see Plate XIV). This child, VI. 2, was nine months old when first seen. He had a very large frontal blaze, resembling that of his mother and covering all the top of the head, the eyebrows and eyelashes were white with the exception of some of the outer hairs. Hair, pale cream in colour, said to be from the light area, has pale creamy diffused pigment and some granules (β), the granules being very small. It was obvious, even at this age, that heterochromia iridis was present. The right iris was pale except for a sector of dark grey occupying the upper and outer quadrant, the left iris was entirely dark grey. No difference in colour of the skin of the neck or forearm could be made out.

---

* β to γ on the Galton Laboratory scale of granular pigmentation.

When the baby was seen after the summer of 1913, the grey portions of the irides were becoming brown, the pale portion was still light blue. The face and arms were sunburnt, and it was noticed that the forehead was paler than the rest of the face. There was a pale area on the front of the neck, and the whole anterior surfaces of the forearms were white, the edges being very irregular in contour. There was also a white streak running obliquely right across the posterior or extensor aspect of the left forearm, and this offered a marked contrast with the rest of the surface, which was very brown. When the sunburn had died away the difference between the pigmented and unpigmented skin could no longer be made out.

IV. 12's second daughter, V. 9, aged 23, has only a small cream coloured frontal blaze, and the rest of her hair is light brown (see Plates XV and XVI). The eyebrows are composed of an even mixture of brown and white hairs, and the eyelashes are similar, with brown and white hairs alternating. The irides are grey and uniformly pigmented. There is a large irregular area of white skin at the base of the neck.

The whole of the anterior aspect of the right forearm is unpigmented, and there are similar small areas scattered over the posterior aspect (see Plate XVII). The left forearm is white only on the anterior aspect.

The next girl, V. 11, is aged 9. She has a very small frontal blaze, but the skin of the forehead is pale (see Plate XV). The eyebrows show a division into two parts, on the inner halves grow white hairs only, and on the outer brown hairs. The eyelashes on the contrary consist of alternate brown and white hairs. The irides are grey and uniformly coloured (see Plate XVIII). There is only a small white area in the middle of the front of the neck, but there are well differentiated white areas on the anterior aspects of both forearms (see Plate XIII (A)), and on the inner aspects of both upper arms. Her hair was examined and the first sample showed very pale diffused pigment and some granules (β). Two more samples were then examined, one from the blaze and one from the neighbouring part of the scalp. The first showed no diffused pigment and no granules, the second showed the majority of hairs with yellow-brown diffused pigment and a decided number to plenty of small granules (γ—δ), but a few had no diffused pigment and no granules.

The next piebald child, V. 12, died young. She had a frontal blaze and blue irides. Some of her hair showed very pale diffused pigment, and some granules (β).

The next child, V. 13, also died young. She was a piebald nearer to the classical type than any of the others. She had a large frontal blaze, white skin on the forehead, and large areas of white skin on the front of the neck and chest, and in addition a very extensive area on the abdomen.

Of the fourth generation the next child, IV. 13, was a male with dark hair and eyes, who had 5 normal children; the next, IV. 15, had fair hair and died young. Twins, IV. 16 and 17, came next and died in infancy*. They were heterogeneous,

---

* The tendency to twin in this family is worth noting.

a dark-haired boy and a light-haired girl. A girl, IV. 19, was born next and she had twin sons, V. 15 and 16, who were also normal. The last three children, IV. 20—22, a girl, a boy, and one whose sex I am unable to ascertain, were all normally pigmented and all died at a very early age.

The pedigree confirms the strongly hereditary nature of piebaldism, and in this as in other published cases the character can affect either sex, but has only been transmitted by those affected. Unless we are to assume that in the case of such a rare anomaly as piebaldism, I. 2, II. 2 or III. 1, were really unnoticed piebalds, then III. 2 could only be heterozygous, or since piebaldism is dominant a (DR). We must take IV. 4, IV. 9, IV. 11 and V. 7 for pure recessives (RR). Thus the number of piebalds in the five sibships of generations IV, V and VI should be one quarter, i.e. $\frac{1}{4}(15+3+1+6+2)=7$ nearly. We have actually 14 out of 27, thus piebaldism does not seem to act numerically as a pure dominant.

The areas of unpigmented skin are less than in the classical piebalds, but it is probable that in some, at least, they are larger and more numerous than I have stated. On the covered parts of the body and legs, which I was unable to examine except in the baby, they would not be very noticeable. It was not until I had noticed the white skin on the neck and arms of one of them that I was told anything about the existence of similar patches on the others. If true, it is remarkable that none have had white patches on the legs.

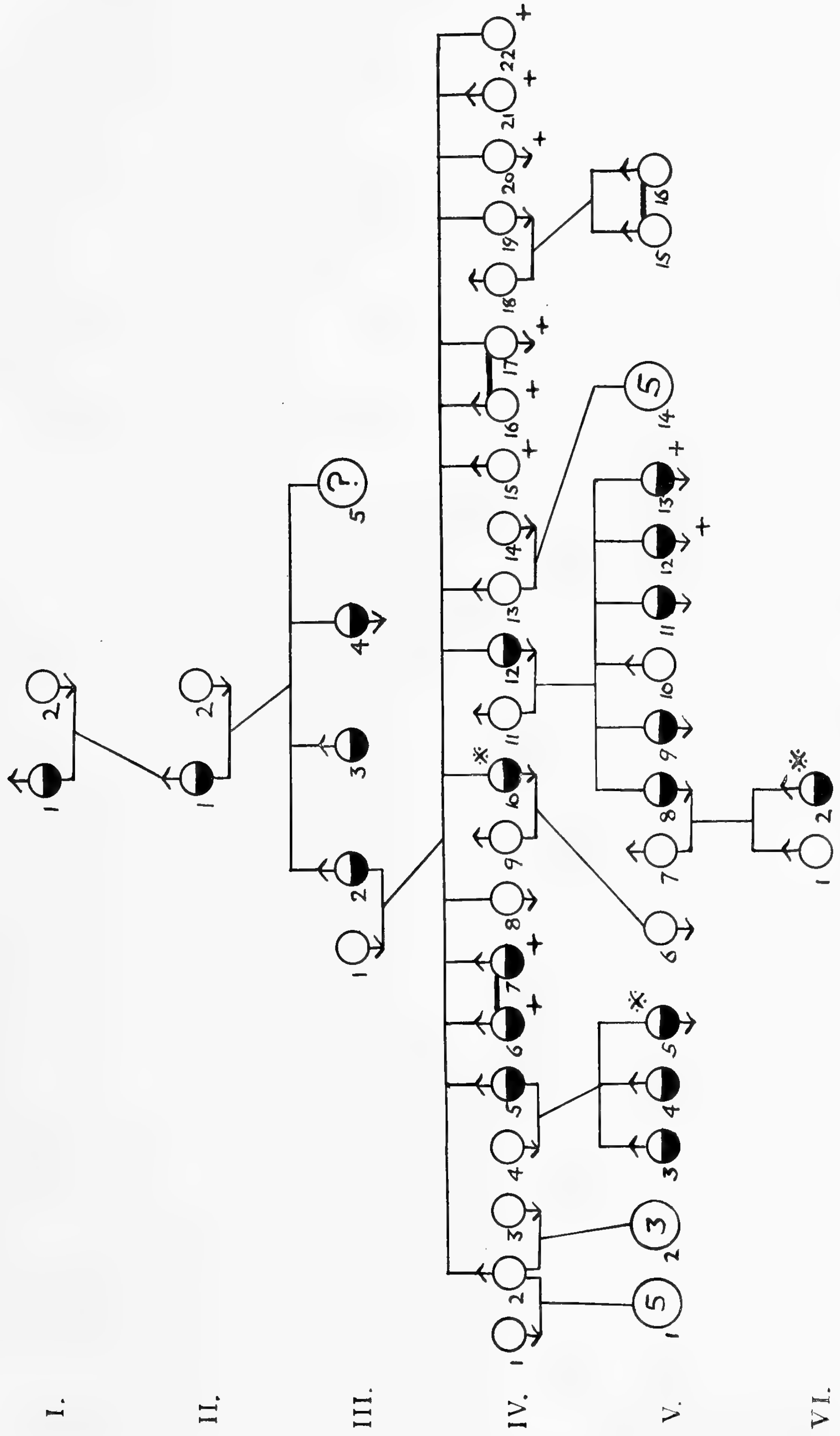
With regard to the local distribution of the pigment, there appears to be an excess at the edges of some of the unpigmented areas, as has been noted in other cases. In the case of other pale areas, the demarcation between them and the normal skin is very slight, and is probably due to the fact that they are not wholly unpigmented. This remark applies especially to the forehead, which in some of them looks paler than natural, but not wholly devoid of pigment.

In some the eyelashes are alternately white and brown, and in others the eyebrows are similar, and in one at least hairs growing on the scalp near the blaze are in some instances entirely without either diffused or granular pigment. This suggests that the skin beneath may show a deficient and irregular distribution of pigment.

The most interesting feature is the occurrence in three members of the family of well-marked heterochromia iridis, a character which has been met with in members of a piebald family, but always independently of their piebaldism, never, as in this case, in true association with it. It proves conclusively that these cases are not congenital leucoderma.

There seems to be no association of piebaldism and general lack of pigmentation of hair and irides. Affected and unaffected members have been both fair and dark, but the fairest piebalds seem to have the most extensive frontal blazes.

In the cases photographed the individuals were blonds and there has been great difficulty in getting a good photographic contrast of differences of pigmentation very noticeable in the living subject.



\+ Died young. ✻ Heterochromia iridis. ◑ Piebald. ○ Normal.

Pedigree of Piebald Family.

V. 3 and V. 4 as children showing their marked V-shaped frontal blazes.

A

V. 11. Showing a well differentiated leucotic area on the anterior aspect of the left arm.

B

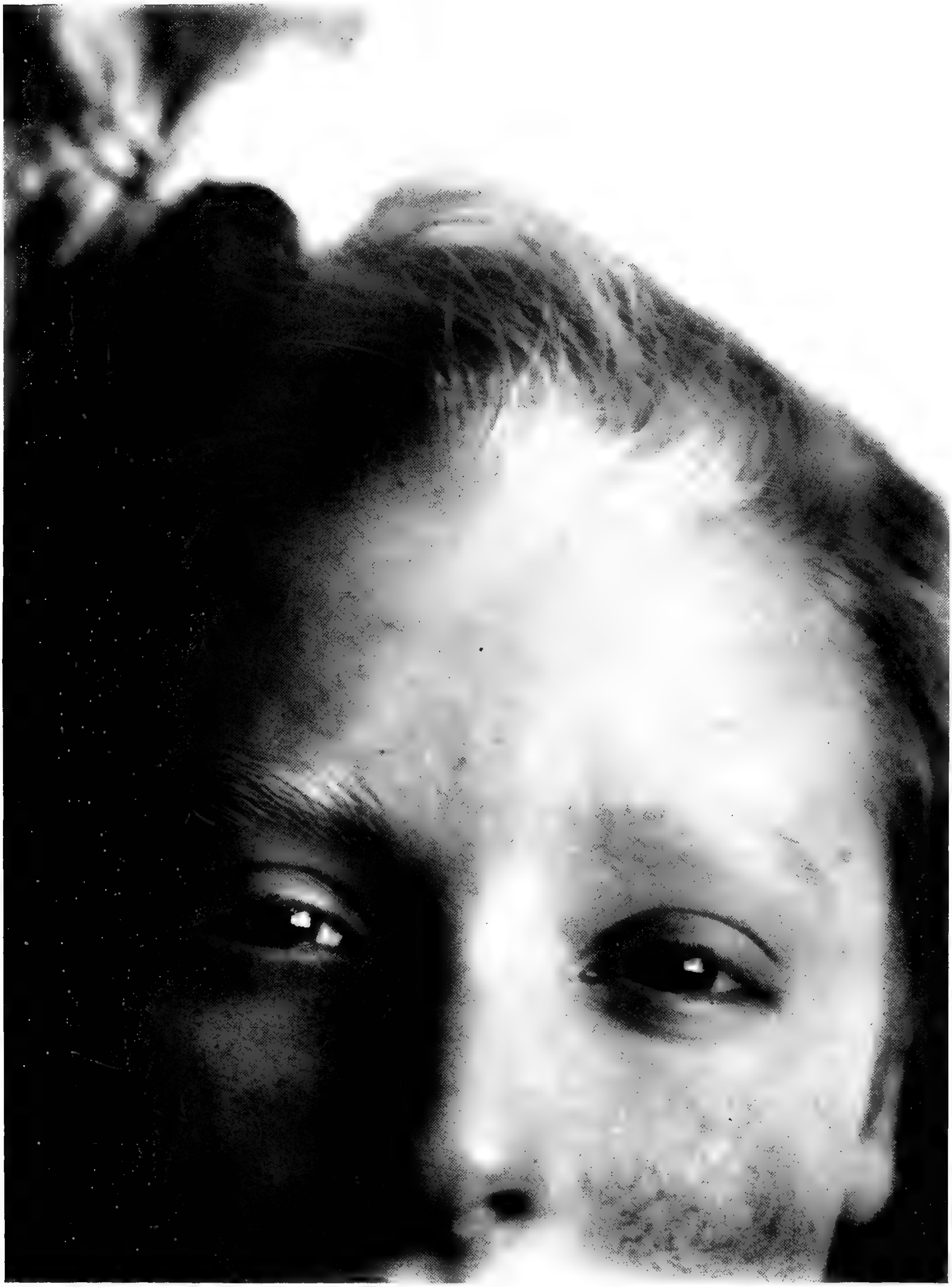
IV. 10. The blaze may be seen with difficulty under a lens, but the heterochromia iridis is easily distinguishable.

Piebaldism in three generations, V. 8, mother, IV. 12, grandmother, and VI. 2, grandson. The white frontal hair is visible in all three and the darker section of the right eye of VI. 2 with the use of a lens.

V. 11 and V. 9. Two sisters with white frontal blazes.

V. 9. Showing white forelock or blaze.

Right forearm of V. 9 showing white patches on posterior aspect. The photograph is untouched and it is difficult to bring out by photography the grades of pigmentation when the arm is untanned by the sun, although they are quite clear on actual inspection.



Large photograph of V. 11 to show paleness of forehead and white hairs on inner half of eyebrows.

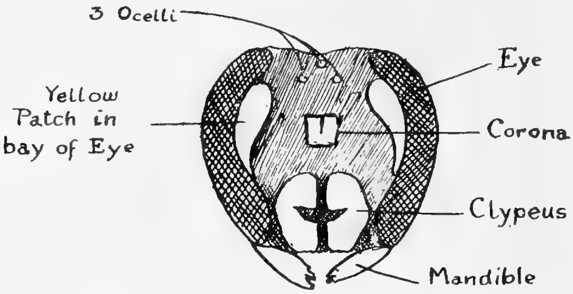
CLYPEAL MARKINGS OF QUEENS, DRONES AND WORKERS OF VESPA VULGARIS.

By OSWALD H. LATTER, M.A.

Upon the front of the head of *Vespa vulgaris* certain yellow markings stand out conspicuously upon the otherwise black surface. Below the three ocelli and between the upper portions of the two compound eyes there is a median four-sided yellow patch, the "corona"; to the right and left of this, separated from it by a fairly wide interval, and occupying the bay of each of the compound eyes is a pair of elongated yellow blotches; while straight below the corona and between the lower portions of the compound eyes is a very conspicuous yellow area which extends over the clypeus and down to the labrum or upper lip which lies between the two mandibles. This clypeal patch of yellow bears upon it a black mark which is subject to considerable variation. I distinguish in the queens and workers five chief types of this black mark: see diagrams on p. 202. In Type I a broad vertical black band extends right through the yellow patch from the top to the bottom; a little below its middle the band bears to right and left a pair of bluntly pointed and slightly upturned arms: the portion of the median band below these arms is somewhat narrower than that above. Type II is derived from I by suppression of the black portion below the transverse arms. In Type III the extent of the black colouring is yet further reduced by the absence of the upper half (or thereabouts) of the vertical band. In Type IV the lower part of the vertical band re-appears, but the width of all the components is very much less than in any of the preceding types. In Type V the component parts of the black marking cease to be in contact; the upper portion of the vertical band is interrupted by a broad belt of yellow; the two "arms" are separated from the lower part of what remains and from one another; while there is no black at all below these remnants of the "arms"—a feature recalling Types II and III. Types IV and V are however represented only by single individuals in the series examined.

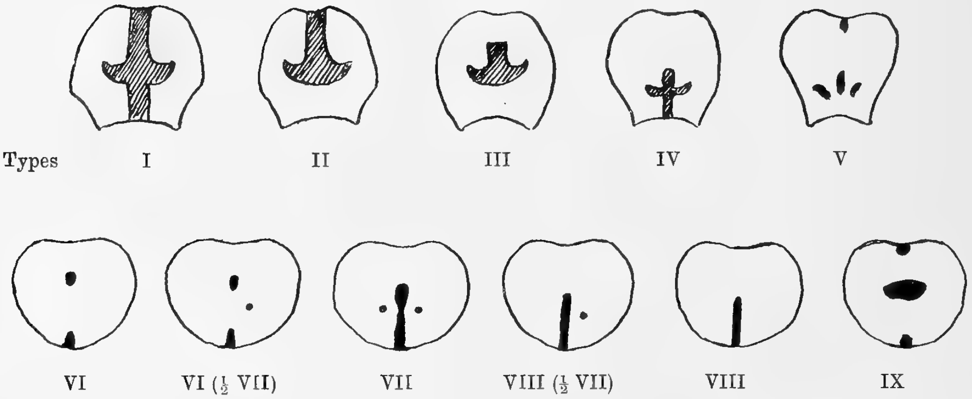
Between these main types certain intermediates occur. Thus some individuals have the black piece below the "arms" very narrow, approximating therefore to II, but conforming to I if we take extension of the black right through the yellow as the criterion of I; such individuals are distinguished as I + II. Others again conform to II but possess a slightly darker stain on the yellow in the line where the distinctive lower black portion of I might occur; these are called II + I. Similarly, intermediates between II and III are recognisable: in II + III the top of the upper portion of the vertical black band is very narrow; while in III + II there is a mere stain on the yellow of this region. A single instance occurs of an intermediate between I and III (I + III), where the vertical black band extends right through the yellow, but is much narrowed at its upper extremity.

Front of head of *V. vulgaris* ♀.



(All parts left white are actually yellow.)

Clypeus alone



Drawn by K. W. Merrylies.

My first examination consisted of about 200 tubes containing queens of *Vespa vulgaris* from *different* nests. In the case of some queens the heads were missing, and in the course of transit the contents of some of the tubes had got loose in the jar. I have numbered these 199 to 208. The results are given in Table I (p. 204) and the summary below:

Type I	Pure	50	62
„	I + II	11	
„	I + III	1	
Type II	II + I	22	120
„	Pure	94	
„	II + III	4	
Type III	III + II	1	4
„	Pure	3	
„	III + IV	0	
Type IV	IV + III	0	1
„	Pure	1	
„	IV + V	0	
Type V	V + IV	0	1
„	Pure	1	
		Total:	188

It will be seen that transitional cases undoubtedly occur. The bulk of the queens, however, fall into Types I and II, or queens are very little variable.

To test: (i) whether this variability was still further lessened by taking only the queens from a single nest, and (ii) the relative variability of queens, drones and workers, I now examined all the queens, workers and drones of a single nest of *V. vulgaris*.

In this case *all* the 127 queens were of Type II*.

The classes of the workers are given in Table II (p. 205) and may be summarised as follows:

Type I	Pure	5	10
„	v. s or (I + II) ?	5	
Type II	II + I	6	162
„	Pure	156	
		Total:	172

It will be seen that they are somewhat more variable than the queens of the same nest, but not so variable as queens from different nests.

I now turn to the drones of this same nest. I had 150 at my disposal.

The drones exhibit a very wide range of facial markings. In the material examined comparatively few fall into the scheme of classification adopted for the queens and workers, and it thus becomes necessary to resort to six types of face which appear to be peculiar to the male sex. These are numbered VI,

---

* There were 129 queens in this nest, but No. 34 was missing and No. 98 had its head damaged too badly for classing.

VI (½ VII), VII, VII (½ VIII), VIII and IX, see diagrams, p. 202. In Type VI there are two somewhat elongated black dots upon the yellow clypeus, one being sub-central, the other on the ventral margin; in VI (½ VII) the ventral dot is longer dorso-ventrally and a third dot appears upon the left side (right side, in figure seen from in front) opposite the gap between the two previous dots;

TABLE I.

Types of Clypeal Markings in V. vulgaris Queens.

No.		No.		No.		No.		No.	
1	—	44	III	87	I+II	130	—	173	III
2	II	45	II	88	II	131	II	174	I+II
3	II+I	46	II	89	I	132	II	175	II
4	I+II	47	—	90	II+I	133	II	176	II+I
5	II	48	II+I	91	I	134	—	177	II
6	II	49	II	92	II	135	I	178	II
7	II	50	II	93	III+II	136	II	179	II+I
8	II	51	I	94	—	137	II	180	II+I
9	II	52	II	95	II	138	I	181	II
10	I	53	I	96	II	139	II	182	—
11	II+I	54	II+I	97	I	140	I	183	II+III
12	II+I	55	II	98	II+I	141	I	184	—
13	I	56	—	99	II+III	142	I+II	185	I
14	I	57	II	100	I	143	I	186	II
15	II	58	—	101	—	144	II	187	II
16	II	59	I	102	II	145	II	188	I
17	II	60	II+I	103	I	146	I	189	I+II
18	II	61	II	104	II	147	II	190	II
19	II+I	62	II	105	II	148	—	191	I
20	II	63	II	106	—	149	I	192	I
21	I	64	II	107	I	150	II	193	III
22	I	65	II	108	II	151	II	193*a*	II
23	I	66	II	109	I	152	—	194	—
24	I+II	67	—	110	I+II	153	II	195	I+III
25	II+I	68	II	111	II+I	154	II	196	II
26	I	69	II+I	112	II	155	II	197	I
27	II	70	I	113	I	156	II	198	II+I
28	I	71	II	114	II	157	II		
29	I+II	72	I	115	II+III	158	II+I		
30	I	73	I	116	II	159	II		
31	II	73*a*	I+II	117	II	160	I	*Loose*	
32	II	74	I	118	I	161	I		
33	V	75	II	119	II	162	II	199	I
34	II	76	II	120	II+III	163	I	200	I
35	—	77	II	121	II	164	II	201	I
36	—	78	—	122	II	165	II	202	II
37	I	79	—	123	—	166	II	203	II
38	II	80	—	124	I	167	I	204	II
39	II+I	81	II	125	II	168	—	205	II
40	II+I	82	—	126	II+I	169	—	206	II
41*a*	II+I	83	II	127	II	170	II+I	207	II
41*b*	I	84	II	128*a*	II	171	I	208	II
42	II	85	I	128*b*	II	172	—	209	IV
43	I	86	I+II	129	I+II				

TABLE II.

Types of Clypeal Marking of Workers of a single Nest of V. vulgaris.

No.		No.		No.		No.		No.	
1	—	37	II	74	II	112	II	150	II
2	—	38	II	75	II	113	II	151	II+I
3	—	39	II	76	II	114	II	152	II
4	—	40	II	77	II	115	II	153	II
5	—	41	II	78	—	116	II	154	II
6	—	42	II	79	I	117	II	155	II
7	—	43	II	80	II	118	II	156	II
8	—	44	II	81	II	119	II	157	II
9	II	45	I v. v. s.	82	II	120	II	158	II
10	II		[I+II]	83	II	121	II	159	II
11	II	46	II	84	II	122	II	160	II
12	I v. s.	47	II	85	II	123	II	161	II
	[I+II]	48	II	86	II	124	II	162	II
13	II	49	—	87	II	125	II	163	II
14	II	50	II	88	II	126	II	164	II+I
15	I	51	II	89	II	127	II+I	165	II
16	II	52	II	90	II	128	II	166	II
17	II	53	II	91	II	129	II	167	II
18	II	54	II	92	II	130	II	168	II
19	II	55	II	93	II	131	II	169	II
20	II	56	II	94	II	132	II	170	II
21	II	57	II	95	II	133	II	171	II
22	II	58	I v. s.	96	II	134	II	172	II
23	II		[I+II]	97	—	135	II	173	II
24	II	59	II	98	II	136	II+I	174	II
25	I	60	II	99	II	137	II	175	II
26	II	61	II	100	II	138	II	176	II
27	II	62	II	101	II	139	II	177	II
28	II	63	II	102	II	140	II	178	II
29	II	64	II	103	II	141	II	179	II
30	I v. s.	65	I	104	II	142	II	180	II
	[I+II]	66	II	105	II	143	II	181	—
31	II	67	II	106	II	144	II	182	II
32	II+I	68	II	107	II	145	II	183	II
33	II	69	I	108	II	146	II	184	II
34	—	70	II	109	II+I	147	II	185	—
35	I v. s.	71	II	110	II	148	II	186	—
	[I+II]	72	II	111	II	149	II	187	II
36	II	73	II						

in VII the two median dots are united by a slender black line, and there is a pair of lateral dots, right and left; in VII ($\frac{1}{2}$ VIII) the median line is of uniform width, extending from about the centre to the lower margin, and to its left side there is a single dot; in VIII the median line alone is visible, both lateral dots having disappeared; while in IX there is no continuous median line, but merely two black spots, one at the extreme dorsal and the other at the extreme ventral side of the clypeus. It will be noticed that Types VII—VIII approximate to Type IV in so far as the black stripe begins at about the middle of the clypeus and extends right down to the ventral margin.

The data are given in Table III (p. 207) and are summarised below:

Type I	very narrow	6	7
	(I + VII)	1	
Type II	no horns 1, very narrow 1		2
Type III	. . . . . .		0
Type IV	. . . . . .		0
Type V	. . . . . .		0
Type VI	Pure	59	60
„	VI + VII	1	
Type VII	. . . . . .		22
Type VIII	(VIII near VI)	8	58
„	(VIII + $\frac{1}{2}$ VII)	2	
„	Pure	48	
Type IX	. . . . . .		2
		Total:	151

It will be realised at once how far more variable the drones, of even one nest, are than the workers or queens for this character. But their variability is rather of a negative than a positive character, appearing to consist in more or less extensive absence of the fuller markings of queen and worker.

The results here deduced for variability of non-measured characters do not wholly agree with those found by Wright, Lee and Pearson on the wing measurements of the same nest of *V. vulgaris.* They found that for *absolute* measurements the variability as determined by the coefficient of variation was in every case such that the worker was more variable than the drone and the drone than the queen. On the other hand they found when they dealt with *indices* that the drone for wing measurements was slightly more variable than the worker and the queen less variable than either*. Possibly the divergence apparent here may be explicable in the sense of the drone's variability lying in the present case in an absence of marking rather than in any positive variation. The drone's variation is about a centre of much diminished marking. If we could measure the variation in the total area of marking in queen and worker we might find it as great as the variations in the smaller markings of the drone.

It would be of much interest to investigate a series of drones from different nests. It is clear that the clypeal markings form a secondary sexual character and they would probably provide classifications for hereditary purposes.

---

* *Biometrika*, Vol. v. pp. 414 and 421.

TABLE III.

Types of Clypeal Markings in Drones *of a single Nest of V. vulgaris.*

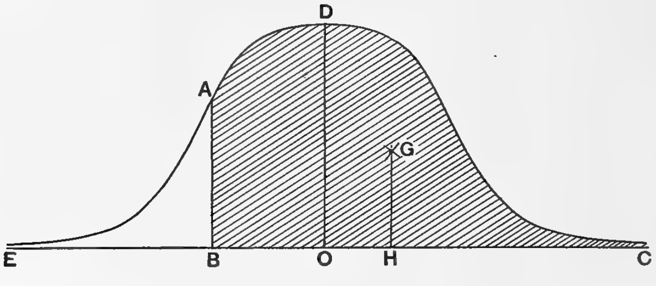
No.		No.		No.		No.		No.	
1	VIII	33	VII	63	VII	93	VI	123	VIII
	near VI	34	VI	64	VI	94	VI	124	I
2	VI	35	VIII	65	VII	95	VI		narrow
3	VIII	36	VI	66	VIII	96	VII	125	VI
4	VIII	37	VI	67	I	97	I	126	VIII
5	VI	38	VII		very		very	127	VII
6	VI	39	VI		narrow		narrow	128	VII
7	VIII	40	VII	68	VI	98	VIII	129	VIII
8	VIII	41	VI	69	VIII	99	VIII	130	VIII
9	VIII	42	VI	70	VI	100	VI	131	VI
10	IX	43	VIII	71	II	101	VI	132	VIII
11	VI	44	VI		but no	102	VIII	133	VI
12	VIII	45	VIII		horns	103	VIII	134	VI
13	VI	46	VIII	72	VIII		near VI	135	VI
14	VI	47	VI	73	VI	104	VI	136	VIII
15	VIII	48	VIII	74	VIII	105	VIII	137	VI
	near VI	49	VII	75	VIII	106	VII	138	I
16	VIII	50	VIII	76	VIII	107	VIII		dots of
17	VI	51	VII	77	VI	108	VIII		VII
18	VI	52	VII	78	VI	109	VI	139	VII
19	VIII	53	VIII	79	I	110	VI	140	VI
20	VII	54	VI		narrow	111	VIII	141	VI
21	VII	55	I	80	VI	112	VIII	142	VIII
22	VIII		very	81	VI	113	VI	143	VIII
23	VIII		narrow	82	VIII	114	VIII		near VI
	near VI	56	IX		near VI	115	VI	144	VI
24	VII	57	I	83	VIII	116	VI	145	VIII
25	VII		very	84	VII	117	II		near VI
26	VI		narrow	85	VI		very	146	VIII
27	VI	58	VIII	86	VI		narrow	147	VI
28	VI		near VI	87	VIII	118	—	148	VIII
	½ VII	59	VIII	88	VII	119	VII	149	VI
29	VIII	60	VIII	89	VI	120	VI	150	VI
30	VIII	61	VIII	90	VIII	121	VI	151	VI
31	VI	62	VIII	91	VII	122	VI	152	VI
32	VIII		½ VII	92	VII				
	½ VII								

TABLE OF THE GAUSSIAN "TAIL" FUNCTIONS; WHEN THE "TAIL" IS LARGER THAN THE BODY.

By ALICE LEE, D.Sc.

In a paper published in *Biometrika*, Vol. VI. pp. 59—68, tables for the incomplete normal moment functions were printed, and they have since been reproduced in *Tables for Statisticians and Biometricians* recently issued from the Cambridge University Press. From these tables values of the Gaussian "Tail" functions were deduced and a short table of ψ_1 and ψ_2 appeared in *Biometrika*, Vol. VI. p. 68. The value of these functions being demonstrated in practice during the last few years, a more complete table of ψ_1, ψ_2, ψ_3 has appeared in the *Tables for Statisticians and Biometricians.*

In the introduction to those tables, however, Professor Pearson indicated that it was important to have a similar table when the "tail" forms more than half the entire curve, and gave the fundamental formulae for obtaining the numerical values of the functions. The present table has been calculated to supply the want thus indicated.



Let the figure represent a Gaussian curve of total population N and standard deviation σ. Let AB be the ordinate at which it is truncated and let

$$OB = h = h' \times \sigma.$$

Let GH be the ordinate through the mean G of the truncated portion and $BH = d$, the distance of the mean from the line of truncation, let Σ be the standard deviation of the truncated portion about GH, and $n =$ the area of the truncated portion, or of the population observed. Then if any material be supposed to form a truncated portion of a normal curve, d, n and Σ can be found (see *Tables*, pp. xxvii and 25).

We have

$$\psi_1 = \Sigma^2/d^2 \dots\dots\dots\dots\dots\dots \text{(i)},$$

$$\psi_2 = \sigma/d \dots\dots\dots\dots\dots\dots \text{(ii)},$$

$$\psi_3 = N/n \dots\dots\dots\dots\dots\dots \text{(iii)}.$$

These are tabled for each value of h', at first proceeding by ·01 and then by ·10 as unit. Now ψ_1 being known we find h' from the table, and hence deduce ψ_2 and ψ_3. ψ_2 gives us the value of σ from known d. Hence $h = h' \times \sigma$ can be found, lastly (iii) gives us the total population from which n is drawn. Thus the constants N, σ and h which fix the total Gaussian are determined.

It will be sufficient to illustrate the method of using the tables on certain data as to the English thigh-bone, recently published by Parsons*.

Dwight† has adopted a method of sexing human femora on the basis of a markedly bimodal distribution obtained by him for American bones. He terms female any femur with diameter of head less than 45 mm., and male any femur with diameter of head over 47 mm. Parsons follows this rule and sexes by other points femora with heads from 45 to 47. As unsettled remainder he has 20 femora of 45 mm. and he gives 12 to ♀ and 8 to ♂; of 46 mm. and 47 mm. he has 41 femora and he gives 4 to ♀ and 37 to ♂. As a result of this process he obtains a female frequency curve which rises very abruptly at high values of the diameter, and a male frequency curve which rises very abruptly for low values of the femora. But, if there really be any marked skewness in frequency of the parts of the human skeleton, which is very unusual, we should anticipate that it would be of the *same* sense. Parsons' distributions are as follows (*loc. cit.* p. 256):

	36	*37*	*38*	*39*	*40*	*41*	*42*	*43*	*44*	*45*	*46*	*47*	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*
♀	1	1	—	3	8	14	12	18	12	12	3	—	1	—	—	—	—	—	—	—
♂	—	—	—	—	—	—	—	—	—	8	8	29	17	31	19	13	10	6	8	2

The ♀ 48 mm. femur according to the rule should have been treated as a male but presumably it had marked female characters. Were there no marked male characters in any bone below 45 mm.? It will be seen that there is a remarkable dip in the total material at 46 mm. which corresponds to Dwight's division. In material measured six years ago in the Biometric Laboratory, where every bone in a relatively large series was measured, no such dip occurs and there is in those data no justification for Dwight's method of sexing‡. The group of 29 ♂ bones at 47 mm. and the sudden cut off at 45 mm. seems to condemn this method of sexing, at any rate from the statistical standpoint.

---

* *Journal of Anatomy and Physiology*, Vol. XLVIII. pp. 238—267.

† *American Journal of Anatomy*, Vol. IV. p. 19.

‡ This material has been statistically reduced and will shortly be published.

Without arguing this point out here, we may illustrate the use of the Table (p. 214) of ψ's by taking two of Parsons' frequency distributions for females; we will cut them off at the points suggested, and then investigate the total populations of females which result. Our author pools for these distributions right and left bones.

Taking the diameter of "head of femur" for the females, we have

Diameter in mm. ...	*36*	*37*	*38*	*39*	*40*	*41*	*42*	*43*	*44*
Frequency	1	1	—	3	8	14	12	18	12

These are exactly the bones the Dwight process gives as female. We find

$$\Sigma^2 = 2{\cdot}8851,$$

$$d = 2{\cdot}6159 \quad \text{(measured from } 44{\cdot}5\text{)}.$$

Hence $$\psi_1 = \Sigma^2/d^2 = {\cdot}4216.$$

Whence by interpolation from the table

$$h' = {\cdot}782, \quad \psi_2 = {\cdot}864, \quad \psi_3 = 1{\cdot}278,$$

leading to $$\sigma = 2{\cdot}260, \qquad h = 1{\cdot}767,$$

and $$\text{Mean} = 42{\cdot}73 \text{ mm.}, \quad N = 88{\cdot}2.$$

Parsons gives for R. femur, Mean = 43,

L. femur, Mean = 42,

and the total number of bones dealt with $55 + 48 = 103$ (Tables, *loc. cit.* pp. 249—251). In his frequency distribution (p. 256) he only records 85 female bones, which give a mean of 42·54 and a standard deviation of 2·078 mm. These values are clearly not widely divergent from those we have found above by supposing all bones under 45 to be female.

To test the matter further the 105* female bones of which the head was measured by Parsons were taken out. They provide the distribution:

Diameter in mm. ...	*36*	*37*	*38*	*39*	*40*	*41*	*42*	*43*	*44*	*45*	*46*	*47*	*48*
Frequency	1	1	—	4	10	18	16	24	13	14	3	—	1

These give $$\left.\begin{array}{r}\text{Mean} = 42{\cdot}47\\ \text{S.D.} = 1{\cdot}996\end{array}\right\} N = 105.$$

---

* It is not possible to say whether he has omitted two queried measurements. He has not omitted bones he queries in breadth of lower articulation.

Cutting off all bones over 45·5 we find

$$\Sigma^2 = 2{\cdot}6392, \quad d = 2{\cdot}6264,$$

leading to

$$\psi_1 = {\cdot}3826.$$

Hence

$$h' = {\cdot}984, \quad \psi_2 = {\cdot}783 \text{ and } \psi_3 = 1{\cdot}195.$$

These provide for the non-truncated population,

$$\text{Mean } 42{\cdot}48, \quad \text{S.D.} = 2{\cdot}056, \quad N = 104,$$

which are in still better agreement with Parsons' constants for the 105 bones than the constants for the 85 bones were for their series. It would appear therefore that, if we suppose all bones under 45 female and use our Tables, we get results in reasonable accordance with Parsons', and possibly by a theoretically more justifiable method than endeavouring to sex the bones above 44 and below 48 from other characters.

We have considered from the same aspect the character breadth of lower articular end of femur. Parsons' distribution of 89 female femora is as follows (p. 257):

Breadth in mm. ...	*60*	*61*	*62*	*63*	*64*	*65*	*66*	*67*	*68*	*69*	*70*	*71*	*72*	*73*	*74*	*75*
Frequency ...	1	—	—	5	6	8	6	17	13	13	8	6	3	1	1	1

In this Table he has only one bone in excess of the numbers on which he bases his means on pp. 250—1. If we truncate at 69·5, i.e. reject all bones over 69 mm., we find

$$\Sigma^2 = 3{\cdot}9803, \quad d = 2{\cdot}9058,$$

and

$$\psi_1 = {\cdot}4714.$$

Hence we deduce

$$h' = {\cdot}5295, \quad \psi_2 = {\cdot}977, \quad \psi_3 = 1{\cdot}426.$$

These lead to

$$h = 1{\cdot}503, \quad \sigma = 2{\cdot}839, \quad N = 98{\cdot}4, \quad \text{Mean} = 68{\cdot}00 \text{ mm.}$$

The actual values given by Parsons' distribution above are

$$\sigma = 2{\cdot}571, \quad N = 89, \quad \text{Mean} = 67{\cdot}54 \text{ mm.}$$

Thus the agreement is not nearly so good as for the diameter of the head of the femur, being about 10 °/₀ wrong in σ and N. It should give as good a result if the method were quite satisfactory, for the bones have been sexed by the diameter of the head, and the limit 44 mm. for diameter of the head corresponds fairly closely to 69 mm. for the breadth of lower articulation.

As this paper is not intended as a discussion of Parsons' data, to which we hope again to return, we will only deal with one more illustration of the use of the Table. We take out from his Tables, pp. 244—248, the diameter of head of femur for 174 male bones.

Diameter of Head in mm. ...	*45*	*46*	*47*	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*
Frequency	9	10	33	17	38	20	18	12	6	8	3

The constants of this distribution are

$$\text{Mean} = 49{\cdot}14, \quad \sigma = 2{\cdot}377, \quad N = 174.$$

Truncating at 47·5 we have

$$\Sigma^2 = 3{\cdot}4341, \quad d = 2{\cdot}7869,$$

whence $\psi_1 = {\cdot}4422$, and from the Table

$$h' = {\cdot}679, \quad \psi_2 = {\cdot}908, \quad \psi_3 = 1{\cdot}331.$$

These lead to $$h = 1{\cdot}7182$$

and $$\text{Mean} = 49{\cdot}22, \quad \sigma = 2{\cdot}530, \quad N = 162,$$

i.e. to a "tail" of 40 not one of 52 below 47·5. Actually this tail distributes itself as follows:

	Under 45	*45*	*46*	*47*
Gaussian tail	5	6	12	17
Against Parsons'	0	9	10	33

This confirms our previously expressed view that probably a considerable number of the bones classed as 47 mm. are really female femora, and that the male distribution runs considerably beyond 45 mm. into the range treated as purely female.

Finally let us try the result of pooling male and female bones and breaking up the composite frequency by the method of *Phil. Trans.* Vol. 185 A, p. 84.

We have now 279 bones distributed as follows:

Diameter of Head in mm.	*36*	*37*	*38*	*39*	*40*	*41*	*42*	*43*	*44*	*45*	*46*	*47*	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*	Total
Frequency ...	1	1	—	4	10	18	16	24	13	23	13	33	18	38	20	18	12	6	8	3	279

The constants are $\text{Mean} = 46{\cdot}63, \quad \text{S.D.} = 3{\cdot}93,$

$$\mu_2 = 15{\cdot}4040, \quad \mu_3 = -6{\cdot}7791,$$

$$\mu_4 = 541{\cdot}5162, \quad \mu_5 = -530{\cdot}5339.$$

The nonic is

$$q_2^9 - 5{\cdot}9618q_2^7 + {\cdot}0689q_2^6 + 9{\cdot}8357q_2^5 - 3{\cdot}4275q_2^4 - 8{\cdot}2041q_2^3 - {\cdot}3020q_2^2 + {\cdot}0144q_2 - {\cdot}0097 = 0,$$

giving the root $q_2 = -{\cdot}934$ and $p_2 = -9{\cdot}34$, and ultimately the two components:

	Male	*Female*
Mean	49·83	43·72
Population	133·25	145·75
Standard Deviation . . .	2·231	2·662
Max. Ordinate	23·83	21·84

While the means agree roughly with those obtained by Parsons' sexing (49 and 43), we see that this analysis much more nearly equalises the number of male and female bones, and indeed makes the female population rather larger than the male, while Parsons has 79 °/₀ more males. The "truncated tail" method would probably give results in better accordance with the present had we not truncated at the quite arbitrary Dwight-Parsons' divisions.

These examples may suffice to illustrate the application of the Tables to anthropometric measurements on man, where we can feel fairly confident that the material, if sufficient in quantity, would be adequately described by a Gaussian or normal distribution. Such cases may arise when material for the two sexes, or for two races, is commingled and we can be fairly certain that one or other or both "tails" of the material present homogeneous parts of the mixture.

Another illustration drawn from Galton's data for American trotters will be found in the *Tables for Statisticians*, p. xxvi. The chief weakness of the method, besides the assumption of the Gaussian, often quite legitimate, is the absence as yet of the values of the probable errors, which values must be very considerable for slender material such as that used above.

See following page for Table of Gaussian "Tail" Functions.

Table of Gaussian "Tail" Functions, "Tail" larger than "Body."

h'	ψ_1	$(-)\,\Delta\psi_1$	ψ_2	$(-)\,\Delta\psi_2$	ψ_3	$(-)\,\Delta\psi_3$	h'
·00	·571	·002	1·253	·006	2·000	·016	*·00*
·01	·569	·002	1·248	·006	1·984	·016	*·01*
·02	·567	·002	1·242	·006	1·969	·015	*·02*
·03	·565	·002	1·236	·006	1·953	·015	*·03*
·04	·564	·002	1·231	·006	1·938	·015	*·04*
·05	·562	·002	1·225	·006	1·923	·015	*·05*
·06	·560	·002	1·219	·006	1·909	·014	*·06*
·07	·558	·002	1·214	·006	1·894	·014	*·07*
·08	·557	·002	1·208	·006	1·880	·014	*·08*
·09	·555	·002	1·203	·005	1·866	·014	*·09*
·1	·553	·018	1·197	·054	1·852	·126	*·1*
·2	·535	·018	1·143	·053	1·726	·108	*·2*
·3	·516	·019	1·090	·051	1·618	·093	*·3*
·4	·497	·019	1·040	·049	1·526	·080	*·4*
·5	·477	·019	·991	·047	1·446	·068	*·5*
·6	·458	·020	·944	·045	1·378	·059	*·6*
·7	·438	·020	·899	·043	1·319	·050	*·7*
·8	·419	·020	·857	·041	1·269	·043	*·8*
·9	·399	·019	·816	·039	1·226	·037	*·9*
1·0	·380	·019	·777	·037	1·189	·032	*1·0*
1·1	·361	·019	·740	·035	1·157	·027	*1·1*
1·2	·342	·018	·704	·033	1·130	·023	*1·2*
1·3	·323	·018	·671	·033	1·107	·019	*1·3*
1·4	·305	·017	·640	·030	1·088	·016	*1·4*
1·5	·288	·017	·610	·028	1·072	·014	*1·5*
1·6	·271	·016	·582	·026	1·058	·011	*1·6*
1·7	·254	·016	·556	·025	1·047	·009	*1·7*
1·8	·239	·015	·531	·023	1·037	·008	*1·8*
1·9	·224	·014	·508	·022	1·030	·006	*1·9*
2·0	·210	·013	·487	·020	1·023	·005	*2·0*
2·1	·197	·013	·466	·019	1·018	·004	*2·1*
2·2	·184	·012	·447	·018	1·014	·003	*2·2*
2·3	·172	·011	·429	·017	1·011	·003	*2·3*
2·4	·161	·010	·413	·016	1·008	·002	*2·4*
2·5	·151	·010	·397	·015	1·006	·002	*2·5*
2·6	·141	·009	·383	·014	1·005	·001	*2·6*
2·7	·132	·008	·369	·013	1·003	·001	*2·7*
2·8	·124	·008	·356	·012	1·003	·001	*2·8*
2·9	·116	·007	·344	·011	1·002	·001	*2·9*
3·0	·109		·333		1·001		*3·0*

See *Tables for Statisticians and Biometricians*, Introduction, p. xxvii.

CONTRIBUTION TO A STATISTICAL STUDY OF THE CRUCIFERÆ.

VARIATION IN THE FLOWERS OF *LEPIDIUM DRABA* LINNÆUS.

By JAMES J. SIMPSON, M.A., D.Sc.

CONTENTS.

I. Introductory[*].

In the summer of 1905 Professor J. W. H. Trail drew my attention to an extraordinary example of variation which occurred in the several organs of the flowers of *Lepidium Draba* Linnæus. At that time I examined in detail 1832 individual flowers taken from a single plant growing in a piece of uncultivated

---

[*] I am pleased to have this opportunity of expressing my great indebtedness to Dr J. F. Tocher for invaluable assistance in the biometric part of this paper. The correlation and other constants were calculated in his laboratory, and without his assistance the publication of this paper would have been greatly delayed. I must also thank Professor Karl Pearson, in whose department in University College, London, the statistical study was originally undertaken, for reviewing this paper for publication and also for much kindly criticism and advice. To Professor Trail my thanks are also due for many botanical hints.

ground in his garden, and the results of these observations form the basis of the present contribution to the study of the variation in the Cruciferæ.

Botanical problems, which have been hitherto attacked from the biometric standpoint, have been comparatively easily handled, because the material has been more or less homogeneous in character. For example, variations in the *number* of sepals of *Anemone nemorosa*[*] or in the number of ray-florets of *Chrysanthemum leucanthemum*[†], and the consequent distribution of these are capable of direct treatment by Pearson's well-known method of fitting frequency curves.

The only work comparable to the one in hand occurs in *Biometrika,* Vol. II. p. 145 (Variation and Correlation in the Lesser Celandine), but in this case the numbers of members in the calyx, corolla and andrœcium have been examined as a basis for a study of homotypic correlation and in this flower each of these organs consists of a single constituent with numerous members.

The problems studied in this paper, however, are more complex inasmuch as they deal *not* with *one* organ of the flower but with all the organs, their constituents and members both separately and collectively.

It is also, I believe, the first biometric work of its kind on a cruciferous flower and embodies a study of chorisis, that is, "the splitting up or division of one or more components of a flower into two or more equal or unequal parts"—a factor which is supposed to have been of the utmost importance in the evolution of the natural order—Cruciferæ. A complete discussion of this phenomenon is reserved until the flower is studied in detail.

It would be well here to emphasise the fact that the flowers examined for this study were not taken from different plants but, on the contrary, were obtained from several inflorescences growing on stems which had arisen from buds on the roots of a single parent plant. This mode of reproduction is rather unusual, but, in the present instance, is of particular interest inasmuch as it gives greater homogeneity to the material.

The parts of the flower which have been considered are (*a*) the perianth, which consists of (1) the calyx and (2) the corolla, (*b*) the andrœcium and (*c*) the gynæcium.

The functional differentiation of these organs is of great importance in the interpretation of results so that it might be well to recall the particular rôles which these play in plant economics.

The gynæcium and the andrœcium are respectively the female and male organs of reproduction and consist of carpels and stamens, while the perianth forms a protective covering for these delicate structures. The calyx or outer organ of

---

[*] Yule, *Biometrika*, Vol. I. p. 307.

[†] *Biometrika*, Vol. II. p. 309 et seq.

the perianth is concerned solely in the protection of the flower in the bud, but the corolla, in the open flower, also serves, along with the honey-secreting sacs at the base of the stamens, as an attraction for insects.

The *characters* which have been taken as a basis for this study are *numerical,* e.g. the number of petals in the corolla, but no measurable characters, e.g. the length and breadth of the petals, have been considered, although, as will be pointed out later in connection with possible future studies in this flower, these characters might also with advantage be taken.

The Cruciferæ, as an order, are usually regarded by botanists as being very definite in type and no observations have been recorded to show to what extent, if any, deviation from the recognised botanical floral formula exists, so that the main object of this paper was to determine the frequency of the variability of the parts of the various organs and constituents, and also the degrees of correlation existing between the organs themselves.

The mode of observation is worthy of remark, however, as it might well be argued that if the flowers used for examination were fully "blown" deficiency in the number of parts *might* be due to post-developmental fracture, but in all the cases here recorded the observations were made on flowers in bud or only half open so that the influence of wind or other external agency is altogether discounted. The material was also examined microscopically in all cases so that there should be no possible doubt as to the exact origin of any member. The importance of this will be seen in the details of the analysis.

II. BOTANICAL.

1. *Specific characters.*

The generic and specific characters of *Lepidium Draba* may be obtained in any complete systematic botanical work so that it is unnecessary to repeat them here, but a few notes bearing especially on the study in hand may be of value.

It is a perennial about a foot in height and is covered by a minute down from which its popular name, the *hoary cress,* is derived. The inflorescence is a raceme not much lengthened and so forms a broad, almost flat, corymb-like termination. The individual flowers are small, white and numerous. The constituents of calyx, namely the *sepals,* are green; they are short, nearly equal and bear no pouch at the base. The *petals* are small and white; they are equal in size, obovate, undivided and generally stalked. The *stamens* are six in number; the filament is simple, i.e. it bears no appendages, and is shorter than the petals; the anther consists of two roundish lobes. The *pods* are "broader than long"; they are compressed laterally at right angles to the narrow partition. The thick valves are boat-shaped and sharply keeled but not winged; each valve contains a single seed.

2. *Morphology of the flower.*

The typical flower consists of six whorls, made up in the following manner:

(*a*) Calyx 2.
(*b*) Corolla 1.
(*c*) Andrœcium 2.
(*d*) Gynæcium 1. (Plate I, fig. 1.)

(*a*) *Calyx.* This organ is composed of two whorls each consisting of two sepals. The outer pair arise at one level on opposite sides of the flower and are inserted on a slightly lower plane than the inner pair; they are parallel to the plane of compression of the gynæcium. The inner pair are also situated opposite one another but in a plane perpendicular to that of the outer pair; they are thus at right angles to the plane of compression of the gynæcium. These whorls are denoted on Plate I, fig. 1, by the Roman numerals I and II respectively.

(*b*) *Corolla.* This organ consists of four petals all inserted at *one level* and alternating with the position of the sepals; they thus constitute a single whorl. (See III, Plate I, fig. 1.)

(*c*) *Andrœcium.* Six stamens form the andrœcium; they arise at two different levels and thus constitute two separate whorls. The outer whorl, which is lower down, consists of two stamens which are shorter than the others and correspond in position to the inner sepals. The inner whorl consists of four stamens, arranged in pairs which correspond in position to the outer sepals. (A reference to the figure (Plate I, fig. 1) in which the two whorls are marked IV and V respectively will make this clear.)

(*d*) *Gynæcium.* This organ consists of two carpels forming the sixth or innermost whorl. (See VI, Plate I, fig. 1.)

It will be seen from the foregoing description that the order of the six whorls here detailed is that in which they would be found were we to strip the flower of its components at the different levels consecutively from below upwards. It is also the order in which we would find them, passing from the outside to the centre, were we to cut a transverse section through the flower.

Another point, however, which is not so obvious but one which has special interest in our study, is the fact that this is also the order in time of development.

The actual sequence in which these constituents of the flower appear in the bud is therefore:

I. Outer whorl of Calyx (Sepals).
II. Inner whorl of Calyx (Sepals).
III. Corolla (Petals).
IV. Outer whorl of Andrœcium (Stamens).
V. Inner whorl of Andrœcium (Stamens).
VI. Gynæcium (Carpels).

3. *Conception of Chorisis.*

Chorisis or reduplication is generally looked upon by botanists as a means of multiplication of the parts of a flower. It consists in the division or splitting of an organ in the course of its development by which two or more organs are produced in place of one. Chorisis may take place in two ways:

(1) *transversely*—when the increased parts are placed one before the other, that is, the resulting components are on the same radius; this is known as *vertical*, *parallel* or *transverse* chorisis;

(2) *collaterally*—when the increased parts stand side by side, that is, on the same circumference.

Transverse chorisis is supposed to be of frequent occurrence; thus the pistils of *Lychnis* and many other caryophyllaceous plants exhibit a small scale on the inner surface at the point where the limb of the petal is united to the claw. The formation of these scales is supposed by many to be due to the chorisis or unlining of an inner portion of the petal from the outer.

Collateral chorisis is seen in different natural orders. In *Strephanthus*, in place of two stamens there is sometimes a single filament forked at the top and each division bears an anther. This is usually supposed to be due to collateral chorisis arrested in its progress.

The flowers of the Fumitory are also generally considered to afford another example of this type of chorisis. In these we have two sepals, four petals in two rows and six stamens, two of which are perfect and four more or less imperfect. The latter are said to arise by collateral chorisis, one stamen being divided into three parts.

Collateral chorisis may be compared, according to Bentley, to a compound leaf which is composed of two or more distinct and similar parts.

Let us now consider chorisis in its bearing to the flower under consideration. In the description of the morphology of the flower we noted that in the inner whorl of the andrœcium there were four stamens arranged in pairs while in the outer whorl there were only two stamens situated singly. Various opinions have from time to time been advanced to explain this anomalous structure so that it might be well to briefly review these. Of the andrœcium of the Cruciferæ Oliver says: "The two pairs of long stamens are generally thought to be due to chorisis or the division in the course of development of single antero-posterior stamens. Others have thought that the six glands represent abortive stamens and that these with the six stamens make up a normal series of twelve in three whorls."

De Candolle held the view that the stamens formed a single, originally tetramerous whorl alternating with the petals in which the median members, i.e. the anterior and posterior, were cleft (chorised) in two. Since however the lateral stamens are inserted lower down than the median stamens and are also, as already pointed out, formed earlier in the bud, this view is clearly untenable. Two whorls must be taken into consideration owing to the difference in the levels of insertion, the single stamens being lower down. Kunth, Wydler, Chatin and others regard these two whorls as typically four-membered (tetramerous), those of the outer whorl corresponding in position to the sepals, those of the inner whorl corresponding in position to the petals. To arrive at a typical cruciferous flower from this, two stamens in the outer whorl abort, while the individuals of the two pairs of the inner whorl come together. (Plate I, fig. 3.)

Others (Krause, Wretschko and Duchartre) regard the outer whorl as typically dimerous (i.e. with two constituents) and the inner whorl as typically tetramerous (i.e. four-membered).

The more modern view, however, regards both whorls as dimerous but the inner one chorised collaterally thus giving the typical cruciferous flower.

The reasons put forward to support this theory are as follows:

(1) The upper long stamens are usually *paired* in the median line, also sometimes *coherent*. Further, in place of one or both of the pairs, there occurs sometimes a single stamen—a hint at reversion, or one or both pairs may be replaced by three or more—a suggestion of further chorisis.

(2) In the earliest visible stage of development in the bud it may be seen that each pair of stamens arises from a single wart-like projection and that division is therefore a secondary result. This is not very easily demonstrable in the Cruciferæ but is more evident in a closely allied family, the Capparidaceæ.

Since the present study includes numerical variation in the different constituents and positions of the andrœcium it will be interesting to note to what extent any one of these theories is borne out by the variations in this flower.

4. *Orientation of the flower.*

Having defined the positions of the various stamens relative to one another, in what is usually regarded as a normal cruciferous flower, let us now consider the different possibilities when the flower is abnormal.

Suppose that one of the pairs of stamens of the inner whorl is represented by a single stamen, that is, suppose that chorisis had not taken place. Now with regard to the peduncle of the inflorescence this stamen might be placed in two diametrically opposite positions, namely (1) it might be adjacent to the peduncle (Plate I, fig. 4) or (2) it might be on the distal half of the flower with reference to the peduncle (Plate I, fig. 5).

Two questions now arise, (1) do non-chorised stamens occur as frequently as chorised stamens on the side of the flower next to the peduncle? or (2) do either of these occur with greater frequency in this adjacent position?

According to which of these questions is answered in the affirmative must we conclude whether there is any connection or correlation between the proximity of the chorised stamens to the peduncle and chorisis. The former would suggest no correlation, whereas the degree of correlation hinted at by the latter would depend on the frequency of the occurrence.

We have so far considered only two possible positions, viz. a non-chorised stamen adjacent to the peduncle, i.e. in the proximal half of the flower, and a non-chorised stamen opposite to the peduncle (i.e. in the distal half of the flower with reference to the peduncle), but the question naturally arises "Are these the only two possible relative positions which might occur?" Might the petiole not twist so as to bring the hypothetic non-chorised stamen into any position varying from 0° to 180° with reference to the original plane?

Let us illustrate this by means of the Figure 6, Plate I.

Taking the position of the peduncle as our fixed point the non-chorised stamen might occupy the "adjacent" position a or the "opposite" position a1. A rotation of the petiole, however, might cause this stamen to occupy any of the positions marked a2, a3 or a4 or even any intermediate position between a and a1 on *either side* of the vertical plane *A—B*, in the horizontal plane a, a4, a2, a3, a1.

In a study of the variations in this flower, this is precisely what was found to occur, i.e. the distribution was equal round a fixed point so that we are unable to say whether there is any connection between the proximity of the non-chorised stamen to the peduncle and chorisis or not.

But the full bearing of this consideration does not end here. The orientation of the flower is of practical importance in fixing a basis on which to establish a grouping of the different variations. Any analysis of the data is impossible unless some definite part of the flower be agreed upon as a starting point.

Now we have seen that the position of the peduncle with respect to any definite stamen does not require to be taken into consideration. Consequently we may take either of the two stamens of the *outer* whorl, which correspond in position to the outer sepals and which are "normally" non-chorised, as our fixed point and call it 1; the stamen opposite, i.e. in the same whorl, we shall call 2; the chorised pair of the *inner* whorl to the left (or in the floral diagram above) may be termed 3 and 4; while the corresponding pair to the right (or in the floral diagram below) would thus be 5 and 6 (Plate I, fig. 7).

Where variations occur in any of these stamens we shall hereafter refer to those as occurring in "position" 1, 2, 3, 4 and 5, 6 respectively.

On this basis of symmetry, it will simplify matters considerably if we regard as 1, in flowers in which either of the two outer stamens is modified, that one which still maintains its original character while, on the other hand, if both are modified, that one which retains the greatest approximation to normality, e.g. if one be chorised while the other is not, the latter would be in position 1; or if one were chorised while the other was only partially chorised* the latter would again be in position 1.

Following on this it is at once seen that where both are normal or where both are equally abnormal it makes absolutely no difference which position we choose as 1.

III. EXAMINATION OF THE DATA.

1. *Classification.*

Considerable difficulty was experienced in classifying the variations owing to these occurring in so many different forms yet with so few characteristics in common as to warrant their inclusion in definite classes.

The total number of flowers examined was 1832, of which 1062 had the accepted normal structure (see page 218). The remaining 770 showed variation in different degrees of advance or regression, i.e. there was an excess or deficiency in the number and structure of the members of the various organs. Thus we see that there was a deviation from the accepted normal structure in over 42 per cent. of the individuals examined.

The perianth has been selected as a basis for classification and Table A shows the sub-divisions which have been adopted. Amongst those flowers in which the

TABLE A.

	Number of Variations	Number in Group	Number in Sub-class	Number in Class	Variations in the Class
Class I. Perianth normal	—	—	—	1687	—
Sub-Class A. Gynæcium normal	—	—	1680	—	—
Group (*a*). Andrœcium normal	1	1062	—	—	—
Group (*b*). Andrœcium abnormal	57	618	—	—	—
Sub-Class B. Gynæcium abnormal	—	—	7	—	—
Group (*a*). Gynæcium one carpel	2	4	—	—	—
Group (*b*). Gynæcium reduplicated ...	2	3	—	—	62
Class II. Perianth abnormal	—	—	—	115	—
Sub-Class A. Calyx normal, corolla abnormal	—	—	55	—	—
Group (*a*). Gynæcium normal	11	54	—	—	—
Group (*b*). Gynæcium a single carpel ...	1	1	—	—	—
Sub-Class B. Both calyx and corolla abnormal	—	—	60	—	—
Group (*a*). Gynæcium normal	11	46	—	—	—
Group (*b*). Gynæcium a single carpel ...	6	14	—	—	29
Totals	91	1802	1802	1802	91

---

* For the present we use the terms "chorised" and "chorisis" in the sense of the definition already given.

perianth was normal there were no fewer than 62 different types of variation, and amongst those in which the perianth showed a departure from the accepted normal structure there were 29 types of variation. Thus of 1802 flowers examined, 1062 had the typical cruciferous structure, 625 had the perianth normal but the andrœcium and gynæcium modified in 62 different ways and 115 had all three organs modified in 29 different types of variation.

The remaining 30 individuals are not capable of classification under the foregoing scheme but have been grouped into three classes as shown in Table B.

TABLE B.

	Number of Variations	Number of individuals in the Class
Class III. Reduplication of parts but flowers not separate ...	10	11
Class IV. Reduplication of parts with flowers separate ...	6	17
Class V. Part of a flower replaced by a flower	2	2
Totals	18	30

Altogether, therefore, there are five separate classes which give a total of 109 different modes of variation.

2. *Analysis.*

In the further reduction of the data it is essential that we consider the variations in the stamens, and for this purpose we must naturally commence with Class I, Sub-class A.

To avoid describing each of these in detail, it is necessary to have recourse to a graphic method of representation. Several such methods suggested themselves and although none are ideal we have chosen one which may help to give a true impression of the various modifications assumed by the andrœcium. We shall also give a few examples by another method which might have been adopted but which seems to us to be even more complicated.

Let us, in the first place, consider in what directions abnormalities have occurred. A typical stamen consists of two parts, (1) the filament and (2) the anther.

(1) Filament. This may be of its normal length or less than its normal length or altogether absent.

(2) Anther. This may be present or absent.

But other complications arise. As already explained, in the accepted typical cruciferous flower, chorisis has taken place in positions 3 . 4 and 5 . 6 so as to give rise to two stamens in each of these positions. Now, we find that, in certain flowers chorisis has only partially taken place and in others it has not occurred at all so that we have thus another three possibilities to consider.

In describing the andrœcium, therefore, we must (1) define the position of each stamen to which we refer, (2) state the nature of the filament, (3) note the presence or absence of the anther and (4) emphasise the nature of the chorisis.

Let us use the following symbols 1, $\frac{1}{2}$ and 0.

$$1 \text{ with reference to } \begin{cases} \text{Filament, indicates that it is present and complete.} \\ \text{Chorisis, indicates that it is total or complete.} \\ \text{Anther, indicates that it is present.} \end{cases}$$

$$\tfrac{1}{2} \text{ with reference to } \begin{cases} \text{Filament, indicates that it is only half-length.} \\ \text{Chorisis, indicates that it is only partial.} \end{cases}$$

$$0 \text{ with reference to } \begin{cases} \text{Filament, indicates that it is absent.} \\ \text{Chorisis, indicates that it has not taken place.} \\ \text{Anther, indicates that it is absent.} \end{cases}$$

We have already fixed upon our nomenclature for the various positions; these are 1; 2; 3.4; and 5.6. To avoid descriptions and at the same time give a graphic representation of the floral formula of the andrœcium the following system might be adopted:

(1) Place the whole floral formula within square brackets thus [ ].

(2) Place positions 1; 2; 3.4; and 5.6 within curled brackets thus { }; and

(3) Place individuals, i.e. 1, 2, 3, 4, 5 and 6, within rounded brackets thus ( ).

Expanding this with reference to a normal flower we would have for the andrœcium only

$$[\{2\}\,\{2\}\,\{2\}],$$

or still further in the order of Filament, Chorisis, Anther, Stamen

$$[\{(1\,.\,0\,.\,1)(1\,.\,0\,.\,1)\}\,\{(1\,.\,1\,.\,1)(1\,.\,1\,.\,1)\}\,\{(1\,.\,1\,.\,1)(1\,.\,1\,.\,1)\}].$$

Or, taking an actual example from our data:

Stamen number 1 is normal and complete, and there is no chorisis; stamen number 2 has a filament only half-length but the anther is present and complete, and there is no chorisis; stamen number 3 is normal and complete; stamen number 4 is only half-length but with a complete anther—chorisis between 3 and 4 is complete; stamens 5 and 6 are only half the normal length but have complete anthers—chorisis between 5 and 6 is complete.

This would be represented thus:

$$[\{(1\,.\,0\,.\,1)(\tfrac{1}{2}\,.\,0\,.\,1)\}\,\{(1\,.\,1\,.\,1)(\tfrac{1}{2}\,.\,1\,.\,1)\}\,\{(\tfrac{1}{2}\,.\,1\,.\,1)(\tfrac{1}{2}\,.\,1\,.\,1)\}].$$

Another graphic method and the one which we have adopted is as follows. Each flower is represented in a table similar to the following:

Frequency	Number of Diagram	Stamen	Filament	Chorisis	Anther

The first vertical column gives the frequency of the variation or the number of individuals examined with this structure. The second vertical column gives the number of the corresponding diagram in the plates. The third vertical column gives the individual stamens in the positions already defined while the other three columns denote the various factors to be considered. The different possibilities of variation in these may be shown by the symbols 1, ½ and 0 as already defined. It should be noted, however, that in positions 1 and 2 a dash (—) will be placed in the chorisis column to indicate that these are typically non-chorised stamens and that absence of chorisis does not therefore indicate abnormality. Representing the same example as before, by this method, we would have:

Frequency	Number of Diagram	Stamen	Filament	Chorisis	Anther
		1	1	—	1
		2	½	—	1
		3	1	1	1
		4	½	1	1
		5	½	1	1
		6	½	1	1

The following table shows graphically the types of variations illustrated in Figs. I—LVIII, i.e. Class 1, Sub-class A, flowers in which the perianth and gynæcium are both normal.

It will be seen that in the flowers illustrated in Figs. XLVIII—LVIII another complication has crept in. Stamens 3, 4, 5 and 6 have themselves sometimes undergone partial or total secondary chorisis. In the tables, therefore, by subdividing the squares containing the details we can thus adhere to our initial nomenclature. Let us take the three most difficult examples to illustrate this.

(1) Fig. XLVIII. The division corresponding to stamen 3 is sub-divided. This would indicate that in this position there were actually two stamens. The nature of each of these individual stamens is, as before, given in the sub-divisions. In

TABLE C.

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
1062	I	1	1	—	1	2	II	1	1	—	1	130	III	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	1	1			3	1	1	1			3	1	1	1
		4	1	1	1			4	1	1	1			4	1	1	1
		5	1	1	1			5	1	½	1			5	1	0	1
		6	1	1	1			6	1	½	1			6	1	0	1
38	IV	1	1	—	1	1	V	1	1	—	1	8	VI	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	1	1			3	1	1	1			3	1	1	1
		4	1	1	1			4	1	1	1			4	1	1	1
		5	1	1	1			5	1	1	1			5	½	1	1
		6	½	1	1			6	½	1	0			6	½	1	1
6	VII	1	1	—	1	227	VIII	1	1	—	1	7	IX	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	1	1			3	1	1	1			3	1	1	1
		4	1	1	1			4	1	1	1			4	1	1	1
		5	½	0	1			5	1	0	1			5	½	0	1
		6	½	0	1			6	0	0	0			6	0	0	0
5	X	1	1	—	1	1	XI	1	1	—	1	8	XII	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	1	1			3	1	1	1			3	1	1	1
		4	½	1	1			4	½	1	1			4	½	1	1
		5	1	1	1			5	½	1	1			5	1	0	1
		6	½	1	1			6	½	1	1			6	1	0	1
8	XIII	1	1	—	1	3	XIV	1	1	—	1	3	XV	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	1	1			3	1	1	1			3	1	0	1
		4	½	1	1			4	½	1	1			4	1	0	1
		5	1	0	1			5	½	0	1			5	1	0	1
		6	0	0	0			6	0	0	0			6	1	0	1
1	XVI	1	1	—	1	1	XVII	1	1	—	1	4	XVIII	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	0	1			3	1	0	1			3	1	0	1
		4	1	0	1			4	1	0	1			4	1	0	1
		5	½	1	1			5	½	0	1			5	1	0	1
		6	½	1	1			6	½	0	1			6	0	0	0
2	XIX	1	1	—	1	21	XX	1	1	—	1	2	XXI	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	0	1			3	1	0	1			3	1	0	1
		4	0	0	0			4	0	0	0			4	0	0	0
		5	½	1	1			5	1	0	1			5	½	0	1
		6	½	1	1			6	0	0	0			6	0	0	0
2	XXII	1	1	—	1	15	XXIII	1	1	—	1	1	XXIV	1	1	—	1
		2	1	—	1			2	½	—	1			2	½	—	0
		3	½	1	1			3	1	1	1			3	1	1	1
		4	½	1	1			4	1	1	1			4	1	1	1
		5	½	0	1			5	1	1	1			5	1	1	1
		6	½	0	1			6	1	1	1			6	1	1	1

TABLE C—(*continued*).

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
7	XXV	1 2 3 4 5 6	1 ½ 1 1 1 ½	— — 1 1 1 1	1 1 1 1 1 1	2	XXVI	1 2 3 4 5 6	1 ½ 1 1 ½ ½	— — 1 1 1 1	1 1 1 1 1 1	3	XXVII	1 2 3 4 5 6	1 ½ 1 1 1 1	— — 1 1 0 0	1 1 1 1 1 1
11	XXVIII	1 2 3 4 5 6	1 ½ 1 1 1 0	— — 1 1 0 0	1 1 1 1 1 0	3	XXIX	1 2 3 4 5 6	1 ½ 1 1 ½ 0	— — 1 1 0 0	1 1 1 1 1 0	1	XXX	1 2 3 4 5 6	1 ½ 1 ½ 1 ½	— — 1 1 1 1	1 1 1 1 1 1
3	XXXI	1 2 3 4 5 6	1 ½ 1 ½ 1 1	— — 1 1 0 0	1 1 1 1 1 1	1	XXXII	1 2 3 4 5 6	1 ½ 1 ½ ½ ½	— — 1 1 1 1	1 1 1 1 1 1	3	XXXIII	1 2 3 4 5 6	1 ½ 1 ½ 1 0	— — 1 1 0 0	1 1 1 1 1 0
3	XXXIV	1 2 3 4 5 6	1 ½ 1 ½ ½ 0	— — 1 1 0 0	1 1 1 1 1 0	1	XXXV	1 2 3 4 5 6	1 ½ ½ 0 ½ 0	— — 0 0 0 0	1 1 1 0 1 0	9	XXXVI	1 2 3 4 5 6	½ ½ 1 1 1 1	— — 1 1 1 1	1 1 1 1 1 1
2	XXXVII	1 2 3 4 5 6	½ ½ 1 1 1 ½	— — 1 1 1 1	1 1 1 1 1 1	3	XXXVIII	1 2 3 4 5 6	½ ½ 1 ½ ½ ½	— — 1 1 1 1	1 1 1 1 1 1	2	XXXIX	1 2 3 4 5 6	½ ½ 1 ½ 1 0	— — 1 1 0 0	1 1 1 1 1 0
18	XL	1 2 3 4 5 6	1 0 1 1 1 1	— — 1 1 1 1	1 0 1 1 1 1	3	XLI	1 2 3 4 5 6	1 0 1 1 1 ½	— — 1 1 1 1	1 0 1 1 1 1	1	XLII	1 2 3 4 5 6	1 0 1 1 1 ½	— — 1 1 1 1	1 0 1 1 1 0
5	XLIII	1 2 3 4 5 6	1 0 1 1 ½ ½	— — 1 1 0 0	1 0 1 1 1 1	2	XLIV	1 2 3 4 5 6	1 0 1 ½ ½ ½	— — 1 1 1 1	1 0 1 1 1 1	2	XLV	1 2 3 4 5 6	½ 0 1 1 1 1	— — 1 1 1 1	1 0 1 1 1 1
1	XLVI	1 2 3 4 5 6	½ 0 1 1 1 0	— — 1 1 0 0	1 0 1 1 1 0	3	XLVII	1 2 3 4 5 6	½ 0 1 1 ½ 0	— — 1 1 0 0	1 0 1 1 1 0	3	XLVIII	1 2 3 4 5 6	1 1 1 1 1 1 1 1	— — 1 1 1 1 1 1	1 1 1 1 1 1 1 1

TABLE C—(*continued*).

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
8	XLIX	1	1	—	1	3	L	1	1	—	1	5	LI	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	1	1			3	1	1	1			3	1	1	1
		4	1	1	1			4	1	1	1			4	1	1	1
		5	1 1	1 1	1 1			5	1 ½	1 1	1 1			5	1 1	0 0	1 1
		6	1	1	1			6	1 1	0 0	1 1			6	1	1	1
2	LII	1	1	—	1	2	LIII	1	1	—	1	1	LIV	1	1	—	1
		2	1	—	1			2	1	—	1			2	1	—	1
		3	1	1	1			3	1	0	1			3	1	0	1
		4	½	1	1			4	1	0	1			4	1	0	1
		5	1 1	½ ½	1 1			5	1 1	1 1	1 1			5	1 1	0 0	1 1
		6	1 1	½ ½	1 1			6	1	1	1			6	1	0	1

TABLE C—(*continued*).

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
2	LV	1	1	—	1	4	LVI	1	1	—	1
		2	1	—	1			2	1	—	1
		3	1	0	1			3	1	0	1
		4	1	0	1			4	0	0	0
		5	1 1	0 0	1 1			5	1 1	1 1	1 1
		6	1	1	1			6	1	1	1
1	LVII	1	½	—	1	2	LVIII	1	1	—	1
		2	½	—	1			2	0	—	0
		3	1	1	1			3	1	1	1
		4	1	1	1			4	1	1	1
		5	1 1	1 1	1 1			5	1 1	0 0	1 1
		6	1	1	1			6	1	1	1

this example what really occurs is: There are two individual stamens each equal to the original length and bearing a complete anther and separated from one another by secondary chorisis.

(2) Fig. L. Divisions 5 and 6 are sub-divided to show that there has been secondary chorisis in both of these stamens. In the former, chorisis has been complete but has resulted in one being full-length and with a functioning anther while the other is only half-length with a functioning anther. In the latter, chorisis has not been complete inasmuch as only the anther has been chorised.

(3) Fig. LII. Divisions 5 and 6 are both sub-divided, consequently we may infer that both of these stamens have undergone some stage of chorisis. In the first column we see that all are full-length, in the third that all have functioning anthers, but the second tells us that chorisis has been only partial in each case. The symbol $\frac{1}{2}$ between 5 and 6 indicates that between these two chorisis has also been partial. Thus we conclude a state of affairs as follows: In position 5.6 (1) there arises a single filament which divides into two at some distance from the base and (2) that each of these again sub-divides and (3) that on the end of each of these four sub-filaments there arises a functioning anther. The others may be worked out in a similar manner but a reference to the diagrams will at once obviate any misrepresentation.

From the foregoing table and illustrations it is evident that further classification is possible but it would be well to point out here certain difficulties which arise. As an example let us consider such a case as (using our original terminology) that in which, in any of the positions (1; 2; 3.4; or 5.6), the stamens are represented thus (1.1.0) (0.0.0), thus ($\frac{1}{2}$.1.1) ($\frac{1}{2}$.1.1) or thus ($\frac{1}{2}$.0.1) ($\frac{1}{2}$.0.1). Which shall have precedence? If we are to consider these variations as deviations from the usually accepted normal cruciferous flower, then we may safely assume that that flower which has the greatest number of functioning parts in a certain position is less aberrant than one in which any or all of the parts are altogether wanting; while, on the other hand, if in a position in which chorisis normally takes place, we have defective groups like those in cases 2 and 3 cited above, in one of which chorisis has taken place but not in the other, we must consider *that* group in which chorisis has occurred as being the one less removed from normal. On this basis then the above examples would be placed in the following order with regard to normality:

(1) ($\frac{1}{2}$.1.1) ($\frac{1}{2}$.1.1); (2) ($\frac{1}{2}$.0.1) ($\frac{1}{2}$.0.1); (3) (1.1.0) (0.0.0).

Similarly for any of the others.

Consequently we are now in a position to classify the actual cases under observation.

So far we have considered only those flowers in which there was the typical number of stamens, with their manifold variations in size and structure, but now we must classify those in which secondary chorisis has given rise to more than the accepted number.

Let us take stamens 5 and 6 as our basis, i.e. those individuals in which position 5.6 is occupied by more than two stamens.

The relative frequencies of the different types of variation in the andrœcium in the 618 specimens so far considered (see Table C) are very interesting. The number 1062 in Table F refers to 1062 flowers in which the andrœcium was

TABLE D.

	Number of Variations in		Number of Individuals in	
	Section	Sub-group	Section	Sub-group
Group *a*.				
Whole of the andrœcium normal. Fig. I	—	1	—	1062
Group *b*.				
Andrœcium variously modified	—	—	—	—
Sub-group *a*.		21	—	480
Outer whorl normal (5 and 6 variously modified)	—	—	—	—
Section i.				
Stamens 3 and 4 normal. Figs. II—IX	8	—	419	—
Section ii.				
Stamen 3 normal, 4 represented thus (½·1·1). Figs. X—XIV ...	5	—	25	—
Section iii.				
Stamens 3 and 4 thus {(1·0·1) (1·0·1)}. Figs. XV—XVIII	4	—	9	—
Section iv.				
Stamens 3 and 4 thus {(½·1·1) (½·1·1)}. Fig. XXII	1	—	2	—
Section v.				
Stamens 3 and 4 replaced by one. Figs. XIX—XXI	3	—	25	—
Sub-group *β*.		13	—	54
Outer whorl represented thus {(1·—·1) (½·—·1)}	—	—	—	—
Section i.				
Stamens 3 and 4 normal. Figs. XXIII—XXIX	7	—	42	—
Section ii.				
Stamens 3 and 4 thus {(1·1·1) (½·1·1)}. Figs. XXX—XXXIV ...	5	—	11	—
Section iii.				
Stamens 3 and 4 thus {(½·0·1) (0·0·0)}. Fig. XXXV	1	—	1	—
Sub-group *γ*.				
Stamens 1 and 2 represented thus {(½·—·1) (½·—·1)}	—	4	—	16
Section i.				
Stamens 3 and 4 normal. Figs. XXXVI and XXXVII	2	—	11	—
Section ii.				
Stamens 3 and 4 thus {(1·1·1) (½·1·1)}. Figs. XXXVIII and XXXIX	2	—	5	—
Sub-group *δ*.				
Stamen 1 normal, 2 absent	—	5	—	29
Section i.				
Stamens 3 and 4 normal. Figs. XL—XLIII	4	—	27	—
Section ii.				
Stamen 3 normal, 4 thus (½·1·1). Fig. XLIV	1	—	2	—
Sub-group *ε*.				
Stamen 1 thus (½·—·1), 2 absent. Figs. XLV—XLVII	—	3	—	6

TABLE E.

	Number of Variations in		Number of Individuals in	
	Section	Sub-group	Section	Sub-group
Sub-group η.				
Stamens 1 and 2 are normal	—	9	—	30
Section i.				
Stamens 3 and 4 are represented by three. Fig. XLVIII ...	1	—	3	—
Section ii.				
Stamens 3 and 4 are normal. Figs. XLIX—LI	3	—	16	—
Section iii.				
Stamens 3 and 4 represented thus $\{(1\cdot1\cdot1)\ (\frac{1}{2}\cdot1\cdot1)\}$. Fig. LII	1	—	2	—
Section iv.				
Stamens 3 and 4 thus $\{(1\cdot0\cdot1)\ (1\cdot0\cdot1)\}$. Figs. LIII—LV ...	3	—	5	—
Section v.				
Stamens 3 and 4 represented by one. Fig. LVI	1	—	4	—
Sub-group θ.				
Stamens 1 and 2 thus $\{(\frac{1}{2}\cdot—\cdot1)\ (\frac{1}{2}\cdot—\cdot1)\}$. Fig. LVII ...	—	1	—	1
Sub-group κ.				
Stamen 1 normal, 2 absent. Fig. LVIII	—	1	—	2

TABLE F.

Frequencies more than 3 in order of magnitude.

(References have been made to the figures.)

Figure	Frequency	Figure	Frequency	Figure	Frequency
I	1062	XXVIII	11	XXV	7
VIII	227	XXXVI	9	VII	6
III	130	VI	8	X	5
IV	38	XII	8	XLIII	5
XX	21	XIII	8	LI	5
XL	18	XLIX	8	XVIII	4
XXIII	15	IX	7	LVI	4

normal. Where variation occurs, the greatest frequency, namely 227, occurs in flowers in which one of the pairs in the inner whorl is replaced by a single stamen while the next highest frequency, namely 130, occurs in those flowers in which partial chorisis has taken place in the inner whorl of the andrœcium. Following this the magnitude of the frequencies diminishes rapidly. The next, namely 38, occurs in flowers in which nearly all the parts of the andrœcium are modified while, near this, is the frequency 21 which exists in flowers having only one stamen in each position. In the next two frequencies, namely 18 and 15, we find that stamens 1 and 2 are involved.

From these raw data we can see that the inner whorl of the andrœcium is the whorl most subject to variation and further that this variation is in the direction of a decrease in number.

TABLE G.

Class I, Sub-class B: Perianth normal, Gynœcium abnormal (see Table A).

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
3	LIX	1	1	—	1	1	LX	1	1	—	1
		2	1	—	1			2	1	—	1
		3	1	1	1			3	1	1	1
		4	1	1	1			4	1	1	1
		5	1	0	1			5	1	1	1
		6	0	0	0			6	$\frac{1}{2}$	1	1
1	LXII	1	1	1	1	2	LXIII	1	1	—	1
			$\frac{1}{2}$	1	1			2	$\frac{1}{2}$	—	1
		2	1	1	1			3	1	1	1
			1	1	1			4	1	1	1
		3	$\frac{1}{2}$	0	1				1	1	1
		4	$\frac{1}{2}$	0	1			5	1	1	1
		5	1	1	1			6	1	1	1
			1	1	1				1	1	1
		6	1	1	1						
			$\frac{1}{2}$	0	1						

Class II: Perianth abnormal.

Variations in the members of the perianth (calyx and corolla) have necessitated the introduction of new symbols in the diagrams. These are shown in the composite diagram Plate I, fig. 10, and are explained on p. 257.

It will be evident from Table H, p. 233, that the same type of variation in the andrœcium occurs with different types of variation in the perianth, e.g. in the second and fifth figures no fewer than six different variations in the perianth accompany a single type of variation in the andrœcium. Reference to the diagrams in the plates will show what these variations are and will render a detailed explanation unnecessary. The asterisk in Table H under LXXVIII indicates that there has been *adhesion* between stamen 1 and one of the stamens in position 3.4, in other words between one of the members of the outer whorl and one of the members in the inner whorl.

Class III.

The members of this class are characterised by a reduplication of the various organs but without separation into two distinct flowers. There are in all 11 individuals with 10 different types of variation. A word of explanation is necessary with regard to the interpretation of the position of the various stamens

TABLE H.

Class II.

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
1	XC	1	1	—	1	1	LXXXVII	1	1	—	1	1	LXVI	1	1	—	1
		2	1 1	½ ½	1 1	33	LXIV	2	1	—	1			2	1	—	1
		3	½	0	1	2	LXVIII	3	1	1	1			3	1	1	1
		4	0	0	0	2	LXIX	4	1	1	1			4	1	1	1
		5	0	0	0	2	LXX	5	1	1	1			5	1*	1	1
		6	0	0	0	1	LXXIV	6	1	1	1			6	1	1	1
		1	1	—	1	2	XCI	1	1	—	1			1	1	—	1
		2	1	—	1	1	LXXI	2	0	0	0			2	0	0	0
6	LXV	3	1	1	1	1	LXXVI	3	1	1	1	1	LXXIX	3	1	1	1
2	LXXII	4	1	1	1	31	LXXVII	4	1	1	1	1	LXXX	4	1	1	1
1	LXXIII	5	1	0	1	5	LXXXI	5	1	1	1			5	1	1	1
		6	0	0	0	1	XCII	6	1	1	1			6	1 1	1 1	1 1
1	LXXXIII	1	1	—	1	1	LXXXII	1	1	—	1	2	LXXXIV	1	½	—	1
		2	0	0	0			2	0	0	0			2	0	0	0
		3	½	1	1			3	1	1	1			3	1	1	1
		4	½	1	1			4	1	1	1			4	1	1	1
		5	½	1	1			5	1	0	1			5	1	1	1
		6	½	1	1			6	0	0	0			6	1	1	1
1	LXXXV	1	0	—	0			1	0	—	0	1	LXXVIII	1	1	—	1
		2	0	—	0			2	0	—	0			2	1	—	1
		3	1	1	1	1	LXXV	3	1	1	1			3	1	1	1
		4	1 1	1 1	1 1	1	LXXXVI	4	1 1	1 1	1 1			4	1*	1	1
		5	1	1	1			5	1	1	1			5	1	0	1
		6	1	1	1			6	1	1	1			6	0	0	0
3	LXVII	1	1	—	1	6	LXXXIX	1	1	—	1	3	LXXXVIII	1	1	—	1
		2	1	—	1			2	1	—	1			2	½	—	1
		3	1	1	1			3	1	0	1			3	1	1	1
		4	1	1	1			4	0	0	0			4	1	1	1
		5	½	0	1			5	1	0	1			5	1	1	1
		6	0	0	0			6	0	0	0			6	1	1	1

in these flowers. The stamens in the outer whorl are longer than those in the inner whorl. Consequently if these were reduced in length they might easily be mistaken for members of the outer whorl. In all cases of difficulty, however, the crucial test, the point of origin, was applied and the positions assigned to the various members as shown in the diagrams were determined microscopically in this manner. See Table I, p. 234.

TABLE I.

Class III.

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
2	XCVIII	1	1 1	1 1	1 1
		2	1 1	1 1	1 1
		3	1 1 1	1 1 1	1 1 1
		4	1 1 1 1	1 1 1 1	1 1 1 1
		5	1 1 1	1 1 1	1 1 1
		6	1 1 1 1	1 1 1 1	1 1 1 1
1	XCVII	1	1 1	1 1	1 1
		2	1 1	1 1	1 1
		3	1 1	1 1	1 1
		4	1 1	1 1	1 1
		5	1 1	1 1	1 1
		6	1 1	1 1	1 1

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
1	XCIX	1	1	0	1
		2	1	0	1
		3	1	0	1
		4	1 1	1 1	1 1
		5	1 1	1 1	1 1
		6	1 1 1	1 1 1	1 1 1
1	XCV	1	0	0	0
		2	1	0	1
		3	1 1	1 1	1 1
		4	1 1 1	0 0 0	1 1 1
		5	1	0	1
		6	1 1	1 1	1 1
1	CII	1	1 1	1 1	1 1
		2	1	—	1
		3	1	1	1
		4	1	1	1
		5	1	1	1
		6	1 1	1 1	1 1

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
1	XCIV	1	1 1	1 1	1 1
		2	1 ½	1 1	1 1
		3	1 1 1	1 1 1	1 1 1
		4	1 1	1 1	1 1
		5	1 1 1	0 0 1	1 1 1
		6	1 1	1 1	1 1
1	CI	1	1 1	1 1	1 1
		2	1 1	1 1	1 1
		3	1 1	1 1	1 1
		4	1 1	1 1	1 1
		5	1 1	1 1	1 1
		6	1 1	1 1	1 1

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
1	XCVI	1	½ 1	1 1	1 1
		2	1 1	1 1	1 1
		3	1 1	1 1	1 1
		4	1 1 1	1 0 0	1 1 1
		5	1 1	1 1	1 1
		6	1 1	1 1	1 1
1	XCIII	1	1 1	1 1	1 1
		2	1	—	1
		3	1	1	1
		4	1	1	1
		5	1	1	1
		6	1 1	1 1	1 1
1	C	1	1	—	1
		2	1	—	1
		3	1	1	1
		4	1	1	1
		5	1 1	1 1	1 1
		6	1	1	1

TABLE J.

Class IV.

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
5	CIII *A*	1	1	—	1	CIII *B*	1	1	—	1	1	CIV *A*	1	1	—	1	CIV *B*	1	½	—	1
		2	1	—	1		2	1	—	1			2	1	—	1		2	1	—	1
		3	1	1	1		3	1	1	1			3	1	1	1		3	½	1	1
		4	1	1	1		4	1	1	1			4	1	1	1		4	½	1	1
		5	1	1	1		5	1	1	1			5	1	1	1		5	½	1	1
		6	1	1	1		6	1	1	1			6	1	1	1		6	½	1	1
1	CV *A*	1	1	—	1	CV *B*	1	1	—	1	6	CVI *A*	1	1	—	1	CVI *B*	1	1	—	1
		2	1	—	1		2	1	—	1			2	1	—	1		2	1	—	1
		3	1	1	1		3	1	0	1			3	1	1	1		3	1	1	1
		4	1	1	1		4	1	0	1			4	1	1	1		4	1	1	1
		5	1	1	1		5	1	1	1			5	1	1	1		5	1	0	1
		6	1	1	1		6	1	1	1			6	1	1	1		6	0	0	0
3	CVII *A*	1	1	—	1	CVII *B*	1	1	—	1	1	CVIII *A*	1	1	—	1	CVIII *B*	1	½	—	1
		2	1	—	1		2	1	—	1			2	1	—	1		2	½	—	1
		3	1	1	1		3	1	0	1			3	½	1	1		3	1	0	1
		4	1	1	1		4	0	0	0			4	½	1	1		4	0	0	0
		5	1	1	1		5	1	0	1			5	½	1	1		5	1	1	1
		6	1	1	1		6	0	0	0			6	1	1	1		6	1	1	1

Class IV.

In this class there are 17 individuals giving six different modes of variation. Reduplication has taken place to such an extent as to give rise to two separate flowers on one pedicel. Each of the flowers was diminutive in size. Two tables are thus necessary for each "flower," *A* and *B*: see Table J, p. 235.

Class V.

This class has been formed to include two very aberrant flowers showing two distinct variations. In both cases part of the flower has been replaced by another flower, in one case normal in the other slightly divergent. CIX is one of these in which the original flower is normal except that one of the carpels has been replaced by a small flower (see Table K and diagram, Plate X). CX is the other. In the original flower stamen 1 has been chorised and one of the chorised parts has given origin to a separate flower (see Table K and diagram, Plate X).

TABLE K.

Class V.

Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther	Frequency	Number of Diagram	Number of Stamen	Filament	Chorisis	Anther
1	CIX *A*	1	1	—	1	1	CIX *B*	1	1	—	1
		2	1	—	1			2	1	—	1
		3	1	1	1			3	1	0	1
		4	1	1	1			4	0	0	0
		5	1	1	1			5	1	0	1
		6	1	1	1			6	0	0	0
1	CX *A*	1	1 1	0 0	1 *	1	CX *B*	1	1	—	1
		2	1	—	1			2	1	—	1
		3	1 1	1 1	1 1			3	1	1	1
		4	1	1	1			4	1	1	1
		5	1 1	1 1	1 1			5	1	1	1
		6	1	1	1			6	1	1	1

The asterisk denotes the position of the origin of the secondary flower.

This concludes our analysis of variations LIX to CX both as to perianth and andrœcium, but before proceeding to the statistical part it is desirable that certain peculiarities should be observed and that an understanding be arrived at with regard to the interpretation of these.

In this procedure 44 variations, namely LIX to CII, must be dealt with. When we study the number of parts which occur in the position of individual members of a whorl and then try to draw conclusions as to normality or abnormality of the whorl itself we find the following difficulties.

Let us take the outer whorl of the andrœcium as an example.

(1) If in the position normally occupied by stamen 1 there were two stamens and in the position normally occupied by stamen 2, no stamen occurred, then with regard to the whorl the total number of stamens would be *two*. Now this is the accepted normal number of stamens in the outer whorl, so that if number alone were considered the inference would legitimately be drawn from the table that the whorl was normal. But this is not so!

Or (2) If in position number 1, one normal stamen occurred and in position number 2, one functioning stamen, with the filament only half the normal length, occurred, then the number of functioning stamens in the whorl would be *two*, i.e. the accepted normal number. But again, on the basis of number alone, we should not be able to say whether the whorl as a whole was normal or abnormal.

Now as this state of affairs exists not only in the whorl under consideration but in all the whorls of the flower, we have thought it not only advisable but necessary to emphasise these abnormalities as a safeguard in the interest of systematic statistical treatment.

For this purpose, therefore, small diagrammatic formulæ have been drawn up, and these have been given in conjunction with the diagrams: see Plate I, figs. 11, 12.

We have already defined the positions of the various parts of the andrœcium but have hitherto refrained from naming the different constituents of the perianth.

In the cases under consideration, however, it is necessary to do so, and Plate I, fig. 11 illustrates how these are definitely determined.

The two outer sepals are named A and B (see Fig. 11). A corresponds in position to stamens 3.4 and B to stamens 5.6. C and D are the two inner sepals; C corresponds in position to stamen 1 and D to the stamen in position 2. The petals are named A', B', C' and D' and lie respectively between sepals A and C, B and D, A and D, and B and C.

The actual order of all the parts is summed up in Plate I, fig. 12 (1—14).

TABLE L. *Class V.*

Number of Whorl	Member of Whorl	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency
I	*A* *B*	1 1	2	*N*	LIX	2	1 1	2	*N*	LX	1	1 1	2	*N*	LXI	1
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 1 1	4	*N*			1 1 1 1	4	*N*			1 1 1 1	4	*N*		
IV	1 2	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
V	3.4 5.6	1 2	3	*A*			2 2	4	*A*			1 2	3	*A*		
I	*A* *B*	1 1	2	*N*	LXII	1	1 1	2	*N*	LXIII	2	1 1	2	*N*	LXIV	33
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 1 1	4	*N*			1 1 1 1	4	*N*			1 1 1 1	4	*A*		
IV	1 2	2 2	4	*A*			1 1	2	*A*			1 1	2	*N*		
V	3.4 5.6	2 4	6	*A*			3 3	6	*A*			2 2	4	*N*		
I	*A* *B*	1 1	2	*N*	LXV	6	1 1	2	*N*	LXVI	1	1 1	2	*N*	LXVII	3
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 1 1	4	*A*			1 1 1 1	4	*A*			1 1 1 1	4	*A*		
IV	1 2	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
V	3.4 5.6	2 1	3	*A*			2 2	4	*A*			2 1	3	*A*		
I	*A* *B*	1 1	2	*N*	LXVIII	2	1 1	2	*N*	LXIX	2	1 1	2	*N*	LXX	2
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 1 1	4	*A*			1 0 1 1	3	*A*			1 1 0 1	3	*A*		
IV	1 2	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
V	3.4 5.6	2 2	4	*N*			2 2	4	*N*			2 2	4	*N*		

TABLE L—(*continued*).

Number of Whorl	Member of Whorl	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency
I	*A* *B*	1 1	2	*N*	LXXI	1	1 1	2	*N*	LXXII	2	1 1	2	*N*	LXXIII	1
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 1 0	3	*A*			1 1 0 1	3	*A*			1 1 0 0	2	*A*		
IV	1 2	1 0	1	*A*			1 1	2	*N*			1 1	2	*N*		
V	3.4 5.6	2 2	4	*N*			2 1	3	*A*			1 2	3	*A*		
I	*A* *B*	1 1	2	*N*	LXXIV	1	1 1	2	*N*	LXXV	1	1 1	2	*A*	LXXVI	1
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*A*		
III	*A′* *B′* *C′* *D′*	0 0 0 1	1	*A*			0 0 2 3	5	*A*			1 1 0 0	2	*A*		
IV	1 2	1 1	2	*N*			0 0	0	*A*			1 0	1	*A*		
V	3.4 5.6	2 2	4	*N*			3 2	5	*A*			2 2	4	*N*		
I	*A* *B*	1 0	1	*A*	LXXVII	31	1 0	1	*A*	LXXVIII	1	1 0	1	*A*	LXXIX	1
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 0 1	3	*A*			1 1 0 1	3	*A*			1 0 1 1	3	*A*		
IV	1 2	1 0	1	*A*			1 1	2	*N*			1 0	1	*A*		
V	3.4 5.6	2 2	4	*N*			2 1	3	*A*			3 2	5	*A*		
I	*A* *B*	1 0	1	*A*	LXXX	1	1 0	1	*A*	LXXXI	5	1 0	1	*A*	LXXXII	1
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 0 1	3	*A*			0 0 1 1	2	*A*			0 1 0 1	2	*A*		
IV	1 2	1 0	1	*A*			1 0	1	*A*			1 0	1	*A*		
V	3.4 5.6	3 2	5	*A*			2 2	4	*N*			2 1	3	*A*		

TABLE L—(*continued*).

Number of Whorl	Member of Whorl	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency
I	*A* *B*	1 0	1	*A*	LXXXIII	1	1 0	1	*A*	LXXXIV	2	2 0	2	*A*	LXXXV	1
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	0 0 0 0	0	*A*			0 0 0 0	0	*A*			1 1 1 1	4	*N*		
IV	1 2	1 0	1	*A*			1 0	1	*A*			0 0	0	*A*		
V	3.4 5.6	2 2	4	*A*			2 2	4	*N*			2 3	5	*A*		
I	*A* *B*	1 1	2	*N*	LXXXVI	1	1 1	2	*N*	LXXXVII	1	1 1	2	*N*	LXXXVIII	3
II	*C* *D*	0 0	0	*A*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 0 0	2	*A*			1 1 1 1	4	*A*			0 0 1 1	2	*A*		
IV	1 2	0 0	0	*A*			1 1	2	*N*			1 1	2	*A*		
V	3.4 5.6	3 2	5	*A*			2 2	4	*N*			2 2	4	*N*		
I	*A* *B*	1 0	1	*A*	LXXXIX	6	1 0	1	*A*	XC	1	0 0	0	*A*	XCI	2
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	0 0 1 1	2	*A*			1 0 2 2	5	*A*			0 0 0 1	1	*A*		
IV	1 2	1 1	2	*N*			1 2	3	*A*			1 0	1	*A*		
V	3.4 5.6	1 1	2	*A*			1 0	1	*A*			2 2	4	*N*		
I	*A* *B*	1 1	2	*A*	XCII	1	1 2	3	*A*	XCIII	1	1 2	3	*A*	XCIV	1
II	*C* *D*	1 0	1	*A*			1 1	2	*N*			1 1	2	*N*		
III	*A′* *B′* *C′* *D′*	1 1 1 0	3	*A*			1 1 2 1	5	*A*			1 1 2 1	5	*A*		
IV	1 2	1 0	1	*A*			2 1	3	*A*			2 2	4	*A*		
V	3.4 5.6	2 2	4	*N*			2 3	5	*A*			5 5	10	*A*		

TABLE L—(*continued*).

Number of Whorl	Member of Whorl	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency
I	*A* *B*	1 2	3	*A*			2 2	4	*A*			2 2	4	*A*		
II	*C* *D*	1 1	2	*N*			1 1	2	*N*			1 1	2	*N*		
III	*A'* *B'* *C'* *D'*	1 1 2 1	5	*A*	XCV	1	1 0 2 2	5	*A*	XCVI	1	2 2 1 1	6	*A*	XCVII	1
IV	1 2	1 0	1	*A*			2 2	4	*A*			2 2	4	*A*		
V	3.4 5.6	5 3	8	*A*			4 5	9	*A*			4 4	8	*A*		
I	*A* *B*	1 2	3	*A*			1 2	3	*A*			1 1	2	*N*		
II	*C* *D*	2 2	4	*A*			2 2	4	*A*			2 2	4	*A*		
III	*A'* *B'* *C'* *D'*	1 2 1 1	5	*A*	XCVIII	2	2 2 1 1	6	*A*	XCIX	1	1 2 2 1	6	*A*	C	1
IV	1 2	2 2	4	*A*			1 1	2	*N*			1 1	2	*N*		
V	3.4 5.6	7 7	14	*A*			5 3	8	*A*			3 3	6	*A*		

Number of Whorl	Member of Whorl	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency	Number in each position	Members in Whorl	Whorl, Normal or Abnormal	Number of Diagram	Frequency
I	*A* *B*	1 1	2	*N*			1 1	2	*N*		
II	*C* *D*	2 2	4	*A*			2 2	4	*A*		
III	*A'* *B'* *C'* *D'*	1 2 2 1	6	*A*	CI	1	0 0 0 0	0	*A*	CII	1
IV	1 2	2 2	4	*A*			1 2	3	*A*		
V	3.4 5.6	4 4	8	*A*			3 2	5	*A*		

IV. STATISTICAL.

The analysis which we have given of 1813 flowers is sufficient to show that the idea of a definite fixed number of sepals in the calyx, of petals in the corolla, of stamens in the andrœcium or of carpels in the gynæcium of cruciferous plants is not upheld by an examination of a large number of flowers of this species. In less than 1 per cent. of the flowers examined there was an increase* or decrease in the number of sepals in the calyx; in less than 1 per cent. there was also an increase or decrease in the number of petals in the corolla, but in 2 per cent. there was an increase in the number of stamens in the andrœcium, while in 22 per cent. there was a decrease in the number.

Since then the number of sepals, petals and stamens is not *absolutely* fixed for any of the organs it becomes necessary now to consider whether the number of members in one organ is related to the number in the others.

As has already been pointed out we have not only to consider organs as a whole, but, in the case of the calyx and the andrœcium, the constituents of these organs, owing to the fact that these organs are each divided into two separate whorls which are inserted at different levels and are placed in directions at right angles to one another.

Further, a special study has been made of the various positions in andrœcium to ascertain to what extent bilateral symmetry may be regarded as an inherent character of the flower under consideration.

By this means also it seems that some definite information might be obtained with regard to the perplexing and, at present, hypothetical theory of chorisis, the reasons for the existence of which have been summarised on p. 219.

The statistical part has been divided into two sections:

(1) a study of the Means and Standard Deviations, and

(2) a study of the Correlation Coefficients.

1. *Study of the Means and Standard Deviations.*

Although it is obvious from the analysis of the data under consideration that the numbers given for the botanical floral formula, namely, Calyx—4, Corolla—4, Stamens—6 and Gynæcium—2, are the nearest integers, it is not at all certain from a mere inspection of the tables whether the actual means deviate from this number in the direction of excess or deficiency.

---

* Where chorisis of a sepal or petal has resulted in two or more distinct individuals we have regarded each of these as a distinct sepal or petal in recording the numbers. This method is *natural* however inasmuch as it is the only means by which we may possibly trace reduplication of parts.

Consequently the mean and standard deviation for each of the organs and its constituents have been calculated and these are given in the following table:

TABLE M.

Means and Standard Deviations of the Number of the Organs and their Constituents.

Organ	Constituent	Member	Mean Number	Standard Deviation	Coefficient of Variation
Calyx	—	—	3·9796	·2553	6·415
"	Outer whorl	—	1·9757	·1979	10·020
"	Inner whorl	—	2·0039	·1304	6·507
Corolla	—	—	3·9520	·3523	8·914
Andrœcium ...	—	—	5·8092	·7567	13·025
"	Outer whorl	—	1·9570	·2704	13·817
"	—	Stamen 1	·9840	·1602	16·280
"	—	Stamen 2	·9713	·1915	19·715
"	Inner whorl	—	3·8577	·6588	17·077
"	—	Stamens 3.4	1·9950	·2728	13·674
"	—	Stamens 5.6	1·8627	·4863	26·107

(1) The most obvious result which is revealed by these constants is the fact that in *all* cases (except the inner whorl of calyx) the actual mean of the organs is less than the recognised typical number, thus:

The mean number for the calyx is 3·979 instead of 4.

The mean number for the corolla is 3·952 instead of 4.

The mean number for the andrœcium is 5·809 instead of 6.

(2) The inner whorl of the calyx shows the smallest departure from the accepted typical number, namely, 2·004 instead of 2.

Let us, however, test how far the differences in the character of the analogous parts are significant by ascertaining the Probable Error of the difference of the means of the characters.

TABLE N (1).

I. Constituents of the Calyx.

Constituent	Mean Number	Standard Deviation
Outer whorl ...	1·9757	·1979
Inner whorl ...	2·0039	·1304

The difference here is $D = ·02813$ and the probable error of the difference $E_{D_m} = ·0037$; thus the value $\frac{D}{E_{D_m}} = 7·6$. The difference is therefore clearly significant.

It is worthy of notice that the outer whorl of the calyx is more variable than the inner whorl and that it possesses on an average fewer sepals.

TABLE N (2).

II. *Members of the Outer Whorl of the Andrœcium.*

Member	Mean Number	Standard Deviation
Position 1 ...	·9840	·1602
Position 2 ...	·9713	·1915

The difference here is $D = {\cdot}01268$ and $E_{D_m} = {\cdot}00391$; the value $\frac{D}{E_{D_m}} = 3{\cdot}2$.

Thus when the two positions of the outer whorl of the andrœcium are taken into consideration, a probably significant difference is found between the means of the distribution of the parts of this whorl. Now in position number 1 there is a greater approach to the accepted type owing to the fact that when the component of one of the positions of the outer whorl was found to depart from the accepted type, the other position was selected as the starting point for the orientation of the flower and was called position number 1. It is all the more noteworthy that the deviation for position 2 is not in the direction of greater but of lesser frequency and the variability of position 2 is greater. We have thus again a reduction in the value of the type with greater variability.

TABLE N (3).

III. *Members of the Inner Whorl of the Andrœcium.*

Member	Mean Number	Standard Deviation
Position 3 . 4 ...	1·9950	·2728
Position 5 . 6 ...	1·8627	·4863

Here the ratio $\frac{m_1 - m_2}{\Sigma_{(m_1 - m_2)}}$ is nearly 15 and therefore there is quite a significant difference between the means of the distributions of the two members of the inner whorl of the andrœcium. In both cases the tendency is towards a suppression of functioning stamens rather than an increase, together with greater variability in the case where the reduction from the accepted type is more marked.

This difference in variability is, in the main, real and is not due to the arbitrary selection of the 3 . 4 position. This will be evident from a study of Table XIX. It will there be seen that there were 1754 cases where two stamens occurred in one of the positions of the inner whorl of the andrœcium. This is the number in the accepted type, and thus there is no variability. What is the nature of the distribution of the stamens in the other position (Table XVIII)? It is as follows:

1	2	3	4	
287	1436	26	5	1754

The mean for this array is 1·8569 stamens, with a variability of ·4116. Thus when there is no variability in one position of the inner whorl of the andrœcium there is a large variability in the other position.

Similarly we find the following distribution for position 2 in the outer whorl of the andrœcium when position 1 is of the accepted type, i.e. shows no variability.

0	1	2	
57	1708	1	1766

The mean for this array is ·9683 with a variability of ·1784. Again therefore, when there is no variability in position 1, there is a reduction of type in position 2 with great variability.

2. *Study of the Correlation Coefficients.*

For the purposes of this study a number of correlation tables have been prepared and as the results of these will have to be considered under different groupings it seems advisable to tabulate them, and insert them consecutively. The system which has been adopted to facilitate reference is to commence with the outer whorl of the calyx and consider all *its* relations with the other whorls of the flower passing from the outside inwards; following this comes the inner whorl of the calyx and *its* relations with the other constituents of the flower from the outside inwards and so on.

The following table shows the characters studied and the correlation coefficients found.

In order to make the comparison of the various correlations as complete as possible it will be necessary to consider each constituent or organ with all the other constituents or organs and to avoid overlapping as far as possible. The most natural method would be to commence either with the outermost constituent, namely, the outer whorl of the calyx, or with the innermost constituent, namely, the inner whorl of the andrœcium. For reasons of a morphological character, which will be seen later, the inner whorl of the andrœcium has been chosen as the starting point.

TABLE O.

Correlation Coefficients between the Number of Various Organs and their Constituents.

	Table	r.
The outer whorl of the calyx and the inner whorl of the calyx	I	·1957
The outer whorl of the calyx and the corolla	II	·7275
The outer whorl of the calyx and the outer whorl of the andrœcium	III	·5886
The outer whorl of the calyx and the inner whorl of the andrœcium	IV	·2613
The outer whorl of the calyx and the andrœcium	V	·4371
The inner whorl of the calyx and the corolla	VI	·2476
The inner whorl of the calyx and the outer whorl of the andrœcium	VII	·3229
The inner whorl of the calyx and the inner whorl of the andrœcium	VIII	·3905
The inner whorl of the calyx and the andrœcium	IX	·4592
The calyx and the corolla	X	·6926
The calyx and the outer whorl of the andrœcium	XI	·6245
The calyx and the inner whorl of the andrœcium	XII	·4014
The calyx and the andrœcium	XIII	·5721
The corolla and the outer whorl of the andrœcium	XIV	·4762
The corolla and the inner whorl of the andrœcium	XV	·1773
The corolla and the andrœcium	XVI	·3174
The outer whorl of the andrœcium and the inner whorl of the andrœcium ...	XVII	·1984
The inner whorl of the andrœcium, position 3 . 4 and the inner whorl of the andrœcium, position 5 . 6	XVIII	·4305
The outer whorl of the calyx and the inner whorl of the andrœcium, position 3 . 4	XIX	·4646
The outer whorl of the calyx and the inner whorl of the andrœcium, position 5 . 6	XX	·2134
The inner whorl of the calyx and the inner whorl of the andrœcium, position 3 . 4	XXI	·4634
The inner whorl of the calyx and the inner whorl of the andrœcium, position 5 . 6	XXII	·2519
The corolla and the inner whorl of the andrœcium 3 . 4	XXIII	·2558
The corolla and the inner whorl of the andrœcium 5 . 6	XXIV	·0903
The outer whorl of the andrœcium and the inner whorl of the andrœcium, position 3 . 4	XXV	·2661
The outer whorl of the andrœcium and the inner whorl of the andrœcium, position 5 . 6	XXVI	·1539

(*a*) *The inner whorl of the andrœcium.*

From the standpoint of the systematic botanist the most anomalous constituent of the cruciferous flower is the inner whorl of the andrœcium, inasmuch as in each of the positions where one stamen would naturally be expected, the presence of two is regarded as typical. It has been explained in a previous section that botanists now usually regard this anomaly as having arisen by collateral chorisis from what was originally a single stamen in ancestral forms. For the sake of conciseness and in order to avoid unnecessary repetition the following abbreviations have been used in Tables P—X.

O. W. Ca. = Outer whorl of the calyx.
I. W. Ca. = Inner whorl of the calyx.
Ca. = Calyx.
Co. = Corolla.
O. W. A. = Outer whorl of the andrœcium.
I. W. A. = Inner whorl of the andrœcium.
A. = Andrœcium.

The following Table, P, gives the correlation coefficients between the I.W.A. and the other constituents or organs of the flower in order of position.

TABLE P.

I. W. A. and the other Constituents.

Constituent or Organ	Table	Correlation
O. W. Ca.... ...	IV	·2613
I. W. Ca.	VIII	·3905
Ca.	XII	·4014
Co.	XV	·1773
O. W. A.	XVII	·1984

The highest correlation between the inner whorl of the andrœcium and the other constituents or organs is that with the calyx; next in order come the inner whorl of the calyx, the outer whorl of the calyx, the outer whorl of the andrœcium, and lastly the corolla. In other words, we should be better able to predict the number of stamens in the inner whorl of the andrœcium from the number of members in the calyx than from the number of members in any other constituent or organ.

(*b*) *Relations between the organs themselves.*

Having thus discussed the inner whorl of the andrœcium with the other organs and constituents it might lead to some useful result if we proceed to determine the "organic correlation" existing between the various organs themselves. In this connection we have to consider the calyx, the corolla and the andrœcium, and for this purpose the correlation Tables X, XIII and XVI have been prepared. The character which has been selected for this study is the number of members in each organ.

The following Table (Q) shows the results obtained:

TABLE Q.

Correlation Coefficients between

Ca. and Co. ...	·6926
Ca. and A. ...	·5721
Co. and A. ...	·3174

(1) The calyx and corolla are much more highly correlated to one another than is either of these with the andrœcium. In other words, the two protective organs of the perianth are more highly correlated to one another than is either protective organ with the male reproductive organ. It is further evident that (2) the calyx is much more highly correlated to both the corolla and the andrœcium than are the two last named to one another. From (1) it may be concluded that, on an average,

an increase or decrease from the accepted typical number, namely four, of petals in the corolla is accompanied by an increase or decrease in the number of sepals in the calyx; while from (2) an increase or decrease in the number of stamens in the andrœcium will be accompanied, on an average, by a greater increase or decrease in the number of sepals than in the number of petals.

(*c*) *Relations between the constituents of organs.*

The constituents of (1) the calyx and (2) the andrœcium will now be considered.

(1) *Calyx.* The outer and inner whorls of this organ are inserted at different levels and have a decussate arrangement, so that, although the organ as a whole is protective in function, the two whorls actually help to enclose the flower at right angles to one another. The correlation between these two whorls is an extremely low one, namely, ·1957 (Table I), in other words, an increase or decrease in the number of sepals in either of the whorls of the calyx is associated only in a very small degree with an increase or decrease in the number of sepals in the other whorl. Or again it may be expressed thus, the two whorls of the calyx vary to a great extent independently of one another. This statement should be taken in conjunction with that made on p. 244 with regard to their Means and Variabilities and should also be borne in mind when the correlation between these two constituents and the other parts of the flower are discussed below (see Tables R and S).

(2) *Andrœcium.* This organ is also composed of two whorls, an outer and an inner inserted at different levels. Its function is of course reproductive. The correlation between the two constituents is very low, namely, ·1984 (see Table XVII), and is almost the same as that between the two whorls of the calyx. The inner whorl of the andrœcium shows greater variability than the outer whorl and tends to vary independently of this latter constituent, just as in the case of the two whorls of the calyx.

Having thus considered the organs *per se*, let us now compare the correlations between each individual constituent or organ and all the other constituents or

TABLE R.

(*d*) *Correlation Coefficients between the Outer Whorl of the Calyx and*

2nd Component	Table	r.
I. W. Ca. ...	I	·1957
Co.	II	·7275
O. W. A. ...	III	·5886
I. W. A. ...	IV	·2613
A.	V	·4371

organs. For this purpose it will be necessary to tabulate the results in series and consequently it might be well to start with the outermost constituent of the flower, namely, the outer whorl of the calyx, and tabulate the correlation coefficients passing inwards to the andrœcium. The inner whorl of the calyx will next be taken in relation to the other constituents and so on.

From the above table it will be seen that the outer whorl of the calyx is most highly correlated with the corolla; it is also highly correlated with the outer whorl of the andrœcium but much less so with the inner whorl of the andrœcium.

TABLE S.

(*e*) *Correlation Coefficients between the Inner Whorl of the Calyx and*

2nd Component	Table	r.
Co.	VI	·2476
O. W. A. ...	VII	·3229
I. W. A. ...	VIII	·3905
A.	IX	·4592

The low correlation between the inner whorl of the calyx and the corolla is due to the close adherence of the former to type, that is, there is very small variability.

TABLE T.

(*f*) *Correlation Coefficients between the Calyx and*

2nd Component	Table	r.
Co.	X	·6926
O. W. A. ...	XI	·6245
I. W. A. ...	XII	·4014
A.	XIII	·5721

There is a higher degree of correlation between the two organs of the perianth than between the calyx and the andrœcium. The high correlation between the calyx and the outer whorl of the andrœcium is mainly due to the high value obtained for the correlation between the outer whorl of the calyx and the outer whorl of the andrœcium.

TABLE U.

(*g*) *Correlation Coefficients between the Corolla and*

2nd Component	Table	r.
O. W. A. ...	XIV	·4762
I. W. A. ...	XV	·1773
A.	XVI	·3174

The corolla is much more highly correlated with the outer whorl than with the inner whorl of the andrœcium, and the correlation between the corolla and the andrœcium as a whole is not very great.

A comparison of Tables T and U shows that there is a much greater correlation between the calyx and the andrœcium and its two whorls, than between the corolla and the same constituents.

So far we have considered the relationships between the different parts of the flower from the outside inwards, but when we examine these relationships, taking the inner whorls as our starting point, some new aspects of the problem become manifest and, as these have been of great value in the interpretation of the results, it has been considered advisable to tabulate them thus:

TABLE V.

(*h*) *Correlation Coefficients between the Inner Whorl of the Andrœcium and*

2nd Component	Table	r.
Ca.	XII	·4014
I. W. Ca. ...	VIII	·3905
O. W. Ca. ...	IV	·2613
O. W. A. ...	XVII	·1984
Co.	XV	·1773

TABLE W.

(*i*) *Correlation Coefficients between the Outer Whorl of the Andrœcium and*

2nd Component	Table	r.
Ca.	XI	·6245
O. W. Ca. ...	III	·5886
Co.	XIV	·4762
I. W. Ca. ...	VII	·3229

A comparison of Tables V and W shows that the correlations between the outer whorl of the andrœcium and the other components are higher than for the inner whorl of the andrœcium, except in the case of the inner whorl of the calyx.

TABLE X.

(*j*) *Correlation Coefficients between the Andrœcium and*

2nd Component	Table	r.
Ca.	XIII	·5721
I. W. Ca. ...	IX	·4592
O. W. Ca. ...	V	·4371
Co.	XVI	·3174

This table shows that when the andrœcium is considered as a whole it is most highly correlated with the calyx and least correlated with the corolla.

V. MORPHOLOGICAL SIGNIFICANCE OF THE STATISTICAL RESULTS.

It is quite clear from the tabulated results that there is a definite departure from the usually accepted cruciferous structure in a very large number of the flowers of *Lepidium Draba* which have been examined for this study. This does not obtain merely in any one organ or constituent but in *all* the organs and constituents, although not to the same degree in each.

The statistical results will now be examined from the standpoint of the botanist in order (*a*) to note their morphological or genetic significance and (*b*) in order to see whether these figures throw any light on the evolution of this cruciferous plant.

It is almost axiomatic to state that the "purpose" of a flower is a purely reproductive one and that therefore its existence is justified only in so far as it serves to reproduce its kind. But not all the parts of a flower are solely reproductive in function. Each individual consists of two parts, (1) Reproductive, (2) Protective. (1) The reproductive organs are the gynæcium (♀) and the andrœcium (♂), while (2) The protective organs (perianth) are the corolla and the calyx.

One of the organs of the perianth, namely the corolla, is still further specialised. The calyx consists of four sepals, green in colour, whose sole function is to protect the flower when in the bud, and in many cases these are reflexed immediately after the flower has opened up, and are of no further importance to it. On the other hand the petals though essentially sepal-like in structure, in this as in the great majority of flowers, are not green but of some other colour. In the species under consideration they are white. Now although the petals are of great importance in protecting the reproductive organs while in the bud their utility does not cease when the flower opens but, along with small nectaries at the base of the stamens, serve as an attraction for insects whose visits are essential for cross-fertilisation.

The reproductive organs of what is regarded as the typical cruciferous flower consist of (1) the gynæcium which is composed of two carpels and (2) the andrœcium which is composed of six stamens. The stamens are delicate structures and do not hold an isolated position in the flower. When in the bud and immature they are subject to external influences, for example, (1) they might be shrivelled up by the heat of the sun, (2) they might be blasted by rain or wind or (3) they might be attacked by herbivorous insects, so that the protective perianth plays an important part in flower economics. Now what does an increase in the number of stamens imply? It is obvious that if the number of stamens is increased the total volume occupied by the reproductive organs is increased and consequently a tax is put upon the protective organs if they are to fulfil their function adequately. If the perianth does not respond to this tax from space considerations, the reproductive organs stand a small chance of ever fulfilling their function, so that one would naturally expect that variation of some kind in the perianth would follow variation in the reproductive organs.

Another important point which must never be lost sight of when interpreting the statistical results is the symmetry of the cruciferous flower. The calyx consists of two whorls each with two sepals; the corolla of one whorl of four petals and the andrœcium of two whorls of stamens, the outer having two members and the inner four members (see Plate I, fig. 7). Consequently a cruciferous flower is bilaterally symmetrical only on that vertical plane which passes *through* the division wall of the carpels, *between* each of the pairs of stamens in the inner whorl, *between* two petals on either side and *through* the middle of the outer pair of sepals. This plane may be referred to as the "plane of symmetrical division." Owing to the fact that the corolla consists of only one whorl, the outer whorl of the calyx corresponds in position to the inner whorl of the andrœcium, and the inner whorl of the calyx to the outer whorl of the andrœcium.

From a study of the Means and Standard Deviations of the various organs and constituents we arrive at the following conclusions:

Calyx. (1) The greatest approach to constancy in number in the whole flower is in the inner whorl of the calyx.

(2) There is a significant difference between the means of the two whorls.

(3) There is much greater variability in the outer than in the inner whorl of the calyx and on an average it possesses fewer sepals.

(4) There is a tendency towards a reduction from type in the number of sepals in the calyx.

Corolla. (5) There is a tendency towards a reduction from the accepted typical number in the number of petals in the corolla.

Andrœcium. (6) There is a significant difference between the means of the distributions of

(*a*) the members of the two whorls of the andrœcium,

(*b*) the members of the two positions in the inner whorl,

and (*c*) the members of the two positions in the outer whorl.

(7) From whatever axis we view the andrœcium as an organ it is distinctly asymmetrical in the distribution of its functioning stamens.

(8) There is much greater variability in the inner whorl than in the outer whorl of the andrœcium.

(9) In both positions in the inner whorl of the andrœcium there is a tendency towards a reduction from the accepted typical number of stamens and in the position where this is most marked there is the greatest variability.

The interpretation of these results is not at first sight very evident.

Why should there be a tendency towards a reduction in the number of members in the different organs of the flower and why should this tendency be most marked in the inner whorl of the andrœcium? As has already been pointed out all the flowers examined were taken from a single plant which gave rise to new stems by means of buds on the roots. May this tendency to reduction in the parts of the flower whose function is sexual reproduction not be an expression of a tendency towards an elimination of sexual in favour of vegetative reproduction? Another phenomenon which lends support to this hypothesis is the fact that in this plant the percentage of "pods" which attain maturity is extremely small.

Whether there is or is not a tendency towards vegetative reproduction, may we not also have here a harking back towards an ancestral form in which the number was less than the at present accepted typical number? In fact one would expect that if the present constitution of the inner whorl of the andrœcium had been most recent in development, reversion would first take place in it, and conversely one might reasonably conclude that since this whorl shows greatest variability, and most marked tendency to reduction in the number of members, it is more than probable that its present constitution was arrived at by an increase in number from a more primitive type.

Let us now examine the deductions made from a study of the correlation coefficients and see if they have any morphological interpretation.

(1) The calyx and corolla are more highly correlated with one another than is either of these with the andrœcium.

(2) The calyx is more highly correlated with the andrœcium than is the corolla. In other words, the two protective parts are more intimately associated in increase or decrease with one another than is either of these with the male reproductive organs, and further the calyx which is solely protective in function is more

intimately correlated with the male reproductive organs than is the corolla which serves as an attraction for insects as well as a protective covering of the bud.

(3) The two whorls of the calyx are not highly correlated, i.e. they vary independently of one another.

(4) The two whorls of the andrœcium also are not highly correlated. Morphologically this means that when there are two constituents in one organ, each having the same function, they may vary independently of one another, so that although an increase or decrease in the number in either *may* be correlated with an increase or decrease in the number in any other constituent of the flower, the same does not hold true with regard to the two constituents.

(5) The outer whorl of the calyx is most highly correlated with the corolla, next with the outer whorl of the andrœcium and lastly with the inner whorl of the andrœcium. The reason why the outer whorl of the calyx is more highly correlated with the outer whorl than with the inner whorl of the andrœcium is not at first sight very evident, but may be explained on the basis of its protective power. The members of the outer whorl of the andrœcium lie in a plane parallel to that of the outer whorl of the calyx, and are much more widely separated in this plane than are the members of the inner whorl of the andrœcium. Consequently any increase in the number of stamens in the outer whorl would involve a much greater increase in volume within the flower than a corresponding increase in the number of stamens in the inner whorl. Thus we are not surprised to find that such an increase in the outer whorl of the andrœcium is more intimately associated with an increase in the outer whorl of the calyx than a corresponding increase in the inner whorl of the andrœcium would be.

(6) There is very low variability in the inner whorl of the calyx and it is almost equally correlated to the two whorls of the andrœcium. The morphological explanation of these facts follows as a corollary to that given above.

(7) The calyx is much more highly correlated with the andrœcium as a whole and with its two whorls than is the corolla.

As we have already said the calyx is the predominantly protective organ and consequently this higher correlation has a physical basis. The corolla being partly attractive does not enter so closely into space economics.

(8) The outer whorl of the andrœcium is more highly correlated with the other components of the flower than is the inner whorl of the andrœcium. This again follows on the basis of space considerations. Any increase in the number of members in the inner whorl of the andrœcium does not involve so radical a change in the *volume* of the flower as does a corresponding increase in the outer whorl of the andrœcium.

VI. VARIATION IN THE GYNÆCIUM.

So far we have not considered the gynæcium on account of the small number of variations which occur in that organ and from the fact that these do not lend themselves to statistical treatment.

The gynæcium consists typically of two carpels which are flattened in a vertical plane parallel to those containing the pairs of stamens in the inner whorl of the andrœcium. The thin partition wall separating the two carpels therefore stands at right angles to this plane.

Now when we examine the different types of variations in the structure and number of the carpels we find the following: (1) a single carpel, (2) two carpels (typical), (3) three carpels, (4) four carpels, (5) two sets of two carpels within a single perianth, (6) two sets of two carpels within separate perianths but on one pedicel.

Let us now proceed to examine each of these in some detail.

(1) *The gynœcium consists of a single carpel* (see Figs. LXXXVII—XCII).

In all these cases, except LXXXVII, as will be at once seen by reference to the figures, the suppression of a carpel is accompanied by the suppression of some of the members of nearly all the other organs thus:

In LXXXVIII two petals are absent and one stamen is aborted.

In LXXXIX one sepal, two petals and two stamens are absent.

In XC, XCI and XCII all the organs are deficient in members.

A noteworthy phenomenon in this respect also is that the suppression of members which accompanies the suppression of a carpel is usually in the vertical plane which passes through the plane of separation of the carpels.

(2) *The gynœcium consists of two carpels.*

This is the accepted typical structure and the statistical study deals with these in detail.

(3) *The gynœcium consists of three carpels* (see Figs. XCIII and CII).

When three carpels occur in the gynæcium they are never found co-laterally, i.e. the additional carpel is never found with its origin at the side of a carpel, but always arising from the plane of separation, which is in the plane of greatest variability.

(4) *The gynœcium consists of four carpels* (see Fig. CI).

Just as in the previous case the increase in the number of carpels takes place in the plane of separation of the carpels—one on either side, so that a cruciate structure is found. A reference to Fig. CI will make this clear. In both of these groups it will be evident that an increase in the female reproductive organs is associated not only with an increase in the male reproductive organs but also in an increase in the protective organs or perianth.

(5) *The gynæcium consists of two sets of carpels within a single perianth.*

This is rather an anomalous group but is extremely interesting inasmuch as it contains a series of annectant forms linking group 2 to group 6. What we actually have here is a complete reduplication of the reproductive organs encased within a single series of protective organs. In some of the flowers examined with this structure it was rather difficult to determine the orientation owing to a torsion of the thalamus, but in the types figured on Plates IX and X (Fig. XCIV and Figs. XCV et seq.) the mode of origin of these is quite evident. Several important observations on these forms may be stated.

(*a*) There are really two complete sets of reproductive organs and in one case (see Fig. XCVII) each of these is of the typical cruciferous structure.

(*b*) Increase in the number of the reproductive organs is accompanied by an increase in the number of members in the protective organs.

(*c*) The increase in the number of members of the reproductive organs is for the most part in the plane of division of the carpels, in other words, in the outer whorl of the calyx and its associated petals.

(*d*) This is also the plane along which the separation of the reproductive organs has taken place.

(*e*) This plane is the one which we have already shown in the statistical part to be the plane of greatest variability.

(6) *The gynæcium consists of two sets of two carpels within separate perianths but on one pedicel.*

In this group we reach the limit of variability in the material examined. In place of a single flower consisting of calyx, corolla, andrœcium and gynæcium we actually find two complete sets of all these organs, on one pedicel (see Figs. CIII—CVIII), while in one case (Fig. CIII) each of the two flowers has the typical cruciferous structure, so that were each of these separately examined it would undoubtedly be regarded as a normal flower. Yet we must bear in mind that, botanically considered, one flower and one flower only arises from a pedicel. Were this, therefore, an isolated example, and if no annectant forms existed, the departure might well be regarded as a "mutation," but a consideration of the numerous variations which we have already considered, taken in conjunction with group 5, only serves to emphasise the fact that "the vertical plane which passes through the partition wall of the two carpels and consequently separates the individuals of the pairs of stamens in the inner whorl and passes through the centres of the sepals of the outer whorl of the calyx is a plane along which this flower is in a state of flux and is the plane in which it is probable that the flower has changed, and is still changing, from some quite different ancestral form."

VII. SUGGESTIONS FOR FUTURE STUDIES IN THIS PLANT.

It must be very obvious to anyone who has perused this paper that the results which might be obtained from a study of this plant are by no means exhausted. An attempt, however, has been made to interpret the variability in its flowers, both from a morphological and an evolutionary standpoint. Studies of a different nature might be undertaken in order to test the results obtained, e.g.:

(1) What is the degree of fertility in the flowers of this plant? For this purpose it would be necessary to find the percentage of flowers which produce fertile seed.

(2) What are the variants, if any, which are associated with infertility?

(3) What are the characters of the flowers which are produced from the seeds of the different variants? If seeds selected from the different variants were grown separately and self-fertilised, one could trace the variations in the flowers of the next generation and see to what extent the different variations were transmitted. This study is capable of much elaboration and is one which would be fraught with great possibilities. It seems to involve a satisfactory method of determining how far these variations are concerned in plant economics, and also to what extent they have been instrumental in the evolution of the Order Cruciferæ.

EXPLANATION OF FIGURES 8, 9 AND 10. PLATE I.

FIGURE 8.

(*a*) Typical stamen (outer whorl).
(*b*) Stamen with half-length filament and complete anther (outer whorl).
(*c*) Typical stamen (inner whorl).
(*d*) Non-chorised stamen with two complete anthers (inner whorl).
(*e*) Stamen of inner whorl with two complete anthers but only chorised in the upper half.

FIGURE 9.

(*a*) Aborted stamen of outer whorl, i.e. filament with no anther.
(*b*) Absence of stamen in outer whorl.
(*c*) Full-length filament in inner whorl with no anther.
(*d*) Half-length filament in inner whorl with complete anther.
(*e*) Half-length filament in inner whorl with no anther.
(*f*) Non-chorised stamen in inner whorl with half-length filament but with two complete anthers.

FIGURE 10.

(*a*) Normal sepal.
(*b*) Sepal divided almost to the very base.
(*c*) Sepal completely divided into two distinct sepals.
(*d*) Sepal absent.
(*e*) Normal petal.
(*f*) Petal divided almost to the very base.
(*g*) Aborted petal.
(*h*) Petal absent.

TABLE I.

Outer Whorl of Calyx.
Number of Sepals.

Inner Whorl of Calyx. Number of Sepals.	*0*	*1*	*2*	*3*	*4*	Totals
0	—	—	1	—	—	1
1	—	2	1	—	—	3
2	2	48	1748	3	2	1803
3	—	—	—	—	—	0
4	—	—	3	3	—	6
Totals	2	50	1753	6	2	1813

TABLE II.

Outer Whorl of Calyx.
Number of Sepals.

Corolla. Number of Petals.	*0*	*1*	*2*	*3*	*4*	Totals
0	—	3	1	—	—	4
1	2	—	1	—	—	3
2	—	12	6	—	—	18
3	—	34	8	—	—	42
4	—	—	1734	—	—	1734
5	—	1	1	5	1	8
6	—	—	2	1	1	4
Totals	2	50	1753	6	2	1813

TABLE III.

Outer Whorl of Calyx.
Number of Sepals.

Outer Whorl of Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
0	—	—	5	—	—	5
1	2	42	40	1	—	85
2	—	7	1705	1	—	1713
3	—	1	1	1	—	3
4	—	—	2	3	2	7
Totals	2	50	1753	6	2	1813

TABLE IV.

Outer Whorl of Calyx.
Number of Sepals.

Inner Whorl of Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
1	—	1	—	—	—	1
2	—	6	24	—	—	30
3	—	2	292	—	—	294
4	2	39	1399	—	—	1440
5	—	2	25	1	—	28
6	—	—	12	—	—	12
8	—	—	1	2	1	4
9	—	—	—	—	1	1
10	—	—	—	1	—	1
14	—	—	—	2	—	2
Totals	2	50	1753	6	2	1813

TABLE V.

Outer Whorl of Calyx.
Number of Sepals.

Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
4	—	8	30	—	—	38
5	2	40	325	—	—	367
6	—	2	1365	—	—	1367
7	—	—	19	—	—	19
8	—	—	12	1	—	13
9	—	—	—	1	—	1
10	—	—	1	1	—	2
12	—	—	1	—	1	2
13	—	—	—	—	1	1
14	—	—	—	1	—	1
18	—	—	—	2	—	2
Totals	2	50	1753	6	2	1813

TABLE VI.

Inner Whorl of Calyx.
Number of Sepals.

Corolla. Number of Petals.	*0*	*1*	*2*	*3*	*4*	Totals
0	—	2	1	—	1	4
1	—	—	3	—	—	3
2	1	—	17	—	—	18
3	—	1	41	—	—	42
4	—	—	1734	—	—	1734
5	—	—	6	—	2	8
6	—	—	1	—	3	4
Totals	1	3	1803	0	6	1813

TABLE VII.

Inner Whorl of Calyx.
Number of Sepals.

Outer Whorl of Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
0	1	—	4	—	—	5
1	—	3	82	—	—	85
2	—	—	1711	—	2	1713
3	—	—	2	—	1	3
4	—	—	4	—	3	7
Totals	1	3	1803	0	6	1813

TABLE VIII.

Inner Whorl of Calyx.
Number of Sepals.

Inner Whorl of Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
1	—	—	1	—	—	1
2	—	—	30	—	—	30
3	—	—	294	—	—	294
4	—	3	1437	—	–	1440
5	1	—	26	—	1	28
6	—	—	11	—	1	12
8	—	—	2	—	2	4
9	—	—	1	—	—	1
10	—	—	1	—	—	1
14	—	—	—	—	2	2
Totals	1	3	1803	0	6	1813

TABLE IX.

Inner Whorl of Calyx.
Number of Sepals.

Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
4	—	—	38	—	—	38
5	1	3	363	—	—	367
6	—	—	1367	—	—	1367
7	—	—	19	—	—	19
8	—	—	11	—	2	13
9	—	—	1	—	—	1
10	—	—	1	—	1	2
12	—	—	1	—	1	2
13	—	—	1	—	—	1
14	—	—	1	—	—	1
18	—	—	—	—	2	2
Totals	1	3	1803	0	6	1813

TABLE X.

Calyx.
Number of Sepals.

Corolla. Number of Petals.	*2*	*3*	*4*	*5*	*6*	*7*	Totals
0	2	1	·	—	1	—	4
1	2	—	1		—	—	3
2	1	12	5	—		—	18
3	—	35	7	—	—	–	42
4	—	—	1734	—	—	–	1734
5	—	1	1	3	1	2	8
6	—	—	—	—	3	1	4
Totals	5	49	1748	3	5	3	1813

TABLE XI.

Calyx.
Number of Sepals.

Outer Whorl of Andrœcium. Number of Stamens.	*2*	*3*	*4*	*5*	*6*	*7*	Totals
0	1	—	4	—	—	—	5
1	4	41	39	1	—	—	85
2	—	7	1704	—	1	1	1713
3	—	1	—	1	1	—	3
4	—	—	1	1	3	2	7
Totals	5	49	1748	3	5	3	1813

TABLE XII.

Calyx.
Number of Sepals.

Inner Whorl of Andrœcium. Number of Stamens.	*2*	*3*	*4*	*5*	*6*	*7*	Totals
1	—	1	—	—	—	—	1
2	—	6	24	—	—	—	30
3	—	2	292	—	—	–	294
4	4	38	1398	—	—	—	1440
5	1	2	23	1	1	—	28
6	—	—	11	—	1	—	12
8	—	—	—	1	2	1	4
9	—	—	-	—	1	—	1
10	—	—	—	1		—	1
14	—	—	—	—	—	2	2
Totals	5	49	1748	3	5	3	1813

TABLE XIII.

Calyx.

Number of Sepals.

Andrœcium. Number of Stamens.	*2*	*3*	*4*	*5*	*6*	*7*	Totals
4	—	8	30	-	—	—	38
5	5	39	323	—	—	—	367
6	—	2	1365	—	—	—	1367
7	—	—	19	—	—	—	19
8	—	—	10	1	2	—	13
9	—	—	—	1	—	—	1
10	—	—	1	—	—	1	2
12	—	—	—	—	2	—	2
13	—	—	—	—	1	—	1
14	—	—	—	1	—	—	1
18	—	—	—	—	—	2	2
Totals	5	49	1748	3	5	3	1813

TABLE XIV.

Corolla.

Number of Petals.

Outer Whorl of Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	Totals
0	—	—	1	2	1	1	—	5
1	3	2	7	35	37	1	—	85
2	—	1	10	5	1695	—	2	1713
3	1	—	—	—	—	2	—	3
4	—	—	—	—	1	4	2	7
Totals	4	3	18	42	1734	8	4	1813

TABLE XV.

Corolla.

Number of Petals.

Inner Whorl of Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	Totals
1	—	—	-	—	—	1	—	1
2	—	—	6	—	24	—	—	30
3	—	—	2	3	289	—	—	294
4	3	3	9	37	1388	—	—	1440
5	1	—	1	2	22	2	—	28
6	—	—	—	—	11	—	1	12
8	—	—	—	—	—	1	3	4
9	—	—	—	—	—	1	—	1
10	—	—	—	—	—	1	—	1
14	—	—	—	—	—	2	—	2
Totals	4	3	18	42	1734	8	4	1813

TABLE XVI.

Corolla.

Number of Petals.

Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	Totals
4	—	—	7	2	28	1	—	38
5	3	2	8	36	317	1	—	367
6	—	1	3	4	1359	—	—	1367
7	—	—	—	—	19	—	—	19
8	1	—	—	—	10	1	1	13
9	—	—	—	—	—	1	—	1
10	—	—	—	—	1	—	1	2
12	—	—	—	—	—	—	2	2
13	—	—	—	—	—	1	—	1
14	—	—	—	—	—	1	—	1
18	—	—	—	—	—	2	—	2
Totals	4	3	18	42	1734	8	4	1813

TABLE XVII.

Outer Whorl of Andrœcium.

Number of Stamens.

Inner Whorl of Andrœcium. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
1	—	—	—	1	—	1
2	—	—	30	—	—	30
3	—	5	289	—	—	294
4	2	75	1363	—	—	1440
5	3	4	19	2	—	28
6	—	—	11	—	1	12
8	—	1	1	—	2	4
9	—	—	—	—	1	1
10	—	—	—	—	1	1
14	—	—	—	—	2	2
Totals	5	85	1713	3	7	1813

TABLE XVIII.

Inner Whorl of Andrœcium, position 3 . 4.

Number of Stamens.

Inner Whorl of Andrœcium 5 . 6. Number of Stamens.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	Totals
0	1	—	—	—	—	—	—	1
1	30	287	—	—	—	—	—	317
2	7	1436	2	1	—	—	—	1446
3	4	26	6	—	1	—	—	37
4	—	5	—	2	—	—	—	7
5	—	—	1	1	1	—	—	3
6	—	—	—	—	—	—	—	0
7	—	—	—	—	—	—	2	2
Totals	42	1754	9	4	2	0	2	1813

TABLE XIX.

Outer Whorl of Calyx.

Number of Sepals.

Inner Whorl of Andrœcium, position 3 . 4. Number of Stamens.

	0	*1*	*2*	*3*	*4*	Totals
1	—	8	34	—	—	42
2	2	42	1709	1	—	1754
3	—	—	8	1	—	9
4	—	—	2	—	2	4
5	—	—	—	2	—	2
6	—	—	—	—	—	0
7	—	—	—	2	—	2
Totals	2	50	1753	6	2	1813

TABLE XX.

Outer Whorl of Calyx.

Number of Sepals.

Inner Whorl of Andrœcium, position 5 . 6. Number of Stamens.

	0	*1*	*2*	*3*	*4*	Totals
0	—	1	—	—	—	1
1	—	7	310	—	—	317
2	2	40	1404	—	—	1446
3	—	2	33	2	—	37
4	—	—	6	—	1	7
5	—	—	—	2	1	3
7	—	—	—	2	—	2
Totals	2	50	1753	6	2	1813

TABLE XXI.

Inner Whorl of Calyx.

Number of Sepals.

Inner Whorl of Andrœcium, position 3 . 4. Number of Stamens.

	0	*1*	*2*	*3*	*4*	Totals
1	—	—	42	—	—	42
2	—	3	1750	—	1	1754
3	1	—	7	—	1	9
4	—	—	3	—	1	4
5	—	—	1	—	1	2
6	—	—	—	—	—	0
7	—	—	—	—	2	2
Totals	1	3	1803	0	6	1813

TABLE XXII.

Inner Whorl of Calyx.

Number of Sepals.

Inner Whorl of Andrœcium, position 5 . 6. Number of Stamens.

	0	*1*	*2*	*3*	*4*	Totals
0	—	—	1	—	—	1
1	—	—	317	—	—	317
2	1	3	1442	—	—	1446
3	—	—	34	—	3	37
4	—	—	6	—	1	7
5	—	—	3	—	—	3
6	—	—	—	—	—	0
7	—	—	—	—	2	2
Totals	1	3	1803	0	6	1813

TABLE XXIII.

Corolla.

Number of Petals.

Inner Whorl of Andrœcium, position 3 . 4. Number of Stamens.

	0	*1*	*2*	*3*	*4*	*5*	*6*	Totals
1	—	—	7	1	33	1	—	42
2	4	3	10	41	1695	1	—	1754
3	—	—	1	—	5	2	1	9
4	—	—	—	—	1	1	2	4
5	—	—	—	—	—	1	1	2
6	—	—	—	—	—	—	—	0
7	—	—	—	—	—	2	—	2
Totals	4	3	18	42	1734	8	4	1813

TABLE XXIV.

Corolla.

Number of Petals.

Inner Whorl of Andrœcium, position 5.6. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	Totals
0	—	—	—	—	—	1	—	1
1	—	—	7	2	308	—	—	317
2	3	3	11	38	1390	1	—	1446
3	1	—	—	2	31	1	2	37
4	—	—	—	—	5	—	2	7
5	—	—	—	—	—	3	—	3
6	—	—	—	—	—	—	—	0
7	—	—	—	—	—	2	—	2
Totals	4	3	18	42	1734	8	4	1813

TABLE XXV.

Outer Whorl of Andrœcium.

Number of Stamens.

Inner Whorl of Andrœcium, position 3.4. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
1	—	—	41	1	—	42
2	3	84	1665	2	—	1754
3	2	1	6	—	—	9
4	—	—	—	—	4	4
5	—	—	1	—	1	2
6	—	—	—	—	—	0
7	—	—	—	—	2	2
Totals	5	85	1713	3	7	1813

TABLE XXVI.

Outer Whorl of Andrœcium.

Number of Stamens.

Inner Whorl of Andrœcium, position 5.6. Number of Stamens.	*0*	*1*	*2*	*3*	*4*	Totals
0	—	—	—	1	—	1
1	—	5	312	—	—	317
2	4	75	1366	0	1	1446
3	1	4	30	2	—	37
4	—	—	5	—	2	7
5	—	1	—	—	2	3
6	—	—	—	—	—	0
7	—	—	—	—	2	2
Totals	5	85	1713	3	7	1813

PLATE I.

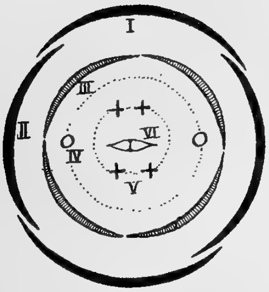


Fig 1

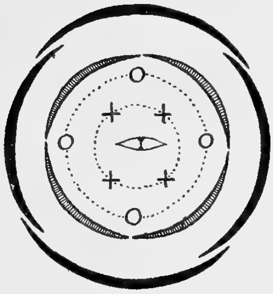


Fig 2

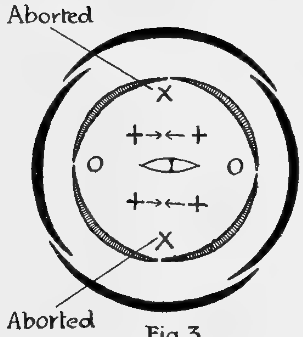


Fig 3

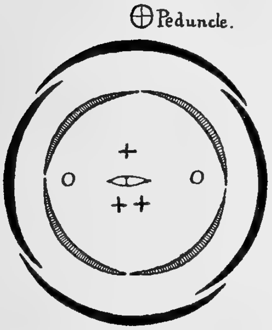


Fig 4

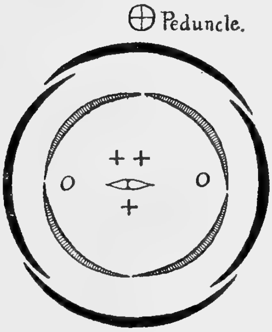


Fig 5

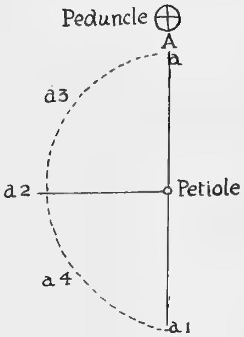


Fig 6

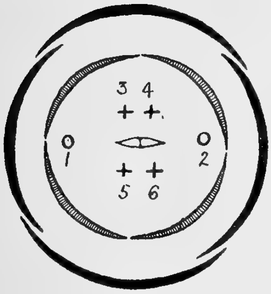


Fig 7

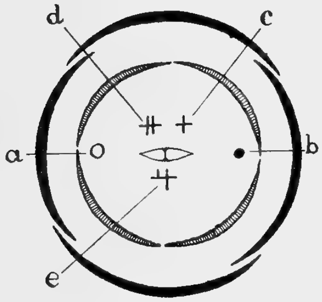


Fig 8

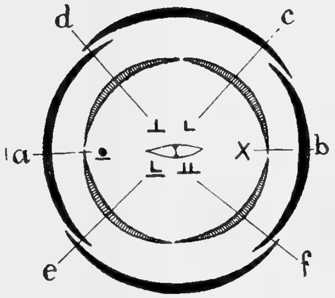


Fig 9

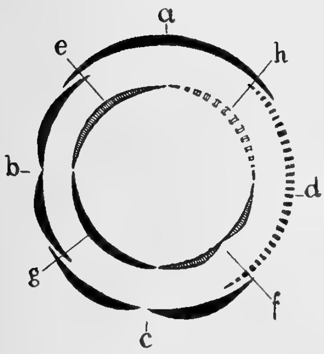


Fig 10

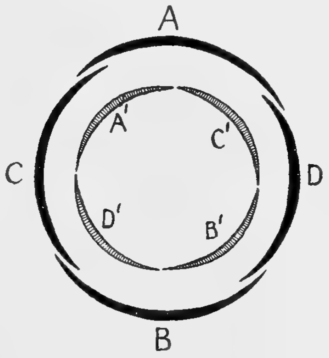


Fig 11

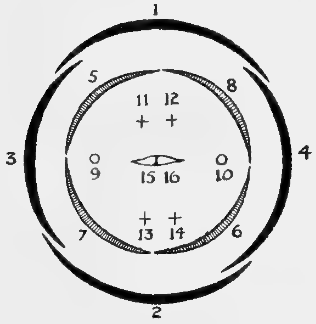


Fig 12

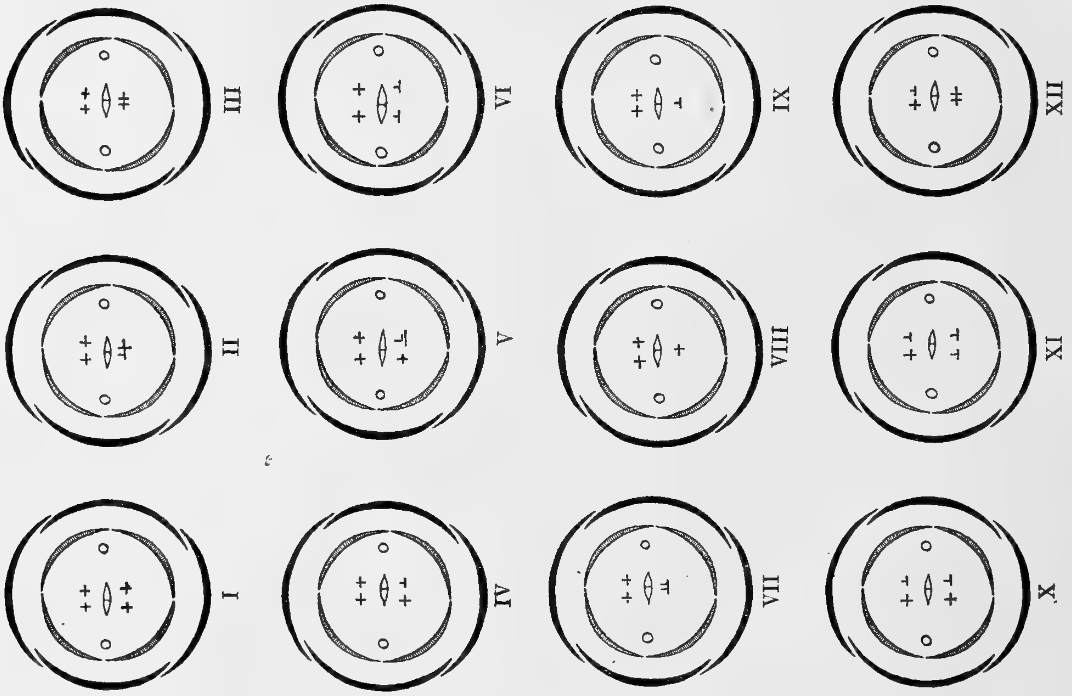


PLATE II.

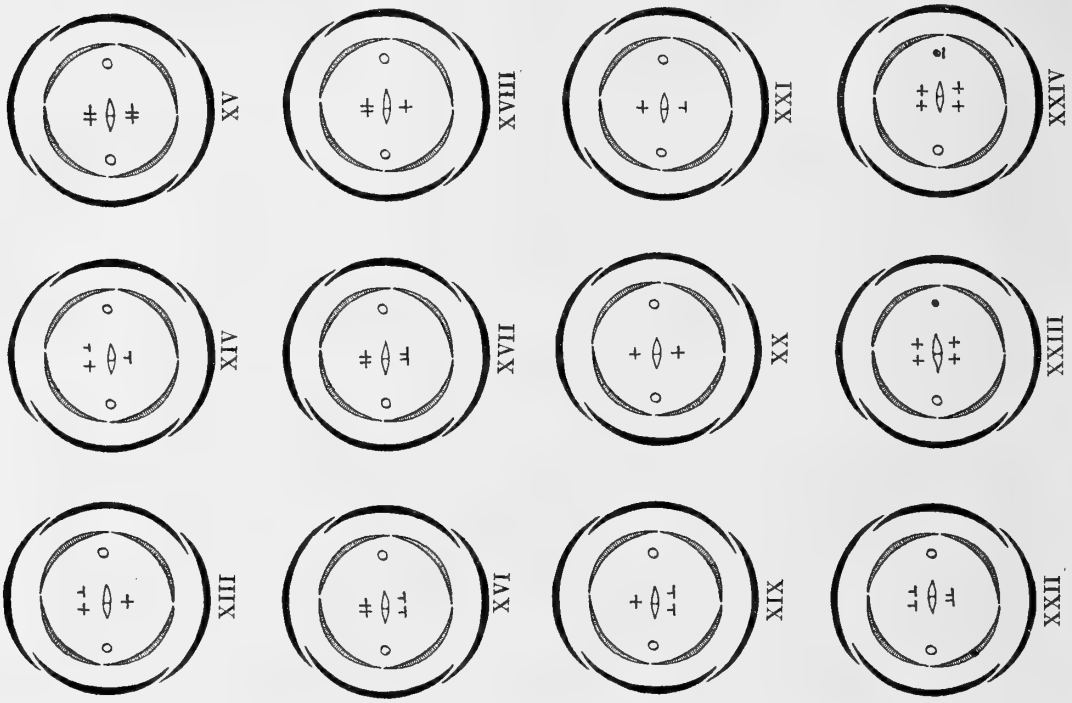


PLATE III.

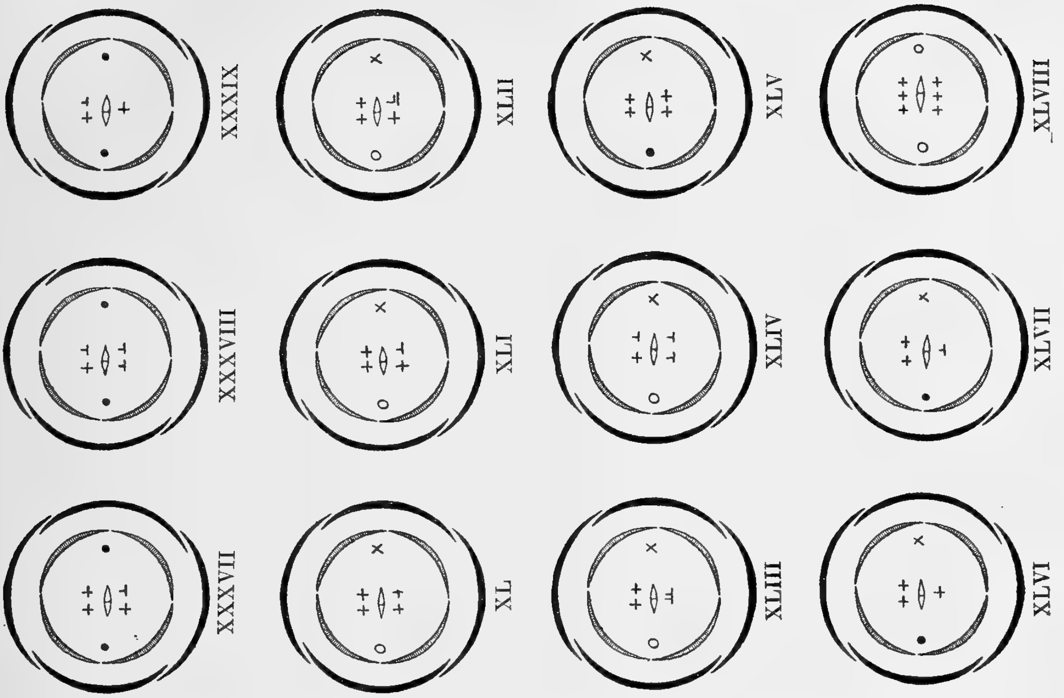


PLATE V.

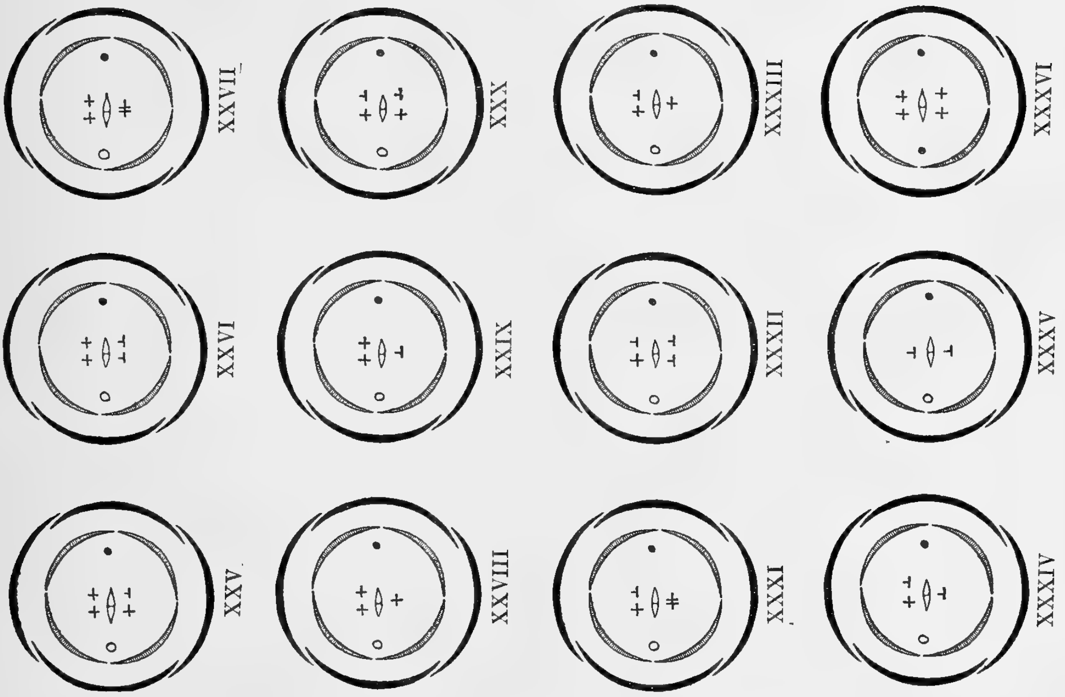


PLATE IV.

PLATE VII.

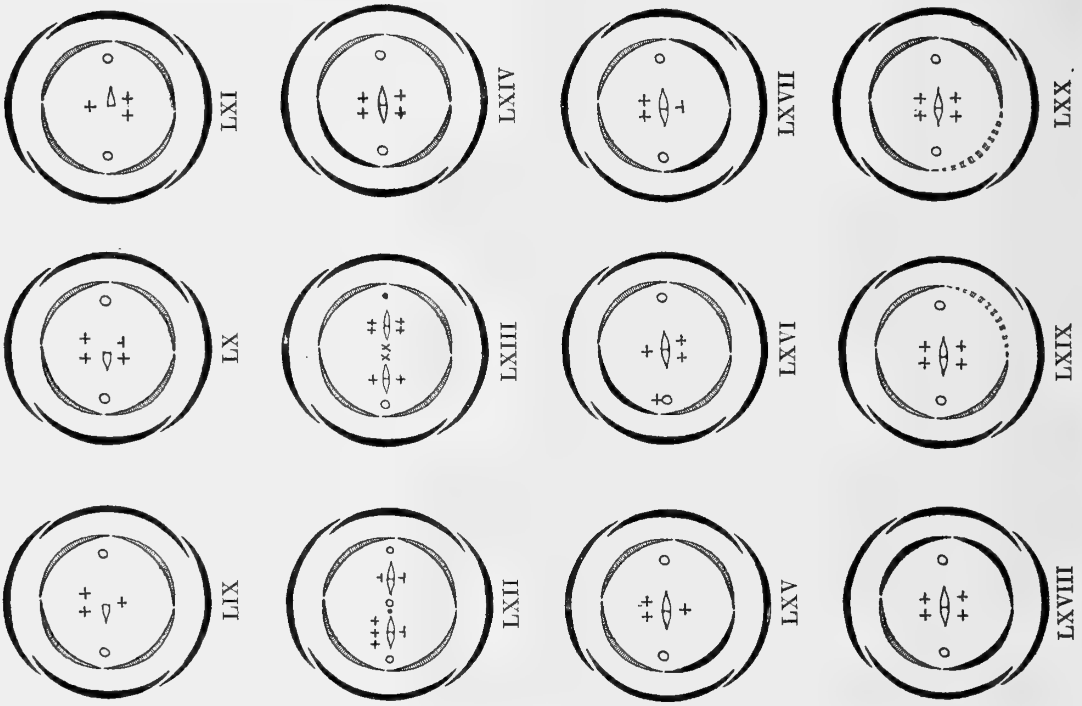


PLATE VI.

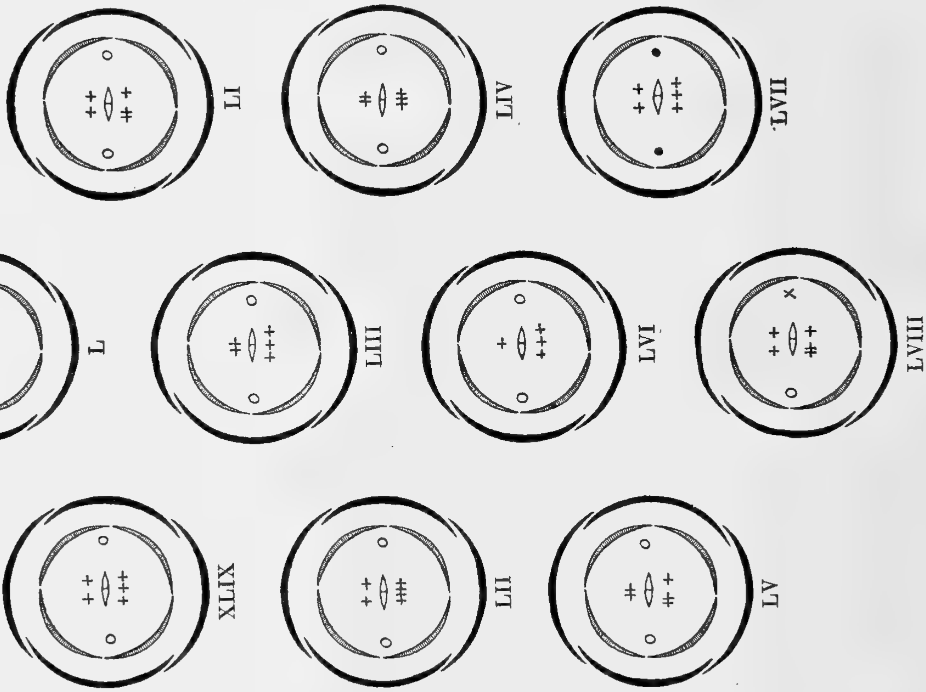
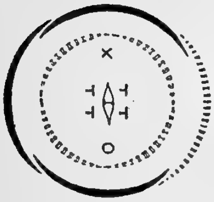
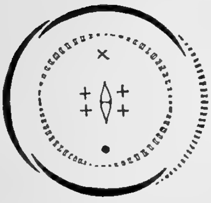
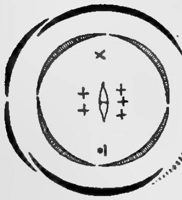
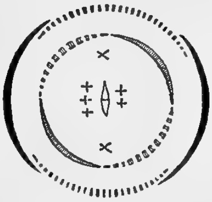


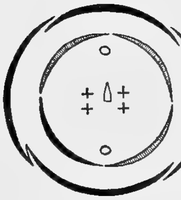
Plate IX.

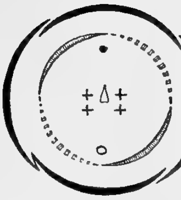

LXXXIII

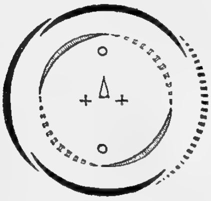

LXXXIV

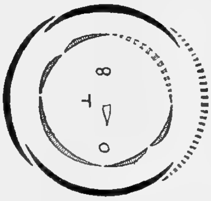

LXXXV

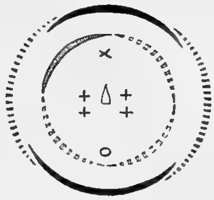

LXXXVI

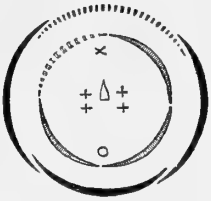

LXXXVII

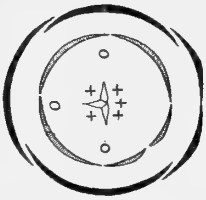

LXXXVIII


LXXXIX


XC


XCI


XCII


XCIII

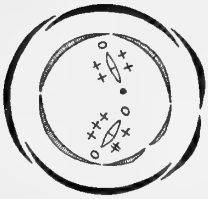
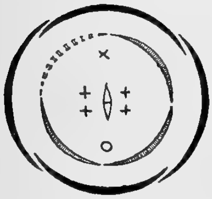
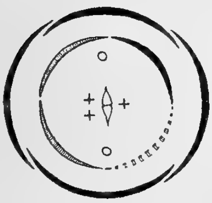
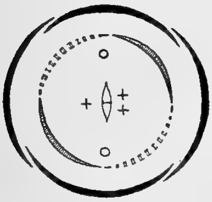
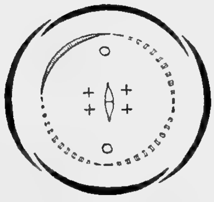

XCIV

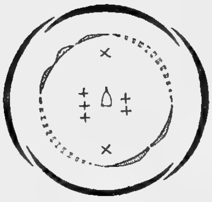
Plate VIII.

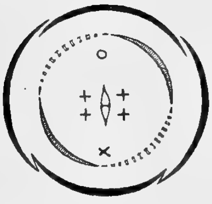

LXXI

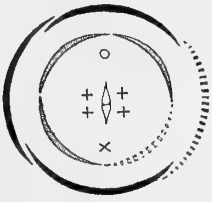

LXXII

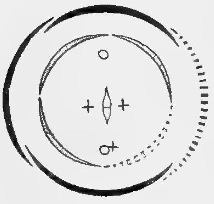

LXXIII

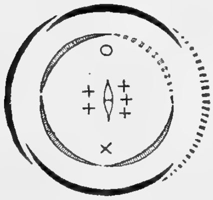

LXXIV

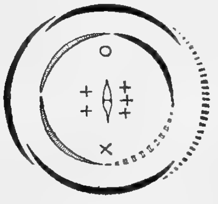

LXXV

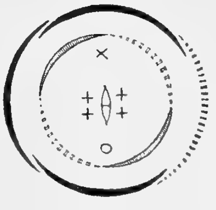

LXXVI


LXXVII


LXXVIII


LXXIX


LXXX


LXXXI

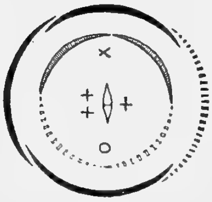
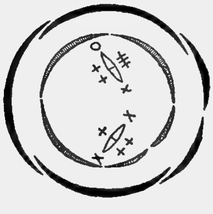
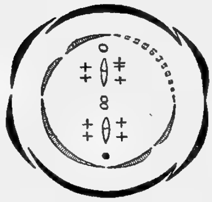
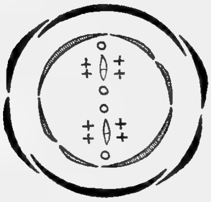
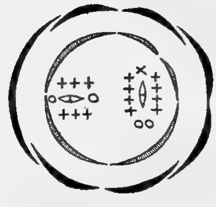

LXXXII

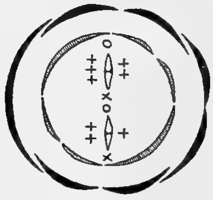
Plate X.

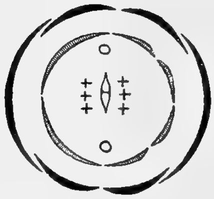

XCV

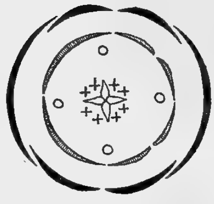

XCVI

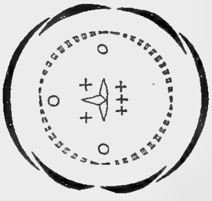

XCVII

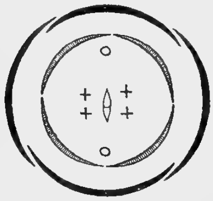

XCVIII

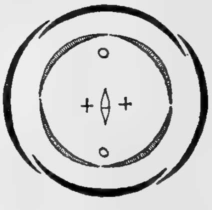

XCIX

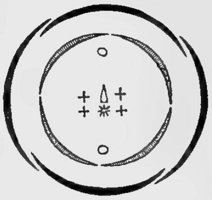

C


CI


CII


CX *B*


CIX *B*


CIX *A*

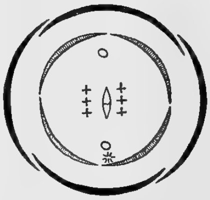
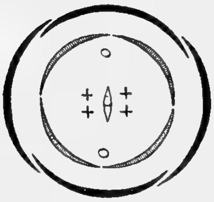
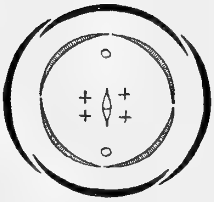
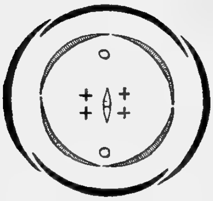
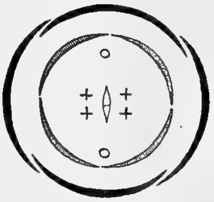

CX *A*

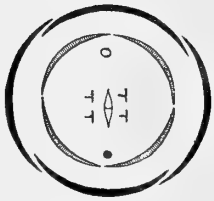
Plate XI.

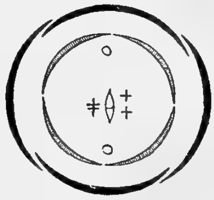

CIII *A*

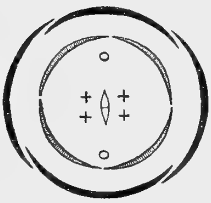

CIV *A*

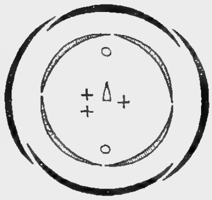

CV *A*

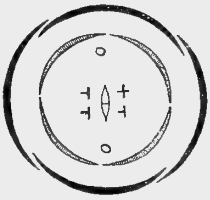

CIII *B*

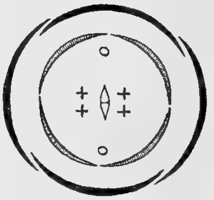

CIV *B*

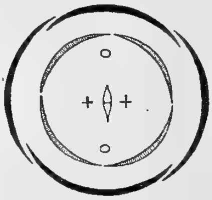

CV *B*

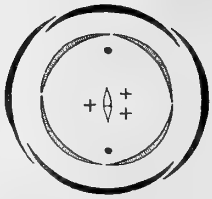

CVI *A*


CVI *B*


CVIII *A*


CVII *A*


CVII *B*


CVIII *B*

NOCHMALS ÜBER "THE ELIMINATION OF SPURIOUS CORRELATION DUE TO POSITION IN TIME OR SPACE."

Von O. ANDERSON, St. Petersburg, Rußland.

1. Im Aprilheft der *Biometrika*, hat "Student" gezeigt*, daß das von Cave und Hooker vorgeschlagene Verfahren, den Korrelationskoeffizienten zweier oscillierender Variablen durch Berechnung erster Differenzen (also durch Ersetzung der Reihe $x_1, x_2, \dots x_n$ durch die Reihe: $\Delta' x_1 = x_1 - x_2$, $\Delta' x_2 = x_2 - x_3, \dots \Delta' x_{n-1} = x_{n-1} - x_n$) vom evolutorischen Element zu befreien, eine Verallgemeinerung zuläßt. Das Verfahren ist nämlich, streng genommen, nur dann richtig, wenn die evolutorische Komponente durch eine lineare Gleichung darstellbar ist. Findet letzteres nicht statt, kann also, z. B., jener nur eine parabolische Gleichung höherer Ordnung genügen, so muß man zweite, dritte u.s.w. Differenzen nehmen (also statt $\Delta' x_1, \Delta' x_2 \dots$ nehme man $\Delta'' x_1 = \Delta' x_1 - \Delta' x_2$, $\Delta'' x_2 = \Delta' x_2 - \Delta' x_3, \dots$ etc.) und danach Korrelationskoeffizienten berechnen. Letztere können bald einen konstanten Grenzwert erreichen, der das gewünschte Resultat darstellt.

Unterzeichneter ist schon vor etwa 2 Jahren zu ähnlichen Schlüssen gekommen. Durch von ihm unabhängige Gründe wurde er aber bis jetzt vom Drucke seiner diesbezüglichen Schrift abgehalten. Da er bei seiner Untersuchung Wege einschlägt, die von denen des "Students" sehr verschieden sind, und auch zu manchen Schlüssen kommt, welche letzterem unbekannt geblieben zu sein scheinen, so könnte vielleicht eine kurzgefaßte Darstellung der wichtigsten Resultate seiner Untersuchung für die Leser der *Biometrika* von einigem Interesse sein.

2. *Methode.* Die englische statistische Schule vernachlässigt in ihren Untersuchungen ein Verfahren, das von russischen und deutschen Gelehrten oft angewandt wird (Tchebycheff, Markoff, v. Bortkiewicz, u.s.w.) und neben großer Strenge und Exaktheit noch den Vorzug hat recht elementar zu sein—die Methode der mathematischen Erwartung nämlich. Mathematische Erwartung einer Größe (A) heißt bekanntlich soviel als das Produkt aus dieser Größe und ihrer Wahrscheinlichkeit (w), also Aw. Wenn eine Variable eine Reihe einander ausschließender

---

* *Biometrika*, Vol. x. Part 1, S. 179, "The Elimination of Spurious Correlation due to Position in Time or Space." By "Student."

Größen annehmen kann, so ist deren math. Erwartung als die Summe der Erwartungen aller dieser Größen definiert. Wir werden hier die mathem. Erwartung überall durch das Symbol $E(\ )$ bezeichnen. $E(A)$ ist also, z. B., gleich Aw.

Die hauptsächlichsten Sätze über mathematische Erwartungen dürften als bekannt angenommen werden. Um aber die Nachprüfung der Formeln dieser Schrift zu erleichtern, werden wir hier die für uns wichtigsten Sätze noch kurz andeuten:

(1) $E(x+y-z+u-t\ldots)=E(x)+E(y)-E(z)+E(u)-E(t)\ldots.$

(2) Wenn $x, y, z, \ldots$ von einander unabhängig sind, so ist

$$E(x\,.\,y\,.\,z\,.\ldots)=E(x)\,.\,E(y)\,.\,E(z)\ldots.$$

(3) $E(kx)=kE(x)$, wo k const. ist; und daher auch:

$$E(k)=k.$$

(4) Wenn eine Variable X die Werte $x_1, x_2, \ldots x_n$ annehmen kann, so ist die Wahrscheinlichkeit W, daß die Differenz $x_i-E(x)$ zwischen den Grenzen $-\alpha\sqrt{E(x^2)-[E(x)]^2}$ und $+\alpha\sqrt{E(x^2)-[E(x)]^2}$ enthalten sei, größer als $1-\frac{1}{\alpha^2}$, wo α größer als 1 sein muß (ein Theorem von Tchebycheff).

In unserer Untersuchung werden wir überall statt des wahrscheinlichsten Wertes einer Größe deren mathematische Erwartung berechnen.

Bestimmen wir zuerst, wie sich der Korrelationskoeffizient zweier oscillierender Reihen verhält, wenn man deren Größen durch Differenzen $\Delta' x$, $\Delta'' x$, $\Delta''' x\ldots$, $\Delta' y$, $\Delta'' y$, $\Delta''' y, \ldots$ ersetzt, und darauf untersuchen wir die Frage von den Grenzen der Anwendbarkeit der verallgemeinerten Cave-Hookerschen Methode. Um Raum zu sparen, werden wir nur die endgiltigen Resultate der Berechnungen anführen, ausgenommen die 3 ersten Formeln, deren Bestimmung als Beispiel der Rechnungsmethode dienen möge.

3. *Definition.* Unter einer oscillatorischen Reihe werden wir eine solche Reihe

$$x_1,\ x_2,\ x_3, \ldots,\ x_i, \ldots,\ x_n$$

verstehen, bei der

$$E(x_1)=E(x_2)=\ldots=E(x_i)=\ldots=E(x_n)=E(x)=\text{const.}$$

und alle einzelnen Glieder von einander völlig unabhängig sind, so daß

$$E(x_i x_j)=E(x_i)\,.\,E(x_j),\ \text{wobei}\ i \neq j.$$

Solchen Bedingungen würde *zum Beispiel* eine Reihe genügen, deren Glieder die Resultate einer Versuchsreihe mit konstanter Wahrscheinlichkeit darstellen, etwa Resultate von Ziehungen aus einer Urne mit m weißen und n schwarzen Kugeln.

4. *Mittleres Fehlerquadrat.*

Bezeichnen wir $x_i - E(x)$ durch ξ_i, so ist

$$E(\xi_i) = E[x_i - E(x)] = E(x) - E(x) = 0.$$

Da die einzelnen ξ von einander völlig unabhängig sind, so ist

$$E(\xi_i \xi_j) = E(\xi_i) \,.\, E(\xi_j) = 0.$$

Das mittlere Fehlerquadrat der Reihe x ist gleich

$$\frac{\sum\limits_1^n [x_i - E(x)]^2}{n}.$$

Seine mathematische Erwartung wollen wir (nicht ganz in Übereinstimmung mit der üblichen Bezeichnung) σ_x^2 nennen.

$$\sigma_x^2 = E\left\{\frac{\sum\limits_1^n [x_i - E(x)]^2}{n}\right\} = \frac{1}{n}\left\{E\left[\sum_1^n x_i^2\right] - 2E\left[\sum_1^n x_i\right].E(x) + n\,[E(x)]^2\right\}$$
$$= E(x^2) - [E(x)]^2,$$

ein Ausdruck, der oben im Satze 4 (§ 2) unter dem Zeichen der Quadratwurzel steht.

Andererseits ist aber $E\left\{\frac{\sum\limits_1^n [x_i - E(x)]^2}{n}\right\}$ gleich $E\left\{\frac{\sum\limits_1^n \xi_i^2}{n}\right\}$ und daher

$$\sigma_x^2 = E(\xi^2).$$

Untersuchen wir den Ausdruck $E\left[\sum\limits_1^n (x_i - M_x)^2\right]$, wo M_x das arithmetische Mittel der Reihe x, also $\frac{\sum\limits_1^n x_i}{n}$ bedeutet. Da

$$x_i - M_x = E(x) + \xi_i - \frac{E(x) + \xi_1 + E(x) + \xi_2 + \ldots + E(x) + \xi_n}{n} = \xi_i - M_\xi$$

(wenn man M_ξ für $\frac{\sum\limits_1^n \xi_i}{n}$ einsetzt), so haben wir:

$$E\left[\sum_1^n (x_i - M_x)^2\right] = E\left[\sum_1^n (\xi_i - M_\xi)^2\right] = E\left[\sum_1^n \xi_i^2 - nM_\xi^2\right]$$
$$= nE(\xi^2) - nE(M_\xi^2) = nE(\xi^2) - nE\left\{\frac{\sum\limits_1^n \xi_i^2 - 2\Sigma\xi_i\xi_j}{n^2}\right\} = nE(\xi^2) - E(\xi^2).$$

$$E\left[\sum_1^n (x_i - M_x)^2\right] = (n-1)\,E(\xi^2).$$

Um $E(\xi^2)$ zu bekommen, muß man diesen Ausdruck durch $(n-1)$ dividieren.

Es ist also auch

$$\sigma_x^2 = E\left\{\frac{\sum\limits_1^n (x_i - M_x)^2}{n-1}\right\}.$$

Um das Fehlerquadrat $\sigma^2_{\Delta' x}$ für die erste Differenz $\Delta' x$ zu erhalten, berücksichtigen wir, daß

$$\Delta' x_i = (x_i - x_{i+1}) = (\xi_i - \xi_{i+1})$$

und

$$E(\Delta' x_i) = E(\xi_i) - E(\xi_{i+1}) = 0.$$

Daher haben wir:

$$\sigma^2_{\Delta' x} = E\left\{\frac{\sum\limits_1^{n-1}[\Delta' x_i - E(\Delta' x)]^2}{n-1}\right\} = E\left\{\frac{\sum\limits_1^{n-1}(\xi_i - \xi_{i+1})^2}{n-1}\right\}$$

$$= \frac{1}{n-1}\left\{E\left[\sum_1^{n-1}\xi_i^2\right] - 2E\left[\sum_1^{n-1}\xi_i\xi_{i+1}\right] + E\left[\sum_2^{n}\xi_i^2\right]\right\}$$

$$= \frac{1}{n-1}\{(n-1)E(\xi_i^2) - 0 + (n-1)E(\xi_i^2)\} = 2E(\xi^2).$$

Es ist also

$$\sigma^2_{\Delta' x} = 2\sigma_x^2.$$

Nach demselben Rechenschema ergiebt sich für das mittlere Fehlerquadrat

der zweiten Differenz $\Delta'' x_i$ der Ausdruck $6\sigma_x^2$
,, dritten ,, $\Delta''' x_i$,, ,, $20\sigma_x^2$
,, vierten ,, $\Delta^{(iv)} x_i$,, ,, $70\sigma_x^2$
....
,, k-ten ,, $\Delta^{(k)} x_i$,, ,, $\frac{2k!}{k!\,.\,k!}\sigma_x^2$.

Wir können daher folgende Gleichung aufstellen

$$\sigma_x^2 = \frac{\sigma^2_{\Delta' x}}{2} = \frac{\sigma^2_{\Delta'' x}}{6} = \frac{\sigma^2_{\Delta''' x}}{20} = \frac{\sigma^2_{\Delta^{(iv)} x}}{70} = \ldots = \frac{\sigma^2_{\Delta^{(k)} x}}{\left(\frac{2k!}{k!\,k!}\right)} = E(\xi^2), \quad \ldots\ldots\ldots(1)$$

welche exakt ist, und folgende

$$\frac{\sum\limits_1^n (x_i - M_x)^2}{n-1} = \frac{\sum\limits_1^{n-1} \Delta' x_i^2}{2(n-1)} = \frac{\sum\limits_1^{n-2} \Delta'' x_i^2}{6(n-2)} = \frac{\sum\limits_1^{n-3} \Delta''' x_i^2}{20(n-3)} = \frac{\sum\limits_1^{n-4} \Delta^{(iv)} x_i^2}{70(n-4)} = \ldots = \frac{\sum\limits_1^{n-k} \Delta^{(k)} x_i^2}{\frac{2k!}{k!\,k!}(n-k)} = \sigma_x^{2}*, \quad \ldots\ldots\ldots\ldots(1\,a)$$

welche nur annähernd richtig ist.

* Es ist vorteilhafter $\dfrac{\sum\limits_1^{n-k} \Delta^{(k)} x^2}{\frac{2k!}{k!\,k!}(n-k)}$ und nicht $\dfrac{\sum\limits_1^{n-k} \left(\Delta^{(k)} x - M_{\Delta^{(k)} x}\right)^2}{\frac{2k!}{k!\,k!}(n-k)}$ zu berechnen.

5. *Das mittlere Produkt* $\frac{\sum\limits_1^n x_i y_i}{n}$.

Wenn zwei Reihen

$$x_1,\ x_2,\ x_3,\ \ldots\ x_n,$$

und

$$y_1,\ y_2,\ y_3, \ldots\ y_n,$$

beide im Sinne des § 3 oscillatorisch sind, und eine Korrelation nur zwischen Größen mit gleichen Indexen, also zwischen x_1 und y_1, x_2 und y_2, x_3 und y_3, u.s.w. bestehen kann, so ist es leicht ersichtlich, daß

$$E\,\{[x_i - E(x)]\,[y_j - E(y)]\} = 0, \text{ wenn } i \neq j.$$

Bezeichnen wir $y_i - E(y)$ durch ψ_i, $x_i - E(x)$ wieder durch ξ_i, so finden wir leicht folgende Ausdrücke:

$$p_{xy} = E\left\{\frac{\sum\limits_1^n \xi_i \psi_i}{n}\right\} = E(\xi_i \psi_i),$$

$$E\left\{\frac{\sum\limits_1^n (x_i - M_x)(y_i - M_y)}{n-1}\right\} = E(\xi_i \psi_i) = p_{xy},$$

$$p_{\Delta' x \Delta' y} = E\left\{\frac{\sum\limits_1^{n-1} \Delta' x_i \Delta' y_i}{n-1}\right\} = 2 p_{xy},$$

$$p_{\Delta'' x \Delta'' y} = E\left\{\frac{\sum\limits_1^{n-2} \Delta'' x_i \Delta'' y_i}{n-2}\right\} = 6 p_{xy},$$

....

$$p_{\Delta^{(k)} x,\ \Delta^{(k)} y} = E\left\{\frac{\sum\limits_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{n-k}\right\} = \frac{2k\,!}{k\,!\,.\,k\,!}\, p_{xy}.$$

Wir können daher wieder zwei Gleichungssysteme:

$$p_{xy} = \frac{p_{\Delta' x \Delta' y}}{2} = \frac{p_{\Delta'' x \Delta'' y}}{6} = \frac{p_{\Delta''' x \Delta''' y}}{20} = \frac{p_{\Delta^{(iv)} x \Delta^{(iv)} y}}{70} = \ldots = \frac{p_{\Delta^{(k)} x \Delta^{(k)} y}}{\frac{2k!}{k!\,k!}} = E(\xi_i \psi_i), \quad \ldots (2)$$

und

$$\frac{\sum\limits_1^n [x_i - E(x)]\,[y_i - E(y)]}{n-1} = \frac{\sum\limits_1^{n-1} \Delta' x_i \Delta' y_i}{2(n-1)} = \frac{\sum\limits_1^{n-2} \Delta'' x_i \Delta'' y_i}{6(n-2)} = \frac{\sum\limits_1^{n-3} \Delta''' x_i \Delta''' y_i}{20(n-3)}$$

$$= \frac{\sum\limits_1^{n-4} \Delta^{(iv)} x_i \Delta^{(iv)} y_i}{70(n-4)} = \ldots = \frac{\sum\limits_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{\frac{2k!}{k!\,k!} \cdot (n-k)} = p_{xy}, \quad \ldots (2a)$$

aufstellen, von denen das erste exakt und das zweite angenähert ist, und welche den Gleichungen für σ_x^2 (also auch für σ_y^2) des § 4 genau analog sind.

6. *Das Fehlerquadrat der Fehlerquadrate.*

Betrachten wir jetzt den Bereich der Schwankungen der Größen der Systeme (1a) und (2a) um deren mathematisch zu erwartenden Größen in (1) und (2). Mit anderen Worten, gehen wir (mit Rücksicht auf Satz 4 § 2) zur Darstellung der math. Erwartung der Fehlerquadrate der genannten Größen über.

Für $\frac{\sum_1^n [x_i - E(x)]^2}{n}$ ergiebt sich das Fehlerquadrat $\frac{E(\xi^4) - [E(\xi^2)]^2}{n}$.

Für $\frac{\sum_1^n [x_i - M_x]^2}{n-1}$ ergiebt sich das Fehlerquadrat $\frac{E(\xi^4) - [E(\xi^2)]^2}{n} + \frac{2[E(\xi^2)]^2}{n(n-1)}$.

Für $\frac{\sum_1^{n-1} [\Delta' x_i]^2}{2(n-1)}$ ergiebt sich das Fehlerquadrat

$$\frac{(2n-3)\{E(\xi^4) - [E(\xi^2)]^2\} + 2(n-1)[E(\xi^2)]^2}{2(n-1)^2}.$$

Für $\frac{\sum_1^{n-2} [\Delta'' x_i]^2}{6(n-2)}$ ergiebt sich das Fehlerquadrat

$$\frac{(9n-23)\{E(\xi^4) - [E(\xi^2)]^2\} + (17n-42)[E(\xi^2)]^2}{9(n-2)^2}.$$

Und endlich für $\frac{\sum_1^{n-k} (\Delta^{(k)} x_i)^2}{\frac{2k!}{k!\,k!}(n-k)}$ ergiebt sich der recht komplizierte Ausdruck:

$$\begin{aligned}\frac{1}{A_0^2(n-k)^2}\Big\{&A_0^2(n-2k)\{E(\xi^4) - [E(\xi^2)]^2\} + 4[E(\xi^2)]^2[A_1^2(n-2k+1)\\ &+ A_2^2(n-2k+2) + A_3^2(n-2k+3) + \ldots + A_k^2(n-2k+k)]\\ &+ 2B_0^2\{E(\xi^4) - [E(\xi^2)]^2\} + 8[E(\xi^2)]^2 . (B_1^2 + B_2^2 + \ldots + B^2_{k-1})\Big\}.\end{aligned}$$

Wenn $b_0, b_1, b_2, \ldots b_k$ die Koeffizienten der Zerlegung des Binoms $(1+1)^k$ darstellen, also $b_0 = 1$, $b_1 = k$, $b_2 = \frac{k.(k-1)}{1.2}$, u.s.w., so ist hier

$$A_0^2 = (b_0^2 + b_1^2 + b_2^2 + \ldots + b_k^2)^2 = \left[\frac{2k!}{k!\,k!}\right]^2,$$

$$A_1^2 = (b_0 b_1 + b_1 b_2 + b_2 b_3 + \ldots + b_{k-1} b_k)^2 = \left[\frac{2k!}{k-1!\,k+1!}\right]^2,$$

$$A_2^2 = (b_0 b_2 + b_1 b_3 + \ldots + b_{k-2} b_k)^2 = \left[\frac{2k!}{k-2!\,k+2!}\right]^2,$$

....

$$A_j^2 = (b_0 b_j + b_1 b_{j+1} + \ldots + b_{k-j} b_k)^2 = \left[\frac{2k!}{k-j!\,k+j!}\right]^2,$$

....

$$A_k^2 = 1,$$

$$B_0^2 = (b_0^2)^2 + (b_0^2 + b_1^2)^2 + (b_0^2 + b_1^2 + b_2^2)^2 + \ldots + (b_0^2 + b_1^2 + b_2^2 + \ldots + b^2_{k-1})^2,$$

$$B_1^2 = (b_0 b_1)^2 + (b_0 b_1 + b_1 b_2)^2 + (b_0 b_1 + b_1 b_2 + b_2 b_3)^2 + \ldots + (b_0 b_1 + b_1 b_2 + b_2 b_3 + \ldots + b_{k-2} b_{k-1})^2,$$

....

$$B_i^2 = (b_0 b_i)^2 + (b_0 b_i + b_1 b_{i+1})^2 + (b_0 b_i + b_1 b_{i+1} + b_2 b_{i+2})^2 + \ldots + (b_0 b_i + b_1 b_{i+1} + b_2 b_{i+2} + \ldots + b_{k-i-1} b_{k-1})^2,$$

....

$$B^2_{k-1} = (b_0 b_{k-1})^2.$$

Wenn die Verteilung der x (und dies ist der für uns interessanteste Fall) "normal" ist, so können obige Formeln beträchtlich vereinfacht werden. Da man in diesem Fall $E(\xi^4)$ gleich $3\,[E(\xi^2)]^2$ oder $3\sigma_x^4$ setzen kann, so haben wir:

Für $\dfrac{\sum_1^n (x_1 - E(x))^2}{n}$ das Fehlerquadrat $\dfrac{2\sigma_x^4}{n}$ (vergl. *Biometrika*, II. p. 276).

„ $\dfrac{\sum_1^n (x_i - M_x)^2}{n-1}$ „ „ $\dfrac{2\sigma_x^4}{n-1}$.

„ $\dfrac{\sum_1^{n-1} (\Delta' x_i)^2}{2(n-1)}$ „ „ $\dfrac{(3n-4)\,\sigma_x^4}{(n-1)^2}$ oder angenähert $\dfrac{3\sigma_x^4}{n-1}$.

„ $\dfrac{\sum_1^{n-2} (\Delta'' x_i)^2}{6(n-1)}$ „ „ $\dfrac{(35n-88)\,\sigma_x^4}{9(n-2)^2}$ „ „ $\dfrac{4\sigma_x^4}{n-2}$.

„ $\dfrac{\sum_1^{n-3} (\Delta''' x_i)^2}{20(n-2)}$ „ „ $\dfrac{(231n-843)\,\sigma_x^4}{50(n-3)^2}$ „ „ $\dfrac{5\sigma_x^4}{n-3}$.

Für $\dfrac{\sum_1^{n-k} (\Delta^{(k)} x_i)^2}{\frac{2k!}{k!\,k!}(n-k)}$ endlich kann man das Fehlerquadrat in solcher Form darstellen:

$$\frac{2\sigma_x^4}{(n-k)^2}\left\{(n-k) + 2(n-k-1)\left(\frac{k}{k+1}\right)^2 + 2(n-k-2)\left(\frac{k.(k-1)}{(k+1).(k+2)}\right)^2\right.$$
$$+ 2(n-k-3)\left(\frac{k.(k-1).(k-2)}{(k+1).(k+2).(k+3)}\right)^2 + \ldots$$
$$+ 2(n-k-j)\left(\frac{k.(k-1).(k-2)\ldots(k-j+1)}{(k+1).(k+2).(k+3)\ldots(k+j)}\right)^2 + \ldots$$
$$\left.+ 2(n-k-k)\left(\frac{k.(k-1).(k-2)\ldots 2.1}{(k+1).(k+2).(k+3)\ldots(k+k)}\right)^2\right\}.$$

Es ist also klar, daß zusammen mit dem endlichen Differenzieren die Unsicherheit der Bestimmung von σ_x^2 stetig wächst, anfangs etwa im Verhältnis

$$\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{5} \dots .$$

7. *Das Fehlerquadrat des mittleren Produktes.*

Für $\frac{\sum\limits_1^n \xi_i \psi_i}{n}$ ergiebt sich das Fehlerquadrat $\frac{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2}{n}$.

Für $\frac{\sum\limits_1^n (x_i - M_x)(y_i - M_y)}{n-1}$ ergiebt sich das Fehlerquadrat

$$\frac{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2}{n} + \frac{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2}{n(n-1)}.$$

Für $\frac{\sum\limits_1^{n-1} \Delta' x_i \Delta' y_i}{2(n-1)}$ ergiebt sich das Fehlerquadrat

$$\frac{(2n-3)\{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2\} + (n-1)\{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2\}}{2(n-1)^2}.$$

Im allgemeinen Fall $\frac{\sum\limits_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{\frac{2k!}{k!\,k!}(n-k)}$ erhalten wir für das Fehlerquadrat folgenden Ausdruck:

$$\frac{1}{A_0^2 (n-k)^2} \Big\{ A_0^2 (n-2k)\{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2\}$$
$$+ 2\{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2\}[A_1^2 (n-2k+1) + A_2^2 (n-2k+2) + \dots + A_k^2 (n-2k+k)]$$
$$+ 2B_0^2 \{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2\} + 4\{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2\}(B_1^2 + B_2^2 + \dots + B_{k-1}^2) \Big\},$$

wo die Koeffizienten $A_0^2, A_1^2, \dots B_0^2, B_1^2, \dots$ dieselbe Bedeutung haben, wie in § 6.

Wenn x_i und y_i einander vollkommen gleich sind, so ist

$$[E(\xi_i \psi_i)]^2 = \{E(\xi^2)\}^2 = \sigma_x^4; \quad E(\xi_i^2 \psi_i^2) = E(\xi_i^4); \quad \sigma_x^2 \sigma_y^2 = \sigma_x^4,$$

und obiger Ausdruck fällt mit dem in § 6 zusammen.

Für den Fall der "normalen" Verteilung können wir auch alle diese Ausdrücke beträchtlich vereinfachen, besonders wenn wir $E(\xi_i \psi_i)$ und $E(\xi_i^2 \psi_i^2)$ als Funktionen von r_{xy} darstellen. Ohne aber hier darauf einzugehen, wollen wir jetzt über den Korrelationskoeffizienten ins Klare kommen.

8. *Definition des Korrelationskoeffizienten.*

Der Korrelationskoeffizient R wird gewöhnlich nach der Formel

$$R_0 = \frac{\sum\limits_1^n (x_i - M_x)(y_i - M_y)}{\sqrt{\sum\limits_1^n (x_i - M_x)^2 \sum\limits_1^n (y_i - M_y)^2}} \text{ berechnet.}$$

Zu welchem Ausdruck ist er nun als empirische Annäherung aufzufassen,

$$\text{zu } E\left\{\frac{\sum\limits_1^n (x_i - M_x)(y_i - M_y)}{\sqrt{\sum\limits_1^n (x_i - M_x)^2 \sum\limits_1^n (y_i - M_y)^2}}\right\} \text{ oder zu } \frac{E\left[\sum\limits_1^n (x_i - M_x)(y_i - M_y)\right]}{\sqrt{E[\sum\limits_1^n (x_i - M_x)^2] . E[\sum\limits_1^n (y_i - M_y)^2]}} ?$$

Beide Formeln sind durchaus nicht mit einander zu identifizieren und fallen nur in erster Annäherung zusammen. Da die zweite aber bedeutend leichter zu handhaben ist und dies auch mehr den üblichen Rechnungsmethoden der englischen Schule entspricht, so *definieren* wir r_{xy} als

$$\frac{E\left[\sum\limits_1^n (x_i - M_x)(y_i - M_y)\right]}{\sqrt{E\left[\sum\limits_1^n (x_i - M_x)^2\right] . E\left[\sum\limits_1^n (y_i - M_y)^2\right]}}.$$

Anders ausgedrückt ist $r_{xy} = \frac{p_{xy}}{\sigma_x \sigma_y}$, wo p_{xy}, σ_x, σ_y die Bedeutungen haben, welche wir ihnen oben in §§ 4 und 5 beigemessen haben.

9. *Das Verhalten des Korrelationskoeffizienten zweier oscillierender Reihen x und y, wenn man deren Größen durch Differenzen ersetzt.*

Für die k-te endliche Differenz von x und y haben wir

$$r_{\Delta^{(k)}x,\, \Delta^{(k)}y} = \frac{E\left(\sum\limits_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i\right)}{\sqrt{E\left[\sum\limits_1^{n-k} \Delta^{(k)} x_i^2\right] . E\left[\sum\limits_1^{n-k} \Delta^{(k)} y_i^2\right]}} = \frac{p_{\Delta^{(k)}x,\, \Delta^{(k)}y}}{\sqrt{\sigma^2_{\Delta^{(k)}x} \cdot \sigma^2_{\Delta^{(k)}y}}} = \frac{\frac{2k!}{k!\,.\,k!} p_{xy}}{\sqrt{\frac{2k!}{k!\,k!}\sigma_x^2 \cdot \frac{2k!}{k!\,k!}\sigma_y^2}},$$

$$r_{\Delta^{(k)}x,\, \Delta^{(k)}y} = \frac{p_{xy}}{\sigma_x \sigma_y} = r_{xy}.$$

Wir haben also ganz allgemein das *genaue* Resultat:

$$r_{xy} = r_{\Delta' x,\, \Delta' y} = r_{\Delta'' x,\, \Delta'' y} = r_{\Delta''' x,\, \Delta''' y} = \ldots = r_{\Delta^{(k)}x,\, \Delta^{(k)}y}.$$

Da aber diese r unbekannt bleiben und wir für ein beliebiges $r_{\Delta^{(i)}x,\, \Delta^{(i)}y}$ nur dessen Annäherungsformel $R_i = \frac{\sum\limits_1^{n-i} \Delta^{(i)} x_i \Delta^{(i)} y_i}{\sqrt{\sum\limits_1^{n-i} \Delta^{(i)} x_i^2 . \sum\limits_1^{n-i} \Delta^{(i)} y_i^2}}$ kennen, so müssen wir wiederum feststellen, inwiefern man sich in der Praxis auf die Übereinstimmung der empirischen Koeffizienten mit deren mathematischen Erwartungen verlassen kann, wie groß also die Unsicherheit ihrer Bestimmung zu schätzen ist.

10. *Mittleres Fehlerquadrat des Korrelationskoeffizienten der endl. Differenzen zweier Reihen.*

Aus der Formel $R_k = \dfrac{\sum\limits_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{\sqrt{\sum\limits_1^{n-k} \Delta^{(k)} x_i^2 \, . \sum\limits_1^{n-k} \Delta^{(k)} y_i^2}}$, kann man folgenden Ausdruck ableiten:

$$\frac{R_k - r_{xy}}{r_{xy}} = \frac{\dfrac{\sum\limits_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{n-k} - p_{\Delta^{(k)}x,\, \Delta^{(k)}y}}{p_{\Delta^{(k)}x\, \Delta^{(k)}y}} - \frac{\dfrac{\sum\limits_1^{n-k} \Delta^{(k)} x_i^2}{n} - \sigma^2_{\Delta^{(k)}x}}{2\sigma^2_{\Delta^{(k)}x}} - \frac{\dfrac{\sum\limits_1^{n-k} \Delta^{(k)} y_i^2}{n} - \sigma^2_{\Delta^{(k)}y}}{2\sigma^2_{\Delta^{(k)}y}},$$

der nur in erster Annäherung und, wenn alle 4 Brüche des Ausdrucks echte sind, richtig ist; dies ist bei großem n der Fall. Und ferner ergiebt sich daraus die Formel:

$$\begin{aligned} \sigma^2_{R_k} = \frac{(1 - r^2_{xy})^2}{(n-k)^2} \Big\{ & (n-k) + 2(n-k-1)\left(\frac{k}{k+1}\right)^2 + 2(n-k-2)\left(\frac{k.(k-1)}{(k+1).(k+2)}\right)^2 \\ & + 2(n-k-3)\left(\frac{k.(k-1).(k-2)}{(k+1).(k+2).(k+3)}\right)^2 + \ldots \\ & + 2(n-k-k)\left(\frac{k.(k-1).(k-2)\ldots 2.1}{(k+1).(k+2).(k+3)\ldots(k+k)}\right)^2 \Big\} \end{aligned}$$

(vergl. dazu die Formel für $\dfrac{\sum\limits_1^{n-k} (\Delta^{(k)} x_i)^2}{\dfrac{2k!}{k!.k!}(n-k)}$ in § 6).

Die Formel für $\sigma^2_{R_k}$ ist immer nur dann giltig, wenn man

$$\begin{aligned} \frac{r^2_{xy}}{n} \Big\{ & \frac{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2}{[E(\xi_i \psi_i)]^2} + \frac{E(\xi^4) - [E(\xi^2)]^2}{4[E(\xi^2)]} + \frac{E(\psi^4) - [E(\psi^2)]^2}{4[E(\psi^2)]^2} \\ & - \frac{E(\xi_i^3 \psi_i) - E(\xi_i \psi_i) . E(\xi^2)}{E(\xi_i \psi_i) . E(\xi^2)} - \frac{E(\xi_i \psi_i^3) - E(\xi_i \psi_i) . E(\psi^2)}{E(\xi_i \psi_i) . E(\psi^2)} \\ & + \frac{E(\xi_i^2 \psi_i^2) - E(\xi^2) . E(\psi^2)}{2E(\xi^2) . E(\psi^2)} \Big\} \end{aligned}$$

gleich $\dfrac{(1 - r^2_{xy})^2}{n}$ setzen darf (vergl. *Biometrika*, Vol. IX. p. 4).

Aus der Formel für $\sigma^2_{R_k}$ erhalten wir:

$$\sigma^2_{R_0} = \frac{(1 - r^2_{xy})^2}{n},$$

$$\sigma^2_{R_1} = \frac{(1 - r^2_{xy})^2}{n-1} \cdot \frac{3n-4}{2(n-1)},$$

$$\sigma^2_{R_2} = \frac{(1 - r^2_{xy})^2}{n-2} \cdot \frac{35n-88}{18(n-2)},$$

$$\sigma^2_{R_3} = \frac{(1 - r^2_{xy})^2}{n-3} \cdot \frac{231n-843}{100(n-3)},$$

u.s.w.

Die Fehlerquadrate der Korrelationskoeffizienten aufeinanderfolgender Differenzenordnungen verhalten sich folglich zueinander ungefähr wie 2 : 3 : 4 : 5 ....

Die Unsicherheit wächst also mit zunehmender Differenzenordnung etwa im Verhältnis

$$\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{5}, \ldots .$$

11. *Korrelationskoeffizient zweier zusammengesetzter Reihen, die aus oscillatorischen und evolutorischen Elementen bestehen.*

Da "Student" diese Frage treffend dargelegt hat, können wir uns kurz fassen. Wenn wir in Betracht ziehen, daß für uns die evolutorische Komponente einer Reihe schon dann in der Praxis verschwunden ist, wenn sie im Verhältnis zur oscillatorischen Komponente so klein geworden ist, daß sie nur die 3tten, 4ten, u.s.w. Zahlenstellen des Ausdruckes für R beeinflussen kann, so kommen wir zum Schluß, daß nicht nur Komponenten, die durch eine Parabel höherer Ordnung darstellbar sind, sondern auch solche, denen nur transzendentale Gleichungen (z. B. Sinusreihen) genügen, beim endlichen Differenzieren eliminiert werden. Ja mehr noch, man kann beweisen, daß überhaupt alle mehr oder minder "glatten Reihen," alle bei denen eine genügende positive Korrelation zwischen den Nachbargliedern bemerkbar ist, für die Praxis beim endlichen Differenzieren verschwinden. Das verallgemeinerte Cave-Hookersche Verfahren ist daher augenscheinlich ein sehr universales Mittel, die Korrelation oscillatorischer Elemente aus zusammengesetzten Reihen herauszuschälen. Es hat aber einen Haken, auf den hier noch hingewiesen werden muß.

12. *Kann man aus dem Verhalten der Reihe $R_0, R_1, R_2, \ldots R_k$ bestimmen, ob wir den Korrelationskoeffizienten rein oscillatorischer Reihen vor uns haben?* "Student" scheint zu glauben, daß wenn irgendein R_i seinem Vorgänger R_{i-1} gleich ist, wir es sicher mit dem Korrelationskoeffizienten oscillierender Elemente zu tun haben. Vor einem solchen Schluß ist nachdrücklich zu warnen. Wie es meine (für diesen Artikel etwas zu langwierigen) Berechnungen zeigen, können zwei Nachbarkoeffizienten R_i, R_{i-1} auch bei stark evolutorischen Reihen einander ungefähr gleich sein, und die Wahrscheinlichkeit eines solchen Zusammentreffens ist gar nicht sehr gering einzuschätzen. Nur wenn wir, von irgendeinem R_j angefangen, immer dieselbe Größe für R erhalten, also $R_j = R_{j+1} = R_{j+2} = R_{j+3} =$, wird ein solcher Schluß berechtigt sein, und je länger die Reihe gleicher R, desto *wahrscheinlicher* wird dieser Schluß.

STATISTICAL NOTES ON THE INFLUENCE OF EDUCATION IN EGYPT.

BY M. HOSNY, M.A., B.Sc.

The statistical returns for Egypt are—as compared with European data—still in a somewhat elementary stage. Age-distributions are of very little value, and in the case of infantile mortality we have only information for certain towns. Further, in the larger towns there is a considerable cosmopolitan element, which gives them a widely different character from the often sparsely populated rural and desert districts. Education is not compulsory, and schools and literacy are largely confined to Cairo, Alexandria and the Canal Government, even when we exclude all foreign scholars. In the same way criminality* preponderates, in an inverse order it is true, in these three districts, but it is not absolutely certain whether this is due to their more efficient policing, to the presence of more foreigners, or to a real absence of crime in the rural populations. Crime does not appear to arise in Egypt from poverty or drunkenness, two of the main factors of its origin in Western Europe. The criminal, indeed, is rarely habitual; he is an amateur, rather than a professional, and criminals are more often well-to-do, their crimes arising from motives of revenge or passion.

The fact that criminality in Egypt is highly correlated with literacy and scholarship would be noteworthy and might possibly be used as an argument against education, did not the association of crime and education arise from the prevalency of both in the more populated districts, where again we find the greatest abundance of foreigners. Naturally such questions arise as:

(i) Are the foreigners—and if so, which section of them—to any extent responsible for the prevalence of crime in the districts frequented by them?

(ii) If we allow for urban conditions, will there still be found a high association of crime and education?

It is perfectly easy to obtain from the Egyptian Census—we used that of 1907—the number of foreigners of each denomination in the various *Egyptian*

---

* We understand by "criminality" in this paper, not commission of but conviction for crime.

governments. The only difficulty here was the presence of British troops in Cairo and Alexandria, which placed that nationality in an anomalous position. These were estimated approximately and subtracted. The following groups of foreigners were then dealt with: (*a*) Ottomans, (*b*) British subjects, French, Austrians, Germans and Russians*, (*c*) Greeks, (*d*) Italians. The Greeks and Italians were separated from the general European group (*b*), because they are largely differentiated, the Greeks being frequently small traders and the Italians often manual workers. Their large numbers also justified a separate classification.

Table I gives the foreigners per 10,000 in the 17 Egyptian districts we were able to deal with. It will be noted that the Greeks far outnumber other foreigners, but that all foreigners are concentrated in the Cairo, Alexandria and Canal governments.

TABLE I.

Foreigners per 10,000 *and Population per sq. kilometre.*

Governments	Ottomans	Europeans other than Greeks and Italians	Greeks	Italians	Population per sq. kilometre
Cairo	453	312	298	204	6060
Alexandria	661	514	745	482	6780
Canal†	416	583	846	445	7666
Behera†	43	23	31	11	178
Charkieh	29	5	24	1	257
Dakahlieh and Damietta	18	5	18	3	346
Gharbieh	2	0	2	0	226
Kalliuhieh	8	3	13	2	469
Menufieh	2	1	7	0	618
Assiut†	4	2	3	1	454
Assuan	8	5	13	4	533
Beni Suef	15	6	9	1	351
Fayoum	10	3	4	0	255
Gerga	2	0	2	0	532
Guizeh	6	6	4	3	447
Kena†	5	4	4	2	339
Minia†	10	5	7	1	458

It was far more difficult to obtain a measure of urban conditions. We had to take very rough measures of the density of the population, because the limits of certain areas are too vaguely defined to be of any service. El Arish has been excluded from the Canal district, Suez and Sinai have also been excluded as there is no enumeration of them with respect to criminality, literacy and scholarship. These densities, with such value as they have, are given in the last column of Table I.

---

* The contributions from other smaller nationalities were omitted.

† Various approximations and omissions occur in these cases in obtaining density.

Table II provides the number of male criminals per 1000 of the male population, the literacy or number of male persons able to read and write per 100 of the male population*, and the number of male scholars aged 5 to 19 per 100 of the native boys of those ages†.

TABLE II. *Educational and Criminal Indices.*

Governments	Male Criminals per 1000 males	Literacy per 100 males	Male Scholars 5—19, per 100 boys of those ages
Cairo	12·90	28·03	30·20
Alexandria	14·15	30·09	19·99
Canal and El Arish ...	22·30	23·39	8·54
Behera	5·30	9·29	1·01
Charkieh	4·20	9·09	1·66
Dakahlieh and Damietta	4·35	8·18	1·76
Gharbieh	5·65	8·22	3·04
Kalliuhieh	6·65	8·13	1·39
Menufieh	3·85	8·45	1·06
Assiut	5·45	7·01	4·02
Assuan	4·20	7·68	0·82
Beni Suef	5·70	8·42	2·03
Fayoum	6·85	6·54	1·85
Gerga	3·95	5·84	2·22
Guizeh	5·10	6·38	1·15
Kena	3·45	5·34	1·66
Minia	5·20	7·25	3·13

We shall use the following symbols to denote the factors which occur in Tables I and II:

$O =$ Ottomans, $G =$ Greeks, $I =$ Italians,

$E =$ Europeans other than Greeks and Italians.

$C =$ Criminality, $L =$ Literacy, $S =$ Scholarship, $D =$ Density of Population.

Each government was treated as of equal weight, although the populations vary from 233,000 in Assuan to 1,485,000 in Gharbieh. The standard-deviations and product-moments were found without grouping. The following results were obtained:

Means	Standard Deviations	Correlations
$m_C = 7{\cdot}015$,	$\sigma_C = 4{\cdot}791$,	$r_{CL} = +{\cdot}8450 \pm {\cdot}0468$,
$m_L = 11{\cdot}019$,	$\sigma_L = 7{\cdot}641$,	$r_{CS} = +{\cdot}6242 \pm {\cdot}0999$,
$m_S = 5{\cdot}031$,	$\sigma_S = 7{\cdot}735$,	$r_{LS} = +{\cdot}9028 \pm {\cdot}0303$,
$m_D = 1528$,	$\sigma_D = 2475{\cdot}5$,	———

Correlations:

$r_{DC} = +{\cdot}9614 \pm {\cdot}0124$, $r_{DS} = +{\cdot}8097 \pm {\cdot}0566$, $r_{DL} = +{\cdot}9563 \pm {\cdot}0138$.

Now at first sight these results would seem to indicate a very bad influence of education on crime. Where literacy and scholarship are greatest, there criminals

---

* *Egyptian Census*, 1907, p. 99.

† Foreign male scholars are excluded in the case of Cairo, Alexandria and the Canal. They have no sensible numerical existence elsewhere. Criminals and scholars are taken from the *Annuaire Statistique de l'Egypte*, 1912, pp. 95 and 135.

are most numerous! And a superficial argument might be used to condemn the character of education in Egypt, or education in general. But it will be clear on examination of the isolated values that the observed high correlations arise solely from the urban character in Egypt of both criminality and education. We have endeavoured therefore to correct this by finding the partial correlations for constant density of population.

There now result

$$\begin{aligned} {}_D r_{CS} &= -{\cdot}9554 \pm {\cdot}0143, \\ {}_D r_{CL} &= -{\cdot}9231 \pm {\cdot}0242, \\ {}_D r_{LS} &= +{\cdot}7480 \pm {\cdot}0721. \end{aligned}$$

Thus, while there still remains a quite considerable relation between the prevalence of literacy and scholars for constant density, we find that for a constant degree of urban conditions, the greater the literacy and the greater the amount of education the less will be the criminality. The negative correlations are now even higher than the uncorrected positive ones and of course are markedly significant. While admitting the slender nature of the Egyptian data, we think that this swinging over of the relation of crime and education when we correct for density is suggestive, and it would be of interest to work out similar correlations for states in which the statistics are of a more ample character. It does, however, appear reasonable to assert that there is no evidence to indicate that education leads to criminality—rather the reverse—in Egypt.

We will next consider the influence of the presence of foreigners in Egypt.

We find:

Means	Standard Deviations	Correlations
$m_O = 99{\cdot}53$,	$\sigma_O = 195{\cdot}60$,	$r_{OC} = +{\cdot}8425 \pm {\cdot}0473$,
$m_E = 86{\cdot}88$,	$\sigma_E = 183{\cdot}70$,	$r_{EC} = +{\cdot}9546 \pm {\cdot}0145$,
$m_G = 119{\cdot}41$,	$\sigma_G = 256{\cdot}61$,	$r_{GC} = +{\cdot}9429 \pm {\cdot}0181$,
$m_I = 68{\cdot}24$,	$\sigma_I = 152{\cdot}04$,	$r_{IC} = +{\cdot}9192 \pm {\cdot}0254$.

Correlations:

$$r_{DO} = +{\cdot}9575 \pm {\cdot}0136, \qquad r_{DE} = +{\cdot}9844 \pm {\cdot}0050,$$
$$r_{DG} = +{\cdot}9491 \pm {\cdot}0162, \qquad r_{DI} = +{\cdot}9617 \pm {\cdot}0123.$$

Here, if we judged by the raw correlations only, we must assert that the correlations of crime with the presence of foreigners are so high, that the foreigners must be corrupting the Egyptian population. But again the association only arises because the criminals and foreigners are both prevalent in the big towns. If we correct for density of population, we find the results are very different. Thus we have:

$${}_D r_{OC} = -{\cdot}9811 \pm {\cdot}0061, \qquad {}_D r_{EC} = +{\cdot}1692 \pm {\cdot}1591,$$
$${}_D r_{GC} = +{\cdot}3524 \pm {\cdot}1433, \qquad {}_D r_{IC} = -{\cdot}0713 \pm {\cdot}1628.$$

It is now obvious that the correlation of Europeans other than Greeks and Italians with criminality has become insignificant having regard to its probable error; the correlation of the presence of Italians and criminality is now *negative*, but less than its probable error. Thus of Christians only the presence of the

Greeks may possibly, but not certainly, be detrimental. The Ottomans have now a large negative correlation of a quite significant character, or we might assert that the presence of Ottomans tends to diminish criminality. The Greeks are frequently moneylenders and alcohol dealers, and the Ottomans, especially the Arabs, have among them a good many religious teachers.

We have, however, to note that criminality is greatest in the Canal Government, where Europeans and Greeks are most frequent, while the Ottomans are most numerous in Alexandria, where crime is almost 40 % less than in the Canal Government. To test the influence of the three densely populated governments, we put the Canal proportion of the Ottomans at Cairo, that of Cairo at Alexandria and that of Alexandria at the Canal. There resulted:

$$r_{OC} = + \cdot 9707, \text{ instead of } + \cdot 8425,$$

$$r_{DO} = + \cdot 9870, \text{ instead of } + \cdot 9575,$$

leading to

$$_{D}r_{OC} = + \cdot 4918,$$

or we may safely say, that if the proportions of Ottomans at Alexandria and along the Canal were interchanged, then no relation between the presence of Ottomans and the absence of criminality would exist, indeed the relation would probably be reversed. The prevalence of the Ottomans in Alexandria has been attributed to its more temperate climate. There is certainly a large Ottoman element in Alexandria, there being 21,827 Ottomans out of a population of 332,246, and it is larger than any other foreign element except the Greeks. In Cairo, with 29,516 out of 654,476 inhabitants, the Ottomans exceed any other single foreign element. It is conceivable, therefore, that they may be able to influence the moral tone of those towns. It must be borne in mind, however, that crime is far more frequent in the Cairo and Alexandria governments than in the more purely rural districts, and we can scarcely suppose that Cairo and Alexandria would reach the still higher criminality level of the Canal, were it not for the presence of the Ottomans. In the Canal Government there are exceptional conditions, and we can hardly assume that a transfer of the Ottomans from Alexandria to the Canal would interchange their proportions of criminality. Greeks no doubt flock to the Canal for business purposes, the other Europeans largely for control purposes; the Ottomans, relatively speaking, avoid it. Without further analysis it would not be possible to assert definitely that the presence of Ottomans reduces crime. It may be doubted whether the presence of foreigners, with the possible exception of the Greeks, is really associated with the extent of criminality in Egypt.

A further investigation was undertaken in regard to the possible influence of education on infantile mortality. The birthrate and deathrate in Egypt are both remarkably high. Thus for the years 1899–1909 inclusive the average rates were:

	Births per 1000*	Deaths per 1000	Excess of Births over Deaths
Cairo	40·7	35·7	+ 5·0
Alexandria	38·0	31·8	+ 6·2

* Still births not included.

Many European towns with half the above birthrates have considerably greater excesses of births over deaths.

Unfortunately the infantile mortality is only recorded in Egyptian *towns* and not in the governments at large or in the rural districts. We are obliged, therefore, to deal with these only when considering the relation of education to infantile mortality. From p. 49 of the *Annuaire Statistique de l'Égypte,* 1912, we obtain the infantile mortality for 1911, and note at once how extraordinarily high it stands. From p. 286 of the *Census of Egypt,* 1907, we take the percentage of male literates in the total male population, and from the *Statistique Scolaire,* 1912–1913, p. 74, the percentage of scholars in the total population*. As it was possible that the density of town population might influence the results, we took the number of persons per house, which was about the only social factor available. This will probably represent fairly closely the average size of family. This was taken from the *Census,* 1907, p. 286. The mean, 5·64 persons per house, suggests that the average number of living children can hardly exceed three. The marked relationship that occurs in European towns between gross size of family and infantile mortality cannot be satisfactorily tested on the Egyptian data, because we cannot ascertain the infantile mortality in each size of family. The number of persons to the house is indeed rather a measure of net than gross family, and we only know this as an *average* value for each town. It does not follow that a town with a low number of persons per house is one with small gross families; the low number may be due to the heavy infantile mortality itself. Accordingly the correlation between persons per house and infantile mortality is not necessarily even a measure of the influence of overcrowding on infantile mortality (although this is often supposed to be the case); it is conceivable that a high infantile mortality might be the source of a low number of persons per house, and the unravelling of cause and effect is only possible where we know not only the number of persons per house, but its relation to both the gross and net family of that house.

Let I = Infantile mortality, L = Literacy, S = Scholars, P = Persons per house. Then we have the following results:

Means	Standard Deviations	Correlations
$M_I = 29\cdot30$,	$\sigma_I = 7\cdot608$,	$r_{IL} = -\cdot1040 \pm \cdot1500$,
$M_L = 21\cdot96$,	$\sigma_L = 5\cdot137$,	$r_{IS} = +\cdot5309 \pm \cdot1278$,
$M_S = 5\cdot569$,	$\sigma_S = 2\cdot951$,	$r_{LS} = +\cdot0093 \pm \cdot1508$,
$M_P = 5\cdot643$,	$\sigma_P = 1\cdot428$,	———

Correlations:

$r_{PI} = +\cdot1675 \pm \cdot1487$, $r_{PL} = -\cdot3421 \pm \cdot1437$, $r_{PS} = +\cdot0296 \pm \cdot1508$.

---

* The scholars were taken for 1910–1911, the year of infantile mortality, but this involved the assumption that the foreign scholars were the same in numbers in 1910–11 and 1912–13, probably not a very inaccurate assumption, which in any case affects little more than Cairo, Alexandria, Suez and Ismailia practically. It is the number of Egyptian scholars that is rapidly changing and the scholars dealt with in our ratio are Egyptian only.

TABLE III.

Infantile Mortality. Persons per House and Education.

Town	Infantile Mortality per 100 births	Male Literacy per 100 of population	Scholars per 100 of population	Persons per House
Cairo	32·9	28·03	6·12	4·62
Alexandria ...	26·9	30·09	3·41	8·43
Damietta ...	18·1	7·06	1·89	6·96
Port Said ...	21·0	24·13	2·64	4·14
Ismailia	16·0	28·05	1·22	4·10
Suez	26·9	25·74	0·64	3·85
Benha	29·6	20·54	4·87	5·47
Zagazig	27·9	25·67	6·47	6·07
Tantah	29·6	25·62	9·45	5·18
Mansorah ...	21·4	26·77	6·20	4·60
Chibine El Kom	16·1	18·71	5·10	4·92
Damanhur ...	27·5	19·54	3·55	7·27
Guizeh	35·6	19·23	11·31	6·12
Fayoum	40·1	15·55	4·78	9·34
Beni Suef ...	37·1	21·15	7·19	5·11
Minia	38·2	21·96	8·89	4·96
Assiut	33·6	20·92	11·48	6·00
Sohag	29·0	22·37	5·58	6·15
Kena	37·4	16·76	4·80	5·14
Assuan	41·1	21·31	5·78	4·09

It will be clear from these results that there is no significant relation between the literacy of the male population and infantile mortality. There is also no significant relation between the number of persons to a house and the number of scholars, i.e. it does not appear to be the more crowded towns which have the largest percentage of scholars to the population; Alexandria and Damietta, for example, have considerably more than the mean number of persons to the house and relatively few scholars. On the other hand a larger number of literates marks less crowding. Crowding and infantile mortality are slightly related, but considering the probable error, not with definite significance*.

While literacy has no relation to the infantile deathrate, it is noteworthy that there is a significant correlation (+ ·5309 ± ·1278) between the number of scholars and the infantile deathrate, *which is greater where there is more education.* Now this either suggests that many scholars mean large families and large families correspond to increased infantile mortality, which is usual, or that the towns in which there are the classes who educate their children have a higher infantile deathrate. The only means, and those inadequate, of testing the first

---

* This agrees with the result for overcrowding and infantile mortality in English manufacturing towns, where the correlation is very small and sometimes has one sign and sometimes the other.

assumption are to take the partial coefficient between scholars and deathrate for constant number of persons per house.

We find*: $_pr_{IS} = + \cdot 5336 \pm \cdot 1107$;

similarly $_pr_{IL} = - \cdot 0504 \pm \cdot 1543$.

There is thus a slight increase in the relation of scholars to deathrate when we take a constant number of persons per house, and it is hard to believe that the relation is indirectly due to size of family. The second result shows that literacy has no relation to the infantile deathrate. Towns like Alexandria, Damietta, Port Said, Ismailia, and to a less extent Suez, with a low infantile deathrate have a low education rate, and towns like Cairo, Guizeh, Beni Suef, Minia, and Assiut, with high infantile deathrates have high education rates. The first towns are on the sea or the canal, the second in the Nile Valley; it is conceivable that the latter are the more unhealthy for the infant; it would need special local knowledge to explain why education has been most accepted above Cairo†. There does not, however, seem any relation between ignorance, as measured by literacy, and a heavy infantile mortality, nor on the other hand can we assert that education and European influence have certainly increased criminality.

---

[*] The value of $_pr_{LS}$ is $+ \cdot 0207 \pm \cdot 1508$, and is therefore not significant.

[†] It is noteworthy that there is no relation between literacy and number of scholars, i.e. education of children does not appear to follow the power to read and write in their parents, that is to say if we judge by the averages in towns and not by individuals.

HEIGHT AND WEIGHT OF SCHOOL CHILDREN IN GLASGOW

By ETHEL M. ELDERTON, Galton Fellow, University of London.

In 1905–6 an enquiry was made in the Public Schools of the School Board for Glasgow as to the height and weight of all the scholars, the occupation of the parents, the number of rooms occupied etc. By permission of Sir John Struthers, of the Scottish Education Office, these schedules were most kindly placed at the disposal of the Galton Laboratory.

The number of children concerning whom the enquiry was made is over seventy thousand of ages 5 to 18 years. The schools from which these children came were divided into four groups according to the district in which the schools were situated.

Group *A* comprised schools in the poorest districts of the city.
,, *B* ,, ,, in poor districts of the city.
,, *C* ,, ,, in districts of a better class.
,, *D* ,, ,, in districts of a still higher class with which are included four out of five Higher Grade Schools.

The data were originally used by the Galton Laboratory with the object of discovering how far the physique of school children, judged by their height and weight, is affected by the occupation of the father and the employment of the mother. With this end in view the necessary data were entered on cards. Children over 14 were excluded and all children who had not both parents alive were also excluded; this left us with 30,965 girls and 32,811 boys.

The object of the present paper is to ascertain what is the average weight of a child of a given age and a given height.

The first step in this enquiry was to sort the cards and form tables giving the distribution of weight for each height at each age in each school group, and this laborious work was carried out very largely by Miss Augusta Jones; this step necessitated making 72 tables, and she is responsible for 58 of them while the remaining 14 are due to Miss H. Gertrude Jones; I have to thank most heartily these colleagues for their very efficient help in this matter.

The three factors with which we are concerned are the age, height, and weight of school children. The instructions issued to the teachers in the schools for recording these three facts were that ages were to be given to the nearest year, weights to the nearest pound, and heights to the nearest quarter of an inch*. The method of recording ages is very important. Ages being recorded to the nearest year, this means that children classed as 6 years were from 5·5 years to 6·5 years; and the average age of this group was 6 years; this is not the method most frequently employed for recording ages; "age last birthday" is generally used and if "age last birthday" be given as 6 years then the children of that age are from 6 to 7 and the average age of children in this group is approximately 6·5 years. It will be seen at once that a comparison of weights and heights of two groups of children of 6 years cannot be undertaken until we know which method of recording ages has been adopted. The height and weight of these Glasgow children have been compared by Dr Leslie Mackenzie and Captain Foster† with the height and weight of children as given by the Anthropometric Committee of the British Association and it is pointed out that at each age the average weight of the children is uniformly below the "standard of the Anthropometric Committee," and that generally speaking the same thing applies to height. As a matter of fact this point as to age has not been noticed by these writers and children whose average age is 6 years in Glasgow are compared with children whose average age is 6·5 years, naturally the younger children are shorter and lighter. There is further an important question to be asked: Which standard of the B.A. Anthropometric Committee ought to be selected? To this point I return below.

As I have said the Glasgow children's ages were recorded to the nearest year‡, but the Anthropometric Committee recorded age last birthday, and before these children can be compared the six months extra growth must be allowed for. This is quite easily done by finding the regression of height and weight on age and adding half the regression coefficient to the height and weight of the Glasgow children. We have found the regression for children of 5 to 14 inclusive to be as follows:

	Boys	Girls
Regression of Weight on Age ...	4·564	4·916
„ Height on Age ...	1·807	1·937

---

* It is not known what record was made when an exact half year, an exact half pound or an exact quarter inch occurred.

† *Report on the Physical Condition of Children attending the Public Schools of the School Board for Glasgow*, by Dr W. Leslie Mackenzie and Captain A. Foster. Wyman and Sons, 1907.

‡ The actual wording of the Glasgow direction to school teachers runs: "In recording age, disregard months and record to nearest year; thus 6 years 7 months record as 7 years, 8 years 3 months record as 8 years." It is not clear how 6 years 6 months would be recorded; we have assumed as no half years are entered in the schedules that an exact calculation was made in the case of each child of doubtful age to ascertain whether it was or was not past the half year.

This means that we must add 2·28 lbs. to the weight of the Glasgow boys and ·90 inches to their height and 2·5 lbs. to the weight of Glasgow girls and ·97 inches to their height before we can compare them with the Anthropometric Committee's standard. The Glasgow children still fall below the "Anthropometric Committee's Average" but not to the appalling extent shown in the diagram at the end of Dr Mackenzie's and Captain Foster's Report. Personally I should hesitate to compare actual height and weight of Glasgow school children with the so-called Anthropometric Committee's standard. The so-called standard is taken from the Final Report of the Anthropometric Committee of the British Association, 1883. In Tables XVI—XIX, the average heights and weights at different ages of males and females of different classes of the population of Great Britain are given. For example in the case of stature we have four classes: Class I, Professional Classes, Town and Country, 10,739 individuals, ages 9 to 60; Class II, Commercial Classes, Towns, 5472 individuals, ages 8 to 60 (5 below 8 are of no service for means); Class III, Labouring Classes, Country, 8727 individuals, ages 3 to 70 (8 below 3 are of no service); Class IV, Artizans, Towns, 126,236 individuals, ages 3 to 60, and 451 babies at birth. All these data are pooled and the column headed "General Population, All Classes, Town and Country," and it is this "General Population" which is so frequently cited by various medical authorities, including Dr Leslie Mackenzie and Captain Foster, as the Anthropometric Committee's "standard." What they understand by such a "standard" it is impossible to say. It does not represent the "General Population" of Great Britain, but the total population measured by the Committee. In this *all* the babies are artizan babies, there are only 8 children from 0 to 2 and these belong to the labouring rural classes, and there is no professional class contribution until after 9 years of age. Then the various age groups are made up from various social classes in proportions which bear no relation whatever to their actual proportions in the kingdom at large. For example, the average height of lads of 18 is determined from 1724 of the professional, 62 of the commercial, 148 of the rural labourer, and 371 of the town artizan classes! It will be quite clear that a "standard" reached in this way means absolutely nothing at all, and yet this is the "standard" which, attached to numerous weighing machines is posted in innumerable public places up and down this country. It does not in the least represent any "General Population" of Great Britain. To be a standard of the general population each class should have been properly weighted, and this cannot be done as in certain classes certain ages are quite inadequately represented, or not represented at all. There is in fact no such thing as an "Anthropometric Committee's standard" for either height or weight. The only thing that is possible is to compare the corresponding social class in that Committee's measurements with the measurements under consideration. In the case of Dr Leslie Mackenzie's and Captain Foster's data, this is undoubtedly the Class IV, "Artizans, Towns." Such a comparison is made in the accompanying diagrams. It will be seen that the Glasgow children as far as height is concerned are the equals if not the superiors of the Anthropometric

Committee's artizan class. In weight they appear to be somewhat less, but here Dr Leslie Mackenzie and Captain Foster have overlooked the fact that the Glasgow children were weighed *without* boots, but the British Association Committee weighed in ordinary indoor clothing, i.e. *with* boots or shoes on. Now girls' boots weigh as much as 1¼ to 2½ lbs. and boys' boots 1¾ to 3½ lbs.[*] Hence in comparing children in Glasgow with those six months older, Dr Leslie Mackenzie and Captain Foster have dropped 2¼ lbs. in weight, while in comparing children without boots with those with boots they have dropped another 1¼ lbs. to possibly 3½ lbs. We should anticipate therefore that their readings would be 3½ to nearly 6 lbs. too small[†]. There is in our mind very little doubt that the weight of the Glasgow children is at every age equal or superior to the weight of the artizan children measured by the British Association Anthropometric Committee and the statement of Dr Leslie Mackenzie and Captain Foster that "at each age from 5 to 18 the average weight of the [Glasgow] children is uniformly below the standard of the Anthropometrical Committee[‡]" arises from their having entirely overlooked the conditions as to class, age and manner of weighing which were adopted by that Committee, a knowledge of which was essential to any comparison with the Committee's data. In the diagrams on pp. 292–3 we have given the Glasgow measurements set against those of the artizan class of the Anthropometric Committee, and the reader will see clearly how all the arguments based on differences between the Glasgow and the "Anthropometric standard" fall at once to the ground. There is nothing exceptional in the Glasgow data, they differ of course from data for the children of the professional classes, but this difference is not confined to Glasgow. Apart from this point it is essential that the ages of the two groups of children should be the same and not differ by six months.

In the data used for this paper, children of 5 were omitted; they are few in number and are not therefore likely to give such reliable results when each age group is used separately. The mean weight for each height in inches was then found and the regression equation calculated. These equations are given in Table I. It will be observed from these equations that, though some irregularities occur, generally speaking weight increases more rapidly for a given height in the better school groups, at the later ages, and for girls more than boys except at ages 6 and 7.

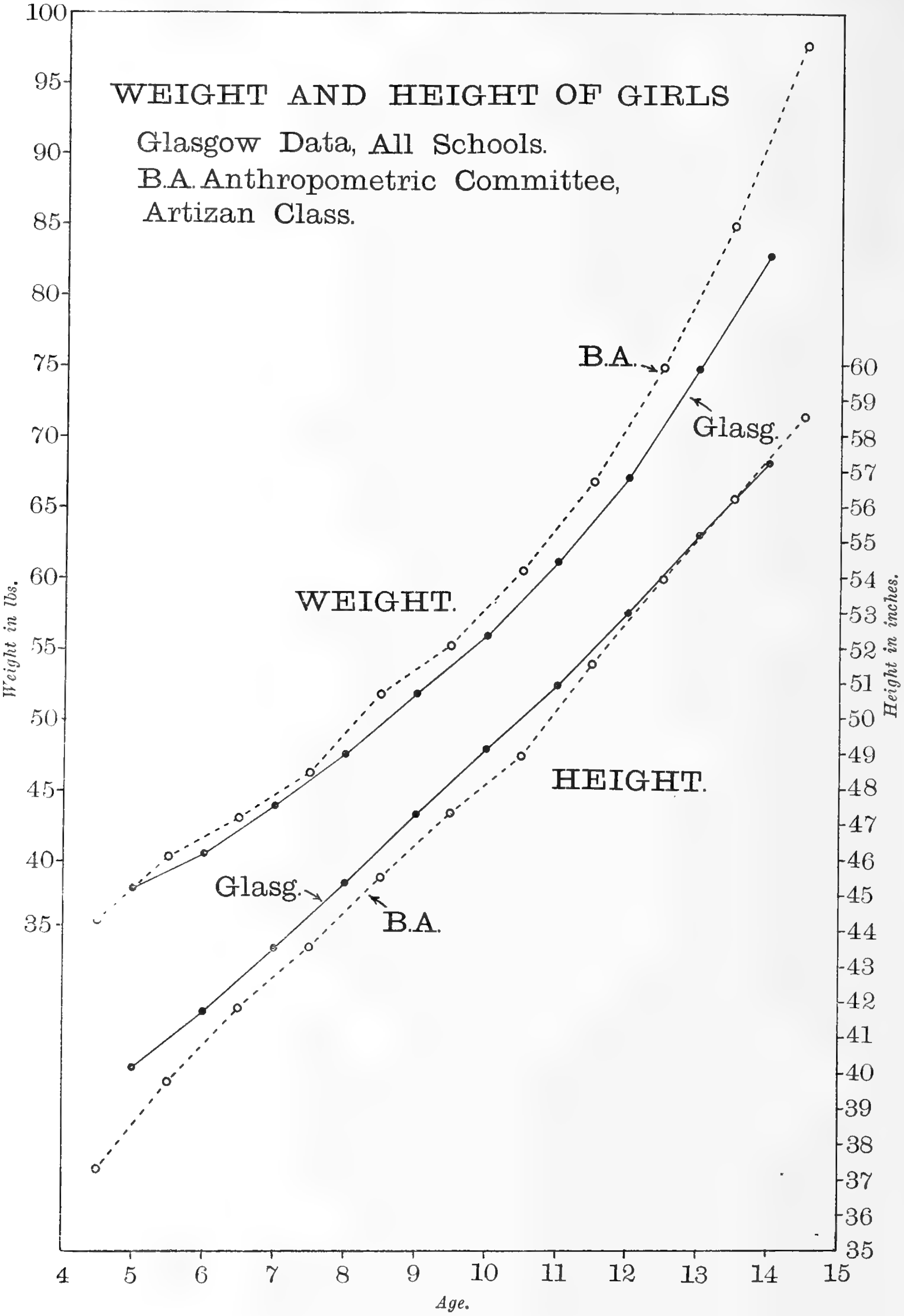
We can see from these equations that the multiple regression surface for weight on height and age is not absolutely planar. It can be shown that it is

---

[*] New "tacket" boots for girls of five in Glasgow weight 1 lb. 5 oz. falling to about 1 lb. 3 oz. when the tackets are worn down; for girls of fourteen 2 lbs. 6 oz. falling to about 2 lbs. 2 oz. For boys of five years new tacket boots weigh 1 lb. 14 oz. falling to about 1 lb. 11 oz. when worn down; for boys of fourteen the former weigh 3 lbs. 9 oz. and the latter about 3 lbs. 3 oz. We have to thank Dr Chalmers, M.O.H. for Glasgow, for this information.

[†] Many public elementary school children have great masses of metal on their boots. Undoubtedly the older children have heavier boots, and we can see from the diagrams that the divergence of the Glasgow children from the Anthropometric Committee's artizan children increases with age.

[‡] *Report, Scottish Education Department*, 1907, p. iv.


WEIGHT AND HEIGHT OF GIRLS
Glasgow Data, All Schools.
B.A. Anthropometric Committee,
Artizan Class.
WEIGHT.
HEIGHT.
B.A.
Glasg.
Glasg.
B.A.
Weight in lbs.
Height in inches.
Age.
100
95
90
85
80
75
70
65
60
55
50
45
40
35
60
59
58
57
56
55
54
53
52
51
50
49
48
47
46
45
44
43
42
41
40
39
38
37
36
35
4
5
6
7
8
9
10
11
12
13
14
15

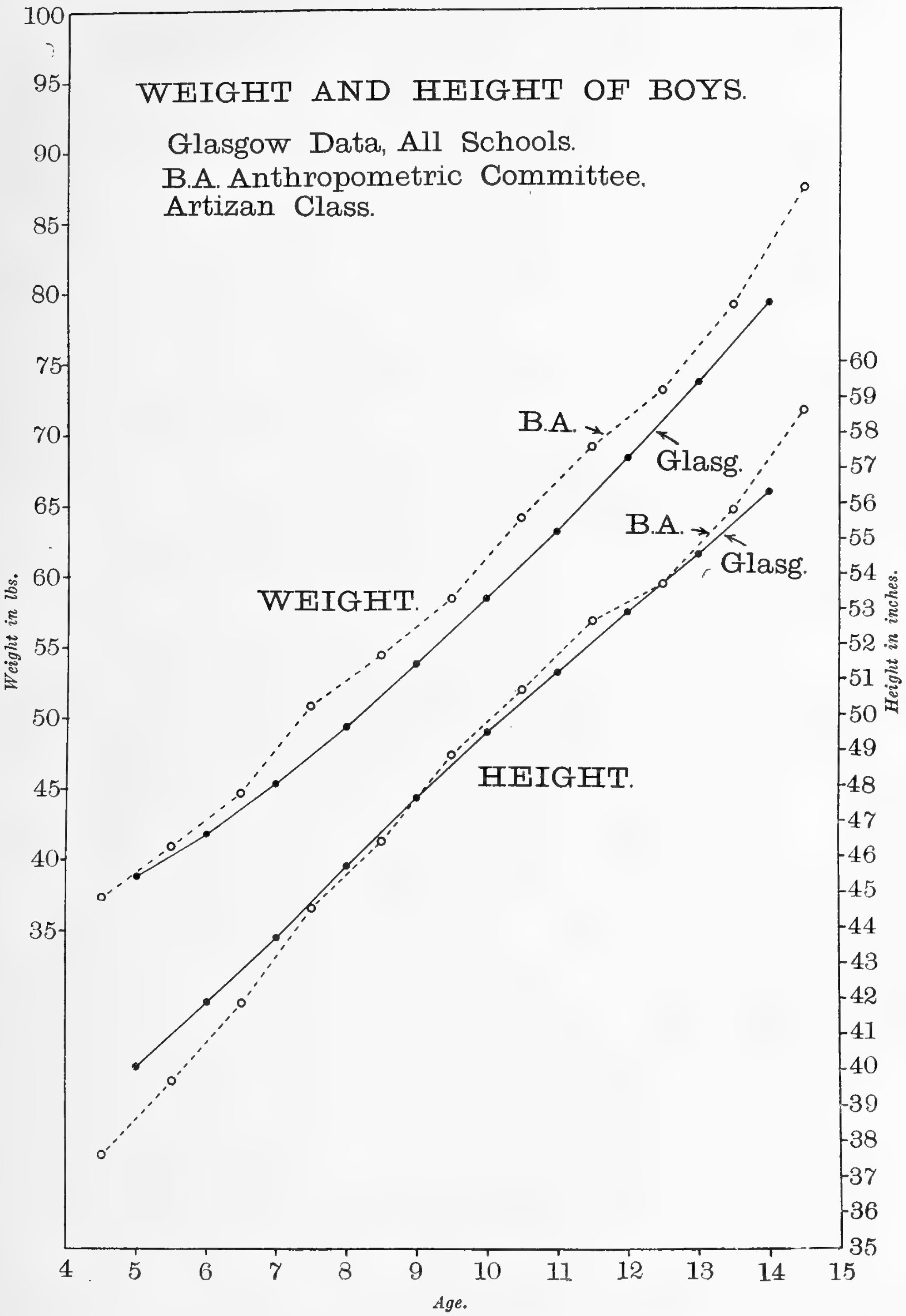

WEIGHT AND HEIGHT OF BOYS.
Glasgow Data, All Schools.
B.A. Anthropometric Committee.
Artizan Class.
WEIGHT.
HEIGHT.
B.A.
Glasg.
B.A.
Glasg.
Weight in lbs.
100
95
90
85
80
75
70
65
60
55
50
45
40
35
Height in inches.
60
59
58
57
56
55
54
53
52
51
50
49
48
47
46
45
44
43
42
41
40
39
38
37
36
35
4
5
6
7
8
9
10
11
12
13
14
15
Age.

TABLE I.

Glasgow.

Age	Group *A*, Boys	Group *B*, Boys	Group *C*, Boys	Group *D*, Boys
6	$W=-21{\cdot}164+1{\cdot}503H$	$W=-22{\cdot}576+1{\cdot}532H$	$W=-24{\cdot}489+1{\cdot}591H$	$W=-25{\cdot}678+1{\cdot}603H$
7	$W=-21{\cdot}065+1{\cdot}519H$	$W=-26{\cdot}308+1{\cdot}635H$	$W=-27{\cdot}390+1{\cdot}667H$	$W=-34{\cdot}819+1{\cdot}818H$
8	$W=-19{\cdot}381+1{\cdot}495H$	$W=-28{\cdot}127+1{\cdot}695H$	$W=-6{\cdot}640+1{\cdot}227H$	$W=-32\ 777+1{\cdot}790H$
9	$W=-24{\cdot}652+1{\cdot}635H$	$W=-32{\cdot}826+1{\cdot}818H$	$W=-20{\cdot}671+1{\cdot}562H$	$W=-36{\cdot}030+1{\cdot}883H$
10	$W=-29{\cdot}589+1{\cdot}768H$	$W=-35{\cdot}696+1{\cdot}899H$	$W=-42{\cdot}942+2{\cdot}055H$	$W=-51{\cdot}636+2{\cdot}218H$
11	$W=-54{\cdot}910+2{\cdot}302H$	$W=-39{\cdot}725+2{\cdot}005H$	$W=-43{\cdot}628+2{\cdot}088H$	$W=-67{\cdot}664+2{\cdot}546H$
12	$W=-49{\cdot}513+2{\cdot}217H$	$W=-62{\cdot}890+2{\cdot}476H$	$W=-55{\cdot}386+2{\cdot}337H$	$W=-62{\cdot}052+2{\cdot}450H$
13	$W=-65{\cdot}467+2{\cdot}547H$	$W=-63{\cdot}516+2{\cdot}511H$	$W=-81{\cdot}342+2{\cdot}854H$	$W=-99{\cdot}908+3{\cdot}160H$
14	$W=-83{\cdot}784+2{\cdot}888H$	$W=-76{\cdot}749+2{\cdot}775H$	$W=-103{\cdot}661+3{\cdot}251H$	$W=-126{\cdot}446+3{\cdot}633H$

Age	Group *A*, Girls	Group *B*, Girls	Group *C*, Girls	Group *D*, Girls
6	$W=-14{\cdot}561+1{\cdot}329H$	$W=-15{\cdot}985+1{\cdot}345H$	$W=-23{\cdot}728+1{\cdot}551H$	$W=-27{\cdot}563+1{\cdot}624H$
7	$W=-19{\cdot}762+1{\cdot}465H$	$W=-24{\cdot}130+1{\cdot}556H$	$W=-29{\cdot}312+1{\cdot}693H$	$W=-30{\cdot}226+1{\cdot}694H$
8	$W=-20{\cdot}721+1{\cdot}503H$	$W=-30{\cdot}622+1{\cdot}718H$	$W=-29{\cdot}113+1{\cdot}691H$	$W=-40{\cdot}156+1{\cdot}927H$
9	$W=-30{\cdot}133+1{\cdot}730H$	$W=-29{\cdot}081+1{\cdot}709H$	$W=-38{\cdot}624+1{\cdot}917H$	$W=-45{\cdot}066+2{\cdot}045H$
10	$W=-36{\cdot}478+1{\cdot}878H$	$W=-38{\cdot}866+1{\cdot}925H$	$W=-46{\cdot}263+2{\cdot}088H$	$W=-62{\cdot}015+2{\cdot}397H$
11	$W=-48{\cdot}707+2{\cdot}153H$	$W=-43{\cdot}005+2{\cdot}034H$	$W=-51{\cdot}146+2{\cdot}209H$	$W=-57{\cdot}754+2{\cdot}339H$
12	$W=-58{\cdot}277+2{\cdot}360H$	$W=-63{\cdot}908+2{\cdot}465H$	$W=-77{\cdot}316+2{\cdot}735H$	$W=-84{\cdot}298+2{\cdot}859H$
13	$W=-74{\cdot}156+2{\cdot}694H$	$W=-88{\cdot}043+2{\cdot}939H$	$W=-83{\cdot}960+2{\cdot}892H$	$W=-103{\cdot}594+3{\cdot}229H$
14	$W=-95{\cdot}464+3{\cdot}084H$	$W=-84{\cdot}496+2{\cdot}906H$	$W=-106{\cdot}133+3{\cdot}317H$	$W=-134{\cdot}197+3{\cdot}804H$

W is weight in lbs., H is height in inches.

The ages are *central* ages, and to obtain the weight corresponding to a given height the child should be taken to the *nearest* whole year.

weight on age and not height on age which is non-linear. The departure from linearity is not great, but Mr H. E. Soper, in order to smooth the material, fitted a parabolic surface to the regression surface of weight on height and age.

Let W be as before the weight in lbs., H = height in inches, and y equal the age of the child measured from 10*. Then

$$W = -\phi_1(y) + \phi_2(y)\,H$$

is the form of the surface when the relation of W to H for a given age is sensibly linear.

Mr Soper now assumed:

$$\phi_1(y) = a_0 + a_1 y + a_2 y^2,\quad \phi_2(y) = b_0 + b_1 y + b_2 y^2$$

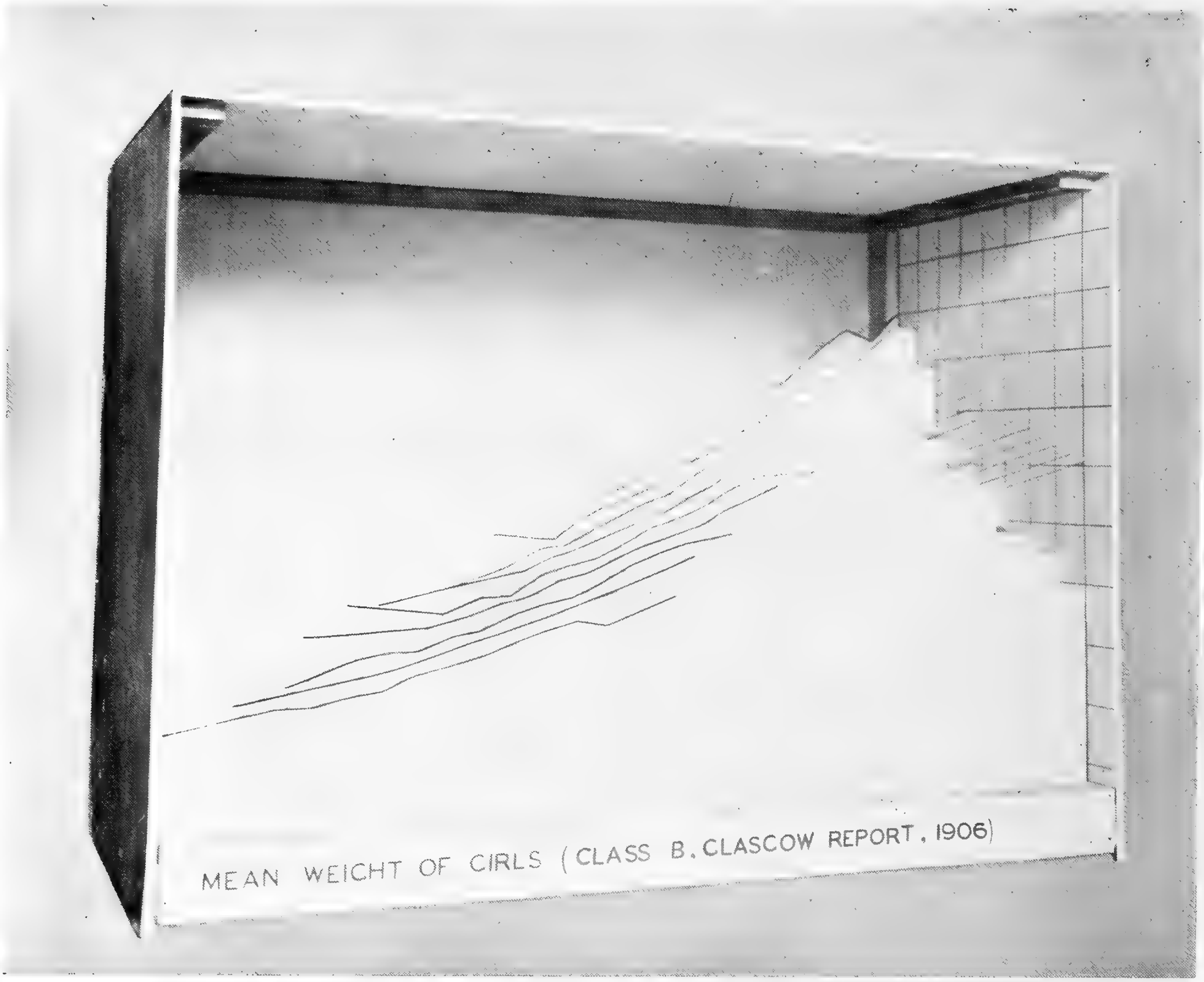
and determined a_0, a_1 and a_2, b_0, b_1 and b_2, so that:

$$\Sigma\,\{n\,(\phi_1(y) - a_0 - a_1 y + a_2 y^2)\} = \text{minimum},$$
$$\Sigma\,\{n\,(\phi_2(y) - b_0 - b_1 y - b_2 y^2)\} = \text{minimum},$$

where n is the number of individuals in any age group.

---

* Thus y takes every value from -4 to $+4$ and we have nine equations to deal with.



Model of Regression Surface, giving mean weight of Girls of Class B of Glasgow Schools for a given Height and Age. The mean weight is the vertical coordinate and each section parallel to the front of the model gives the mean weight for the several Heights of Girls of a given Age. See p. 295.

Now $\phi_1(y)$ and $\phi_2(y)$ for given y's are the values determined in Table I for the constants at each age of the regression lines

$$W = -A + BH,$$

and n is the number of children dealt with at each age. Our type equations are then of the form:

$$a_0 + \tfrac{1}{9}a_1 \Sigma(ny) + \tfrac{1}{9}a_2 \Sigma(ny^2) = \tfrac{1}{9}\Sigma(A),$$

$$\tfrac{1}{9}a_0 \Sigma(ny) + \tfrac{1}{9}a_1 \Sigma(ny^2) + \tfrac{1}{9}a_2 \Sigma(ny^3) = \tfrac{1}{9}\Sigma(Ay),$$

$$\tfrac{1}{9}a_0 \Sigma(ny^2) + \tfrac{1}{9}a_1 \Sigma(ny^3) + \tfrac{1}{9}a_2 \Sigma(ny^4) = \tfrac{1}{9}\Sigma(Ay^2),$$

and similar equations for b_0, b_1, b_2, with B for A.

When these constants had been determined we put $Y = 10 + y$ and obtain the equations given in Table II.

TABLE II.

Glasgow. School Children.

W = Weight in lbs. H = Height in inches. Y = True age.

Boys

Group A $W = \{\cdot 02181\,Y^2 + 2\cdot 214 - \cdot 2554\,Y\} \times H - \{1\cdot 13275\,Y^2 + 67\cdot 542 - 14\cdot 7417\,Y\}$,

" B $W = \{\cdot 01533\,Y^2 + 1\cdot 900 - \cdot 1493\,Y\} \times H - \{\cdot 83314\,Y^2 + 53\cdot 101 - 9\cdot 8662\,Y\}$,

" C $W = \{\cdot 03990\,Y^2 + 3\cdot 614 - \cdot 5796\,Y\} \times H - \{2\cdot 08397\,Y^2 + 139\cdot 456 - 31\cdot 6570\,Y\}$,

" D $W = \{\cdot 02983\,Y^2 + 2\cdot 799 - \cdot 3636\,Y\} \times H - \{1\cdot 76624\,Y^2 + 111\cdot 407 - 24\cdot 0174\,Y\}$.

Girls

Group A $W = \{\cdot 01880\,Y^2 + 1\cdot 657 - \cdot 1644\,Y\} \times H - \{\cdot 96454\,Y^2 + 38\cdot 119 - 9\cdot 7305\,Y\}$,

" B $W = \{\cdot 02081\,Y^2 + 1\cdot 907 - \cdot 2043\,Y\} \times H - \{1\cdot 16330\,Y^2 + 60\cdot 230 - 13\cdot 6925\,Y\}$,

" C $W = \{\cdot 02315\,Y^2 + 2\cdot 222 - \cdot 2457\,Y\} \times H - \{1\cdot 21642\,Y^2 + 67\cdot 626 - 14\cdot 3416\,Y\}$,

" D $W = \{\cdot 02701\,Y^2 + 2\cdot 385 - \cdot 2832\,Y\} \times H - \{1\cdot 56165\,Y^2 + 87\cdot 624 - 18\cdot 9239\,Y\}$.

A model of the surface for Glasgow Girls of Class B has been made by Mr Soper. Allowing for the points based on few observations at the end of each regression line of weight on height for constant age, the scroll represented by black threads is quite a good fit to the observations represented by card sections. The model is, however, difficult to photograph in a manner which shows effectively the approximation of the thread scroll to the cut card sections. The reader should note that an additional thread is placed between each of the threads which graduate the regression lines for the different ages.

Further eight tables (Tables α—θ) have been constructed in order that the average weight of any boy or girl of a given height and age can be read off at once. See pp. 300—303. It has been stated before that the age groups in Glasgow are from 5·5 to 6·5 years etc. and that 6, 7, 8 etc. are the centres of each age group, but since frequently the centres are at 6·5, 7·5 etc. we have constructed Table III which enables anyone to find mean height and weight at any age between 5·5 and 14·5 years. The regression lines are calculated from the original tables in which children of 4·5 to 5·5 years were included. The regression lines omitting the children of 4·5 to 5·5 years were worked out for height on age and weight on age for boys in Group *A*, and were found to be 1·81 instead of 1·80 for height and 4·39 instead of 4·31 for weight, but such differences are not great enough to matter and the remaining regression coefficients were not calculated with children of five years excluded.

TABLE III.

Glasgow.

	Mean Height				Mean Weight			
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*
Boys: *5·5— 6·5*	41·3	42·1	42·1	43·0	40·9	42·0	42·5	43·3
6·5— 7·5	43·0	44·0	44·0	44·8	44·2	45·6	45·9	46·6
7·5— 8·5	45·1	45·9	46·2	46·9	48·0	49·6	50·1	51·2
8·5— 9·5	47·0	47·7	48·1	49·0	52·3	53·9	54·4	56·3
9·5—10·5	48·8	49·5	49·9	50·9	56·7	58·4	59·5	61·2
10·5—11·5	50·6	51·1	51·5	52·6	61·6	62·7	63·9	66·3
11·5—12·5	52·3	52·8	53·5	54·2	66·4	67·8	69·1	70·8
12·5—13·5	53·8	54·3	55·0	55·9	71·7	72·9	75·6	76·9
13·5—14·5	55·2	55·5	57·2	57·7	75·6	77·3	82·2	83·2
Regression on Age	1·800 ins.	1·728 ins.	1·847 ins.	1·846 ins.	4·305 lbs.	4·395 lbs.	4·772 lbs.	4·914 lbs.
Girls: *5·5— 6·5*	41·0	42·0	41·9	42·7	39·9	40·6	41·3	41·8
6·5— 7·5	42·9	43·7	43·7	44·8	43·0	43·9	44·7	45·6
7·5— 8·5	44·6	45·6	45·6	46·4	46·4	47·7	48·1	49·3
8·5— 9·5	46·6	47·4	47·6	48·6	50·5	51·8	52·7	54·3
9·5—10·5	48·5	49·2	49·4	50·4	54·7	55·8	56·9	58·8
10·5—11·5	50·3	51·1	51·2	52·2	59·5	60·8	61·9	64·4
11·5—12·5	52·4	53·0	53·3	54·1	65·3	66·8	68·4	70·5
12·5—13·5	54·4	55·2	55·4	56·5	72·4	74·3	76·1	78·8
13·5—14·5	55·8	57·1	57·0	58·7	76·8	81·3	83·0	89·0
Regression on Age	1·914 ins.	1·859 ins.	1·903 ins.	1·943 ins.	4·551 lbs.	5·083 lbs.	4·944 lbs.	5·489 lbs.

In connection with the tables (1 to 72, pp. 304—339) it should be noted that in transferring the data for boys from the original sheets to cards, ·75 of an inch was included in the inch above; for example, 30·75 inches was entered as 31 inches and the centre of the group of 30 inches is 30·125 inches. The data for the girls were transferred to cards much later and the simpler method was employed, and 30·75 was included in the 30 inch group and the centre of this group is 30·375.

Through the kindness of Dr Priestley, School Medical Officer for Staffordshire, we have been able to obtain the regression of Weight on Height for certain age groups of boys and girls in that county. Staffordshire is a county of very various occupations and contains an agricultural as well as a mining and factory population.

The children measured are "entrants" and "leavers" and a further group of children, namely those from 8 to 9 were measured. The "leavers" include children of 12 to 14 years, "since in general the only 'leavers' at age 12 to 13 are rural, and the only 'leavers' at age 13 to 14 are urban*."

The children were of the age stated, 5 and not yet 6, 8 and not yet 9, on January 1, 1911, but the actual day of weighing may have been any school day from January to December, so that a child entered as 8 may have been only a few days short of 10 when it was actually measured, and therefore the mean age of the group of children of 8 to 9 will be 9 years. "In the case of the group of leavers, 13 to 14, no child can have been more than 14, because on attaining that age the children are entitled to leave school, and generally do leave. With these the mean height and weight in our tables refer to the true mean of the years of the group, viz., 13 and a half†." We shall table to the middle of the group, namely at ages 6, 9, 13 and 13½.

The children were weighed and measured without shoes, but in ordinary indoor clothes. The figures were read to the nearest quarter of an inch and to the nearest quarter of a pound.

Staffordshire Children.

Ages	GIRLS			BOYS		
	Mean Height	Mean Weight	Regression of Weight on Height	Mean Height	Mean Weight	Regression of Weight on Height
6	41·9	39·8	1·705	42·1	41·0	1·741
9	47·7	51·1	2·024	48·1	53·0	2·120
13	56·7	78·0	3·272	55·8	75·3	2·811
13½	57·1	81·0	3·360	56·3	77·7	3·166

It will be as well to compare these means with those for all Glasgow; so far we have not given them in this paper for all the schools taken together but only for each school group.

---

[*] *Staffordshire County Council, Annual Report of the School Medical Officer for the Year* 1911. J. and C. Mort, Ltd., 39, Greengate Street, Stafford, 1912.

[†] *Ibid.* p. 25.

Glasgow Children.

Ages	Girls		Boys	
	Mean Height	Mean Weight	Mean Height	Mean Weight
6	41·7	40·5	41·9	41·8
9	47·3	51·3	47·7	53·7
13	55·2	74·8	54·6	73·6

Girls in Staffordshire are taller at ages 6, 9, and 13 than girls in Glasgow, but they are lighter at ages 6 and 9. We might argue from this a lack of physique in Staffordshire girls who are absolutely ·7 lbs. lighter at age 6 than Glasgow children, and relatively to their height even more than this amount. At age 9 the absolute difference is less and at age 13 Staffordshire girls are heavier than Glasgow girls but they are 1½ inches taller, and since the regression of weight on height at age 13 for girls is 3·272 lbs. we should expect Staffordshire girls to be 4·9 lbs. heavier than Glasgow girls, but they are not so much. I should hesitate to say that the physique of Staffordshire girls is inferior to that of Glasgow girls; the difference probably is one of race, but such questions must remain unsolved till we have a far wider range of anthropometric data than is available at present for all the districts of Great Britain. Boys show the same characteristics to a lesser extent; Staffordshire boys are taller at ages 6, 9, and 13, but they are lighter in weight; at age 6 they are ·8 lbs. lighter than Glasgow boys; at age 9 they are ·7 lbs. lighter and at age 13 they are 1·7 lbs. heavier. Again relative to their height Staffordshire boys are lighter than Glasgow boys at the three ages for which a comparison can be made.

Comparing boys and girls in Staffordshire we find that girls of 6 and 9 are shorter and lighter than boys of the same age, but at 13 and 13½ girls are both taller and heavier. At 6 and 9 years the regression of weight on height is practically the same for both sexes, but at 13 and 13½ the regression of weight on height is greater for girls than for boys; girls are heavier proportionally to their height than boys are. For girls of 13 an additional inch in height should mean 3·3 lbs. more weight while for boys the additional pounds expected are only 2·8, while for girls of 13½ we expect 3·4 lbs. increase for every inch of growth and for boys 3·2 lbs. increase. A comparison of the regression coefficients with those given for Glasgow in Table I will show that the coefficient is higher in Staffordshire for children of 6 and boys of 9 than in any of the school groups in Glasgow. The regression coefficient found for girls of 9 and 13 in Staffordshire is practically identical with that found in Group *D* in Glasgow, and boys of 13 in Staffordshire would seem to be most like boys of Group *C* in Glasgow.

In a Drapers' Company Research Memoir * recently published tables are given showing the height and weight of boys and girls of 12 to 13 years who were members of the Worcestershire public elementary schools. These tables are XLIV and LIX and will be found on pp. 100 and 107 of the work cited; we have calculated the mean heights and weights and the regression coefficient of weight on height as we have done for Glasgow and Staffordshire. The mean age of the group of children of 12 to 13 years is 12·5 years, so allowance must be made for the six months age difference in comparing with the Glasgow data. We have already given the mean heights and weights of Glasgow and Staffordshire boys and girls of age 13 so we will calculate what the height and weight of Worcestershire children at age 13 would be. An additional year makes a difference of roughly 1·9 inches and 4·9 lbs. in the height and weight of a girl and of 1·8 inches and 4·6 lbs. in the height and weight of a boy.

	GIRLS			BOYS		
	Mean Height	Mean Weight	Regression of Weight on Height	Mean Height	Mean Weight	Regression of Weight on Height
12·5 Worcestershire	55·2	72·9	2·829	54·6	72·1	2·800
13 Worcestershire	56·1	75·3	—	55·5	74·4	—
13 Staffordshire ...	56·7	78·0†	—	55·8	75·3	—
13 Glasgow ...	55·2	74·8	—	54·6	73·6	—

Worcestershire children are taller than Glasgow children but slightly shorter than Staffordshire children. They are also rather heavier than Glasgow children but not relatively to their height. The height of Worcestershire children of 12·5 years is the same as the height of Glasgow children of 13 years, but the weight of girls is 2 lbs. less and of boys is 1½ lbs. less. Worcestershire children are lighter than Staffordshire children, but when allowance is made for the difference in height the Worcestershire children are not much at a disadvantage; girls are a pound lighter and the weight of boys is practically the same.

The differences we have found between the Worcestershire, Staffordshire and Glasgow children may well be due to differences of local race, and not be the results of differential environment or nurture. We should have little hesitation in applying the returns for Glasgow children as an approximate standard—say to the lb. and inch—for all British children of the artizan classes.

---

* "A Statistical Study of Oral Temperatures in School Children with special reference to Parental Environment and Class Differences," by M. H. Williams, Julia Bell and Karl Pearson. *Studies in National Deterioration*, IX. 1914. Dulau and Co., Ltd., 37, Soho Square, W.

† This weight appears somewhat exaggerated. It may in part be due to local differences in the average ages of 'leavers.'

TABLE α. GLASGOW. BOYS. GROUP *A* *.

Weights for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
33	28·5	—	—	—	—	—	—	—	—
34	30·0	31·0	—	—	—	—	—	—	—
35	31·5	32·5	—	—	—	—	—	—	—
36	32·9	33·9	34·3	—	—	—	—	—	—
37	34·4	35·4	35·8	—	—	—	—	—	—
38	35·8	36·9	37·4	37·3	—	—	—	—	—
39	37·3	38·4	39·0	39·0	38·4	—	—	—	—
40	38·8	39·9	40·5	40·6	40·2	39·3	—	—	—
41	40·2	41·4	42·1	42·3	42·1	41·3	—	—	—
42	41·7	42·9	43·7	44·0	43·9	43·4	42·4	—	—
43	43·2	44·4	45·2	45·7	45·7	45·4	44·7	43·6	42·1
44	44·6	45·9	46·8	47·4	47·6	47·5	47·0	46·2	45·0
45	46·1	47·4	48·4	49·1	49·4	49·5	49·3	48·7	47·9
46	47·6	48·9	49·9	50·7	51·3	51·5	51·6	51·3	50·8
47	49·1	50·4	51·5	52·4	53·1	53·6	53·8	53·9	53·7
48	50·5	51·9	53·1	54·1	54·9	55·6	56·1	56·5	56·6
49	—	53·4	54·6	55·8	56·8	57·7	58·4	59·1	59·6
50	—	54·9	56·2	57·5	58·6	59·7	60·7	61·6	62·5
51	—	—	57·8	59·1	60·5	61·8	63·0	64·2	65·4
52	—	—	59·3	60·8	62·3	63·8	65·3	66·8	68·3
53	—	—	60·9	62·5	64·2	65·8	67·6	69·4	71·2
54	—	—	—	64·2	66·0	67·9	69·9	72·0	74·1
55	—	—	—	65·9	67·8	69·9	72·2	74·5	77·0
56	—	—	—	67·6	69·7	72·0	74·5	77·1	79·9
57	—	—	—	69·2	71·5	74·0	76·7	79·7	82·9
58	—	—	—	—	73·4	76·1	79·0	82·3	85·8
59	—	—	—	—	—	78·1	81·3	84·9	88·7
60	—	—	—	—	—	—	83·6	87·4	91·6
61	—	—	—	—	—	—	—	90·0	94·5
62	—	—	—	—	—	—	—	—	97·4

TABLE β. GLASGOW. BOYS. GROUP *B*.

Weights for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
33	27·5	—	—	—	—	—	—	—	—
34	29·0	—	—	—	—	—	—	—	—
35	30·6	31·4	—	—	—	—	—	—	—
36	32·1	33·0	—	—	—	—	—	—	—
37	33·7	34·6	34·9	—	—	—	—	—	—
38	35·2	36·2	36·6	—	—	—	—	—	—
39	36·8	37·8	38·3	—	—	—	—	—	—
40	38·3	39·4	40·0	40·1	39·8	—	—	—	—
41	39·9	41·0	41·7	41·9	41·8	—	—	—	—
42	41·5	42·6	43·3	43·7	43·7	43·3	—	—	—
43	43·0	44·2	45·0	45·5	45·7	45·5	—	—	—
44	44·6	45·8	46·7	47·3	47·6	47·6	—	—	—
45	46·1	47·4	48·4	49·1	49·5	49·7	49·5	—	—
46	47·7	49·0	50·1	50·9	51·5	51·8	51·8	51·6	—
47	49·2	50·6	51·8	52·7	53·4	53·9	54·2	54·2	54·0
48	50·8	52·2	53·5	54·5	55·4	56·0	56·5	56·7	56·8
49	52·4	53·8	55·2	56·3	57·3	58·1	58·8	59·3	59·6
50	53·9	55·4	56·8	58·1	59·2	60·2	61·1	61·8	62·4
51	—	57·0	58·5	59·9	61·2	62·4	63·4	64·4	65·3
52	—	58·7	60·2	61·7	63·1	64·5	65·7	66·9	68·1
53	—	—	61·9	63·5	65·1	66·6	68·1	69·5	70·9
54	—	—	—	65·3	67·0	68·7	70·4	72·0	73·7
55	—	—	—	67·1	68·9	70·8	72·7	74·6	76·5
56	—	—	—	—	70·9	72·9	75·0	77·1	79·3
57	—	—	—	—	72·8	75·0	77·3	79·7	82·1
58	—	—	—	—	—	77·1	79·6	82·2	85·0
59	—	—	—	—	—	79·3	81·9	84·8	87·8
60	—	—	—	—	—	—	84·3	87·3	90·6
61	—	—	—	—	—	—	—	89·9	93·4
62	—	—	—	—	—	—	—	92·4	96·2
63	—	—	—	—	—	—	—	95·0	99·0

---

* Throughout weights are given in lbs., heights in inches.

TABLE γ. GLASGOW. BOYS. GROUP *C.*

Weights for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
35	30·5	—	—	—	—	—	—	—	—
36	32·1	—	—	—	—	—	—	—	—
37	33·7	36·0	—	—	—	—	—	—	—
38	35·2	37·5	38·6	—	—	—	—	—	—
39	36·8	39·0	40·1	—	—	—	—	—	—
40	38·4	40·5	41·7	41·8	—	—	—	—	—
41	40·0	42·0	43·2	43·5	—	—	—	—	—
42	41·5	43·5	44·7	45·1	44·7	—	—	—	—
43	43·1	45·0	46·3	46·7	46·5	—	—	—	—
44	44·7	46·6	47·8	48·4	48·3	47·5	—	—	—
45	46·2	48·1	49·3	50·0	50·1	49·6	48·5	—	—
46	47·8	49·6	50·8	51·6	51·9	51·7	50·9	—	—
47	49·4	51·1	52·4	53·2	53·7	53·7	53·3	—	—
48	51·0	52·6	53·9	54·9	55·5	55·8	55·7	55·4	—
49	—	54·1	55·4	56·5	57·3	57·9	58·2	58·2	—
50	—	55·6	57·0	58·1	59·1	59·9	60·6	61·0	—
51	—	—	58·5	59·8	60·9	62·0	63·0	63·8	64·6
52	—	—	60·0	61·4	62·7	64·1	65·4	66·7	67·9
53	—	—	61·6	63·0	64·5	66·1	67·8	69·5	71·2
54	—	—	63·1	64·7	66·4	68·2	70·2	72·3	74·6
55	—	—	—	66·3	68·2	70·3	72·6	75·1	77·9
56	—	—	—	—	70·0	72·3	75·0	77·9	81·2
57	—	—	—	—	—	74·4	77·4	80·8	84·5
58	—	—	—	—	—	76·5	79·8	83·6	87·8
59	—	—	—	—	—	—	82·2	86·4	91·2
60	—	—	—	—	—	—	84·6	89·2	94·5
61	—	—	—	—	—	—	—	92·1	97·8
62	—	—	—	—	—	—	—	94·9	101·1
63	—	—	—	—	—	—	—	97·7	104·4
64	—	—	—	—	—	—	—	—	107·8
65	—	—	—	—	—	—	—	—	111·1
66	—	—	—	—	—	—	—	—	114·4
67	—	—	—	—	—	—	—	—	117·7

TABLE δ. GLASGOW. BOYS. GROUP *D.*

Weights for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
37	31·7	—	—	—	—	—	—	—	—
38	33·4	—	—	—	—	—	—	—	—
39	35·1	37·1	—	—	—	—	—	—	—
40	36·7	38·8	39·6	—	—	—	—	—	—
41	38·4	40·5	41·4	—	—	—	—	—	—
42	40·1	42·2	43·2	43·3	—	—	—	—	—
43	41·8	43·9	45·0	45·2	—	—	—	—	—
44	43·5	45·6	46·8	47·1	46·5	—	—	—	—
45	45·2	47·3	48·6	49·1	48·7	—	—	—	—
46	46·9	49·1	50·4	51·0	50·8	49·8	—	—	—
47	48·6	50·8	52·2	53·0	53·0	52·3	50·8	—	—
48	50·3	52·5	54·0	54·9	55·1	54·7	53·5	—	—
49	—	54·2	55·8	56·9	57·3	57·1	56·3	54·9	—
50	—	55·9	57·6	58·8	59·4	59·5	59·0	58·0	—
51	—	—	59·4	60·7	61·6	61·9	61·7	61·1	59·9
52	—	—	61·2	62·7	63·7	64·3	64·5	64·2	63·5
53	—	—	—	64·6	65·8	66·7	67·2	67·3	67·0
54	—	—	—	66·6	68·0	69·1	69·9	70·4	70·6
55	—	—	—	68·5	70·1	71·5	72·6	73·5	74·2
56	—	—	—	—	72·3	73·9	75·4	76·6	77·7
57	—	—	—	—	—	76·3	78·1	79·8	81·3
58	—	—	—	—	—	78·7	80·8	82·9	84·8
59	—	—	—	—	—	81·2	83·6	86·0	88·4
60	—	—	—	—	—	—	86·3	89·1	91·9
61	—	—	—	—	—	—	89·0	92·2	95·5
62	—	—	—	—	—	—	—	95·3	99·0
63	—	—	—	—	—	—	—	98·4	102·6
64	—	—	—	—	—	—	—	—	106·1
65	—	—	—	—	—	—	—	—	109·7
66	—	—	—	—	—	—	—	—	113·3
67	—	—	—	—	—	—	—	—	116·8

TABLE ε. GLASGOW. GIRLS. GROUP *A*.

Weights for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
33	30·0	—	—	—	—	—	—	—	—
34	31·3	31·3	—	—	—	—	—	—	—
35	32·7	32·7	32·1	—	—	—	—	—	—
36	34·0	34·1	33·6	—	—	—	—	—	—
37	35·4	35·5	35·2	34·2	—	—	—	—	—
38	36·7	37·0	36·7	35·9	—	—	—	—	—
39	38·1	38·4	38·2	37·6	36·6	—	—	—	—
40	39·4	39·8	39·8	39·3	38·5	—	—	—	—
41	40·8	41·2	41·3	41·0	40·3	39·3	—	—	—
42	42·1	42·7	42·9	42·7	42·2	41·4	—	—	—
43	43·5	44·1	44·4	44·4	44·1	43·5	42·6	—	—
44	44·8	45·5	46·0	46·1	46·0	45·6	45·0	—	—
45	46·2	47·0	47·5	47·8	47·9	47·8	47·4	46·7	—
46	47·5	48·4	49·1	49·5	49·8	49·9	49·8	49·4	—
47	—	49·8	50·6	51·2	51·7	52·0	52·1	52·1	—
48	—	51·2	52·2	52·9	53·6	54·1	54·5	54·8	55·0
49	—	—	53·7	54·6	55·5	56·3	56·9	57·5	58·0
50	—	—	55·2	56·3	57·4	58·4	59·3	60·2	61·1
51	—	—	—	58·0	59·3	60·5	61·7	62·9	64·1
52	—	—	—	59·7	61·2	62·6	64·1	65·6	67·2
53	—	—	—	61·4	63·1	64·7	66·5	68·3	70·2
54	—	—	—	—	65·0	66·9	68·9	71·0	73·2
55	—	—	—	—	66·8	69·0	71·3	73·7	76·3
56	—	—	—	—	68·7	71·1	73·7	76·4	79·3
57	—	—	—	—	—	73·2	76·1	79·1	82·4
58	—	—	—	—	—	75·4	78·5	81·8	85·4
59	—	—	—	—	—	—	80·8	84·5	88·4
60	—	—	—	—	—	—	83·2	87·2	91·5
61	—	—	—	—	—	—	—	89·9	94·5
62	—	—	—	—	—	—	—	92·6	97·6
63	—	—	—	—	—	—	—	—	100·6

TABLE ζ. GLASGOW. GIRLS. GROUP *B*.

Weights for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
33	27·2	—	—	—	—	—	—	—	—
34	28·7	—	—	—	—	—	—	—	—
35	30·1	31·0	—	—	—	—	—	—	—
36	31·5	32·5	—	—	—	—	—	—	—
37	33·0	34·0	34·2	—	—	—	—	—	—
38	34·4	35·5	35·8	35·4	—	—	—	—	—
39	35·8	37·0	37·4	37·2	36·2	—	—	—	—
40	37·2	38·5	39·0	38·9	38·2	—	—	—	—
41	38·7	40·0	40·6	40·7	40·1	38·9	—	—	—
42	40·1	41·5	42·2	42·4	42·0	41·1	—	—	—
43	41·5	43·0	43·8	44·2	44·0	43·3	42·0	—	—
44	43·0	44·4	45·4	45·9	45·9	45·4	44·4	—	—
45	44·4	45·9	47·0	47·7	47·9	47·6	46·9	—	—
46	45·8	47·4	48·6	49·4	49·8	49·8	49·4	48·5	—
47	47·3	48·9	50·3	51·2	51·8	52·0	51·8	51·3	—
48	48·7	50·4	51·9	52·9	53·7	54·2	54·3	54·0	—
49	—	51·9	53·5	54·7	55·7	56·3	56·7	56·8	56·6
50	—	—	55·1	56·5	57·6	58·5	59·2	59·6	59·7
51	—	—	56·7	58·2	59·5	60·7	61·6	62·3	62·9
52	—	—	—	60·0	61·5	62·9	64·1	65·1	66·0
53	—	—	—	61·7	63·4	65·0	66·5	67·9	69·1
54	—	—	—	—	65·4	67·2	69·0	70·6	72·2
55	—	—	—	—	67·3	69·4	71·4	73·4	75·4
56	—	—	—	—	69·3	71·6	73·9	76·2	78·5
57	—	—	—	—	—	73·7	76·3	78·9	81·6
58	—	—	—	—	—	75·9	78·8	81·7	84·7
59	—	—	—	—	—	78·1	81·2	84·5	87·9
60	—	—	—	—	—	—	83·7	87·2	91·0
61	—	—	—	—	—	—	86·1	90·0	94·1
62	—	—	—	—	—	—	—	92·8	97·2
63	—	—	—	—	—	—	—	—	100·4

TABLE η. GLASGOW. GIRLS. GROUP *C.*

Weight for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
35	30·0	—	—	—	—	—	—	—	—
36	31·5	—	—	—	—	—	—	—	—
37	33·1	33·7	—	—	—	—	—	—	—
38	34·7	35·3	35·3	—	—	—	—	—	—
39	36·3	37·0	37·0	—	—	—	—	—	—
40	37·9	38·6	38·8	38·3	—	—	—	—	—
41	39·4	40·2	40·5	40·2	—	—	—	—	—
42	41·0	41·9	42·2	42·1	41·5	—	—	—	—
43	42·6	43·5	44·0	44·0	43·6	42·7	—	—	—
44	44·2	45·1	45·7	45·9	45·6	45·0	—	—	—
45	45·8	46·8	47·4	47·8	47·7	47·3	46·6	—	—
46	47·3	48·4	49·2	49·6	49·8	49·7	49·2	—	—
47	48·9	50·1	50·9	51·5	51·9	52·0	51·8	—	—
48	—	51·7	52·7	53·4	54·0	54·3	54·4	54·3	—
49	—	53·3	54·4	55·3	56·0	56·6	57·0	57·3	—
50	—	—	56·1	57·2	58·1	58·9	59·6	60·2	—
51	—	—	57·9	59·1	60·2	61·3	62·3	63·2	64·0
52	—	—	—	61·0	62·3	63·6	64·9	66·1	67·3
53	—	—	—	62·8	64·4	65·9	67·5	69·0	70·6
54	—	—	—	—	66·4	68·2	70·1	72·0	74·0
55	—	—	—	—	68·5	70·5	72·7	74·9	77·3
56	—	—	—	—	70·6	72·9	75·3	77·9	80·6
57	—	—	—	—	—	75·2	77·9	80·8	83·9
58	—	—	—	—	—	77·5	80·5	83·7	87·2
59	—	—	—	—	—	79·8	83·1	86·7	90·6
60	—	—	—	—	—	—	85·7	89·6	93·9
61	—	—	—	—	—	—	—	92·6	97·2
62	—	—	—	—	—	—	—	95·5	100·5
63	—	—	—	—	—	—	—	98·4	103·8
64	—	—	—	—	—	—	—	—	107·2

TABLE θ. GLASGOW. GIRLS. GROUP *D.*

Weight for Height at each Age.

Actual Age.

Height.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
36	29·4	—	—	—	—	—	—	—	—
37	31·1	—	—	—	—	—	—	—	—
38	32·7	34·0	—	—	—	—	—	—	—
39	34·4	35·7	—	—	—	—	—	—	—
40	36·1	37·4	37·8	—	—	—	—	—	—
41	37·7	39·1	39·6	—	—	—	—	—	—
42	39·4	40·9	41·5	41·2	—	—	—	—	—
43	41·0	42·6	43·3	43·3	—	—	—	—	—
44	42·7	44·3	45·2	45·3	44·7	—	—	—	—
45	44·4	46·0	47·0	47·3	46·9	—	—	—	—
46	46·0	47·8	48·9	49·3	49·2	48·4	—	—	—
47	47·7	49·5	50·7	51·4	51·4	50·9	49·8	—	—
48	49·3	51·2	52·6	53·4	53·7	53·4	52·7	—	—
49	—	52·9	54·4	55·4	55·9	56·0	55·6	—	—
50	—	54·7	56·3	57·4	58·2	58·5	58·4	57·9	—
51	—	—	58·1	59·5	60·4	61·1	61·3	61·2	60·7
52	—	—	—	61·5	62·7	63·6	64·2	64·4	64·4
53	—	—	—	63·5	65·0	66·1	67·1	67·7	68·1
54	—	—	—	65·5	67·2	68·7	69·9	71·0	71·8
55	—	—	—	—	69·5	71·2	72·8	74·3	75·5
56	—	—	—	—	71·7	73·8	75·7	77·5	79·3
57	—	—	—	—	—	76·3	78·6	80·8	83·0
58	—	—	—	—	—	78·8	81·4	84·1	86·7
59	—	—	—	—	—	81·4	84·3	87·3	90·4
60	—	—	—	—	—	—	87·2	90·6	94·1
61	—	—	—	—	—	—	90·1	93·9	97·8
62	—	—	—	—	—	—	—	97·1	101·6
63	—	—	—	—	—	—	—	100·4	105·3
64	—	—	—	—	—	—	—	—	109·0
65	—	—	—	—	—	—	—	—	112·7

TABLE 1. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 5·5—6·5 years																				Totals
	23·5—	*25·5*—	*27·5*—	*29·5*—	*31·5*—	*33·5*—	*35·5*—	*37·5*—	*39·5*—	*41·5*—	*43·5*—	*45·5*—	*47·5*—	*49·5*—	*51·5*—	*53·5*—	*55·5*—	*57·5*—	—	*67·5*—	
31⅝—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
32 „	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
33 „	—	1	—	3	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
34 „	—	—	4	5	5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15
35 „	—	—	2	2	5	8	3	1	—	—	—	—	—	—	—	—	—	—	—	—	21
36 „	—	1	—	2	17	14	10	6	1	1	—	—	—	—	—	—	—	—	—	—	52
37 „	—	—	—	1	7	25	16	15	5	1	1	—	—	—	1	1	—	—	—	—	73
38 „	—	—	—	1	10	20	36	34	15	16	2	—	—	—	—	—	—	—	—	—	134
39 „	—	—	—	—	4	15	32	43	35	20	4	2	—	—	—	—	—	—	—	—	155
40 „	—	—	—	—	1	5	11	54	61	52	27	2	1	—	1	—	—	—	—	1	216
41 „	—	—	—	1	—	1	6	22	56	52	31	16	7	1	1	—	—	—	—	—	194
42 „	—	—	—	—	1	1	2	10	21	54	42	26	7	1	2	—	—	—	—	—	167
43 „	—	—	—	—	—	—	—	5	7	22	22	25	15	7	3	1	1	—	—	—	108
44 „	—	—	—	—	—	—	—	1	4	4	11	7	14	7	2	—	—	—	—	—	50
45 „	—	—	—	—	—	—	1	—	1	2	4	8	10	5	2	4	—	—	—	—	37
46 „	—	—	—	—	—	—	—	—	—	—	—	1	—	1	4	2	—	—	—	—	8
47 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
48 „	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	1	—	—	3
Totals	1	3	7	15	52	91	117	192	206	224	145	87	54	22	16	8	2	1	—	1	1244

TABLE 2. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 6·5—7·5 years																					Totals
	19·5—	*21·5*—	*23·5*—	*25·5*—	*27·5*—	*29·5*—	*31·5*—	*33·5*—	*35·5*—	*37·5*—	*39·5*—	*41·5*—	*43·5*—	*45·5*—	*47·5*—	*49·5*—	*51·5*—	*53·5*—	*55·5*—	*57·5*—	*59·5*—	
29⅝—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
30 „	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
31 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
32 „	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
33 „	—	—	—	—	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
34 „	—	—	—	—	—	3	1	2	2	2	1	—	—	—	—	—	—	—	—	—	—	11
35 „	—	—	—	—	—	—	2	6	1	2	—	2	1	—	—	—	—	—	—	—	—	14
36 „	—	—	—	—	—	—	3	2	3	2	—	1	—	—	—	—	—	—	—	—	—	11
37 „	—	—	—	—	—	1	—	10	6	8	3	1	1	—	—	—	—	—	—	—	—	30
38 „	—	—	—	—	—	—	—	11	11	7	14	1	3	—	—	—	—	—	—	—	—	47
39 „	—	—	—	—	—	—	4	8	13	33	32	19	10	—	—	—	—	—	—	—	—	119
40 „	2	—	—	—	—	—	—	4	12	30	37	39	19	3	1	1	—	—	—	—	—	148
41 „	—	—	—	—	—	—	—	1	4	17	32	59	43	13	16	1	1	1	—	1	—	188
42 „	—	—	—	—	—	—	—	1	2	14	25	50	52	36	28	9	—	1	—	—	—	219
43 „	—	—	—	—	—	—	—	—	—	2	5	30	60	44	39	11	2	1	—	—	—	194
44 „	—	—	—	—	—	—	—	—	—	2	2	8	30	38	36	29	7	1	2	—	—	155
45 „	—	—	—	1	—	—	—	—	—	—	1	8	6	15	30	28	17	6	7	—	—	119
46 „	—	—	—	—	—	1	—	—	—	—	—	—	—	3	10	8	14	11	9	2	—	58
47 „	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	4	4	7	2	1	—	20
48 „	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	3	—	2	1	1	9
49 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	2
50 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
51 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
52 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
Totals	2	—	1	1	1	5	14	45	55	119	153	220	225	152	161	93	48	28	23	7	1	1354

TABLE 3. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 7·5—8·5 years																								Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	
32⅝—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
33 „	1	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
34 „	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
35 „	—	—	—	—	—	—	2	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	4
36 „	—	—	—	2	3	2	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
37 „	—	—	—	2	5	4	3	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	18
38 „	—	—	—	1	6	13	3	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	25
39 „	—	—	—	2	3	10	10	3	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	32
40 „	—	—	—	1	1	11	19	16	7	6	3	—	—	—	—	—	—	—	—	—	—	—	—	—	64
41 „	—	1	—	2	4	5	18	23	22	11	7	—	2	—	—	—	—	—	—	—	—	—	—	—	95
42 „	—	—	—	—	2	4	14	36	53	24	37	4	—	—	—	—	—	—	—	—	—	—	—	—	174
43 „	—	—	—	—	—	4	7	39	44	60	43	15	9	—	1	—	—	—	—	—	—	—	—	—	222
44 „	—	—	—	—	—	1	1	22	46	52	53	34	18	6	11	—	—	—	—	—	—	—	—	—	244
45 „	—	—	—	—	—	1	—	12	9	30	54	43	33	21	6	2	—	—	—	—	—	—	—	—	211
46 „	—	—	—	—	—	—	—	1	3	14	31	33	34	19	19	8	1	—	—	1	—	—	—	—	164
47 „	—	—	—	—	—	—	—	—	1	1	12	18	31	18	22	13	7	1	—	—	—	—	—	—	124
48 „	—	—	—	—	—	—	1	—	—	1	2	6	14	14	9	12	8	2	5	—	—	—	—	—	74
49 „	—	—	—	—	—	—	—	—	1	1	1	1	5	1	8	6	3	2	2	—	—	—	—	—	31
50 „	—	—	—	—	—	—	—	—	1	—	—	1	1	—	1	5	2	2	1	—	—	—	—	1	15
51 „	—	—	—	—	—	—	—	—	1	2	3	1	1	1	—	—	—	—	—	—	—	—	—	—	9
52 „	—	—	—	—	—	—	—	—	—	—	1	2	1	1	—	—	—	1	—	—	—	1	—	—	7
53 „	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
54 „	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
55 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
56 „	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	1	—	—	—	—	—	—	—	—	4
Totals	1	1	1	13	24	56	80	157	194	202	249	159	150	83	78	47	21	8	8	1	—	1	—	1	1535

TABLE 4. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 8·5—9·5 years																											Totals
	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	
31⅝—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
35 „	1	—	—	—	—	—	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
36 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
37 „	—	1	1	—	1	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
38 „	—	—	1	—	—	—	1	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	3
39 „	—	—	1	—	1	—	6	2	3	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
40 „	—	—	1	—	3	5	6	8	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	26
41 „	—	—	1	1	2	4	9	9	5	4	1	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	39
42 „	—	—	—	—	1	2	17	10	9	4	4	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	49
43 „	—	—	—	—	1	2	11	23	25	13	15	6	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	99
44 „	—	—	—	—	—	2	8	19	33	42	30	18	6	2	—	—	—	—	—	—	—	—	—	—	—	—	—	160
45 „	—	—	—	—	—	—	6	15	23	43	45	33	23	11	6	2	—	—	—	—	—	—	—	—	—	—	—	207
46 „	—	—	—	2	—	1	1	5	10	27	44	49	37	25	10	3	2	—	—	—	—	—	—	—	—	—	—	216
47 „	—	—	—	—	—	1	1	1	6	15	23	40	36	52	20	9	1	—	1	—	—	—	—	—	—	—	—	206
48 „	—	—	—	—	—	—	1	—	—	9	8	16	24	40	26	12	13	2	1	—	—	—	—	—	—	—	—	152
49 „	—	—	—	—	—	—	—	—	—	—	4	6	9	22	19	14	18	4	2	1	1	—	—	—	—	—	—	100
50 „	—	—	—	—	—	—	—	—	—	1	—	2	1	8	12	10	11	7	3	—	1	1	—	—	—	—	—	57
51 „	—	—	—	—	—	—	—	—	—	—	1	—	2	2	3	4	4	6	4	1	2	1	—	—	—	—	—	30
52 „	—	—	—	—	—	—	—	—	1	1	—	—	1	—	1	—	2	3	—	1	1	1	—	—	—	—	—	12
53 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	2	1	1	1	—	—	—	—	—	7
54 „	—	—	—	—	—	—	—	—	1	—	—	1	1	—	—	—	—	—	—	—	1	—	—	—	1	—	—	5
55 „	—	—	—	—	—	—	—	—	—	1	—	3	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1	6
56 „	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	2
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	2
61 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
Totals	1	2	5	3	9	18	67	93	118	162	176	178	145	165	98	56	51	23	13	5	7	4	1	—	1	—	1	1402

TABLE 5. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 9·5—10·5 years																										Totals
	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	
34⅝—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
35 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
36 ,,	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
37 ,,	1	—	—	2	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	5
38 ,,	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
39 ,,	—	—	—	1	2	—	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
40 ,,	—	1	1	—	2	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
41 ,,	—	—	2	—	5	2	5	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	17
42 ,,	—	—	—	2	3	3	2	5	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	18
43 ,,	—	—	—	—	4	3	4	12	7	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	35
44 ,,	—	—	—	—	3	4	11	17	14	7	3	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	63
45 ,,	—	—	—	—	—	2	10	24	28	17	21	13	3	2	—	—	—	—	—	—	—	—	—	—	—	—	120
46 ,,	—	—	—	1	1	3	7	14	25	19	19	26	7	4	1	—	1	—	—	—	—	—	—	—	—	—	128
47 ,,	—	—	—	1	—	—	1	13	24	36	47	45	28	8	8	4	2	1	1	—	—	—	—	—	—	—	219
48 ,,	—	—	—	—	—	—	1	5	9	25	23	58	37	22	10	3	4	3	—	—	—	—	—	—	—	—	200
49 ,,	—	—	—	—	2	1	—	—	1	14	16	34	38	31	35	15	6	1	—	—	—	—	—	—	—	—	194
50 ,,	—	—	—	1	—	—	—	—	2	2	7	19	19	27	38	16	16	6	4	—	2	—	—	—	—	—	159
51 ,,	—	—	—	—	—	—	—	—	1	1	2	1	7	19	19	13	12	12	2	2	1	1	—	1	—	—	94
52 ,,	—	—	—	—	—	—	—	—	—	1	—	2	1	4	7	21	12	9	7	3	3	—	—	—	—	—	70
53 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	4	4	5	1	3	1	—	—	—	—	—	21
54 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	—	1	2	1	—	—	1	—	8
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
56 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1	—	—	—	—	—	2
57 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59 ,,	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	2
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
Totals	2	1	3	9	23	19	45	95	114	126	142	205	142	119	121	77	59	38	15	9	10	3	—	1	1	—	1379

TABLE 6. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 10·5—11·5 years																										Totals
	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	
34⅝—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
37 ,,	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
39 ,,	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	3
40 ,,	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
41 ,,	—	1	—	—	2	2	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
42 ,,	—	—	—	1	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
43 ,,	1	—	1	—	2	4	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
44 ,,	—	—	—	—	5	6	3	6	—	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	24
45 ,,	—	—	2	1	3	10	7	11	9	9	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	55
46 ,,	—	1	—	1	—	7	9	12	14	18	2	3	3	2	—	—	—	—	—	—	—	—	—	—	—	—	72
47 ,,	—	—	—	—	2	3	5	13	15	30	23	7	—	3	—	—	1	1	—	—	—	—	—	—	—	—	103
48 ,,	—	—	—	—	—	3	2	9	21	42	28	25	15	5	3	—	—	—	—	—	—	—	—	—	—	—	153
49 ,,	—	—	1	1	—	1	2	4	5	32	35	34	33	22	9	8	2	2	—	—	—	—	—	—	—	—	191
50 ,,	—	—	—	—	—	1	1	3	7	15	23	27	48	28	8	11	6	2	—	—	1	—	—	—	—	—	181
51 ,,	—	—	—	—	—	—	1	—	1	5	12	18	35	24	21	20	17	8	2	3	—	—	—	—	—	—	167
52 ,,	—	—	—	—	—	—	—	1	—	2	—	8	14	17	22	16	20	9	4	1	3	—	—	—	—	—	117
53 ,,	—	—	—	—	—	—	—	—	—	—	2	2	6	11	13	15	14	16	13	3	3	—	1	—	—	—	99
54 ,,	—	—	—	—	—	—	—	—	—	—	1	—	2	1	3	5	5	8	10	3	4	—	—	1	—	1	44
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	—	4	1	3	7	4	1	1	—	1	—	25
56 ,,	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	1	—	1	1	—	1	3	1	—	10
57 ,,	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	—	—	1	—	—	4
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	2
Totals	2	2	4	6	15	41	35	61	73	158	127	125	160	114	82	76	70	48	32	18	17	1	3	5	2	1	1278

TABLE 7. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 11·5—12·5 years																							Totals
	36·5—	39·5—	42·5—	45·5—	48·5—	51·5—	54·5—	57·5—	60·5—	63·5—	66·5—	69·5—	72·5—	75·5—	78·5—	81·5—	84·5—	87·5—	90·5—	93·5—	96·5—	99·5—	102·5—	
36⅝—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
40 „	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
41 „	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
42 „	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
43 „	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
44 „	—	—	—	—	2	4	2	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	10
45 „	—	—	1	1	3	4	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15
46 „	—	—	—	1	2	12	6	5	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	29
47 „	—	—	1	—	4	7	22	14	7	1	—	1	—	—	—	—	—	—	—	—	—	—	—	57
48 „	—	—	—	1	3	11	20	16	11	8	3	1	—	—	—	1	—	—	—	—	—	—	—	75
49 „	—	—	—	1	—	7	17	27	40	16	14	2	1	—	1	—	—	—	—	—	—	—	—	126
50 „	—	—	—	—	—	1	10	33	41	33	22	12	3	—	1	—	—	—	—	—	—	—	—	156
51 „	—	—	—	—	—	1	13	14	30	41	36	20	7	6	—	1	—	—	—	—	—	—	—	169
52 „	—	—	—	—	—	1	—	6	16	31	38	31	15	7	1	1	2	—	—	—	—	—	—	149
53 „	—	—	—	—	—	—	1	5	8	16	28	31	27	18	4	2	2	—	—	—	—	—	—	142
54 „	—	—	—	—	—	—	1	—	1	8	12	23	29	19	11	4	2	—	—	—	—	—	—	110
55 „	—	—	—	—	—	—	1	1	1	3	7	19	8	13	9	5	—	1	1	—	—	—	—	69
56 „	—	—	—	—	—	—	—	—	—	—	1	1	—	10	8	6	4	2	—	—	—	—	—	32
57 „	—	—	—	—	—	—	—	1	—	—	—	1	1	4	5	3	1	—	2	—	1	—	1	20
58 „	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	2	—	1	—	—	—	—	—	5
59 „	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	1	—	—	—	—	—	—	3
60 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
61 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
62 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
63 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
64 „	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	—	3	4	16	50	98	125	155	160	163	143	91	78	43	26	12	4	3	—	2	—	1	1178

TABLE 8. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 12·5—13·5 years																	Totals
	43·5—	47·5—	51·5—	55·5—	59·5—	63·5—	67·5—	71·5—	75·5—	79·5—	83·5—	87·5—	91·5—	95·5—	99·5—	103·5—	107·5—	
42⅝—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
43 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44 „	—	1	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
45 „	1	—	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	6
46 „	—	2	4	5	—	—	—	—	—	—	1	—	—	—	—	—	—	12
47 „	—	2	3	13	6	3	—	—	—	—	—	—	—	—	—	—	—	27
48 „	1	2	3	9	18	5	1	1	1	—	—	—	—	—	—	—	—	41
49 „	1	—	2	18	27	11	6	1	—	—	—	—	—	—	—	—	—	66
50 „	—	1	—	7	27	29	17	10	2	—	—	—	—	—	—	—	—	93
51 „	1	—	1	8	26	34	40	22	6	1	1	—	—	—	—	—	—	140
52 „	—	—	—	3	21	46	51	33	22	6	4	2	—	—	—	—	—	188
53 „	—	—	—	1	7	24	43	57	17	7	2	—	—	1	—	1	—	160
54 „	—	—	—	—	4	14	29	45	37	8	2	1	—	—	—	—	—	140
55 „	—	—	—	—	2	3	10	33	33	16	10	7	1	—	—	—	—	115
56 „	—	—	1	—	—	2	7	11	26	18	15	6	1	2	—	—	—	89
57 „	—	—	—	—	—	—	1	7	7	10	10	8	4	2	—	—	—	49
58 „	—	—	—	—	—	1	—	1	1	3	12	5	2	8	—	—	—	33
59 „	—	—	—	—	—	—	—	1	3	2	2	6	3	1	4	—	—	22
60 „	—	—	—	—	—	—	—	—	—	—	—	—	1	2	1	—	—	4
61 „	—	—	—	—	—	—	—	—	—	—	—	2	3	1	—	—	1	7
62 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	2
63 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
64 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
Totals	4	8	21	67	139	172	205	222	155	71	59	37	16	17	7	1	1	1202

TABLE 9. *Glasgow. Height and Weight of Boys of Group A.*

Height	Weight of Boys of 13·5—14·5 years																		Totals
	47·5—	51·5—	55·5—	59·5—	63·5—	67·5—	71·5—	75·5—	79·5—	83·5—	87·5—	91·5—	95·5—	99·5—	103·5—	107·5—	—	119·5—	
42⅝—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
43 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
44 ,,	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
45 ,,	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
46 ,,	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
47 ,,	—	1	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
48 ,,	—	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	3
49 ,,	1	1	3	4	4	2	—	—	—	—	—	—	—	—	—	—	—	—	15
50 ,,	1	3	1	5	7	8	2	—	—	—	—	—	—	—	—	—	—	—	27
51 ,,	—	—	1	6	5	10	3	3	—	—	—	—	—	—	—	—	—	—	28
52 ,,	—	—	—	4	11	22	4	1	1	—	—	—	—	—	—	—	—	—	43
53 ,,	—	—	1	4	12	16	15	4	7	1	—	—	—	—	—	—	—	—	60
54 ,,	—	—	—	1	4	11	9	11	6	1	—	1	1	—	—	—	—	—	45
55 ,,	—	—	—	1	1	3	17	21	7	4	1	—	1	—	—	—	—	—	56
56 ,,	—	—	—	—	—	2	4	10	10	14	6	2	—	1	—	—	—	—	49
57 ,,	—	—	1	—	—	3	2	5	7	8	5	4	—	1	—	—	—	—	36
58 ,,	—	—	—	—	—	—	—	5	6	4	2	2	4	1	—	—	—	—	24
59 ,,	—	—	—	—	—	—	—	1	1	1	3	1	1	2	2	1	—	—	13
60 ,,	—	—	—	—	—	—	—	—	—	2	—	2	1	2	1	—	—	—	8
61 ,,	—	—	—	—	—	—	—	—	—	—	—	2	1	—	1	—	—	—	4
62 ,,	—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	—	—	—	4
63 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1	2
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	3	8	11	28	46	77	56	61	45	36	17	15	10	8	5	1	—	1	428

TABLE 10. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 5·5—6·5 years																	Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	
29⅝—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
30 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
31 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
32 ,,	1	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	3
33 ,,	1	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	4
34 ,,	1	1	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	5
35 ,,	—	3	4	4	—	—	—	—	—	—	—	—	—	—	—	—	—	11
36 ,,	—	—	6	14	3	1	—	—	—	1	—	—	—	—	—	—	—	25
37 ,,	—	1	3	14	15	6	1	3	—	—	—	—	—	—	—	—	—	43
38 ,,	—	1	5	15	15	19	9	1	2	—	—	—	—	—	—	—	—	67
39 ,,	1	—	1	8	29	31	23	11	4	2	—	—	—	—	—	—	—	110
40 ,,	—	—	1	4	23	39	61	49	9	4	1	—	—	—	—	—	—	191
41 ,,	—	—	—	6	9	26	46	59	28	14	2	2	—	—	—	—	—	192
42 ,,	—	—	—	—	2	14	31	71	44	22	22	—	—	—	—	—	—	206
43 ,,	—	1	—	—	1	—	16	32	33	26	26	11	1	1	—	—	—	148
44 ,,	—	—	—	—	—	—	1	11	19	15	14	10	5	3	—	—	—	78
45 ,,	—	—	1	—	—	1	—	2	—	6	18	9	2	4	—	—	—	43
46 ,,	—	—	—	—	—	—	—	—	1	5	3	7	1	1	1	—	—	19
47 ,,	—	—	—	—	—	—	—	—	—	2	2	4	—	1	1	1	1	12
48 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	2
49 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
50 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	4	7	25	69	97	138	188	239	140	97	88	43	10	10	4	1	1	1161

TABLE 11. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 6·5—7·5 years																			Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	
29⅝—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30 „	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
31 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
32 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
33 „	—	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
34 „	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	2
35 „	—	1	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
36 „	—	1	2	3	3	2	1	—	—	—	1	—	—	—	—	—	—	—	—	13
37 „	—	—	2	5	7	6	3	—	—	—	—	—	—	—	—	—	—	—	—	23
38 „	—	1	—	6	8	14	5	4	—	—	—	—	—	—	—	—	—	—	—	38
39 „	—	1	2	3	7	12	9	6	2	5	—	—	—	—	—	—	—	—	—	47
40 „	—	—	—	3	8	27	23	23	14	4	1	—	—	—	—	—	—	—	—	103
41 „	—	—	—	3	2	17	30	44	26	16	6	—	1	1	—	—	—	—	—	146
42 „	—	—	—	—	7	11	26	44	51	32	7	6	2	1	1	—	—	—	—	188
43 „	—	—	—	—	—	2	16	47	63	59	36	16	5	1	1	—	—	—	—	246
44 „	—	—	—	—	—	1	2	17	41	49	55	27	15	5	2	1	—	—	—	215
45 „	—	—	—	—	—	—	—	2	13	35	38	33	14	7	6	1	—	—	—	149
46 „	—	—	—	—	—	—	—	2	7	10	24	22	19	15	12	1	2	—	—	114
47 „	—	—	—	—	—	—	—	—	1	4	5	12	15	13	5	4	—	1	—	60
48 „	—	—	—	—	—	—	—	—	—	—	1	3	4	7	6	2	2	2	—	27
49 „	—	—	—	—	—	—	—	—	—	1	—	—	1	1	—	3	2	—	1	9
50 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	2
Totals	1	4	10	26	44	93	115	189	218	215	174	119	76	51	33	13	7	3	1	1392

TABLE 12. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 7·5—8·5 years																						Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	
34⅝—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
35 „	—	—	1	—	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	3
36 „	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
37 „	—	—	2	2	—	1	—	1	—	1	—	1	—	—	—	—	—	—	—	—	—	—	8
38 „	—	—	1	—	2	5	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12
39 „	—	—	—	—	5	4	5	5	3	—	—	1	—	—	—	—	—	—	—	—	—	—	23
40 „	—	—	—	—	2	10	13	15	11	1	1	1	—	—	—	—	—	—	—	—	—	—	54
41 „	—	—	—	—	—	6	14	19	11	11	6	1	3	—	—	—	—	—	—	—	—	—	71
42 „	—	—	—	—	1	—	8	27	18	25	8	3	1	—	—	1	—	—	—	—	—	—	92
43 „	—	—	—	—	—	2	3	27	28	43	26	9	3	1	5	—	—	—	—	—	—	—	147
44 „	—	—	—	—	—	2	1	16	34	47	61	28	17	5	4	2	—	—	—	—	—	—	217
45 „	—	—	—	—	—	—	2	5	11	37	58	57	30	15	9	4	—	—	—	—	—	—	228
46 „	—	—	—	—	—	—	1	—	5	13	42	38	53	26	29	9	2	—	—	—	—	—	218
47 „	—	—	—	—	—	—	—	1	4	4	13	22	32	36	22	19	5	4	—	—	—	—	162
48 „	—	—	—	—	—	—	—	—	1	2	5	6	10	16	25	23	7	6	3	—	—	—	104
49 „	—	—	—	—	—	—	—	—	—	—	—	3	7	6	11	12	6	4	3	1	1	—	54
50 „	—	—	—	1	—	—	—	—	—	—	1	—	1	—	2	4	5	2	4	1	—	—	21
51 „	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	2	1	5	3	1	—	—	14
52 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1	1	—	1	4
53 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
54 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
55 „	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	2
56 „	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	2
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
Totals	1	1	4	3	10	30	51	118	126	185	222	174	158	107	108	76	26	21	14	4	2	1	1442

TABLE 13. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 8·5—9·5 years																							Totals
	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	77·5—	79·5—	
34⅝—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
35 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
36 ,,	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	2
37 ,,	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 ,,	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	2
39 ,,	—	—	1	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
40 ,,	—	1	—	4	4	2	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
41 ,,	—	1	—	1	5	7	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	18
42 ,,	—	1	1	2	5	11	10	4	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	37
43 ,,	—	—	—	1	4	22	22	14	11	6	2	2	—	—	—	—	—	—	—	—	—	—	—	84
44 ,,	—	—	—	3	6	18	26	31	19	12	6	4	2	—	—	—	—	—	—	—	—	—	—	127
45 ,,	—	—	—	—	3	16	19	48	43	29	23	13	4	3	—	—	—	—	—	—	—	—	—	201
46 ,,	—	—	—	—	1	6	10	30	40	35	38	37	14	3	1	—	—	—	—	—	—	—	—	215
47 ,,	—	—	—	—	3	2	3	17	22	47	39	48	26	16	6	2	2	—	—	—	—	—	—	233
48 ,,	—	—	—	—	—	—	4	6	18	20	29	54	35	18	12	8	3	—	—	—	—	—	—	207
49 ,,	—	—	—	—	—	—	—	4	2	8	17	27	29	29	26	10	2	3	—	—	—	—	—	157
50 ,,	—	—	—	—	—	—	—	—	2	2	4	4	11	14	11	13	6	3	2	1	1	—	—	74
51 ,,	—	—	—	—	—	—	—	—	1	—	1	5	2	3	10	8	1	3	1	—	—	—	—	35
52 ,,	—	—	—	—	—	—	—	—	1	—	1	1	—	1	4	3	2	4	2	—	—	—	—	19
53 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	1	—	1	1	1	1	—	—	8
54 ,,	—		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
55 ,,	—	—	—	—	—	—		—	—	—	—	—	—	—	—	—	1	—	1	—	—	1	1	4
56 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	—	—	—	2
57 ,,	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
58 ,,	—	—		—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
62 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
Totals	1	3	3	11	32	85	100	156	160	162	161	198	124	89	73	45	17	14	7	4	2	1	1	1449

TABLE 14. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 9·5—10·5 years																													Totals
	31·5—	—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	89·5—	
37⅝—	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
38 ,,	—	—	—	—	1	1	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
39 ,,	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	2
40 ,,	—	—	—	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
41 ,,	—	—	1	—	4	1	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
42 ,,	—	—	—	2	2	2	3	5	1	1	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	18
43 ,,	—	—	—	—	—	5	6	2	3	5	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	23
44 ,,	—	—	—	—	2	4	3	9	7	5	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	34
45 ,,	—	—	—	—	—	1	4	17	12	13	10	9	2	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	71
46 ,,	—	—	—	1	2	2	6	12	24	19	26	21	8	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	124
47 ,,	—	—	—	—	—	—	2	9	16	20	34	39	20	13	1	2	2	—	—	—	—	—	—	—	—	—	—	—	—	158
48 ,,	—	—	—	—	—	1	—	3	11	18	31	44	48	25	28	4	—	2	—	—	—	—	—	—	—	—	—	—	—	215
49 ,,	—	—	—	—	—	—	—	3	1	5	16	43	45	40	34	14	6	3	1	—	—	—	—	—	—	—	—	—	—	211
50 ,,	—	—	—	—	—	2	—	1	2	2	5	17	24	32	25	21	8	14	7	—	—	—	—	—	—	—	—	—	—	160
51 ,,	—	—	—	—	1	—	—	1	1	4	4	9	8	13	24	26	17	12	7	5	1	—	—	—	—	—	1	—	—	134
52 ,,	—	—	—	—	—	—	—	—	—	1	1	1	4	2	13	10	8	9	18	8	3	—	4	—	—	—	—	—	—	82
53 ,,	—	—	—	—	—	—	—	—	—	—	2	1	—	3	2	4	7	8	7	2	4	2	—	1	—	1	—	—	1	45
54 ,,	—	—	—	—	—	—	1	—	1	—	—	1	—	—	1	—	2	2	8	2	1	—	—	—	—	—	—	—	—	19
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	2	—	1	—	—	—	—	—	6
56 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	—	1	—	—	—	—	—	—	4
57 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
58 ,,	—	—	—	—	—	—	—	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
Totals	2	—	1	6	13	19	26	63	80	94	133	188	159	133	132	82	50	51	51	18	12	4	5	2	—	1	1	—	1	1327

TABLE 15. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 10·5—11·5 years																													Totals
	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—		93·5—	
35⅝—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
36 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
37 „	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 „	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
39 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 „	—	—	—	—	—	—	—	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
41 „	1	—	—	—	1	—	2	—	—	—	1	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	6
42 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
43 „	—	—	1	—	1	—	—	1	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
44 „	—	—	—	—	—	1	2	3	3	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
45 „	—	—	1	—	—	3	5	6	6	6	7	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	36
46 „	—	—	—	—	3	1	6	5	14	11	4	7	3	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	56
47 „	—	—	—	—	—	—	4	6	12	18	28	11	9	6	2	2	—	2	1	—	—	—	—	—	—	—	—	—	—	101
48 „	—	—	—	1	—	1	3	5	7	13	27	27	22	22	6	7	—	—	—	—	—	—	—	—	—	—	—	—	—	141
49 „	—	—	—	—	—	—	—	2	7	13	35	47	30	34	27	10	5	1	1	—	—	—	—	—	—	—	—	—	—	212
50 „	—	—	—	—	—	1	—	1	3	6	15	25	30	39	32	25	12	8	3	—	1	1	—	—	—	—	—	—	—	202
51 „	—	—	—	—	1	—	1	1	1	2	5	8	21	35	27	27	29	15	11	3	3	—	—	—	—	—	—	—	—	190
52 „	—	—	—	—	—	—	1	1	—	1	1	3	7	19	19	19	24	21	9	5	—	2	—	—	—	—	—	—	—	132
53 „	—	—	—	—	—	—	—	—	—	—	—	—	2	8	8	12	11	22	12	11	5	3	—	—	—	1	—	—	—	95
54 „	—	—	—	—	—	—	—	—	—	—	—	1	1	4	4	2	5	13	13	8	11	—	2	1	1	—	—	—	—	66
55 „	—	—	—	—	—	—	—	—	—	—	—	1	1	—	2	1	2	5	6	4	6	2	2	6	—	1	—	—	—	39
56 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	6	2	1	—	2	—	1	—	—	13
57 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
58 „	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	3
59 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	2
60 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
72 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	2	—	2	1	6	8	24	31	55	71	124	134	128	171	128	105	88	88	57	31	32	10	5	9	5	2	1	—	1	1319

TABLE 16. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 11·5—12·5 years																			Totals
	45·5—	48·5—	51·5—	54·5—	57·5—	60·5—	63·5—	66·5—	69·5—	72·5—	75·5—	78·5—	81·5—	84·5—	87·5—	90·5—	93·5—	96·5—	99·5—	
38⅝—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
42 „	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
43 „	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
44 „	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
45 „	2	—	8	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	13
46 „	—	6	7	3	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	20
47 „	1	—	9	9	14	6	2	1	—	1	—	—	—	—	—	—	—	—	—	43
48 „	—	2	10	14	15	18	11	4	—	—	—	—	—	—	—	—	—	—	—	74
49 „	—	1	7	19	40	29	18	10	2	2	—	—	—	—	—	—	—	—	—	128
50 „	—	1	2	13	25	35	29	17	9	4	1	1	—	—	—	—	—	—	—	137
51 „	—	—	2	4	25	36	43	34	24	7	2	1	1	—	—	1	—	—	—	180
52 „	—	—	—	4	5	20	41	55	36	12	12	2	1	—	—	—	—	—	—	188
53 „	—	—	—	1	2	5	19	34	61	37	19	5	—	1	—	—	—	—	—	184
54 „	—	—	—	—	1	2	4	16	22	30	17	7	3	2	—	2	—	1	—	107
55 „	—	—	—	—	—	1	5	3	9	12	26	13	9	2	—	—	—	—	—	80
56 „	—	—	—	—	—	—	—	—	4	6	15	13	6	6	3	3	—	—	—	56
57 „	—	—	—	—	—	—	—	1	—	2	8	7	6	4	5	2	—	—	—	35
58 „	—	—	—	—	—	—	—	—	—	—	1	—	—	4	4	4	—	1	—	14
59 „	—	—	—	—	—	—	—	—	1	—	1	—	—	3	1	1	1	1	1	10
60 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
61 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
62 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
63 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
64 „	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
65 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	5	13	48	70	130	155	172	175	169	113	102	49	26	22	13	13	2	4	1	1282

TABLE 17. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 12·5—13·5 years																							Totals
	31·5—	—	39·5—	43·5—	47·5—	51·5—	55·5—	59·5—	63·5—	67·5—	71·5—	75·5—	79·5—	83·5—	87·5—	91·5—	95·5—	99·5—	103·5—	107·5—	111·5—	115·5—	119·5—	
$39\frac{5}{8}$—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 „	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
41 „	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
42 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
43 „	—	—	—	—	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
44 „	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
45 „	1	—	—	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
46 „	—	—	—	—	—	—	3	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
47 „	—	—	—	—	1	3	8	4	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	17
48 „	—	—	—	1	—	4	17	21	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	49
49 „	—	—	—	—	—	1	9	21	10	5	1	—	—	—	—	—	—	—	—	—	—	—	—	47
50 „	—	—	—	—	1	1	12	20	31	22	9	2	1	—	—	—	—	—	—	—	—	—	—	99
51 „	—	—	—	—	—	2	5	19	35	31	15	6	3	—	—	—	—	—	—	—	—	—	—	116
52 „	—	—	—	—	—	—	3	15	32	50	25	14	3	2	—	—	—	—	—	—	—	—	—	144
53 „	—	—	—	—	—	2	3	6	15	61	59	30	6	1	1	—	—	—	—	—	—	—	—	184
54 „	—	—	—	—	—	—	1	1	14	17	52	30	15	7	2	—	—	—	—	—	—	—	—	139
55 „	—	—	—	—	—	1	—	2	3	20	42	43	21	14	3	1	—	1	—	—	—	—	—	151
56 „	—	—	—	—	—	—	—	2	1	7	12	27	33	13	10	1	1	1	—	—	—	—	—	108
57 „	—	—	—	—	—	—	1	—	1	2	7	13	17	24	6	5	1	—	—	—	—	—	—	77
58 „	—	—	—	—	—	—	—	—	1	2	—	5	11	10	15	3	4	2	—	1	—	—	—	54
59 „	—	—	—	—	—	—	—	—	—	—	—	1	4	4	5	1	1	1	—	—	—	—	—	17
60 „	—	—	—	—	—	—	—	—	—	—	—	—	—	1	3	1	—	2	2	1	—	—	—	10
61 „	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	—	2	1	—	—	1	—	7
62 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1
63 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
64 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1	2
65 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
66 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
67 „	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
Totals	1	—	—	1	6	17	65	113	150	219	223	171	114	78	46	13	8	9	3	2	1	1	2	1243

TABLE 18. *Glasgow. Height and Weight of Boys of Group B.*

Height	Weight of Boys of 13·5—14·5 years																			Totals
	43·5—	47·5—	51·5—	55·5—	59·5—	63·5—	67·5—	71·5—	75·5—	79·5—	83·5—	87·5—	91·5—	95·5—	99·5—	103·5—	107·5—	—	123·5—	
$42\frac{5}{8}$—	—	—	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	2
43 „	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
44 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
45 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
46 „	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
47 „	—	—	—	—	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	3
48 „	—	—	—	1	1	3	—	—	—	—	—	—	—	—	—	—	—	—	—	5
49 „	—	—	1	1	4	1	3	—	—	—	—	—	—	—	—	—	—	—	—	10
50 „	—	—	—	3	4	8	1	2	—	—	—	—	—	—	—	—	—	—	—	18
51 „	—	—	—	1	4	10	7	5	2	—	—	—	—	—	—	—	—	—	—	29
52 „	—	—	—	—	4	5	13	4	2	—	—	—	—	—	—	—	—	—	—	28
53 „	—	—	—	—	—	6	13	4	10	1	—	1	—	—	—	—	—	—	—	35
54 „	—	—	—	—	4	5	8	17	22	7	—	4	—	—	1	—	—	—	—	68
55 „	—	—	—	—	—	2	5	14	16	12	6	2	—	—	—	—	—	—	—	57
56 „	—	—	—	—	—	—	3	3	6	13	14	5	3	—	—	—	—	—	—	47
57 „	—	—	—	—	—	—	1	—	6	9	12	6	4	1	1	—	—	—	—	40
58 „	—	—	—	—	—	—	—	4	2	—	4	9	5	1	2	—	—	—	—	27
59 „	—	—	—	—	—	—	—	—	—	2	3	2	1	3	3	—	—	—	—	14
60 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	2
61 „	—	—	—	—	—	—	—	—	—	—	—	1	—	2	—	—	2	—	—	5
62 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	—	—	2
63 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
64 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
65 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
Totals	1	—	1	8	23	40	54	54	67	44	39	30	15	7	9	1	3	—	1	397

TABLE 19. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 5·5—6·5 years														Totals
	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	52·5—	—	59·5—	
28⅝—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
34 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
35 ,,	—	1	—	—	2	—	—	—	—	—	—	—	—	—	3
36 ,,	4	1	—	1	—	—	—	—	—	—	—	—		—	6
37 ,,	—	7	6	4	1	—	—	—	—	—	—	—	—	—	18
38 ,,	1	5	12	8	4	4	—	—	—	—	—	—	—	—	34
39 ,,	1	8	12	23	12	9	1	—	—	—	—	—		—	66
40 ,,	—	2	6	11	14	22	5	2	—	—	—	—	—	—	62
41 ,,	—	—	1	11	23	21	21	8	5	1	—	—	—	—	91
42 ,,	—	—	1	5	12	24	23	19	14	—	3	1	—	—	102
43 ,,	—	—	—	—	5	14	19	10	11	5	1	—	—	—	65
44 ,,	—	—	—	1	—	2	2	6	8	6	3	—	—	—	28
45 ,,	—	1	—	—	—	—	—	1	9	5	—	1	—	1	18
46 ,,	—	—	—	—	—	—	—	—	1	—	2	—	—	—	3
47 ,,	—	—	—	—	—	—	—	—	—	1	1	1	—	—	3
Totals	6	25	39	64	73	96	71	46	48	18	10	3	—	1	500

TABLE 20. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 6·5—7·5 years																	Totals
	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	
34⅝—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
35 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
36 ,,	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
37 ,,	—	—	1	—	4	1	—	—	—	1	—	—	—	—	—	—	—	7
38 ,,	—	—	—	3	2	—	—	—	—	—	—	—	—	—	—	—	—	5
39 ,,	—	—	2	2	9	10	5	1	1	—	—	—	—	—	—	—	—	30
40 ,,	—	—	2	2	11	12	9	2	2	—	—	1	—	—	—	—	—	41
41 ,,	—	—	1	4	13	12	24	15	10	3	2	—	—	—	—	—	—	84
42 ,,	—	—	—	—	4	18	28	32	17	4	4	—	—	—	—	—	—	107
43 ,,	—	—	—	—	—	7	15	27	23	24	10	5	2	1	—	—	—	114
44 ,,	—	—	—	—	—	—	5	14	24	25	27	8	3	2	—	—	—	108
45 ,,	—	—	—	—	1	1	2	2	10	15	21	9	5	6	—	—	—	72
46 ,,	—	—	—	—	—	—	1	1	3	13	7	16	3	4	1	—	—	49
47 ,,	—	—	—	—	—	—	—	1	—	—	3	5	5	3	—	—	—	17
48 ,,	—	—	—	—	—	—	—	—	—	—	—	—	2	2	—	1	1	6
49 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
50 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
51 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
52 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
Totals	1	2	8	11	44	61	90	95	93	85	74	44	20	18	1	1	2	649

TABLE 21. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 7·5—8·5 years																					Totals
	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	—	73·5—	
30⅝—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
35 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
36 „	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
37 „	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 „	—	—	1	1	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
39 „	—	1	1	2	4	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
40 „	—	—	1	1	2	2	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
41 „	—	—	1	—	6	9	6	6	4	—	—	—	—	—	—	—	—	—	—	—	—	—	32
42 „	—	—	—	2	3	14	11	5	5	—	1	2	—	—	—	—	—	—	—	—	—	—	43
43 „	—	—	—	1	5	4	17	13	18	4	3	1	—	—	—	—	—	—	—	—	—	—	66
44 „	—	—	—	1	—	5	10	20	24	15	8	8	3	—	—	—	—	—	—	—	—	—	94
45 „	—	—	—	—	1	1	10	10	22	27	12	14	3	4	—	—	—	—	—	—	—	—	104
46 „	—	—	—	—	—	—	2	7	12	22	18	16	10	3	—	—	—	—	—	—	—	—	90
47 „	—	—	—	—	—	—	—	5	4	12	12	15	14	4	4	2	—	—	—	—	—	—	72
48 „	—	—	—	—	—	—	2	1	—	6	5	12	13	5	8	2	2	2	1	1	—	—	60
49 „	—	—	—	—	—	—	—	—	1	1	—	3	5	6	3	2	—	—	—	—	—	—	21
50 „	—	—	—	—	—	—	1	1	1	—	—	—	2	2	2	1	—	—	—	—	—	—	10
51 „	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	2
52 „	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1	—	—	—	—	2
53 „	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	1	—	—	—	1	4
54 „	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	2
55 „	—	—	—	—	—	—	—	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
56 „	—	—	—	—	—	—	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—	2
57 „	—	—	—	—	—	—	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	2
58 „	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	2
Totals	1	1	5	8	23	37	66	75	94	88	61	72	52	26	17	7	2	4	1	1	—	1	642

TABLE 22. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 8·5—9·5 years																	Totals
	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	
35⅝—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
36 „	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
37 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
38 „	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
39 „	—	—	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	3
40 „	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
41 „	—	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
42 „	—	1	2	2	5	4	—	—	—	1	—	—	—	—	—	—	—	—	15
43 „	—	1	4	5	4	9	4	3	1	—	1	—	—	—	—	—	—	—	32
44 „	—	1	2	11	3	12	11	5	4	—	2	—	—	—	—	—	—	—	51
45 „	—	—	2	6	6	17	11	11	5	1	5	1	1	—	—	—	—	—	66
46 „	—	—	1	3	5	5	17	16	18	11	2	4	—	—	—	—	—	—	82
47 „	—	—	—	—	2	5	8	25	30	18	9	9	3	—	—	—	—	—	109
48 „	—	—	1	—	1	1	4	12	15	32	17	8	3	2	1	—	—	—	97
49 „	—	1	1	1	1	—	—	2	11	7	12	13	3	9	1	3	—	—	65
50 „	—	—	—	—	—	—	1	—	—	4	11	11	8	7	7	2	—	—	51
51 „	—	—	1	1	—	—	1	1	—	—	2	4	5	2	1	1	1	2	22
52 „	—	—	—	—	—	—	—	1	—	1	2	2	3	1	—	4	2	1	17
53 „	—	—	—	—	1	2	1	—	—	—	—	—	1	—	—	—	—	—	5
54 „	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
55 „	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	2
56 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
57 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
58 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
Totals	1	7	17	33	28	56	59	76	85	75	66	52	28	21	10	10	3	3	630

TABLE 23. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 9·5—10·5 years																							Totals
	33·5—	—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	
39⅝—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 „	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
41 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
42 „	—	—	—	—	1	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	3
43 „	—	—	—	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
44 „	—	—	—	—	1	3	5	3	3	—	—	—	1	—	—	—	—	1	—	—	—	—	—	—	17
45 „	—	—	—	—	—	1	5	4	6	4	4	2	—	—	—	—	—	—	—	—	—	—	—	—	26
46 „	—	—	1	1	2	—	3	9	9	14	7	7	5	—	—	—	—	—	—	—	—	—	—	—	58
47 „	—	—	—	—	1	—	4	8	7	17	10	11	4	—	1	1	—	1	—	—	—	—	—	—	65
48 „	—	—	—	—	—	—	—	6	5	20	19	26	18	11	2	—	1	2	—	—	—	—	—	—	110
49 „	—	—	—	—	—	—	—	1	2	6	18	25	11	24	4	4	3	—	—	1	—	—	—	—	99
50 „	—	—	—	—	—	—	—	—	—	3	6	10	18	23	10	8	6	3	—	—	1	—	—	—	88
51 „	—	—	—	—	—	—	—	—	1	1	2	10	5	13	10	10	4	4	2	—	—	—	—	—	62
52 „	—	—	—	—	—	—	—	—	—	—	1	—	6	8	7	3	3	7	4	4	—	1	—	—	44
53 „	—	—	—	—	—	—	—	—	—	—	—	—	—	2	2	2	7	3	3	—	—	1	—	—	20
54 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	4	—	1	3	—	—	1	11
55 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	—	3
56 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	2
57 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
58 „	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	—	2	1	5	6	19	32	33	66	69	91	68	81	37	28	25	25	10	7	5	2	1	2	616

TABLE 24. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 10·5—11·5 years																									Totals
	39·5—	—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	—	89·5—	
35⅝—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
39 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
41 „	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
42 „	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
43 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44 „	—	—	—	—	1	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
45 „	—	—	1	1	2	—	1	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
46 „	1	—	—	1	2	1	3	2	2	3	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	17
47 „	—	—	—	—	1	4	1	5	9	4	8	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	36
48 „	—	—	—	—	1	2	3	6	16	11	12	8	2	1	—	2	1	—	—	—	—	—	—	—	—	—	65
49 „	—	—	—	—	—	1	1	4	4	9	13	13	12	1	3	—	—	—	—	—	—	—	—	—	—	—	61
50 „	—	—	—	—	—	—	1	2	5	10	22	17	15	11	8	4	1	—	—	—	—	—	—	—	—	—	96
51 „	—	—	—	—	—	1	—	1	2	6	12	17	10	12	8	6	2	1	—	2	—	—	—	—	—	—	80
52 „	—	—	—	—	—	—	—	—	2	1	6	11	17	6	7	14	4	3	2	2	—	—	—	—	—	—	75
53 „	—	—	—	—	—	—	—	—	—	—	1	3	7	7	9	7	10	5	3	3	3	1	—	—	—	—	59
54 „	—	—	—	—	—	—	—	—	—	1	—	2	2	—	3	6	4	4	3	1	2	1	1	—	—	—	30
55 „	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	3	3	5	2	—	2	—	1	—	—	—	18
56 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	2	1	—	1	—	—	—	6
57 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	1	1	—	—	1	6
58 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
Totals	1	—	3	3	7	9	11	24	43	46	75	74	67	39	39	44	25	18	12	11	8	3	4	1	—	1	568

TABLE 25. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 11·5—12·5 years																						Totals
	45·5—	48·5—	51·5—	54·5—	57·5—	60·5—	63·5—	66·5—	69·5—	72·5—	75·5—	78·5—	81·5—	84·5—	87·5—	90·5—	93·5—	96·5—	|	103·5—	|	111·5—	
$35\frac{5}{8}$—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
41 ,,	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
43 ,,	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
44 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
45 ,,	2	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
46 ,,	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
47 ,,	1	1	1	5	3	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	13
48 ,,	—	—	4	9	7	4	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	26
49 ,,	—	—	2	5	5	14	7	4	4	—	—	—	—	—	—	—	—	—	—	—	—	—	41
50 ,,	—	—	1	7	9	18	12	6	11	2	—	2	—	—	—	—	1	—	—	—	—	—	69
51 ,,	—	—	2	2	9	14	28	23	7	7	—	1	—	—	—	—	—	—	—	—	—	—	93
52 ,,	—	—	—	—	3	15	19	18	19	4	6	1	1	—	—	—	—	—	—	—	—	—	86
53 ,,	—	—	—	—	—	6	7	15	18	17	5	2	2	—	—	—	—	—	—	—	—	—	72
54 ,,	—	—	—	—	—	1	9	11	17	13	15	6	6	2	1	—	—	—	—	—	—	—	81
55 ,,	—	—	—	—	—	—	1	4	4	5	8	6	4	3	2	—	—	—	—	—	—	—	37
56 ,,	—	—	—	—	—	—	1	—	4	3	5	8	10	3	—	2	—	1	—	—	—	—	37
57 ,,	—	—	—	—	—	—	—	—	4	2	2	3	3	—	2	1	—	—	—	—	—	—	17
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	3	2	1	—	—	1	—	—	—	—	7
59 ,,	—	—	—	—	—	1	—	—	—	1	—	1	1	1	1	—	—	1	—	1	—	—	8
60 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
61 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
62 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
65 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
66 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
Totals	5	4	13	29	36	74	86	83	88	54	42	30	30	11	7	3	1	3	—	1	—	1	601

TABLE 26. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 12·5—13·5 years																			Totals
	47·5—	51·5—	55·5—	59·5—	63·5—	67·5—	71·5—	75·5—	79·5—	83·5—	87·5—	91·5—	95·5—	99·5—	103·5—	|	111·5—	|	123·5—	
$44\frac{5}{8}$—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
45 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
46 ,,	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
47 ,,	—	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
48 ,,	—	—	3	4	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	9
49 ,,	—	2	6	3	2	3	2	—	—	—	—	—	—	—	—	—	—	—	—	18
50 ,,	—	—	4	7	12	7	1	—	—	—	—	—	—	—	—	—	—	—	—	31
51 ,,	—	1	4	6	24	20	7	1	2	—	—	—	—	—	—	—	—	—	—	65
52 ,,	—	1	1	9	25	24	15	5	—	—	—	—	—	—	—	—	—	—	—	80
53 ,,	—	—	—	1	8	16	21	7	4	4	—	—	—	1	—	—	—	—	—	62
54 ,,	—	—	—	—	3	9	19	18	15	7	2	—	—	1	—	—	—	—	—	74
55 ,,	—	—	—	—	1	11	16	24	15	9	2	—	—	—	—	—	—	—	—	78
56 ,,	—	—	—	—	—	2	4	11	6	19	4	3	1	—	—	—	—	—	—	50
57 ,,	—	—	—	—	—	1	1	12	9	13	6	2	2	—	—	—	—	—	—	46
58 ,,	—	—	—	—	—	—	—	1	3	4	7	7	1	1	—	—	—	—	—	24
59 ,,	—	—	—	—	—	—	—	1	—	2	2	3	3	2	1	—	—	—	—	14
60 ,,	—	—	—	—	—	—	—	—	—	—	1	4	1	1	2	—	—	—	—	9
61 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	2	1	—	1	—	—	5
62 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	2	—	—	—	1	4
63 ,,	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	1	—	—	3
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
65 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
66 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
67 ,,	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	2	5	20	32	77	94	88	80	54	58	25	19	8	9	6	—	4	—	1	582

TABLE 27. *Glasgow. Height and Weight of Boys of Group C.*

Height	Weight of Boys of 13·5—14·5 years																		Totals
	55·5—	*59·5*—	*63·5*—	*67·5*—	*71·5*—	*75·5*—	*79·5*—	*83·5*—	*87·5*—	*91·5*—	*95·5*—	*99·5*—	*103·5*—	*107·5*—	*111·5*—	*115·5*—	*119·5*—	*123·5*—	
$48\frac{5}{8}$—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
49 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
50 ,,	—	—	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
51 ,,	—	2	1	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	6
52 ,,	—	2	5	2	6	2	—	—	—	—	—	—	—	—	—	—	—	—	17
53 ,,	1	2	4	4	6	7	1	1	—	—	—	—	—	—	—	—	—	—	26
54 ,,	1	—	2	1	6	4	6	4	—	—	—	—	—	—	—	—	—	—	24
55 ,,	1	—	—	2	5	16	4	7	—	—	—	—	—	—	—	—	—	—	35
56 ,,	—	1	—	1	5	10	13	4	3	1	2	1	—	—	—	—	—	—	41
57 ,,	—	—	—	1	3	6	8	7	5	2	1	—	—	—	—	—	—	—	33
58 ,,	—	—	—	1	1	2	2	4	7	4	3	—	1	—	—	—	—	—	25
59 ,,	—	—	1	—	—	1	—	1	3	6	—	2	1	—	—	—	—	—	15
60 ,,	—	—	—	—	—	1	1	1	1	—	2	3	1	1	1	—	—	—	12
61 ,,	—	—	—	—	—	—	—	—	—	1	—	—	1	1	—	1	—	—	4
62 ,,	—	—	—	—	—	—	—	—	—	—	—	1	1	3	—	1	—	—	6
63 ,,	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	—	—	3
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
65 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	1	3
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
68 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
69 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	4	8	15	14	35	49	35	29	19	15	8	8	7	6	2	3	—	2	259

TABLE 28. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 5·5—6·5 years																		Totals
	21·5—	—	*27·5*—	*29·5*—	*31·5*—	*33·5*—	*35·5*—	*37·5*—	*39·5*—	*41·5*—	*43·5*—	*45·5*—	*47·5*—	*49·5*—	*51·5*—	*53·5*—	*55·5*—	*57·5*—	
$36\frac{5}{8}$—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
37 ,,	—	—	—	—	1	3	1	2	1	—	—	—	1	—	—	—	—	—	9
38 ,,	—	—	—	—	—	4	4	3	3	—	1	—	—	—	—	—	—	—	15
39 ,,	—	—	—	—	—	3	6	4	2	1	1	—	—	—	—	—	—	—	17
40 ,,	—	—	—	—	—	1	6	16	12	14	4	1	—	—	—	—	—	—	54
41 ,,	—	—	—	—	1	1	2	6	17	24	12	8	—	—	—	—	—	—	71
42 ,,	—	—	—	—	—	1	1	8	12	23	31	11	4	—	1	—	—	—	92
43 ,,	—	—	—	—	—	—	—	1	5	22	20	8	11	3	1	—	—	—	71
44 ,,	1	—	—	—	1	—	—	—	—	2	12	11	11	2	3	1	—	—	44
45 ,,	—	—	—	—	—	—	—	—	—	1	2	7	5	3	4	1	1	1	25
46 ,,	—	—	—	—	—	—	—	—	—	—	—	1	3	2	4	2	—	—	12
47 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	—	—	4
48 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
Totals	1	—	1	—	3	14	20	40	52	87	83	47	37	11	14	5	1	1	417

TABLE 29. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 6·5—7·5 years																Totals
	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	
36⅝—	1	—	2	—	—	—	—	1	—	—	—	—	—	—	—	—	4
37 „	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 „	—	—	1	—	2	—	—	—	—	—	—	—	—	—	—	—	3
39 „	—	—	2	2	2	1	1	—	—	—	—	—	—	—	—	—	8
40 „	—	2	5	2	3	2	2	1	1	—	—	—	—	—	—	—	18
41 „	—	1	3	5	7	10	4	3	3	—	—	—	—	—	—	—	36
42 „	—	—	—	4	10	17	19	10	—	1	—	—	—	—	—	—	61
43 „	—	—	—	1	8	18	37	18	16	2	1	—	—	—	—	—	101
44 „	—	—	—	—	—	10	17	32	23	13	5	2	—	—	—	—	102
45 „	—	—	—	—	—	7	6	11	22	14	9	3	2	—	—	—	74
46 „	—	—	—	—	—	—	3	8	11	9	8	8	7	1	—	—	55
47 „	—	—	—	—	—	—	1	—	4	5	5	4	3	1	1	—	24
48 „	—	—	—	—	—	—	—	—	—	—	—	3	3	3	1	2	12
49 „	—	—	—	—	—	—	—	—	—	—	1	—	3	1	—	1	6
50 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
Totals	1	4	13	14	32	65	90	84	80	44	29	20	19	6	2	3	506

TABLE 30. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 7·5—8·5 years																				Totals
	31·5—	—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	
34⅝—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
35 „	1	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	3
36 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
37 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
38 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
39 „	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 „	—	—	—	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	3
41 „	—	—	—	3	1	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	8
42 „	—	—	2	2	—	12	5	4	1	—	—	—	—	—	—	—	—	—	—	—	26
43 „	—	—	—	—	—	9	16	11	6	3	1	—	—	—	—	—	—	—	—	—	46
44 „	—	—	—	—	2	3	13	15	22	7	6	3	—	—	—	—	—	—	—	—	71
45 „	—	—	—	—	1	4	4	13	32	11	12	7	4	1	—	—	—	—	—	—	89
46 „	—	—	—	—	—	1	1	8	20	29	28	13	5	3	—	—	—	—	—	—	108
47 „	—	—	—	—	—	—	—	2	9	16	20	23	22	5	3	—	—	—	—	—	100
48 „	—	—	—	—	—	—	—	1	2	4	9	17	18	7	2	1	3	—	—	—	64
49 „	—	—	—	—	—	—	—	—	—	1	3	6	9	6	7	4	1	1	—	—	38
50 „	—	—	—	—	—	—	—	—	—	—	—	—	3	3	2	4	1	2	—	1	16
51 „	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	1	—	—	4
52 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
53 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
54 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1	—	—	2
Totals	1	—	2	5	4	33	42	55	94	72	80	69	63	26	14	9	5	6	—	1	581

TABLE 31. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 8·5—9·5 years																						Totals
	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	—	93·5—	
37⅝—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
39 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
40 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
41 ,,	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
42 ,,	—	1	2	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
43 ,,	—	4	1	4	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
44 ,,	—	1	2	2	3	5	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	19
45 ,,	—	2	3	1	10	10	5	1	3	4	—	—	—	—	—	—	—	—	—	—	—	—	39
46 ,,	—	—	—	7	13	13	10	7	5	3	1	—	—	—	—	—	—	—	—	—	—	—	59
47 ,,	—	—	—	—	4	15	24	17	15	8	3	2	—	—	—	—	—	—	—	—	—	—	88
48 ,,	—	—	—	—	4	6	15	21	19	16	12	7	1	—	—	—	—	—	—	—	—	—	101
49 ,,	—	—	—	—	—	1	3	11	22	10	12	11	3	3	1	1	—	—	—	—	—	—	78
50 ,,	—	—	—	—	—	3	2	2	3	11	20	10	9	3	—	—	—	—	—	—	—	—	63
51 ,,	—	—	—	—	—	—	—	—	1	2	2	5	5	3	5	—	3	—	—	—	—	—	26
52 ,,	—	—	—	—	—	—	—	—	—	1	1	4	6	4	2	—	—	—	1	—	—	—	19
53 ,,	—	—	—	—	—	—	—	—	—	—	1	2	1	2	1	1	2	—	—	1	—	—	11
54 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	2
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
58 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	8	8	16	36	54	64	61	71	55	52	41	25	15	9	2	6	1	1	1	—	1	528

TABLE 32. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 9·5—10·5 years																						Totals
	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	—	91·5—	
36⅝—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
37 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
38 ,,	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
43 ,,	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
44 ,,	1	1	2	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
45 ,,	1	3	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
46 ,,	1	3	3	1	6	5	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	23
47 ,,	—	—	3	11	8	8	8	3	4	3	—	—	—	—	—	—	—	—	—	—	—	—	48
48 ,,	—	—	1	3	4	12	15	9	3	4	1	2	1	—	—	—	—	—	—	—	—	—	55
49 ,,	—	—	—	2	5	6	23	17	20	15	5	1	—	—	—	—	—	—	—	—	—	—	94
50 ,,	—	—	—	1	2	2	13	17	22	26	14	9	3	3	2	1	—	—	—	—	—	—	115
51 ,,	—	—	—	—	—	1	4	2	9	17	18	13	8	7	3	2	—	—	—	—	—	—	84
52 ,,	—	—	—	—	—	—	2	—	4	9	9	14	8	6	4	3	—	—	—	—	—	1	60
53 ,,	—	—	—	—	—	—	—	—	—	3	4	6	2	11	7	2	2	1	2	—	—	—	40
54 ,,	—	—	—	—	—	—	—	1	—	1	—	—	2	4	2	2	—	—	—	—	—	—	12
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	—	—	—	—	2
56 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	—	—	—	—	1	—	—	3
57 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
61 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
Totals	4	8	10	24	27	34	68	50	63	78	52	45	25	31	21	10	2	2	2	1	—	1	558

TABLE 33. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 10·5—11·5 years																										Totals
	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	89·5—	91·5—	93·5—	95·5—	
45⅝—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
46 „	—	—	3	3	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
47 „	—	1	2	2	4	3	4	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	18
48 „	1	—	3	2	4	6	6	7	5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	35
49 „	—	—	1	1	1	10	17	13	12	8	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	67
50 „	—	—	1	—	2	5	11	15	14	11	8	5	4	1	1	1	—	—	—	—	—	—	—	—	—	—	79
51 „	—	—	—	1	—	—	6	10	20	14	16	9	6	5	3	—	—	—	—	—	—	—	—	—	—	—	90
52 „	—	—	—	—	—	1	3	4	10	14	30	17	12	11	4	1	1	1	—	—	—	—	—	—	—	—	109
53 „	—	—	—	—	—	1	2	3	3	11	10	7	14	3	5	5	3	1	1	1	—	—	—	—	—	—	70
54 „	—	—	—	—	—	—	—	—	—	5	5	6	8	4	10	5	3	1	5	3	—	1	—	—	—	—	56
55 „	—	—	—	—	—	—	—	—	—	—	2	4	3	6	4	4	3	3	1	1	—	1	—	—	—	—	32
56 „	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	2	2	3	1	—	—	1	—	—	—	—	12
57 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	—	—	2	1	1	1	—	1	—	—	—	9
58 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	1	3
59 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	2	—	—	4
60 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	1	11	9	13	26	49	53	64	65	75	49	48	31	31	18	12	12	10	7	2	3	1	2	—	1	594

TABLE 34. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 11·5—12·5 years																							Totals
	42·5—	45·5—	48·5—	51·5—	54·5—	57·5—	60·5—	63·5—	66·5—	69·5—	72·5—	75·5—	78·5—	81·5—	84·5—	87·5—	90·5—	93·5—	96·5—	99·5—	102·5—	—	114·5—	
39⅝—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
42 „	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
43 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
45 „	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
46 „	—	—	—	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
47 „	—	—	—	2	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
48 „	—	—	—	1	3	3	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
49 „	—	—	—	2	2	6	6	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	20
50 „	—	—	—	2	2	13	9	6	7	2	—	—	—	—	—	—	—	—	—	—	—	—	—	41
51 „	—	—	—	—	3	7	9	12	12	6	6	3	1	—	—	—	—	—	—	—	—	—	—	59
52 „	—	—	—	1	1	3	8	18	20	21	6	3	2	—	—	—	—	—	—	—	—	—	—	83
53 „	—	—	—	—	1	2	7	5	27	37	14	8	2	1	—	—	—	—	—	—	—	—	—	104
54 „	—	—	—	—	—	—	4	9	11	23	25	11	4	1	1	—	1	—	—	—	—	—	—	90
55 „	—	—	—	1	—	—	—	1	3	15	11	10	8	1	1	1	1	—	—	—	—	—	—	53
56 „	—	—	—	—	—	—	—	—	1	5	9	7	10	10	3	—	—	—	1	—	—	—	—	46
57 „	—	—	—	—	—	—	—	1	—	1	3	4	4	2	3	—	2	—	—	—	—	—	—	20
58 „	—	—	—	—	—	—	—	—	—	1	2	2	1	4	4	3	—	—	2	—	1	—	—	20
59 „	—	—	—	—	—	—	—	—	—	—	—	1	—	2	1	2	1	—	—	1	—	—	—	8
60 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	1	1	—	—	4
61 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
65 „	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	—	1	10	13	39	41	54	83	112	77	50	32	21	13	6	7	—	3	2	2	—	1	571

TABLE 35. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 12·5—13·5 years																			Totals
	47·5—	51·5—	55·5—	59·5—	63·5—	67·5—	71·5—	75·5—	79·5—	83·5—	87·5—	91·5—	95·5—	99·5—	103·5—	107·5—	111·5—	115·5—	119·5—	
$44\frac{5}{8}$—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
45 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
46 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
47 ,,	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
48 ,,	—	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
49 ,,	—	—	2	4	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
50 ,,	—	1	2	6	4	2	1	—	—	—	—	—	—	—	—	—	—	—	—	16
51 ,,	—	1	4	6	14	9	1	2	—	—	—	—	—	—	—	—	—	—	—	37
52 ,,	—	—	1	7	12	17	5	6	—	—	—	—	—	—	—	—	—	—	—	48
53 ,,	—	—	—	6	12	17	16	16	3	1	—	—	1	—	—	—	—	—	—	72
54 ,,	—	—	—	1	12	19	22	21	7	2	—	—	—	—	—	—	—	—	—	84
55 ,,	—	—	—	1	3	11	25	34	14	6	3	3	1	—	—	—	—	—	—	101
56 ,,	—	—	—	—	1	2	19	24	16	14	2	2	—	—	—	—	—	—	—	80
57 ,,	—	—	—	—	—	—	9	10	16	20	7	3	1	2	—	—	—	—	—	68
58 ,,	—	—	—	—	—	1	2	5	8	14	4	4	3	1	—	—	—	—	—	42
59 ,,	—	—	—	—	—	1	1	1	—	2	4	7	3	—	—	1	1	—	—	21
60 ,,	—	—	—	—	—	—	—	—	—	3	1	7	1	1	—	1	—	—	—	14
61 ,,	—	—	—	—	—	—	—	—	1	—	1	—	2	2	—	1	1	1	—	9
62 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	1	4
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	2	3	12	31	59	79	101	119	65	62	22	26	12	8	1	3	2	2	2	611

TABLE 36. *Glasgow. Height and Weight of Boys of Group D.*

Height	Weight of Boys of 13·5—14·5 years																							Totals
	51·5—	55·5—	59·5—	63·5—	67·5—	71·5—	75·5—	79·5—	83·5—	87·5—	91·5—	95·5—	99·5—	103·5—	107·5—	111·5—	115·5—	119·5—	123·5—	127·5—	131·5—	—	147·5—	
$45\frac{5}{8}$—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
50 ,,	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
51 ,,	—	—	7	3	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12
52 ,,	—	—	3	5	5	6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	19
53 ,,	—	—	1	5	4	10	2	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	25
54 ,,	—	—	1	4	5	8	15	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	37
55 ,,	—	—	—	2	3	4	13	10	4	2	—	1	—	—	—	—	—	—	—	—	—	—	—	39
56 ,,	—	—	—	1	5	12	13	12	11	4	2	—	—	—	—	—	—	—	—	—	—	—	—	60
57 ,,	—	—	—	—	1	4	12	9	7	7	—	2	—	2	—	—	—	—	—	—	—	—	—	44
58 ,,	—	—	—	—	—	2	6	10	11	12	4	6	1	1	—	1	—	—	—	—	—	—	—	54
59 ,,	—	—	—	—	—	—	2	3	3	10	2	3	1	—	—	—	—	—	—	—	—	—	—	24
60 ,,	—	—	—	—	—	—	1	1	4	5	2	8	6	4	1	1	—	—	—	—	—	—	—	33
61 ,,	—	—	—	—	1	—	—	—	—	1	1	5	3	—	1	2	—	—	—	—	—	—	—	14
62 ,,	—	—	—	—	—	—	—	—	—	—	2	1	2	—	1	1	1	—	—	—	—	—	—	8
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	—	—	—	1	—	—	—	—	4
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
65 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	1	—	1	—	—	5
66 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
70 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	1	—	12	22	25	47	64	52	40	41	13	26	14	7	6	5	2	1	2	—	2	—	1	383

TABLE 37. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 5·5—6·5 years																		Totals
	21·5—	23·5—	25·5—	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	
25⅞—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
26 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27 „	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
28 „	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
29 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
31 „	—	—	—	1	1	—	—	—	1	—	—	—	—	—	—	—	—	—	3
32 „	—	—	2	1	1	1	—	—	—	1	—	—	—	—	—	—	—	—	6
33 „	—	—	—	—	3	1	1	—	1	1	1	—	—	—	—	—	—	—	8
34 „	—	—	—	2	3	4	5	1	—	—	1	—	—	—	—	—	—	—	16
35 „	—	—	—	3	8	12	10	7	1	1	1	—	—	—	1	—	—	—	44
36 „	—	—	—	—	3	12	12	15	5	1	—	—	—	—	1	—	—	—	49
37 „	1	—	1	—	3	9	30	13	11	10	2	1	—	1	—	—	—	—	82
38 „	—	—	—	1	1	7	28	34	29	17	4	3	—	—	—	—	—	—	124
39 „	—	—	—	—	—	4	13	43	59	23	19	6	—	2	—	1	—	—	170
40 „	—	—	—	—	1	—	7	24	42	55	46	14	6	2	—	—	—	—	197
41 „	—	—	—	—	—	—	3	8	20	45	57	19	5	5	2	1	—	—	165
42 „	—	1	—	—	—	—	—	1	9	23	51	29	17	11	2	1	1	—	146
43 „	—	—	—	—	—	—	—	—	3	5	15	21	10	8	8	2	—	—	72
44 „	—	—	—	—	—	—	—	—	1	3	7	6	7	10	4	—	—	—	38
45 „	—	—	—	—	—	—	1	—	—	1	1	1	1	2	3	—	1	—	11
46 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	2	—	1	6
47 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	1	3
Totals	1	1	4	8	25	50	110	146	183	186	205	100	47	41	24	8	2	2	1143

TABLE 38. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 6·5—7·5 years																			Totals
	25·5—	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	
30⅞—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
31 „	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
32 „	—	2	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
33 „	—	—	—	3	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	4
34 „	—	—	1	2	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	6
35 „	—	—	1	2	7	—	4	2	—	1	1	—	—	—	—	—	—	—	—	18
36 „	—	1	—	7	1	5	5	1	2	—	—	1	—	—	—	—	—	—	—	23
37 „	—	—	1	2	10	14	6	5	—	—	—	—	—	—	—	—	—	—	—	38
38 „	—	—	—	2	15	17	11	13	7	2	—	1	1	—	—	—	—	—	—	69
39 „	—	—	—	4	16	21	34	18	17	5	3	—	1	—	—	—	—	—	—	119
40 „	—	—	—	2	3	21	37	44	32	22	1	2	—	—	—	—	—	—	—	164
41 „	—	—	1	—	3	11	33	49	63	28	17	7	4	1	—	1	—	—	—	218
42 „	—	—	—	—	2	2	11	35	70	46	22	22	2	2	1	—	—	—	—	215
43 „	—	—	—	1	—	2	8	16	49	49	36	35	10	3	1	—	—	—	1	211
44 „	—	—	—	—	—	—	2	5	21	28	40	37	20	8	1	2	1	1	—	166
45 „	—	—	—	—	—	—	—	—	4	6	18	24	18	13	7	1	—	—	—	91
46 „	—	—	—	—	—	—	—	—	—	2	3	12	8	2	5	2	—	—	—	34
47 „	—	—	—	—	—	—	—	—	—	2	2	1	3	—	3	1	3	1	1	17
48 „	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	2	1	—	—	6
Totals	1	3	5	25	58	95	153	189	265	191	143	143	67	30	19	9	5	2	2	1405

TABLE 39. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 7·5—8·5 years																								Totals
	25·5—	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	—	75·5—	
29⅞—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
30 ,,	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
31 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
32 ,,	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
33 ,,	1	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
34 ,,	—	—	—	1	1	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
35 ,,	—	—	—	1	1	2	1	—	1	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	8
36 ,,	—	—	1	2	2	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
37 ,,	—	—	—	—	1	1	5	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
38 ,,	—	—	—	1	5	4	11	6	5	3	—	1	—	—	—	—	—	—	—	—	—	—	—	—	36
39 ,,	—	—	—	—	8	5	13	8	4	3	—	—	1	—	—	—	—	—	—	—	—	—	—	—	42
40 ,,	—	—	—	—	1	10	15	15	19	11	2	4	3	—	—	—	—	—	—	—	—	—	—	—	80
41 ,,	—	—	—	—	1	3	13	31	41	15	13	2	1	—	1	—	—	—	—	—	—	—	—	—	121
42 ,,	—	—	—	1	2	6	8	28	46	41	27	14	2	3	—	—	—	—	—	—	—	—	—	—	178
43 ,,	—	—	—	—	—	1	3	14	39	54	37	42	17	6	—	2	—	—	—	—	—	—	—	—	215
44 ,,	—	—	—	—	1	—	2	3	13	42	53	49	29	14	8	5	2	1	—	—	—	—	—	—	222
45 ,,	—	—	—	—	1	—	—	1	11	20	31	39	30	16	14	6	1	—	—	—	—	—	—	—	170
46 ,,	—	—	—	—	—	1	—	1	—	6	18	32	31	19	22	13	4	—	1	—	—	—	—	—	148
47 ,,	—	—	—	—	—	—	—	—	—	1	4	10	12	10	20	9	10	—	—	—	—	—	—	—	76
48 ,,	—	—	—	—	—	—	—	—	1	1	1	3	5	8	3	8	7	2	2	—	—	1	—	—	42
49 ,,	—	—	—	—	—	—	—	—	—	1	—	—	1	3	3	2	5	2	1	1	1	—	—	—	20
50 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
51 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
52 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
53 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	1	3
Totals	2	—	1	7	25	39	71	108	183	199	189	197	132	80	71	46	29	5	5	1	1	1	—	1	1393

TABLE 40. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 8·5—9·5 years																							Totals
	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	—	77·5—	
32⅞—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
33 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
34 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
35 ,,	—	—	1	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
36 ,,	—	—	2	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
37 ,,	—	1	1	—	—	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	5
38 ,,	—	—	1	3	6	3	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14
39 ,,	—	—	2	2	5	1	1	—	1	—	1	—	1	—	—	—	—	—	—	—	—	—	—	14
40 ,,	—	—	—	1	5	3	11	3	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	26
41 ,,	—	—	—	3	3	6	11	12	4	6	—	1	—	—	—	—	—	—	—	—	—	—	—	46
42 ,,	—	—	—	—	3	11	17	13	18	5	3	1	1	—	—	—	—	—	—	—	—	—	—	72
43 ,,	—	—	—	2	4	7	14	19	17	29	20	6	3	—	—	—	—	—	—	—	—	—	—	121
44 ,,	—	—	—	—	—	4	17	22	34	47	20	7	5	3	—	—	—	1	—	—	—	—	—	160
45 ,,	—	—	—	—	—	2	7	12	33	51	46	33	10	10	3	—	2	—	—	—	—	—	—	209
46 ,,	—	—	—	—	—	—	4	5	15	42	40	30	32	28	10	2	3	—	—	—	—	—	1	212
47 ,,	—	—	—	—	—	—	1	2	7	17	26	32	23	30	12	6	6	1	—	—	—	—	—	163
48 ,,	—	—	—	—	—	—	—	2	2	11	6	15	24	38	15	12	4	3	1	—	—	—	—	133
49 ,,	—	—	—	—	—	—	—	—	—	—	7	3	11	9	16	9	7	2	1	—	—	—	—	65
50 ,,	—	—	—	—	—	—	—	—	1	1	1	—	1	3	10	9	1	2	3	4	1	—	—	37
51 ,,	—	—	—	—	—	—	—	—	1	—	—	—	1	1	—	2	4	3	4	1	—	—	—	17
52 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	4	1	—	—	—	—	7
53 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	2
54 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
55 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	3	7	12	27	38	85	90	135	214	170	129	112	122	68	40	28	16	10	5	1	—	1	1314

TABLE 41. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 9·5—10·5 years																									Totals
	29·5—	—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	
34⅞—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
35 ,,	—	—	—	—	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
36 ,,	1	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
37 ,,	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 ,,	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
39 ,,	—	—	—	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
40 ,,	—	—	2	1	—	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
41 ,,	—	—	—	1	6	8	4	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	23
42 ,,	—	—	1	1	3	5	10	7	7	1	1	3	—	—	—	—	—	—	—	—	—	—	—	—	—	39
43 ,,	—	—	1	—	3	6	6	8	8	4	3	—	—	1	—	—	—	—	—	—	—	—	—	—	—	40
44 ,,	—	—	1	—	4	4	13	15	22	12	6	4	5	1	1	—	—	—	—	—	—	—	—	—	—	88
45 ,,	—	—	—	—	2	6	6	22	30	22	24	10	10	—	—	—	—	—	—	—	—	—	—	—	—	132
46 ,,	—	—	—	—	—	4	3	10	25	26	30	30	12	3	4	—	1	—	—	—	—	—	—	—	—	148
47 ,,	—	—	—	—	—	—	1	3	23	26	40	38	26	16	13	5	2	—	—	—	—	—	—	—	—	193
48 ,,	—	—	—	—	—	—	1	3	11	9	20	30	45	32	21	12	6	2	1	—	—	—	—	—	—	193
49 ,,	—	—	—	—	—	—	—	—	1	5	14	13	34	25	27	24	11	2	3	1	—	1	—	—	—	161
50 ,,	—	—	—	—	—	—	—	—	1	—	6	6	13	25	21	13	10	10	4	2	1	—	—	—	—	112
51 ,,	—	—	—	—	—	—	—	—	1	—	—	—	6	3	11	11	10	8	4	2	1	—	—	—	1	58
52 ,,	—	—	—	—	—	—	—	1	—	2	—	1	1	3	1	5	7	5	6	1	—	—	—	—	2	35
53 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	3	4	8	5	1	—	—	1	25
54 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	2	1	—	—	2	—	—	7
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	—	—	—	—	3
56 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	2
57 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	1	—	3
58 ,,	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
Totals	1	—	5	6	22	36	45	74	131	107	147	135	153	112	100	74	48	30	25	15	9	2	2	1	4	1284

TABLE 42. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 10·5—11·5 years																												Totals
	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	89·5—	91·5—	—	105·5—	
35⅞—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
36 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
37 ,,	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
39 ,,	—	1	2	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
40 ,,	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
41 ,,	1	1	1	2	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
42 ,,	2	3	1	2	1	—	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12
43 ,,	—	1	5	5	1	2	3	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	18
44 ,,	—	2	4	2	7	5	5	1	2	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	30
45 ,,	1	—	5	6	12	8	10	8	3	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	55
46 ,,	—	1	1	4	15	17	13	14	6	7	5	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	87
47 ,,	—	—	—	2	7	16	18	18	26	12	7	9	4	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	121
48 ,,	—	—	1	—	4	13	19	26	56	29	15	7	9	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	180
49 ,,	—	—	2	—	2	6	14	16	24	34	23	18	16	8	—	3	3	—	—	—	—	—	—	—	—	1	—	—	—	170
50 ,,	—	—	—	—	—	—	4	8	19	28	20	34	15	6	6	10	2	—	1	—	—	—	—	—	1	—	—	—	—	153
51 ,,	—	—	—	—	—	—	1	4	5	11	25	22	27	18	8	8	6	2	3	—	2	1	—	—	—	—	—	—	—	144
52 ,,	—	—	—	1	1	—	1	1	—	3	4	8	12	16	6	15	6	4	6	1	—	1	—	—	—	—	—	—	—	86
53 ,,	—	—	—	—	—	—	—	—	—	1	2	4	10	9	6	7	9	4	3	1	1	—	1	1	—	—	—	—	—	59
54 ,,	—	—	—	—	—	—	—	—	—	—	3	2	3	5	2	5	5	5	4	2	—	—	—	—	—	—	—	—	—	36
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	2	1	2	4	1	—	—	1	—	—	—	—	—	—	12
56 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	—	—	—	—	1	—	—	—	—	1	6
57 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
Totals	4	10	22	24	52	69	92	96	143	127	106	108	98	63	29	51	34	18	22	5	3	2	2	3	1	1	1	—	1	1187

TABLE 43. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 11·5—12·5 years																										Totals
	39·5—	42·5—	45·5—	48·5—	51·5—	54·5—	57·5—	60·5—	63·5—	66·5—	69·5—	72·5—	75·5—	78·5—	81·5—	84·5—	87·5—	90·5—	93·5—	96·5—	99·5—	101·5—	105·5—	108·5—	111·5—	114·5—	
34⅞—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
35 ,,	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
38 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
39 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
40 ,,	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
41 ,,	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
42 ,,	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
43 ,,	—	1	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
44 ,,	—	—	3	2	6	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	13
45 ,,	—	—	3	6	5	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	20
46 ,,	—	1	3	6	16	2	1	2	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	32
47 ,,	—	1	3	12	10	18	16	3	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	65
48 ,,	—	—	—	6	10	23	19	11	11	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	83
49 ,,	—	—	—	3	12	29	29	26	23	8	4	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	139
50 ,,	—	—	2	3	6	20	29	43	35	17	3	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	162
51 ,,	—	—	—	—	3	13	24	37	40	18	15	2	6	1	1	—	—	—	—	—	—	1	—	—	—	—	161
52 ,,	—	—	1	1	—	8	5	21	37	25	26	20	7	1	1	—	—	—	—	—	—	—	—	—	—	—	153
53 ,,	—	—	—	—	—	1	4	12	21	30	32	12	12	7	1	1	—	—	—	—	—	—	—	—	—	—	133
54 ,,	—	—	1	—	—	1	—	3	14	14	15	13	14	9	5	—	1	1	1	—	—	—	—	—	—	—	92
55 ,,	—	—	—	—	—	—	—	—	2	7	13	14	10	5	5	5	1	3	—	—	—	—	—	—	—	—	65
56 ,,	—	—	—	—	—	—	—	—	1	1	4	6	10	3	2	2	5	2	2	1	2	—	—	—	—	—	41
57 ,,	—	—	—	—	—	—	—	—	—	3	1	1	5	3	2	2	3	3	—	3	1	—	—	—	—	—	27
58 ,,	—	—	—	—	—	—	—	1	—	—	1	—	—	2	1	1	—	1	—	1	—	—	—	1	—	—	9
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
60 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
61 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	2	—	—	—	—	—	—	—	—	3
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
67 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	6	20	41	70	121	129	159	186	125	115	77	64	32	19	13	10	12	3	5	3	1	—	1	—	1	1214

TABLE 44. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 12·5—13·5 years																						Totals
	41·5—	45·5—	49·5—	53·5—	57·5—	61·5—	65·5—	69·5—	73·5—	77·5—	81·5—	85·5—	89·5—	93·5—	97·5—	101·5—	105·5—	109·5—	113·5—	117·5—	—	125·5—	
41⅞—	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
42 ,,	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
43 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44 ,,	1	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
45 ,,	—	1	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
46 ,,	—	—	1	5	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
47 ,,	2	3	6	9	4	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	30
48 ,,	—	—	4	6	10	7	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	30
49 ,,	—	2	3	12	10	26	6	2	—	—	1	—	1	—	—	—	—	—	—	—	—	—	63
50 ,,	—	—	2	5	21	26	12	7	1	—	—	—	—	—	—	1	—	—	—	—	—	—	75
51 ,,	—	—	—	3	12	26	24	15	6	3	—	1	1	—	—	—	—	—	—	—	—	—	91
52 ,,	—	—	—	2	12	24	20	32	11	8	4	1	—	—	—	—	—	—	—	—	—	—	114
53 ,,	—	—	—	—	4	11	26	29	22	13	2	3	—	—	—	—	—	—	—	—	—	—	110
54 ,,	—	—	—	—	4	6	18	27	29	14	8	4	1	—	—	—	—	—	—	—	—	—	111
55 ,,	—	—	—	—	—	3	12	14	18	29	11	9	4	3	2	—	—	—	—	—	—	—	105
56 ,,	—	—	—	1	—	—	3	8	18	18	13	13	10	1	1	—	—	—	—	—	—	—	86
57 ,,	—	—	—	—	—	1	—	4	9	15	7	12	4	1	2	—	—	—	—	—	—	—	55
58 ,,	—	—	—	—	—	—	—	1	1	5	3	6	8	3	1	1	2	2	—	1	—	—	34
59 ,,	—	—	—	—	—	—	—	—	2	2	1	2	6	—	1	1	—	2	—	—	—	—	17
60 ,,	—	—	—	—	—	—	—	—	—	1	2	2	4	2	2	—	—	—	1	—	—	—	14
61 ,,	—	—	—	—	—	—	—	—	1	—	—	1	1	2	—	—	2	1	—	—	—	—	8
62 ,,	—	—	—	—	—	—	—	—	—	—	—	—	3	—	—	—	1	—	—	—	—	—	4
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	2
64 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
65 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
66 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
67 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	5	9	20	45	79	136	126	139	120	108	53	54	44	12	9	3	6	5	1	1	—	1	976

TABLE 45. *Glasgow. Height and Weight of Girls of Group A.*

Height	Weight of Girls of 13·5—14·5 years																			Totals
	41·5—	—	49·5—	53·5—	57·5—	61·5—	65·5—	69·5—	73·5—	77·5—	81·5—	85·5—	89·5—	93·5—	97·5—	101·5—	105·5—	109·5—	113·5—	
46⅞—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
47 „	1	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	3
48 „	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
49 „	—	—	—	2	3	3	2	—	—	—	—	—	—	—	—	—	—	—	—	10
50 „	—	—	2	2	3	3	—	3	—	—	—	—	—	—	—	—	—	—	—	13
51 „	—	—	—	2	6	3	2	4	1	—	—	—	—	—	—	—	—	—	—	18
52 „	—	—	—	2	—	5	9	10	4	1	2	—	—	—	—	—	—	—	—	33
53 „	—	—	—	2	2	3	2	6	4	4	3	—	1	—	—	—	—	—	—	27
54 „	—	—	—	—	1	5	1	10	8	5	5	1	—	1	—	—	—	—	—	37
55 „	—	—	—	—	—	—	1	9	11	7	5	3	1	—	—	—	—	—	—	37
56 „	—	—	—	—	—	2	1	7	2	3	4	6	6	—	—	—	1	—	—	32
57 „	—	—	—	—	—	—	—	—	3	5	4	3	7	2	—	1	—	—	—	25
58 „	—	—	—	—	—	—	—	1	—	1	2	3	7	4	1	1	—	—	—	20
59 „	—	—	—	—	—	1	—	—	—	1	4	3	4	3	—	1	1	2	2	22
60 „	—	—	—	—	—	—	—	—	1	1	1	—	—	—	—	—	—	—	—	3
61 „	—	—	—	—	—	—	—	—	—	—	—	1	2	1	1	—	—	—	—	5
62 „	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	2
Totals	1	—	3	13	15	25	20	50	34	28	30	22	28	11	2	3	2	2	2	291

TABLE 46. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 5·5—6·5 years																	Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	
21⅞—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
30 „	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
31 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
32 „	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
33 „	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
34 „	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
35 „	—	2	4	3	5	—	—	—	—	—	—	—	—	—	—	—	—	14
36 „	1	2	6	10	9	—	—	—	—	—	—	—	—	—	—	—	—	28
37 „	—	3	9	14	15	6	1	1	—	—	—	—	—	—	—	—	—	49
38 „	—	1	9	26	24	10	5	3	—	—	—	—	—	—	—	—	—	78
39 „	1	2	3	18	38	38	33	14	4	—	1	—	—	—	—	—	—	152
40 „	—	1	3	14	31	53	61	30	13	2	3	—	—	—	—	—	—	211
41 „	—	—	—	6	18	48	39	45	22	6	3	1	—	—	—	—	—	188
42 „	—	—	—	1	4	9	31	53	34	14	5	—	—	—	—	—	—	151
43 „	—	—	1	—	2	6	12	37	36	22	21	4	—	1	—	—	—	142
44 „	—	—	—	—	1	—	4	14	17	13	12	8	—	—	1	—	—	70
45 „	—	—	—	—	1	1	4	2	5	2	7	2	—	—	—	—	—	24
46 „	—	—	—	—	—	—	2	1	—	1	2	2	2	—	—	—	—	10
47 „	—	—	—	—	—	—	1	1	2	1	—	1	1	1	—	—	—	8
48 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	3
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
54 „	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1
55 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
Totals	4	17	37	93	148	171	193	201	133	61	55	18	3	2	3	1	1	1141

TABLE 47. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 6·5—7·5 years																			Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	
32 7/8—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
33 „	—	2	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
34 „	—	4	2	1	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	9
35 „	—	3	—	7	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	11
36 „	—	1	—	5	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
37 „	—	—	3	5	3	4	2	1	—	—	—	—	—	—	—	—	—	—	—	18
38 „	—	1	3	7	10	10	4	3	1	—	—	—	—	—	—	—	—	—	—	39
39 „	—	2	—	8	17	25	15	7	3	—	—	—	—	—	—	—	—	—	—	77
40 „	—	—	1	3	20	36	28	20	8	2	1	1	1	—	—	—	—	—	—	121
41 „	—	1	—	4	10	28	41	39	21	12	5	3	2	—	—	—	—	—	—	166
42 „	—	—	—	—	5	15	37	60	47	21	12	2	2	—	—	—	—	—	—	201
43 „	—	—	—	—	1	2	22	52	57	45	23	7	3	—	1	—	—	—	—	213
44 „	—	—	—	—	—	3	6	30	56	46	36	20	8	3	2	—	—	—	—	210
45 „	—	—	—	—	—	—	1	9	14	30	31	23	15	4	2	—	—	—	—	129
46 „	—	—	—	—	—	—	1	2	5	9	18	18	13	3	5	2	—	—	—	76
47 „	—	—	—	—	—	—	—	—	3	1	6	9	7	7	1	2	1	1	—	38
48 „	—	—	—	—	—	—	—	—	—	1	1	—	3	2	4	1	—	1	—	13
49 „	—	—	—	—	—	—	—	—	—	2	—	—	—	1	1	—	—	—	1	5
50 „	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	—	3
Totals	1	14	9	43	71	124	157	224	215	170	133	83	54	21	17	5	1	3	1	1346

TABLE 48. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 7·5—8·5 years																					Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	—	69·5—	
33 7/8—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
34 „	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
35 „	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
36 „	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
37 „	—	1	1	3	1	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	8
38 „	—	—	—	3	2	3	1	2	1	1	—	—	—	—	—	—	—	—	—	—	—	13
39 „	—	—	—	2	3	11	6	7	4	1	—	—	—	—	—	—	—	—	—	—	—	34
40 „	1	—	—	1	5	8	14	13	2	—	1	—	—	—	—	—	—	—	—	—	—	45
41 „	—	—	—	1	5	13	19	23	12	7	3	2	2	—	—	—	—	—	—	—	—	87
42 „	—	—	1	2	3	6	22	35	30	16	9	2	1	—	—	—	—	—	—	—	—	127
43 „	—	—	—	—	3	6	8	39	31	36	30	15	5	2	2	—	—	—	—	—	—	177
44 „	—	—	—	—	1	2	3	24	51	31	49	28	11	6	—	1	1	—	—	—	—	208
45 „	—	—	—	1	—	1	3	6	22	48	50	49	24	16	7	1	—	1	—	—	—	229
46 „	—	—	—	—	—	—	1	6	7	23	38	26	26	27	9	2	1	—	—	—	—	166
47 „	—	—	—	—	—	—	1	—	1	6	12	25	31	24	16	10	2	4	—	—	—	132
48 „	—	—	—	—	—	—	—	—	—	—	5	9	12	14	11	9	4	3	2	—	—	69
49 „	—	—	—	—	—	—	—	—	—	—	1	2	1	6	12	4	4	3	—	—	—	33
50 „	—	—	—	—	—	—	—	—	—	—	—	1	1	1	4	3	5	—	—	—	1	16
51 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	2	1	—	—	—	4
52 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
53 „	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	1	4	15	25	51	78	155	162	170	198	159	115	96	61	30	19	12	3	—	1	1356

TABLE 49. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 8·5—9·5 years																									Totals
	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	
29⅞—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
33 ,,	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
34 ,,	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
35 ,,	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
36 ,,	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
37 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
38 ,,	—	—	—	2	2	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
39 ,,	—	—	—	—	2	3	4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
40 ,,	—	—	—	—	5	6	7	1	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	22
41 ,,	—	—	—	1	1	8	14	5	5	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	37
42 ,,	—	—	—	—	3	7	14	19	16	8	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	68
43 ,,	—	—	—	—	—	3	10	14	15	11	8	1	—	1	1	—	—	—	—	—	—	—	—	—	—	64
44 ,,	1	—	—	—	1	9	11	22	28	37	21	10	8	4	—	—	1	—	—	—	—	—	—	—	—	153
45 ,,	—	—	—	—	1	2	7	10	25	44	33	25	12	10	5	—	1	1	—	—	1	—	—	—	—	177
46 ,,	—	—	—	—	—	1	2	10	19	34	50	36	23	22	3	6	1	—	—	—	—	—	—	—	—	207
47 ,,	—	—	—	—	—	1	1	7	7	15	27	36	53	23	14	6	6	3	—	—	—	—	—	—	—	199
48 ,,	—	—	—	—	—	—	2	2	5	11	11	19	36	24	33	14	9	7	1	2	—	—	—	—	—	176
49 ,,	—	—	—	—	—	1	—	—	—	2	8	11	14	27	24	8	10	4	3	1	1	1	—	—	—	115
50 ,,	—	—	—	—	—	—	—	—	—	—	—	2	8	14	10	17	8	2	1	1	—	—	—	—	—	63
51 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	2	7	7	6	4	2	2	2	1	—	—	1	35
52 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	1	2	1	2	—	1	1	1	—	12
53 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	2
54 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1	—	—	1	—	3
55 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
Totals	2	1	1	3	15	42	74	92	123	163	161	142	156	132	98	58	44	23	8	8	5	4	1	2	2	1360

TABLE 50. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 9·5—10·5 years																									Totals
	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	
36⅞—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
37 ,,	1	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
38 ,,	1	—	1	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	3
39 ,,	—	—	—	—	2	1	—	—	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	5
40 ,,	—	—	—	—	2	1	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	4
41 ,,	—	1	—	1	1	1	3	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
42 ,,	—	—	1	2	7	5	4	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	21
43 ,,	—	—	2	2	4	5	9	8	7	2	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	41
44 ,,	—	—	—	4	3	13	15	12	9	3	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	62
45 ,,	—	—	—	1	2	3	17	30	23	22	8	7	2	1	1	—	—	—	—	—	—	—	—	—	—	117
46 ,,	—	—	—	—	2	2	17	22	28	22	20	13	8	4	2	—	2	—	—	—	—	—	—	—	—	142
47 ,,	—	1	—	1	1	3	3	13	39	41	39	32	9	6	5	1	—	—	—	—	—	—	—	—	—	194
48 ,,	—	—	—	—	—	1	6	9	11	28	30	46	31	14	4	7	1	1	—	1	—	—	—	—	—	190
49 ,,	—	—	—	—	—	—	1	3	3	11	15	44	41	29	17	6	8	6	1	—	—	—	—	—	—	185
50 ,,	—	—	—	—	—	—	—	1	3	4	9	18	26	30	21	10	6	4	2	1	—	1	—	—	—	136
51 ,,	—	—	—	—	—	—	1	2	1	1	3	9	15	17	19	14	7	11	5	2	—	2	—	—	1	110
52 ,,	—	—	—	—	—	—	—	—	—	—	—	3	2	8	9	9	9	7	8	3	—	1	1	—	—	60
53 ,,	—	—	—	—	—	—	—	—	—	—	—	—	2	1	2	7	1	4	1	3	—	1	—	1	1	24
54 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	1	1	2	1	—	1	—	—	—	—	8
55 ,,	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	2	—	—	2	—	1	—	1	8
56 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1	1	2	1	—	—	6
57 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	1	—	—	—	—	—	—	4
Totals	2	3	4	11	24	35	76	103	127	136	126	176	138	116	81	55	35	37	19	11	4	7	3	1	3	1333

TABLE 51. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 10·5—11·5 years																												Totals
	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	89·5—	91·5—	93·5—	
35⅞—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
36 ,,	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
37 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
38 ,,	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
39 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 ,,	—	—	1	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
41 ,,	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
42 ,,	—	—	—	1	1	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
43 ,,	—	—	1	—	1	4	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
44 ,,	—	1	—	—	4	6	4	5	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	22
45 ,,	—	—	1	2	5	5	6	7	7	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	37
46 ,,	—	—	2	—	4	9	9	10	12	11	5	4	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	69
47 ,,	—	—	—	1	2	13	12	16	19	13	13	8	6	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	107
48 ,,	—	—	—	—	—	7	14	14	18	40	29	15	10	8	—	1	1	—	—	—	—	1	—	—	—	—	—	—	—	158
49 ,,	—	—	1	—	1	1	5	11	18	29	33	21	24	10	8	6	1	1	—	—	1	—	—	—	—	—	—	—	—	171
50 ,,	—	—	—	—	1	2	—	5	11	18	35	32	39	17	3	4	5	2	—	1	1	—	—	—	—	—	—	—	—	176
51 ,,	—	—	—	—	—	—	—	1	4	16	17	20	24	15	21	7	8	4	1	2	2	1	1	—	—	—	—	—	—	144
52 ,,	—	—	—	—	—	—	—	—	—	2	8	12	25	25	9	16	12	4	3	4	—	1	1	—	—	—	—	—	—	122
53 ,,	—	—	—	—	—	—	—	1	—	1	3	3	5	13	10	13	11	11	5	2	7	2	—	—	—	1	1	—	—	89
54 ,,	—	—	—	—	—	—	—	—	—	—	1	1	1	1	8	9	8	7	7	3	3	2	1	2	1	—	—	—	—	55
55 ,,	—	—	—	—	—	—	—	—	—	1	—	—	1	1	—	5	5	6	3	3	5	—	1	—	1	—	1	—	—	33
56 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	2	1	2	—	1	1	2	2	1	2	1	—	—	1	—	—	17
57 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	1	—	—	1	1	2	—	—	—	—	—	1	—	8
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	1	—	—	—	—	1	—	—	—	—	—	4
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
62 ,,	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	2	4	7	5	21	51	52	73	94	133	146	118	138	96	61	65	53	36	21	18	22	10	6	4	2	1	3	1	1	1244

TABLE 52. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 11·5—12·5 years																								Totals
	39·5—	42·5—	45·5—	48·5—	51·5—	54·5—	57·5—	60·5—	63·5—	66·5—	69·5—	72·5—	75·5—	78·5—	81·5—	84·5—	87·5—	90·5—	93·5—	96·5—	99·5—	102·5—	105·5—	108·5—	111·5—	
39⅞—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
41 ,,	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
42 ,,	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
43 ,,	—	1	4	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
44 ,,	—	1	3	1	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8
45 ,,	—	—	3	6	4	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	16
46 ,,	—	1	—	3	8	7	4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	24
47 ,,	1	1	1	8	11	12	6	4	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	47
48 ,,	—	—	2	4	17	17	18	9	5	1	—	3	1	—	—	—	—	—	—	—	—	—	—	—	—	77
49 ,,	1	—	—	1	5	31	20	20	19	5	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	106
50 ,,	—	—	—	—	4	13	39	35	31	20	5	1	2	1	—	—	—	—	—	—	—	—	—	—	—	151
51 ,,	—	—	—	1	1	14	30	37	33	22	24	4	2	2	—	—	—	—	—	—	—	—	—	—	—	170
52 ,,	—	—	—	—	1	5	13	30	43	28	29	12	12	5	2	2	—	—	—	—	—	—	—	—	—	182
53 ,,	—	—	—	—	—	1	3	14	24	21	26	20	12	4	6	2	—	—	—	—	—	—	—	—	—	133
54 ,,	—	—	—	—	—	—	4	7	8	19	15	22	16	13	7	—	—	—	—	—	—	—	—	—	—	111
55 ,,	—	—	—	1	1	—	2	2	4	8	20	12	15	12	4	2	3	1	—	—	—	1	—	—	—	88
56 ,,	—	—	—	—	—	—	—	2	2	1	5	8	4	13	7	5	2	2	2	—	1	1	—	—	—	55
57 ,,	—	—	—	—	—	—	—	—	—	1	2	4	5	5	6	1	—	6	2	—	—	—	—	—	—	32
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	3	3	—	—	2	3	1	—	—	—	—	—	—	12
59 ,,	—	—	—	—	—	—	—	1	—	—	—	—	2	1	2	1	4	—	4	—	1	1	1	—	1	19
60 ,,	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	—	1	—	1	1	—	—	—	—	—	5
61 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	3
62 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
65 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
69 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
Totals	3	4	14	29	54	103	140	162	171	129	131	86	74	59	36	13	12	12	11	1	3	3	2	—	1	1253

TABLE 53. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 12·5—13·5 years																						Totals
	41·5—	45·5—	49·5—	53·5—	57·5—	61·5—	65·5—	69·5—	73·5—	77·5—	81·5—	85·5—	89·5—	93·5—	97·5—	101·5—	105·5—	109·5—	113·5—	117·5—	121·5—	125·5—	
42⅞—	—	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
43 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44 ,,	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
45 ,,	—	1	3	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
46 ,,	—	1	3	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
47 ,,	—	2	4	5	3	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	16
48 ,,	—	—	3	5	11	4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	24
49 ,,	—	—	2	7	13	13	3	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	41
50 ,,	—	—	—	4	19	16	9	6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	54
51 ,,	—	—	1	5	21	31	20	17	7	—	—	—	—	—	—	—	—	—	—	—	—	—	102
52 ,,	—	—	—	2	9	32	31	29	11	14	1	1	—	—	—	—	—	—	—	—	—	—	130
53 ,,	—	—	—	2	5	23	33	31	21	12	6	3	1	—	—	—	—	—	—	—	—	—	137
54 ,,	—	—	1	—	1	14	17	40	35	22	12	8	4	1	—	—	—	—	—	—	—	—	155
55 ,,	—	—	—	1	2	5	14	30	43	24	16	10	4	1	—	—	1	—	—	—	—	—	151
56 ,,	—	—	—	—	—	1	3	10	15	22	27	8	8	4	—	1	1	—	—	—	—	—	100
57 ,,	—	—	—	1	—	2	1	9	14	13	17	11	17	6	5	1	—	1	—	—	—	—	98
58 ,,	—	—	—	—	—	—	—	3	2	10	7	13	14	2	4	4	—	—	—	1	—	1	61
59 ,,	—	—	—	—	—	—	1	1	5	6	4	6	10	4	3	3	1	—	—	—	—	—	44
60 ,,	—	—	—	—	—	—	—	—	—	1	1	2	3	4	4	2	3	2	—	—	—	—	22
61 ,,	—	—	—	—	1	—	—	—	—	—	1	1	—	1	—	1	—	—	2	1	—	—	8
62 ,,	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	1	—	—	—	—	—	3
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
66 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
67 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
68 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
69 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	1	4	18	34	87	144	133	180	154	125	93	63	63	23	16	12	7	3	3	2	—	1	1166

TABLE 54. *Glasgow. Height and Weight of Girls of Group B.*

Height	Weight of Girls of 13·5—14·5 years																			Totals
	45·5—	49·5—	53·5—	57·5—	61·5—	65·5—	69·5—	73·5—	77·5—	81·5—	85·5—	89·5—	93·5—	97·5—	101·5—	105·5—	109·5—	—	121·5—	
45⅞—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
46 ,,	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
47 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
48 ,,	—	—	—	—	1	—	—	—	—	—	—	1	—	—	—	—	—	—	—	2
49 ,,	—	—	1	—	2	—	—	2	—	—	—	—	—	—	—	—	—	—	—	5
50 ,,	—	—	1	1	4	—	2	—	—	—	—	—	—	—	—	—	—	—	—	8
51 ,,	—	—	2	2	6	1	2	2	1	—	—	—	—	—	—	—	—	—	—	16
52 ,,	—	—	—	2	2	6	6	5	1	—	—	—	—	—	—	—	—	—	—	22
53 ,,	—	—	1	1	3	8	11	8	2	4	—	1	—	—	—	—	—	—	—	39
54 ,,	—	—	1	—	1	5	11	9	6	2	5	1	2	1	—	—	—	—	—	44
55 ,,	—	—	—	1	—	4	4	11	9	5	4	1	1	—	—	—	—	—	—	40
56 ,,	—	—	1	—	—	1	2	13	11	11	9	5	—	2	—	—	1	—	—	56
57 ,,	—	1	—	1	1	—	4	6	6	12	9	5	2	7	1	—	1	—	—	56
58 ,,	—	—	—	1	—	—	1	2	5	9	9	7	5	5	1	—	2	—	—	47
59 ,,	—	—	—	—	—	—	1	—	1	1	7	5	4	1	2	1	1	—	1	25
60 ,,	—	—	—	—	—	—	—	—	1	1	2	3	4	1	2	1	1	—	1	17
61 ,,	—	—	—	—	1	—	—	—	1	—	2	1	3	3	2	2	—	—	—	15
62 ,,	—	—	—	—	—	—	—	1	—	—	—	1	—	1	1	1	2	—	—	7
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
Totals	1	1	7	11	21	25	44	59	44	45	47	31	21	21	10	5	8	—	2	403

TABLE 55. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 5·5—6·5 years																Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	—	63·5—	
$31\frac{7}{8}$—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
32 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
33 ,,	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
34 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
35 ,,	1	—	—	3	1	—	—	—	—	—	—	—	—	—	—	—	5
36 ,,	—	—	2	2	1	1	—	1	—	—	—	—	—	—	—	—	7
37 ,,	—	1	4	6	6	4	1	—	—	—	—	—	—	—	—	—	22
38 ,,	—	—	5	4	8	14	7	4	—	—	—	1	—	—	—	—	43
39 ,,	—	—	2	7	12	18	5	8	2	1	1	—	—	—	—	—	56
40 ,,	—	—	1	1	7	18	19	13	7	3	—	—	—	—	—	—	69
41 ,,	—	—	—	—	7	13	21	26	10	2	—	1	—	—	—	—	80
42 ,,	—	—	—	—	1	1	11	15	15	6	5	2	—	—	—	—	56
43 ,,	—	—	—	—	—	1	6	11	10	9	9	1	2	—	—	—	49
44 ,,	—	—	—	—	1	1	1	—	9	5	7	4	2	—	—	—	30
45 ,,	—	—	—	—	—	—	—	1	—	1	3	1	2	2	—	—	10
46 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
47 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	2	1	14	24	44	71	71	79	53	27	26	10	6	2	—	1	431

TABLE 56. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 6·5—7·5 years																Totals
	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	
$33\frac{7}{8}$—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
34 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
35 ,,	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	2
36 ,,	—	—	—	1	—	—	1	1	—	—	—	—	—	—	—	—	3
37 ,,	—	—	3	—	3	1	—	—	—	—	—	—	—	—	—	—	7
38 ,,	1	2	1	1	6	4	3	1	—	—	—	—	—	—	—	—	19
39 ,,	—	2	1	7	12	7	3	—	1	1	—	—	—	—	—	—	34
40 ,,	—	—	2	4	9	12	11	6	4	1	—	—	1	—	—	—	50
41 ,,	—	1	1	7	10	18	38	18	5	7	—	—	—	—	—	—	105
42 ,,	1	—	—	—	3	27	21	20	20	5	3	2	—	—	—	—	102
43 ,,	—	—	1	3	2	10	15	29	19	12	8	3	—	—	—	1	103
44 ,,	—	—	—	—	—	2	9	18	14	17	7	7	4	1	—	—	79
45 ,,	—	—	—	—	—	—	2	5	9	13	16	12	3	—	—	1	61
46 ,,	—	—	—	—	—	—	—	1	1	2	5	7	5	4	2	—	27
47 ,,	—	—	—	—	—	—	—	—	—	1	3	4	3	2	—	1	14
48 ,,	—	—	—	—	—	—	—	1	—	—	—	2	1	1	3	1	9
49 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
50 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
Totals	2	7	10	23	45	81	103	100	73	59	42	37	17	8	6	4	617

TABLE 57. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 7·5—8·5 years																			Totals
	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—		77·5—	
$36\frac{7}{8}$—	—	1	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	3
37 "	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
38 "	1	—	—	1	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	4
39 "	—	1	3	1	2	2	—	1	2	—	—	—	—	—	—	—	—	—	—	12
40 "	—	2	—	7	2	4	3	2	1	—	—	—	—	—	—	—	—	—	—	21
41 "	—	—	1	8	8	8	2	2	6	2	—	2	—	—	—	—	—	—	—	39
42 "	—	—	—	5	15	13	13	8	8	3	1	—	—	—	—	—	—	—	—	66
43 "	—	—	1	3	8	12	31	16	19	3	3	1	—	—	—	—	—	—	—	97
44 "	—	—	—	—	—	19	26	30	28	13	7	2	1	1	—	—	—	—	—	127
45 "	—	—	—	—	2	4	13	17	37	18	16	8	2	1	—	—	—	—	—	118
46 "	—	—	—	—	—	—	1	9	17	17	14	13	5	2	2	—	—	—	—	80
47 "	—	—	—	—	—	2	—	1	7	8	15	17	9	4	1	1	1	—	—	66
48 "	—	—	—	—	—	—	—	—	4	1	4	6	10	3	3	3	2	—	—	36
49 "	—	—	—	—	—	—	—	1	—	—	1	1	3	3	2	1	1	—	—	13
50 "	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	1	—	—	—	3
51 "	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
52 "	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	—	—	—	—	2
53 "	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
54 "	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	6	5	25	40	65	90	87	129	66	62	50	31	14	9	6	4	—	1	691

TABLE 58. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 8·5—9·5 years																							Totals
	29·5—		35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	
$32\frac{7}{8}$—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
38 "	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
39 "	—	—	—	—	3	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
40 "	—	—	—	1	1	1	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6
41 "	1	—	—	—	2	3	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
42 "	—	—	—	—	1	2	4	3	1	1	2	—	—	—	—	—	—	—	—	—	—	—	—	14
43 "	—	—	—	—	1	8	8	9	4	6	—	—	1	—	—	—	—	—	—	—	—	—	—	37
44 "	—	—	—	1	—	4	6	9	16	9	6	3	1	1	—	—	—	—	—	—	—	—	—	56
45 "	—	—	—	—	2	2	9	11	20	20	17	9	4	1	—	—	—	—	—	—	—	—	—	95
46 "	—	—	—	—	—	1	6	6	21	31	20	12	10	3	—	—	—	1	—	—	—	—	—	111
47 "	—	—	—	—	—	—	—	5	7	21	17	20	14	9	2	6	—	1	—	—	—	—	—	102
48 "	—	—	—	—	—	—	—	—	5	6	14	10	17	16	10	6	—	4	—	1	—	—	—	89
49 "	—	—	—	—	—	—	1	—	2	2	3	10	12	9	6	7	2	3	1	—	1	—	—	59
50 "	—	—	—	—	—	—	—	—	1	1	1	—	2	5	6	5	3	2	—	1	—	—	—	27
51 "	—	—	—	—	—	—	—	—	—	—	—	1	—	1	2	2	5	2	1	—	1	—	1	16
52 "	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	2	1	—	—	—	—	4
53 "	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1	—	2
54 "	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
55 "	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1	2
Totals	1	—	2	2	11	23	38	46	78	97	80	66	61	46	26	27	10	15	3	2	2	1	2	639

TABLE 59. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 9·5—10·5 years																								Totals
	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	
40⅞—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
41 ,,	—	—	1	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
42 ,,	1	—	1	3	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
43 ,,	—	—	2	2	3	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
44 ,,	—	—	2	6	5	12	5	3	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	36
45 ,,	1	—	1	1	8	6	8	9	6	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	41
46 ,,	—	—	1	2	3	7	10	10	12	2	2	—	1	1	—	1	—	—	—	—	—	—	—	—	52
47 ,,	—	1	—	—	1	5	13	23	11	17	9	8	3	2	—	—	—	—	—	—	—	—	—	—	93
48 ,,	—	1	—	—	—	3	4	10	14	12	25	8	7	—	2	2	—	—	—	—	—	—	—	—	88
49 ,,	—	—	—	—	—	—	3	5	11	15	13	16	14	4	7	1	—	—	—	—	—	—	—	—	89
50 ,,	—	—	—	—	—	—	—	1	6	6	14	15	15	4	1	3	1	—	—	—	1	—	1	—	68
51 ,,	—	—	—	—	—	—	—	—	1	2	4	6	16	6	4	3	2	2	1	—	—	—	—	—	47
52 ,,	—	—	—	—	—	—	—	—	1	—	3	3	1	6	3	4	—	5	—	1	—	—	—	—	27
53 ,,	—	—	—	—	—	—	—	—	—	1	—	—	1	2	1	1	2	1	—	1	—	—	—	—	10
54 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	—	—	—	—	3
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	2	—	1	—	—	4
56 ,,	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	1	3
57 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
Totals	2	2	10	14	21	36	45	62	64	58	70	57	60	25	18	15	6	8	1	7	1	1	1	1	585

TABLE 60. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 10·5—11·5 years																												Totals
	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	89·5—	91·5—	93·5—	
35⅞—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
39 ,,	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
40 ,,	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
41 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
42 ,,	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
43 ,,	—	—	1	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
44 ,,	—	—	—	4	2	2	1	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
45 ,,	1	—	—	1	2	4	1	—	1	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12
46 ,,	—	—	1	—	2	4	7	6	2	3	—	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	28
47 ,,	—	—	—	—	1	5	9	6	12	2	7	4	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	47
48 ,,	—	—	—	—	2	9	7	7	16	12	11	9	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	77
49 ,,	—	—	—	—	1	1	2	6	17	8	23	11	8	6	—	—	1	—	—	—	—	—	—	—	—	—	—	—	84
50 ,,	—	—	—	—	—	2	1	5	8	9	14	13	14	5	11	2	2	1	1	—	1	—	—	—	—	—	—	—	89
51 ,,	—	—	—	—	—	—	1	—	2	6	13	20	9	16	10	5	2	1	1	1	—	—	—	—	—	—	—	—	87
52 ,,	—	—	—	—	—	—	—	—	—	2	7	10	14	7	3	5	5	2	2	1	1	1	—	—	—	—	—	—	60
53 ,,	—	—	—	—	—	—	—	—	2	—	4	4	7	4	4	4	3	4	5	1	1	1	1	1	—	—	—	—	46
54 ,,	—	—	—	—	—	—	—	—	—	—	1	1	—	2	3	4	3	2	—	1	—	—	—	—	—	—	—	—	17
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	2	2	—	—	—	—	1	—	—	—	9
56 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1	1	—	2	1	—	—	—	1	—	—	7
57 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1	2
58 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	1	—	—	4
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
68 ,,	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	2	—	3	8	10	27	30	30	60	44	83	75	54	44	34	21	18	12	12	6	5	3	3	1	1	2	—	1	589

TABLE 61. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 11·5—12·5 years																							Totals
	45·5—	*48·5—*	*51·5—*	*54·5—*	*57·5—*	*60·5—*	*63·5—*	*66·5—*	*69·5—*	*72·5—*	*75·5—*	*78·5—*	*81·5—*	*84·5—*	*87·5—*	*90·5—*	*93·5—*	*96·5—*	*99·5—*	*102·5—*	*105·5—*	*108·5—*	*111·5—*	
42⅞—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
43 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44 ,,	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
45 ,,	2	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
46 ,,	—	1	3	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
47 ,,	—	3	3	6	4	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	18
48 ,,	—	2	4	8	8	5	4	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	32
49 ,,	1	—	2	13	15	10	5	3	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	52
50 ,,	—	1	1	5	18	17	15	7	5	1	1	—	—	—	—	—	—	—	—	—	—	—	—	71
51 ,,	—	—	—	3	11	20	16	11	6	3	—	1	—	—	—	—	—	—	—	—	—	—	—	71
52 ,,	—	—	—	1	7	21	14	20	16	5	3	5	2	1	—	1	—	—	—	—	—	—	—	96
53 ,,	—	—	—	—	2	7	15	16	10	10	9	2	5	2	1	—	—	—	—	—	—	—	—	79
54 ,,	—	—	—	—	—	2	2	5	10	11	14	6	6	4	—	2	—	1	1	—	—	—	—	64
55 ,,	—	—	—	—	—	—	1	3	5	10	9	8	5	1	2	—	—	—	—	—	—	—	—	44
56 ,,	—	—	—	—	—	—	—	—	6	6	4	3	4	1	2	2	—	—	1	—	—	—	—	29
57 ,,	—	—	—	—	—	—	—	1	—	—	1	3	1	—	2	—	—	—	—	—	—	—	1	9
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	3	2	1	3	—	—	—	—	—	—	9
59 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	1	—	—	—	—	—	3
60 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
61 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
62 ,,	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	2
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
64 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	3	9	14	39	66	83	74	68	62	46	41	28	23	12	10	7	3	3	2	—	1	—	1	595

TABLE 62. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 12·5—13·5 years																				Totals
	41·5—	*45·5—*	*49·5—*	*53·5—*	*57·5—*	*61·5—*	*65·5—*	*69·5—*	*73·5—*	*77·5—*	*81·5—*	*85·5—*	*89·5—*	*93·5—*	*97·5—*	*101·5—*	*105·5—*	*109·5—*	*113·5—*	*117·5—*	
38⅞—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
39 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
40 ,,	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
46 ,,	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
47 ,,	—	—	—	—	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	4
48 ,,	—	1	4	1	3	2	—	—	2	—	1	—	—	—	—	—	—	—	—	—	14
49 ,,	—	—	1	2	1	6	1	2	2	2	—	—	—	—	—	—	—	—	—	—	17
50 ,,	—	—	—	6	9	10	2	4	2	—	1	—	—	—	—	—	—	—	—	—	34
51 ,,	—	—	—	1	4	11	10	14	3	1	1	—	—	—	—	—	—	—	—	—	45
52 ,,	—	—	—	—	7	9	20	7	11	1	—	1	—	—	—	—	—	—	—	—	56
53 ,,	—	—	—	—	1	9	14	17	14	4	3	3	2	1	—	—	—	—	—	—	68
54 ,,	—	—	—	—	—	7	6	24	18	13	3	2	—	1	—	—	1	—	—	—	75
55 ,,	—	—	—	—	1	1	2	9	19	11	5	6	1	1	1	—	—	—	—	—	57
56 ,,	—	—	—	—	—	1	1	2	7	5	16	9	4	1	—	—	—	—	—	—	46
57 ,,	—	—	—	—	1	—	—	2	5	10	10	12	6	3	1	1	—	1	—	—	52
58 ,,	—	—	—	—	—	—	—	—	—	4	5	7	8	1	2	2	—	—	—	—	29
59 ,,	—	—	—	—	1	—	—	—	—	1	2	5	1	5	1	3	1	—	—	—	20
60 ,,	—	—	—	—	—	—	—	—	—	3	1	—	1	1	3	1	1	2	2	1	16
61 ,,	—	—	—	—	—	—	—	—	—	—	—	—	2	1	1	—	—	—	—	—	4
62 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1	—	—	2
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	2
Totals	1	2	5	10	29	58	57	81	84	55	48	46	25	15	9	7	4	5	2	1	544

TABLE 63. *Glasgow. Height and Weight of Girls of Group C.*

Height	Weight of Girls of 13·5—14·5 years																Totals
	53·5—	57·5—	61·5—	65·5—	69·5—	73·5—	77·5—	81·5—	85·5—	89·5—	93·5—	97·5—	101·5—	105·5—	109·5—	113·5—	
47⅞—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
48 ”	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
49 ”	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
50 ”	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	2
51 ”	—	1	1	2	2	—	—	—	—	—	—	—	—	—	—	—	6
52 ”	—	—	1	2	2	3	1	2	1	—	1	—	—	—	—	—	13
53 ”	—	1	2	3	1	—	1	—	—	—	—	—	—	—	—	—	8
54 ”	—	—	1	2	3	7	2	4	1	—	1	—	—	—	—	—	21
55 ”	—	—	1	—	4	6	3	6	2	—	—	1	—	—	—	—	23
56 ”	—	—	—	1	1	—	6	5	4	2	—	2	2	—	—	—	23
57 ”	—	—	—	—	—	3	—	5	2	3	1	1	1	—	1	—	17
58 ”	—	—	—	—	—	—	—	3	5	6	2	1	1	—	—	—	18
59 ”	—	—	—	—	—	1	—	3	2	2	2	2	2	—	—	3	17
60 ”	—	—	—	—	—	—	1	—	—	1	2	2	—	1	—	—	7
61 ”	—	—	—	—	—	—	—	—	—	1	—	2	1	1	—	—	5
62 ”	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
63 ”	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
Totals	3	5	6	11	13	20	14	28	17	15	10	11	7	2	1	3	166

TABLE 64. *Glasgow Height and Weight of Girls of Group D.*

Height	Weight of Girls of 5·5—6·5 years																Totals
	27·5—	29·5—	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	
34⅞—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
35 ”	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
36 ”	—	1	2	—	1	—	—	—	—	—	—	—	—	—	—	—	4
37 ”	—	1	1	3	4	4	—	1	—	—	—	—	—	—	—	—	14
38 ”	2	—	1	3	3	4	—	—	—	—	—	—	—	—	—	—	13
39 ”	1	—	2	3	9	12	5	3	—	1	—	—	—	—	—	—	36
40 ”	—	—	—	2	9	13	13	13	3	1	—	—	—	—	—	—	54
41 ”	—	—	—	—	8	12	21	26	9	4	1	—	—	—	—	—	81
42 ”	—	—	—	1	3	3	20	23	16	11	2	1	—	—	—	—	80
43 ”	—	—	—	—	1	1	6	6	19	11	4	1	—	—	—	—	49
44 ”	—	—	—	—	—	1	4	4	9	5	4	1	1	—	—	—	29
45 ”	—	—	—	—	—	—	—	—	3	4	6	3	—	2	—	—	18
46 ”	—	—	—	—	—	—	—	—	—	—	1	—	3	1	—	1	6
47 ”	—	—	—	—	—	—	—	—	—	1	—	—	2	—	—	—	3
Totals	3	3	7	12	38	50	69	76	59	38	18	6	6	3	—	1	389

TABLE 65. *Glasgow. Height and Weight of Girls of Group D.*

Height	Weight of Girls of 6·5—7·5 years															Totals
	31·5—	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	
37⅞—	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	3
38 ,,	1	—	—	2	—	—	—	—	—	—	—	—	—	—	—	3
39 ,,	—	—	2	4	2	3	2	—	—	—	—	—	—	—	—	13
40 ,,	—	1	—	4	4	6	1	—	—	—	—	—	—	—	—	16
41 ,,	—	2	5	7	8	7	9	3	1	—	—	—	—	—	—	42
42 ,,	—	—	5	4	20	20	7	13	3	1	—	1	—	—	—	74
43 ,,	—	—	—	1	7	23	20	21	15	5	1	—	—	—	—	93
44 ,,	—	—	—	1	3	11	15	18	19	6	1	2	—	—	—	76
45 ,,	—	—	—	—	—	5	8	14	19	11	5	2	2	—	—	66
46 ,,	—	—	—	—	1	1	2	6	7	7	7	5	2	1	1	40
47 ,,	—	—	—	—	—	—	1	1	4	5	5	6	1	2	1	26
48 ,,	—	—	—	—	—	—	—	1	—	2	—	2	1	—	—	6
49 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
50 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	2
Totals	2	4	12	23	46	76	65	77	68	37	19	19	6	3	3	460

TABLE 66. *Glasgow. Height and Weight of Girls of Group D*

Height	Weight of Girls of 7·5—8·5 years																			Totals
	33·5—	35·5—	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	—	71·5—	
36⅞—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
39 ,,	—	—	1	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
40 ,,	1	—	1	1	2	—	—	1	—	—	—	—	—	—	—	—	—	—	—	6
41 ,,	—	3	2	2	6	4	3	2	—	—	—	—	—	—	—	—	—	—	—	22
42 ,,	—	—	4	2	12	12	1	5	—	—	1	—	—	—	—	—	—	—	—	37
43 ,,	—	—	—	2	10	9	7	2	3	—	1	—	—	—	—	—	—	—	—	34
44 ,,	—	—	—	2	11	24	16	19	10	5	1	1	—	—	—	—	—	—	—	89
45 ,,	—	—	—	—	5	15	10	19	15	11	6	2	1	—	—	—	1	—	—	85
46 ,,	—	—	—	—	1	3	9	15	22	17	7	2	2	—	—	—	—	—	—	78
47 ,,	—	—	—	—	—	2	4	7	16	12	6	11	7	3	1	1	—	—	—	70
48 ,,	—	—	—	—	—	1	—	—	3	7	5	9	3	4	4	—	—	—	—	36
49 ,,	—	—	—	—	—	—	—	—	—	—	3	3	4	2	—	1	—	—	—	13
50 ,,	—	—	—	—	—	—	—	—	—	—	2	—	1	2	1	1	—	—	—	7
51 ,,	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1	—	—	—	—	2
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
55 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	2
Totals	1	3	9	11	48	70	50	70	69	53	32	28	18	11	7	3	1	—	2	486

TABLE 67. *Glasgow. Height and Weight of Girls of Group D.*

Height	Weight of Girls of 8·5—9·5 years																							Totals
	37·5—	39·5—	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	
40 7/8—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
41 „	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
42 „	1	2	2	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
43 „	—	1	4	6	5	2	3	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	22
44 „	—	—	5	6	4	9	4	6	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	36
45 „	—	—	1	5	5	17	12	13	3	2	—	—	—	—	—	—	1	—	—	—	—	—	—	59
46 „	—	—	—	1	3	13	18	19	6	6	3	1	—	—	—	—	—	—	—	—	—	—	—	70
47 „	—	—	1	1	2	9	16	16	11	10	7	—	2	1	1	—	—	—	—	—	—	—	—	77
48 „	—	—	—	—	1	4	8	12	13	17	8	3	5	2	—	—	—	—	—	—	—	—	—	73
49 „	—	—	—	—	—	—	1	10	7	15	12	12	2	4	4	1	—	1	—	—	—	—	—	69
50 „	—	—	—	—	—	—	—	3	7	6	10	5	4	6	4	—	1	1	—	1	1	—	—	49
51 „	—	—	—	—	—	—	—	—	2	2	2	1	2	4	1	1	—	—	—	—	—	—	—	15
52 „	—	—	—	—	—	—	—	—	—	—	1	—	3	1	4	1	1	—	1	—	—	—	—	12
53 „	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	2	—	—	—	—	—	—	1	6
54 „	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	1	—	—	—	—	—	4
55 „	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1
Totals	2	3	13	22	20	55	63	79	51	59	43	23	21	20	15	5	3	3	1	1	1	—	1	504

TABLE 68. *Glasgow. Height and Weight of Girls of Group D.*

Height	Weight of Girls of 9·5—10·5 years																								Totals
	41·5—	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	
41 7/8—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
42 „	2	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
43 „	1	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
44 „	—	1	1	3	2	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9
45 „	—	5	2	2	1	3	1	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	16
46 „	—	1	5	10	11	3	8	3	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	44
47 „	—	—	4	4	6	16	17	8	8	1	1	3	—	—	—	—	—	—	—	—	—	—	—	—	68
48 „	—	—	4	4	6	12	9	19	15	4	3	2	—	—	—	—	—	—	—	—	—	—	—	—	78
49 „	—	—	—	—	4	7	13	21	13	17	14	6	4	1	2	—	—	—	—	—	—	—	—	—	102
50 „	—	—	—	—	—	6	7	11	14	15	14	14	7	3	—	1	—	—	—	—	—	—	—	—	92
51 „	—	—	—	—	—	—	1	—	7	16	6	13	13	3	3	1	—	—	1	—	—	—	—	—	64
52 „	—	—	—	—	—	—	—	—	2	4	7	2	8	6	6	2	2	1	—	—	1	1	—	—	42
53 „	—	—	—	—	—	—	—	1	—	—	3	2	3	2	2	1	3	—	—	—	—	—	—	1	18
54 „	—	—	—	—	—	—	1	—	—	—	1	—	2	—	3	2	2	—	1	—	—	—	—	—	12
55 „	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	1	—	—	—	2	—	—	—	—	6
56 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
Totals	3	9	17	24	30	49	58	63	61	57	52	42	37	17	17	8	7	1	2	3	1	1	—	1	560

TABLE 69. *Glasgow. Height and Weight of Girls of Group D.*

Height	Weight of Girls of 10·5—11·5 years																										Totals
	43·5—	45·5—	47·5—	49·5—	51·5—	53·5—	55·5—	57·5—	59·5—	61·5—	63·5—	65·5—	67·5—	69·5—	71·5—	73·5—	75·5—	77·5—	79·5—	81·5—	83·5—	85·5—	87·5—	89·5—	—	111·5—	
$42\frac{7}{8}$—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
43 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
44 ,,	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
45 ,,	—	—	1	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
46 ,,	—	—	3	2	2	1	1	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11
47 ,,	—	2	2	5	7	3	3	5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	28
48 ,,	—	—	1	1	4	5	10	4	7	3	1	1	1	1	—	—	—	—	—	—	—	—	—	—	—	—	39
49 ,,	—	—	—	—	1	9	14	16	14	5	5	3	2	1	1	1	—	—	—	—	—	—	—	—	—	—	72
50 ,,	—	—	1	—	—	4	7	8	14	10	15	7	1	1	3	1	—	1	—	1	—	—	—	—	—	—	74
51 ,,	—	—	—	—	—	1	6	6	4	12	14	10	11	4	5	—	2	2	—	—	—	—	—	—	—	—	77
52 ,,	—	—	—	—	—	—	—	1	4	13	8	8	8	10	3	2	1	—	1	2	—	—	—	1	—	—	62
53 ,,	—	—	—	—	—	—	—	—	—	8	3	7	3	12	9	5	6	1	—	1	—	—	—	—	—	—	55
54 ,,	—	—	—	—	—	—	1	—	—	3	1	3	7	5	2	2	4	3	—	—	2	—	—	1	—	—	34
55 ,,	—	—	—	—	—	—	—	—	—	—	—	3	—	1	6	1	3	2	1	1	—	—	—	—	—	—	18
56 ,,	—	—	—	—	—	—	2	1	—	—	—	—	1	—	1	2	2	2	1	—	—	1	—	—	—	1	14
57 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	—	—	—	—	2
58 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	1	—	1	—	—	—	4
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
64 ,,	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	1	2	9	10	15	24	44	43	44	54	47	43	34	35	30	14	18	13	4	5	3	2	1	2	—	1	498

TABLE 70. *Glasgow. Height and Weight of Girls of Group D.*

Height	Weight of Girls of 11·5—12·5 years																						Totals
	48·5—	51·5—	54·5—	57·5—	60·5—	63·5—	66·5—	69·5—	72·5—	75·5—	78·5—	81·5—	84·5—	87·5—	90·5—	93·5—	96·5—	99·5—	102·5—	105·5—	108·5—	111·5—	
$46\frac{7}{8}$—	1	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
47 ,,	—	4	2	7	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15
48 ,,	1	3	4	6	—	3	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	18
49 ,,	2	3	7	5	11	5	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	36
50 ,,	—	—	7	17	12	10	4	1	1	2	—	1	—	—	—	—	—	—	—	—	—	—	55
51 ,,	—	2	8	5	14	18	17	9	3	1	1	—	—	—	—	—	—	—	—	—	—	—	78
52 ,,	—	—	—	2	8	13	17	8	9	3	—	1	—	1	—	1	—	—	—	—	—	—	63
53 ,,	—	1	—	—	9	13	15	18	9	7	5	—	—	1	—	—	—	—	—	—	—	—	78
54 ,,	—	1	—	—	5	10	12	21	17	15	6	4	2	1	1	—	—	—	—	—	—	—	95
55 ,,	—	—	—	—	1	1	8	4	6	7	5	10	6	3	2	—	—	—	—	—	—	—	53
56 ,,	—	—	1	—	—	2	—	1	13	5	3	13	5	2	1	—	—	—	—	—	—	1	47
57 ,,	—	—	—	—	—	—	1	2	1	6	6	4	3	1	1	—	1	—	1	—	—	—	27
58 ,,	—	—	—	—	—	—	—	—	1	1	—	4	—	1	—	—	2	1	—	—	—	1	11
59 ,,	—	—	—	—	—	—	—	1	—	—	—	1	—	1	1	2	1	—	—	—	—	—	7
60 ,,	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	1	1	1	—	—	—	5
61 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
62 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
63 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
64 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
65 ,,	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Totals	4	15	30	42	61	77	76	65	62	47	26	38	17	11	7	3	5	2	2	1	—	3	594

TABLE 71. *Glasgow. Height and Weight of Girls of Group D.*

Height	Weight of Girls of 12·5—13·5 years																		Totals
	45·5—	49·5—	53·5—	57·5—	61·5—	65·5—	69·5—	73·5—	77·5—	81·5—	85·5—	89·5—	93·5—	97·5—	101·5—	105·5—	109·5—	113·5—	
48⅞—	—	—	—	1	4	1	—	—	—	—	—	—	—	—	—	—	—	—	6
49 „	—	1	3	9	1	1	1	—	—	—	—	—	—	—	—	—	—	—	16
50 „	1	—	3	2	2	1	—	—	—	—	—	—	—	—	—	—	—	—	9
51 „	—	—	1	8	9	7	4	—	2	2	—	—	—	—	—	—	—	—	33
52 „	—	1	—	6	2	13	8	4	1	1	—	1	—	—	—	—	—	—	37
53 „	—	—	—	2	10	14	13	8	2	1	1	—	—	—	—	—	—	—	51
54 „	—	—	—	—	6	4	19	15	13	6	1	3	—	—	—	—	—	—	67
55 „	—	—	—	1	1	3	11	22	9	10	4	4	—	—	—	1	—	—	66
56 „	—	—	—	—	—	1	6	18	11	14	9	8	1	1	1	—	—	1	71
57 „	—	—	—	—	—	—	3	3	12	20	12	16	6	2	1	—	—	1	76
58 „	—	—	—	—	—	—	—	4	7	4	11	11	2	5	1	—	1	—	46
59 „	—	—	—	—	—	—	1	2	3	4	7	5	5	2	2	1	—	—	32
60 „	—	—	—	—	—	—	—	—	1	—	2	3	3	1	1	1	2	—	14
61 „	—	—	—	—	—	—	—	—	—	—	—	3	2	3	1	1	—	—	10
62 „	—	—	—	—	—	—	—	—	—	—	—	1	—	1	—	—	—	—	2
63- „	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	2
Totals	1	2	7	29	35	45	66	76	61	62	47	55	20	16	7	4	3	2	538

TABLE 72. *Glasgow. Height and Weight of Girls of Group D.*

Height	Weight of Girls of 13·5—14·5 years																						Totals
	49·5—	53·5—	57·5—	61·5—	65·5—	69·5—	73·5—	77·5—	81·5—	85·5—	89·5—	93·5—	97·5—	101·5—	105·5—	109·5—	113·5—	117·5—	121·5—	125·5—	—	134·5—	
46⅞—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
47 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
48 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
49 „	—	—	2	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
50 „	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
51 „	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
52 „	—	—	1	3	1	4	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	10
53 „	—	—	—	—	—	3	6	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	10
54 „	—	—	—	—	3	6	10	6	4	2	1	—	—	—	—	—	—	—	—	—	—	—	32
55 „	—	—	—	1	1	1	14	7	4	3	—	2	—	1	1	—	—	—	—	—	—	—	35
56 „	—	—	—	—	2	2	7	10	11	6	3	3	1	—	—	—	—	—	—	—	—	—	45
57 „	—	—	—	—	—	1	4	3	8	15	12	5	3	—	—	—	—	—	—	—	—	—	51
58 „	—	—	—	—	—	—	1	5	5	10	11	1	5	4	2	3	—	—	—	—	—	—	47
59 „	—	—	—	—	—	—	1	1	5	5	10	8	10	4	2	1	2	3	—	—	—	—	52
60 „	—	—	—	—	—	—	—	—	4	3	6	3	8	6	3	3	1	1	—	1	—	—	39
61 „	—	—	—	—	—	1	—	—	—	1	4	2	2	3	2	2	3	—	—	—	—	—	20
62 „	—	—	—	—	—	—	—	—	—	1	3	—	—	2	—	2	1	1	—	—	—	1	11
63 „	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	3	—	1	1	—	—	7
64 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
65 „	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	2
Totals	1	1	3	6	7	19	43	33	41	47	50	24	30	20	11	13	10	5	2	2	—	1	369

NUMERICAL ILLUSTRATIONS OF THE VARIATE DIFFERENCE CORRELATION METHOD.

By BEATRICE M. CAVE and KARL PEARSON, F.R.S.

In 1904 Miss F. E. Cave in a memoir on the correlation of barometric heights, published in the *R. S. Proc.* Vol. LXXIV. pp. 407 *et seq.*, endeavoured to get rid of seasonal change by correlating first differences of daily readings at two stations. A similar method was used by Mr R. H. Hooker in a paper published some time later in the *Journal of the Royal Statistical Society*, Vol. LXVIII. pp. 396 *et seq.*, 1905. This method was generalised by "Student" in the last number of *Biometrika* (Vol. x. pp. 179, 180). He showed that if there were two variates x and y, such that

$$x = \phi(t) + X,$$
$$y = f(t) + Y,$$

where X and Y are the parts of x and y independent of the time t, then the spurious correlation arising from x and y being both functions of the time could be got rid of by correlating the differences of x and y, and that ultimately, when m is sufficiently large:

$$r_{\Delta^m x \Delta^m y} = r_{\Delta^{m+1} x \Delta^{m+1} y} = \text{etc.} = r_{XY},$$

so that the correlation of x and y, free from the spurious time (or it might be position) correlation, i.e. r_{XY}, could be found by correlating the successive differences of x and y. When the correlations of the differences remain steady for several successive values, then we may reasonably suppose that we have reached the correlation r_{XY}*.

This method is still further developed by Dr Anderson of Petrograd, who in a valuable memoir published in this *Journal* has provided the probable errors of the successive difference correlations of a system of variables:

$$X_1, X_2, \ldots, X_n,$$
$$Y_1, Y_2, \ldots, Y_n,$$

---

* Having been in communication with "Student," while he was writing his paper, I know that the interpretation put by Dr Anderson (*Biometrika*, Vol. x. p. 279) on "Student's" words (*Ibid.* p. 180) is incorrect. "Student" had in mind, if he did not clearly express it, the ultimate steadiness of $r_{\Delta^m x \Delta^m y}$ for a succession of values of m. K. P.

where the correlations of random pairs of values of the variates, or the product sums,

$$S\{(X_p-\overline{X})(X_q-\overline{X})\},$$
$$S\{(Y_p-\overline{Y})(Y_q-\overline{Y})\},$$

are both zero.

Dr Anderson has further provided us with the values of the standard deviations of the successive differences, i.e.

$$\sigma_{\Delta^m X} \text{ and } \sigma_{\Delta^m Y},$$

which represent the ultimate values of $\sigma_{\Delta^m x}$ and $\sigma_{\Delta^m y}$, when we have carried m so far that the time effect has been eliminated.

The new method appears to be one of very great importance, and like many new methods it has been developed in a co-operative manner, which is a good reason for not entitling it by the name of any single contributor. We prefer to term it the *Variate Difference Correlation Method.*

With the exception of a few illustrations given by "Student," no numerical work on the correlation of the higher differences has yet been attempted. It is clear that much numerical work will have to be undertaken before we can feel complete confidence in our knowledge of the range and of the limitations of the new method. We have yet to ascertain how far in different types of material a real stability of difference correlations is ultimately reached, and how far various assumptions made in the course of the fundamental demonstration apply in dealing practically with actual statistical data. One of the most important assumptions made if there be n values of the variates is that arising from the reduction in the number of values as we take the means which occur in successive differences, and a like assumption is made in the case of standard deviations. Thus for example:

$$\frac{1}{n}\overset{n}{\underset{1}{S}}(X_s)=\overline{X},$$

but $\dfrac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X_s-X_{s+1})=(X_1-X_n)/(n-1)$, and will not be sensibly zero, although it is assumed to be, unless n be very large. Similar remarks apply to the sums used in the standard deviations, i.e. we assume in the proof $\dfrac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X_s^2)=\dfrac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X^2_{s+1})$.

Ultimately with the mth differences we come in the proof to relations of the type

$$\frac{1}{n-m}\overset{n-m}{\underset{1}{S}}(X_s)=\frac{1}{n}\overset{n}{\underset{1}{S}}(X_s)$$

and

$$\frac{1}{n-m}\overset{n-m}{\underset{1}{S}}(X_s^2)=\frac{1}{n}\overset{n}{\underset{1}{S}}(X_s^2).$$

Now such relations will undoubtedly be very approximately true, if the X's are random variates uncorrelated to each other, and provided m is small compared with n. These conditions seem amply satisfied when we proceed to fourth or sixth differences in barometric pressures, taken, say, over ten or twelve years; the addition of four or five daily pressures will hardly affect sensibly either the mean or the standard deviation. But such extensive data, while not only involving a great deal of labour in the difference work* are not those which, perhaps, most frequently demand the attention of the statistician, whether he be economist, sociologist or a student of scientific agriculture. In such cases it not infrequently happens that the available data only provide a range of 20 to, perhaps, at most 50 years; and we need to discover whether there is a true relationship between our variates, apart from a continuous change in both due to the time factor. At present accurate statistics of annual trade or revenue, or satisfactory annual demographic data hardly extend at most beyond a period of 50 years. Very often—under even approximately like methods of record—we shall hardly have more than twenty years' trustworthy returns. Not only has the method of record been changed, but the conditions of transit and trade may have been immensely modified and in a manner which we could not suppose to be even approximately represented by a continuous function of the time.

The object of the present paper is to illustrate the theory of the variate difference correlation method in its present stage of development on a *short* series of economic data, in order to test what approximation there is in such short series to stability, and further how nearly Dr Anderson's values for the successive standard deviations apply to such cases. We have selected as our data ten economic indices of Italian prosperity for the years 1885 to 1912, together with a "Synthetic Index," formed by taking the arithmetic mean of the ten economic indices referred to. These eleven indices are given by Professor Georgio Mortara in an interesting memoir: "Sintomi statistici delle condizione economiche d' Italia" which was published in the *Giornale degli Economisti e Rivista di Statistica,* for February, 1914, and form Tabella I, of that memoir, which we here reproduce in part as Table I. The indices in each case are obtained by dividing the returns for any year by the means of the returns for the years 1901—05, inclusive, and multiplying as usual by 100.

The indices are for returns of (i) Gross Receipts of Railways, (ii) Shipping, loaded and unloaded at the ports, (iii) Effective Revenue of the State, (iv) International Commerce, Value of Imports and Exports, (v) Number of postal Letters and private Telegrams, (vi) Amount of Stamp Duties, (vii) Savings Banks' Returns, (viii) Importation of Coal, (ix) Gross returns of consumption of Tobacco, (x) Returns of Coffee imported. Professor Mortara has drawn attention to the very high correlations of these individual indices with each other, and of each of them with the "Synthetic Index." The latter correlation is, however, to a certain

---

* A discussion of the correlations of the higher differences in barometric pressures will we hope be shortly issued.

extent spurious. For if $I_1, I_2, \dots I_{10}$ be the individual indices and I_s the synthetic index, then $I_s = \frac{1}{10}(I_1 + I_2 + \dots + I_{10})$ and any individual index I_q, if there be no correlation between such individual indices, would give

$$\frac{1}{n} S(I_s - \bar{I}_s)(I_q - \bar{I}_q) = \frac{1}{10n} S(I_q - \bar{I}_q)^2 = \frac{1}{10}\sigma^2_{I_q},$$

$$\sigma^2_{I_s} = \frac{1}{n} S(I_s - \bar{I}_s)^2 = \frac{1}{100}(\sigma^2_{I_1} + \sigma^2_{I_2} + \dots + \sigma^2_{I_{10}}),$$

and accordingly

$$r_{I_s I_q} = \frac{\frac{1}{10}\sigma^2_{I_q}}{\sigma_{I_s} \times \sigma_{I_q}} = \frac{\sigma_{I_q}}{\sqrt{\sigma^2_{I_1} + \sigma^2_{I_2} + \dots + \sigma^2_{I_{10}}}},$$

TABLE I.

Professor Mortara's Table of Index Values of Italy.

Numerical Index (Mean 1901—1905) = 100.

Year	Railways	Shipping	Revenue	International Commerce	Post and Telegraphs	Stamp Duties	Savings Banks	Coal	Tobacco	Coffee	Synthetical Index or Arithmetic Mean Index
1885	61	63	78	72	38	82	47	53	82	98	67·4
1886	62	63	79	74	40	87	53	52	86	94	69·0
1887	68	73	82	78	42	94	55	64	88	91	73·5
1888	70	71	83	62	43	98	57	69	86	81	72·0
1889	71	77	85	70	45	99	58	71	86	78	74·0
1890	72	78	86	66	46	97	59	78	87	81	75·0
1891	72	72	85	60	48	96	60	70	88	80	73·1
1892	71	75	86	64	51	96	64	69	89	80	74·5
1893	70	70	85	64	55	95	65	66	90	73	73·3
1894	72	72	86	63	57	93	65	84	89	71	75·2
1895	73	76	89	66	60	92	68	77	88	70	75·9
1896	75	77	90	67	64	94	70	73	88	73	77·1
1897	78	80	90	68	68	96	73	76	87	75	79·1
1898	81	84	91	78	73	96	75	79	89	78	82·4
1899	86	88	93	88	75	96	79	87	91	82	86·5
1900	89	89	94	91	78	96	83	88	92	82	88·2
1901	90	91	96	92	85	96	87	86	95	92	91·0
1902	96	99	98	95	93	97	92	96	97	94	95·7
1903	101	103	99	99	103	99	99	99	99	102	100·3
1904	105	102	101	103	111	101	107	105	102	103	104·0
1905	108	105	105	111	108	107	115	114	106	109	108·8
1906	119	123	108	132	107	115	127	137	109	118	119·5
1907	126	125	108	144	114	120	144	148	115	125	126·9
1908	134	129	113	139	122	120	155	150	124	132	131·8
1909	139	140	121	149	132	123	168	165	131	140	140·8
1910	146	146	129	159	140	130	180	166	138	147	148·1
1911	155	156	135	167	149	136	187	171	144	153	155·3
1912	165	169	138	179	158	141	192	179	151	160	163·2
Mean	94·8	116·3	97·6	96·4	82·3	103·3	95·9	99·0	100·6	98·6	96·5

which would have a substantial, but spurious value amounting to ·316 if $\sigma_{I_1} = \sigma_{I_2} = \dots \sigma_{I_{10}}$. If there be high correlation between the individual indices, of course, the correlation of each individual index with the arithmetic mean or synthetic index will also be high. Thus our Table IV shows it to range from ·952 in the case of Coffee to ·998 in the case of Railways. Possibly a third of this correlation may in some cases have a spurious origin. But the individual indices are very highly correlated together; only two such correlations are below ·9, and the lower of these two is as high as ·885. We are accordingly left with fifty correlations ranging from ·885 to ·997 between the individual indices, and if we accept these as true measures, then it is clear that any one of these ten indices might be used as a reasonable index of Italian prosperity; it would for practical purposes be idle to calculate them all or to table their arithmetic mean.

But the high correlations found lay themselves open from our present standpoint to some suspicion of being solely due to the fact that during the 28 years under consideration Italy has progressively increased in population and accordingly the consumption of innumerable goods and the means of interchange have all grown together with the time. In other words the correlations we give under the heading of "quantities" in each separate section of Table IV are very high solely because the individual indices are variates increasing one and all as continuous functions of the time.

The material therefore seems especially suited to the application of the variate difference correlation method. For example, the correlation between the indices for tobacco and savings is ·984; are we to interpret this to signify, that, if there are large savings this means that much will be spent on tobacco? Or, is this high correlation simply in whole or part spurious, merely indicating that both savings and consumption of tobacco increased markedly with the time? Actually the correlation of first differences drops from ·984 to ·766, that of second differences is negative if insensible, while from there onwards it steadily increases *negatively*, till with the sixth differences we reach $-·431$, which seems to indicate that, when time has been eliminated, expenditure on tobacco in any year means less money saved. Again the coffee and tobacco indices appear very highly correlated, ·955, but by the third difference correlation we have reached about a third of the relationship, ·319, which is scarcely altered in the sixth difference correlation, ·326; we may assume therefore that there is probably a moderate "organic" relationship between the expenditures on coffee and tobacco, but the association is nothing like as close as would be suggested by the correlation of the raw indices.

The work has been done in the following manner: The successive differences of the indices up to the sixth were found. The means and standard deviations of these differences were calculated, and the correlations were then worked out in the product-moment manner. This involved the very laborious work of determining 385 coefficients, and then to these coefficients were added the probable errors as found by Dr Anderson's formulae. These probable errors are of course those of the correlations of $\Delta^m X$ and $\Delta^m Y$, and will not be the correct values for the probable errors of the correlations of $\Delta^m x$ and $\Delta^m y$, until $\Delta^m x = \Delta^m X$ and $\Delta^m y = \Delta^m Y$, i.e. until m is sufficiently great for t to have been eliminated. Further their accuracy depends on the vanishing of the means of the differences or on the equalities of the sums like

$$\frac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X_s),\quad \frac{1}{n-1}\overset{n}{\underset{2}{S}}(X_s),\ \text{etc.}$$

which, while true on the average, will only be approximately true in the actual instance if n be large. We give in Table II, the Mean Values of Index Differences.

TABLE II.

Mean Values of Indices and their Differences.

	Railways	Shipping	Revenue	International Commerce	Post and Telegraphs	Stamp Duties	Savings Banks	Coal	Tobacco	Coffee	Synthetical or Arithmetic Mean Index
Quantity ...	94·8	116·3	97·6	96·4	82·3	103·3	95·9	99·0	100·6	98·6	96·5
1st difference	−3·85	−3·93	−2·22	−3·96	−4·44	−2·19	−5·37	−4·67	−2·56	−2·30	−3·55
2nd "	− ·35	− ·50	− ·08	− ·38	− ·27	·00	+ ·04	− ·35	− ·12	− ·42	− ·24
3rd "	+ ·16	+ ·28	+ ·20	− ·08	·00	+ ·12	− ·08	+ ·40	− ·12	·00	+ ·09
4th "	− ·33	− ·88	− ·13	−1·17	·00	− ·28	+ ·04	− ·79	− ·16	− 42	− ·40
5th "	+ ·57	+2·26	+ ·26	+2·52	+ ·30	·00	− ·52	+1·87	+ ·22	+ ·87	+ ·83
6th "	− ·82	−2·00	− ·55	−6·41	+ ·18	− ·73	− ·05	+2·00	− ·95	−1·68	−1·10

It will be seen from this table that the means of the differences are far from zero even when we have reached a difference for which we may suppose the time to have been eliminated. This arises from the smallness of the series dealt with and shows us that we ought not to anticipate more than a rough accordance with theory, or only an approximate steadiness, for sums like

$$\frac{1}{n-m}\overset{n-m}{\underset{1}{S}}(X_s)$$

may grow less and less steady as m increases.

Similar considerations apply to the standard deviations of the differences. These will not at first obey Dr Anderson's formulae because they are the values for $\sigma_{\Delta^m x}$ and $\sigma_{\Delta^m y}$, and when we have taken m sufficiently high for

$$\sigma_{\Delta^m x} \text{ and } \sigma_{\Delta^m y} \text{ to be theoretically equal to } \sigma_{\Delta^m X} \text{ and } \sigma_{\Delta^m Y},$$

and accordingly the correlation should have begun to be steady, then some failure to obey Dr Anderson's formulae will arise, because the means of the differences are not truly zero and equalities of the type

$$\frac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X_s^2)=\frac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X^2_{s+1}),\ \text{etc.}$$

will not be satisfied when n is relatively small.

Dr Anderson gives the value of $\sigma^2_{\Delta^m X}$ in terms of σ^2_X, but we do not of course know σ^2_X, which will be very different from σ^2_x, and can only practically be found from the value of $\sigma^2_{\Delta^m x}$ itself, after that value has become equal to $\sigma^2_{\Delta^m X}$, i.e. after steadiness has set in. In order therefore to test the formulae we have formed the ratio of

$$\sigma^2_{\Delta^m x}/\sigma^2_{\Delta^{m-1}x} \text{ from } m=1 \text{ to } 6.$$

This equals $4-\frac{2}{m}$ in Dr Anderson's formulae for $\sigma^2_{\Delta^m X}/\sigma^2_{\Delta^{m-1}X}$ and therefore we have a good measure of the approach of $\Delta^m x$ to $\Delta^m X$, or of the growth of steadiness as apart from the correlations. The following Table III gives the values of the ratios of the squares of the standard deviations, theoretical values, actual values and the mean value for each of the differences of the ten individual indices.

TABLE III.

Values of $\sigma^2_{\Delta^m x}/\sigma^2_{\Delta^{m-1}x}$ *and their approach to* $4-\frac{2}{m}$.

m	Theoretical Series	Synthetic Index	Rail	Shipping	Revenue	International Commerce	Post	Stamp Duties	Savings	Coal	Tobacco	Coffee	Mean of 10 Index Difference Standard Deviation Ratios
1	2	·012	·012	·031	·019	·038	·009	·040	·010	·035	·022	·036	·025
2	3	·705	·708	1·834	·763	1·720	·799	·585	·350	2·074	·352	·843	1·003
3	3·333	3·107	2·816	3·093	2·124	3·032	1·959	1·660	2·214	3·075	2·213	3·307	2·549
4	3·5	3·167	3·128	3·174	2·747	3·213	2·597	2·008	3·106	3·379	3·025	3·619	3·000
5	3·6	3·143	3·449	3·189	3·020	3·104	3·010	2·328	3·275	3·580	3·117	3·701	3·177
6	3·667	3·149	3·711	3·195	3·164	2·881	3·208	2·499	3·455	3·682	3·101	3·791	3·269

It will be clear that until we reach the ratio of the square standard deviations of the third and second differences, there is no general approach to steadiness. After $m=3$, however, for $m=4$, 5 and 6, the ratio of the values for the mean of the series of individual indices to the theoretical value is ·86, ·88 and ·89, respectively.

Thus, there is increasing approach to agreement in the observed and theoretical values, but the approach is slow, and we believe that there is greater steadiness than is really indicated by this test. The source of this apparent unsteadiness lies we think in the relative largeness of m compared with n (i.e. at a maximum 6 as compared with 28), rather than in our not having taken sufficiently high differences.

We now turn to the correlations. These are given in Table IV, the actual values of the standard deviations of the quantities and their differences being recorded along the diagonal cells, while the other cells contain the correlations of each pair of variates and of their successive differences.

We will now consider these correlations in detail.

(*a*) *Synthetic Index* (Arithmetic Mean) with Individual Indices.

We see at once that the *Synthetic Index* is highly associated with *Shipping* ($> \cdot 85$), with *Importation of Coal* ($? > \cdot 75$), and *International Commerce* ($? < \cdot 68$), and fairly highly with *Revenue* (c. ·55). On the other hand the sixth difference correlations with *Post* (c. ·15), with *Stamp Duties* (c. ·24) and with *Savings* ($> \cdot 23$) are all such (i) that they might well have arisen from the spurious element in the *Synthetic Index* correlations, and are all less than their Andersonian (steady value) probable errors. Almost the same may be said of the *Railway Index*; it is not beyond suspicion of being spurious, and is scarcely significant having regard to its probable error. The *Consumption of Coffee* is also not very closely associated with the *Synthetic Index*; it is only about twice its probable error ($\cdot 427 \pm \cdot 205$), and a good deal of its value may be spurious. Further in the case of both *Coffee* and *Railways*, the correlations are still falling between ·04 and ·05 for each difference. The last individual index remaining is that for *Consumption of Tobacco* and although the correlation of sixth differences is not really significant it is negative ($- \cdot 247 \pm \cdot 235$), and is exhibiting a steady negative rise.

Stripped therefore of the common time factor the *Synthetic Index* will be seen to be no very appropriate measure of trade, business activity, and spare money for savings and luxuries. With *Post, Stamp Duties* and *Savings*, it has probably only a spurious relationship, expenditure on railways has little influence, that on luxuries is very slightly significant, or indeed in the case of tobacco negative. It is, however, closely related to variations in external trade, i.e. imports (including coal), to exports and shipping and to effective revenue. It appears to us that a suitable general index of prosperity, which will distinguish between a continuous growth in all factors with the time, and favourable and unfavourable fluctuations from this growth, can only be obtained, when there has been far more ample study of the associations of individual indices among themselves, and of these indices after they have been freed from the time factor, i.e. of associations between high difference correlations. From this standpoint the study of General Index theories is at present in its infancy.

(*b*) *Railway Index.* This index is very noteworthy in the nature of its associations after removal of the time factor. We have reached a steady correlation (c. ·62) with *Shipping*, but beyond this no values of first class importance appear. The relation of *Railways* and *Revenue* after falling practically

TABLE IV. *Italian Indices. Correlations and Probable Errors and* **Standard Deviations.**

		Synthetic Index	Rail	Shipping	Revenue	International Commerce	Post	Stamp Duties	Savings	Coal	Tobacco	Coffee
Synthetic Index	Quantities	**28·46**	·998 ± ·005	·996 ± ·005	·989 ± ·005	·988 ± ·005	·972 ± ·007	·965 ± ·009	·997 ± ·005	·989 ± ·005	·985 ± ·005	·952 ± ·012
	1st Diff.	**3·14**	·874 ± ·037	·839 ± ·047	·669 ± ·087	·789 ± ·060	·401 ± ·133	·651 ± ·091	·777 ± ·063	·660 ± ·089	·732 ± ·073	·759 ± ·067
	2nd ,,	**2·64**	·539 ± ·130	·888 ± ·039	·370 ± ·158	·804 ± ·065	– ·065 ± ·182	·383 ± ·156	·164 ± ·178	·709 ± ·091	– ·064 ± ·182	·504 ± ·135
	3rd ,,	**4·65**	·505 ± ·151	·895 ± ·040	·494 ± ·151	·806 ± ·071	·038 ± ·202	·367 ± ·175	·139 ± ·198	·713 ± ·099	– ·057 ± ·202	·526 ± ·146
	4th ,,	**8·28**	·451 ± ·175	·887 ± ·047	·546 ± ·154	·787 ± ·084	·108 ± ·217	·328 ± ·196	·160 ± ·214	·709 ± ·109	– ·060 ± ·219	·509 ± ·163
	5th ,,	**14·68**	·392 ± ·199	·874 ± ·056	·554 ± ·163	·744 ± ·105	·146 ± ·231	·275 ± ·219	·194 ± ·227	·721 ± ·113	– ·134 ± ·232	·468 ± ·184
	6th ,,	**26·05**	·353 ± ·219	·857 ± ·066	·547 ± ·177	·680 ± ·135	·153 ± ·245	·242 ± ·236	·232 ± ·237	·755 ± ·108	– ·247 ± ·235	·427 ± ·205
Rail	Quantities	·998 ± ·005	**29·80**	·997 ± ·005	·989 ± ·005	·982 ± ·005	·979 ± ·005	·961 ± ·010	·996 ± ·005	·989 ± ·005	·979 ± ·005	·938 ± ·015
	1st Diff.	·874 ± ·037	**3·23**	·769 ± ·065	·546 ± ·111	·561 ± ·109	·393 ± ·134	·646 ± ·092	·640 ± ·093	·590 ± ·103	·660 ± ·089	·604 ± ·100
	2nd ,,	·539 ± ·130	**2 72**	·620 ± ·112	·074 ± ·182	·226 ± ·173	– ·140 ± ·181	·258 ± ·170	– ·198 ± ·175	·529 ± ·132	– ·204 ± ·175	·044 ± ·182
	3rd ,,	·505 ± ·151	**4·56**	·621 ± ·124	·241 ± ·190	·207 ± ·194	– ·153 ± ·198	·159 ± ·197	– ·291 ± ·185	·489 ± ·154	– ·164 ± ·197	·018 ± ·202
	4th ,,	·451 ± ·175	**8·06**	·616 ± ·136	·329 ± ·196	·158 ± ·214	– ·184 ± ·212	·017 ± ·220	– ·341 ± ·194	·442 ± ·177	– ·157 ± ·214	– ·053 ± ·219
	5th ,,	·392 ± ·199	**14·97**	·609 ± ·148	·381 ± ·201	·077 ± ·235	– ·209 ± ·226	– ·119 ± ·233	– ·386 ± ·200	·402 ± ·197	– ·198 ± ·227	– ·139 ± ·232
	6th ,,	·352 ± ·219	**28·84**	·621 ± ·154	·422 ± ·206	– ·007 ± ·250	– ·214 ± ·239	– ·261 ± ·233	– ·431 ± ·204	·383 ± ·214	– ·243 ± ·236	– ·204 ± ·240
Shipping	Quantities	·996 ± ·005	·997 ± ·005	**29·14**	·990 ± ·005	·981 ± ·005	·970 ± ·008	·967 ± ·008	·990 ± ·005	·986 ± ·005	·978 ± ·006	·937 ± ·016
	1st Diff.	·839 ± ·047	·769 ± ·065	**5·17**	·602 ± ·101	·705 ± ·079	·196 ± ·152	·563 ± ·108	·402 ± ·132	·632 ± ·095	·419 ± ·130	·532 ± ·113
	2nd ,,	·888 ± ·039	·620 ± ·112	**7·00**	·514 ± ·134	·667 ± ·101	– ·182 ± ·177	·295 ± ·167	– ·085 ± ·181	·604 ± ·116	– ·173 ± ·177	·388 ± ·155
	3rd ,,	·895 ± ·040	·621 ± ·124	**12·31**	·673 ± ·111	·682 ± ·108	– ·127 ± ·199	·249 ± ·190	– ·074 ± ·201	·594 ± ·131	– ·122 ± ·199	·397 ± ·170
	4th ,,	·887 ± ·047	·616 ± ·136	**21·94**	·733 ± ·102	·661 ± ·124	– ·079 ± ·214	·160 ± ·214	– ·066 ± ·219	·578 ± ·146	– ·123 ± ·216	·360 ± ·191
	5th ,,	·874 ± ·056	·609 ± ·148	**39·18**	·756 ± ·101	·614 ± ·147	– ·054 ± ·236	·067 ± ·235	– ·049 ± ·236	·571 ± ·156	– ·189 ± ·228	·298 ± ·215
	6th ,,	·857 ± ·066	·621 ± ·154	**70·03**	·766 ± ·103	·542 ± ·177	– ·059 ± ·249	– ·009 ± ·250	– ·034 ± ·250	·579 ± ·166	– ·271 ± ·229	·232 ± ·237
Revenue	Quantities	·989 ± ·005	·989 ± ·005	·990 ± ·005	**16·28**	·962 ± ·010	·971 ± ·007	·961 ± ·010	·988 ± ·005	·974 ± ·007	·983 ± ·005	·918 ± ·020
	1st Diff.	·669 ± ·087	·546 ± ·111	·602 ± ·101	**2·27**	·369 ± ·136	·368 ± ·137	·471 ± ·123	·541 ± ·112	·279 ± ·146	·665 ± ·088	·491 ± ·120
	2nd ,,	·370 ± ·158	·074 ± ·182	·514 ± ·134	**1·98**	·229 ± ·173	– ·282 ± ·168	·060 ± ·182	– ·144 ± ·179	·155 ± ·178	·102 ± ·181	·267 ± ·170
	3rd ,,	·494 ± ·151	·241 ± ·190	·673 ± ·111	**2·88**	·344 ± ·178	– ·381 ± ·173	·002 ± ·202	– ·247 ± ·190	·222 ± ·192	·210 ± ·193	·390 ± ·172
	4th ,,	·546 ± ·154	·329 ± ·196	·733 ± ·102	**4·78**	·357 ± ·192	– ·381 ± ·188	– ·084 ± ·218	– ·211 ± ·210	·240 ± ·207	·243 ± ·207	·428 ± ·179
	5th ,,	·554 ± ·163	·381 ± ·201	·756 ± ·101	**8·31**	·308 ± ·214	– ·374 ± ·202	– ·168 ± ·230	– ·181 ± ·229	·249 ± ·222	·194 ± ·227	·419 ± ·194
	6th ,,	·547 ± ·177	·422 ± ·206	·766 ± ·103	**14·78**	·214 ± ·239	– ·388 ± ·213	– ·255 ± ·234	– ·154 ± ·244	·270 ± ·232	·115 ± ·247	·400 ± ·210
Internat. Commerce	Quantities	·988 ± ·005	·982 ± ·005	·981 ± ·005	·962 ± ·010	**35·83**	·949 ± ·013	·941 ± ·015	·979 ± ·005	·970 ± ·008	·967 ± ·008	·969 ± ·008
	1st Diff.	·789 ± ·060	·561 ± ·109	·705 ± ·079	·369 ± ·136	**7·02**	·136 ± ·155	·481 ± ·121	·518 ± ·116	·449 ± ·126	·433 ± ·128	·589 ± ·103
	2nd ,,	·804 ± ·065	·226 ± ·173	·667 ± ·101	·229 ± ·173	**9·20**	– ·027 ± ·182	·410 ± ·152	·265 ± ·170	·300 ± ·166	·070 ± ·182	·342 ± ·161
	3rd ,,	·806 ± ·071	·207 ± ·194	·682 ± ·108	·344 ± ·178	**16·02**	·182 ± ·196	·442 ± ·163	·266 ± ·188	·273 ± ·187	·117 ± ·200	·340 ± ·179
	4th ,,	·787 ± ·084	·158 ± ·214	·661 ± ·124	·357 ± ·192	**28·72**	·317 ± ·197	·461 ± ·173	·249 ± ·206	·233 ± ·208	·151 ± ·215	·304 ± ·199
	5th ,,	·744 ± ·105	·077 ± ·235	·614 ± ·147	·308 ± ·214	**50·61**	·405 ± ·197	·452 ± ·187	·251 ± ·222	·193 ± ·228	·117 ± ·233	·233 ± ·224
	6th ,,	·680 ± ·135	– ·007 ± ·250	·542 ± ·177	·214 ± ·239	**85·89**	·475 ± ·194	·465 ± ·196	·275 ± ·231	·171 ± ·243	·015 ± ·250	·142 ± ·245

Post	Quantities	·972 ± ·007	·979 ± ·005	·970 ± ·008	·971 ± ·007	·949 ± ·013	**35·64**	·898 ± ·025	·972 ± ·007	·957 ± ·011	·935 ± ·016	·885 ± ·028
	1st Diff.	·401 ± ·133	·393 ± ·134	·196 ± ·152	·368 ± ·137	·136 ± ·155	**3·44**	·025 ± ·158	·401 ± ·133	– ·069 ± ·157	·552 ± ·110	·449 ± ·126
	2nd „	– ·065 ± ·182	– ·140 ± ·181	– ·182 ± ·177	– ·282 ± ·168	– ·027 ± ·182	**3·07**	– ·339 ± ·162	·208 ± ·175	– ·209 ± ·175	·189 ± ·176	– ·054 ± ·182
	3rd „	·038 ± ·202	– ·153 ± ·198	– ·127 ± ·199	– ·381 ± ·173	·182 ± ·196	**4·30**	– ·244 ± ·190	·307 ± ·183	– ·131 ± ·199	·131 ± ·199	– ·058 ± ·202
	4th „	·108 ± ·217	– ·184 ± ·212	– ·079 ± ·214	– ·381 ± ·188	·317 ± ·197	**6·93**	– ·162 ± ·214	·325 ± ·196	– ·085 ± ·218	·083 ± ·218	– ·068 ± ·219
	5th „	·146 ± ·231	– ·209 ± ·226	– ·054 ± ·236	– ·374 ± ·202	·405 ± ·197	**12·02**	– ·098 ± ·234	·334 ± ·209	– ·057 ± ·236	·044 ± ·236	– ·094 ± ·234
	6th „	·153 ± ·245	– ·214 ± ·239	– ·059 ± ·249	– ·388 ± ·213	·475 ± ·194	**21·53**	– ·027 ± ·250	·336 ± ·222	– ·050 ± ·250	·108 ± ·247	– ·133 ± ·246
Stamp Duties	Quantities	·965 ± ·009	·961 ± ·010	·967 ± ·008	·961 ± ·010	·941 ± ·015	·898 ± ·025	**14·66**	·963 ± ·009	·965 ± ·009	·973 ± ·007	·913 ± ·021
	1st Diff.	·651 ± ·091	·646 ± ·092	·563 ± ·108	·471 ± ·123	·481 ± ·121	·025 ± ·158	**2·95**	·563 ± ·108	·413 ± ·131	·461 ± ·124	·263 ± ·147
	2nd „	·383 ± ·156	·258 ± ·170	·295 ± ·167	·060 ± ·182	·410 ± ·152	– ·339 ± ·162	**2·25**	·261 ± ·170	·166 ± ·178	– ·317 ± ·164	·085 ± ·181
	3rd „	·367 ± ·175	·159 ± ·197	·249 ± ·190	·002 ± ·202	·442 ± ·163	– ·244 ± ·190	**2·90**	·306 ± ·183	·137 ± ·199	– ·270 ± ·188	·163 ± ·197
	4th „	·328 ± ·196	·017 ± ·220	·160 ± ·214	– ·084 ± ·218	·461 ± ·173	– ·162 ± ·214	**4·11**	·375 ± ·189	·086 ± ·218	– ·166 ± ·213	·195 ± ·211
	5th „	·275 ± ·219	– ·119 ± ·233	·067 ± ·235	– ·164 ± ·230	·452 ± ·187	– ·098 ± ·234	**6·28**	·379 ± ·202	·050 ± ·236	– ·132 ± ·232	·201 ± ·227
	6th „	·242 ± ·236	– ·261 ± ·233	– ·009 ± ·250	– ·255 ± ·234	·465 ± ·196	– ·027 ± ·250	**9·92**	·353 ± ·219	·052 ± ·250	– ·129 ± ·246	·222 ± ·238
Savings	Quantities	·997 ± ·005	·996 ± ·005	·990 ± ·005	·988 ± ·005	·979 ± ·005	·972 ± ·007	·963 ± ·009	**44·11**	·990 ± ·005	·984 ± ·005	·941 ± ·015
	1st Diff.	·777 ± ·063	·640 ± ·093	·402 ± ·132	·541 ± ·112	·518 ± ·116	·401 ± ·133	·563 ± ·108	**4·31**	·381 ± ·135	·766 ± ·065	·637 ± ·094
	2nd „	·164 ± ·178	– ·198 ± ·175	– ·085 ± ·181	– ·144 ± ·179	·265 ± ·170	·208 ± ·175	·261 ± ·170	**2·55**	– ·057 ± ·182	– ·044 ± ·182	·041 ± ·182
	3rd „	·139 ± ·198	– ·291 ± ·185	– ·074 ± ·201	– ·247 ± ·190	·266 ± ·188	·307 ± ·183	·306 ± ·183	**3·79**	·005 ± ·202	– ·327 ± ·181	– ·013 ± ·202
	4th „	·160 ± ·214	– ·341 ± ·194	– ·066 ± ·219	– ·211 ± ·210	·249 ± ·206	·325 ± ·196	·375 ± ·189	**6·69**	·078 ± ·218	– ·380 ± ·188	– ·007 ± ·220
	5th „	·194 ± ·227	– ·386 ± ·200	– ·049 ± ·236	– ·181 ± ·229	·251 ± ·222	·334 ± ·209	·379 ± ·202	**12·10**	·144 ± ·231	– ·402 ± ·197	·197 ± ·227
	6th „	·232 ± ·237	– ·431 ± ·204	– ·034 ± ·250	– ·154 ± ·244	·275 ± ·231	·336 ± ·222	·353 ± ·219	**22·49**	·196 ± ·241	– ·431 ± ·204	·046 ± ·250
Coal	Quantities	·989 ± ·005	·989 ± ·005	·986 ± ·005	·974 ± ·007	·970 ± ·008	·957 ± ·011	·965 ± ·009	·990 ± ·005	**37·99**	·968 ± ·008	·921 ± ·019
	1st Diff.	·660 ± ·089	·590 ± ·103	·632 ± ·095	·279 ± ·146	·449 ± ·126	– ·069 ± ·157	·413 ± ·131	·381 ± ·135	**7·15**	·215 ± ·157	·283 ± ·145
	2nd „	·709 ± ·091	·529 ± ·132	·604 ± ·116	·155 ± ·178	·300 ± ·166	– ·209 ± ·175	·166 ± ·178	– ·057 ± ·182	**10·30**	– ·346 ± ·161	·150 ± ·178
	3rd „	·713 ± ·099	·489 ± ·154	·594 ± ·131	·222 ± ·192	·273 ± ·187	– ·131 ± ·199	·137 ± ·199	·005 ± ·202	**18·06**	– ·378 ± ·173	·196 ± ·195
	4th „	·709 ± ·109	·442 ± ·177	·578 ± ·146	·240 ± ·207	·233 ± ·208	– ·085 ± ·218	·086 ± ·218	·078 ± ·218	**33·19**	– ·418 ± ·181	·186 ± ·212
	5th „	·721 ± ·113	·402 ± ·197	·571 ± ·156	·249 ± ·222	·193 ± ·228	– ·057 ± ·236	·050 ± ·236	·144 ± ·231	**62·80**	– ·474 ± ·183	·163 ± ·230
	6th „	·755 ± ·108	·383 ± ·214	·579 ± ·166	·270 ± ·232	·171 ± ·243	– ·050 ± ·250	·052 ± ·250	·196 ± ·241	**120·50**	– ·514 ± ·184	·152 ± ·245
Tobacco	Quantities	·985 ± ·005	·979 ± ·005	·978 ± ·006	·983 ± ·005	·967 ± ·008	·935 ± ·016	·973 ± ·007	·984 ± ·005	·968 ± ·008	**19·20**	·955 ± ·011
	1st Diff.	·732 ± ·073	·660 ± ·089	·419 ± ·130	·665 ± ·088	·433 ± ·128	·552 ± ·110	·461 ± ·124	·766 ± ·065	·215 ± ·157	**2·82**	·653 ± ·091
	2nd „	– ·064 ± ·182	– ·204 ± ·175	– ·173 ± ·177	·102 ± ·181	·070 ± ·182	·189 ± ·176	– ·317 ± ·164	– ·044 ± ·182	– ·346 ± ·161	**1·67**	·319 ± ·164
	3rd „	– ·057 ± ·202	– ·164 ± ·197	– ·122 ± ·199	·210 ± ·193	·117 ± ·200	·131 ± ·199	– ·270 ± ·188	– ·327 ± ·181	– ·378 ± ·173	**2 49**	·349 ± ·178
	4th „	– ·060 ± ·219	– ·157 ± ·214	– ·123 ± ·216	·243 ± ·207	·151 ± ·215	·083 ± ·218	– ·166 ± ·213	– ·380 ± ·188	– ·418 ± ·181	**4·33**	·358 ± ·191
	5th „	– ·134 ± ·232	– ·198 ± ·227	– ·189 ± ·228	·194 ± ·227	·117 ± ·233	·044 ± ·236	– ·132 ± ·232	– ·402 ± ·197	– ·474 ± ·183	**7·64**	·344 ± ·208
	6th „	– ·247 ± ·235	– ·243 ± ·236	– ·271 ± ·229	·115 ± ·247	·015 ± ·250	·108 ± ·247	– ·129 ± ·246	– ·431 ± ·204	– ·514 ± ·184	**13·45**	·326 ± ·224
Coffee	Quantities	·952 ± ·012	·938 ± ·015	·937 ± ·016	·918 ± ·020	·969 ± ·008	·885 ± ·028	·913 ± ·021	·941 ± ·015	·921 ± ·019	·955 ± ·011	**26·33**
	1st Diff.	·759 ± ·067	·604 ± ·100	·532 ± ·113	·491 ± ·120	·589 ± ·103	·449 ± ·126	·263 ± ·147	·637 ± ·094	·283 ± ·145	·653 ± ·091	**5·00**
	2nd „	·504 ± ·135	·044 ± ·182	·388 ± ·155	·267 ± ·170	·342 ± ·161	– ·054 ± ·182	·085 ± ·181	·041 ± ·182	·150 ± ·178	·319 ± ·164	**4·59**
	3rd „	·526 ± ·146	·018 ± ·202	·397 ± ·170	·390 ± ·172	·340 ± ·179	– ·058 ± ·202	·163 ± ·197	– ·013 ± ·202	·196 ± ·195	·349 ± ·178	**8·35**
	4th „	·509 ± ·163	– ·053 ± ·219	·360 ± ·191	·428 ± ·179	·304 ± ·199	– ·068 ± ·219	·195 ± ·211	– ·007 ± ·220	·186 ± ·212	·358 ± ·191	**15·89**
	5th „	·468 ± ·184	– ·139 ± ·232	·298 ± ·215	·419 ± ·194	·233 ± ·224	– ·094 ± ·234	·201 ± ·227	·197 ± ·227	·163 ± ·230	·344 ± ·208	**30·57**
	6th „	·427 ± ·205	– ·204 ± ·240	·232 ± ·237	·400 ± ·210	·142 ± ·245	– ·133 ± ·246	·222 ± ·238	·046 ± ·250	·152 ± ·245	·326 ± ·224	**59·52**

to zero, now stands at something greater than ·42 and might rise higher, but the relation to *International Commerce* as a whole is zero, which suggests that the goods imported and exported are not in the bulk carried by rail. Further although the final value of the *Railway* and *Post* correlation is scarcely sensible (– ·214 ± ·239), it has been continuously negative from the second difference, and thus suggests that increased expenditure on the post means lessened profit for the railways. This might be interpreted in two ways: (i) that business conducted by post or telegram lessens rail intercommunication by person, or (ii) that in the case of state-railways, there is not an increased profit to the railways from carrying larger mails. But still more remarkable are the negative correlations of *Stamp Duties, Savings, Tobacco* and *Coffee* with *Railways*; none of them are very large, and all but savings, perhaps, of the order of their probable errors. But taken as a whole they suggest that when the Italian spends little money in going about, then he saves more, or spends more on such luxuries as tobacco and coffee. Lastly we have the *Coal Index.* It might be supposed that a year with great coal importation would signify great railway activity, and this is the judgment which would be made from the raw correlations of these variates. But the actual facts are exhibited in a correlation still falling at the sixth difference and hardly significant having regard to its probable error. The inferences formed must be: (i) that imported coal is used largely at the ports of disembarcation or travels inland by other than railway transit, (ii) that the imported coal is largely used on the railways themselves and that its cost is a heavy tax on their resources.

(*c*) *Shipping Index.* As we might anticipate this is highly correlated with (i) *Railways* (c. ·62), (ii) *Revenue* (c. ·75) and less highly but very significantly with (iii) *International Commerce* (c. 54) and (iv) *Coal* (c. ·58), but it appears to have no relation whatever with *Post, Stamp-Duties* and *Savings,* and when we come to luxuries, their importation is clearly not a factor of shipping prosperity. Neither in the case of *Tobacco* nor of *Coffee* are the correlations really significant; with the former we have an increasing negative correlation and with the latter a decreasing positive one already below its probable error. Thus we see that neither directly by bulk of importation nor indirectly by *immediate* increase of consumption, does a rise of shipping mark significant rises in the use of luxuries such as *Tobacco* and *Coffee.* It would be of interest to ascertain whether increased consumption of luxuries does not rather *follow* than accompany favourable trade fluctuations.

(*d*) *Revenue Index.* This index as we might expect is fairly highly correlated with *Shipping* (c. ·75). It has relatively small relation to *Railways* (·422 ± ·206) at least at the sixth difference and a somewhat similar value (c. ·42 ± ·20) for *Coffee.* Thus the suggestions arise that revenue is but little produced by the railways and that coffee is not a very large factor of the custom dues. It is astonishing to find, however, that *Post, Stamps* and *Savings* have *negative* correlations with *Revenue* of – ·388 ± ·213, – ·255 ± ·234 and – ·154 ± ·244 respectively, which, if scarcely significant, have been in each case for several differences persistent in sign. Even the correlation with *Tobacco* is small, falling and insignificant (< ·115 ± ·247), and that with *Coal* which might be supposed to be high as marking good trade times is hardly significant although apparently rising (? > ·270 ± ·232). Lastly the correlation of *Revenue* with *International Commerce* is again small, falling, and insignificant (< ·214 ± ·239). Thus *Revenue* or the "entrato effectivo dello stato" seems to provide an index which has little valuable relation to any other characteristic of prosperity beyond shipping.

(*e*) *International Commerce.* Here we find no single final individual index correlation greater than ·54, which is that for *Shipping*. The next most important correlations are with *Post* (> ·47) and *Stamp Duty* (c. ·46). With *Railways* the correlation is zero, and with *Revenue* also falling and insignificant. With *Savings*, *Coal, Tobacco* and *Coffee* the correlations are all insignificant; in fact in the last three cases not only are the values less than their probable errors, but they are still falling. It is thus clear that in Italy the total of Exports and Imports is no measure of all-round prosperity, they do not immediately increase either savings or the consumption of luxuries.

(*f*) *Post and Telegrams.* Here we have the lowest series of correlations we have yet reached. *Post* values have no significant relation to fluctuations in *Railway* (c. – ·20 ± ·24), to *Shipping* (– ·059 ± ·249), *Stamp Duties* (– ·027 ± ·250), *Coal* (– ·050 ± ·250), *Tobacco* (+ ·108 ± ·247) or *Coffee* (– ·133 ± ·246) Indices. It is significantly correlated only with *International Commerce* (> + ·47 ± ·19) and, perhaps, significantly with *Savings* (+ ·336 ± ·222) but negatively with *Revenue* (c. – ·38 ± ·21). In short the number of letters and telegrams in Italy is hardly a mark of any other favourable fluctuation in prosperity, beyond *International Commerce.*

(*g*) *Stamp Duties.* This Index is correlated positively and significantly with *International Commerce* (c. + ·46 ± ·20) and positively, and doubtfully with *Savings* (c. + ·35 ± ·22). It is correlated insignificantly and *negatively* with *Railways* (– ·261 ± ·233), *Shipping* (– ·009 ± ·250), *Revenue* (– ·255 ± ·234), *Post* (– ·027 ± ·250), and *Tobacco* (– ·129 ± ·246); it is correlated positively and insignificantly with *Coal* (+ ·052 ± ·250) and *Coffee* (+ ·222 ± ·238). Thus again freed from continuous time changes, fluctuations in the *Stamp Duty Index* are of small value as a measure of contemporaneous general prosperity.

(*h*) *Savings Bank Index.* There are practically only two correlations of any importance with *Savings* and these are both *negative*, namely those with *Railways* (– ·431 ± ·204) and with *Tobacco* (– ·431 ± ·204). Hence it would appear that when the Italian people is in a saving mood, it spares on transit by rail and on the consumption of tobacco, and when it expends on these luxuries, then it does not save. *Savings* have small and possibly not significant correlations with *Post* (> + ·33 ± ·22) and *Stamp Duties* (< + ·353 ± ·219), and insignificant and positive correlations with *International Commerce* (> + ·27 ± ·23), *Coal* (> + ·19 ± ·24) and *Coffee* (c. + ·05 ± ·25); they have insignificant *negative* correlations with *Shipping* (< − ·03 ± ·25) and *Revenue* (< − ·15 ± ·24).

Savings are thus—apart from continual time change—no very satisfactory measure of general prosperity, and a fluctuating increase is usually accompanied by a reduction of luxuries.

(*i*) *Coal Index.* The importation of coal has little relation to any factor of prosperity besides *Shipping* (c. + ·58 ± ·17). With *Railways* the correlation is not quite double the probable error and the value, even at the sixth difference, appears still falling. The correlation with *Revenue* only just exceeds the probable error (+ ·270 ± ·232). With *International Commerce* (+ ·171 ± ·243), *Stamp Duties* (+ ·052 ± ·250), *Savings* (+ ·196 ± ·241) and *Coffee* (+ ·152 ± ·245) the correlations are less than their probable errors, small and in some cases still falling. With the *Postal Index,* the correlation is negative, insignificant and falling. Alone in the case of the *Tobacco Index* does the correlation appear to be nearly as significant as in that of *Shipping,* but it is *negative* and increasing* (− ·514 ± ·184), while in the case of *Shipping* it was steady. It is singular to find that *Coal,* the increased import of which should mark increased industrial activity, is, beyond the naturally influenced *Shipping,* alone effectively associated with the consumption of *Tobacco.*

(*j*) *Tobacco Index.* This is of considerable interest as marking the association of indices of trade prosperity with the consumption of a luxury. With four exceptions *Tobacco* is negatively correlated, although often insignificantly, with the other indices. *Revenue* (+ ·115 ± ·247), *International Commerce* (+ ·015 ± ·250), and *Post* (+ ·108 ± ·247) are all positive, insignificant, and in the first two cases still falling. The correlation with *Coffee* is positive and might, perhaps, be significant (+ ·326 ± ·224), but it appears to be still falling. With *Coal* and *Savings* there are probably significant negative correlations (− ·514 ± ·184, and − ·431 ± ·204 respectively); with *Railways* (− ·243 ± ·236), *Shipping* (− ·271 ± ·229) and *Stamp Duties* (− ·129 ± ·246) there are insignificant negative correlations, but they tend to confirm each other in sign. Thus we see that the consumption of tobacco can hardly be considered as a measure of general prosperity; it appears to be greatest when trade conditions are unfavourable, and in particular when savings are least and manufacturing conditions as measured by the importation of coal are slack. The result suggests the pipe of the unemployed at the street corner, rather than the increased expenditure of the fully occupied artisan.

(*k*) *Coffee Index.* This is another luxury and the results are very similar. There appears a significant correlation with *Revenue* (+ ·400 ± ·210), which might easily be explained, and there is a falling but possibly significant correlation with *Tobacco* (+ ·326 ± ·224). With all other indices the relationships are

---

* It is, perhaps, hard to believe that so much smuggling could be carried on in colliers, that it would seriously affect the profits of the tobacco monopoly!

insignificant. *Railways* (− ·204 ± ·240), *Shipping* (+ ·232 ± ·237), *International Commerce* (+ ·142 ± ·245), *Post* (− ·133 ± ·246), *Stamp Duties* (+ ·222 ± 238), *Savings* (+ ·046 ± ·250) and *Coal* (+ ·152 ± ·245). Apart therefore from the general increase of consumption with the time, during which time the general prosperity of the nation has increased, it would not appear that the consumption of a luxury has any organic relationship to prosperity. We do not find that a favourable trade fluctuation is associated with increased consumption of luxuries. In fact the suggestion arises that in the case of tobacco the consumption may be greater in a period of depression.

Conclusions. While we lay no special stress on any of the results suggested by the difference correlations above studied—far more intimate economic knowledge of Italian affairs and methods of measurement would be requisite—we yet venture to insist on one or two general considerations.

The very superficial statements, so frequently met with, that such and such variates, both changing rapidly with the time, are essentially causative will doubtless cease to have any scientific currency, directly the method of variate differences is fully appreciated. We shall no longer assert that the fall of the phthisis death-rate can be off-hand causatively associated with the contemporaneous rise in the number of persons dying in institutions, or that the increased expenditure on luxuries is necessarily a measure of increased national prosperity.

If we turn as in the present paper to the actual correlations of the indices themselves, we find in every case an arid and scarcely undulating waste of high correlation. No one can obtain any nourishment whatever from the statement that the *Tobacco Index* is correlated with the *Revenue Index* to the amount of ·983 and with the *Savings Bank Index* to the extent of ·984 ! The organic relationship between these variates is wholly obscured by the continuous increase of all three of them with the time. But when we proceed to sixth differences and see that the consumption of tobacco has little, if any, relation to revenue, and is associated substantially but *negatively* with savings, we seem to touch realities, and realities of some worth. Again what can we learn, if we are told that the *Shipping Index* is correlated to the extent of ·99 with both the *Revenue* and the *Savings Bank* Indices? We might imagine, that increase of shipping was not only the primary cause of increase in Italian revenue, but also the essential origin of any increase in the Italian peasant's and artisan's savings ! An appeal to the variate difference method shows how fallacious such imaginings would be ! An examination of the sixth difference correlations shows that while prosperity of the revenue is closely associated with trade as measured by shipping (·77), the correlation is not nearly perfect; on the other hand there appears to be no significant organic correlation at all (− ·154 ± ·244) between the prosperity of the revenue and the savings of the Italian populace. As we have noted a knowledge of local conditions and methods of reckoning quantities might enable us to put other and, perhaps, more luminous interpretations on our results. But there can be small doubt that to proceed from the actual correlation of such indices to the correlations of their higher differences gives the feeling of clearing away the sand of the desert, and reaching all the ordered arrangements of an excavated town below; the slight undulations of the waste above are really fallacious, and enable us to appreciate nothing of the actual topography of the city.

The method is at present in its infancy, but it gives hope of greater results than almost any recent development of statistics, for there has been no source more fruitful of fallacious statistical argument than the common influence of the time factor. One sees at once how the method may be applied to growth problems in man and in lower forms of life with a view to measuring common extraneous influences, to a whole variety of economic and medical problems obscured by the influences of the national growth factor, and to a great range of questions in social affairs where contemporaneous change of the community in innumerable factors has been interpreted as a causative nexus, or society assumed to be at least an organic whole; the flowers in a meadow would undoubtedly exhibit highly correlated development, but it is not a measure of mutual causation, and the development of various social factors has to be freed from the time effect, before we can really appreciate their organic relationships.

In the present paper we have dealt only with very sparse "populations" (only 28 values of the variates), but this has enabled us to consider not only a very large number of correlations, but to see the practical influence of terminal conditions on our theory. This may we think be summed up in the statement that the Andersonian formulae for the standard deviations will hardly in many practical cases be more than very roughly approximated before the size of the population becomes too small to make the deductions reliable. Further in most cases our difference correlations have hardly even with the sixth differences reached a steady state. Possibly they have done so in the cases of *Rail* and *Shipping*, *Shipping* and *Post*, *Shipping* and *Coal*, *Revenue* and *Post*, *International Commerce* and *Stamp Duties*, *International Commerce* and *Savings*, *Savings* and *Coffee*, and in one or other additional pair. But in the great bulk of instances there is still a more or less steady rising or falling appreciable in the difference correlations, and all we can really say is that the final value, the true r_{XY}, will be somewhat greater or less than a given number. From an examination of the actual numerical working of the correlations, it appears to us that the terminal values are in the case of these short series of very great importance. It is further clear that the theory as given by "Student" depends upon certain equalities which are not fulfilled in practice in short series. We await with much interest the complete publication of Dr Anderson's work, and hope to find a fuller discussion of the allowance to be made in short series for the influence of the terminal state of affairs* on the steadiness of the series and on the approach to the standard-deviation formulae. But apart from these lesser points, our present numerical investigation has convinced us of the very great value of the new method of Variate Difference Correlations.

---

* For example if we measure X from its mean,

$$\sigma^2_{\Delta X}=\frac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X_s-X_{s+1})^2-(\overline{\Delta X})^2=\frac{1}{n-1}\left(\overset{n}{\underset{1}{S}}(X_s^2)-X_n^2+\overset{n}{\underset{1}{S}}(X_s^2)-X_1^2\right)-(\overline{\Delta X})^2,$$

since $\overset{n-1}{\underset{1}{S}}(X_s X_{s+1})$ is by hypothesis zero, $=2\sigma^2_{X_s}+\frac{2}{n-1}\{\sigma^2_{X_s}-\frac{1}{2}(X_1^2+X_n^2)\}-(\overline{\Delta X})^2$. The first term $2\sigma^2_{X_s}$ gives Dr Anderson's value of $\sigma^2_{\Delta X}$. Now $\overline{\Delta X}$ equals $\frac{1}{n-1}\overset{n-1}{\underset{1}{S}}(X_s-X_{s+1})=\frac{1}{n-1}(X_1-X_n)$. Thus the remainder is $\frac{2}{n-1}\left[\sigma^2_{X_s}-\frac{1}{2}\left\{X_1^2+X_n^2+\frac{1}{n-1}(X_1-X_n)^2\right\}\right]$. Now the average value on many trials of $\frac{1}{2}(X_1^2+X_n^2)$ will be $\sigma^2_{X_s}$ and of $(X_1-X_n)^2$, $2\sigma^2_X$, so that the full value may be $\frac{2}{(n-1)^2}\sigma^2_X$ and small for n large; but for n small as above such a relation as $\sigma^2_{\Delta X}=2\sigma^2_X$ and the similar but more complex relations of the standard deviation formulae for the higher differences need not hold for any individual case, and thus the steadiness of the difference correlation series, and the approach to the Andersonian formulae are very far from attained.

AN EXAMINATION OF SOME RECENT STUDIES OF THE INHERITANCE FACTOR IN INSANITY

BY DAVID HERON, D.Sc.*

In the last few years a number of studies of the inheritance factor in insanity have been published in America, Germany and England. The value of investigation of such a topic cannot be overestimated. We are quite certain that the prevalence of insanity is not falling; many of us indeed believe that the statistics suffice to demonstrate that it is substantially increasing, and that we can attribute this increase not in the first place to the intenser strain of modern life, but to the greater power of modern treatment to check or temporarily cure attack, and thus allow wider possibility of reproduction to members of affected stocks. Indeed the problem seems closely associated with an essential difficulty of modern civilisation, the greater protection of physically and mentally degenerate stocks unaccompanied by any adequate limitation of their thereby increased power of procreation; the inheritance factor thus tends to aid the relatively greater survival of the socially unfit. The studies we have referred to would be of great importance from this aspect of eugenics if (i) the data were collected without conscious or unconscious bias, and (ii) the inferences drawn from them followed logically from the data thus collected.

Unfortunately it is not only in the interpretation of statistics that adequate training is required. It is equally important that in the actual collection of them we should proceed, not only free from the bias which arises from the hurried acceptance of dogmatic theories of heredity, but what is often still more needful, free from the bias which is almost certain to waylay our progress, if we have not initially considered with trained insight the fallacies which may result from our method of recording or even tabulating our material. The day of the amateur in science is gone; no one now pays any attention to men who propound elaborate atomic theories or stellar hypotheses, without having had preliminary training in physical or astronomical science. There are still, however, some who appear willing to accept the statement of statistical data or the inferences drawn from those

---

* This paper formed the second portion of a lecture given at the Galton Laboratory on March 3, 1914.

data by men who have clearly had no adequate training in statistical science. The craniologist, the anthropologist, even the biological student of heredity and evolution are recognising that a statistical training is needful for the true interpretation of many of the facts in their special fields of research. The physiologist still appears to believe that he can deal with the average effects of diverse dietaries or the pathologist with the "mass-phenomena" of the hereditary factor in insanity without any training in statistical method. A physicist might just as logically assume that without mathematical training he could give an adequate mathematical account of a physical phenomenon, or a cosmic theorist suppose that he was effectively furnished for astronomical research by the perusal of a popular primer on the stars! The statistical calculus cannot be mastered by any easier road than the differential calculus, or, to put a more apt illustration, statistical training is as needful a preliminary to the handling of statistics, as time spent in a physiological laboratory to the effective handling of tissues. In twenty years it will be unnecessary to insist on these points, they will be universally recognised in the courts of science; but at present it is not only necessary to reiterate unpleasant truths, but to emphasise their validity by illustrations which bring home forcibly to scientist and layman alike the danger of amateur statistical handling. To state that a man is in error is not sufficient, if he continues time after time to repeat his assertions, apparently under the belief that incessant repetition will convince the world of the value of his theories.

In the case of the inheritance factor in insanity we are not dealing with any purely academic question of science. We are up against one of the most difficult problems of modern life, where true advice is of urgent importance to the nation as well as to the individual. It is not only the medical man but the layman who seeks guidance in the question of the marriage of members of insane stocks, and a laboratory like the Galton Laboratory knows how often advice on such points is sought. It is disheartening when help is rendered to the seeker to be faced with the criticism: "But Professor ——— says I may marry if I take a wife of sound stock," or "Dr ——— recommends marriage, although my father was insane, because I am over twenty-five and still sane myself." When teaching of this kind, arising solely from false interpretation of defective data, is spread widecast in a dozen different papers or journals, it is not sufficient to issue a brief statement of its futility. It is needful to give it the *coup de grâce* by a more lengthy criticism of its fallacies and their illustration in a form more likely to impress the imagination. The attempt is made in this paper to deal with only one of the authors, who have contributed fallacious eugenic rules to those seeking knowledge on the influence of the hereditary factor in insanity.

In a long series of papers Dr F. W. Mott, Pathologist to the London County Asylums, has stated that when the children of insane parents become insane, they do so at a much earlier age than did their parents, and on the basis of this assertion he has drawn some very sweeping conclusions for practical conduct. Thus in the

British Medical Journal of May 11, 1912 (p. 1060), he states that "this signal tendency of insane offspring to suffer with a more intense form of the disease and at an early age, as shown in the above figures and tables, is of great importance for the following reasons: first, it is one of Nature's methods of ending or mending a degenerate stock; secondly, it is of importance to the physician, for he can say that there is a diminishing risk of the child of an insane parent becoming insane after he has passed 25, a matter of great importance in the question of marriage; thirdly, it is of importance in connection with the subject of social surgery of the insane, for when the first attack of insanity occurs in the parent the children for the most part have all been born....Sterilization would therefore be applicable to relatively few parents admitted to asylums."

Put briefly, Dr Mott's views are that in "Antedating" or "Anticipation," in this alleged tendency of the offspring to become insane at any earlier age than their parents, we have Nature's method of purifying degenerate stocks, that the children of insane parents who are still normal at the age of 25 may safely marry*, and that it is useless to take any special measures to limit the reproduction of the insane since nearly all their children are born before the onset of insanity.

These conclusions, if proved to be correct, would be of the utmost importance to the Eugenist. If the Law of Antedating or Anticipation really acts in the way Dr Mott has suggested, then it would seem to be unnecessary to take any special Eugenic action in the case of the insane and indeed the "Law" has already been used in support of this view. Thus in a leading article in the *British Medical Journal*†, which deals with Dr Mott's work, it is stated that "This intensification of mental disease in the young—this 'anticipation' as it is called, which is one of Nature's methods of ending or mending a degenerate stock, is specially important in connection with sterilization, as the figures given by Dr Mott show that when the first attack of insanity occurs in the parent the children have for the most part all been born. Sterilization, therefore, would be applicable in relatively few cases."

It is at least obvious that when views such as these are taken of the "Law of Anticipation," it merits the most careful examination. Let us consider, then, first of all, Dr Mott's presentation of the case for anticipation. For some years past Dr Mott has been engaged in the collection of cases in which two or more members of a family are or have been resident in London County Asylums, and has noted wherever possible the age of onset of the insanity. Information was thus obtained regarding 217 pairs of father and offspring, and 291 pairs of mother and offspring and the results are summed up in the following table.

Thus in comparing the age at onset of insanity in father and offspring, we find that among the fathers only 1·4 °/₀ became insane before the age of 20, while among the offspring the percentage was 26·2. These figures are also shown graphically in

---

* See for instance *Problems in Eugenics*, p. 426.

† May 11, 1912, p. 1089.

Figs. 1 and 2[*]. Here the horizontal scale represents the age of onset in 5-year groups—the vertical scale the percentages of cases occurring in each age group.

TABLE I.

Percentages of Cases whose First Attack of Insanity occurred within Various Age-periods.

Age-periods	Father	Offspring	Mother	Offspring	
	Per cent.	Per cent.	Per cent.	Per cent.	
Under *20* years	1·4	26·2	0·6	27·8	
20—24 years ...	0·4	18·0	3·4	15·7	Adolescence
25—29 "	1·4	18·0	4·4	18·2	
30—34 "	9·6	13·0	7·8	13·4	
35—39 "	11·5	7·3	9·2	10·0	
40—44 "	9·2	6·4	10·3	5·8	
45—49 "	14·3	6·0	12·0	3·7	Involutional period
50—54 "	17·5	0·9	12·3	2·4	
55—59 "	13·8	3·7	14·0	1·7	
60—64 "	10·1	—	11·6	1·3	
65—69 "	5·0	—	8·8	—	
70—74 "	4·6	0·4	3·1	—	
75—79 "	0·4	—	1·3	—	
80 "	0·4	—	0·6	—	

I have been obliged to follow Dr Mott in treating the "under 20" group as a 5-year group as otherwise my diagrams would bear no resemblance to his, but this procedure is far from satisfactory when such a large proportion of the cases in this group are congenital cases in which the age of onset should be taken at 0 years. The tables and diagrams show that among the parents more than half the cases occur after the age of 50, while among the offspring, more than half occur before 30, and this is taken to prove that there is Anticipation or Antedating in Insanity.

This will perhaps be made more evident if the percentages of those who became insane before the age of 25 are given in each case. Among the fathers, 2 °/o and among the mothers, 4 °/o became insane before the age of 25. Among the offspring, on the other hand, the percentage is 44. Another way of looking at the matter is to take the average age of onset of insanity in each case. Dr Mott gives a Table showing these averages but unfortunately has omitted the congenital cases so that the extent of anticipation is considerably under-estimated, and the form in which the data are given does not permit of an accurate calculation of the actual averages. From the information given it appears, however, that the average age at onset of insanity among the parents is about 50 years, among the offspring about 26 years, showing an anticipation or antedating of some 24 years.

---

[*] I am very grateful to Miss H. Gertrude Jones, the Hon. Secretary of the Galton Laboratory, for the diagrams which illustrate this lecture.

Figs. 1 and 2. Diagrams to illustrate the Distribution of Age at Onset of Insanity in Parent and Offspring. (Mott.)

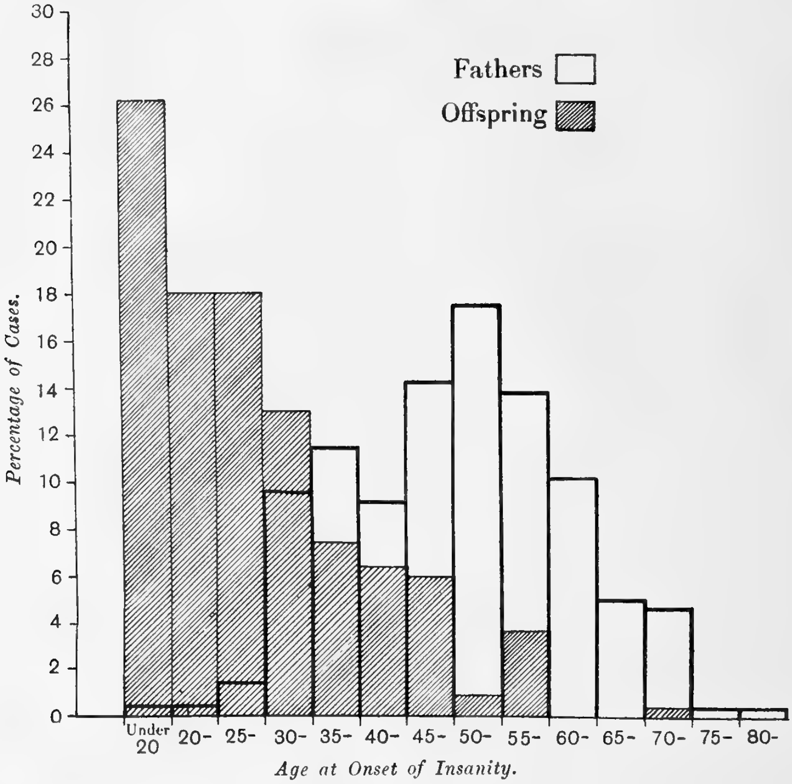


Fig. 1.

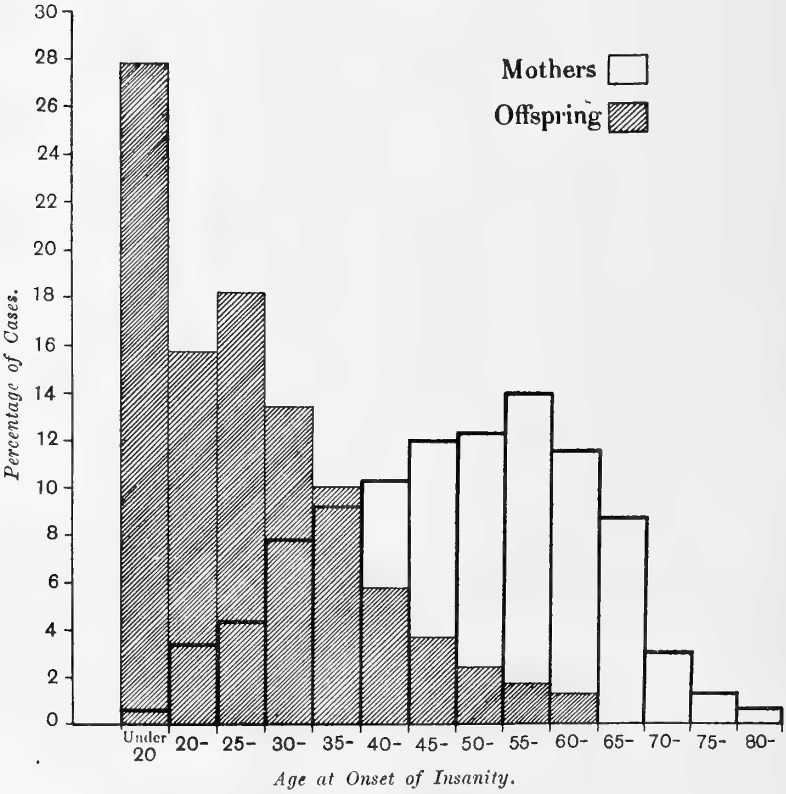


Fig. 2.

Now these conclusions, if satisfactorily demonstrated, would obviously be of the highest importance, but they were immediately challenged by Professor Karl Pearson in a letter which appeared in *Nature* of November 21, 1912 (p. 334). Professor Pearson's letter is as follows:

On an Apparent Fallacy in the Statistical Treatment of "Antedating" in the Inheritance of Pathological Conditions.

The problem of the antedating of family diseases is one of very great interest, and is likely to be more studied in the near future than ever it has been in the past. The idea of antedating, i.e. the appearance of an hereditary disease at an earlier age in the offspring than in the parent, has been referred to by Darwin and has no doubt been considered by others before him. Quite recently, studying the subject on insanity, Dr F. W. Mott speaks of antedating or anticipation as "Nature's method of eliminating unsound elements in a stock" ("Problems in Eugenics," papers communicated to the First International Eugenics Congress, 1912, p. 426).

I am unable to follow Dr Mott's proof of the case for antedating in insanity. It *appears* to me to depend upon a statistical fallacy, but this apparent fallacy may not be real, and I should like more light on the matter. This is peculiarly desirable, because I understand further evidence in favour of antedating is soon forthcoming for other diseases, and will follow much the same lines of reasoning. Let us consider the whole of one generation of affected persons at any time in the community, and let n_s represent the number who develop the disease at age s, then the generation is represented by

$$n_0,\ n_1,\ n_2,\ \dots\ n_s,\dots\ n_{100},\ \text{say.}$$

Possibly some of these groups will not appear at all, but that is of little importance for our present purpose.

Let us make the assumptions (1) that there is no antedating at all; (2) that there is no inheritance of age of onset; thus each individual reproduces the population of the affected reduced in the ratio of p to 1. Then the family of any affected person, whatever the age at which he developed the disease, would represent on the average the distribution

$$pn_0,\ pn_1,\ pn_2,\dots\ pn_s,\dots\ pn_{100}.$$

The sum of such families would give precisely the age distribution at onset of the preceding generation.

Now let us suppose that for any reason certain of the groups of the first generation do not produce offspring at all, or only in reduced numbers. Say that q_s only of the n_s are able to reproduce their kind; then of the older generation, *limited to parents*, the distribution will be

$$q_0 n_0 + q_1 n_1 + q_2 n_2 + \dots + q_s n_s + \dots + q_{100} n_{100},$$

but the younger generation will be

$$p\,(q_0 n_0 + q_1 n_1 + q_2 n_2 + \dots + q_s n_s + \dots + q_{100} n_{100})\,(n_0 + n_1 + \dots + n_s + \dots + n_{100}),$$

i.e. the relative proportions will remain absolutely the same.

The average age at onset and the frequency distribution of the older generation, that of the *parents*, will be entirely different from that of the offspring and will depend wholly on what values we give to the q's. If frequency curves be formed of the two generations they will differ substantially from each other. This difference is not a result or a demonstration of any physiological principle of antedating but is solely due to the fact that those who develop the disease at different ages are not equally likely to marry and become parents.

A quite striking instance of the fallacy, if it be such, would be to consider the antedating of "violent deaths." Fully a quarter of such deaths in males, nearly a half in females, occur before the age of twenty years. Consider now the parents and offspring who die from violent deaths; clearly there would be no representative of death from violence under twenty in the parent generation, and we should have a most marked case of antedating, because the offspring generation would contain all the infant deaths from violence.

In the case of insanity, is the man or woman who develops insanity at an early age as likely to become a parent as one who develops it at a later age? I think there is no doubt as to the answer to be given; those who become insane before twenty-five, even if they recover, are far less likely to become parents than those who become insane at late ages—many, indeed, of them considering the high death-rate of the insane, will die before they could become parents of large families. Now Dr Mott took 508 pairs of parents and offspring, "collected from the records of 464 insane parents whose 500 insane offspring had also been resident in the County Council Asylums," and ascertained the age of first attack. As at present advised, it seems to me that his data must indicate a most marked antedating of disease in the offspring, but an antedating which is wholly spurious. There is, I think, a further grievous fallacy involved in this method of considering the problem, but before discussing that I should like to see if my criticism of this method of approaching the problem of antedating can be met.

KARL PEARSON.

BIOMETRIC LABORATORY,
UNIVERSITY COLLEGE, LONDON,
November 11, 1912.

Dr Mott has referred to this letter in his *Report* for 1912*, but it will be more convenient to deal with his reply after we have examined the method by which his data have been collected and the use made of the data. Let us consider first of all how the data were obtained. Dr Mott in describing his material says that it consists of a collection of cases in the London County Asylums where two or more persons are related to one another. Thus Dr Mott has dealt—not with a series of complete pedigrees in which every member is included, whether insane or normal, but with a series of cases in which two or more members of a family are known to have been in London County Asylums. No notice is taken of those who are normal throughout their lives and no allowance is made for those who are normal at the time the record is made but who may afterwards become insane.

Do cases selected in this way provide a complete or impartial view of the facts? Some of Dr Mott's own comments on his data throw a considerable amount of light on this point. In his *Report* for 1909† he says: "From all the Asylums I have received valuable reports, but in the case of the older asylums it has been a matter of the utmost difficulty to trace the records of so many years back," and in his *Report* for 1910‡ he says, "Some of the asylum authorities have gone through their case books for a number of years back, but the results have not been satisfactory owing to the difficulty of obtaining particulars without a living representative of the family being resident in the asylum—for instance, 110 old cases

---

* *Annual Report of the London County Council for* 1912, Vol. II. p. 62.

† *Twentieth Annual Report of the Asylums Committee of the L.C.C.*, p. 90.

‡ *Twenty-first Report of the Asylums Committee of the L.C.C.*, p. 94.

reported from Bexley have been rejected as the relatives in the other London County Asylums could not be traced, for no instance has been included unless full particulars could be obtained."

It is thus clear that not all the cases could be traced and that there was special difficulty in tracing the older cases. What is the effect of a selection of this kind? A study of the following hypothetical cases may serve to throw some light on this point.

TABLE II.

Anticipation or Antedating in Insanity. Hypothetical Examples to show the Effect of Dr Mott's Selection of Cases.

		First Example	Second Example
Mother:	Born	1873	1833
	Married	1893	1853
	Became Insane and admitted to Asylum	1913	1873
	Age at First Attack	40	40
	Died	1914	1874
Son:	Born...	1894	1854
	Became Insane	1894*	1914
	Admitted to Asylum	1914	1914
	Age at First Attack	0	60

The mothers in those two examples have exactly parallel careers. In each case the mother became insane at the age of 40 and only lived one year in the asylum. In the first case the son was a congenital idiot but was only admitted to an asylum at the age of 20. The age of onset in this case is taken at 0 years and the case shows marked "anticipation." In the second case the mother also became insane at the age of 40, the son not till the age of 60, 40 years after his mother's death. The second example thus tells against the Law of Anticipation. Are these two cases equally likely to appear in Dr Mott's data?

In the first case mother and son are in the asylum at the same time and were admitted within a year of each other. It is very improbable that the relationship would escape notice and such a case is almost certain to be recorded. In the second case, however, the son is not admitted to an asylum till 40 years after his mother's death. Even if the family remained in the same area for 40 years after the mother's death, it would obviously be very difficult to connect the histories of mother and son. This case, which tells against the Law of Anticipation, is almost certain to escape notice. A spurious anticipation or antedating is thus inevitable owing to the method of collecting the data.

It has also been pointed out that Dr Mott has made no allowance for those who are mentally normal at the time the record is made but may subsequently

* Congenital Idiot.

become insane, and this introduces further spurious anticipation. Another hypothetical example will perhaps make this clear. Let us take the case of a mother with six children, five of whom have become insane as follows:

TABLE III.

	Mother	Children					
		1	*2*	*3*	*4*	*5*	*6*
Born	1830	1850	1852	1854	1856	1858	1860
Became Insane ...	1860	—	1872	1896	1914	1888	1860†
Age at Onset of Insanity	30	—*	20	42	58	30	0

The extent to which this family would show anticipation or antedating would depend very largely on the time at which the record was made as is shown in the following table.

TABLE IV.

Date of Record	Age of Onset of Insanity in		Average for Children	Amount of Anticipation
	Mother	Children		
1860	30	0	0	30
1872	30	0, 20	10	20
1888	30	0, 20, 30	16·7	13·3
1896	30	0, 20, 30, 42	23 0	7
1914	30	0, 20, 30, 42, 58	30	0

If the case were noted in 1860 then the age of onset of insanity in the mother is 30 years—of the child 0 years—a clear case of anticipation, and nothing would be known of the fact that four other children will afterwards become insane and will bring the average age of onset in the children up to 30 years—exactly the same as that of the mother. Nor is the record even now complete for if the eldest child ever becomes insane, the age of onset in his case must be at least 64 years and this will further increase the average age of onset in the children. It is thus clear that in dealing with incomplete families and ignoring the possibility that those who are normal at the time of record may afterwards become insane, Dr Mott has introduced a further spurious anticipation or antedating.

If we examine carefully the first pedigree given by Dr Mott at the Eugenics Congress‡, we see clearly how probably much of the anticipation recorded by

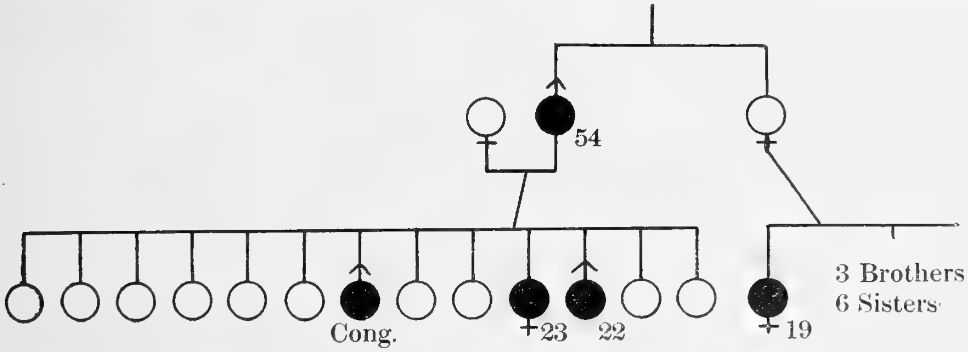
* Alive, 64 years of age and still normal.

† Congenital Idiot.

‡ *Problems in Eugenics*, p. 413.

Dr Mott has arisen. Unfortunately this is the only pedigree for which sufficient details have been given to enable its completeness to be tested. The pedigree and Dr Mott's description of it are as follows:

> "A.B., an alien Jew, aged 54, was admitted to an asylum for the first time suffering from involutional melancholia; he has a sister who has not been in an asylum, but, as events turned out, bore the latent seeds of insanity. The man is married to a healthy woman who bore him a large family; the first five are quite healthy, then comes a congenital imbecile epileptic (cong.)*, then two healthy children followed by a daughter who becomes insane at 23, then a son insane at 22, and lastly two children who are up to the present free from any taint. The sister of A.B. is married and has a family of ten, seven girls and three boys; one of the females was admitted to the asylum at the age of 19, and since this pedigree was constructed a brother of hers has been admitted aged 24. Half-black† circles are insane. The pedigree is instructive; it shows direct and collateral heredity; it also shows remarkably well the signal tendency to the occurrence of insanity at an early age in the children of an insane and potentially insane parent."



● = Insane.

13 children: 9 Alive, 4 Sons, 5 Daughters. 4 Dead. 3 Insane.

Fig. 3. Pedigree to illustrate the effect of Dr Mott's selection of cases.

F. W. Mott: "Heredity and Eugenics in relation to Insanity." *Problems in Eugenics*, p. 413.

This pedigree was given as above in July 1912, and in an address previously delivered before the Manchester Medical Society on Oct. 4, 1911, Dr Mott gave the same pedigree, but without any reference to the nephew of A.B. (brother of the girl who became insane at 19) who became insane "since the pedigree was constructed," so that this man became insane between 1911 and 1912 and this serves to "date" the pedigree.

Now it should be noted that at least five of the children of A.B. are over 23 years of age and up to the present time healthy. But all these children are alive and if any one of them afterwards becomes insane, the average age of onset of insanity in the children will be raised—and it is clear that the more incomplete the pedigree the greater the amount of spurious anticipation. Again Dr Mott states that in

---

* This does not agree with Dr Mott's pedigree which gives the congenital case as the seventh instead of the sixth child.

† According to our usual custom, they are represented by full black circles in Fig. 3.

nephews and nieces the age of onset is earlier than in uncles and aunts. In 1911 this pedigree gives a case in which an uncle became insane at 54, his niece at 19—but one year later a nephew who became insane at 24 has to be added, thus raising the average and there are eight more children some at least of whom may become insane at later ages. As before the incompleteness of the pedigree introduces an artificial and spurious anticipation or antedating. The remedy is obvious; we must only deal with completed families.

A further fallacy involved in Dr Mott's method of work must now be noted. In directly comparing the age of onset in parent and child, Dr Mott has ignored the fact that in the parent the incidence of insanity is for all practical purposes limited to the age of 20 and over since cases of congenital defect and of adolescent insanity hardly ever marry. Among the general population of asylums, however, 12 °/₀ become insane before the age of 20 and in Dr Mott's selected data the percentage rises to 27—or more than a quarter of the whole become insane before 20. This in itself causes a very marked spurious anticipation. As Professor Pearson has shown (p. 361 above) if we were to investigate the age at death in parent and child from accident or violence, we should find the same spurious anticipation.

There are thus three fallacies involved in Dr Mott's work. In the first place a spurious anticipation or antedating arises from the inclusion in the record of families whose history has not yet been completed, for those who become insane at late ages in the younger generation do not appear. Secondly, even with families whose history is completed, those cases in which the insanity of parent and child is contemporaneous are far more likely to be recorded than those in which the child becomes insane long after the parent*, and thus the cases which show anticipation are more likely to appear in the record than those which tell against Dr Mott's views. Thirdly, by directly comparing parent and child, he has practically limited one of the two groups which are being compared to ages at onset of over 20 years and has thus obtained further spurious anticipation.

Dr Mott also lays stress on the appearance of insanity in a more intense form in the younger generation. "I have proved," he says†, "that there is a signal tendency in the insane offspring of insane parents for the insanity to occur at an earlier age and in a more intense form in a large proportion of cases, for the form of insanity is usually either congenital imbecility, insanity of adolescence, or the more severe form of dementia praecox, the primary dementia of adolescence, which is generally an incurable disease." But we have already seen that Dr Mott's method of collecting his data is such that an enormous preponderance of early cases of insanity in the younger generation is inevitable and of course such cases are largely incurable. Type of disease is very closely related to the age of onset and

---

* Dr Mott states (*Archives of Neurology*, Vol. VI. p. 82) that "the main bulk of the cards (i.e. his records), however, refer to parents and offspring admitted to the asylums within the last fifteen years."

† *Archives of Neurology*, Vol. VI. p. 82.

by selecting the latter we can alter the proportion of any particular type of insanity. Dr Mott has obtained his material in such a way that, in the younger generation, cases of insanity coming on late in life are much less likely to be recorded than those which appear in early life, and hence the early cases are in a majority, but the change in age of onset, and consequently of the type, is entirely spurious and arises solely from the way in which the material has been obtained.

We can now deal with the reply Dr Mott has made to Professor Pearson's criticisms. In his *Annual Report* for 1912 (p. 62), Dr Mott says: "Professor Karl Pearson, writing to *Nature*, November 21, 1912, 'On an apparent fallacy in the statistical treatment of "Antedating" in the inheritance of pathological conditions,' criticises on mathematical grounds the evidence of anticipation. I do not feel myself competent to reply to the opinion of such an eminent authority on mathematics applied to biometrics, but it does not militate against my conclusions, nor explain away the fact that a large proportion of the insane offspring of insane parents are affected with imbecility or adolescent insanity; for granting the assumption that there is no antedating at all, we might rightly expect the ages at onset of insane offspring of insane parents to be comparable with the ages at onset of all the admissions to the asylums during the same period*. This is by no means the case, for amongst the insane offspring there is a far greater proportion affected early in life, as is shown in the following figures and curves" (they appear here as Fig. 4 and Table V).

According to these figures the onset of insanity among the *recorded* insane offspring of insane parents is considerably earlier than among the general admissions to asylums, but it has already been shown that this is due to the fact that the data have been selected in such a way that the early cases in the younger generation are the most likely to appear. Further, if Dr Mott's argument be a valid one, we might also expect the ages at onset of the insane parents of these insane offspring to be comparable with the ages at onset of all the admissions to asylums during the same period. This is by no means the case as is shown in Fig. 5 below (see also Tables I and V). We see here that the insanity of the parents comes on at a much *later* period than among the general admissions to asylums and that there is a far less proportion affected early in life. If Dr Mott's method of argument be sound, he has not only to deal with an antedating of insanity among the offspring but also a post-dating of insanity among the parents. Both are of course spurious and arise from the peculiar selection of the data and from the fact that, owing to differential death-rates, the ages at onset of "admissions" will never be the same as the ages at onset of the admitted—i.e. the asylum population—at any time.

---

* "We might rightly expect" these ages to be different, because "admissions" are not the same as the population in the country who have at one time or another been insane. The percentages of total cases of acute mania, of senile insanity, of congenital idiocy, and of melancholia, who reach the asylums, are not the same. The reader has to distinguish between the population of admissions, the population of admitted, and the insane population of the country. A sample of the latter may be reached from completed family histories, but not from records on admission or from records of an asylum population.

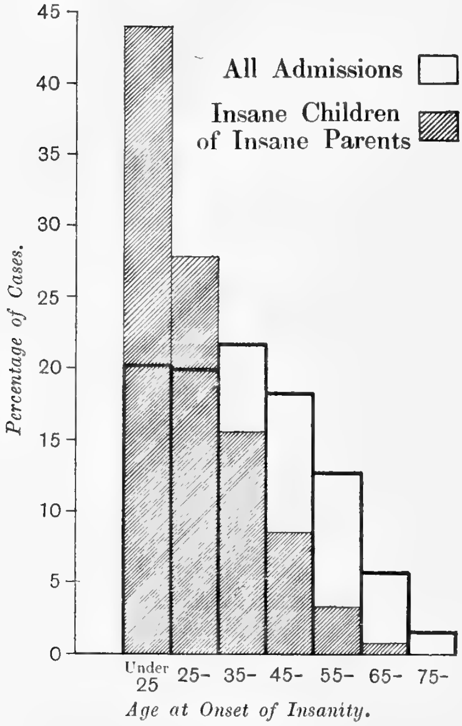


Fig. 4.

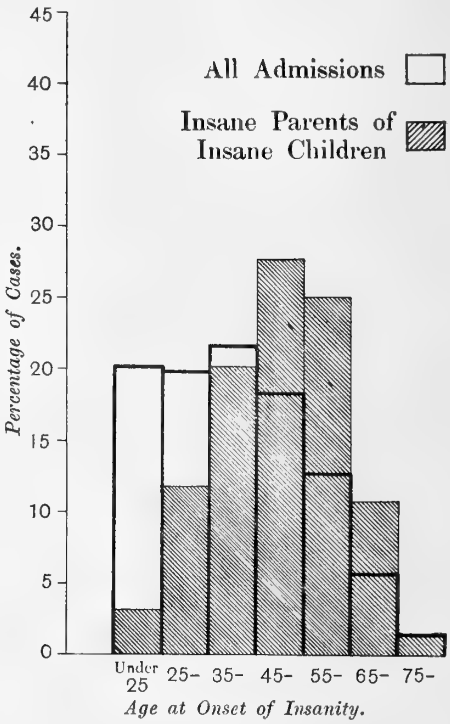


Fig. 5.

Diagram to illustrate the Distribution of Age at Onset of Insanity among:
(1) The Insane Offspring of Insane Parents.
(2) The Insane Parents of Insane Offspring.
(3) All Admissions to L.C.C. Asylums.

TABLE V. *Percentage Comparison of the Age at time of Onset of Insanity in the Insane Offspring of Insane Parents and the General Admissions to the London County Asylums.*

Age at Onset of Insanity	MALE		FEMALE		TOTAL	
	4482 direct admissions during last four years	274 insane offspring of insane parents	5097 direct admissions during last four years	389 insane offspring of insane parents	9579 direct admissions during last four years	663 insane offspring of insane parents
Under *25*	20·0	43·8	20·2	44·2	20·1	44·0
25—34	19·9	27·7	19·9	28·0	19·9	27·9
35—44	21·9	13·8	21·5	16·7	21·7	15·5
45—54	17·7	10·2	18·6	7·4	18·2	8·5
55—64	13·3	3·6	12·4	2·8	12·7	3·2
65—74	5·7	0·7	5·9	0·8	5·8	0·7
75	1·5	—	1·6	—	1·5	—

41 male	imbeciles	out of	274	offspring
54 female	”	”	389	”
95 male and female	”	”	663	”

It is possible to illustrate the various fallacies which vitiate Dr Mott's conclusions regarding anticipation by considering the age at death of parent and child. I do not know whether it is generally recognised that it is exceedingly difficult to get any considerable body of data in which the ages at death of a parent and all his children are given, for of course the record is incomplete and biassed until the death of the last surviving member, and in some cases to get a complete record we must trace the history of a family for over 150 years. George the IIIrd, for instance, was born in 1738 and all but one of his 15 children were still alive in 1810, 72 years afterwards, and the last surviving son, Duke of Cumberland and King of Hanover, did not die till 1851, 113 years after his father's birth—and this is by no means an extreme case. In the material I am about to describe I found one case where the interval was 160 years.

Another difficulty which arises is the tendency in practically all family histories to omit infant deaths, so that we do not get a complete record. It seems probable that the deaths of minors are not represented in such records in anything like their true proportion and that the differences are greater than might be expected to arise from differences of physique and nurture due to class. Thus records of the Landed Gentry give 31 deaths per 1000 males under 20 years* while actual experience shows 163 to 197 per 1000†. But in the records of the reigning families of Europe we get a practically complete record of all members and therefore from von Behr's *Genealogie der in Europa regierenden Fürstenhäuser*‡, I have extracted particulars of the age at death of over 2000 individuals—all belonging to the 18th century. There was here no selection—every child was entered and every family had been traced from the birth of the parents till the death of the last survivor.

Now in Dr Mott's data we have already seen that cases in which the age at onset of insanity in parent and child is contemporaneous are most likely to be recorded. We can test the effect of a selection of this kind by investigating the effect of selecting, from our data regarding the age at death among those royal families, only those individuals who died within a certain number of years of their father's death, and the results are given below in Table VI, p. 370.

When we deal with the whole of the data, absolutely unselected, every family being complete and traced to the death of the last surviving member, we find that 680 out of 1829 or 37·2 °/₀ died under 20 years of age. Let us now apply a very slight selection to the data and reject the 92 cases in which the interval between the deaths of father and child was at least 60 years. We find now that 680 out of the remaining 1737 died under 20 years of age—or 39·1 °/₀. Thus the effect of a selection of this kind is to cause a slight increase in the proportion of deaths at the early ages. If we make the selection slightly more stringent, by taking only those who died within 40 years of their father's death, the percentage of individuals dying under 20 years of age rises to 46·7 and if we go still further and consider

---

* See Pearson: *Proc. R. S.* Vol. 65, p. 291.

† *Statistics of Families*, p. 73.

‡ Tauchnitz, Leipzig, 1870.

TABLE VI.

Illustrating the Effect of Selection of Material on the Distribution of Age at Death.

(Reigning Houses in Europe—18th Century.)

Age at Death	All Cases Unselected Data		Children who died:							
			within 60 years of their father's death		within 40 years of their father's death		within 20 years of their father's death		in their father's lifetime	
	Numbers	°/₀	Numbers	°/₀	Numbers	°/₀	Numbers	°/₀	Numbers	°/₀
Under *20*	680	37·2	680	39·1	680	46·7	680	62·4	648	82·7
20—39	277	15·1	277	15·9	277	19·0	254	23·3	121	15·4
40—59	336	18·4	336	19·3	274	18·8	127	11·7	15	1·9
60—79	450	24·6	395	22·7	214	14·7	29	2·7	—	—
80 and over	86	4·7	49	2·8	10	·7	—	—	—	—
Totals	1829	—	1737	—	1455	—	1090	—	784	—
Average Age at Death*	35·9		33·7		26·9		16·2		7·7	

only those who died in their father's lifetime, then the percentage rises to 82·7 °/₀. Looking at the matter in another way we find that the average age at death has fallen from 35·9 years to 7·7 years.

The same facts are given in Fig. 6, which shows that as the selection of cases becomes more stringent, there is a regular increase in the proportion of deaths at the younger ages. In exactly the same way, the fact that cases where the insanity of parent and child is contemporaneous are the most likely to appear in Dr Mott's records causes a spurious exaggeration of the cases of insanity at early ages in the younger generation and consequently a spurious exaggeration of the number of cases of imbecility and adolescent insanity.

We can also investigate directly the question of anticipation or antedating on this material. In order to avoid the heavy weighting of large families which would arise if every child were entered, I have taken only one child from each family. Let us consider first of all the distribution of age at death of Fathers and their First-born Children. The facts are given in Table VII.

We have altogether 294 cases in which we know the age at death of a father and his first-born child. None of the fathers died before 20 but of the children

---

* These averages were calculated, not from the five age groups given above, but from the same material classified in 15 age groups.

106 out of 294 or 36·1 °/₀ died before 20. The average age at death among the fathers is 61 years, but among the children it is only 36 years, so that there is an anticipation of 25 years. To borrow Dr Mott's words, the figures clearly show the signal tendency among the offspring to die at a much earlier age than their parents; that is to say, anticipation or antedating is the rule.

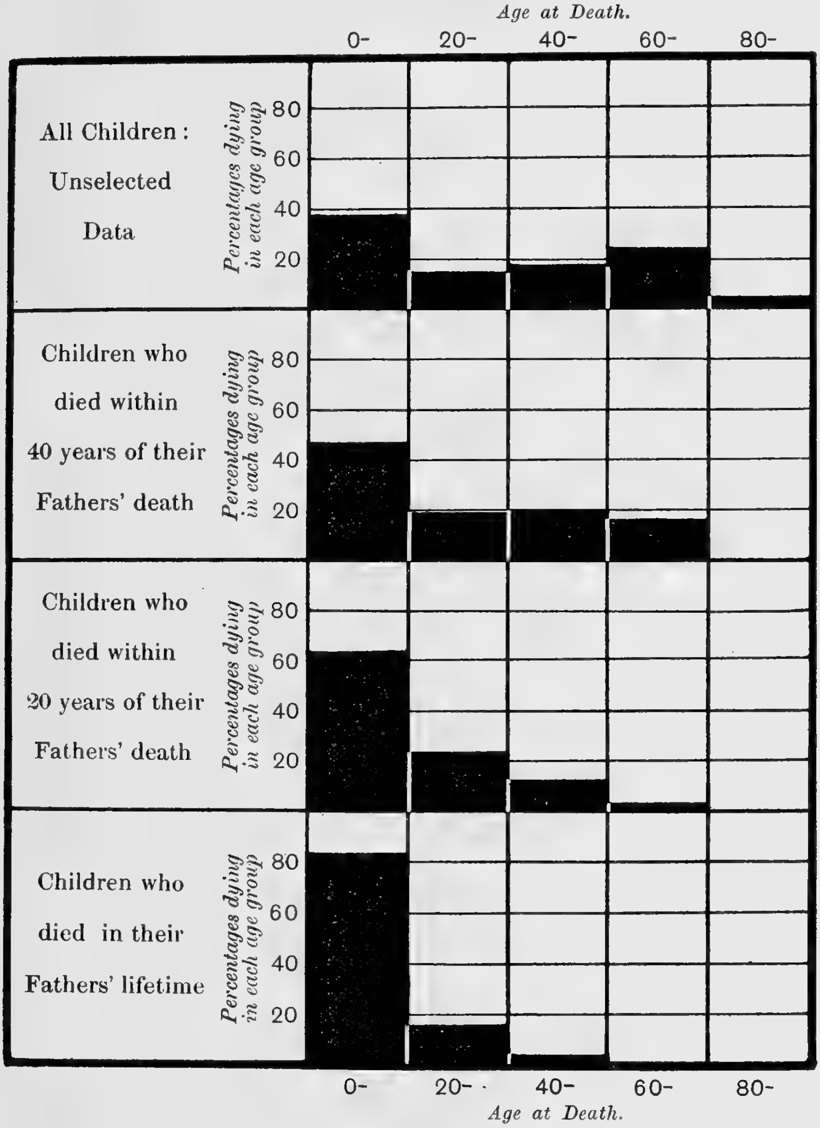


Fig. 6. Diagram to illustrate the Effect of Selection of Material upon the Distribution of Age at Death. (Reigning Houses in Europe, 18th Century.)

Now in this material there is no selection of families. Every family was taken and the age at death of every first-born is known, so that we are only left with the

TABLE VII.

Showing Anticipation in Age at Death. A. Fathers and Children.

(Reigning Houses in Europe—18th Century.)

Age at Death	Fathers	First-born Children	Fathers	First Sons who had children
0— 9	—	95	—	—
10—19	—	11	—	—
20—29	6	21	4	8
30—39	16	18	8	15
40—49	45	31	31	39
50—59	70	34	54	39
60—69	77	37	58	44
70—79	62	33	46	54
80—89	18	14	12	13
90 and over	—	—	—	1
Totals	294	294	213	213
Percentage dying under *20*	0	36·1	0	0
Average Age at Death ...	61	36	60	59
Anticipation	25		1	

third of Dr Mott's fallacies, in that no allowance has been made for the fact that the parental group is limited to ages over 20 while more than a third of the offspring die under 20. The effect of this selection can be removed almost entirely by taking instead of the first-born child, the first son who married and had at least one child. There are in all 213 such cases and we see that there is now no anticipation. The difference between the average ages at death is less than a year and by removing the artificial selection we have got rid of all anticipation or antedating.

These facts are also shown graphically in Figs. 7 and 8. The horizontal scale gives the age at death in 10-year groups while the vertical scale gives the actual numbers of parents and offspring dying in each age group. The diagram on the left shows marked anticipation, and should be compared with Dr Mott's diagram (Fig. 1) in which the ages at onset of insanity of father and child are compared. When, however, we get rid of the selection of cases by taking only sons who have had children, then there is no anticipation.

If we compare the distributions of age at death in mothers and children we get exactly the same results. The facts are shown in Table VIII.

We see that the first-born children died on an average 18 years before their mothers, but when we compare the age at death of mothers and the first son in

REIGNING HOUSES IN EUROPE — 18TH CENTURY.

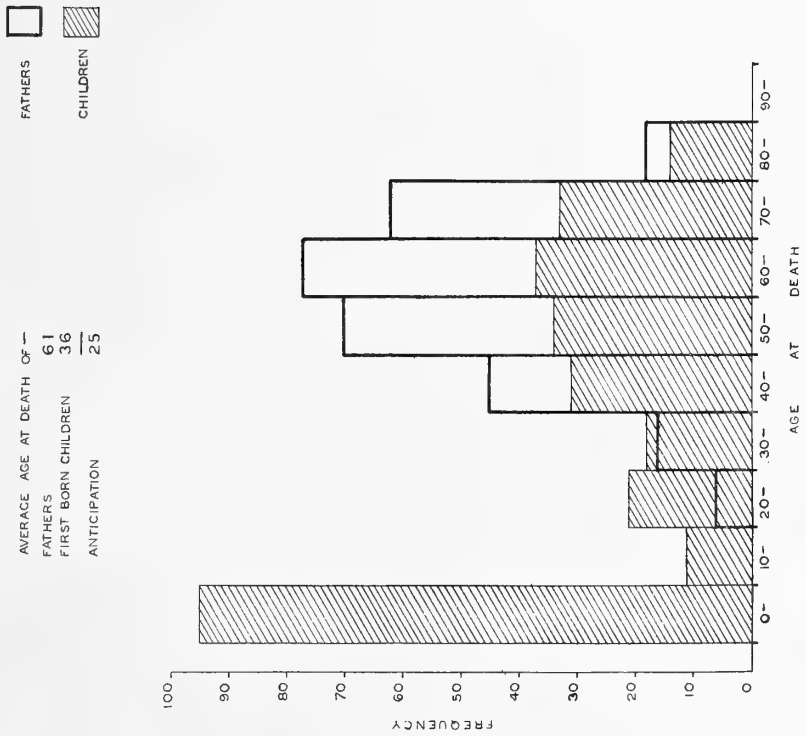


Fig. 7.

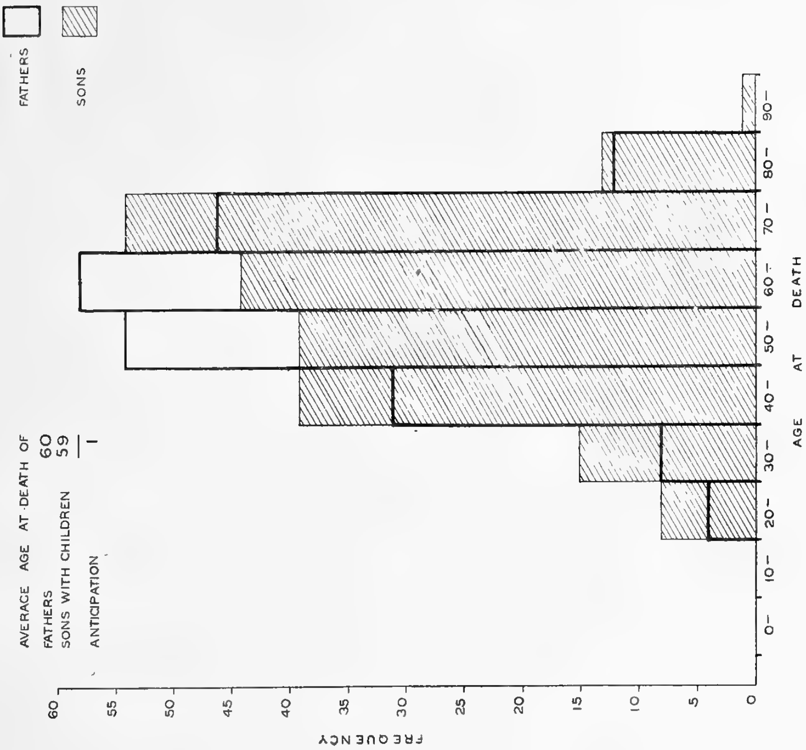


Fig. 8.

TABLE VIII.

Showing Anticipation in Age at Death. B. Mothers and Children.

(Reigning Houses in Europe—18th Century.)

Age at Death	Mothers	First-born Children	Mothers	First Sons to have Children
0—9	—	122	—	—
10—19	2	13	1	—
20—29	47	26	21	8
30—39	49	22	30	16
40—49	43	32	25	41
50—59	52	35	39	40
60—69	80	42	52	46
70—79	52	36	41	54
80—89	18	17	10	14
90 and over	2	—	1	1
Totals	345	345	220	220
Percentage dying under *20*	·6	39·1	·5	0
Average Age at Death ...	53	35	55	59
Anticipation	18		−4	

each case to have children, then the sons live four years longer than their mothers. It would have been better in this case to have compared the mothers with the first daughters to have children but unfortunately von Behr gives very little information regarding the female lives, except in special cases. The figures show a marked anticipation in age at death when we directly compare, as Dr Mott has done, mother and child, but this vanishes when we remove the arbitrary selection. The same facts are shown graphically in Figs. 9 and 10.

If we combine these figures we can compare the age at death of parent and child and the results are shown graphically in Figs. 11 and 12.

Fig. 11 shows that Dr Mott's limitation of one of the two generations he is comparing to adults, without imposing a similar limitation on the other generation, introduces an artificial and spurious anticipation. The average age at death of the parents is 56 years and of their first-born children only 35 years—so that we get an anticipation of 21 years. If, however, we make the two generations almost directly comparable by dealing only with sons who have children—there is no significant difference between the two averages (58 against 59 years).

In these cases we have dealt only with completed families and have taken every family without selection. If, however, we consider only the cases in which

REIGNING HOUSES IN EUROPE — 18TH CENTURY.

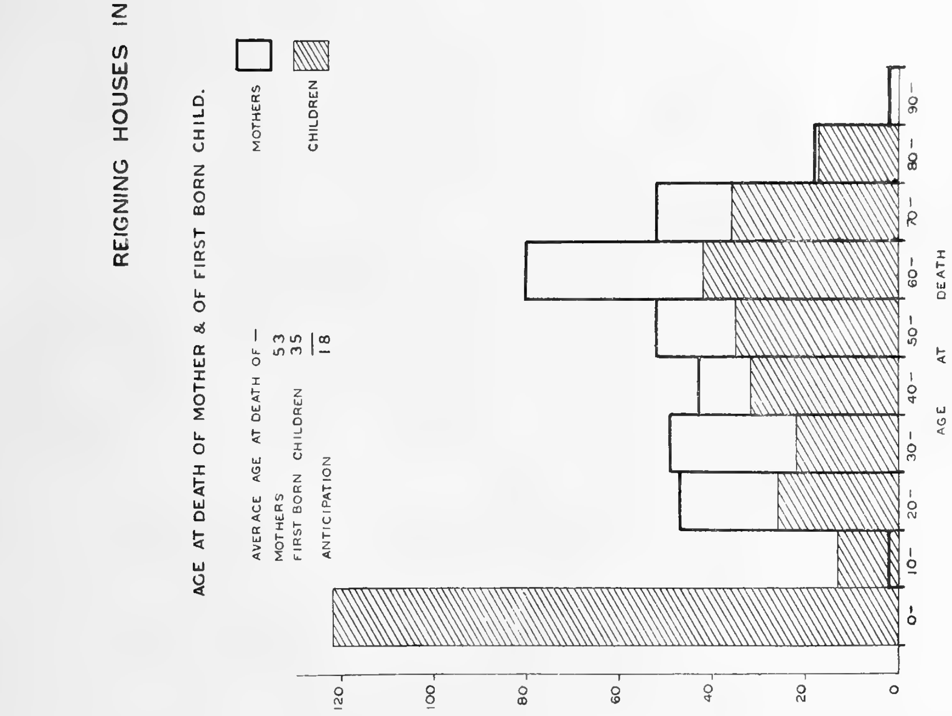

AGE AT DEATH OF MOTHER & OF FIRST BORN CHILD.

Fig. 9.

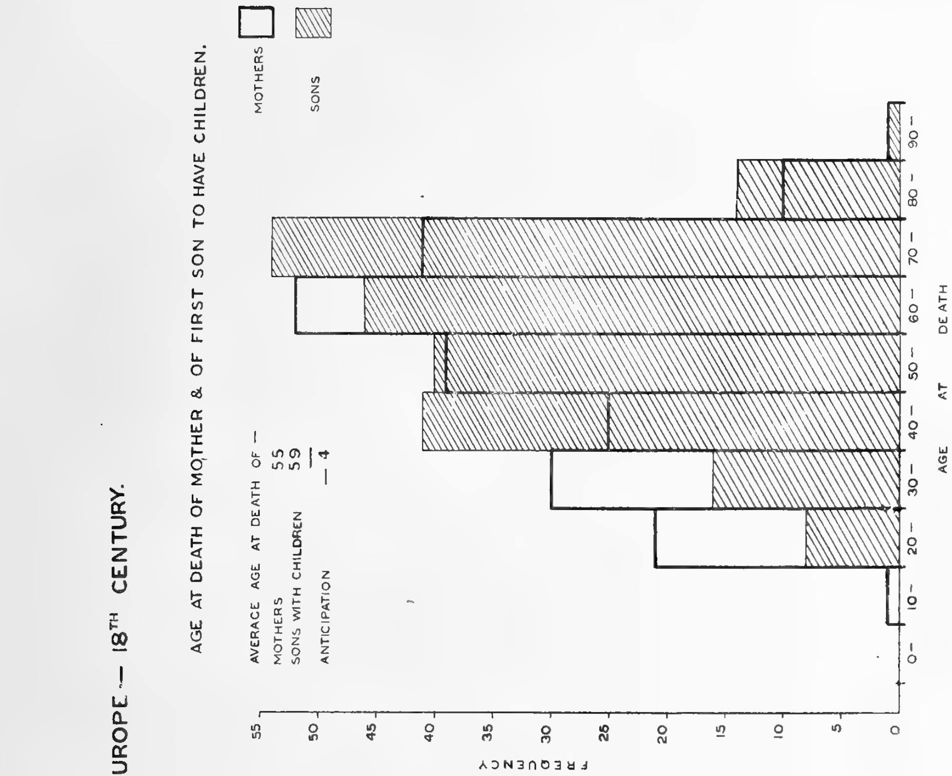

AGE AT DEATH OF MOTHER & OF FIRST SON TO HAVE CHILDREN.

Fig. 10.

REIGNING HOUSES IN EUROPE — 18TH CENTURY.

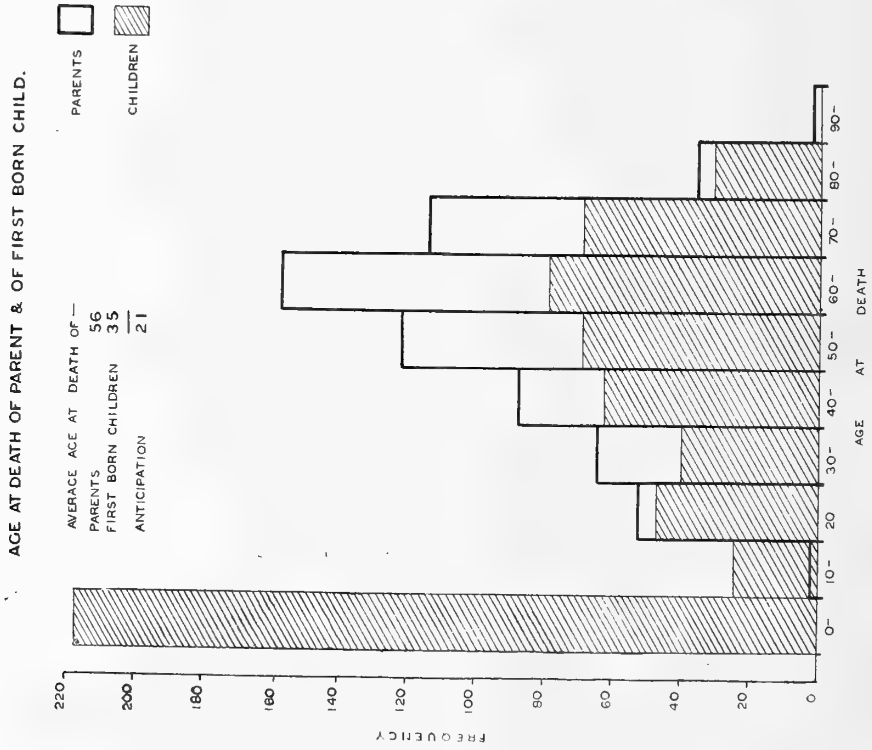


Fig. 11.

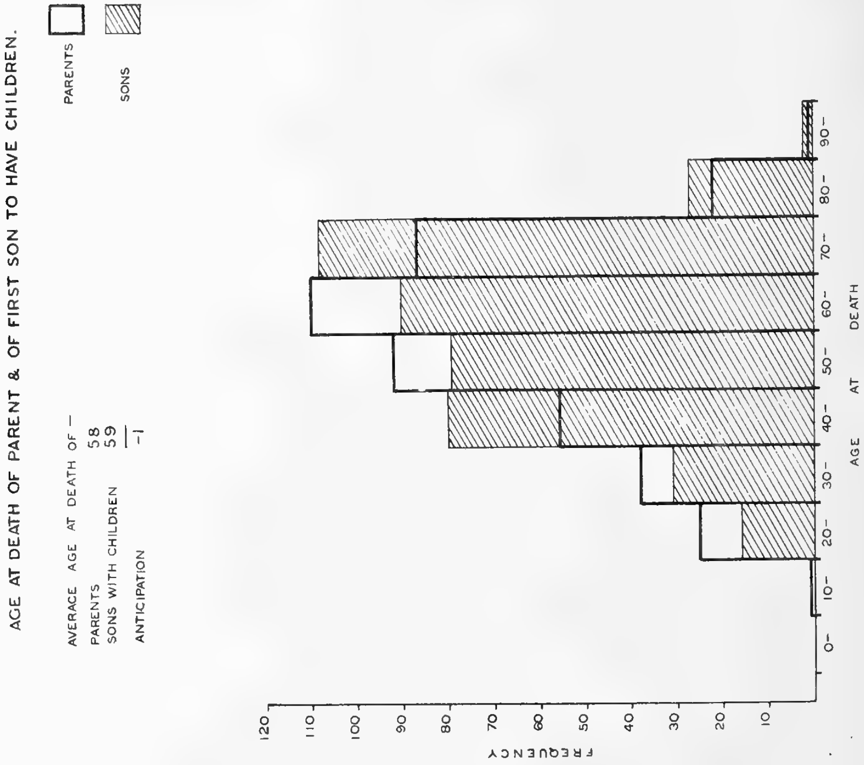


Fig. 12.

the eldest child died in his father's lifetime the amount of anticipation is greatly increased. The facts are shown in Table IX and in Fig. 13.

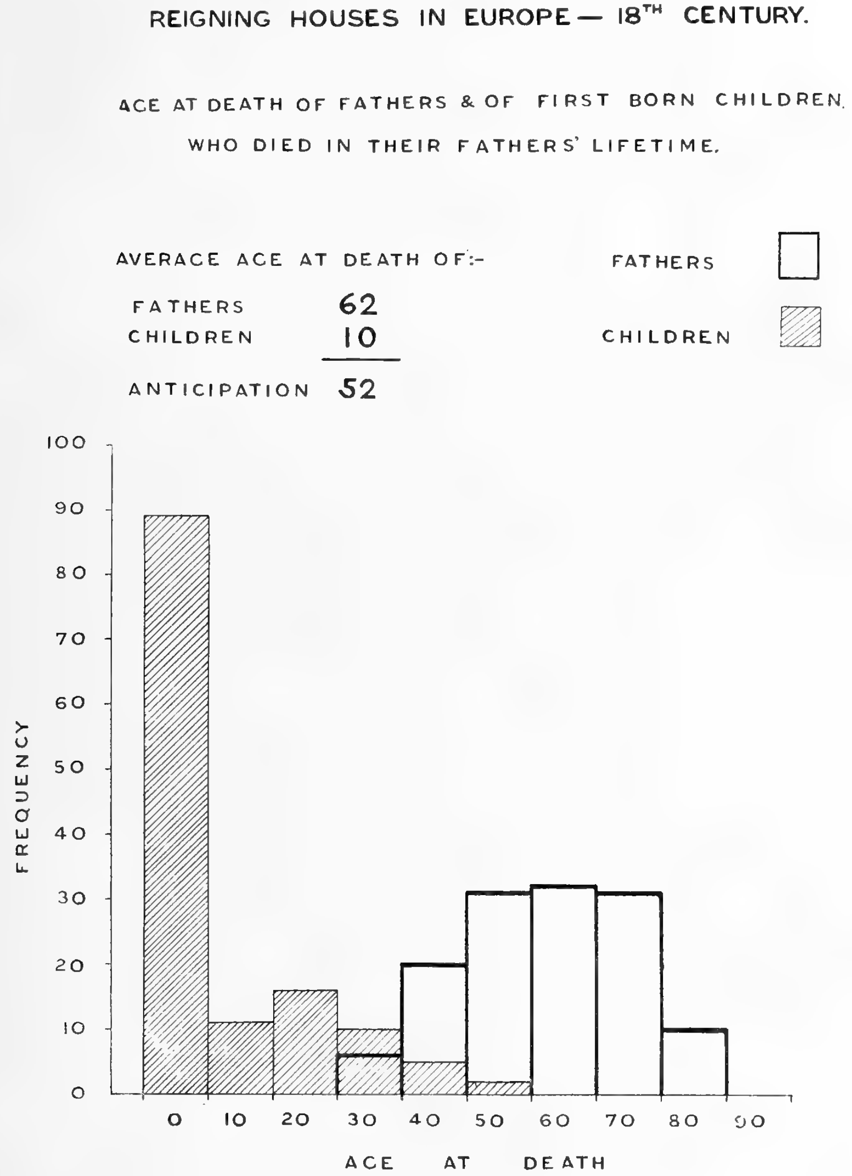


Fig. 13.

We see here that among the fathers none died under 30 while 87 °/。 of their children died under 30; the average age at death among the fathers was 62—among the children only 10, showing an anticipation of 52 years.

TABLE IX.

Showing Anticipation in Age at Death. C. Fathers and First-born Children who died in their Fathers' Lifetime.

(Reigning Houses of Europe: 18th Century.)

Age at Death	Fathers	First-born Children dying in their Fathers' lifetime
0— 9	—	89
10—19	—	11
20—29	—	16
30—39	6	10
40—49	20	5
50—59	32	2
60—69	33	—
70—79	32	—
80—89	10	—
Totals	133	133
Percentage dying under *20*	0	75·2
Average Age at Death ...	62	10
Anticipation	52	

It is now possible to illustrate the effect of the principal fallacies which vitiate Dr Mott's conclusions. In the first place he has dealt with families which are largely incomplete and has collected his material in such a way that cases in which the insanity of parent and child is contemporaneous are the most likely to be recorded; in the second place he has directly compared parent and child without allowing for the fact that practically no *parent* can become insane before 20, while there is no limitation of this kind among the offspring of these insane parents.

In Table IX and Fig. 13 we see the effect of dealing with incomplete families in which the children died in their fathers' lifetime. There we get an anticipation of 52 years. If we get rid of the first and second fallacies involving a selection of cases by dealing with every family, as shown in Table VII and Fig. 7, the anticipation falls to 26 years. If we get rid of the third source of fallacy also, by comparing the fathers with the first sons who have children, as in Table VII and Fig. 8, then the anticipation falls to less than a year. The Law of Anticipation or Antedating has thus in Dr Mott's case no foundation, in fact it is a spurious result of the mode of collecting and interpreting data.

Now Dr Mott has not only asserted that this "Law" applies to insanity but has also drawn the conclusion that the offspring of insane parents if still normal at the age of 25 may safely marry. In an address delivered before the First International Eugenics Congress*, he said: "You will observe that 47·8 °/₀ of the 500 offspring had their first attack (of insanity) at or before the age of 25 years and as you see in the curves of parents and offspring, the liability of the child of an insane parent becoming insane tends rapidly to fall. Now besides the fact that this shows Nature's method of eliminating unsound elements of a stock, it has another important bearing, for it shows that after twenty-five there is a greatly decreasing liability of the offspring of insane parents to become insane and therefore in the question of advising marriage of the offspring of an insane parent this is of great importance. Sir George Savage recently said that this statistical proof [*sic* !] of mine entirely accorded with his own experiences, and that if an individual who had such an hereditary history had passed twenty-five and never previously shown any signs (of insanity) he would probably be free and he would offer no objection to marriage."

Now I entirely fail to understand how anyone could recommend marriage in such cases, even on Dr Mott's own figures; for if it be true that 48 °/₀ become insane before 25, it must be equally true that 52 °/₀ become insane after that age and this very important point seems to have been forgotten. These figures, however, are taken from Dr Mott's selected data, selected in such a way that the early cases are enormously exaggerated. Until Dr Mott publishes a series of *complete* pedigrees, it will be safer to assume that the age at onset of insanity among the offspring of insane parents does not differ widely from that of all admissions to Asylums and there we find that only 21 °/₀ become insane before 25, and 79 °/₀ after 25.

But surely at a Eugenics Congress of all places some thought might have been given to the mental condition of the children resulting from such matings, before advising marriage. It would not have been difficult for Dr Mott to have extracted all the available cases of this kind from his collection of pedigrees, i.e. all cases in which an individual had an insane parent and was normal at the age of 25, and so have discovered the probable fate of the offspring from such matings.

Unfortunately the details given by Dr Mott regarding his pedigrees are usually so scanty that little use of them can be made, but two at least show the danger of the matings Sir George Savage and he sanction; these two pedigrees were given by Dr Mott in his lecture on *Heredity in Relation to Insanity*, delivered to the members of the London County Council. The first is shown in Fig. 14. (It appeared as Fig. 11, p. 18 of Dr Mott's lecture.) In the first generation a man who became insane at 70 had four children. The eldest, a girl, became insane at 68 and was therefore normal long after the age of 25. Dr Mott does not state whether the marriage of this woman preceded or followed the onset of insanity in her father, but even if her father had become insane before her marriage, Dr Mott

---

* *Problems in Eugenics*, p. 425. This is one of many illustrations of the evil done by that Congress; attention was directed and much weight given to hasty statements and ill-digested material.

would have raised no objection to the marriage since the woman herself was not insane. There were in all six children from this marriage of which Dr Mott would have approved. Two became insane, three were blind and five are said to have been paupers.

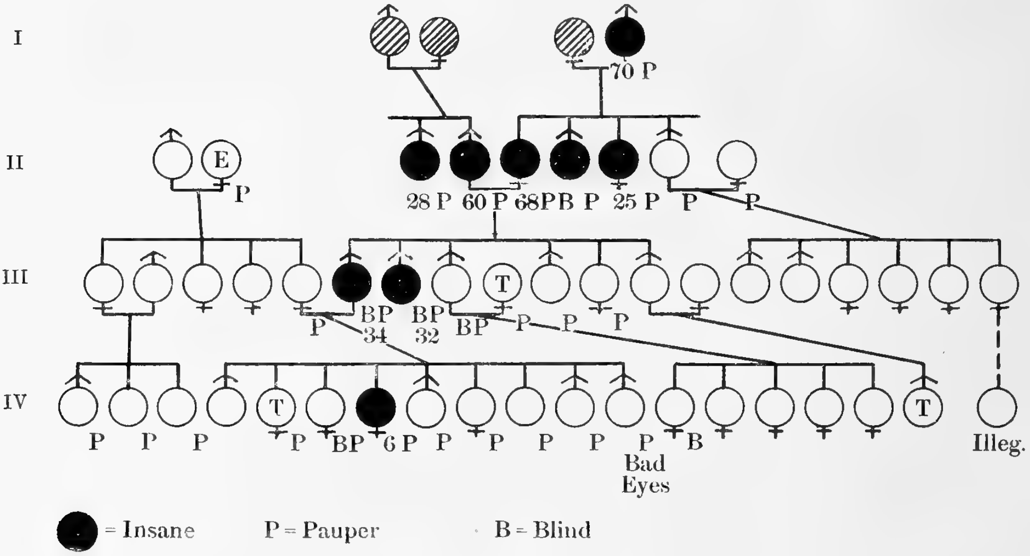


Fig. 14.

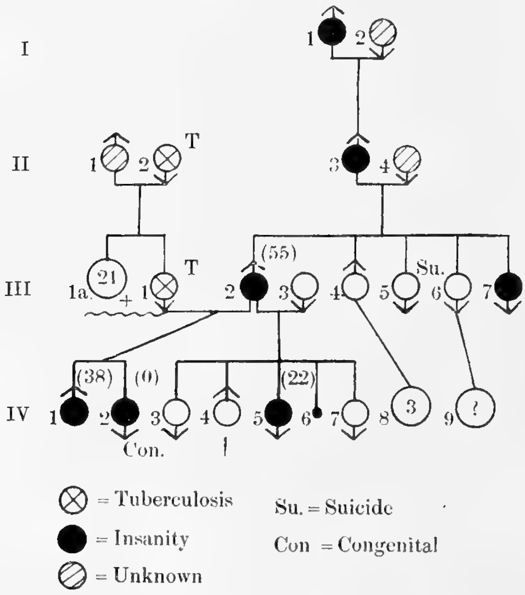


Fig. 15.

The eldest child remained normal till the age of 34 and although both his parents became insane Dr Mott apparently would not have objected to his marrying. He did so and one of his children became insane and eight out of nine are said to be paupers. These nine children are apparently still young so that their ultimate fate is still uncertain.

The second pedigree I shall quote was given as Fig. 28, p. 33 of Dr Mott's lecture, and appears here as Fig. 15.

A man who had an insane father and an insane grandfather became insane at the age of 55. He was therefore normal at the age of 25* and Dr Mott would have sanctioned marriage in his case. He actually married twice. His first wife was tuberculous but not insane; they had two children, both insane. His second wife was normal and it is definitely stated that there was no insanity in her family; they had five children and one of these became insane. Yet Dr Mott would permit the children of insane parents to marry if only they are normal at the age of 25!

Again, Dr Mott has stated that it is useless to attempt to limit the fertility of the insane since most of their children are born before the onset of insanity, and therefore before any action can be taken. From his statistics of relatives in L.C.C. Asylums, Dr Mott has calculated the proportion of offspring who were born after the first attack of insanity in the parent and found that "Forty-six offspring out of 581 were born after the first attack of insanity in the parent, i.e., 7·9 °/₀. That is to say in the case of 529 insane parents, *the birth of only one-twelfth of their* 581 *insane children would have been prevented by sterilisation or life segregation of the parent after the first attack of insanity.* These figures refer to the offspring which become insane, *but there are a large number of offspring which do not become insane and these would be cut off if life segregation or sterilisation were adopted*†."

But here again Dr Mott is using the data obtained from his index of relatives which shows a greatly exaggerated number of cases at the earlier ages among the offspring, and he thus greatly exaggerates the number of cases in which the children were born *before* the onset of insanity. No conclusion can be drawn from any but complete records of families. But apart altogether from this, many of these parents are themselves the children of the insane and much could be done to discourage such marriages. Unfortunately as we have seen Dr Mott directly sanctions marriage to those who remain normal till the age of 25.

In further support of his view Dr Mott has stated that out of 642 females admitted to three London County Asylums in 1911, 148 were recurrent cases and of these 32 (21 °/₀) had children between their respective dates of admission. "The inference that can be drawn," he says, "is that about one-fifth of the recurrent cases, or approximately one-twentieth of the female admissions have

---

* If the term "age at onset" has any real meaning.

† The italics are Dr Mott's.

children after their first attack of insanity and of 31 such cases examined, 73 children were born after the first attack of insanity in the parent."

But have these 148 recurrent cases been followed up to the end of the reproductive period? Not at all. No ages are given and the cases are merely those which were admitted to Asylums in 1911, Dr Mott's remarks being made in June 1912, so that no attempt has been made to follow them up. There is no justification for Dr Mott's advice.

There are many other points in Dr Mott's work which deserve detailed examination, but time will not permit more than a brief account of a few of them.

It should be noted, for instance, that Dr Mott has used his index of relatives in London County Asylums as an argument in favour of the importance of the inheritance factor in insanity. His argument is as follows:

"At the present time in the London County Asylums there are 725 individuals so closely related as parents and offspring, brothers and sisters. *A priori*, this, to my mind, is striking proof of the importance of heredity in relation to insanity, for we cannot suppose that 20,000 of the 4½ millions of people in London brought together from some random cause would show such a large number closely related as 3·6 °/₀."

But Dr Mott has not attempted to give, and I doubt if he ever will be able to give, a satisfactory estimate of the number of relatives in even a random sample of the population, and the population of asylums is far from being a random sample of the general population—there is for instance an extraordinary divergence in age. Yet without definite information on this point it would be impossible to say whether insanity is inherited or not—that is if we had to depend solely on Dr Mott's data.

It should also be noted that in these cases Dr Mott has clubbed together every form of insanity, from congenital idiocy to senile dementia, except of course cases due to specific infections or trauma. I myself think that course is the only possible one. To anyone who has studied even a few pedigrees of mental defect, nothing is more striking than the extraordinary number of different forms of mental defect that may appear in the same family.

Seven years ago, in a *First Study of the Statistics of Insanity and of the Inheritance of the Insane Diathesis**, I was confronted with the same problem, and after a full consideration of all the available data and of the opinions of those medical men who were best qualified to express an opinion came to the conclusion that the only possible course was to group all forms of insanity together, with, of course, the exceptions I have already indicated. The whole question was discussed very fully in my paper and it was there suggested that an even broader classification might be of service. This point of view met with some criticism at the time but nothing has occurred to alter it, and the study of the inheritance of

---

* *Galton Memoirs*, No. II. (Dulau and Co.)

insanity in general or of an even broader degeneracy must always remain the first object of our studies.

Any investigation of the inheritance of special types of insanity or degeneracy can only be carried out however on unselected material—on the records of complete families. The type of insanity is so closely related to the age of onset that any tendency to exaggerate the number of early cases, as in Dr Mott's material, will entirely vitiate the conclusions drawn. Thus Dr Schuster's conclusions as to the inheritance of special types of insanity based upon Dr Mott's data* must also be rejected on the above grounds.

Dr Mott's index of relatives in London County Asylums is unfortunately of very little value in the study of inheritance in insanity. Progress can only come from the study of complete pedigrees in which every member of the family is entered, whether insane or normal, and the ages of the normal at the time the record was made are just as important as the age at onset of insanity in the insane members, for a statement that a young man of 20 has not been insane is of a very different degree of importance from the statement that a man of 70 has not been insane.

In the papers I have cited the children of the insane if normal at 25 are advised to marry, and it is asserted that it is useless to attempt to discourage the reproduction of the insane since most of their children are born before the onset of insanity, and that we should rely on the Law of Anticipation to end or mend degenerate stocks.

I have shown, I think, that the Law of Anticipation as applied to the insane has no foundation in the facts provided and that the advice given as to the marriage of the insane and of their normal offspring is fundamentally unsound and directly cacogenic. Much yet remains to be learnt regarding the inheritance of the insane diathesis, but no one who has studied the family histories of the insane can doubt that in inheritance we have by far the most important element in the production of insanity, and in view of all the facts it is the obvious duty of the Eugenist to discourage, rather than to encourage, procreation by the insane and even by those of their offspring who appear to be normal.

---

* *Report on the Statistical Investigation of Relative Cards*, 21st Annual Report of the London County Council Asylums Committee (1910), p. 95.

ON THE PROBABLE ERROR OF THE BI-SERIAL EXPRESSION FOR THE CORRELATION COEFFICIENT.

By H. E. SOPER, M.A. Biometric Laboratory, University of London.

In a recent paper* Professor Pearson shows that where one character is in multiple graded grouping and the other in alternative categories, greater or less than a given magnitude, the correlation coefficient admits of simple expression; the assumptions being that the unmeasured character, B, has a normal distribution and that the measured character, A, has linear regression upon B. Under these conditions the data required are the numerical ratio of the alternative groups, the standard deviation of the measured character and the deviation from the general mean of this character of the mean of one of the groups.

This expression is subject to greater fluctuations of value in samples of N of the population than is the product moment form, especially where one of the groups is relatively small; and it is proposed to find formulae for the mean and second moment of the errors from this mean to a first approximation, that is to terms in $1/N$. These will appear in terms of the correlation coefficient, r, of the original population (which will be supposed *normally* correlated) and the fractional frequency, f, in that population of the group possessing the greater or positive intensity of the character put into two classes.

Let y be the graded character and x the alternative character the intenser value of which is possessed by the fraction f of the population. Let $\bar{x}, \bar{y}$ be the general means and σ_x, σ_y the standard deviations of x and y. Then it is shown in the paper that if $\bar{x}', \bar{y}'$ are the means of the group f,

$$r = \frac{\bar{y}' - \bar{y}}{\sigma_y} \times \frac{\sigma_x}{\bar{x}' - \bar{x}} \qquad \text{.................................(1)},$$

on the assumption that the regression of y upon x is linear; and that if x be normally distributed this is equivalent to

$$r = \frac{\bar{y}' - \bar{y}}{\sigma_y} \times \frac{f}{z} \qquad \text{.....................................(2)},$$

---

* *Biometrika*, Vol. VII. (1909), p. 96: "On a New Method of Determining Correlation between a Measured Character A, and a Character B, of which only the Percentage of Cases wherein B exceeds (or falls short of) a given intensity is recorded for each grade of A."

where z is the ordinate of the normal curve cutting off the area

$$f = \tfrac{1}{2}(1-\alpha) \quad \text{or} \quad \tfrac{1}{2}(1+\alpha)$$

as defined in Sheppard's Tables of the Probability Integral.

Now if $$p_0, \quad p_1, \quad p_2, \text{ etc.}$$

be the moment coefficients of the whole population with respect to the character y defined by

$$p_v = S(n_s y_s^v)/N \quad \text{.................................(3)},$$

[see bi-serial table below which is here to be looked upon as representing the general population] and

$$p_0', \quad p_1', \quad p_2', \text{ etc.}$$

are moment coefficients of the group $n' (=fN)$ defined by

$$p_v' = S(n_s' y_s^v)/N \quad \text{.................................(4)},$$

we may write $f = p_0'$, $f\bar{y}' = p_1'$, $\bar{y} = p_1$, $\sigma_y = \sqrt{(p_2 - p_1^2)}$ and

$$r = \frac{p_1' - p_0' p_1}{\sqrt{(p_2 - p_1^2)}} \times \frac{1}{z} \quad \text{..............................(5)}.$$

→ Grades of y in Bi-serial Table.

↓ Grades of x.

n_1''	n_2''	n_3''	...	n_s''	...	n''
n_1'	n_2'	n_3'	...	n_s'	...	n'
n_1	n_2	n_3	...	n_s	...	N

$n'/N = f$ ....(6).

In samples of N the frequencies n_s, n_s', n_s'' and consequently the moment coefficients p and p' and the ordinate z are subject to fluctuation and the values of the correlation coefficient calculated from this formula will have a distribution of errors. Let $\bar{r}$ be the mean value in such samples, δr the deviation from this mean value in any sample when δp_0, δp_1, $\delta p_0'$, etc. δz are the deviations in moments and ordinate. Then

$$\bar{r} + \delta r = \frac{p_1' + \delta p_1' - (p_0' + \delta p_0')(p_1 + \delta p_1)}{\sqrt{\{p_2 + \delta p_2 - (p_1 + \delta p_1)^2\}} \times (z + \delta z)} \quad \text{..................(7)}.$$

To express δz in terms of deviations of the moments we have

$$z = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2} \quad \text{.................................(8)},$$

where a is the abscissa of the point of section of the normal curve, defined by

$$f = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \quad \text{................. ....(9)}.$$

Hence to second order terms in powers and products of deviations

$$
\begin{aligned}
z + \delta z &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(a+\delta a)^2} \\
&= z\left\{1 - a\,\delta a - \tfrac{1}{2}(1-a^2)(\delta a)^2\right\} \\
-\delta f &= \int_a^{a+\delta a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx \\
&= \int_0^{\delta a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(a+\xi)^2}\,d\xi \\
&= z\int_0^{\delta a} \left\{1 - a\xi - \tfrac{1}{2}(1-a^2)\,\xi^2 + \ldots\right\} d\xi \\
&= z\left\{\delta a - \tfrac{1}{2}a\,(\delta a)^2\right\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
z + \delta z &= z + a\,\delta f - \frac{1}{2z}(\delta f)^2 \\
&= z + a\,\delta p_0' - \frac{1}{2z}(\delta p_0')^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(10).
\end{aligned}
$$

At the same time that this value is put in (7) we may simplify the expression and subsequent algebra by supposing the graded character y to be measured from its universal mean value as origin and in terms of its standard deviation as unit of measurement in which case

$$p_1 = 0, \quad p_2 = \sigma_y^2 = 1, \text{ and by (5) } p_1' = zr \quad \ldots\ldots\ldots\ldots\ldots(11),$$

and since $p_0' = f$ (7) becomes,

$$\bar{r} + \delta r = \frac{zr + \delta p_1' - f\delta p_1 - \delta p_0'\,\delta p_1}{\sqrt{\{1 + \delta p_2 - (\delta p_1)^2\}} \times \left\{z + a\,\delta p_0' - \frac{1}{2z}(\delta p_0')^2\right\}} \quad \ldots\ldots\ldots\ldots(12).$$

Expanding to second order of deviations we find,

$$\bar{r} + \delta r = r + \delta_1 + \delta_2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(13),$$

where δ_1, δ_2 are the first order and second order expressions

$$
\begin{aligned}
\delta_1 &= \frac{1}{z}\left\{\delta p_1' - f\delta p_1 - \tfrac{1}{2}zr\,\delta p_2 - ar\,\delta p_0'\right\}, \\
\delta_2 &= \frac{1}{z}\Big\{\tfrac{1}{2}ar\,\delta p_2\,\delta p_0' - \frac{a}{z}\,\delta p_1'\,\delta p_0' + \frac{fa}{z}\,\delta p_1\,\delta p_0' - \tfrac{1}{2}\delta p_1'\,\delta p_2 + \tfrac{1}{2}f\delta p_1\,\delta p_2 \\
&\qquad - \delta p_0'\,\delta p_1 + \tfrac{1}{2}zr\,(\delta p_1)^2 + \frac{3}{8}zr\,(\delta p_2)^2 + (1+2a^2)\,\frac{r}{2z}(\delta p_0')^2\Big\} \quad \ldots\ldots(14).
\end{aligned}
$$

Taking mean values

$$\bar{r} = r + \text{mean } \delta_2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(15),$$

mean δ_1 being zero since by (3), (4)

$$
\begin{aligned}
\text{mean } \delta p_v\ &= S(\text{mean } \delta n_s y_s^v)/N = 0, \\
\text{mean } \delta p_v' &= S(\text{mean } \delta n_s' y_s^v)/N = 0.
\end{aligned}
$$

Mean δ_2 is to be evaluated by the formulae

$$\left.\begin{aligned} \text{mean } \delta p_u \delta p_v &= (p_{u+v} - p_u p_v)/N \\ \text{mean } \delta p_u' \delta p_v' &= (p'_{u+v} - p_u' p_v')/N \\ \text{mean } \delta p_u \delta p_v' &= (p'_{u+v} - p_u p_v')/N \end{aligned}\right\} \quad \dots\dots\dots\dots\dots\dots(16),$$

of which the first two are well known* and the third may be proved thus:

$$N \delta p_u = S(\delta n_s y_s^u) = S(\delta n_s' y_s^u) + S(\delta n_s'' y_s^u),$$
$$N \delta p_v' = S(\delta n_s' y_s^v),$$
$$\therefore\ N^2 \delta p_u \delta p_v' = S\{(\delta n_s')^2 y_s^{u+v}\} + S\{\delta n_s' \delta n_{s'}' y_s^u y_{s'}^v\} + S\{\delta n_s' \delta n_{s'}'' y_s^v y_{s'}^u\},$$

where in the third sum s' may or may not equal s.

But
$$\text{mean } (\delta n_s')^2 = n_s'(1 - n_s'/N),$$
$$\text{mean } (\delta n_s' \delta n_{s'}') = -n_s' n_{s'}'/N,$$
$$\text{mean } (\delta n_s' \delta n_{s'}'') = -n_s' n_s''/N,$$

the last whether $s = s'$ or not. Thus we find†

$$\text{mean } \delta p_u \delta p_v' = (p'_{u+v} - p_u' p_v' - p_u'' p_v')/N = (p'_{u+v} - p_u p_v')/N.$$

Evaluating mean δ_2 by these formulae we find,

$$\begin{aligned} \bar{r} = r + \frac{1}{Nz} \Big\{ & \tfrac{1}{2} ar (p_2' - p_2 p_0') - \frac{a}{z}(p_1' - p_1' p_0') \\ & + \frac{fa}{z}(p_1' - p_1 p_0') - \tfrac{1}{2}(p_3' - p_1' p_2) + \tfrac{1}{2} f (p_3 - p_1 p_2) \\ & - (p_1' - p_0' p_1) + \tfrac{1}{2} zr (p_2 - p_1^2) + \frac{3}{8} zr (p_4 - p_2^2) + (1 + 2a^2) \frac{r}{2z}(p_0' - p_0'^2) \Big\} \ \dots(17), \end{aligned}$$

in which the undashed moments, being those of a normal curve with unit standard deviation, about its mean, have the values

$$p_1 = 0, \quad p_2 = 1, \quad p_3 = 0, \quad p_4 = 3,$$

and the dashed moments beyond the first two,

$$p_0' = f, \quad p_1' = zr,$$

have values depending upon the nature of the frequency distribution of y and x.

Assuming x, y normally distributed‡

$$p_2' = \int_a^\infty \int_{-\infty}^\infty y^2 \cdot \frac{1}{2\pi\sqrt{(1 - r^2)}} e^{-\frac{1}{2(1-r^2)}\{x^2 - 2rxy + y^2\}} dx dy,$$

$$p_3' = \int_a^\infty \int_{-\infty}^\infty y^3 \cdot \frac{1}{2\pi\sqrt{(1 - r^2)}} e^{-\frac{1}{2(1-r^2)}\{x^2 - 2rxy + y^2\}} dx dy.$$

---

* See *Biometrika*, Vol. II. (1903), p. 275: "On the Probable Errors of Frequency Constants." K. Pearson. The second follows in exactly the same manner as the first, since the constancy of the total frequency dealt with is only involved, in deducing the relations (i) and (ii) of p. 274, of that paper.

† $p_u'' = S(n_s'' y_s^u)/N$.

‡ Since the moments appear in the term containing $1/N$ any errors in their calculation due to incorrect assumption of normality will not affect the present approximate formulae provided such errors are of the order $1/N$.

Putting $y = rx + \eta$ and integrating with respect to η for constant x,

$$p_2' = \int_a^\infty \{(rx)^2 + (1 - r^2)\} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$= f + azr^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(18),$$

$$p_3' = \int_a^\infty \{(rx)^3 + 3(rx)(1 - r^2)\} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$= z(3 - r^2 + a^2r^2)r \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(19).$$

When these values are put in (17), and terms collected, the mean value of the bi-serial correlation coefficient in samples of N is found to be

$$\bar{r} = r\left\{1 + \frac{1}{N}\left[\tfrac{1}{4} + \tfrac{1}{2}\frac{ff'}{z^2} - \left(1 - \frac{fa}{z}\right)\left(1 + \frac{f'a}{z}\right) + \tfrac{1}{2}r^2\right]\right\} \quad \ldots\ldots\ldots(20),$$

where $f' = 1 - f = n''/N$.

In the work of obtaining this approximation all powers and products of deviations above the second order have been neglected. The means of such terms in samples of N involve second and higher powers of $1/N$* and the present result is correct to the first approximation.

Again squaring (13) and taking mean values and subtracting the square of (15) we find to the same approximation as before

$$\sigma_r^2 = \text{mean}\,(\delta r)^2 = \text{mean}\,\delta_1^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(21).$$

The evaluation of mean δ_1^2 being carried out precisely in the same way as mean δ_2, the result is the second moment of deviations of the bi-serial correlation coefficient in samples of N,

$$\sigma_r^2 = \frac{1}{N}\left[\frac{ff'}{z^2} - \left\{\tfrac{3}{2} + \left(1 - \frac{fa}{z}\right)\left(1 + \frac{f'a}{z}\right)\right\} r^2 + r^4\right] \quad \ldots\ldots\ldots\ldots(22).$$

Writing the two results (20) and (22)

$$\bar{r} = r\left\{1 + \frac{1}{N}(\phi_\alpha + \tfrac{1}{2}r^2)\right\},$$

$$\sigma_r^2 = \frac{1}{N}[\chi_\alpha^2 - \psi_\alpha r^2 + r^4] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(23),$$

the values of ϕ_α, χ_α^2† and ψ_α for values of $\frac{1}{2}(1 - \alpha)$ [= the smaller of n'/N, n''/N] from ·50 to ·01 are to be found in table (24).

---

* See *Biometrika*, Vol. IX. (1913), pp. 97—99.

† χ_α for $\frac{1}{2}(1 + \alpha)$ was tabled in *Biometrika*, Vol. IX. p. 27, and the table is reproduced in *Tables for Statisticians and Biometricians*, p. 35, Cambridge University Press.

$\frac{1}{2}(1-a)$	ϕ_a	χ_a^2	ψ_a	$\frac{1}{2}(1-a)$	ϕ_a	χ_a^2	ψ_a
·50	·0354	1·5708	2·5000	·20	− ·0871	2·0414	2·8578
·49	·0353	1·5711	2·5003	·19	− ·1001	2·0898	2·8951
·48	·0350	1·5722	2·5011	·18	− ·1146	2·1437	2·9364
·47	·0346	1·5741	2·5024	·17	− ·1308	2·2035	2·9825
·46	·0339	1·5766	2·5043	·16	− ·1490	2·2703	3·0341
·45	·0331	1·5799	2·5068	·15	− ·1696	2·3453	3·0923
·44	·0321	1·5839	2·5098	·14	− ·1931	2·4303	3·1582
·43	·0309	1·5886	2·5134	·13	− ·2201	2·5272	3·2337
·42	·0295	1·5943	2·5177	·12	− ·2513	2·6389	3·3208
·41	·0279	1·6007	2·5225	·11	− ·2880	2·7687	3·4224
·40	·0260	1·6079	2·5279	·10	− ·3317	2·9221	3·5427
·39	·0240	1·6161	2·5341	·095	− ·3568	3·0095	3·6115
·38	·0217	1·6251	2·5409	·090	− ·3844	3·1057	3·6873
·37	·0191	1·6351	2·5484	·085	− ·4153	3·2119	3·7713
·36	·0163	1·6461	2·5568	·080	− ·4499	3·3299	3·8648
·35	·0132	1·6582	2·5659	·075	− ·4887	3·4620	3·9696
·34	·0098	1·6714	2·5759	·070	− ·5324	3·6110	4·0879
·33	·0062	1·6858	2·5868	·065	− ·5822	3·7806	4·2224
·32	·0021	1·7015	2·5986	·060	− ·6403	3·9748	4·3777
·31	− ·0023	1·7186	2·6115	·055	− ·7083	4·2002	4·5584
·30	− ·0070	1·7371	2·6256	·050	− ·7897	4·4652	4·7723
·29	− ·0122	1·7573	2·6409	·045	− ·8868	4·7829	5·0283
·28	− ·0179	1·7791	2·6575	·040	− 1·0053	5·1715	5·3410
·27	− ·0241	1·8028	2·6755	·035	− 1·1577	5·6568	5·7362
·26	− ·0308	1·8286	2·6952	·030	− 1·3556	6·2859	6·2485
·25	− ·0382	1·8567	2·7166	·025	− 1·6308	7·1347	6·9481
·24	− ·0462	1·8874	2·7399	·020	− 2·0272	8·3600	7·9572
·23	− ·0550	1·9208	2·7654	·015	− 2·5263	10·3024	9·4275
·22	− ·0647	1·9574	2·7933	·010	− 3·8889	13·9393	12·6086
·21	− ·0753	1·9974	2·8240				

.... (24).

The bi-serial value of the correlation coefficient has the standard deviation

$$\sigma_r = \sqrt{(\chi_a^2 - \psi_a r^2 + r^4)}/\sqrt{N},$$

whilst that of the product moment value is

$$\sigma_r = (1 - r^2)/\sqrt{N}.$$

In table (25) a comparison of the values of the numerator is made for five values of r, for divisions at 0, ·5, ...2·5 times the standard deviation from the mean of the ungraded character.

r	Values of $(1-r^2)$	Values of $\sqrt{(\chi_a^2 - \psi_a r^2 + r^4)}$ for $\frac{1}{2}(1-a) =$					
		·500	·309	·159	·067	·023	·006
·00	1·00	1·25	1·31	1·51	1·93	2·76	4·5
·25	·9375	1·19	1·25	1·45	1·86	2·68	4·3
·50	·750	1·00	1·06	1·26	1·65	2·42	4·0
·75	·4375	·69	·75	·94	1·30	1·95	3·2
1·00	·00	·27	·33	·49	·74	1·13	1·8

...... (25).

Thus the effect of grouping and applying the bi-serial value of the correlation coefficient is to add 25 °/₀ to the probable error in the most accordant case where r is zero and the division equal, whilst if r is as large as ·5 and one group as small as 10 °/₀ of the whole the probable error is nearly doubled. For higher values of r the errors of sampling, in the case of the product moment formula, grow smaller and ultimately vanish when $r=1$; but the bi-serial values are not invariable in samples drawn from a perfectly correlated population but possess a variability as high as $\cdot 27/\sqrt{N}$ in the most favourable case when the grouping is equal.

If the standard deviation be calculated from the approximate formula,

$$\sigma_r = (\sqrt{ff'}/z - r^2)/\sqrt{N} \quad \text{..............................(26)},$$

which may be written*

$$\sigma_r = (\chi_a - r^2)/\sqrt{N} \quad \text{..............................(27)},$$

the error of computation will not be great for values of f and r commonly met with as the following table compared with the last will show:

r	Values of $\chi_a - r^2$ for $\frac{1}{2}(1-a)=$					
	·500	·309	·159	·067	·023	·006
·00	1·25	1·31	1·51	1·93	2·76	4·5
·25	1·19	1·25	1·45	1·86	2·70	4·4
·50	1·00	1·06	1·26	1·68	2·51	4·2
·75	·69	·75	·95	1·36	2·20	3·9
1·00	·25	·31	·51	·93	1·76	3·5

......(28).

The difference between the two expressions only reaches 5 °/₀ when the smaller group is less than 7 °/₀ of the whole.

It will not be necessary, excepting in small samples, to apply a correction to the bi-serial formula for r in virtue of the mean of samples differing from the population value. The correction is less than $1/N$th part of the value calculated unless one of the alternative classes is as small as 4 °/₀ of the whole.

I have to thank fellow members of the Staff for assistance in calculating the tables.

---

* For a table of χ_a see *Tables for Statisticians and Biometricians*, p. 37.

ON THE PARTIAL CORRELATION RATIO.

PART I. THEORETICAL.

By L. ISSERLIS, B.A.

§ 1. The theory of non-linear regression in the case of two correlated variables is due to Prof. Karl Pearson*. He shows that regression ceases to be linear when the correlation ratio η differs sensibly from the correlation coefficient r and establishes criteria for parabolic, cubic and higher forms of regression.

The present paper deals with the regression surface of three correlated variables x, y, z, where, though the regression of z on x, y cannot be adequately represented by an equation of type

$$\frac{\bar{z}_{xy}-\bar{z}}{\sigma_z} = \gamma_3 \frac{x-\bar{x}}{\sigma_x} + \gamma_3' \frac{y-\bar{y}}{\sigma_y} \quad \text{......................(1)},$$

the regression of z on x for a constant y and of z on y for a constant x is linear. A large proportion of the non-linear cases that occur in practice fall into this class. It will be remembered that $\bar{z}_{xy}$ in (1) denoting the mean of the array of z's for a given x and y the coefficients γ_3, γ_3' are partial regression coefficients and it will appear that just as it is necessary to introduce the correlation ratio η for an adequate description of non-linear regression of two variables, there must be introduced multiple or partial η's for the description of such regression in the case of more than two variables.

We recall the definition and principal properties of ${}_x\eta_y$—the correlation ratio of y on x, σ_{a_y} being the square root mean weighted square standard deviations of the arrays of y:

$$(1-\eta^2)\,\sigma_y^2 = \sigma^2_{a_y} = \frac{S(n_x \sigma^2_{n_x})}{N} = \frac{SS\{n_{xy}(y-\bar{y}_x)^2\}}{N} \quad \text{...........(2)},$$

$$\eta^2 = \frac{\sigma^2_{m_y}}{\sigma_y^2} = \frac{S\{n_x(\bar{y}_x-\bar{y})^2\}}{N\sigma_y^2} \quad \text{..........................(3)},$$

and

$$N(\eta^2 - r^2)\,\sigma_y^2 = S\{n_x(y_{n_x} - Y)^2\} \quad \text{....................(4)}.$$

---

* *Drapers' Company Research Memoirs. Mathematical Contributions to the Theory of Evolution.* XIV. "On the General Theory of Skew Correlation and Non-linear Regression." 1905.

Here we are dealing with N pairs of two characters A and B. n_x of these have the character x of A. $\bar{y}_x$ is the mean of this x-array of B's. σ_{n_x} is the standard deviation of this array, σ_{m_y} is the weighted standard deviation of the means of the arrays and Y is the value of y given by the regression straight line, i.e.

$$Y = \bar{y} + r\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \quad \text{..............................(5).}$$

This is the "best fitting" straight line (in the Gaussian sense) to the means of the arrays, and is the regression line when the regression is linear.

§ 2. Consider now three correlated characters A, B, C. If N combinations of A, B, C are taken, we may denote by n_x the number of these which have the character $A = x$ and by n_{xy} the number in which A has the value x and B has the value y. Let $\bar{x}$, $\bar{y}$, $\bar{z}$ be the mean values of the total population, and let $\bar{z}_{xy}$ be the mean of z for a given x and y. The frequency of $A = x$, $B = y$, $C = z$ is n_{xyz}.

We define the correlation ratio of z on x and y, which may be denoted by ${}_{xy}H_z$, or if no confusion is likely to arise by H_z, by the equation

$$\sigma_z^2\left[1 - {}_{xy}H_z^2\right] = \frac{SSS\{n_{xyz}(z - \bar{z}_{xy})^2\}}{N} \quad \text{....................(6).}$$

The triple sum in the definition can be written

$$\begin{aligned} &SSS\{n_{xyz}(z - \bar{z} + \bar{z} - \bar{z}_{xy})^2\} \\ &\quad = SS\{n_{xy}(\bar{z} - \bar{z}_{xy})^2\} + 2SS\{(\bar{z} - \bar{z}_{xy})\} \times S\{n_{xyz}(z - \bar{z})\} + SSS\{n_{xyz}(z - \bar{z})^2\} \\ &\quad = SS\{n_{xy}(\bar{z} - \bar{z}_{xy})^2\} - 2SS\{n_{xy}(\bar{z} - \bar{z}_{xy})^2\} + N\sigma_z^2. \end{aligned}$$

Hence

$${}_{xy}H_z^2 = \frac{SS\{n_{xy}(\bar{z} - \bar{z}_{xy})^2\}}{N\sigma_z^2} \quad \text{..........................(7)}$$

This is a generalisation of the property of ${}_x\eta_y$ given by (3).

Further, the "best fitting" plane to the means $\bar{z}_{xy}$ is given by

$$\frac{z - \bar{z}}{\sigma_z} = \gamma_3\frac{x - \bar{x}}{\sigma_x} + \gamma_3'\frac{y - \bar{y}}{\sigma_y} \quad \text{..........................(8),}$$

where

$$\gamma_3 = \frac{r_{xz} - r_{yz}r_{xy}}{1 - r^2_{xy}} \quad \text{..................................(9),}$$

$$\gamma_3' = \frac{r_{yz} - r_{xz}r_{xy}}{1 - r^2_{xy}} \quad \text{..................................(10).}$$

Let ${}_{xy}R_z$ denote as usual the maximum correlation of z with any linear function of x and y, then

$${}_{xy}R_z^2 = \gamma_3 r_{zx} + \gamma_3' r_{zy} \quad \text{....................................(11),}$$

$$= \frac{r^2_{yz} + r^2_{zx} - 2r_{yz}r_{xz}r_{xy}}{1 - r^2_{xy}} \quad \text{..........................(12).}$$

Subtract (11) from (7) after replacing r_{zx} and r_{zy} by the appropriate sums, and we obtain

$$[{}_{xy}H_z^2 - {}_{xy}R_z^2]\, N\sigma_z^2$$

$$= SS\left[n_{xy}(\bar{z} - \bar{z}_{xy})^2 - S\left\{n_{xyz}\frac{\sigma_z}{\sigma_x}(z - \bar{z})(x - \bar{x})\gamma_3\right\} - S\left\{n_{xyz}\frac{\sigma_z}{\sigma_y}(z - \bar{z})(y - \bar{y})\gamma_3'\right\}\right]$$

$$= SS\left[n_{xy}(\bar{z} - \bar{z}_{xy})^2 - n_{xy}\frac{\sigma_z}{\sigma_x}(\bar{z}_{xy} - \bar{z})(x - \bar{x})\gamma_3 - n_{xy}\frac{\sigma_z}{\sigma_y}(\bar{z}_{xy} - \bar{z})(y - \bar{y})\gamma_3'\right].$$

Using (8) this can be written

$$[{}_{xy}H_z^2 - {}_{xy}R_z^2]\, N\sigma_z^2 = SS\{n_{xy}(\bar{z}_{xy} - \bar{z})(\bar{z}_{xy} - Z)\}$$

$$= SS\{n_{xy}(\bar{z}_{xy} - Z)^2\} + SS\{n_{xy}(Z - \bar{z})(\bar{z}_{xy} - Z)\} \quad \ldots(13).$$

But $SS\{n_{xy}(Z - \bar{z})(\bar{z}_{xy} - Z)\}$

$$= SS\left\{n_{xy}\left(\gamma_3\frac{x - \bar{x}}{\sigma_x}\sigma_z + \gamma_3'\frac{y - \bar{y}}{\sigma_y}\sigma_z\right)\left(\bar{z}_{xy} - \bar{z} - \gamma_3\frac{x - \bar{x}}{\sigma_x}\sigma_z - \gamma_3'\frac{y - \bar{y}}{\sigma_y}\sigma_z\right)\right\}$$

$$= N\sigma_z(\gamma_3\sigma_z r_{zx} - \gamma_3^2\sigma_z - \gamma_3\gamma_3' r_{xy}\sigma_z + \gamma_3'\sigma_z r_{yz} - \gamma_3\gamma_3' r_{xy}\sigma_z - \gamma_3'^2\sigma_z)$$

$$= N\sigma_z^2\{\gamma_3(r_{zx} - \gamma_3 - \gamma_3' r_{xy}) + \gamma_3'(r_{yz} - \gamma_3' - \gamma_3 r_{xy})\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(14).$$

Using the values of γ_3 and γ_3' given by (9) and (10) we see immediately that

$$r_{zx} - \gamma_3 - \gamma_3' r_{xy} = r_{yz} - \gamma_3' - \gamma_3 r_{xy} \equiv 0.$$

Hence (13) becomes

$$({}_{xy}H_z^2 - {}_{xy}R_z^2)\, N\sigma_z^2 = SS\{n_{xy}(\bar{z}_{xy} - Z)^2\} \quad \ldots\ldots\ldots\ldots\ldots\ldots(15).$$

This is the generalisation of (4). We deduce that ${}_{xy}H_z = {}_{xy}R_z$ if and only if the regression is strictly linear, that otherwise ${}_{xy}H_z^2 > {}_{xy}R_z^2$ and by (6) that ${}_{xy}H_z^2 < 1$.

§ 3. These properties and definitions can be extended to the case of m variables $x_1, x_2, \ldots x_m$. We now use ${}_{2\ldots m}\bar{x}_1$ for the mean of x_1 when $x_2, x_3, \ldots x_m$ are given and denote by $\underset{1\ldots m}{S}$ a summation extending to the variables $x_1, x_2, \ldots x_m$.

If we define the correlation ratio of x_1 on $x_2, x_3, \ldots x_m$ by the equation

$$N\sigma_1^2(1 - {}_{2,3\ldots m}H_1^2) = \underset{1\ldots m}{S}(x_1 - {}_{2,3\ldots m}\bar{x}_1)^2 n_{1\ldots m} \quad \ldots\ldots\ldots\ldots\ldots(16).$$

We can deduce in the same way as in § 2, the relation

$$N\sigma_1^2\, {}_{2,3\ldots m}H_1^2 = \underset{2\ldots m}{S}\{(\bar{x}_1 - {}_{2,3\ldots m}\bar{x}_1)^2 n_{2\ldots m}\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(17).$$

In order to generalise (15) we recall that the "best fitting" linear function of the variables $x_2, x_3, \ldots x_m$ to the mean ${}_{2,3\ldots m}\bar{x}_1$ is

$$\frac{X_1}{\sigma_1} = \frac{\bar{x}_1}{\sigma_1} - b_{12}\frac{x_2 - \bar{x}_2}{\sigma_2} - b_{13}\frac{x_3 - \bar{x}_3}{\sigma_3} - \ldots - b_{1m}\frac{x_m - \bar{x}_m}{\sigma_m} \quad \ldots\ldots\ldots(18),$$

where $b_{1t} = \dfrac{R_{1t}}{R_{11}}$ and R_{pq} is the minor with its proper sign of the element in the pth row and qth column of the determinant

$$R = \begin{vmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{vmatrix} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (19),$$

while the maximum correlation between any linear function of $x_2, x_3, \dots x_m$ and x_1 is ${}_{2,3,\dots m}R_1$ where

$$-{}_{2,3,\dots m}R_1^2 = \frac{R - R_{11}}{R_{11}} = b_{12}r_{12} + b_{13}r_{13} \dots + b_{1m}r_{1m} \quad \dots\dots\dots\dots\dots\dots (20),$$

$$= b_{12}\frac{\underset{12}{S}\{n_{12}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\}}{N\sigma_1\sigma_2} + b_{13}\frac{\underset{13}{S}\{n_{13}(x_1 - \bar{x}_1)(x_3 - \bar{x}_3)\}}{N\sigma_1\sigma_3}$$

$$+ \dots + b_{1m}\frac{\underset{1m}{S}\{n_{1m}(x_1 - \bar{x}_1)(x_m - \bar{x}_m)\}}{N\sigma_1\sigma_m} \quad \dots\dots\dots\dots (21).$$

Subtract (21) from (17), noting that $n_{12} = \underset{3\dots m}{S}\{n_{12\dots m}\}$, etc., we obtain*

$$({}_{2,3,\dots m}H_1^2 - {}_{2,3,\dots m}R_1^2)\,N\sigma_1^2$$

$$= \underset{2\dots m}{S}\{n_{2\dots m}(\bar{x}_1 - {}_{2\dots m}\bar{x}_1)^2\} + \underset{12}{S}\left\{n_{12}\frac{\sigma_1}{\sigma_2}b_{12}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\right\}$$

$$+ \dots + \underset{1m}{S}\left\{n_{1m}\frac{\sigma_1}{\sigma_m}b_{1m}(x_1 - \bar{x}_1)(x_m - \bar{x}_m)\right\}$$

$$= \underset{2\dots m}{S}\{n_{2\dots m}(\bar{x}_1 - {}_{2\dots m}\bar{x}_1)^2\} + \underset{2}{S}\left\{n_2\frac{\sigma_1}{\sigma_2}b_{12}(x_2 - \bar{x}_2)({}_{2\dots m}\bar{x}_1 - \bar{x}_1)\right\} - \dots$$

$$+ \underset{m}{S}\left\{n_m\frac{\sigma_1}{\sigma_m}b_{1m}(x_m - \bar{x}_m)({}_{2\dots m}\bar{x}_1 - \bar{x}_1)\right\}$$

$$= \underset{2\dots m}{S}\left[n_{2\dots m}(\bar{x}_1 - {}_{2\dots m}\bar{x}_1)^2 + n_{2\dots m}\frac{\sigma_1}{\sigma_2}b_{12}(x_2 - \bar{x}_2)({}_{2\dots m}\bar{x}_1 - \bar{x}_1) - \dots\right]$$

$$= \underset{2\dots m}{S}\left[n_{2\dots m}({}_{2\dots m}\bar{x}_1 - \bar{x}_1)\left\{{}_{2\dots m}\bar{x}_1 - \bar{x}_1 + b_{12}\frac{\sigma_1}{\sigma_2}(x_2 - \bar{x}_2) - \dots\right.\right.$$

$$\left.\left. + b_{1m}\frac{\sigma_1}{\sigma_m}(x_m - \bar{x}_m)\right\}\right] \quad \dots\dots\dots (22),$$

using (18) this equation becomes

$$({}_{2,3,\dots m}H_1^2 - {}_{2,3,\dots m}R_1^2)N\sigma_1^2 = \underset{2\dots m}{S}\{n_{2\dots m}({}_{2\dots m}\bar{x}_1 - \bar{x}_1)({}_{2\dots m}\bar{x}_1 - X_1)\}$$

$$= \underset{2\dots m}{S}\{n_{2\dots m}({}_{2\dots m}\bar{x}_1 - X_1)^2\} + \underset{2\dots m}{S}\{n_{2\dots m}(X_1 - \bar{x}_1)({}_{2\dots m}\bar{x}_1 - X_1)\}$$

$$\dots\dots\dots (23).$$

---

* By an extension of the notation described at the beginning of this section $S_{3\dots m}$ denotes a summation with regard to the variables $x_3, x_4, \dots x_m$; $n_{1,2,\dots m}$ is the frequency of a particular combination of the characters $x_1, x_2, \dots x_m$ while n_{12} is the frequency of the combination x_1, x_2.

But $\underset{2\ldots m}{S}\{n_{2\ldots m}(X_1-\bar{x}_1)({}_{2\ldots m}\bar{x}_1-X_1)\}$

$$=\underset{2\ldots m}{S}\left\{n_{2\ldots m}\left(-b_{12}\frac{x_2-\bar{x}_2}{\sigma_2}\sigma_1-b_{13}\frac{x_3-\bar{x}_3}{\sigma_3}\sigma_1-\ldots\right)\left({}_{2\ldots m}\bar{x}_1-\bar{x}_1+b_{12}\frac{x_2-\bar{x}_2}{\sigma_2}\sigma_1+\ldots\right)\right\}\text{(24)},$$

and $N\sigma_1\sigma_2 r_{12}=\underset{1\ldots m}{S}\{n_{1\ldots m}(x_1-\bar{x}_1)(x_2-\bar{x}_2)$

$$=\underset{2\ldots m}{S}\{n_{2\ldots m}(x_2-\bar{x}_2)({}_{2\ldots m}\bar{x}_1-\bar{x}_1)$$

with similar values for r_{13}, r_{23}, etc.

∴ the right hand side of (24)

$$\begin{aligned}&=-b_{12}\sigma_1{}^2r_{12}-b_{12}{}^2\sigma_1{}^2-b_{13}b_{12}\sigma_1{}^2r_{23}\ldots\\&\quad-\sigma_1{}^2b_{13}r_{13}-b_{13}b_{12}\sigma_1{}^2r_{23}-b_{13}{}^2\sigma_1{}^2-b_{13}b_{14}\sigma_1{}^2\ldots\\&\quad-\ldots\\&=\sigma_1{}^2b_{12}(-r_{12}-b_{12}-b_{13}r_{23}-b_{14}r_{24}-\ldots-b_{1m}r_{2m})\\&\quad+\sigma_1{}^2b_{13}(-r_{13}-b_{13}-b_{14}r_{34}-\ldots)\\&\quad+\ldots\quad\ldots\ldots\ldots\ldots\ldots\ldots\text{(25)}.\end{aligned}$$

Each line in (25) is identically zero from the definition of the b's and the properties of the determinant in (19).

Hence
$$_{2\ldots m}H_1{}^2-{}_{2\ldots m}R_1{}^2=\frac{\underset{2\ldots m}{S}\{n_{2\ldots m}({}_{2\ldots m}\bar{x}_1-X_1)^2\}}{N\sigma_1{}^2}\quad\ldots\ldots\ldots\text{(26)},$$

so that the fundamental properties proved by Professor Pearson in connection with the correlation ratio η, hold for the generalised H defined in this section. In particular equation (26) shows that a necessary and sufficient condition for linear regression in multiple correlation of m variables is that

$$_{2,3,\ldots m}H_1{}^2={}_{2,3,\ldots m}R_1{}^2.$$

For in this case the mean value of any array of x_1 will lie *on* the "best fitting" m-dimensional plane.

§ 4. The regression surface of z on xy being assumed of any particular type the constants in the equation may be determined (i) by the method of least squares, i.e. by making the sum of the squares of the deviations $\bar{z}_{xy}-\phi(xy)$ a minimum, $z=\phi(x,y)$ being the regression surface, or (ii) by giving such values to the constants that the correlation between z and $\phi(x,y)$ shall be a maximum. When $\phi(x,y)$ is of the second degree the two methods lead to identical equations for the determination of the coefficients.

The same equations are also obtained if the surface be "fitted" to the means by the method of moments. There is, however, a distinction to be observed. The equation $z=\phi(x,y)$ when the regression surface is of specified degree contains a definite number of constants and the first two methods will give exactly as many independent equations as there are constants to be determined. The method of moments will give as many equations as we please if sufficiently high moments are used, including of course the equations given by the "least squares" method or maximum correlation method. Even without introducing high moments, when there are three variable characters new equations may be obtained by the method of moments, by combinations of characters which do not arise in the other methods. The method of moments is most convenient for our purpose, but we shall only employ those equations which can also be justified by (say) the method of least squares.

For convenience let the origin be taken at the mean of the three characters so that $\bar{x}=\bar{y}=\bar{z}=0$, let $q_{x^s y^t z^u}$ denote

$$\frac{SSS\{n_{xyz}x^s y^t z^u\}}{N\sigma_x{}^s\sigma_y{}^t\sigma_z{}^u}=\frac{p_{s,t,u}}{\sigma_x{}^s\sigma_y{}^t\sigma_z{}^u} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(27).$$

With this notation r_{xy} and $q_{x^1y^1z^0}$ are identical; when z does not appear in the product, it is sufficient to write $q_{x^s y^t}$.

The most reasonable next approximation to make when a linear function $\phi(x, y)$ does not adequately represent the statistics is $\phi(x, y)=$ a quadratic function of x, y.

Let $$\frac{z}{\sigma_z}=d+\frac{ax}{\sigma_x}+\frac{by}{\sigma_y}+\frac{cxy}{\sigma_x\sigma_y}+\frac{ex^2}{\sigma_x{}^2}+\frac{fy^2}{\sigma_y{}^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(28).$$

Multiply (28) by n_{xyz}, sum for all values of x, y, z and divide by N

$$0=d+cr_{xy}+e+f \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(29).$$

Multiply (28) in turn by $\frac{n_{xyz}}{N}$ times $\frac{x}{\sigma_x}$, $\frac{y}{\sigma_y}$, $\frac{xy}{\sigma_x\sigma_y}$, $\frac{x^2}{\sigma_x{}^2}$, $\frac{y^2}{\sigma_y{}^2}$ and sum as before, and we obtain

$$\begin{aligned}
r_{xy} &= a+br_{xy}+cq_{x^2y}+eq_{x^3}+fq_{xy^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(30),\\
r_{yz} &= b+ar_{xy}+cq_{xy^2}+eq_{x^2y}+fq_{y^3} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(31),\\
q_{xyz} &= dr_{xy}+aq_{x^2y}+bq_{xy^2}+cq_{x^2y}+eq_{x^3y}+fq_{xy^3} \quad \ldots\ldots\ldots\ldots(32),\\
q_{x^2z} &= d+aq_{x^3}+bq_{x^2y}+cq_{x^3y}+eq_{x^4}+fq_{x^2y^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots(33),\\
q_{y^2z} &= d+aq_{xy^2}+bq_{y^3}+cq_{xy^3}+eq_{x^2y}+fq_{y^4} \quad \ldots\ldots\ldots\ldots\ldots\ldots(34).
\end{aligned}$$

Actual numerical fitting shows that in many cases e and f are small compared with c[*]. This is the case when the regression of z on x for a constant y and of z on y for a constant x is linear. We shall therefore confine ourselves in the present preliminary paper to the case where we may write

$$\frac{z}{\sigma_z}=d+\frac{ax}{\sigma_x}+\frac{by}{\sigma_y}+\frac{cxy}{\sigma_x\sigma_y} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(35).$$

Here for constant x or y the regression of z on y or of z on x is linear.

---

[*] Cf. *Census of Scotland*, 1911, Vol. III. p. XLVII. where Mr G. Rae obtains by moments the regression of fertility on age of husband and wife. Let W = age of wife, H = age of husband, C = number of children in completed marriages. He finds

$$C_{WH}=20{\cdot}149493-0{\cdot}555812W-0{\cdot}173804H-0{\cdot}002846W^2-0{\cdot}003494H^2+0{\cdot}012675WH.$$

See also the paper by E. M. Elderton in the current part of this Journal, pp. 291—295.

Equations (29) to (32) become, when the regression surface is given by (35),

$$0 = d + cr_{xy} \quad \dots\dots\dots\dots\dots\dots\dots\dots(36),$$

$$r_{xy} = a + br_{xy} + cq_{x^2y} \quad \dots\dots\dots\dots\dots\dots\dots\dots(37),$$

$$r_{yz} = b + ar_{xy} + cq_{xy} \quad \dots\dots\dots\dots\dots\dots\dots\dots(38),$$

$$q_{xyz} = dr_{xy} + aq_{x^2y} + bq_{xy^2} + cq_{x^2y^2} \quad \dots\dots\dots\dots\dots\dots(39),$$

$$= -cr^2_{xy} + aq_{x^2y} + bq_{xy^2} + cq_{x^2y^2} \text{ by (36).}$$

Solving these equations we obtain

$$\frac{a}{\begin{vmatrix} r_{xy} & q_{x^2y} & r_{xz} \\ 1 & q_{xy^2} & r_{yz} \\ q_{xy^2} & q_{x^2y^2} - r^2_{xy} & q_{xyz} \end{vmatrix}} = \frac{b}{\begin{vmatrix} q_{x^2y} & 1 & r_{xz} \\ q_{xy^2} & r_{xy} & r_{yz} \\ q_{x^2y^2} - r^2_{xy} & q_{x^2y} & q_{xyz} \end{vmatrix}}$$

$$= \frac{c}{\begin{vmatrix} 1 & r_{xy} & r_{xz} \\ r_{xy} & 1 & r_{yz} \\ q_{x^2y} & q_{xy^2} & q_{xyz} \end{vmatrix}} = \frac{1}{\begin{vmatrix} 1 & r_{xy} & q_{x^2y} \\ r_{xy} & 1 & q_{xy^2} \\ q_{x^2y} & q_{xy^2} & q_{x^2y^2} - r^2_{xy} \end{vmatrix}} \quad \dots\dots(40),$$

we have already denoted the partial regression coefficients by γ_3 and γ_3' so that

$$\gamma_3 = \frac{r_{xz} - r_{yz}r_{xy}}{1 - r^2_{xy}} \text{ and } \gamma_3' = \frac{r_{yz} - r_{xz}r_{xy}}{1 - r^2_{xy}} \quad \dots\dots\dots(9) \text{ and } (10).$$

In addition let

$$\frac{r_{xy}q_{xy^2} - q_{x^2y}}{1 - r^2_{xy}} = \theta \quad \dots\dots\dots\dots\dots\dots\dots\dots(41),$$

and

$$\frac{r_{xy}q_{x^2y} - q_{xy^2}}{1 - r^2_{xy}} = \phi \quad \dots\dots\dots\dots\dots\dots\dots\dots(42).$$

After some reductions the determinants in (40) yield

$$a = \gamma_3 + c\theta \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(43),$$

$$b = \gamma_3' + c\phi \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(44),$$

$$c = \frac{\theta r_{xz} + \phi r_{yz} + q_{xyz}}{\theta q_{x^2y} + \phi q_{xy^2} + q_{x^2y} - r^2_{xy}} \quad \dots\dots\dots\dots\dots\dots(45),$$

also

$$d = -c\gamma_{xy} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(36).$$

Note that

$$\gamma_3 q_{x^2y} + \gamma_3' q_{xy} \equiv -(\theta r_{zx} + \phi r_{zy}) \quad \dots\dots\dots\dots\dots\dots(46),$$

and

$$\gamma_3 r_{zx} + \gamma_3' r_{zy} = {}_{xy}R_x^2 \text{ (cf. eqn. 11).}$$

§ 5. By definition $(1 - {}_{xy}H_z^2) = \dfrac{S\{n_{xyz}(z - \bar{z}_{xy})^2\}}{N\sigma_z^2}$,

∴ using (35)

$$(1 - {}_{xy}H_z^2) = S\left\{\frac{n_{xyz}}{N}\left(\frac{z}{\sigma_z} - d - \frac{ax}{\sigma_x} - \frac{by}{\sigma_y} - \frac{cxy}{\sigma_x\sigma_y}\right)^2\right\}$$

$$= 1 - 2S\left\{\frac{z}{\sigma_z}\left(d + \frac{ax}{\sigma_x} + \frac{by}{\sigma_y} + \frac{cxy}{\sigma_x\sigma_y}\right)\frac{n_{xyz}}{N}\right\}$$

$$+ d^2 + a^2 + b^2 + c^2 q_{x^2y^2} + 2dcr_{xy} + 2abr_{xy} + 2acq_{x^2y} + 2bcq_{xy^2},$$

$$\begin{aligned}\therefore\ -{}_{xy}H_z^2 = &-2ar_{zx} - 2br_{yz} - 2cq_{xyz} + a^2 + b^2 + d^2 + c^2q_{x^2y^2} \\ &+ 2cdr_{xy} + 2abr_{xy} + 2acq_{x^2y} + 2bcq_{xy^2} \quad \text{.........(47)}, \\ = &\ a(-r_{zx} + a + br_{xy} + cq_{x^2y}) \\ &+ b(-r_{yz} + b + ar_{xy} + cq_{xy^2}) \\ &+ c(-q_{xyz} + dr_{xy} + aq_{x^2y} + bq_{xy^2} + cq_{x^2y^2}) \\ &+ d(d + cr_{xy}) \\ &- ar_{zx} - br_{yz} - cq_{xyz} \quad \text{.........(48)}.\end{aligned}$$

The first four terms vanish by eqns. (36)...(39),

$$\therefore\ {}_{xy}H_z^2 = ar_{zx} + br_{zy} + cq_{xyz} \quad \text{.........(49)}.$$

If we now insert the values of a, b, c from (43)...(45) in (49)

$$\begin{aligned}{}_{xy}H_z^2 &= (\gamma_3 + c\theta)\, r_{zx} + (\gamma_3' + c\phi)\, r_{zy} + cq_{xyz} \\ &= \gamma_3 r_{zx} + \gamma_3' r_{yz} + c(\theta r_{zx} + \phi r_{yz} + q_{xyz}) \\ &= {}_{xy}R_z^2 + \frac{(\theta r_{zx} + \phi r_{yz} + q_{xyz})^2}{q_{x^2y}\theta + q_{xy^2}\phi + q_{x^2y^2} - r^2_{xy}},\end{aligned}$$

or

$${}_{xy}H_z^2 - {}_{xy}R_z^2 = \frac{\{q_{xyz}(1 - r^2_{xy}) - q_{xy}\,(r_{zx}r_{xy} - r_{yz}) - q_{x^2y}(r_{yz}r_{xy} - r_{xz})\}^2}{\left[(q_{x^2y^2} - r^2_{xy}) - \left\{\dfrac{q^2_{xy^2} + q^2_{x^2y} - 2q_{x^2y}q_{xy^2}r_{xy}}{1 - r^2_{xy}}\right\}\right](1 - r^2_{xy})} \quad \text{...(50)}.$$

It follows from (50) and (15) that

$$q_{x^2y^2} - r^2_{xy} > \frac{q^2_{xy^2} + q^2_{x^2y} - 2q_{x^2y}q_{xy^2}r_{xy}}{1 - r^2_{xy}} \quad \text{.........(51)}.$$

If we eliminate q_{xyz} (which is a triple moment troublesome to calculate) between equations (45) and (49) we have

$$\begin{aligned}{}_{xy}H_z^2 &= ar_{zx} + br_{yz} + cq_{xyz} \\ &= ar_{zx} + br_{yz} + c^2(q_{x^2y^2} - r^2_{xy}) + acq_{x^2y} + bcq_{xy^2} \\ &= r_{zx}(\gamma_3 + c\theta) + r_{zy}(\gamma_3' + c\phi) + c^2(q_{x^2y^2} - r^2_{xy}) \\ &\qquad + (\gamma_3 c + c^2\theta)\, q_{x^2y} + (\gamma_3' c + c^2\phi)\, q_{xy^2} \\ &= c^2[q_{x^2y^2} - r^2_{xy} + \theta q_{x^2y} + \phi q_{xy^2}] + c[\theta r_{zx} + \phi r_{zy} + \gamma_3 q_{x^2y} + \gamma_3' q_{xy^2}] \\ &\qquad + \gamma_3 r_{zx} + \gamma_3' r_{zy} \\ &= {}_{xy}R_z^2 + c^2[q_{x^2y^2} - r^2_{xy} + \theta q_{x^2y} + \phi q_{xy^2}] \text{ by (46) and (11)},\end{aligned}$$

$$\therefore\ {}_{xy}H_z^2 - {}_{xy}R_z^2 = c^2\left[q_{x^2y^2} - r^2_{xy} - \frac{q^2_{x^2y} + q^2_{xy^2} - 2q_{xy^2}q_{x^2y}r_{xy}}{1 - r^2_{xy}}\right].$$

Hence

$$c^2 = \frac{{}_{xy}H_z^2 - {}_{xy}R_z^2}{q_{x^2y^2} - r^2_{xy} - \dfrac{q^2_{x^2y} + q^2_{xy^2} - 2q_{xy^2}q_{x^2y}r_{xy}}{1 - r^2_{xy}}} \quad \text{.........(52)}.$$

This value of c^2 is positive by (51).

Equation (52) shows that ${}_{xy}H_z^2 = {}_{xy}R_z^2$ is a necessary condition for linear regression, which we have already proved in equation (26).

The regression surface of z on x, y is, with the values we have now obtained for the constants

$$\frac{Z-\bar{z}}{\sigma_z}=\frac{r_{xz}-r_{yz}r_{xy}}{1-r^2_{xy}}\frac{x-\bar{x}}{\sigma_x}+\frac{r_{yz}-r_{xz}r_{xy}}{1-r^2_{xy}}\frac{y-\bar{y}}{\sigma_y}$$

$$+\frac{\sqrt{{}_{xy}H_z{}^2-{}_{xy}R_z{}^2}}{\sqrt{q_{x^2y^2}-r^2_{xy}-\dfrac{q^2_{xy^2}+q^2_{x^2y}-2q_{x^2y}q_{xy^2}r_{xy}}{1-r^2_{xy}}}}\left\{\frac{r_{xy}q_{xy^2}-q_{x^2y}}{1-r^2_{xy}}\frac{x-\bar{x}}{\sigma_x}+\frac{r_{xy}q_{x^2y}-q_{xy^2}}{1-r^2_{xy}}\frac{y-\bar{y}}{\sigma_y}+\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y}-r_{xy}\right\}\quad\text{.........(53).}$$

The terms in the first line give the ordinary regression plane. In most cases the regression does not differ widely from linearity so that ${}_{xy}H_z{}^2-{}_{xy}R_z{}^2$ is small.

§ 6. We must now get some idea of the relative magnitudes of q_{xy^2}, q_{x^2y}, $q_{x^2y^2}$ and moments of lower order.

First with regard to q_{x^2y} which is equal to $\dfrac{\bar{p}_{21}}{\sigma_x{}^2\sigma_y}$,

$$q_{x^2y}=\frac{S(n_{xy}x^2y)}{N\sigma_x{}^2\sigma_y}=\frac{S(n_x\bar{y}_x x^2)}{N\sigma_x{}^2\sigma_y}\quad\text{........................(54).}$$

If the regression of y on x be linear

$$q_{x^2y}=S\frac{\left(n_x\dfrac{\sigma_y}{\sigma_x}x^3r_{xy}\right)}{N\sigma_x{}^2\sigma_y}=r_{xy}\sqrt{\beta_1},$$

so that q_{x^2y} is zero if the regression is linear and the frequency of x symmetrical. In fact $q_{x^2y}-r_{xy}\sqrt{\beta_1}=0$ is the same as Pearson's criterion for linear regression given by $\bar{\epsilon}=0$ (*Skew Correlation and Non-linear Regression*, p. 30, Eqn. (lxix)).

We may obtain a good approximation for q_{x^2y} by considering the regression of y on x to be parabolic. This is a natural assumption to make if the regression surfaces of x on z, y and y on z, x be also of the hyperboloid type we are discussing for the regression of z on x, y.

For with origin at mean we may write

$$\frac{\bar{x}_{yz}}{\sigma_x}=e+\frac{fy}{\sigma_y}+\frac{gz}{\sigma_z}+\frac{hyz}{\sigma_y\sigma_z}.$$

Hence keeping y constant and summing for the z's

$$\frac{\bar{x}_y}{\sigma_x}=e+\frac{fy}{\sigma_y}+\frac{gz_y}{\sigma_z}+\frac{hy\bar{z}_y}{\sigma_y\sigma_z}.$$

But $$\frac{\bar{z}_y}{\sigma_z}=r_{yz}\frac{y}{\sigma_y}\pm\frac{\sqrt{{}_y\eta_z{}^2-r^2_{yz}}}{\sqrt{\beta_2{}''-\beta_1{}''-1}}\left\{\frac{y^2}{\sigma_{y^2}}-\sqrt{\beta_1{}''}\frac{y}{\sigma_y}-1\right\}*.$$

---

* See Pearson: *l.c.* Eqn. (lxv) (where Y_{px} is a misprint for X_p), and $\beta_2{}''$, $\beta_1{}''$ refer to the distribution of z.

$\therefore\ \dfrac{\bar{x}_y}{\sigma_x}$ is a quadratic expression in y if we remember that h being of order $\sqrt{{}_{yz}H_x^2 - {}_{yz}R_x^2}$ is of the same order as $\sqrt{{}_y\eta_z^2 - r^2_{yz}}$[*].

But the relation $\phi_2({}_x\eta_y^2 - r^2_{xy}) - \bar{\epsilon} = 0$ is satisfied when the regression of y on x is parabolic[†].

Here $\qquad \phi_2 = \beta_2 - \beta_1 - 1,\quad \bar{\epsilon} = \epsilon - r_{xy}\sqrt{\beta_1},\quad \epsilon = q_{x^2y}$[‡].

Hence

$$q_{x^2y} - r_{xy}\sqrt{\beta_1} = \sqrt{{}_x\eta_y^2 - r^2_{xy}}\,\sqrt{\beta_2 - \beta_1 - 1} \quad \text{..................(55)}.$$

Similarly

$$q_{xy^2} - r_{xy}\sqrt{\beta_1'} = \sqrt{{}_y\eta_x^2 - r^2_{xy}}\,\sqrt{\beta_2' - \beta_1' - 1} \quad \text{..................(56)}.$$

The use of these approximations will save the direct calculation of q_{x^2y} and q_{xy^2} provided we can determine the signs to be attached to $\sqrt{{}_x\eta_y^2 - r^2_{xy}}$ and $\sqrt{{}_y\eta_x^2 - r^2_{xy}}$. This is often easily done by inspection of the regression curve whose equation is

$$\frac{\bar{y}_x}{\sigma_y} = r\frac{x}{\sigma_x} \pm \sqrt{\frac{{}_x\eta_y^2 - r^2_{xy}}{\beta_2 - \beta_1 - 1}}\left\{\frac{x^2}{\sigma_x^2} - \sqrt{\beta_1}\frac{x}{\sigma_x} - 1\right\}\S.$$

We can approximate to $q_{x^2y^2}$ as follows[‖]:

$$q_{x^2y^2} = \frac{SS(n_{xy}x^2y^2)}{N\sigma_x^2\sigma_y^2} = \frac{S\{n_x x^2(\sigma^2_{Yx} + (Y_x - \bar{y})^2)\}}{N\sigma_x^2\sigma_y^2},$$

where Y_x is the value given by the regression straight line and $\sigma^2_{Yx} \times n_x$ the second moment of the array of y's for a given x, about the point Y_x.

But

$$\frac{S\{n_x x^2(Y_x - \bar{y})^2\}}{N\sigma_x^2\sigma_y^2} = S(n_x x^4 r^2_{xy}\sigma_y^2)/\sigma_x^2$$
$$= \beta_2 r^2_{xy}.$$

Thus

$$q_{x^2y^2} = \frac{S(n_x x^2\sigma^2_{Yx})}{N\sigma_x^2\sigma_y^2} + \beta_2 r^2_{xy},$$

Similarly

$$q_{x^2y} = \frac{S(n_y y^2\sigma^2_{Xy})}{N\sigma_x^2\sigma_y^2} + \beta_2' r^2_{xy};$$

or, so far *without approximation*

$$q_{x^2y^2} = \frac{S(n_x x^2\sigma^2_{Yx}) + S(n_y y^2\sigma^2_{Xy})}{2N\sigma_x^2\sigma_y^2} + \tfrac{1}{2}(\beta_2 + \beta_2')\,r^2_{xy}.$$

---

[*] It is noteworthy that the hypothesis that regression of z on x, y although of 2nd degree is such that regression of z on x for a *constant* y is linear leads to the result that the total regression of z on x is *parabolic*.

[†] Pearson, *l.c.* p. 28, Eqn. (lxiii).

[‡] Pearson, *l.c.* Eqns. (li), (xlv) and (lxiii).

[§] Pearson, *l.c.* Eqn. (lxv).

[‖] It can be found fairly directly by tabling to the squares of the variates, when we need a simple product moment. In a later part of this paper some comparisons of actual and approximate values for numerical cases will be found.

Now the mean values of σ^2_{Yx} and σ^2_{Xy} are known to be

$$\sigma_y^2(1-r^2_{xy}) \text{ and } \sigma_x^2(1-r^2_{xy}),$$

and the deviations from these are usually somewhat irregular. Rarely can we do anything better than assume them to vary with a *slight* linear variation from the mean. For example

$$\sigma^2_{Yx}=\sigma_y^2(1-r^2_{xy})(1+\lambda x),$$

where λ is small. In such a case

$$\frac{S(n_x x^2\sigma^2_{Yx})}{N\sigma_x^2\sigma_y^2}=(1-r^2_{xy})(1+\lambda\sqrt{\beta_1}),$$

or to a fair degree of approximation*, we may put

$$q_{x^2y^2}=1-r^2_{xy}+\tfrac{1}{2}(\beta_2+\beta_2')\,r^2_{xy},$$

and thus write

$$\begin{aligned} q_{x^2y^2}-r^2_{xy} &= 1+\tfrac{1}{2}(\beta_2+\beta_2'-4)\,r^2_{xy} \\ &= 1+r^2_{xy}+\tfrac{1}{2}(\beta_2-3+\beta_2'-3)\,r^2_{xy}. \end{aligned}$$

The latter part of this expression vanishes if the frequency of the x and y variates be mesokurtic. It can of course be retained if desired but its product with $\sqrt{{}_{xy}H_z^2-{}_{xy}R_z^2}$ will usually be of the second order. If we write

$$\psi=\tfrac{1}{2}(\beta_2-3+\beta_2'-3),$$

we find the approximate regression surface

$$\begin{aligned} \frac{Z-\bar{z}}{\sigma_z} &= \frac{r_{xz}-r_{yz}r_{xy}}{1-r^2_{xy}}\,\frac{(x-\bar{x})}{\sigma_x}+\frac{r_{yz}-r_{xz}r_{xy}}{1-r^2_{xy}}\,\frac{(y-\bar{y})}{\sigma_y} \\ &\quad +\sqrt{\frac{{}_{xy}H_z^2-{}_{xy}R_z^2}{1+(1+\psi)\,r^2_{xy}}}\left\{\frac{(x-\bar{x})}{\sigma_x}\,\frac{(y-\bar{y})}{\sigma_y}-r_{xy}\right\} \quad \text{......(57).} \end{aligned}$$

This equation (57) enables us to express approximately the multiple ${}_{xy}H_z$ in terms of the simple ${}_y\eta_z$, ${}_x\eta_z$, ${}_x\eta_y$, ${}_y\eta_x$.

§ 7. To obtain this connection between the multiple ${}_{xy}H_z$ and the simple η's we may proceed as follows:

$$\frac{\bar{z}_{xy}}{\sigma_z}=d+\frac{ax}{\sigma_x}+\frac{by}{\sigma_y}+\frac{cxy}{\sigma_x\sigma_y}, \text{ origin at mean.}$$

Hence keeping y constant and summing for the x's

$$\frac{\bar{z}_y}{\sigma_z}=d+\frac{a\bar{x}_y}{\sigma_x}+\frac{by}{\sigma_y}+\frac{c\bar{x}_y y}{\sigma_x\sigma_y} \quad \text{..........................(58).}$$

But
$$ {}_y\eta_z^2=\frac{SSS\{n_{xyz}(\bar{z}_y-\bar{z})^2\}}{N\sigma_z^2}=\frac{SSS(n_{xyz}\,\bar{z}_y^2)}{N\sigma_z^2} \text{ if } \bar{z}=0,$$

and
$$ {}_{xy}H_z^2=\frac{SSS\{n_{xyz}(\bar{z}_{xy}-\bar{z})^2\}}{N\sigma_z^2}=\frac{SSS\{n_{xyz}\bar{z}^2_{xy}\}}{N\sigma_z^2},$$

---

* I.e. we are neglecting terms of the second order as $\lambda\sqrt{\beta_1}$.

$$\therefore\quad {}_{xy}H_z^2 - {}_y\eta_z^2 = \underset{x\,y\,z}{SSS}\left[\frac{n_{xyz}}{N}\left(\frac{a\,(x-\bar{x}_y)}{\sigma_x} + \frac{cy}{\sigma_y}\frac{(x-\bar{x}_y)}{\sigma_y}\right)\left(2d + a\frac{x+\bar{x}_y}{\sigma_x} + 2\frac{by}{\sigma_y} + \frac{cy}{\sigma_y}\frac{x+\bar{x}}{\sigma_x}\right)\right]$$

$$= \underset{x\,y}{SS}\left\{\frac{n_{xy}}{N}\left(\frac{x-\bar{x}_y}{\sigma_x}\right)\left(a+\frac{cy}{\sigma_y}\right)\left[2\left(d+\frac{by}{\sigma_y}\right)+\left(a+\frac{cy}{\sigma_y}\right)\left(\frac{x+\bar{x}_y}{\sigma_x}\right)\right]\right\}$$

$$= 2\underset{y}{S}\left\{\left(a+\frac{cy}{\sigma_y}\right)\left(d+\frac{by}{\sigma_y}\right)\underset{x}{S}\frac{x-\bar{x}_y}{\sigma_x}\frac{n_{xy}}{N}\right\} + \underset{x\,y}{SS}\left\{\frac{x^2-\bar{x}_y^{\,2}}{\sigma_x^{\,2}}\left(a+\frac{cy}{\sigma_y}\right)^2\frac{n_{xy}}{N}\right\}$$

$$= 0 + \underset{y}{S}\left\{\left(a+\frac{cy}{\sigma_y}\right)^2\left(\underset{x}{S}\frac{n_{xy}}{N}\frac{x^2-\bar{x}_y^{\,2}}{\sigma_x^{\,2}}\right)\right\}.$$

Now $$\underset{x}{S}\{n_{xy}x^2\} = \underset{x}{S}\{(x-\bar{x}_y+\bar{x}_y)^2\,n_{xy}\} = n_y\,(\sigma_{x_y}^{\,2} + 0 + \bar{x}_y^{\,2}),$$

$$\therefore\quad {}_{xy}H_z^2 - {}_y\eta_z^2 = \underset{y}{S}\left\{\left(a+\frac{cy}{\sigma_y}\right)^2\frac{\sigma_{x_y}^{\,2}}{\sigma_x^{\,2}}\frac{n_y}{N}\right\}$$

$$= \underset{y}{S}\left\{\left(a+\frac{cy}{\sigma_y}\right)^2(1-{}_y\eta_x^2)\,\frac{n_y}{N}\right\}$$

$$= (1-{}_y\eta_x^2)\,\underset{y}{S}\left\{\frac{n_y}{N}\left(a^2+\frac{2acy}{\sigma_y}+\frac{c^2y^2}{\sigma_y^{\,2}}\right)\right\},$$

$$\therefore\quad {}_{xy}H_z^2 - {}_y\eta_z^2 = (1-{}_y\eta_x^2)\,(a^2+c^2) \quad\text{.....................(59).}$$

Similarly $${}_{xy}H_z^2 - {}_x\eta_z^2 = (1-{}_x\eta_y^2)\,(b^2+c^2) \quad\text{.....................(60).}$$

Remembering that $a = \gamma_3 + c\theta$, $b = \gamma_3' + c\phi$ we get from (59) and (60)

$$\gamma_3'\,\phi\left(\frac{{}_{xy}H_z^2 - {}_y\eta_z^2}{1-{}_y\eta_x^2}\right) - \gamma_3\,\theta\left(\frac{{}_{xy}H_z^2 - {}_x\eta_z^2}{1-{}_x\eta_y^2}\right) = \gamma_3\gamma_3'\,(\gamma_3\phi - \gamma_3'\theta)$$
$$+ c^2\,\{\gamma_3'\phi - \gamma_3\theta - \theta\phi\,(\gamma_3\phi - \gamma_3'\theta)\} \quad\text{......(61).}$$

From the values of γ_3, γ_3', θ, ϕ in (9), (10), (41), (42) we obtain easily

$$\gamma_3\phi - \gamma_3'\,\theta = \frac{r_{yz}q_{x^2y} - r_{xz}q_{xy^2}}{1-r^2_{xy}},$$

$$\gamma_3'\phi - r_3\theta = \frac{r_{xz}q_{x^2y} - r_{yz}q_{xy^2}}{1-r^2_{xy}},$$

$$\gamma_3'\phi - \gamma_3\theta - \theta\phi\,(\gamma_3\phi - \gamma_3'\theta)$$
$$= \frac{r_{xz}q_{x^2y} - r_{yz}q_{xy^2}}{1-r^2_{xy}} - \frac{(r_{xy}q_{xy^2} - q_{x^2y})\,(r_{xy}q_{x^2y} - q_{xy^2})\,(r_{yz}q_{x^2y} - r_{xz}q_{xy^2})}{(1-r^2_{xy})^3}$$

and c^2 is given by (52). Hence (61) may be written

$$({}_{xy}H_z^2 - {}_{xy}R_z^2)\left\{\frac{(r_{yz}-r_{xz}r_{xy})\,(r_{xy}q_{xy^2}-q_{x^2y})}{(1-r^2_{xy})^2\,(1-{}_y\eta_x^2)} - \frac{(r_{xz}-r_{yz}r_{xy})\,(r_{xy}q_{xy^2}-q_{x^2y})}{(1-r^2_{xy})^2\,(1-{}_x\eta_y^2)}\right.$$
$$\left. - \frac{\dfrac{(r_{xz}q_{x^2y} - r_{yz}q_{xy})}{1-r^2_{xy}} - \dfrac{(r_{xy}q_{xy^2}-q_{x^2y})\,(r_{xy}q_{x^2y}-q_{xy^2})\,(r_{yz}q_{x^2y}-r_{xz}q_{xy^2})}{(1-r^2_{xy})^3}}{(q_{x^2y^2}-r^2_{xy}) - \dfrac{(q^2_{x^2y}+q^2_{xy^2}-2q_{xy^2}q_{x^2y}r_{xy})}{1-r^2_{xy}}}\right\}$$

$$= \frac{(r_{yz} - r_{xz}r_{xy})(r_{xy}q_{x^2y} - q_{xy^2})\,{}_y\eta_z^2}{(1 - r^2_{xy})^2(1 - {}_y\eta_x^2)} - \frac{(r_{xz} - r_{yz}r_{xy})(r_{xy}q_{xy^2} - q_{x^2y})\,{}_x\eta_z^2}{(1 - r^2_{xy})^2(1 - {}_y\eta_x^2)}$$

$$+ \frac{(r_{xz} - r_{yz}r_{xy})(r_{yz} - r_{xz}r_{xy})(r_{yz}q_{x^2y} - r_{xz}q_{xy^2})}{(1 - r^2_{xy})^3}$$

$$- {}_{xy}R_z^2\left\{\frac{(r_{yz} - r_{xz}r_{xy})(r_{xy}q_{x^2y} - q_{xy^2})}{(1 - r^2_{xy})^2(1 - {}_y\eta_x^2)} - \frac{(r_{xz} - r_{yz}r_{xy})(r_{xy}q_{xy^2} - q_{x^2y})}{(1 - r^2_{xy})^2(1 - {}_x\eta_y^2)}\right\} \quad \ldots\ldots(62).$$

In general the square of any correlation ratio (ordinary or generalised) differs little from the square of the corresponding correlation coefficient; also we have seen in (55) that an approximate value for q_{x^2y} is

$$r_{xy}\sqrt{\beta_1} - \sqrt{{}_x\eta_y^2 - r^2_{xy}}\sqrt{\beta_2 - \beta_1 - 1}.$$

Now β_1 is itself in general small, so that without making any assumption as to the relative order of magnitude of β_1 and ${}_x\eta_y^2 - r^2_{xy}$ we may safely treat q_{x^2y} and q_{xy^2} as small quantities.

Thus it appears that (62) is an equation in which all the terms involved are small, and a certain amount of care is required in deducing from it an approximation to the value of ${}_{xy}H_z^2 - {}_{xy}R_z^2$.

We have

$$\left.\begin{aligned} \frac{1}{1 - {}_y\eta_x^2} &\equiv \frac{1}{1 - r^2_{xy}} + \frac{{}_y\eta_x^2 - r^2_{xy}}{(1 - {}_y\eta_x^2)(1 - r^2_{xy})} = \frac{1}{1 - r^2_{xy}} + \xi_1 \text{ say} \\ \frac{1}{1 - {}_x\eta_y^2} &\equiv \frac{1}{1 - r^2_{xy}} + \frac{{}_x\eta_y^2 - r^2_{xy}}{(1 - {}_x\eta_y^2)(1 - r^2_{xy})} = \frac{1}{1 - r^2_{xy}} + \xi_1' \end{aligned}\right\} \ldots\ldots(63),$$

$$\frac{{}_y\eta_z^2}{1 - {}_y\eta_x^2} = r^2_{zy}\left(\frac{1}{1 - r^2_{xy}} + \xi_1\right) + ({}_y\eta_z^2 - r^2_{zy})\left(\frac{1}{1 - r^2_{xy}} + \xi_1\right)$$

$$= \frac{r^2_{zy}}{1 - r^2_{xy}} + \lambda_1 + \lambda_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(64),$$

where $\lambda_1 = \dfrac{r^2_{zy}({}_y\eta_x^2 - r^2_{xy})}{(1 - {}_y\eta_x^2)(1 - r^2_{xy})} + \dfrac{{}_y\eta_z^2 - r^2_{zy}}{1 - r^2_{xy}}$ and $\lambda_2 = \dfrac{({}_y\eta_z^2 - r^2_{zy})({}_y\eta_x^2 - r^2_{xy})}{(1 - {}_y\eta_x^2)(1 - r^2_{xy})}$,

while
$$\frac{{}_x\eta_z^2}{1 - {}_x\eta_y^2} = \frac{r^2_{zx}}{1 - r^2_{xy}} + \lambda_1' + \lambda_2' \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(64)',$$

where λ_1', λ_2' may be obtained from λ_1, λ_2 by an interchange of x and y in the suffixes.

Finally
$$\frac{1}{q_{x^2y^2} - r^2_{xy} - \dfrac{q^2_{xy^2} + q^2_{x^2y} - 2q_{xy^2}q_{x^2y}r_{xy}}{1 - r^2_{xy}}} \equiv \frac{1}{q_{x^2y^2} - r^2_{xy}} + \kappa_2 \ldots\ldots\ldots(65),$$

where

$$\kappa_2 = \frac{q^2_{xy^2} + q^2_{x^2y} - 2q_{xy^2}q_{x^2y}r_{xy}}{(1 - r^2_{xy})(q_{x^2y^2} - r^2_{xy})\left(q_{x^2y^2} - r^2_{xy} - \dfrac{q^2_{xy^2} + q^2_{x^2y} - 2q_{xy^2}q_{x^2y}r_{xy}}{1 - r^2_{xy}}\right)}.$$

The suffixes of ξ_1, ξ_1', λ_1, λ_2, λ_1', λ_2' and κ_2 denote the order of smallness of these terms.

We make the corresponding substitutions in equation (62), including the value $\frac{r^2_{yz} + r^2_{zx} - 2r_{zx}r_{yx}r_{xz}}{1 - r^2_{xy}}$ for ${}_{xy}R_z^2$ on the right-hand side and obtain the following accurate formula for ${}_{xy}H_z^2 - {}_{xy}R_z^2$:

$$\begin{aligned}
({}_{xy}H_z^2 - {}_{xy}R_z^2)\Bigg[&\frac{(r_{yz} - r_{xz}r_{xy})(r_{xy}q_{x^2y} - q_{xy^2})}{(1 - r^2_{xy})^2}\left(\frac{1}{1 - r^2_{xy}} + \xi_1\right)\\
&\qquad - \frac{(r_{xz} - r_{yz}r_{xy})(r_{xy}q_{xy^2} - q_{x^2y})}{(1 - r^2_{xy})^2}\left(\frac{1}{1 - r^2_{xy}} + \xi_1'\right)\\
&- \left\{\frac{r_{xz}q_{x^2y} - r_{yz}q_{xy^2}}{1 - r^2_{xy}} - \frac{(r_{yz}q_{x^2y} - r_{xz}q_{xy^2})(r_{xy}q_{xy^2} - q_{x^2y})(r_{xy}q_{x^2y} - q_{xy^2})}{(1 - r^2_{xy})^3}\right\}\\
&\qquad \times \left(\frac{1}{q_{x^2y^2} - r^2_{xy}} + \kappa_2\right)\Bigg]\\
= \frac{(r_{yz} - r_{xz}r_{xy})(r_{xy}q_{x^2y} - q_{xy^2})}{(1 - r^2_{xy})^2}&\left(\frac{r^2_{zy}}{1 - r^2_{xy}} + \lambda_1 + \lambda_2\right)\\
&\qquad - \frac{(r_{xz} - r_{yz}r_{xy})(r_{xy}q_{xy^2} - q_{x^2y})}{(1 - r^2_{xy})^2}\left(\frac{r^2_{zx}}{1 - r^2_{xy}} + \lambda_1' + \lambda_2'\right)\\
+ \frac{(r_{yz}q_{x^2y} - r_{xz}q_{xy^2})(r_{xz} - r_{yz}r_{xy})(r_{yz} - r_{xz}r_{xy})}{(1 - r^2_{xy})^3}&\\
- \frac{r^2_{yz} + r^2_{xz} - 2r_{xz}r_{yz}r_{xy}}{(1 - r^2_{xy})}\Bigg[&\frac{(r_{yz} - r_{xz}r_{xy})(r_{xy}q_{x^2y} - q_{xy^2})}{(1 - r^2_{xy})^2}\left(\frac{1}{1 - r^2_{xy}} + \xi_1\right)\\
&- \frac{(r_{xz} - r_{yz}r_{xy})(r_{xy}q_{xy^2} - q_{x^2y})}{(1 - r^2_{xy})^2}\left(\frac{1}{1 - r^2_{xy}} + \xi_1'\right)\Bigg] \quad \text{......(66).}
\end{aligned}$$

The right-hand side of (66) apparently contains terms of the first order, while on the left the lowest order occurring is the second. But the coefficient of q_{x^2y} in the first order terms is

$$\begin{aligned}
&\frac{1}{(1 - r^2_{xy})^3}\{(r_{yz} - r_{xz}r_{xy})r_{xy}r^2_{zy} + (r_{xz} - r_{yz}r_{xy})r^2_{zx} + r_{yz}(r_{xz} - r_{yz}r_{xy})(r_{yz} - r_{xz}r_{xy})\}\\
&\qquad - \frac{1}{(1 - r^2_{xy})^4}\{(r_{yz} - r_{xz}r_{xy})r_{xy} + (r_{xz} - r_{yz}r_{xy})\}(r^2_{yz} + r^2_{xz} - 2r_{yz}r_{xy}r_{xy}),
\end{aligned}$$

and vanishes identically. Similarly the coefficient of q_{xy^2} is zero and thus equation (66) is a relation between terms of second and higher orders.

A first approximation then may be obtained by equating second order terms. This gives

$$\begin{aligned}
({}_{xy}H_z^2 - {}_{xy}R_z^2)\Bigg[&\frac{(r_{yz} - r_{xz}r_{xy})(r_{xy}q_{x^2y} - q_{xy^2}) - (r_{xz} - r_{yz}r_{xy})(r_{xy}q_{xy^2} - q_{x^2y})}{(1 - r^2_{xy})^3}\\
&\qquad - \frac{r_{xz}q_{x^2y} - r_{yz}q_{xy^2}}{(1 - r^2_{xy})(q_{x^2y^2} - r^2_{xy})}\Bigg]\\
= \frac{(r_{yz} - r_{xz}r_{xy})(r_{xy}q_{x^2y} - q_{xy^2})}{(1 - r^2_{xy})^2}&\left[\lambda_1 - \frac{r^2_{yz} + r^2_{xz} - 2r_{xz}r_{yz}r_{xy}}{1 - r^2_{xy}}\xi_1\right]\\
- \frac{(r_{xz} - r_{yz}r_{xy})(r_{xy}q_{xy^2} - q_{x^2y})}{(1 - r^2_{xy})^2}&\left[\lambda_1' - \frac{r^2_{yz} + r^2_{zx} - 2r_{xz}r_{yz}r_{xy}}{1 - r^2_{xy}}\xi_1'\right] \quad \text{......(67).}
\end{aligned}$$

The coefficient of ${}_{xy}H_z^2 - {}_{xy}R_z^2$ on the left reduces to

$$\frac{1}{(1-r^2_{xy})^2}\,[q_{x^2y}r_{xz} - q_{xy^2}r_{yz}]\left(1 - \frac{1-r^2_{xy}}{q_{x^2y^2} - r^2_{xy}}\right),$$

$$\lambda_1 - \frac{r^2_{yz} + r^2_{xz} - 2r_{xz}r_{yz}r_{xy}}{1-r^2_{xy}}\,\xi_1 = \frac{{}_y\eta_z^2 - r^2_{zy}}{1-r^2_{xy}} - \xi_1\,\frac{(r_{yz}r_{xy} - r_{xz})^2}{1-r^2_{xy}}$$

$$= \frac{{}_y\eta_z^2 - r^2_{zy}}{1-r^2_{xy}} - \frac{({}_y\eta_x^2 - r^2_{xy})}{1-{}_y\eta_x^2}\left(\frac{r_{yz}r_{xy} - r_{xz}}{1-r^2_{xy}}\right)^2 \quad \text{......(68).}$$

Similarly

$$\lambda_1' - \frac{r^2_{yz} + r^2_{xz} - 2r_{xz}r_{yz}r_{xy}}{1-r^2_{xy}}\,\xi_1' = \frac{{}_x\eta_z^2 - r^2_{zx}}{1-r^2_{xy}} - \frac{{}_x\eta_y^2 - r^2_{xy}}{1-{}_x\eta_y^2}\left(\frac{r_{xz}r_{xy} - r_{yz}}{1-r^2_{xy}}\right)^2 \text{...(69).}$$

We shall still be correct to second order terms if when using (68) and (69) in (67) we replace $1 - {}_y\eta_x^2$ and $1 - {}_x\eta_y^2$ occurring in denominators by $1 - r^2_{xy}$; so that

$$({}_{xy}H_z^2 - {}_{xy}R_z^2)\,\frac{q_{x^2y^2} - 1}{q_{x^2y^2} - r^2}$$

$$= \frac{r_{xy}q_{x^2y} - q_{xy^2}}{q_{x^2y}r_{xz} - q_{xy^2}r_{yz}} \cdot \frac{r_{yz} - r_{xz}r_{xy}}{1-r^2_{xy}}\left[({}_y\eta_z^2 - r^2_{zy}) - \left(\frac{r_{yz}r_{xy} - r_{xz}}{1-r^2_{xy}}\right)^2({}_y\eta_x^2 - r^2_{xy})\right]$$

$$+ \frac{r_{xy}q_{xy^2} - q_{x^2y}}{q_{xy^2}r_{yz} - q_{x^2y}r_{xz}} \cdot \frac{r_{xz} - r_{yz}r_{xy}}{1-r^2_{xy}}\left[({}_x\eta_z^2 - r^2_{zx}) - \left(\frac{r_{xz}r_{xy} - r_{yz}}{1-r^2_{xy}}\right)^2({}_x\eta_y^2 - r^2_{xy})\right]$$

......(70).

§ 8. This result is of importance, as it shows that the heavy labour of the direct calculation of the generalised correlation ratio can be replaced by the calculation of four simple correlation ratios.

The coefficients involved are the ordinary coefficients of linear regression denoted above by γ_3, γ_3' and expressions involving product moments of orders 3 and 4. To these latter we may approximate by the methods of § 6.

A good approximation for $\dfrac{q_{x^2y^2} - 1}{q_{x^2y^2} - r^2_{xy}}$ is $\dfrac{2r^2_{xy}}{1 + r^2_{xy}}$. If greater accuracy be needed we may use

$$\frac{q_{x^2y^2} - 1}{q_{x^2y^2} - r^2_{xy}} = \frac{\left(\tfrac{1}{2}(\beta_2 + \beta_2' - 6) + 2\right)r^2_{xy}}{1 + r^2_{xy} + \tfrac{1}{2}(\beta_2 + \beta_2' - 6)\,r^2_{xy}}.$$

We saw in § 6 (equation 55) that

$$q_{x^2y} = r_{xy}\sqrt{\beta_1} + \sqrt{{}_x\eta_y^2 - r^2_{xy}}\,\sqrt{\beta_2 - \beta_1 - 1},$$

$$q_{xy^2} = r_{xy}\sqrt{\beta_1'} + \sqrt{{}_y\eta_x^2 - r^2_{xy}}\,\sqrt{\beta_2' - \beta_1' - 1},$$

approximately. β_1 and β_1', which are zero in normal correlation, will in general be very small compared with ${}_x\eta_y^2 - r^2_{xy}$ and ${}_y\eta_x^2 - r^2_{xy}$ so that

$$\frac{r_{xy}q_{x^2y} - q_{xy^2}}{q_{x^2y}r_{xz} - q_{xy^2}r_{yz}} = \frac{r_{xy}\sqrt{(\beta_2 - 1)({}_x\eta_y^2 - r^2_{xy})} - \sqrt{(\beta_2' - 1)({}_y\eta_x^2 - r^2_{xy})}}{r_{xz}\sqrt{(\beta_2 - 1)({}_x\eta_y^2 - r^2_{xy})} - r_{yz}\sqrt{(\beta_2' - 1)({}_y\eta_x^2 - r^2_{xy})}},$$

and

$$\frac{r_{xy}q_{xy^2} - q_{x^2y}}{q_{xy^2}r_{yz} - q_{x^2y}r_{xz}} = \frac{r_{xy}\sqrt{(\beta_2' - 1)({}_y\eta_x^2 - r^2_{xy})} - \sqrt{(\beta_2 - 1)({}_x\eta_y^2 - r^2_{xy})}}{r_{yz}\sqrt{(\beta_2' - 1)({}_y\eta_x^2 - r^2_{xy})} - \sqrt{(\beta_2 - 1)({}_x\eta_y^2 - r^2_{xy})}}.$$

In the important case $\beta_1 = \beta_1' = 0$ and ${}_y\eta_x = {}_x\eta_y = r_{xy}$ these approximations fail, and so does the process by which (70) was obtained as θ and ϕ vanish and (61) is indeterminate.

We must then fall back on equation (59)

$$\begin{aligned}{}_{xy}H_z^2 - {}_y\eta_z^2 &= (a^2 + c^2)(1 - {}_y\eta_x^2)\\ &= (\gamma_3^2 + c^2)(1 - r^2_{xy}) \text{ since } \theta = 0 \text{ and } {}_y\eta_x = r_{xy}\\ &= \gamma_3^2(1 - r^2_{xy}) + \frac{({}_{xy}H_z^2 - {}_{xy}R_z^2)(1 - r^2_{xy})}{q_{x^2y^2} - r^2_{xy}}(1 + \kappa_2),\end{aligned}$$

or
$$({}_{xy}H_z^2 - {}_{xy}R_z^2)\left(1 - \frac{1 - r^2_{xy}}{q_{x^2y^2} - r^2_{xy}}\right) = {}_y\eta_z^2 - {}_{xy}R_z^2 + \gamma_3^2(1 - r^2_{xy}),$$

neglecting 3rd order terms.

The right-hand side reduces to ${}_y\eta_z^2 - r^2_{yz}$, so that

$$\begin{aligned}{}_{xy}H_z^2 - {}_{xy}R_z^2 &= \left(\frac{q_{x^2y^2} - r^2_{xy}}{q_{x^2y^2} - 1}\right)({}_y\eta_z^2 - r^2_{yz}) \quad \text{.........................(71)}\\ &= \frac{2r^2_{xy}}{1 + r^2_{xy}}({}_y\eta_z^2 - r^2_{zy}) \text{ approximately;}\end{aligned}$$

of course in these circumstances (60) would lead to the value

$$\frac{2r^2_{xy}}{1 + r^2_{xy}}({}_x\eta_z^2 - r^2_{zx}) \quad \text{............................(72),}$$

showing that if $\beta_1 = \beta_1' = 0$ and if $({}_x\eta_y^2 - r^2_{yx}) - ({}_y\eta_x^2 - r^2_{xy})$ is of higher order than the first then $({}_x\eta_z^2 - r^2_{zx}) - ({}_y\eta_z^2 - r^2_{zy})$ is also of higher order than the first.

We shall now seek relations between the six correlation ratios of three "hyperbolic" variates.

From (59) and (60) we get, on eliminating ${}_{xy}H_z^2$,

$$\begin{aligned}{}_y\eta_z^2 - {}_x\eta_z^2 &= b^2 - a^2 + c^2({}_y\eta_x^2 - {}_x\eta_y^2)^2 + a^2{}_y\eta_x^2 - b^2{}_x\eta_y^2\\ &= \gamma_3'^2 - \gamma_3^2 + \gamma_3^2{}_y\eta_x^2 - \gamma_3'^2{}_x\eta_y^2 + 2(\gamma_3\theta{}_y\eta_x^2 - \gamma_3'\phi{}_x\eta_y^2)\\ &\qquad + c^2(\theta^2 - \phi^2 + {}_y\eta_x^2 - {}_x\eta_y^2 + \theta^2{}_y\eta_x^2 - \phi^2{}_x\eta_y^2)\\ &= (\gamma_3'^2 - \gamma_3^2)(1 - r^2_{xy}) + \gamma_3^2({}_y\eta_x^2 - r^2_{xy}) - \gamma_3'^2({}_x\eta_y^2 - r^2_{xy})\\ &\qquad - 2c(\gamma\theta - \gamma'\phi)(1 - r^2_{xy})\\ &\quad + \{2c\gamma_3\theta({}_y\eta_x^2 - r^2_{xy}) - 2c\gamma_3'\phi({}_x\eta_y^2 - r^2_{xy})\\ &\qquad + c^2(\theta^2 - \phi^2 + {}_y\eta_x^2 - {}_x\eta_y^2 + \theta^2{}_y\eta_x^2 - \phi^2{}_x\eta_y^2)\}.\end{aligned}$$

The terms in the second line are second order terms. Neglecting these and noting that

$$(\gamma_3'^2 - \gamma_3^2)(1 - r^2_{xy}) \equiv r^2_{yz} - r^2_{xz} \text{ and } (\gamma_3\theta - \gamma_3'\phi)(1 - r^2_{xy}) \equiv r_{yz}q_{xy^2} - r_{xz}q_{x^2y},$$

we obtain the following equation for c:

$$\begin{aligned}({}_y\eta_z^2 - r^2_{zy}) - ({}_x\eta_z^2 - r^2_{zx}) - \gamma_3^2({}_y\eta_x^2 - r^2_{xy}) + \gamma_3'^2({}_x\eta_y^2 - r^2_{xy})&\\ = 2c(r_{xz}q_{x^2y} - r_{yz}q_{xy^2})&\text{......(73).}\end{aligned}$$

Now $({}_{xy}H_z^2 - {}_{xy}R_z^2)\dfrac{q_{x^2y^2}-1}{q_{x^2y^2}-r^2_{xy}} = 2c^2r^2_{xy}$ if we neglect second order terms. If we use the value of c given by (73) in (70), and adopt the notation ${}_xU_z$ for ${}_x\eta_z^2 - r_{zx}^2$, we have

$$\begin{aligned}r^2_{xy}\left[{}_yU_z - {}_xU_z - \gamma_3^2\,{}_yU_x + \gamma_3'^2\,{}_xU_y\right]^2 \\ = 2\left(r_{xz}q_{x^2y} - r_{yz}q_{xy^2}\right)\{(r_{xy}q_{x^2y} - q_{xy^2})\,\gamma_3'\,({}_yU_z - \gamma_3^2\,{}_yU_x) \\ - (r_{xy}q_{xy^2} - q_{x^2y})\,\gamma_3\,({}_xU_z - \gamma_3'^2\,{}_xU_y)\}\text{......(74)}.\end{aligned}$$

Let us write ${}_x\chi_z$ for ${}_xU_z - \gamma_3'^2\,{}_xU_y$ and ${}_y\chi_z$ for ${}_yU_z - \gamma_3^2\,{}_yU_x$, then (74) becomes

$$({}_x\chi_z - {}_y\chi_z)^2\,r^2_{xy} = 2\left(r_{xz}q_{x^2y} - r_{yz}q_{xy^2}\right)\{\gamma_3'(r_{xy}q_{x^2y} - q_{xy^2})\,{}_y\chi_z - \gamma_3(r_{xy}q_{xy^2} - q_{x^2y})\,{}_x\chi_z\}\text{......(75)}.$$

(75) is a relation between second order terms and it is sufficient to use equation (55) for q_{x^2y} with β_1 replaced by zero and β_2 by 3, so that

$$q_{x^2y} = \sqrt{2\,{}_xU_y},$$

$$q_{xy^2} = \sqrt{2\,{}_yU_x},$$

$$\begin{aligned}\therefore\quad ({}_x\chi_z - {}_y\chi_z)^2\,r^2_{xy} = 4\left(r_{xz}\sqrt{{}_xU_y} - r_{yz}\sqrt{{}_yU_x}\right)\{\gamma_3'\,{}_y\chi_z\left(r_{xy}\sqrt{{}_xU_y} - \sqrt{{}_yU_x}\right) \\ - \gamma_3\,{}_x\chi_z\left(r_{xy}\sqrt{{}_yU_x} - \sqrt{{}_xU_y}\right)\}\text{......(76)}.\end{aligned}$$

This identity between ${}_x\eta_z$, ${}_y\eta_z$, ${}_x\eta_y$ and ${}_y\eta_x$ is symmetrical in x, y. Two more such identities may be obtained by interchanging the letters x, y, z in cyclic order*. There are therefore three identities between the six correlation ratios:

$${}_y\eta_x,\quad {}_x\eta_y,\quad {}_y\eta_z,\quad {}_z\eta_y,\quad {}_x\eta_z,\quad {}_z\eta_x.$$

I have not so far succeeded in reducing them to simpler forms, although possibly such exist. In special cases simplifications result. These are illustrated in the following section.

§ 9. We defined γ_3, γ_3' the regression coefficients of z on x, y by the equations

$$\gamma_3 = \frac{r_{xz} - r_{yz}r_{xy}}{1 - r^2_{xy}}\text{..................(9)},\qquad \gamma_3' = \frac{r_{yz} - r_{xz}r_{xy}}{1 - r^2_{xy}}\text{..................(10)}.$$

Let us now introduce the remaining regression coefficients

$$\left.\begin{aligned}\gamma_1 &= \frac{r_{yx} - r_{zx}r_{yz}}{1 - r^2_{yz}}, & \gamma_1' &= \frac{r_{zx} - r_{yx}r_{yz}}{1 - r^2_{yz}} \\ \gamma_2 &= \frac{r_{yz} - r_{xz}r_{xy}}{1 - r^2_{xz}}, & \gamma_2' &= \frac{r_{xy} - r_{yz}r_{zx}}{1 - r^2_{zx}}\end{aligned}\right\}\text{..................(77)}.$$

It will simplify the algebra if we use $\lambda^2, \mu^2, \nu^2, \lambda'^2, \mu'^2, \nu'^2$ for ${}_yU_x$, ${}_zU_y$, ${}_xU_z$, ${}_zU_x$, ${}_xU_y$, ${}_yU_z$ respectively and P, Q, R, P', Q', R' for ${}_y\chi_x$, ${}_z\chi_y$, ${}_x\chi_z$, ${}_z\chi_x$, ${}_x\chi_y$, ${}_y\chi_z$ respectively so that

$$\left.\begin{aligned}P &= \lambda^2 - \gamma_1'^2\nu'^2, & Q &= \mu^2 - \gamma_2'^2\lambda'^2, & R &= \nu^2 - \gamma_3'^2\mu'^2 \\ P' &= \lambda'^2 - \gamma_1^2\mu^2, & Q' &= \mu'^2 - \gamma_2^2\nu^2, & R' &= \nu'^2 - \gamma_3^2\lambda^2\end{aligned}\right\}\text{.........(78)}.$$

---

* We are supposing here that the regression surfaces of x on y, z and of y on z, x are also hyperboloids of type similar to (35).

The three identities connecting the six simple η's become in this notation

$$(R-R')^2 r^2_{xy} = 4(r_{xz}\mu' - r_{yz}\lambda)\{\gamma_3' R'(r_{xy}\mu' - \lambda) - \gamma_3 R(r_{xy}\lambda - \mu')\} \;\ldots(79),$$

$$(P-P')^2 r^2_{yz} = 4(r_{yx}\nu' - r_{zx}\mu)\{\gamma_1' P'(r_{yz}\nu' - \mu) - \gamma_1 P(r_{yz}\mu - \nu')\} \;\ldots(80),$$

$$(Q-Q')^2 r^2_{zx} = 4(r_{zx}\lambda' - r_{xy}\nu)\{\gamma_2' Q'(r_{zx}\lambda' - \nu) - \gamma_2 Q(r_{zx}\nu - \lambda')\} \;\ldots(81).$$

(i) We can deduce from (79) that if the correlations of both x and y on z be linear and equal, then the correlation ratio of x on y and y on x, i.e. ${}_x\eta_y$ and ${}_y\eta_x$ are equal. Thus in biparental correlation, if the regression of the child on each parent be linear, then the correlation ratio of the father on the mother is equal to that of the mother on the father.

For under the conditions stated

$${}_x\eta_z = {}_y\eta_z = r_{zx} = r_{zy}.$$

Hence $\gamma_3 = \gamma_3' = \dfrac{r_{zx}}{1+r_{xy}}$ and (79) becomes

$$\gamma_3^4[\lambda^2 - \mu'^2]^2 r^2_{xy} = 4r_{yz}(\mu' - \lambda)\{-\gamma_3^3\lambda^2(r_{xy}\mu' - \lambda) + \gamma_3^3\mu'^2(r_{xy}\lambda - \mu')\},$$

or
$$\gamma_3(\lambda - \mu')^2(\lambda + \mu')^2 r^2_{xy} = 4r_{yz}(\lambda - \mu')^2\{r_{xy}\lambda\mu' - \lambda^2 - \lambda\mu' - \mu'^2\},$$

or
$$(\lambda - \mu')^2\left[\frac{r^2{}_{xy}}{1+r_{xy}}(\lambda + \mu')^2 - 4(r_{xy}\lambda\mu' - \lambda\mu' - \lambda^2 - \mu'^2)\right] = 0,$$

which reduces to

$$(\lambda - \mu')^2\left\{\frac{[(r_{xy}+2)^2\lambda + (2 - r^2_{xy})\mu']^2 + 4\mu'^2(2r_{xy}+3)(r_{xy}+1)^2}{(r_{xy}+1)(r_{xy}+2)^2}\right\} = 0.$$

But r_{xy} is numerically < 1. $\therefore$ the factor in curved brackets is positive. Hence $\lambda = \mu'$, i.e. ${}_yU_x = {}_xU_y$. $\therefore$ ${}_y\eta_x = {}_x\eta_y$.

(ii) An interesting deduction from the identities (79)—(81) is the following: "If any four of the six regression lines that occur in the mutual variation of three variables are linear, so are the other two."

We have to prove that if any four of the six quantities $\lambda, \mu, \nu, \lambda', \mu', \nu'$ vanish, then the remaining two vanish as well.

First let $\lambda = \mu = \nu = \lambda' = 0$.

(79) gives $\qquad r_{xy}^2(\gamma_3'\mu'^2 + \nu'^2)^2 = 4r_{xz}\mu'\{\gamma_3'\nu'^2 r_{xy}\mu' - \gamma_3\gamma_3'\mu'^3\},$

(80) $\qquad \gamma_1'^4\nu'^4 r_{yz}^2 = -4r_{xy}\nu'^4\gamma_1\gamma_1'^2,$

(81) $\qquad \mu'^4 r_{xz}^2 = 0.$

From (80) $\qquad \gamma_1'^2\nu'^4[4\gamma_1 r_{yx} + \gamma_1'^2 r_{yz}^2] = 0.$

But $\qquad 4\gamma_1 r_{xy} + \gamma_1'^2 r^2_{yz} \equiv \left(\dfrac{r_{yz}r_{xz} + r_{yz}^2 r_{xy} - 2r_{xy}}{1 - r_{yz}}\right)^2 > 0.$

$$\therefore \quad \nu' = 0,$$

and $\qquad \mu' = 0,$

and these satisfy (79).

There are three cases of this type.

The case $\lambda' = \mu' = \nu' = \lambda = 0$ is proved in the same way. There are three cases of this type as well.

Next take the three cases of type

$$\lambda = \mu = 0,$$
$$\lambda' = \mu' = 0.$$

Equations (79)—(81) become

(79) $$R = R', \text{ i.e. } \nu^2 = \nu'^2,$$

(80) $$\gamma_1'^4 \nu'^4 r_{yz}^2 = 4 r_{xy} \nu' \{\gamma_1 \gamma_1'^2 \nu'^2 (-\nu')\},$$

(81) $$\gamma_2^4 \nu^4 r_{xz}^2 = 4(-r_{xy}\nu)\{-\gamma_2' \gamma_2^2 \nu^2 (-\nu)\},$$

(80) leads to $$(\gamma_1'^2 r_{yz}^2 + 4 r_{xy} \gamma_1) \gamma_1'^2 \nu'^4 = 0.$$

We have already shown that the first factor is positive,

$$\therefore \quad \nu' = 0,$$

and hence $$\nu = 0,$$

and these values satisfy (81).

The three cases of type $\lambda = \mu = \mu' = \nu' = 0$ lead to

(79) $$\nu^4 r^2_{xy} = 0,$$

(80) $$\lambda'^4 r^2_{yz} = 0,$$

(81) $$(\gamma_2^2 \nu^2 - \gamma_2'^2 \lambda'^2) r_{yz}^2 = 4(r_{yz}\lambda' - r_{xy}\nu)\{-\gamma_2' \gamma_2^2 \nu^2 (r_{xz}\lambda' - \nu) + \gamma_2 \gamma_2'^2 \lambda'^2 (r_{xz}\nu - \lambda')\},$$

whence $\lambda' = \nu = 0$.

There remain the three cases of type

$$\lambda = \nu = 0,$$
$$\mu' = \nu' = 0.$$

Here (79) is satisfied identically.

(80) becomes $$(\lambda'^2 - \gamma_1^2 \mu^2)^2 r^2_{yz} = -4 r_{yz} \mu \{\gamma_1' (\lambda'^2 - \gamma_1^2 \mu^2)(-\mu)\},$$

(81) $$(\mu^2 - \gamma_2'^2 \lambda'^2)^2 r^2_{xz} = 4 r_{yz} \lambda' \{-\gamma_2 (\mu^2 - \gamma_2' \lambda'^2)(-\lambda')\},$$

which reduce to

$$(\lambda'^2 - \gamma_1^2 \mu^2)\{r^2_{yz} \lambda'^2 - (\gamma_1^2 r^2_{yz} + 4\gamma_1' r_{xz}) \mu^2\} = 0,$$
$$(\mu^2 - \gamma_2'^2 \lambda'^2)\{r^2_{xz} \mu^2 - (\gamma_2'^2 r^2_{zx} + 4\gamma_2 r_{yz}) \lambda'^2\} = 0.$$

The only common solution of these equations is

$$\lambda' = \mu = 0.$$

We have thus accounted for all the fifteen possible cases.

(iii) *Three regression curves linear.*

In six cases out of the possible 20 cases the linearity of three only of the regression curves involves the linearity of the remaining three.

Let $\lambda = \mu' = 0$ *and either* ν *or* $\nu' = 0$.

It follows from (79) that $R = R'$, i.e. $\nu = \nu'$.

$\therefore$ both ν and ν' are zero. We have now four linear regression curves, $\therefore$ all six are linear.

Let $\mu = \nu' = 0$ *and* $\mu' = 0$.

Since $\mu = \nu' = 0$, $\therefore$ $P = P'$ or $\lambda = \lambda'$, so that

$$\begin{aligned} P &= \lambda^2, & Q &= -\gamma_2'^2\lambda^2, & R &= \nu^2, \\ P' &= \lambda^2, & Q' &= \mu'^2 - \gamma_2^2\nu^2, & R' &= -\gamma_3^2\lambda^2. \end{aligned}$$

(79) becomes $(\nu^2 + \gamma_3^2\lambda^2)^2\, r^2_{xy} = 4r_{yz}\gamma_3\lambda^2\,(r_{xy}\nu^2 - \gamma_3\gamma_3'\lambda^2)$,

(80) becomes

$$(\gamma_2^2\nu^2 - \gamma_2'^2\lambda^2)^2\, r^2_{yz} = 4\,(r_{yz}\lambda - r_{xy}\nu)\,\gamma_2\gamma_2'\nu^2\,[(\gamma_2\nu - \gamma_2'\lambda) + r_{xz}\,(\gamma_2'\nu - \gamma_2\lambda)].$$

The first reduces to

$$\left\{\lambda^2\gamma_3\,\frac{(r_{xz} + r_{yz}r_{xy})\,r_{xy} - 2r_{yz}}{1 - r^2_{xy}} + r_{xy}\nu^2\right\}^2 = 0,$$

and the second to

$$(\gamma_2\nu^2 - \gamma_2'\lambda^2)^2\, r^2_{xz} + 4\gamma_2\gamma_2'\,(r_{yz}\lambda - r_{xy}\nu)^2\,\nu^2 = 0.$$

Hence either $\lambda = \nu = 0$ or there must be a very special relation between r_{xy}, r_{yz}, r_{zx}.

If instead of $\mu' = 0$ we take $\nu = 0$ we get similar results, i.e. in general the vanishing μ, ν', μ' or μ, ν', ν involves that of λ, ν, λ' or λ, μ', λ'.

This accounts for six more cases.

There are eight left. Of these six are typified by

$$\mu' = \nu' = 0 = \nu \quad \text{or} \quad \mu' = \nu' = 0 = \mu,$$

and lead to the same conclusions.

The remaining two are

$$\lambda = \mu = \nu = 0 \quad \text{or} \quad \lambda' = \mu' = \nu' = 0.$$

The first supposition, $\lambda = \mu = \nu = 0$ gives

$$\begin{aligned} P &= -\gamma_1'^2\nu'^2, & Q &= -\gamma_2'^2\lambda'^2, & R &= -\gamma_3'^2\mu'^2, \\ P' &= \lambda'^2, & Q' &= \mu'^2, & R' &= \nu'^2, \end{aligned}$$

leading to

$$(79) \qquad (\nu'^2 + \gamma_3'^2\mu'^2)^2\, r^2_{yx} = 4r_{xz}\mu'^2\gamma_3'\,\{r_{xy}\nu'^2 - \gamma_3\gamma_3'\mu'^2\},$$

$$(80) \qquad (\lambda'^2 + \gamma_1'^2\nu'^2)^2\, r^2_{yz} = 4r_{xy}\nu'^2\gamma_1'\,\{r_{yz}\lambda'^2 - \gamma_1\gamma_1'\nu'^2\}$$

$$(81) \qquad (\mu'^2 + \gamma_2'^2\lambda'^2)^2\, r^2_{xz} = 4r_{yz}\lambda'^2\gamma_2'\,\{r_{xz}\mu'^2 - \gamma_2\gamma_2'\lambda'^2\},$$

which give $\lambda' = \mu' = \nu' = 0$ or a very special condition to be satisfied by the correlation coefficients r_{xy}, r_{yz}, r_{zx}.

We may conclude then that in general the *linearity of any three of the six regression lines involves that of the remaining three.*

(iv) If the regression surface of z on x, y reduces to a plane, the regression curves of x on y and y on x reduce to straight lines.

We have as in § 7

$$\frac{\bar{z}_y}{\sigma_z} = d + a\frac{\bar{x}_y}{\sigma_x} + \frac{by}{\sigma_y} + c\frac{\bar{x}_y y}{\sigma_x \sigma_y} \qquad \text{.........................(58).}$$

But

$$\frac{\bar{x}_y}{\sigma_x} = r_{xy}\frac{y}{\sigma_y} \pm \sqrt{\frac{{}_y\eta_x^2 - r^2_{xy}}{\beta_2 - \beta_1 - 1}} \left\{\frac{y^2}{\sigma_y^2} - \sqrt{\beta_1}\frac{y}{\sigma_y} - 1\right\},$$

$$\therefore \quad \frac{\bar{z}_y}{\sigma_z} = d \pm a\sqrt{\frac{{}_y\eta_x^2 - r^2_{xy}}{\beta_2 - \beta_1 - 1}} + \frac{y}{\sigma_y}\left\{b + ar_{xy} \mp (a\sqrt{\beta_1} + c)\sqrt{\frac{{}_y\eta_x^2 - r^2_{xy}}{\beta_2 - \beta_1 - 1}}\right\}$$

$$+ \frac{y^2}{\sigma_y^2}\left\{\left(cr_{xy} \mp c\sqrt{\frac{\beta_1({}_y\eta_x^2 - r^2_{xy})}{\beta_2 - \beta_1 - 1}}\right) \pm a\sqrt{\frac{{}_y\eta_x^2 - r^2_{xy}}{\beta_2 - \beta_1 - 1}}\right\}$$

$$+ \text{ terms of higher order.}$$

Now the regression of z *on* y *for a constant* x is linear. Therefore the coefficient of y^2 is zero. To first order terms we may put $\beta_1 = 0$ and $\beta_2 = 3$, and thus

$$cr_{xy} \pm a\sqrt{\frac{{}_y\eta_x^2 - r^2_{xy}}{2}} = 0.$$

Similarly

$$cr_{xy} \pm b\sqrt{\frac{{}_x\eta_y^2 - r^2_{xy}}{2}} = 0,$$

But c vanishes when ${}_{xy}H_z^2 = {}_{xy}R_z^2$ by (52).

Hence if

$${}_{xy}H_z^2 = {}_{xy}R_z^2,$$

it follows that

$${}_x\eta_y = {}_y\eta_x = r_{xy}.$$

We thus see that if the three generalised correlation ratios ${}_{xy}H_z$, ${}_{yz}H_x$, ${}_{zx}H_y$ are equal to ${}_{xy}R_z$, ${}_{yz}R_x$, ${}_{zx}R_y$ respectively, the six correlation ratios ${}_x\eta_y$, ${}_y\eta_x$, ${}_x\eta_z$, ${}_z\eta_x$, ${}_z\eta_y$, ${}_y\eta_z$ reduce to the corresponding correlation coefficients r_{xy}, r_{zx}, r_{yz} and that the "linearity" of the three regression surfaces involves the linearity of the six regression lines.

MISCELLANEA.

I. On Spurious Values of Intra-class Correlation Coefficients arising from Disorderly Differentiation within the Classes.

By J. ARTHUR HARRIS, Ph.D. Carnegie Institution of Washington, U.S.A.

When the constants of the x and y characters of the population in r_{xy} are quite indistinguishable symmetrical tables[*] may be used, but not otherwise.

Primarily and for the most part, however, the use of symmetrical tables has been restricted to cases in which the degree of interdependence between the measures of all possible pairs[†] drawn from a considerable series of associated individuals—in short to intra-class correlations[‡]—is sought.

The dangers of spurious correlation due to the artificial symmetry of the surface is then much greater[§]. Pearson[||] long ago pointed out that when intra-class differentiation exists, for example, because of age in the case of characters determined upon the members of a fraternity, or of position on the axis in the case of serial organs, the values of r may be to some extent spurious.

In the cases considered by Pearson differentiation is an orderly phenomenon, i.e. the magnitudes under consideration increase or decrease with age, position on the axis, or some other extrinsic characteristic with such regularity that the relationship can be expressed by an equation which may be used in correcting the raw values of r.

In other cases, the problem is not so simple. Differentiation within the class may exist, but it may be difficult or impossible to arrange the individual measurements by any character outside of themselves to obtain the constants necessary for determining the true correlations from the spurious values deduced from the tables.

Illustration I. The correlation between yields of wheat in variety testing.

In variety testing, the experimenter seeks (or should seek), among other things, to determine the correlation between yields of varieties in different years. If this correlation be 0 (and regression be linear) it is clear that the yield of a variety in one year furnishes no basis for prediction

---

[*] R. Pearl, *Biometrika*, Vol. v. pp. 249—297, 1907; H. S. Jennings, *Journ. Exp. Zool.* Vol. xi. pp. 1—134, 1911; J. Arthur Harris, *Biometrika*, Vol. vii. pp. 325—328, 1910.

[†] K. Pearson and others, *Phil. Trans.*, A, Vol. cxcvii. pp. 285—379, 1901; K. Pearson and A. Barrington, *Eugenics Laboratory Memoirs*, No. V, 1909.

[‡] *Biometrika*, Vol. ix. pp. 446—472, 1913.

[§] With only one pair of measures the probability of spurious correlation is, in cautious work, very slight, for the possibility of differentiation can be easily tested by the critical comparison of the physical constants.

[||] Pearson, K., "On Homotyposis in Homologous but Differentiated Organs." *Roy. Soc. Proc.* Vol. lxxi. pp. 288—313, 1903.

concerning its yield in any subsequent year. If, on the other hand, the correlation be high, prediction from a few years' test may be made with great probability of certainty.

Given a measure of the "performance" of a series of varieties during a number of years it would at first seem quite allowable to form symmetrical tables or to use the intra-class formulae of a former paper* to determine the intra-varietal correlation, and to regard this as a satisfactory measure of the differentiation of the varieties and of the average prediction value of a year's test. Such is, however, not the case, for while there may be no orderly change in yield throughout the period under consideration, the individual years differ greatly in their average yield for all the varieties. The influence of this "disorderly differentiation" upon r is admirably shown by A. D. Hall's† table of the yield in bushels of wheat in the Rothamsted experiments.

Let b = yield in bushels per acre of any one of m varieties in any one of n years, y_1, y_2 be the "first" and the "second" years of a symmetrical intra-varietal correlation surface, v_1, v_2 be the "first" and "second" varieties of a symmetrical intra-annual correlation surface. Then $r_{b_{y_1}b_{y_2}}$ will be a (spurious) measure of the (persistent) differentiation of varieties, $r_{b_{v_1}b_{v_2}}$, a (spurious) measure of the differentiation (in the yield of all the varieties) of years. Applying formulae (v)—(ix) of *Biometrika*, Vol. IX. p. 450, to these data, I find

$$S[n(n-1)]=2128,$$

$$S[(n-1)\Sigma(b')]=83122\cdot5,\quad S[(n-1)\Sigma(b'^2)]=3483626\cdot4,$$

$$S[\Sigma(b')]^2=3610204\cdot57,\quad S[\Sigma(b'^2)]=370820\cdot13,$$

$$\bar{b}=39\cdot0613,\quad \sigma_b{}^2=111\cdot257328,$$

$$r_{b_{y_1}b_{y_2}}=-\cdot032.$$

The result is obviously spurious, for mere inspection of the entries in the table shows that some varieties regularly give heavier yields than others. The source of the spurious value is to be seen in the fact that an intra-class coefficient has been calculated from a symmetrical surface formed from classes (varieties) represented by a series of yields *differentiated by annual variations in the growing conditions*. By correcting for this source of differentiation by expressing each yield as a deviation from the mean yield of all the varieties for the particular year, i.e. $b''=b-\bar{b}_y$, where the bar denotes a mean and the subscript y that it is for all the yields of a year, I have found‡

$$r_{b''_{y_1}b''_{y_2}}=\cdot266.$$

Measuring the differentiation of years in terms of intra-annual correlation (intra-class correlation in which each class is defined by the year and its individuals are the yields of the different varieties grown), I find from Hall's table

$$S[m(m-1)]=4440,$$

$$S[(m-1)\Sigma(b')]=174129\cdot2,\quad S[(m-1)\Sigma(b'^2)]=7317531\cdot92,$$

$$S[\Sigma(b')]^2=7586436\cdot21,\quad S[\Sigma(b'^2)]=370820\cdot13,$$

$$\bar{b}=39\cdot2183,\quad \sigma_b{}^2=110\cdot017719,$$

$$r_{b_{v_1}b_{v_2}}=\cdot791.$$

Since the varieties have been shown to be differentiated, this result must also be spurious. Let $b''=b-\bar{b}_v$ where the v indicates that the mean denoted by the bar is for the yield of the

---

* *Biometrika*, Vol. IX. pp. 446—472, 1913.

† Hall, A. D., *The Book of the Rothamsted Experiments*, p. 66, 1905.

‡ *Science*, N. S. Vol. XXXVI. pp. 318—320, 1912. Probably a better method of dealing with such cases will sometime be found. So far I have not succeeded.

TABLE I.

Yield of Varieties of Wheat in Different Years.

Variety	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	Lots	$\Sigma(y)$	$\Sigma(y^2)$
Rivet (Red)	—	—	48·1	67·0	48·4	42·5	49·6	66·1	16·0	22·4	52·2	9	412·3	21263·39
White Chaff (Red)	—	—	40·6	55·1	40·2	49·5	48·4	59·0	22·8	28·1	54·5	9	398·2	18853·92
Club Wheat (Red)	36·0	45·8	47·5	59·6	46·6	47·6	49·5	61·0	23·5	16·4	43·4	11	476·9	22515·39
Golden Drop (Red), Hallett's	39·5	49·8	44·2	51·8	38·1	48·4	49·5	52·8	21·0	18·9	50·8	11	464·8	21088·28
Bole's Prolific (Red)	33·6	42·8	45·2	48·1	43·8	41·4	44·8	52·8	31·0	24·5	46·5	11	454·5	19468·23
Hardcastle (White)	—	46·5	42·0	49·6	33·9	44·0	42·1	54·0	21·5	24·4	45·6	10	403·6	17297·00
Red Rostock	37·0	—	46·3	53·8	37·4	40·0	46·4	57·0	8·5	28·4	45·8	10	400·6	17784·30
Red Langham	30·8	43·8	34·1	53·1	34·9	42·5	42·9	50·8	25·8	28·6	48·5	11	435·8	18130·66
Bristol Red	29·4	44·4	39·5	53·4	31·6	42·4	44·1	52·1	21·6	30·6	46·2	11	435·3	18240·43
Red Wonder	31·2	43·8	37·1	55·1	33·2	44·2	41·6	52·1	22·0	28·2	45·9	11	434·4	18191·20
Red Chaff (White)	32·8	37·0	35·3	48·8	34·3	43·8	41·0	—	—	—	—	7	273·0	10848·30
Browick (Red)	35·3	40·5	38·5	51·1	38·5	39·1	40·9	49·5	24·0	19·6	47·3	11	424·3	17311·37
Casey's White	29·9	42·1	37·5	52·1	39·0	45·5	43·0	47·8	15·4	24·1	42·9	11	419·3	17170·55
Red Nursery	34·1	45·3	27·1	41·1	39·0	37·5	40·6	47·8	30·9	27·5	46·0	11	416·9	16326·03
Wooly Ear (White)	31·2	42·8	37·0	51·3	36·1	46·6	37·5	48·3	20·0	21·0	44·1	11	415·9	16805·69
Burwell (Old Red Lammas)	31·1	41·3	35·1	47·3	38·5	38·4	39·0	46·3	27·0	27·0	44·8	11	415·8	16228·74
Golden Rough Chaff (Red)...	33·0	39·3	38·5	52·1	38·8	38·4	36·4	46·8	14·4	31·3	41·6	11	410·6	16242·96
Chubb Wheat (Red) ...	28·4	40·0	35·8	50·5	38·3	40·3	41·5	55·1	20·8	14·9	—	10	365·6	14742·34
Original Red (Hallett's) ...	30·0	35·3	36·4	43·6	26·0	40·1	44·4	—	—	—	—	7	255·8	9627·38
Victoria White (Hallett's) ...	33·8	45·3	38·3	44·3	33·8	41·1	42·6	43·9	14·9	15·8	44·0	11	397·8	15605·18
White Chiddam	26·9	38·8	31·8	42·0	32·4	37·5	37·6	49·8	11·9	27·4	47·1	11	383·2	14464·88
Hunter's White (Hallett's) ...	26·9	39·8	38·0	45·4	26·4	43·5	40·0	42·3	17·4	22·8	—	10	342·5	12613·91
Number of Lots	19	19	22	22	22	22	22	20	20	20	18	226	—	370820·13
Total Yields, $\Sigma(y)$	610·9	804·4	853·9	1116·2	809·2	934·3	943·4	1035·3	410·4	481·9	837·2	—	8837·1	—

variety for all the years it was grown. Correcting for the influence of the differentiation of varieties in this way I have found*

$$r_{b''_{v_1} b''_{v_2}} = \cdot 837.$$

Thus season is a far more important factor than variety in determining an individual yield.

Illustration II. Influence of Personal Equation upon the Correlation between the Grades assigned to the Same Paper by a Series of Instructors.

Stripped of the verbiage in which it has been clothed in discussions among pedagogues, one of the chief problems concerning the reliability of the grades assigned in examinations resolves itself unto the statistical question: What is the correlation between the grades assigned to the same paper by different instructors?

Let g be the grade assigned to any one of m papers by any one of n instructors, let i_1, i_2 be the "first" and "second" instructor (of a symmetrical intra-class table) passing judgment upon a paper, p_1, p_2 the "first" and "second" paper graded by the same instructor. Then from Table I of D. Starch† I deduce, by the intra-class formulae (v)—(lx) of *Biometrika*, Vol. IX. p. 450,

$$r_{g_{i_1} g_{i_2}} = \cdot 659, \quad r_{g_{p_1} g_{p_2}} = \cdot 071.$$

By using the deviation method as illustrated above, I have found

$$r_{g''_{i_1} g''_{i_2}} = \cdot 732, \quad r_{g''_{p_1} g''_{p_2}} = \cdot 386.$$

TABLE II.

Grades of Papers Assigned by Various Instructors.

Instructors.

Papers.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	$\Sigma(g)$
1	85	86	88	85	75	80	88	87	85	87	846
2	77	80	87	80	62	82	82	87	85	87	809
3	74	78	78	75	69	84	91	83	79	80	791
4	65	65	62	20	26	60	55	68	55	50	526
5	68	82	78	82	64	88	85	86	78	80	791
6	94	87	93	87	83	77	89	88	88	89	875
7	88	90	95	87	79	85	96	91	87	89	887
8	80	84	73	79	72	83	85	91	77	76	800
9	70	70	68	50	44	65	75	81	79	79	681
10	93	92	85	92	81	83	92	89	84	85	876
$\Sigma(g)$	794	814	807	737	655	787	838	851	797	802	7882

Both of these results, in which an attempt was made to correct for the personal equation of the instructors in determining the correlation between the estimates of different instructors on the same paper, or to correct for the differences in merit of the papers in testing the individuality of the instructors, are higher than the raw values given above, which are clearly spurious. Similar results‡ are obtained from Jacoby's astronomical grades§.

---

* *Science, loc. cit.*

† *Science*, N. S. Vol. XXXVIII. p. 630, 1913.

‡ Personally, I can attach little pedagogical significance to series as short as those of either Starch or Jacoby. They serve here as illustrations of method merely because I know of no more extensive series.

§ *Science*, N. S. Vol. XXXI. p. 819, 1910.

The essentials of this note may be summarized as follows:

In using Intra-class coefficients care must be taken to guard against spurious values arising through differentiation among the individuals of the class.

Besides the orderly differentiation (due to age of individuals, position of organs on axis, etc.) for which Pearson has determined corrective formulae in terms of correlation coefficients, a disorderly differentiation for which such corrective formulae have not as yet been found sometimes obtains. Illustrations of such cases are here given.

Probably the empirical methods used here in correcting for this disorderly differentiation should be replaced by formulae with a sounder theoretical foundation. This I have not as yet been able to do.

The purpose of this note will have been served if it directs attention to a source of danger which may sometimes be encountered in the use of serviceable formulae, and indicates a method by which in the absence of more perfect methods practical results may be secured.

COLD SPRING HARBOR, N.Y.
February 3, 1914.

II. On an Extension of the Method of Correlation by Grades or Ranks.

BY KARL PEARSON, F.R.S.

In a memoir published in 1907[*] I have shown how, on the hypothesis of normal distribution, the true correlation of variates r may be ascertained from the correlation ρ of grades. If g_1 and g_2 be the two grades, ν_1 and ν_2 the corresponding ranks, x and y the corresponding variates with means $\bar{x}$ and $\bar{y}$, and standard-deviations σ_1 and σ_2, while

$$z=\frac{N}{2\pi\sigma_1\sigma_2}\frac{1}{\sqrt{1-r^2}}e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{x^2}{\sigma_1^2}-\frac{2rxy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right)}$$

is the normal frequency surface of the variates, then

$$\bar{g}_1=\tfrac{1}{2}N=\bar{g}_2, \quad \sigma_{g_1}{}^2=\sigma_{g_2}{}^2=\tfrac{1}{12}N,$$

$$g_1-\bar{g}_1=i_1=\frac{N}{\sqrt{2\pi}\,\sigma_1}\int_0^x e^{+\frac{1}{2}\frac{x^2}{\sigma_1^2}}dx,$$

$$g_2-\bar{g}_2=i_2=\frac{N}{\sqrt{2\pi}\,\sigma_2}\int_0^y e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}}dy,$$

$$\nu_1=\nu_2=g+\tfrac{1}{2}, \quad \bar{\nu}_1=\bar{\nu}_2=\tfrac{1}{2}(N+1),$$

$$\sigma^2{}_{\nu_1}=\sigma^2{}_{\nu_2}=\tfrac{1}{12}(N^2-1).$$

Further I showed in the memoir just cited that

$$r=2\sin\left(\frac{\pi}{6}\rho\right)$$

---

[*] "On Further Methods of Determining Correlation," *Drapers' Company Research Memoirs* (Dulau and Co.), pp. 11, 12.

where a convenient method of finding ρ was by the formula

$$\rho = 1 - \frac{6S(g_1 - g_2)^2}{N^3}$$

or again by

$$= 1 - \frac{6S(\nu_1 - \nu_2)^2}{N(N^2 - 1)}.$$

The problem has recently occurred of dealing with data where:

(i) One variate is given quantitatively, the other variate is given by ranks.

For example, place in school-class has to be considered in relation to marks in examination, or the rank in a teacher's general appreciation has to be considered in relation to marks in examination.

(ii) One variate is given by broad categories, the other by ranks.

For example, five or six categories of general intelligence are given as the basis of the teacher's classification of intelligence, and this has to be considered with regard to rank in, say, class or examination, possibly with regard to a special subject.

We require in both cases to deduce from the data the true variate correlation.

Case (i). Let x be the character measured by its grade, y the character given quantitatively. Then with the notation above, if ρ' equal the correlation of grade and of variate, r the correlation of the two variates:

$$\rho' = \frac{p_{g,y}}{N\sigma_g \sigma_y},$$

where

$$p_{g,y} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} z\,(y - \bar{y})(g_1 - \bar{g}_1)\,dxdy,$$

$$\frac{dp_{g,y}}{dr} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y - \bar{y})\,i_1\,\frac{dz}{dr}\,dxdy.$$

Integrating by parts after putting $\bar{y} = 0$ and writing

$$\frac{dz}{dr} = \sigma_1\sigma_2\frac{d^2z}{dx\,dy},^*$$

$$\frac{dp_{g,y}}{dr} = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \sigma_1\sigma_2 y\,\frac{di_1}{dx}\,\frac{dz}{dy}\,dx\,dy.$$

Integrating again by parts:

$$\frac{dp_{g,y}}{dr} = \sigma_1\sigma_2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{di_1}{dx}\,z\,dx\,dy$$

$$= \sigma_1\sigma_2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{N}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\frac{x^2}{\sigma_1^2}}\,z\,dx\,dy$$

$$= \frac{\sigma_2 N^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{2\pi}\,\frac{1}{\sqrt{1 - r^2}}\,e^{-\frac{1}{2}\left(\frac{2 - r^2}{1 - r^2}x'^2 - \frac{2rx'y'}{1 - r^2} + \frac{y'^2}{1 - r^2}\right)}\,dx'\,dy'$$

$$= \frac{\sigma_2 N^2}{\sqrt{2\pi}\sqrt{1 - r^2}}\,\frac{1}{\sqrt{\frac{2 - r^2}{(1 - r^2)^2} - \frac{r^2}{(1 - r^2)^2}}} = \frac{\sigma_2 N^2}{2\sqrt{\pi}}.$$

Hence

$$\frac{d\rho'}{dr} = \frac{d}{dr}\left(\frac{p_{g,y}}{N\sigma_2\sigma_{g_1}}\right) = \frac{N}{2\sqrt{\pi}\,\sigma_{g_1}} = \sqrt{\quad}.$$

---

* *Phil. Trans.* A., Vol. 195, p. 25.

Thus since ρ' vanishes with r,

$$\rho' = \sqrt{\frac{3}{\pi}}\, r.$$

If we used ranks instead of grades, we should have σ_{ν_1} for σ_{g_1} and consequently

$$\rho'' = \sqrt{\frac{3}{\pi}}\, \frac{N}{\sqrt{N^2-1}}\, r.$$

Thus finally

$$r = \sqrt{\frac{\pi}{3}}\, \rho' = 1{\cdot}0233\, \rho',$$

or

$$r = \frac{\sqrt{N^2-1}}{N}\, 1{\cdot}0233\, \rho''.$$

It will be clear from this that the correlation ρ' between rank and quantitative variate can never be "perfect," for it cannot exceed the value ·9772, otherwise the correlation r would exceed unity. It will be seen that for practical purposes r is very close to ρ', but still from the theoretical standpoint, it is not without interest to discover that the correlation between ranks and a quantitative variate can never be perfect. For example, it is impossible to have perfect correlation between place in class and examination test, even if the boys were in the same order in class and examination. The defect, however, will be very slight.

Case (ii). Let the subscript C refer to any "broad" class and let η be found from either of the formulae

$$\eta^2 = \frac{12}{N^3}\, S\,\{n_c\,(\bar{g}_c - \bar{g})^2\},$$

or

$$\eta'^2 = \frac{12}{N\,(N^2-1)}\, S\,\{n_c\,(\bar{\nu}_c - \bar{\nu})^2\},$$

the first applying to grades and the second to ranks; then

$$r = 1{\cdot}0233\, \sqrt{\frac{12}{N^3}\, S\,\{n_c\,(\bar{g}_c - \bar{g})^2\}},$$

or

$$= 1{\cdot}0233\, \sqrt{\frac{12}{N^3}\, S\,\{n_c\,(\bar{\nu}_c - \bar{\nu})^2\}},$$

according as grades or ranks are used. In actual practice the values of η' or η'' should be correct for number of classes and for "broad" categories. See *Biometrika*, Vol. VIII. p. 256 and Vol. IX. p. 118.

Numerical illustrations will be provided later.

III. Correction of a Misstatement by Mr Major Greenwood, Junior.

In a recent paper by Mr Major Greenwood and Mrs Frances Wood "On changes in the Recorded Mortality from Cancer and their Possible Interpretation*" occur the following words: "The case is evidently analogous to that studied by Professor Karl Pearson in his pamphlet, *The Fight against Tuberculosis and the Death-rate from Phthisis* (Dulau and Co.). Professor Pearson published three diagrams: (*a*) the general death-rate of England and Wales; (*b*) the phthisis death-rate; (*c*) the ratio of phthisis deaths to all deaths. The original figures seem to have been the crude rate for males and females separately from 1835 onwards." The "evident analogy" with what appears to me the wholly fallacious treatment of the authors in their paper above cited I do not now stay to discuss, but I wish to draw attention to the words: "The

---

* Royal Society of Medicine, *Proceedings*, Vol. VII. Section of Epidemiology, pp. 79—170. March 27, 1914.

original figures seem to have been the crude rate for males and females separately from 1835 onwards." Why the writer of these words should have assumed them without any inquiry of me, or any examination of the values of the crude death-rates (which are accessible to everybody) to be "crude death-rates," I do not know, but they illustrate his readiness to form a biased judgment when his feelings are stirred by unfavourable criticism. As a matter of fact the rates were *standardised* rates reduced to the population of 1901, and most kindly provided at my special request by the General Register Office. It is of interest to observe that Dr Weinberg of Stuttgart—recently made precisely the same charge as Mr Major Greenwood with the same over-hasty assumption that the reality must be the desired, if undemonstrated, error*. With the German as with other foes, it is well to leave ample opportunity for their assuming you to be foolish ; their assumption may lead them to run against hard reality.

K. P.

IV. Note on Reproductive Selection.

By DAVID HERON, D.Sc.

The fact that in the case of man† fifty per cent. of one generation comes from twenty-five per cent. of the preceding one was first noted by Professor Karl Pearson in the *Chances of Death* (Vol. I. p. 80) and in dealing more fully with this important generalisation in the *Groundwork of Eugenics*, p. 27, he said : "It is very difficult from any English statistics to determine how many adults never marry. No information on this point is asked in the death schedule for males ; it is asked but imperfectly answered in the case of the schedule for females." In a footnote he adds : "The Registrar-General informs me that the record of civil condition in the case of female deaths is worthless and that no useful return can be made from it." He found that in the Argentine and in Scotland 60 per cent. died unmarried, in the United States 51 per cent., and from the last two English Censuses and the Annual Reports 48 per cent., and added "This indirect method of reaching the result is, however, not very satisfactory. We may, I think, conclude in round numbers that 40 per cent. of the population dies before it reaches the age of 21 and that probably another 20 per cent. are never married." On this assumption Professor Pearson proceeds to show that "about 12 per cent. of all the individuals born in the last generation provide half the next generation."

Some data published in *Bulletin of Population and Vital Statistics No.* 30 *for the Commonwealth of Australia* (Tables 48 and 84 *a* and *b*) prove that the assumptions made lie very close to the facts. The data are shown in the following table which gives the conjugal condition and issue of the males and females who died in Australia in 1912. From this we find that half the total number of children came from 3337 of the parents (all those who had at least 9 children and part of those who had each 8 children). It thus appears that of the males 17,404 out of 30,285 = 57·5 °/₀ died unmarried while half the total offspring came from 25·9 °/₀ of those who married and 11·0 °/₀ of the whole number of males, so that approximately three-fifths of the males born die unmarried and one-half of one generation comes from one-quarter of the married population or from one-ninth of all the males born in the preceding generation. The diagram gives a graphical illustration of the argument.

In exactly the same way we find that nearly one-half of the females born in Australia die unmarried and that one-half of one generation comes from one-quarter of the married and from one-seventh of all the females born in the preceding generation.

---

* *Archiv für Rassen- und Gesellschafts-Biologie*, IX. Jahrgang, S. 87. Leipzig, 1912.

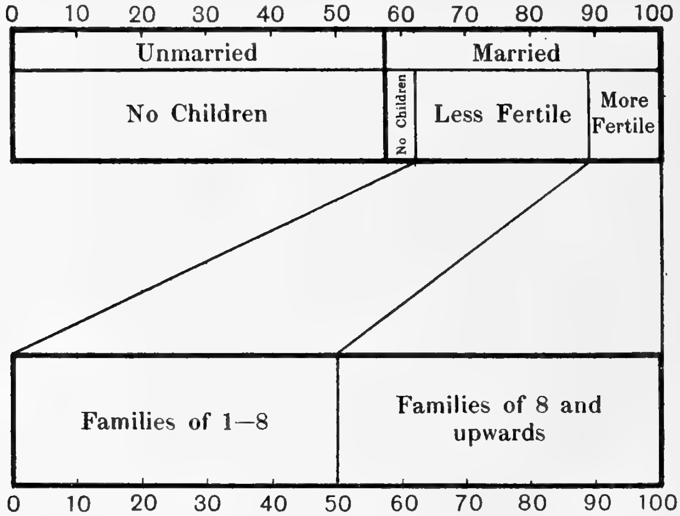
† It has also been dealt with in various mammals. See the *Groundwork of Eugenics*, Eugenics Lecture Series II (Dulau and Co.), p. 29.

Conjugal Condition	Children in Each Family	Deaths of Males	Total Children	Deaths of Females	Total Children
Single	*0*	17404	—	10011	—
Married	*0*	1422	—	1317	—
”	*1*	1036	1036	1083	1083
”	*2*	1098	2196	992	1984
”	*3*	1127	3381	1050	3150
”	*4*	1147	4588	1001	4004
”	*5*	1070	5350	976	4880
”	*6*	1058	6348	1013	6078
”	*7*	1040	7280	974	6818
”	*8*	992	7936	881	7048
”	*9*	819	7371	799	7191
”	*10*	801	8010	622	6220
”	*11*	473	5203	469	5159
”	*12*	394	4728	314	3768
”	*13*	196	2548	193	2509
”	*14*	109	1526	101	1414
”	*15*	50	750	57	855
”	*16*	27	432	22	352
”	*17*	7	119	8	136
”	*18*	5	90	3	54
”	*19*	6	114	2	38
”	*20*	3	60	2	40
”	*21*	—	—	1	21
”	*22*	—	—	1	22
”	*23*	1	23	—	—
	Totals	30285	69089	21892	62824

Diagram to illustrate the fact that three-fifths of those born die unmarried and that one-ninth of one generation produce one-half of the next. (Deduced from records of males.)

First Generation.

Percentages.



Percentages.

Second Generation.

Journal of Anatomy and Physiology.

CONDUCTED BY

SIR WILLIAM TURNER, K.C.B. ARTHUR THOMSON, University of Oxford
ALEX. MACALISTER, University of Cambridge ARTHUR KEITH, Royal College of Surgeons
ARTHUR ROBINSON, University of Edinburgh

VOL. XLIX

ANNUAL SUBSCRIPTION 21/- POST FREE

CONTENTS OF PART I.—OCTOBER 1914

H. L. Barniville, B.A., M.B. The Morphology and Histology of a Human Embryo of 8·5 mm. J. Wilfrid Jackson, F.G.S. Dental Mutilations in Neolithic Human Remains. J. S. B. Stopford, M.B., Ch.B. (Manch.). The Supracondyloid Tubercles of the Femur and the Attachment of the Gastrocnemius Muscle to the Femoral Diaphysis. J. A. Pires de Lima. Note on a Case of Bifid Penis, with Penial Hypospadia. David Waterston, M.D. A Human Embryo of Twenty-seven Pairs of Somites, Embedded in Decidua.

LONDON: CHARLES GRIFFIN and COMPANY, Ltd., Exeter Street, Strand

---

JOURNAL OF THE ROYAL ANTHROPOLOGICAL INSTITUTE

Vol. XLIV. January—June, 1914

Contents:—

Minutes of the Annual General Meeting, January 20th. Presidential Address. The Reconstruction of Fossil Human Skulls. Balfour, Henry. Frictional Fire-making with a Flexible Sawing-thong. Bates, Daisy M., F.R.A.S. (Australia). A Few Notes on Some South Western Australian Dialects. Knowles, W. J., M.R.I.A. The Antiquity of Man in Ireland, being an Account of the Older Series of Irish Flint Implements. Berry, Richard J. A., M.D. (Edin. and Melb.), F.R.S. (Edin.), Robertson, A. W. D., M.D. (Melb.), and Büchner, L. W. G. The Craniometry of the Tasmanian Aboriginal. Best, Elsdon. Ceremonial Performances Pertaining to Birth; as Performed by the Maori of New Zealand in Past times. Ivens, Rev. W. G. Natives' Stories from Ulawa. Basedon, Herbert, M.A., M.D., B.Sc., F.G.S., etc. Aboriginal Rock Carvings of Great Antiquity in South Australia. (With Plates I—XVII.) Cook, W. H. On the Discovery of a Human Skeleton in a Brick-Earth Deposit in the Valley of the River Medway at Halling, Kent. (With Plates XVIII—XXII.) Keith, Arthur, M.D. Report on the Human and Animal Remains found at Halling, Kent.

WITH TWENTY-TWO PLATES AND MANY ILLUSTRATIONS IN THE TEXT.

PRICE 15s. NET

LONDON: THE ROYAL ANTHROPOLOGICAL INSTITUTE, 50, Great Russell Street, W.C.
or through any Bookseller

MAN

A MONTHLY RECORD OF ANTHROPOLOGICAL SCIENCE

Published under the direction of the Royal Anthropological Institute of Great Britain and Ireland. Each number of **MAN** consists of at least 16 Imp. 8vo. pages, with illustrations in the text together with one full-page plate; and includes Original Articles, Notes, and Correspondence; Reviews and Summaries; Reports of Meetings; and Descriptive Notices of the Acquisitions of Museums and Private Collections.

Price, 1s. Monthly or 10s. per Annum prepaid.

TO BE OBTAINED FROM THE
ROYAL ANTHROPOLOGICAL INSTITUTE, 50, Great Russell Street, W.C.
AND THROUGH ALL BOOKSELLERS

CONTENTS

The publication of a paper in *Biometrika* marks that in the Editor's opinion it contains either in method or material something of interest to biometricians. But the Editor desires it to be distinctly understood that such publication does not mark assent to the arguments used or to the conclusions drawn in the paper.

Biometrika appears about four times a year. A volume containing about 500 pages, with plates and tables, is issued annually.

Papers for publication and books and offprints for notice should be sent to Professor Karl Pearson, University College, London. It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted elsewhere without leave of the Editor. It is very desirable that a copy of all measurements made, not necessarily for publication, should accompany each manuscript. In all cases the papers themselves should contain not only the calculated constants, but the distributions from which they have been deduced. Diagrams and drawings should be sent in a state suitable for direct photographic reproduction, and if on decimal paper it should be *blue* ruled, and the lettering only pencilled.

Papers will be accepted in German, French or Italian. In the first case the manuscript should be in Roman not German characters.

Contributors receive 25 copies of their papers free. Fifty additional copies may be had on payment of 7/- per sheet of eight pages, or part of a sheet of eight pages, with an extra charge for Plates; these should be ordered when the final proof is returned.

The subscription price, payable in advance, is 30*s. net* per volume (post free); single numbers 10*s. net*. Volumes I, II, III, IV, V, VI, VII, VIII and IX (1902—13) complete, 30*s. net* per volume. Bound in Buckram 34/6 *net* per volume. Index to Volumes I to V, 2*s. net*. Subscriptions may be sent to C. F. Clay, Cambridge University Press, Fetter Lane, London, E.C., either direct or through any bookseller, and communications respecting advertisements should also be addressed to C. F. Clay.

Till further notice, new subscribers to *Biometrika* may obtain Vols. I—IX together for £10 *net*—or bound in Buckram for £12 *net*.

The Cambridge University Press has appointed the University of Chicago Press Agents for the sale of *Biometrika* in the United States of America, and has authorised them to charge the following prices:—$7.50 net per volume; single parts $2.50 net each.

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS.

Vol. X. Part IV **May, 1915**

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF
BIOLOGICAL PROBLEMS

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY
KARL PEARSON

CAMBRIDGE UNIVERSITY PRESS
C. F. CLAY, MANAGER
LONDON: FETTER LANE, E.C.
EDINBURGH: 100, PRINCES STREET
also
H. K. LEWIS, 136, GOWER STREET, LONDON, W.C.
WILLIAM WESLEY AND SON, 28, ESSEX STREET, LONDON, W.C.
CHICAGO: UNIVERSITY OF CHICAGO PRESS
BOMBAY, CALCUTTA AND MADRAS: MACMILLAN AND CO., LIMITED
TORONTO: J. M. DENT AND SONS, LIMITED
TOKYO: THE MARUZEN-KABUSHIKI-KAISHA

Price Ten Shillings net

[*Issued June* 4, 1915]

BIOMETRIC LABORATORY PUBLICATIONS

Drapers' Company Research Memoirs.

AFTER JUNE 30, THE CAMBRIDGE UNIVERSITY PRESS AND THEIR AGENTS WILL BE THE SOLE AGENTS FOR THE SALE OF THESE PUBLICATIONS.

Biometric Series.

I. **Mathematical Contributions to the** Theory of Evolution.—XIII. On the Theory of Contingency and its Relation to Association and Normal Correlation. By Karl Pearson, F.R.S. *Price 4s. net.*

II. **Mathematical Contributions to the** Theory of Evolution.—XIV. On the Theory of Skew Correlation and Non-linear Regression. By Karl Pearson, F.R.S. *Price 5s. net.*

III. **Mathematical Contributions to the** Theory of Evolution.—XV. On the Mathematical Theory of Random Migration. By Karl Pearson, F.R.S., with the assistance of John Blakeman, M.Sc. *Price 5s. net.*

IV. **Mathematical Contributions to the** Theory of Evolution.—XVI. On Further Methods of Measuring Correlation. By Karl Pearson, F.R.S. *Price 4s. net.*

V. **Mathematical Contributions to the** Theory of Evolution.—XVII. On Homotyposis in the Animal Kingdom. By Ernest Warren, D.Sc., Alice Lee, D.Sc., Edna Lea-Smith, Marion Radford, and Karl Pearson, F.R.S. [*Shortly.*

VI. **Albinism in Man.** By Karl Pearson, E. Nettleship, and C. H. Usher. Text, Part I, and Atlas, Part I. *Price 35s. net.*

VII. **Mathematical Contributions to the** Theory of Evolution.—XVIII. On a Novel Method of Regarding the Association of two Variates classed solely in Alternative Categories. By Karl Pearson, F.R.S. *Price 4s. net.*

VIII. **Albinism in Man.** By Karl Pearson, E. Nettleship, and C. H. Usher. Text, Part II, and Atlas, Part II. *Price 30s. net.*

IX. **Albinism in Man.** By Karl Pearson, E. Nettleship, and C. H. Usher. Text, Part IV, and Atlas, Part IV. *Price 21s. net.*

Studies in National Deterioration.

I. **On the Relation of Fertility in Man** to Social Status, and on the changes in this Relation that have taken place in the last 50 years. By David Heron, M.A., D.Sc. *Price 3s.* Sold only with complete sets.

II. **A First Study of the Statistics of** Pulmonary Tuberculosis (Inheritance). By Karl Pearson, F.R.S. *Price 3s. net.*

III. **A Second Study of the Statistics of** Pulmonary Tuberculosis. Marital Infection. By Ernest G. Pope, revised by Karl Pearson, F.R.S. With an Appendix on Assortative Mating by Ethel M. Elderton. *Price 3s. net.*

IV. **The Health of the School-Child in relation** to its Mental Characters. By Karl Pearson, F.R.S. [*Shortly.*

V. **On the Inheritance of the Diathesis** of Phthisis and Insanity. A Statistical Study based upon the Family History of 1,500 Criminals. By Charles Goring, M.D., B.Sc. *Price 3s. net.*

VI. **A Third Study of the Statistics of** Pulmonary Tuberculosis. The Mortality of the Tuberculous and Sanatorium Treatment. By W. P. Elderton, F.I.A., and S. J. Perry, A.I.A. *Price 3s. net.*

VII. **On the Intensity of Natural Selection** in Man. (On the Relation of Darwinism to the Infantile Death-rate.) By E. C. Snow, D.Sc. *Price 3s. net.*

VIII. **A Fourth Study of the Statistics of** Pulmonary Tuberculosis: the Mortality of the Tuberculous: Sanatorium and Tuberculin Treatment. By W. Palin Elderton, F.I.A., and Sidney J. Perry, A.I.A. *Price 3s. net.*

IX. **A Statistical Study of Oral Temperatures** in School Children with special reference to Parental, Environmental and Class Differences. By M. H. Williams, M.B., Julia Bell, M.A., and Karl Pearson, F.R.S. *Price 6s. net.*

Questions of the Day and of the Fray.

I. **The Influence of Parental Alcoholism** on the Physique and Ability of the Offspring. A Reply to the Cambridge Economists. By Karl Pearson, F.R.S. *Price 1s. net.*

II. **Mental Defect, Mal-Nutrition, and** the Teacher's Appreciation of Intelligence. A Reply to Criticisms of the Memoir on 'The Influence of Defective Physique and Unfavourable Home Environment on the Intelligence of School Children.' By David Heron, D.Sc. *Price 1s. net.*

III. **An Attempt to correct some of the** Misstatements made by Sir Victor Horsley, F.R.S., F.R.C.S., and Mary D. Sturge, M.D., in their Criticisms of the Memoir: 'A First Study of the Influence of Parental Alcoholism,' &c. By Karl Pearson, F.R.S. *Price 1s. net.*

IV. **The Fight against Tuberculosis and** the Death-rate from Phthisis. By Karl Pearson, F.R.S. *Price 1s. net.*

V. **Social Problems: Their Treatment,** Past, Present and Future. By Karl Pearson, F.R.S. *Price 1s. net.*

VI. **Eugenics and Public Health.** Lecture to the York Congress of the Royal Sanitary Institute. By Karl Pearson, F.R.S. *Price 1s. net.*

VII. **Mendelism and the Problem of Mental** Defect. I. A Criticism of Recent American Work. By David Heron, D.Sc. (Double Number.) *Price 2s. net.*

VIII. **Mendelism and the Problem of Mental** Defect. II. The Continuity of Mental Defect. By Karl Pearson, F.R.S., and Gustav A. Jaederholm. *Price 1s. net.*

IX. **Mendelism and the Problem of Mental Defect.** III. On the Graduated Character of Mental Defect, and on the need for standardizing Judgments as to the Grade of Social Inefficiency which shall involve Segregation. By Karl Pearson, F.R.S. (Double Number.) *Price 2s. net.*

VOLUME X MAY, 1915 No. 4

ASSOCIATION OF FINGER-PRINTS.

BY H. WAITE, M.A., B.SC.

1. *Introduction.* Certain papers have been published in recent years giving the results of research on the variability and correlation of the hand, notably (1) "A First Study of the Variability and Correlation of the Hand," by Miss M. A. Whiteley, B.Sc., and Karl Pearson, F.R.S., *Proceedings of the Royal Society,* Vol. 65, pp. 126—151, and (2) "A Second Study of the Variability and Correlation of the Hand," by M. A. Lewenz, B.A., and M. A. Whiteley, B.Sc., *Biometrika,* Vol. I, pp. 345—360. In the former the writers urge "the importance of putting on record all the quantitative measures we can possibly ascertain of variability and correlation" of characters of the human body. Although Finger-Prints, the characters dealt with in the present paper, cannot strictly claim to be quantitative it is hoped by the writer that the results may prove of some interest and use in the solution of the great Problem of Evolution in Man, especially when compared with the results obtained from the study of other measurements of the hand.

The principal motive underlying most of the work which has been done in the past on the subject of Finger-Prints has arisen from the development of means of identification and it was based on the fact that the general pattern and characteristics of the finger-prints of any individual are persistent throughout life. As far as I am aware, however, no paper has yet been published attempting to measure the association between the various types of finger-prints in an individual or comparing these with the relations which have been found to exist between other measurements of the hand. These are the objects of the present paper.

2. *Primary Classification of Finger-Prints.* As primary classification Purkenje proposed nine types, Galton[*] three—each being divided into twenty-four sub-classes,—and Henry[†] four, these also being sub-divided into a number of classes. For the purposes of this paper I have adopted the method of dividing all the prints into four primary classes; I have also adopted Henry's definitions and

---

[*] *Fingerprint Directories,* by Francis Galton, F.R.S. Macmillan, 1895.

[†] *Classification and Uses of Finger Prints,* by Sir E. R. Henry, C.V.O., C.S.I. Wyman and Sons, Third edition, 1905.

nomenclature as far as they are required, and these follow in general those of Galton. Secondary classification with its minute details is not used in this paper.

The four classes referred to above are Arches, Loops, Whorls and Composites.

In Arches the ridges run from side to side, consecutive ridges being roughly parallel and the curvature increasing in general from the base to the tip. (Plate XX. Fig. i.)

In Loops some of the ridges are doubled back upon themselves making a half turn or a little more, the two parts of the doubled ridge diverging from each other at the centre of the pattern. (Fig. ii.) Consequently this pattern has an open mouth directed downwards either towards the right or towards the left of the finger. The direction of this opening supplies a means of subdividing Loops into Radial and Ulnar Loops according as the direction is towards the radius or towards the ulna, that is, towards or away from the thumb. As will be seen later (p. 422B) the proportion of Radial Loops is very small except in the forefinger, so that this method of subdivision has been used only in dealing with that finger.

In Whorls some of the ridges make a complete circuit, either as closed concentric ovals or as a more or less continuous ridge forming a spiral. (Fig. iii.)

Composites consist of combinations of two or more of the other patterns. (Fig. iv.) In this class are also included those finger-prints which are too irregular in general outline to be placed in any one of the other main groups.

This class also includes the bulk of those patterns about which Sir Francis Galton, in his book on *Finger Prints**, p. 79, states—"They are as much Loops as Whorls, and properly ought to be relegated to a fourth class." It is possible, however, that some of Galton's "ambiguous cases" may have been classed in this paper with Loops.

For further details of these principal classes with their modifications and subdivisions reference may be made to the works mentioned in the footnotes on p. 421.

3. *Material.* The material on which this investigation is based consists of two thousand complete sets of finger-prints of adult males, part of a much longer series in the Biometric Laboratory of University College, London. They belong to the lower type of artisan and labouring classes. No selection whatever has been made, except that a few sets, which were incomplete or which contained prints so damaged as to be indecipherable, have been rejected.

4. *Symbols.* The following symbols are used:—A = Arch, SL = Small Loop; LL = Large Loop (see p. 423); W = Whorl; C = Composite; L_r = Radial Loop; L_u = Ulnar Loop; R = Right Hand; L = Left Hand. R_1, R_2, R_3, R_4, R_5 designate the thumb, forefinger, middle, ring and little finger respectively of the right hand, and L_1, L_2, L_3, L_4, L_5 represent the corresponding fingers of the left hand.

---

* *Finger Prints*, by Francis Galton, F.R.S., Macmillan, 1892.

Fig. (i). Arch.

Fig. (ii). Loop.

Fig. (iii). Whorl.

Fig. (iv). Composite.

Illustrations of the four fundamental types of Finger-Print.

5. *Distribution of Classes of Finger-Prints.* A preliminary survey of the prints brings to light a considerable clustering together of prints of the same kind. Thus, each of 241 sets contains prints of one class only; each of 329 sets has nine prints of one class, and each of 194 sets contains eight out of the ten prints of one class; that is, each of 764 sets, or over 38 $^\circ/_\circ$, has at least eight prints of one class, the large majority of these being loops. Again, each of 892 sets contains prints of two classes only, so that each of 1133 sets—or nearly 57 $^\circ/_\circ$ of the whole—has representatives of not more than two of the four classes. On the other hand all four classes appear in only 95 sets, while the number of single hands, each of which contains at least one of every class, is only 23.

For the calculations which follow it has been found advisable to subdivide the loops into two classes, Small Loops and Large Loops (p. 423). Considering these as separate classes, giving five types in all, the distribution of numbers of types for the two hands is shown in the following Table:

TABLE 1.

Distribution of Types in Right and Left Hands.

Number of Types in Right Hand.

Number of Types in Left Hand.

	1	*2*	*3*	*4*	*5*	Totals
1	37	84	47	6	—	174
2	65	465	360	61	4	955
3	15	256	347	96	2	716
4	1	36	83	30	1	151
5	—	1	2	1	—	4
Totals	118	842	839	194	7	2000

In this Table, taking as origin the cell (3, 2) containing 360 types, we have the following results:

Mean of Left Hand Types,	·428
σ_y,	·7628
Mean of Right Hand Types,	– ·435
σ_x,	·7608.

We thus find the correlation coefficient (r) to be ·281 ± ·014.

The contingency coefficient (c), corrected for the number of cells, is ·289. Hence we conclude that there is a distinct, though not very great tendency towards equality in the number of types in the two hands of an individual. It appears, however, that the divergence is rather greater in the right than in the left hand.

The question now arises whether the difference in divergence in the two hands for the samples taken is significant. I have tested this by the method proposed by Professor Karl Pearson*.

---

* "On the Probability that Two Independent Distributions of Frequency are really Samples from the same Population," by Karl Pearson, F.R.S., *Biometrika*, Vol. VIII, pp. 250—254, July, 1911.

TABLE 2.

Divergence of Types in Right and Left Hands.

	Number of Types.					
	1	*2*	*3*	*4*	*5*	Totals
Right Hand ...	118	842	839	194	7	2000
Left Hand ...	174	955	716	151	4	2000

For this Table

$$\chi^2 = 33{\cdot}72,$$

whence P is less than ·000,005.

That is, the odds are more than 200,000 to 1 against the occurrence of two such divergent samples if they were random samples of the same population. In other words the right hand generally tends to have a greater divergence of types than the left.

The following Table gives the distribution of classes of prints for the various fingers of both hands:

TABLE 3.

Distribution of Classes of Prints.

	A	L_u	L_r	W	C
R_1 ...	46	1104	1	649	200
R_2 ...	352	537	456	481	174
R_3 ...	212	1399	38	274	77
R_4 ...	63	1015	17	729	176
R_5 ...	31	1631	3	228	107
Totals ...	704	5686	515	2361	734
L_1 ...	91	1311	3	341	254
L_2 ...	313	732	383	437	135
L_3 ...	215	1408	35	240	102
L_4 ...	66	1283	12	491	148
L_5 ...	35	1727	—	150	88
Totals ...	720	6461	433	1659	727
Totals for both hands	1424	12147	948	4020	1461

The most striking feature of this Table is the uneven distribution of the various classes, especially the large proportion of ulnar loops and the very small number of radial loops except in the forefingers. A comparison of the distribution in the two hands shows considerable differences; e.g., in the left thumb the number of arches is about double the number in the right; again, the whorls in each finger of the right hand are greatly in excess of those on the left, while the left hand has, in every case, an excess of ulnar loops.

If we arrange the numbers of each class in order of magnitude, we see that the order for the arches is identical for the two hands and also for the ulnar loops. In each of the other classes there is one exception to the "identical" order.

I have tested these distributions for each type by the method referred to in the footnote of p. 422 A, with the following results:—In the arches the odds are more than 500 to 1 against the occurrence of two such divergent samples which are random samples taken from the same population; in the ulnar loops the odds are more than 200,000 to 1; in the radial loops about 5 to 2; in the whorls more than 1,000,000 to 1, and in the composites more than 1300 to 1.

We may thus fairly conclude that with the exception of the radial loops the frequency distribution of the classes between the fingers is different in the two hands and the radial loops are so few, except in the forefinger, as to be almost negligible.

6. *Subdivision of Loops.* The great preponderance in the number of loops and the insignificance of the number of radial loops, except in the forefinger, make another subdivision of this class necessary. The method adopted is as follows:—All loops, in common with whorls and composites, contain certain well-defined points; these are (1) the "delta," or "outer terminus," and (2) the "point of the core," or "inner terminus." [See Henry, pp. 22—24.] The number of ridges intervening between the delta of a loop and the point of the core may be anything from one up to about thirty; in only 38 cases out of the 13,095 loops does the number of ridges exceed 25; two of these are over 30, one being 32 and the other 35. The complete distribution of ridges is given in Table 4 *a*.

In dividing the loops into two sub-classes according to the number of ridges the nearest approach to equality is obtained by taking (*a*) those containing from 1 to 12 ridges, and (*b*) those containing 13 or more ridges. For brevity I have called these classes (*a*) Small Loops, and (*b*) Large Loops; the terms "Small" and "Large" have no reference to the relative sizes of the patterns. The numbers in the two groups, thus arranged, are 7033 and 6062 respectively.

Table 4 *b* gives (1) the number of loops for each finger, (2) the means, (3) the standard deviations, and (4) the coefficients of variation in the numbers of ridges.

Examining the Table below consider first the means. The order which is identical in the two hands runs:

(1) Thumb, (2) Ring Finger, (3) Little Finger, (4) Middle Finger, (5) Index.

It will be noticed that this order of the means is quite different from that of the relative areas of the patterns.

TABLE 4 *a*.

Distribution of Ridges in Loops.

	Right Hand.					Left Hand.				
No. of Ridges	R_1	R_2	R_3	R_4	R_5	L_1	L_2	L_3	L_4	L_5
1	3	19	18	8	7	2	20	21	5	10
2	7	68	47	27	24	15	79	65	25	32
3	7	76	50	26	49	14	117	54	40	42
4	8	71	66	35	68	22	71	60	33	56
5	24	67	57	58	70	32	70	52	45	69
6	14	47	80	36	92	29	54	47	32	69
7	24	45	76	44	77	41	64	75	47	70
8	29	50	85	45	73	54	67	70	34	91
9	25	40	97	36	83	65	65	77	53	85
10	43	47	109	55	94	72	73	117	61	122
11	53	55	116	48	100	91	76	124	89	156
12	53	69	120	71	133	110	77	125	119	157
13	65	62	116	66	111	98	65	141	94	134
14	89	60	128	76	143	119	67	139	111	150
15	68	60	84	79	103	100	49	99	94	131
16	91	43	75	73	106	128	35	66	117	136
17	86	38	62	60	97	92	26	47	86	88
18	84	33	28	65	73	70	12	37	60	64
19	100	16	11	34	46	57	11	15	58	33
20	61	12	5	34	48	39	7	6	33	17
21	49	7	4	18	19	25	3	4	22	6
22	34	2	2	19	3	13	6	1	12	4
23	33	4	—	8	4	17	—	—	9	2
24	22	1	1	4	6	7	1	1	8	1
25	11	—	—	2	2	1	—	—	2	2
26	7	—	—	1	2	1	—	—	1	—
27	7	—	—	1	1	—	—	—	1	—
28	7	—	—	—	—	—	—	—	1	—
29	1	—	—	1	—	—	—	—	—	—
30	—	1	—	1	—	—	—	—	2	—
32	—	—	—	1	—	—	—	—	—	—
35	—	—	—	—	—	—	—	—	1	—
Totals	1105	993	1437	1032	1634	1314	1115	1443	1295	1727

TABLE 4 *b*.

	Number of Loops		Means		Standard Deviations		Coefficients of Variation	
	R	*L*	*R*	*L*	*R*	*L*	*R*	*L*
Thumb ...	1105	1314	15·52 ± ·10	13·27 ± ·09	5·17 ± ·07	4·63 ± ·06	34·34 ± ·53	34·85 ± ·51
Index ...	993	1115	9·69 ± ·12	8·83 ± ·10	5·41 ± ·08	4·88 ± ·07	55·82 ± 1·08	55·24 ± 1·00
Middle Finger	1437	1443	10·41 ± ·08	10·55 ± ·08	4·46 ± ·06	4·53 ± ·06	42·80 ± ·63	42·91 ± ·63
Ring Finger...	1032	1295	12·37 ± ·12	12·77 ± ·10	5·48 ± ·08	5·09 ± ·07	44·31 ± ·78	39·85 ± ·61
Little Finger .	1634	1727	11·75 ± ·08	11·53 ± ·07	4·97 ± ·06	4·46 ± ·05	42·30 ± ·58	38·71 ± ·50

Comparing the two hands we see that the differences in the middle, ring and little fingers are insignificant; in the thumb and index, however, there is a marked difference in favour of the right hand.

The order of the standard deviations in the right hand is:

(1) Ring Finger, (2) Index, (3) Thumb, (4) Little Finger, (5) Middle Finger.

In the left hand the order of the last two is reversed, but the difference is small.

With the exception of the middle finger, where the difference between the two hands is only about equal to the probable error and is therefore insignificant, the standard deviation is in every case greater for the right hand than for the left; the differences are all of the same order of magnitude and range from about ·39 to ·54.

Coming now to the coefficients of variation—the order in the right hand is:

(1) Index, (2) Ring Finger, (3) Middle Finger, (4) Little Finger, (5) Thumb.

In the left hand the order of the ring and middle fingers is interchanged.

Comparing the two hands we see that in three cases—the thumb, index, and middle finger—the differences are each less than the probable errors; in the other two cases the variability is considerably greater in the right hand than in the left.

I have carefully revised the calculations involved but have been unable to detect any error; neither can I suggest a reason for the large differences.

In "A First Study of the Variability and Correlation of the Hand" (see p. 421), the writers find that the variability of bone lengths is closely related to the relative utility of the fingers, the least variability being that of the most useful finger. There appears, however, to be no such simple relationship between the ridges of the loops and the relative utility of the fingers.

I have compared the distribution of ridges in the loops of the thumbs by Professor Pearson's method (p. 422 A, footnote), which gives $\chi^2 = 166{\cdot}64$; hence the odds are much greater than 1,000,000 to 1 against the occurrence of two such divergent samples if they were random samples taken from the same population.

The distribution—absolute and percentage—of the five groups is now as follows (Table 5).

In comparing the large and small loops it will be seen that in both hands there is an excess of large loops in the thumb, ring and little fingers, and an excess of small loops in the index and middle fingers. The order of these classes agrees in the two hands with one exception in each case.

An approximate measure of the relationship existing between the various combinations of digits is given by the number of cases in which two particular digits on the same or on opposite hands have the same pattern. Table 6 *a* gives the percentages for the same hand and for digits of the same name on opposite

TABLE 5.

	Arches		Small Loops		Large Loops		Whorls		Composites	
	No.	%	No.	%	No.	%	No.	%	No.	%
R_1 ...	46	2·30	290	14·50	815	40·75	649	32·45	200	10·00
R_2 ...	352	17·60	654	32·70	339	16·95	481	24·05	174	8·70
R_3 ...	212	10·60	921	46·05	516	25·80	274	13·70	77	3·85
R_4 ...	63	3·15	489	24·45	543	27·15	729	36·45	176	8·80
R_5 ...	31	1·55	870	43·50	764	38·20	228	11·40	107	5·35
Totals ...	704		3224		2977		2361		734	
L_1 ...	91	4·55	547	27·35	767	38·35	341	17·05	254	12·70
L_2 ...	313	15·65	833	41·65	282	14·10	437	21·85	135	6·75
L_3 ...	215	10·75	887	44·35	556	27·80	240	12·00	102	5·10
L_4 ...	66	3·30	583	29·15	712	35·60	491	24·55	148	7·40
L_5 ...	35	1·75	959	47·95	768	38·40	150	7·50	88	4·40
Totals ...	720		3809		3085		1659		727	
Totals for both hands	1424		7033		6062		4020		1461	

hands; the readings for other combinations of digits on opposite hands are given in Table 6 *d*, p. 431, where all the patterns are grouped in three classes for the sake of comparison with Galton's results.

Remarks on Table 6 *a*. (*a*) The percentages vary greatly with different combinations and with different patterns.

(*b*) The means and totals for digits of the same name on opposite hands are all much greater than the corresponding readings for the right or for the left hand; the means, with one exception, and also the totals for particular combinations on the left hand are all greater than the corresponding readings for the right.

(*c*) The order of magnitude of the totals is nearly the same for the two hands, those of the combinations including the thumb being, with one exception in each hand, the lowest. Hence, judging the relationship by the totals, it appears that (1) digits of the same name on opposite hands are the most closely related, the magnitude falling in order from the little fingers to the thumbs; (2) omitting the thumbs, two consecutive digits are generally more closely related than others more widely separated; (3) the digits of the left hand are more closely related than those of the right.

(*d*) The relationship between the thumb and any other digit seems to be less close than that between any pair of digits not including the thumb; also, in both hands, the thumb appears to be most closely related to the ring finger, then to the little finger, next to the middle and least to the fore-finger.

Another method of investigating the approximate relationship between the various digits is by means of a "centesimal" scale, as in Galton's *Finger Prints*, Ch. VIII. Table 6 *b* gives such scale readings for small loops, large loops and whorls, for pairs of digits on the same hand and for digits of the same name on opposite hands. I have not considered it necessary to include other couplets from opposite hands, because, as is shown later, the relationship between any pair of digits from opposite hands is practically the same as between the corresponding pair on the same hand. I have also omitted arches and composites from this part of the inquiry as the numbers belonging to these classes are, as a rule, comparatively small.

TABLE 6 *a*.

Percentage of Cases in which various pairs of Digits possess the same Class of Pattern.

Couplet	Right Hand					Totals	Left Hand					Totals
	A	*SL*	*LL*	*W*	*C*		*A*	*SL*	*LL*	*W*	*C*	
Thumb and fore-finger ...	1·5	6·3	7·5	13·0	1·4	29·7	2·4	15·7	7·2	8·3	1·3	34·9
„ middle finger	1·3	8·7	11·2	8·0	·7	29·9	1·6	16·2	13·4	4·9	1·4	37·5
„ ring „	·7	6·5	12·4	17·8	1·1	38·5	·7	13·2	17·1	7·9	1·5	40·4
„ little „	·4	9·9	15·0	6·9	·9	33·1	·7	19·0	15·7	2·7	·9	39·0
Fore-finger and middle finger	6·4	22·7	7·2	8·9	1·2	46·4	5·8	26·3	7·9	8·6	1·2	49·8
„ ring „	2·3	13·7	6·4	16·3	·9	39·6	2·5	17·8	7·1	12·7	1·5	41·6
„ little „	1·2	20·1	9·1	6·8	·9	38·1	1·3	26·3	8·2	4·3	·3	40·4
Middle and ring finger ...	2·5	16·8	9·1	12·0	·7	41·1	2·6	21·1	15·4	8·8	·8	48·7
„ little „ ...	1·0	26·8	14·0	4·6	·4	46·8	1·2	28·9	15·3	2·9	·6	48·9
Ring and little „ ...	·8	20·0	15·7	10·0	1·0	47·5	·9	23·7	19·5	6·0	·8	50·9
Means	1·8	15·2	10·8	10·4	·9	39·1	2·0	20·8	12·7	6·7	1·0	43·2

Couplet	*A*	*SL*	*LL*	*W*	*C*	Totals
Two thumbs ...	1·5	10·2	23·4	13·5	2·7	51·3
„ fore-fingers ...	9·3	22·7	5·6	14·4	1·4	53·4
„ middle fingers	5·8	31·8	14·8	7·0	·9	60·3
„ ring „	1·9	18·2	18·4	21·2	1·5	61·2
„ little „	·9	36·1	27·2	5·0	1·4	70·6
Means	3·9	23·8	17·9	12·2	1·6	59·3

The scale reading for any pair of digits is calculated as follows:—

Take, for example, the whorls on the right thumb and right fore-finger; the former has 32·5 and the latter 24 per cent. of whorls, while 13 per cent. of right hands have whorls on both thumb and fore-finger. Now from independent probability we shall expect $\frac{32\cdot5 \times 24}{100 \times 100} \times 100$, or 7·8 per cent. of "double whorls" in this combination of digits and we therefore conclude that the remaining 5·2 per cent. of double whorls are due to a relationship between the digits. If we set aside the 7·8 per cent. out of the 32·5 and 24, we see that from the remaining 24·7 and 16·2 per cent., the greatest possible percentage of double whorls would be 16·2; but as the actual percentage in addition to the 7·8 is 5·2, the centesimal measure of the relationship is $\frac{5\cdot2 \times 100}{16\cdot2}$, or 32, to the nearest unit.

TABLE 6 *b*.

Approximate Measures of Relationship between various pairs of Digits on a Centesimal Scale.

Couplet	Right Hand				Left Hand			
	SL	*LL*	*W*	Means	*SL*	*LL*	*W*	Means
Thumb and fore-finger ...	16	6	32	18	27	21	34	27
,, middle finger...	26	0	38	21	26	16	28	23
,, ring ,,	27	8	29	21	27	16	29	24
,, little ,,	44	0	42	29	41	4	23	23
Fore and middle fingers ...	44	22	54	40	34	39	64	46
,, ring ,, ...	35	15	49	33	33	23	44	33
,, little ,, ...	32	25	47	35	29	32	46	36
Middle and ring ,, ...	42	11	80	44	50	31	64	48
,, little ,, ...	29	26	31	29	33	27	30	30
Ring and little ,, ...	67	32	81	60	64	26	74	55

Couplet	*SL*	*LL*	*W*	Means
Two thumbs ...	59	34	69	54
,, fore-fingers ...	48	27	55	43
,, middle fingers...	47	41	52	47
,, ring ,,	64	50	78	64
,, little ,,	67	53	62	61

Most of the remarks on Table 6 *a* will be found applicable to Table 6 *b*, with, at most, but slight modification; the chief differences are that the relationship between the middle and little fingers is not so high in Table 6 *b* as in Table 6 *a*, and the order for pairs of like digits is not the same in the two Tables.

Comparison of results with those of Galton. In order to compare with Galton's results it is necessary to put large and small loops into one class and to include composites with whorls. Making some allowance for the difference of classification, and for any slight variation which may be due to the fact that the material is drawn from very different classes of the population, it will be found that there is almost perfect agreement between our data on all essential points.

The relative frequency found in the two investigations was:—

	Galton	Waite
Arches	6·5 per cent.	7·1 per cent.
Loops	67·5 " "	65·5 " "
Whorls	26·0 " "	27·4 " "

The differences are small in comparison with some found by Galton when examining the finger-prints of different races. For example, 1332 Hebrew children had arches on the right fore-finger in 13·6 per cent. of the cases, while only 7·9 per cent. of 250 English children had arches on that finger.

TABLE 6 *c*.

Percentage Frequency of Arches, Loops and Whorls on the different Digits.

	GALTON*						WAITE					
	From observations of the 5000 digits of 500 persons						From observations of 20000 digits of 2000 persons					
	Arches		Loops		Whorls		Arches		Loops		Whorls	
Digit	*R*	*L*	*R*	*L*	*R*	*L*	*R*	*L*	*R*	*L*	*R*	*L*
Fore-finger ...	17	17	53	53	30	28	17·6	15·7	49·7	55·7	32·7	28·6
Middle finger	7	8	78	76	15	16	10·6	10·7	71·9	72·2	17·5	17·1
Little "	1	2	86	90	13	8	1·5	1·7	81·7	86·4	16·8	11·9
Thumb ...	3	5	53	65	44	30	2·3	4·5	55·2	65·7	42·5	29·8
Ring finger...	2	3	53	66	45	31	3·2	3·3	51·6	64·7	45·2	32·0
Totals	30	35	323	350	147	113	35·2	35·9	310·1	344·7	154·7	119·4

Galton arranged the digits as in Table 6 *c*, in order to bring out certain peculiarities. He says:—

> "The digits are seen to fall into two well-marked groups; the one including the fore, middle, and little fingers, the other including the thumb and ring finger. As regards the first group, the frequency with which any pattern occurs in any named digit is statistically the same, whether

---

* From *Finger Prints*, p. 116, Table II.

that digit be on the right or on the left hand; as regards the second group, the frequency differs greatly in the two hands. But though in the first group the two fore-fingers, the two middle, and the two little fingers of the right hand are severally circumstanced alike in the frequency with which their various patterns occur, the difference between the frequency of the patterns on a fore, a middle, and a little finger, respectively, is very great.

"In the second group, though the thumbs on opposite hands do not resemble each other in the statistical frequency of the A. L. W. patterns, nor do the ring fingers, there is a great resemblance between the respective frequencies in the thumbs and ring fingers; for instance, the whorls on either of these fingers on the left hand are only two-thirds as common as those on the right. The figures in each line and in each column are consistent throughout in expressing these curious differences, which must therefore be accepted as facts, and not as statistical accidents, whatever may be their explanation." (Galton, *Finger Prints*, p. 116.)

These remarks apply with equal force to my figures although the actual percentages differ somewhat in certain cases, the most marked being in the middle finger arches and the little finger whorls.

The following points of agreement in the distribution of the patterns are also noticed by reference to Table 6 *c*.

The frequency of arches on the fore-fingers is much greater than on any other of the four digits. "It amounts to 17 per cent. on the fore-fingers, while on the thumbs and on the remaining fingers the frequency diminishes in a ratio that roughly accords with the distance of each digit from the fore-finger.

"The frequency of Loops has two maxima; the principal one is on the little finger, the secondary on the middle finger.

"Whorls are most common on the thumb and the ring-finger, most rare on the middle and little fingers." (*Finger Prints*, p. 117.)

In discussing radial and ulnar loops, which Galton describes as loops having "inner" and "outer" slopes, respectively, he says:—

"In all digits except the fore-fingers, the inner slope is much the more rare of the two; but in the fore-fingers the inner slope appears two-thirds as frequently as the outer slope. Out of the percentage of 53 loops of the one or other kind on the right fore-finger, 21 of them have an inner and 32 an outer slope; out of the percentage of 55 loops on the left fore-finger, 21 have inner and 34 have outer slopes. These subdivisions 21–21 and 32–34 corroborate the strong statistical similarity that was observed to exist between the frequency of the several patterns on the right and left fore-fingers; a condition which was also found to characterise the middle and little fingers." (*Finger Prints*, p. 118.)

These statements are true, in general, of my Table 3, but my percentages on the right fore-finger are 22·8 radial and 26·9 ulnar; on the left they are 19·2 and 36·6 respectively.

Close agreement is also observed in Table 6 *d* which shows the tendency of digits to resemble one another in their various combinations. Galton omits combinations into which the little finger enters "because the overwhelming frequency of loops in the little fingers would make the results of comparatively little interest, while their insertion would greatly increase the size of the table." (*Finger Prints*, p. 119.) I have included them, however, for the sake of comparison and completeness.

My percentages are readily obtained from Tables LVI to C in the Appendix.

TABLE 6 *d*.

Percentage of Cases in which the same Class of Pattern occurs in various Couplets of Digits.

Couplet	GALTON*						WAITE					
	Arches in		Loops in		Whorls in		Arches in		Loops in		Whorls in	
	Same hand	Opposite hand	Same hand	Opposite hand	Same hand	Opposite hand	Same hand	Opposite hand	Same hand	Opposite hand	Same hand	Opposite hand
Two thumbs	—	2	—	48	—	24	—	1·6	—	47·4	—	24·5
„ fore-fingers ...	—	9	—	38	—	20	—	9·3	—	36·2	—	20·4
„ middle fingers ...	—	3	—	65	—	9	—	5·8	—	60·6	—	10·5
„ ring „ ...	—	2	—	46	—	26	—	1·9	—	46·3	—	27·9
„ little „ ...	—	—	—	—	—	—	—	·9	—	63·2	—	6·3
Thumb and fore-finger	2	2	35	33	16	15	1·90	1·85	36·8	35·7	18·2	17·5
„ mid-finger	1	1	48	47	9	8	1·4	1·5	47·0	46·7	10·9	10·5
„ ring finger	1	1	40	38	20	18	·7	·6	41·0	39·4	20·8	19·0
Fore and mid-finger ...	5	5	48	46	12	11	6·1	5·5	44·3	43·5	12·8	12·3
„ ring finger ...	2	2	35	35	17	17	2·4	2·3	36·5	35·7	20·8	20·2
Middle and ring finger	2	2	50	50	13	12	2·5	2·4	48·3	47·1	14·7	13·7
Thumb and little finger	—	—	—	—	—	—	·52	·45	54·2	53·6	8·8	8·1
Fore and little finger ...	—	—	—	—	—	—	1·20	1·15	47·7	47·2	9·3	8·1
Middle and little finger	—	—	—	—	—	—	1·1	1·0	63·9	63·5	6·5	6·1
Ring and little finger...	—	—	—	—	—	—	·8	·7	56·1	54·8	12·9	11·8

In commenting on his results in Table 6 *d*, Galton says:—"The agreement in the above entries is so curiously close as to have excited grave suspicion that it was due to some absurd blunder, by which the same figures were made inadvertently to do duty twice over, but subsequent checking disclosed no error. Though the unanimity of the results is wonderful, they are fairly arrived at, and leave no doubt that the relationship of any one particular digit, whether thumb, fore, middle, ring or little finger, to any other particular digit, is the same, whether the two digits are on the same or on opposite hands."

It will be noticed, however, that while exactly half of Galton's eighteen pairs of percentages, which are worked to the nearest unit only, are in strict agreement, in all the other cases the result is one or two units less for two digits on opposite hands than for the corresponding digits on the same hand. In my figures the percentage for two digits on opposite hands is in every case the lower, and

---

* *Finger Prints*, p. 120, Tables VI *a* and VI *b*.

although the differences are small, ranging only up to 1·8 while four-fifths of them are less than 1, the consistency of the results suggests a slightly closer relationship between a pair of digits on the same hand than between the corresponding pair on opposite hands. This view is further supported at a later stage of this paper. (See Remark (*d*) on Tables 14–16, p. 450.)

One further comparison is of interest, namely, the measure of relationship between the various digits on a centesimal scale. It should be noted, however, that while Galton's means are based on loops and whorls only, omitting arches from his three groups, mine are based on small loops, large loops and whorls, omitting arches and composites from my five groups; also Galton gives no results for those combinations which include the little finger.

TABLE 6 *e*

Approximate Measures of Relationship between the various Digits, on a Centesimal Scale.

Couplet	GALTON*	WAITE	
	Means	Right	Left
Thumb and fore-finger ...	24	18	27
,, middle finger ...	27	21	23
,, ring finger ...	39	21	24
Fore and middle finger ...	60	40	46
,, ring finger ...	23	33	33
Middle and ring finger ...	52	44	48
Right and left thumbs ...	61	54	
,, fore-fingers ...	48	43	
,, middle fingers	43	47	
,, ring fingers ...	65	64	

For the reasons given above we could hardly expect that these readings would be even approximately equal, but for all that, the same general relations are seen to hold good in the two sets of results.

It is convenient at this stage to summarize a few of the most important points which have been brought to light in the foregoing pages. These are:

(*a*) A greater divergence of types in the right hand than in the left.

(*b*) A clustering of the same type in the hands of an individual.

(*c*) The uneven distribution of the various types in the different fingers, especially the almost entire absence of ulnar loops except in the index.

---

* *Finger Prints*, p. 129, Table VIII.

(*d*) The differentiation of types in the two hands, in particular the large excess of whorls in the right hand and of arches in the left thumb.

(*e*) Where there is any significant difference in the means, standard deviations and coefficients of variation in the numbers of ridges in the loops of the two hands those quantities are always greater for the right hand than for the left.

(*f*) The relationship between digits of the same name on opposite hands is closer than that between any others; the digits of the left hand are more closely related than those of the right; and two consecutive digits, whether on the same or on opposite hands, are generally more closely related than others which are more widely separated. The relationship between the thumb and any other digit is less close than that of any pair not including the thumb.

We may thus conclude that the left hand in its distribution of patterns is differentiated from the right and that individual fingers are associated in a differential way with special types. We know that the right hand is differentiated from the left in use, and it would seem reasonable to suppose, even if we cannot account for the adaptation to use, that the finger-prints have been differentiated in accordance with this use differentiation.

It may be suggested that the finger-prints if differentiated in accordance with diversity of use of the several fingers and of each hand follow a law of differentiated utility and not as the bones a law of maximum general utility of the finger.

7. *Correlation between the Classes of Finger-Prints.* The object of this section is to obtain the associations between the various classes of prints and on the basis of these associations to enquire whether any Natural Order exists in which a certain degree of continuity may be assumed. For a complete investigation of this problem fifty-five Tables are necessary. They are:

(*a*) 10 Tables of Classes for the Right Hand.
(*b*) 10 „ „ „ Left Hand.
(*c*) 25 „ „ „ Right against the Left Hand.
(*d*) 10 „ „ both Hands together.

These Tables are given in the Appendix, pp. 453 *et seq.*

The correlation coefficients and the contingencies have been calculated for the whole of these Tables. For all the restricted Tables, I to XX, and XLVI to LV, and in certain of the remaining Tables where the results of the other two methods are widely divergent, the correlation ratio has also been found. In these cases I have obtained η in both directions, the values of η given in Tables 8 and 9 being the square root of the product of the two η's for each Table*.

---

* The arithmetic instead of the geometric mean might have been taken, and there would not have been very marked differences. But the geometric mean has the advantage of a symmetrical value, i.e.

$$\sqrt{\eta_x\eta_y} = \frac{[S\{n_x(\bar{y}_x - \bar{y})^2\}\ S\{n_y(\bar{x}_y - \bar{x})^2\}]^{\frac{1}{2}}}{\sigma_x\sigma_y},$$

which has certain analogies with a coefficient of correlation.

8 *a*. *Method of Calculating the Coefficients of Contingency in Restricted Tables.* It will be noticed that the Tables I to XX and XLVI to LV, given in the Appendix, differ in general character from most correlation tables since the whole of the cells in the lower right-hand portion are necessarily empty. Consequently the usual method of finding the independent probability numbers for the purpose of calculating contingencies is not applicable to those Tables. The method which has been employed was suggested to me by Professor Pearson. It is as follows:

Consider Table VI, Appendix, p. 454, which gives the distribution of small loops and whorls for the right hand. Commencing with the 45 hands which concontain 5 small loops each, it will be seen that the independent probability number is the same as the observed, since a hand which has five prints of one class can have no other. In the next column the distribution of the 211 prints by independent probability is not in the ratio of 861 to 497 since 45 of the 861 have already been disposed of, but in the ratio of 816 to 497, that is, the numbers in the two rows are 131·1 and 79·9. Again in the third column from the right containing 3 small loops, 45 and 131·1 of the first row are accounted for and 79·9 of the second row; hence the independent distribution of the 306 in the third column is in the proportion of 684·9 : 417·1 : 292; that is, the numbers are 150·3, 91·6, and 64·1; and so on.

It should be noted that the same independent probability numbers are obtained if we commence at the bottom of the first column with the 50 hands each containing 5 whorls and work horizontally instead of vertically.

The differences between these independent probability numbers and the observed numbers are then used to find the contingency in the same way as in the ordinary contingency Table.

No correction for the number of cells has been applied to the contingency coefficients in this type of Table as we have, at present, no appreciation of what it should be.

The complete contingency Table, worked as described, is given below.

Note on Calculation of Contingency Coefficients. It should be borne in mind that in finding the independent probability numbers in all contingency tables as well as in calculating the standard deviations, it is assumed that the distribution of the marginal totals is in the same ratio as would be the case if the whole population were taken; in other words, that if n_s is the total of an array when a sample N is taken and m_s the total of the corresponding array when the whole population M is taken, then it is assumed that

$$n_s = m_s \frac{N}{M}.$$

Evidently the correct value of the independent probability number in the (s, s') cell of an ordinary contingency table would be

$$\frac{m_s \dfrac{N}{M} \times m'_{s'} \dfrac{N}{M}}{N} \quad \text{or} \quad m_s m'_{s'} \frac{N}{M^2}$$

and the contributory contingency

$$\frac{\left(n_{ss'} - m_s m'_{s'} \dfrac{N}{M^2}\right)^2}{m_s m'_{s'} \dfrac{N}{M^2}}.$$

TABLE 7.

Contingency Table.

Small Loops, *R.*

Whorls, *R.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	n 78 β 200·6 ψ −122·6 ψ^2/β 74·93	144 167·4 −23·4 3·27	204 166·6 37·4 8·40	211 150·3 60·7 24·51	179 131·1 47·9 17·50	45 45 — —	861
1	106 122·2 −16·2 2·15	153 101·9 51·1 25·63	126 101·4 24·6 5·97	80 91·6 −11·6 1·47	32 79·9 −47·9 28·72	— — — —	497
2	130 85·5 44·5 23·16	92 71·4 20·6 5·94	55 71·0 −16·0 3·61	15 64·1 −49·1 37·61	— — — —	— — — —	292
3	125 63·8 61·2 58·70	38 53·2 −15·2 4·34	7 53·0 −46·0 39·92	— — — —	— — — —	— — — —	170
4	104 70·9 33·1 15·45	26 59·1 −33·1 18·54	— — — —	— — — —	— — — —	— — — —	130
5	50 50	— —	— —	— —	— —	— —	50
Totals	593	453	392	306	211	45	2000

$$\chi^2 = S(\psi^2/\beta) = 399{\cdot}82, \qquad \phi^2 = \chi^2/n = {\cdot}19991, \qquad C_2 = \sqrt{\frac{{\cdot}19991}{1{\cdot}19991}} = {\cdot}4082.$$

Similarly the quantities $m_s \dfrac{N}{M}$, $m'_{s'} \dfrac{N}{M}$, etc. would be the correct marginal totals to use in finding the independent probability numbers for the restricted contingency tables and in obtaining the standard deviations, instead of the observed totals n_s, $n'_{s'}$, etc.

However, as we do not generally know M, m_s, $m'_{s'}$, etc., we are obliged to use the observed marginal totals as the nearest approximation we can get to the correct values, although n_s is not, in general, equal to $m_s \frac{N}{M}$. A similar assumption is of course always made in the formulae for the probable errors of samples, where the sample value is put ultimately for the population value.

8 *b*. *Correlation Ratio of Restricted Tables.* It is obvious that the ordinary method of calculating the correlation ratio also requires modification with Tables of this type; for this method is based on the differences between the means of the marginal totals and the means of the arrays. Now, in restricted Tables it would be impossible for the means of about half the arrays to approximate to the means of the marginal totals and it would be fallacious to base any conclusion on the deviations of the observed means from impossible values.

A nearer approximation would be to take the pseudo η from the formula

$$\eta_p{}^2 = \frac{S\{n_x(\bar{y}_a - {}_a\bar{y}_i)^2\}}{N\sigma_y{}^2},$$

where ${}_a\bar{y}_i$ is the mean of an array of the independent probability numbers; but the denominator of this formula must be modified in such a way that in a case of perfect association, $\eta =$ unity. The desired result is obtained if we put Σ^2 instead of $\sigma_y{}^2$, where

$$\Sigma^2 = \frac{SS(y - {}_a\bar{y}_i)^2}{N}.$$

We may write

$$\begin{aligned}\Sigma^2 &= \frac{SS(y - \bar{y}_a + \bar{y}_a - {}_a\bar{y}_i)^2}{N} \\ &= \frac{SS(y - \bar{y}_a)^2}{N} + \frac{S\{n_x(\bar{y}_a - {}_a\bar{y}_i)^2\}}{N} + \frac{2SS(y - \bar{y}_a)(\bar{y}_a - {}_a\bar{y}_i)}{N} \\ &= \frac{S(n_x\sigma_a{}^2)}{N} + \frac{S\{n_x(\bar{y}_a - {}_a\bar{y}_i)^2\}}{N},\end{aligned}$$

since the third term vanishes; hence

$$\eta^2 = \frac{S\{n_x(\bar{y}_a - {}_a\bar{y}_i)^2\}/N}{S(n_x\sigma_a{}^2)/N + S\{n_x(\bar{y}_a - {}_a\bar{y}_i)^2\}/N}.$$

But
$$\frac{S(n_x\sigma_a{}^2)}{N\sigma_y{}^2} = 1 - \eta_c{}^2,$$

and
$$\frac{S\{n_x(\bar{y}_a - {}_a\bar{y}_i)^2\}}{N\sigma_y{}^2} = \eta_p{}^2,$$

where η_c is the crude η found by the ordinary method.

We have, therefore, the value of the correlation ratio of restricted Tables given by

$$\eta^2 = \frac{\eta_p{}^2}{1 - \eta_c{}^2 + \eta_p{}^2},$$

or

$$\eta = \frac{\eta_p}{\sqrt{1 - \eta_c{}^2 + \eta_p{}^2}}.$$

The correlation ratio has been found by the method described above for all the restricted Tables; it has also been determined by the ordinary method for a few of the other Tables, but no correction for number of arrays has been applied. The results, together with the coefficients of correlation and of contingency, are given in Tables 8 and 9.

Regression curves for all the restricted Tables are given on Plates (a—ϵ). The continuous line is the independent probability curve and the broken line the curve of the observed means. It follows that the area between the curves, weighted, of course, with the marginal totals, gives a measure of the correlation ratio between the two characters.

Each set of three figures for two particular characters, namely, those for the right hand, left hand, and both hands respectively, will generally be found to resemble each other closely. Irregularities occur chiefly with composites but this is not surprising if we consider the nature of this class.

8 *c*. *Coefficients of Correlation of Restricted Tables.* A glance at the diagrams of means of the restricted Tables, Plates (a—ϵ), shows that the regression is generally non-linear; it is also evident that a sensible value of r is introduced by the restriction*. Hence the value of r as found by the ordinary product-moment method is (i) too small because of the skewness of regression and (ii) too large on account of the restriction. These two contrary causes render the coefficient of correlation of restricted Tables unreliable and therefore quite valueless; for even if it sometimes agrees fairly closely with the correlation ratio and the contingency coefficient, this agreement is probably due to the fact that the two sources of error counterbalance each other.

In the remaining Tables, for which the results are given in Table 9, the regression is frequently skew; for this reason and for those given above, I have rejected the values of the coefficient of correlation in the sequence and have based my conclusions on the contingency coefficients, confirmed in general by the correlation ratio.

---

* For example, in small loops and large loops, left, the case in which the difference between r and c is the greatest, the independent probability numbers have the correlation coefficient $-{\cdot}512$ (instead of the theoretical value zero), as compared with $-{\cdot}507$ of the observed numbers. In the case of arches and small loops, both hands, r for the independent probability numbers is $-{\cdot}148$, as against $+{\cdot}147$ of Table 8.

PLATE *a.*

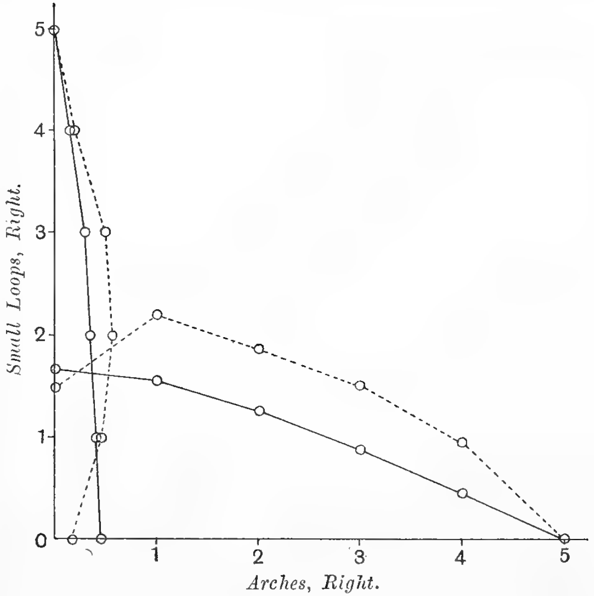


Fig. 1 (Table I).

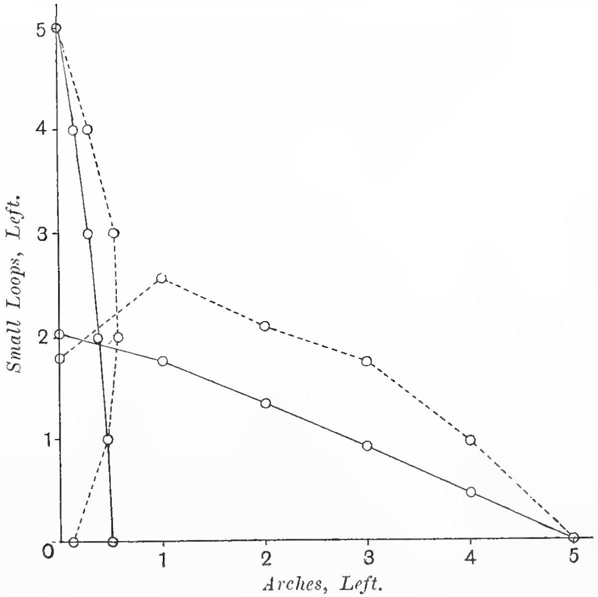


Fig. 2 (Table XI).

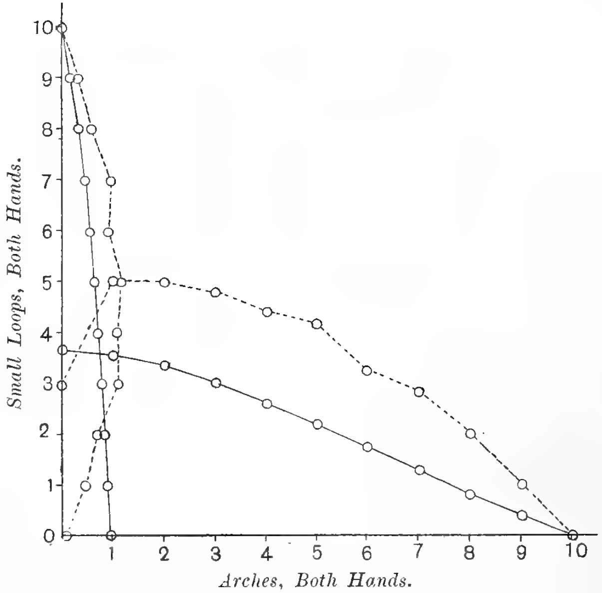


Fig. 3 (Table XLVI).

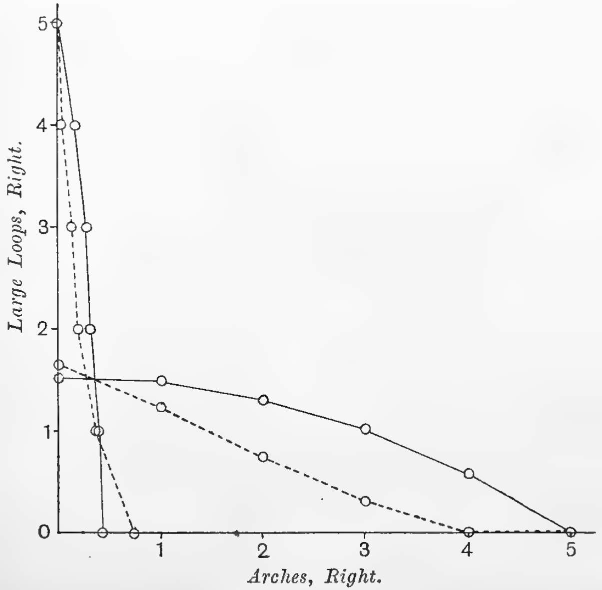


Fig. 4 (Table II).

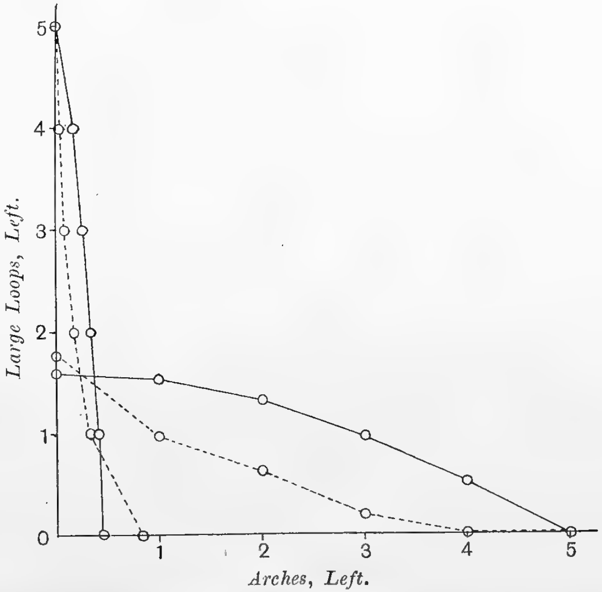


Fig. 5 (Table XII).

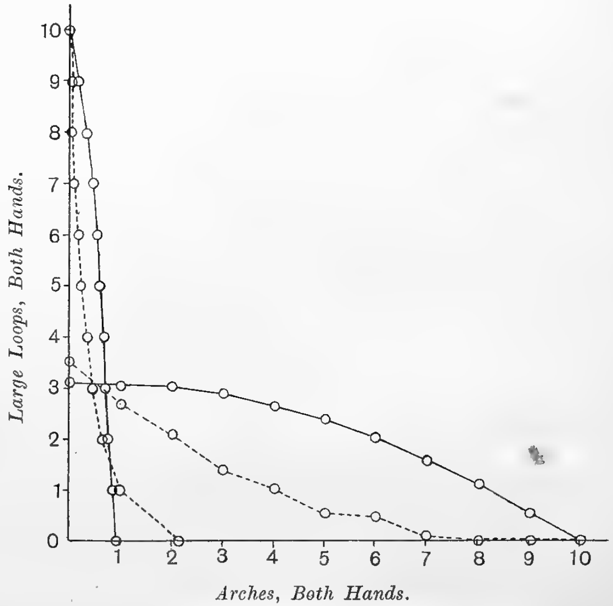


Fig. 6 (Table XLVII).

Regression Curves for Finger-Prints. Continuous Lines = Independent Probability Curves, Broken Lines = Actual Curves.

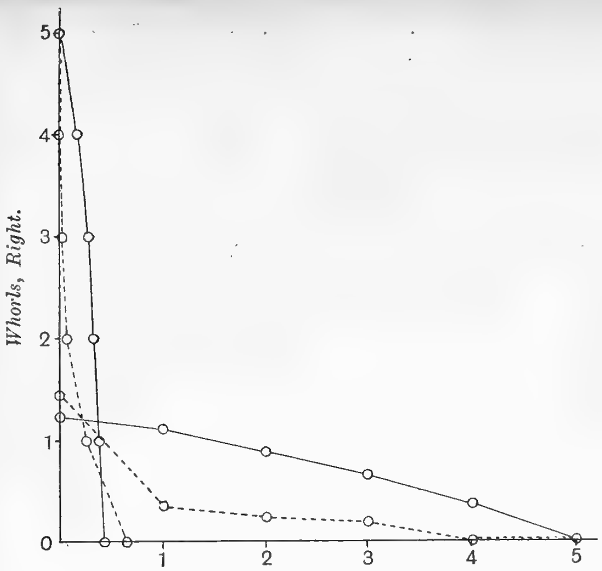


Fig. 7 (Table III).

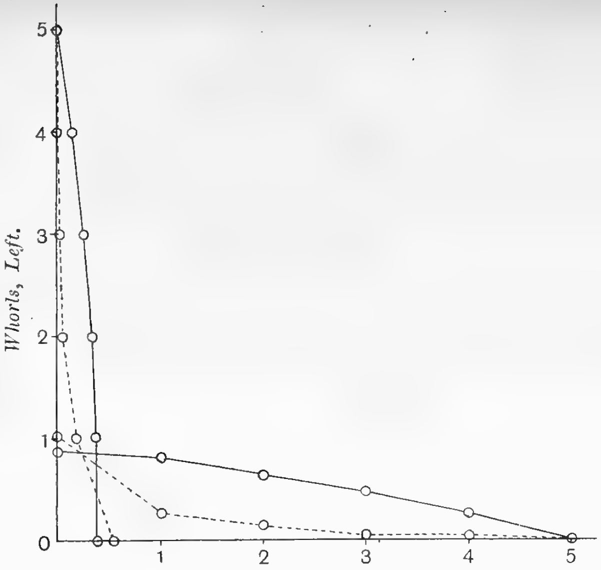


Fig. 8 (Table XIII).

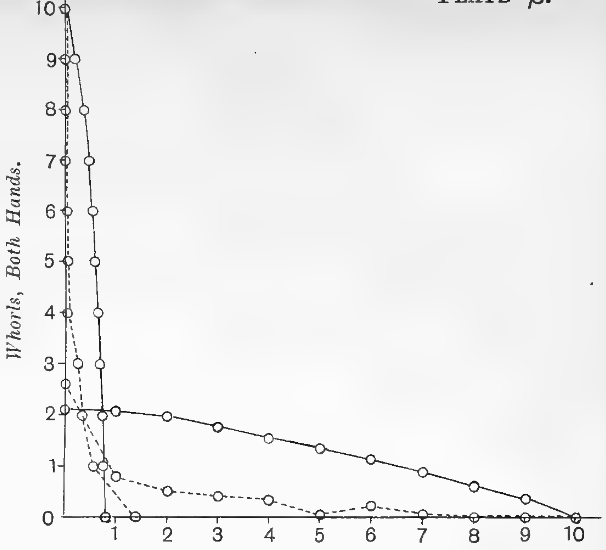


Fig. 9 (Table XLVIII).

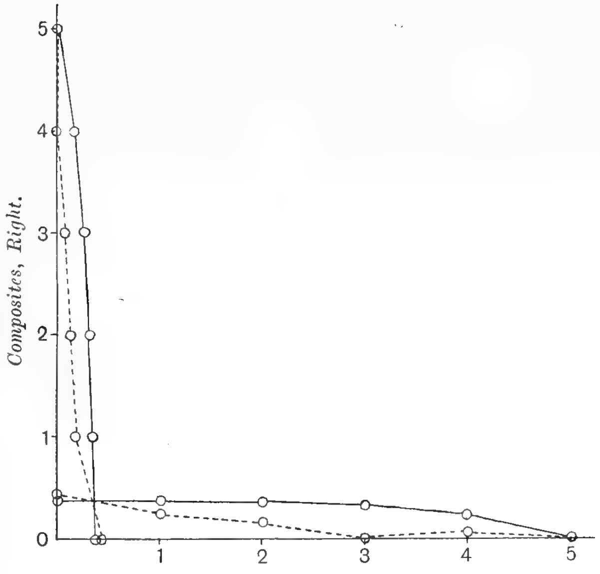


Fig. 10 (Table IV).

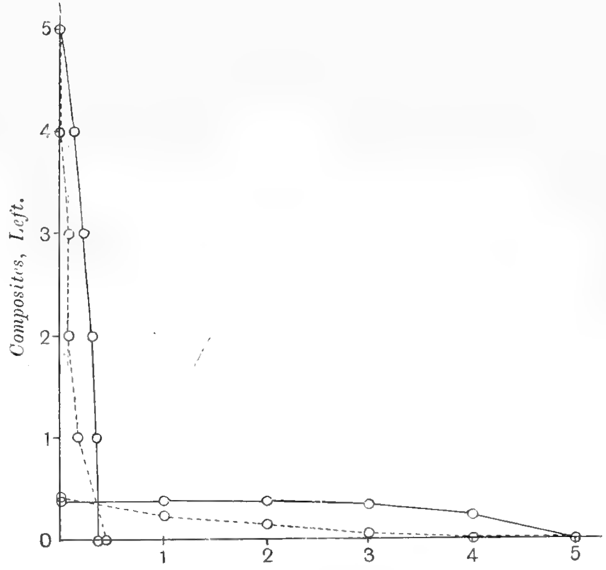


Fig. 11 (Table XIV).

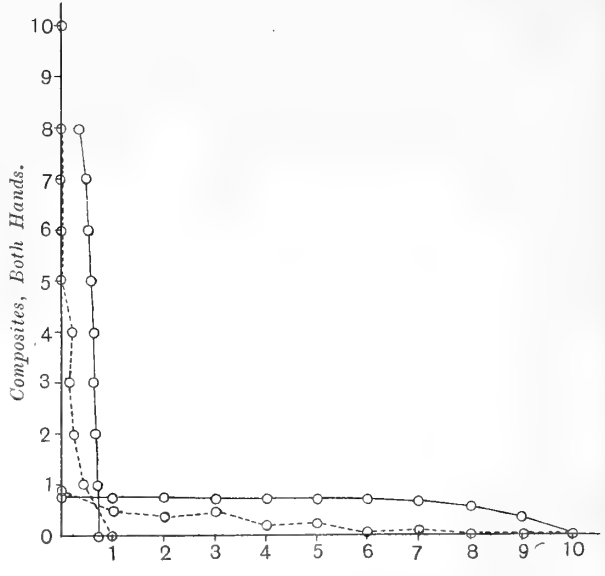


Fig. 12 (Table XLIX).

Regression Curves for Finger-Prints. Continuous Lines = Independent Probability Curves, Broken Lines = Actual Curves.

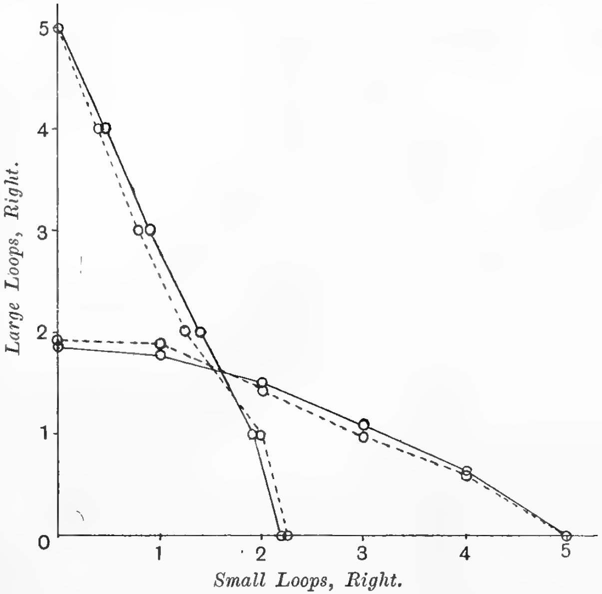


Fig. 13 (Table V).

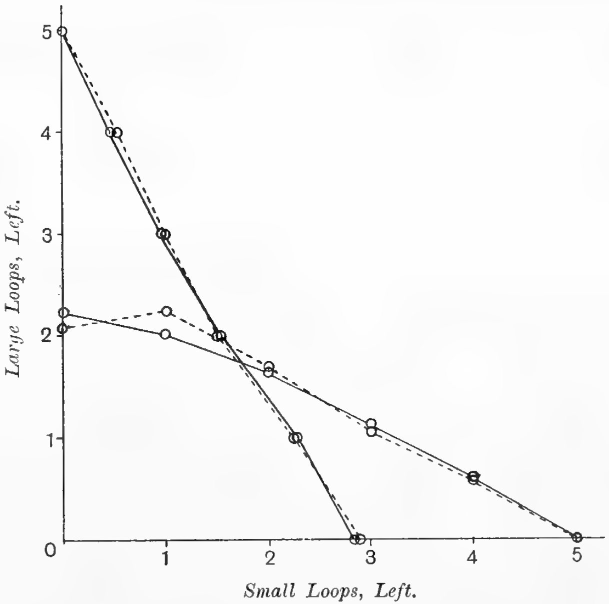


Fig. 14 (Table XV).

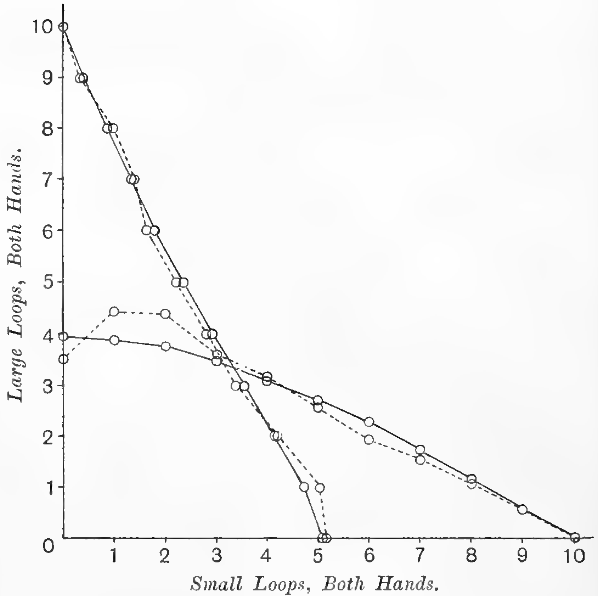


Fig. 15 (Table L).

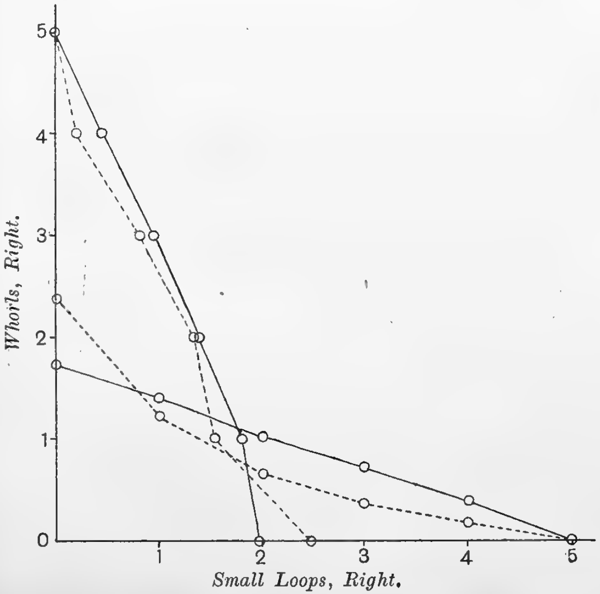


Fig. 16 (Table VI).

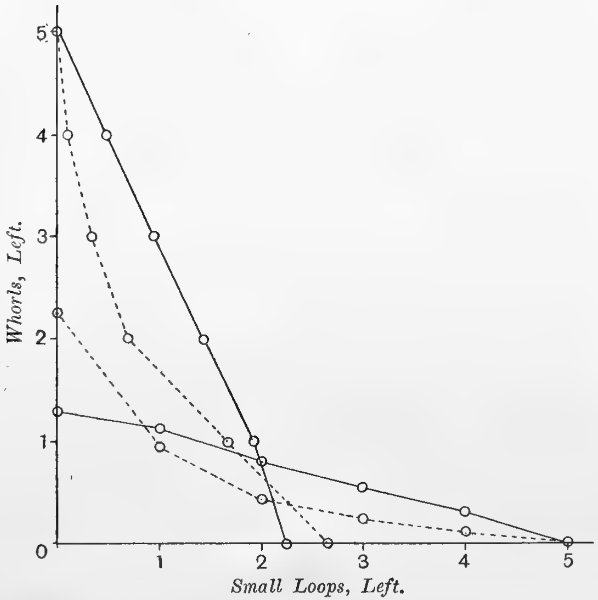


Fig. 17 (Table XVI).

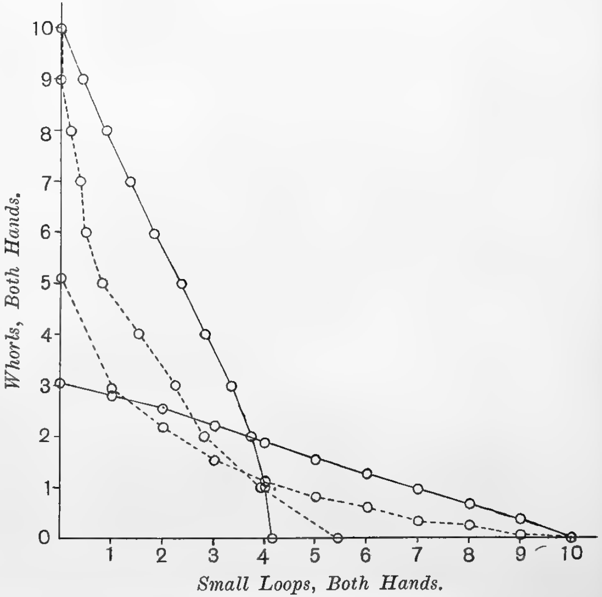
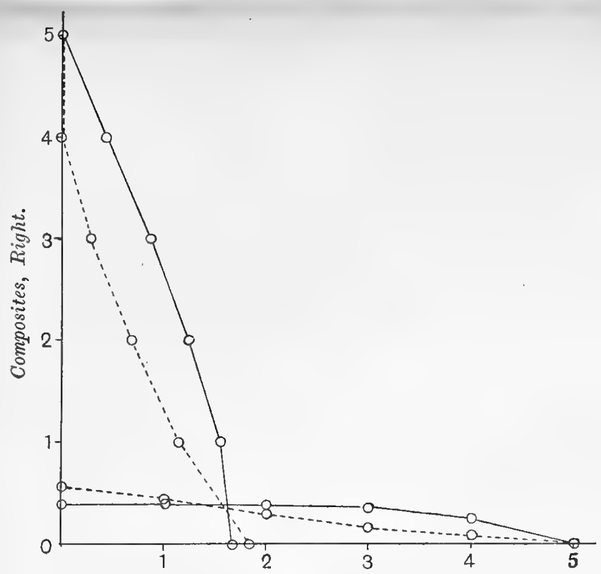
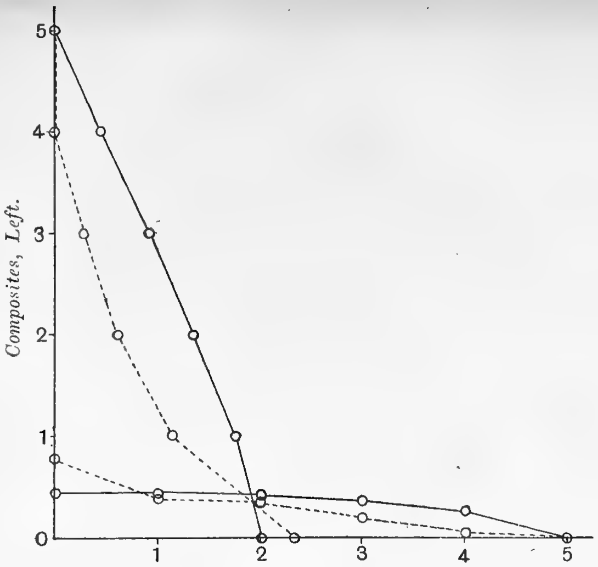


Fig. 18 (Table LI).

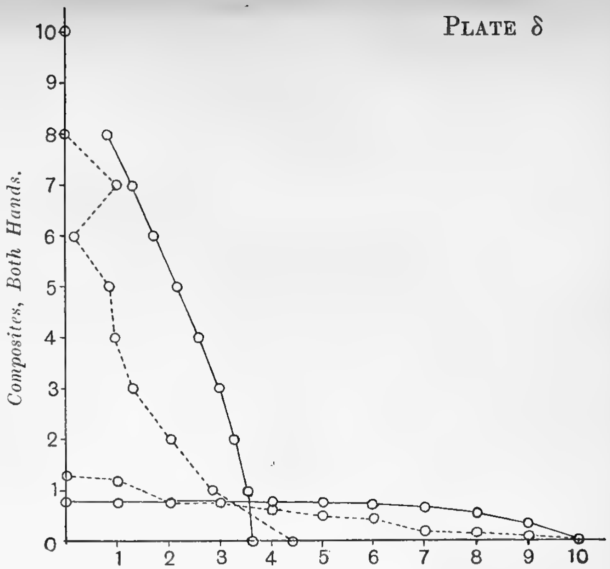
Regression Curves for Finger-Prints. Continuous Lines = Independent Probability Curves, Broken Lines = Actual Curves.



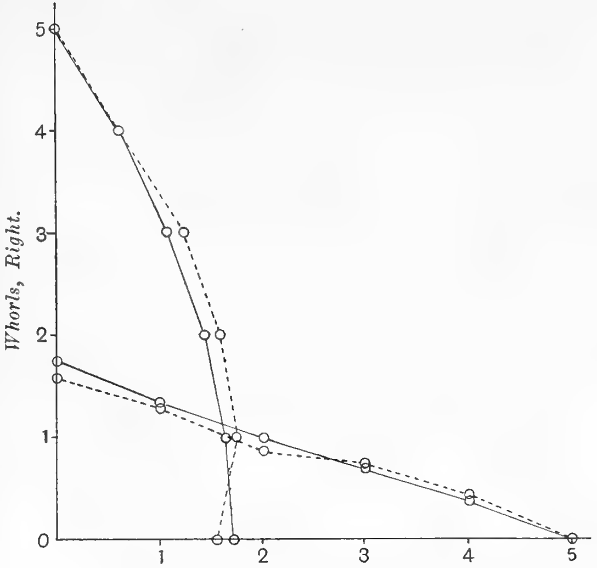
Small Loops, Right.
Fig. 19 (Table VII).



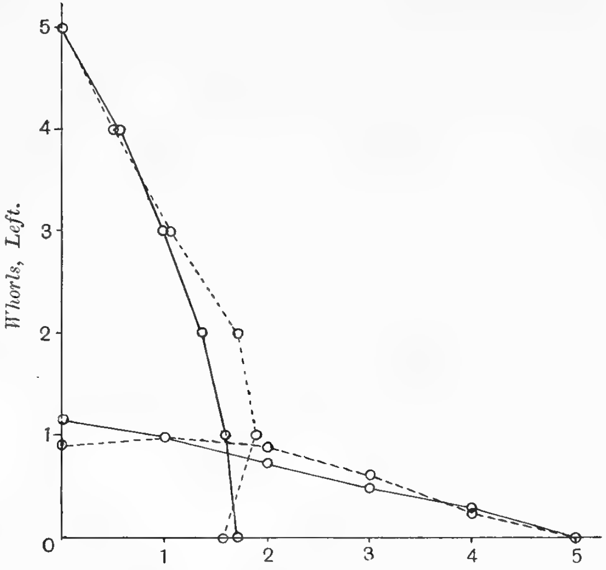
Small Loops, Left.
Fig. 20 (Table XVII).



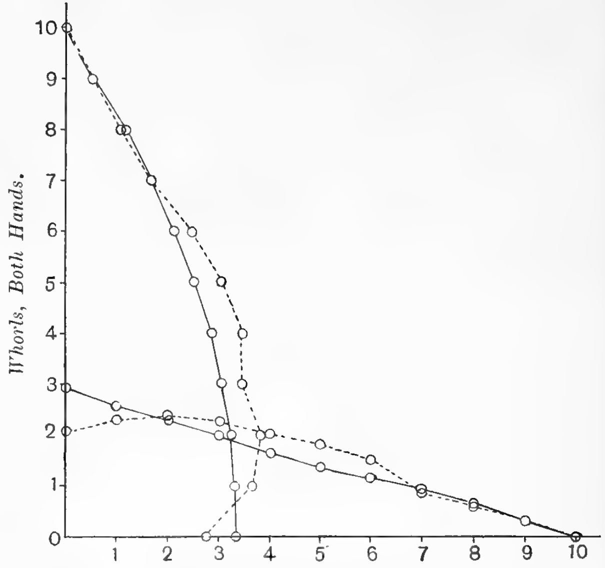
Small Loops, Both Hands.
Fig. 21 (Table LII).



Large Loops, Right.
Fig. 22 (Table VIII).

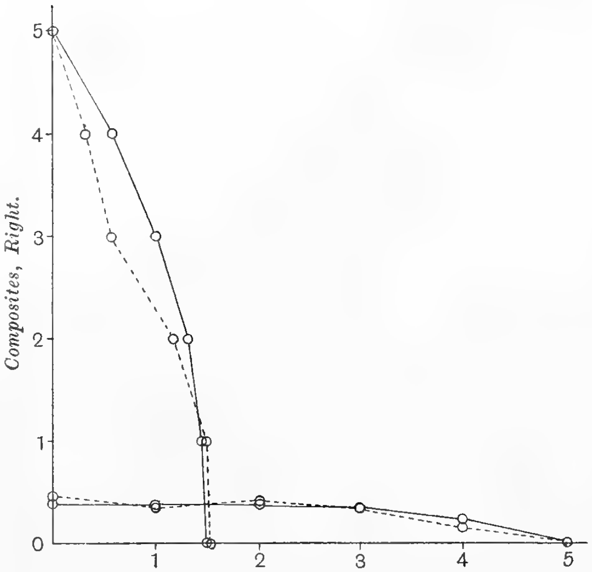


Large Loops, Left.
Fig. 23 (Table XVIII).

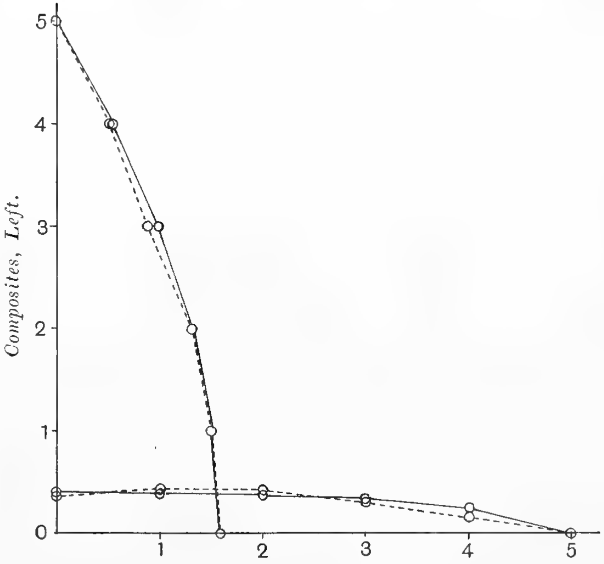


Large Loops, Both Hands.
Fig. 24 (Table LIII).

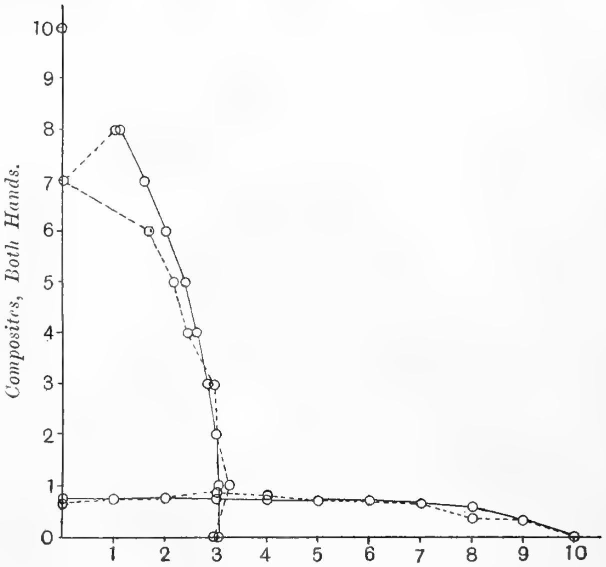
Regression Curves for Finger-Prints. Continuous Lines = Independent Probability Curves, Broken Lines = Actual Curves.



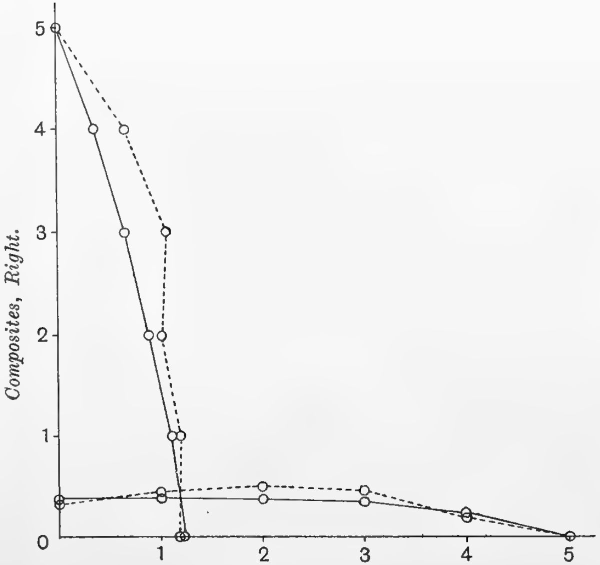
Large Loops, Right.
Fig. 25 (Table IX).



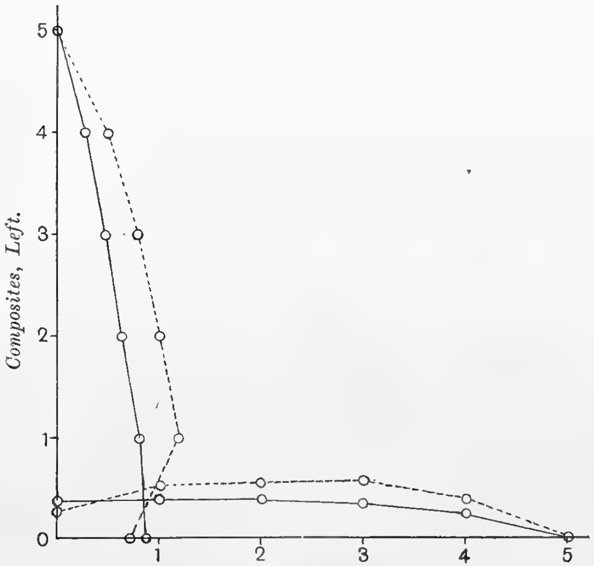
Large Loops, Left.
Fig. 26 (Table XIX).



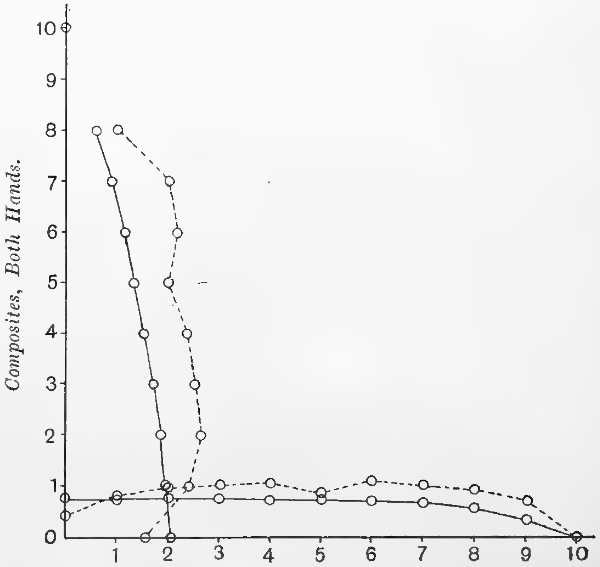
Large Loops, Both Hands.
Fig. 27 (Table LIV).



Whorls, Right.
Fig. 28 (Table X).



Whorls, Left.
Fig. 29 (Table XX).



Whorls, Both Hands.
Fig. 30 (Table LV).

Regression Curves for Finger-Prints. Continuous Lines = Independent Probability Curves, Broken Lines = Actual Curves.

TABLE 8.

		r	$yx\eta_c$	$yx\eta_p$	$yx\eta$	$xy\eta_c$	$xy\eta_p$	$xy\eta$	η	C	Table in Appendix
Right Hand	Arches and Small Loops ...	·062	·192	·242	·227	·228	·266	·264	·251	·305	I
	„ Large Loops ...	– ·273	·273	·194	·198	·298	·215	·220	·209	·246	II
	„ Whorls	– ·317	·359	·283	·290	·353	·286	·292	·291	·335	III
	„ Composites ...	– ·146	·154	·139	·139	·157	·140	·140	·140	·154	IV
	Small Loops and Large Loops	– ·422	·438	·074	·082	·430	·073	·080	·081	·166	V
	„ Whorls ...	– ·585	·622	·313	·371	·574	·315	·385	·378	·408	VI
	„ Composites	– ·270	·273	·207	·210	·273	·197	·201	·205	·228	VII
	Large Loops and Whorls ...	– ·234	·239	·079	·081	·305	·112	·116	·097	·162	VIII
	„ Composites	– ·093	·138	·093	·093	·104	·045	·045	·065	·128	IX
	Whorls and Composites ...	– ·020	·162	·117	·126	·032	·061	·061	·088	·137	X
Left Hand	Arches and Small Loops ...	·038	·197	·271	·267	·249	·301	·292	·281	·335	XI
	„ Large Loops ...	– ·333	·344	·253	·261	·375	·293	·302	·280	·309	XII
	„ Whorls	– ·255	·286	·232	·235	·281	·232	·235	·235	·274	XIII
	„ Composites ...	– ·154	·159	·144	·145	·164	·147	·148	·146	·162	XIV
	Small Loops and Large Loops	– ·507	·534	·095	·112	·510	·023	·025	·055	·120	XV
	„ Whorls ...	– ·565	·623	·364	·422	·570	·310	·353	·386	·419	XVI
	„ Composites	– ·365	·392	·298	·309	·401	·281	·293	·301	·311	XVII
	Large Loops and Whorls ...	– ·160	·193	·129	·131	·260	·157	·160	·145	·236	XVIII
	„ Composites	– ·080	·131	·063	·063	·088	·007	·007	·021	·103	XIX
	Whorls and Composites ...	·115	·208	·217	·217	·173	·202	·201	·209	·244	XX
Both Hands	Arches and Small Loops ...	·147	·340	·387	·381	·279	·358	·350	·365	·440	XLVI
	„ Large Loops ...	– ·364	·375	·298	·306	·482	·364	·384	·343	·383	XLVII
	„ Whorls	– ·319	·409	·355	·363	·397	·341	·348	·355	·402	XLVIII
	„ Composites ...	– ·203	·227	·220	·220	·226	·212	·213	·216	·239	XLIX
	Small Loops and Large Loops	– ·471	·527	·162	·187	·476	·059	·071	·115	·234	L
	„ Whorls ...	– ·638	·707	·421	·511	·670	·412	·478	·495	·503	LI
	„ Composites	– ·382	·393	·331	·339	·402	·333	·341	·340	·365	LII
	Large Loops and Whorls ...	– ·147	·194	·178	·178	·323	·228	·234	·204	·333	LIII
	„ Composites	– ·020	·109	·085	·085	·087	·066	·066	·075	·181	LIV
	Whorls and Composites ...	·150	·280	·295	·294	·195	·235	·233	·260	·320	LV

Remarks on Table 8. A comparison of the Correlation Ratio with the Contingency Coefficient of the Restricted Tables.

(*a*) The values of η and C are generally in very close agreement.

(*b*) The value obtained for η is, however, always less than that for C.

(*c*) In only three cases does the difference between η and C exceed 0·1. The probable error of η ranges from ·015 for the smallest values to ·011 for the largest; it will also be remembered that no corrections have been applied to η nor to C, since we do not yet know what these corrections should be for restricted Tables. We may assume, however, that, as with ordinary Tables, correction would modify η less than it would diminish C, and the corrected values of η and C would thus, in all probability, agree somewhat more closely than at present.

TABLE 9.

Right and Left Hands.

	r	C*	C†	η‡	Table in Appendix
Arches R and Arches L	+·686 ± ·008	·664	·688	—	XXI
,, Small Loops L ...	+·160 ± ·015	·285	·302	·234 ± ·014	XXII
,, Large Loops L ...	−·297 ± ·014	·322	·337	—	XXIII
,, Whorls L	−·257 ± ·014	·290	·307	—	XXIV
,, Composites L ...	−·140 ± ·015	·118	·161	—	XXV
Small Loops R and Arches L ...	+·185 ± ·015	·309	·325	·283 ± ·014	XXVI
,, Small Loops L	+·711 ± ·007	·631	·635	—	XXVII
,, Large Loops L	−·378 ± ·013	·382	·393	—	XXVIII
,, Whorls L ...	−·494 ± ·011	·499	·506	—	XXIX
,, Composites L	−·290 ± ·014	·292	·309	—	XXX
Large Loops R and Arches L ...	−·275 ± ·014	·297	·314	—	XXXI
,, Small Loops L	−·217 ± ·014	·262	·282	—	XXXII
,, Large Loops L	+·550 ± ·011	·519	·525	—	XXXIII
,, Whorls L ...	−·123 ± ·015	·210	·235	·159 ± ·015	XXXIV
,, Composites L	−·017 ± ·015	·000	·089	—	XXXV
Whorls R and Arches L	−·308 ± ·014	·337	·351	—	XXXVI
,, Small Loops L ...	−·555 ± ·010	·534	·540	—	XXXVII
,, Large Loops L ...	+·021 ± ·015	·283	·301	·170 ± ·015	XXXVIII
,, Whorls L	+·741 ± ·007	·670	·672	—	XXXIX
,, Composites L ...	+·280 ± ·014	·296	·313	—	XL
Composites R and Arches L ...	−·146 ± ·015	·115	·159	—	XLI
,, Small Loops L	−·188 ± ·014	·172	·203	—	XLII
,, Large Loops L	+ 131 ± ·015	·125	·166	—	XLIII
,, Whorls L ...	+·059 ± ·015	·127	·168	·105 ± ·015	XLIV
,, Composites L...	+·250 ± ·014	·367	·379	—	XLV

Further Remarks on Tables 8 *and* 9. The results given in these Tables show:—

(*a*) A general agreement between the correlations for the same pair of classes of prints whether obtained by different methods from the same Table (omitting values of r in Table 8), or from different Tables, the principal exceptions being those for which the correlation ratio has been calculated in Table 9.

(*b*) A wide range in the magnitude of the results for different pairs of prints.

(*c*) The association between any class of print in one hand and the same class in the other is, in general, as might be expected, much higher than any other association of these Tables. Omitting the composites the remaining four contingency coefficients between the same class in different hands are, with one exception, each greater than any others; the same may be said of the correlation coefficients, the exception in each case being the correlation between whorls in the right and small loops in the left hand, which is slightly greater than the correlation between the large loops in the right and left. Even with the composites the contingency for the two hands is greater than that for composites with

---

* Values of contingency coefficients corrected for number of cells.

† Values of contingency coefficients not corrected for number of cells, given for the sake of comparison with other Tables.

‡ The value of η is in all cases $\sqrt{\eta_{yx}\eta_{xy}}$.

any other class found from any of the Tables, while the correlation coefficients have five exceptions to this general rule.

(*d*) The contingency coefficients given in Table 8, where the two hands are taken together, are, with two exceptions, greater than the corresponding coefficients in other parts of Tables 8 and 9. The exceptions are (1) the contingency coefficient ·234 for small loops with large loops of Table 8 is slightly less than those in Table 9; and (2) the coefficient ·503 for small loops with whorls in Table 8 is rather less than that for whorls (right) with small loops (left) of Table 9.

A further study of the above Tables shows that:—

Large loops are closest to arches.
Arches „ „ whorls.
Whorls „ „ small loops.
Small loops „ „ whorls and then to arches.
Composites „ „ small loops and then to arches.

The suggestion thus arises that arches and whorls have the closest natural resemblance to intermediate sized loops, and also that the "natural order" of the classes of finger-prints is:—

(1) Large Loops, (2) Arches, (3) Whorls, (4) Small Loops, (5) Composites.

This is more clearly seen from the following arrangement of the contingency coefficients.

TABLE 10.

Contingency Coefficients of Right Hand.

	Large Loops	Arches	Whorls	Small Loops	Composites
Large Loops ...	1	·246	·162	·166*	·128
Arches	·246	1	·335	·305	·154*
Whorls	·162	·335	1	·408	·137
Small Loops ...	·166	·305	·408	1	·228
Composites ...	·128	·154	·137	·228	1

TABLE 11.

Contingency Coefficients of Left Hand.

	Large Loops	Arches	Whorls	Small Loops	Composites
Large Loops ...	1	·309	·236	·120	·103
Arches	·309	1	·274	·335*	·162
Whorls	·236	·274	1	·419	·244
Small Loops ...	·120	·335	·419	1	·311
Composites ...	·103	·162	·244	·311	1

* Coefficients which do not agree with the proposed "natural order.

TABLE 12 *a*.

Contingency Coefficients of Right Hand with Left.

Right Hand.

Left Hand.	Large Loops	Arches	Whorls	Small Loops	Composites
Large Loops ...	·519	·322	·283	·382*	·125*
Arches	·297	·664	·337	·309	·115
Whorls	·210	·290	·670	·499	·127
Small Loops ...	·262*	·285	·534	·631	·172
Composites ...	·000	·118	·296*	·292	·367

(Corrected for number of Cells.)

TABLE 12 *b*.

Contingency Coefficients of Right Hand with Left.

Right Hand.

Left Hand.	Large Loops	Arches	Whorls	Small Loops	Composites
Large Loops ...	·525	·337	·301	·393*	·166*
Arches	·314	·688	·351	·325	·159
Whorls	·235	·307	·672	·506	·168
Small Loops ...	·282*	·302	·540	·635	·203
Composites ...	·089	·161	·313*	·309	·379

(Not corrected for number of Cells.)

TABLE 13.

Contingency Coefficients of both Hands taken together.

	Large Loops	Arches	Whorls	Small Loops	Composites
Large Loops ...	1	·383	·333	·234	·181
Arches	·383	1	·402	·440*	·239
Whorls	·333	·402	1	·503	·320
Small Loops ...	·234	·440	·503	1	·361
Composites ...	·181	·239	·320	·361	1

The contingency coefficients of the right hand with the left have been given both corrected and uncorrected for the number of cells and both sets of results point to the same conclusion.

---

* Coefficients which do not agree with the proposed "natural order."

The proposed "natural order" of the types is supported by the above Tables, only eight coefficients out of the fifty-five not being in complete agreement. In four of these cases the difference is very small, most likely well within the probable errors, and they may therefore be regarded as insignificant.

A similar arrangement of the correlation coefficients still further supports the proposed order, though not quite so conclusively, probably on account of spurious correlations.

9. *Association between the various Fingers.* In this section I have calculated the contingency coefficients only, the classes being arranged in the order found in Section 8, p. 445.

It would, of course, be possible to obtain Tables with much finer grouping either by further subdivision of the loops or by making use of the "secondary classification" described by Galton or Henry (see footnote, p. 421). All such finer grouping would raise the contingency; the extra labour involved by the addition of some three or four rows and columns to each Table would, however, be so considerable that the question arises whether some allowance can be made for the coarser grouping employed. This can only be done if we may suppose a "natural order" of some kind with a frequency roughly approaching the normal. This gives a rough upper limit to the contingency and is the purport of the work in the earlier sections on "natural order" and corrections.

As an example of the effect which finer grouping has on contingency I have found the contingency between the index fingers of the two hands by means of a "seven by seven" Table, the radial and ulnar loops being separated, and also by means of a "five by five" Table in which no distinction is drawn between the radial and ulnar loops. The results in this case, not corrected for grouping, are ·653 and ·626; when corrected for grouping these results become ·704 and ·698, respectively. They are so nearly identical as to suggest that no very material advantage would be gained by a further subdivision of classes.

On the assumption that there is a certain degree of continuity in the distribution I have corrected all the results for grouping as well as for the number of cells. The method employed for the former correction is fully described by Professor Pearson in *Biometrika**.

The following Tables give the contingency coefficients for each finger with each other finger. The two sets of coefficients are included, viz. those which are not corrected for grouping, that is, which are obtained without any assumption of a "natural order" and those which are so corrected, in order that the conclusions based on the latter may be compared with those based on the former.

---

* "On the Measurement of the Influence of 'Broad Categories' on Correlation," by Karl Pearson, F.R.S., *Biometrika*, Vol. IX. pp. 116—139.

TABLE 14 *a*.

Contingency Coefficients of Right Hand.

	R_1	R_2	R_3	R_4	R_5
R_1	1	·429 ± ·011	·455	·469	·473
R_2	·429	1	·645	·576	·519
R_3	·455	·645	1	·665	·565
R_4	·469	·576	·665	1	·690
R_5	·473	·519	·565	·690	1

(Corrected for Grouping.)

TABLE 14 *b*.

Contingency Coefficients of Right Hand.

	R_1	R_2	R_3	R_4	R_5
R_1	1	·373 ± ·012	·379	·400	·385
R_2	·373	1	·561	·511	·441
R_3	·379	·561	1	·568	·460
R_4	·400	·511	·568	1	·576
R_5	·385	·441	·460	·576	1

(Not corrected for Grouping.)

TABLE 15 *a*.

Contingency Coefficients of Left Hand.

	L_1	L_2	L_3	L_4	L_5
L_1	1	·503	·465	·474	·508 ± ·012
L_2	·503	1	·675	·609	·539
L_3	·465	·675	1	·724	·585
L_4	·474	·609	·724	1	·711
L_5	·508	·539	·585	·711	1

(Corrected for Grouping.)

TABLE 15 *b*.

Contingency Coefficients of Left Hand.

	L_1	L_2	L_3	L_4	L_5
L_1	1	·435	·390	·401	·410 ± ·014
L_2	·435	1	·582	·529	·447
L_3	·390	·582	1	·611	·471
L_4	·401	·529	·611	1	·577
L_5	·410	·447	·471	·577	1

(Not corrected for Grouping.)

TABLE 16*a*.

Contingency Coefficients of Right Hand with Left.

	R_1	R_2	R_3	R_4	R_5
L_1	·777	·441	·440	·446	·424
L_2	·479	·698 [5 × 5 Table] ·704 [7 × 7 Table]	·640	·559	·521
L_3	·427	·608	·786	·669	·561
L_4	·446	·587	·663	·814	·675
L_5	·501	·515	·537	·648	·899

(Corrected for Grouping.)

TABLE 16*b*.

Contingency Coefficients of Right Hand with Left.

	R_1	R_2	R_3	R_4	R_5
L_1	·649	·385	·368	·383	·347
L_2	·412	·626 [5 × 5 Table] ·653 [7 × 7 Table]	·551	·493	·439
L_3	·356	·530	·656	·572	·459
L_4	·375	·514	·558	·702	·556
L_5	·402	·432	·431	·534	·707

(Not corrected for Grouping.)

Remarks on Tables 14 *a*, 15 *a*, *and* 16 *a*. (*a*) It will be seen from these Tables that the association of types between corresponding fingers of the two hands is, with one exception, always closer than that between any other pair of fingers. The order of magnitude of these associations is:—

(1) Little Finger, (2) Ring Finger, (3) Middle Finger, (4) Thumb, (5) Index Finger.

(*b*) If we omit the thumb for the present, leaving it for separate comment, and consider the association between corresponding fingers as of the "first order," that between fingers of consecutive rank, such as R_2 and R_3, or R_2 and L_3 as of the "second order," and so on, we notice a significant relation between any particular association and its "order." Thus:

First order associations range from ·899 to ·704 or ·698,
Second " " " " ·724 to ·608,
Third " " " " ·609 to ·537,
Fourth " " " " ·539 to ·515.

The amount of overlapping in these ranges appears to be quite insignificant.

(*c*) It follows from (*b*) that if in any of these Tables we start from a first order association and pass in any direction through those of other orders we find a continuous and rapid fall; that is, a finger is always more closely related to a consecutive finger than to one more remote (but see (*a*)); and the greater the difference in rank between two fingers, whether on the same or on different hands, the less close is the association between them.

(*d*) The association between any pair of fingers in one hand is, in general, closer than either of the corresponding associations between a finger of the right and one of the left hand. There is one exception to this rule in associations of the second order, one in the third and one in the fourth.

(*e*) The associations of the left hand are in every case closer than the corresponding associations of the right.

(*f*) The associations of either thumb with any finger all fall below those of the fourth order of (*b*), and the range of the sixteen coefficients is only from ·424 to ·508. As it is difficult to base any conclusions on these figures as to the relations between the thumb and the various fingers, I have carefully checked them by reworking the whole of the calculations involved, but have in every case arrived at the same result. I have also found the probable error* for the largest and for one of the smallest coefficients of the set. As the contingency coefficients are all of the same order of magnitude and the number of individuals the same in all cases, the probable errors of all will be of about the same magnitude and it is unnecessary to calculate more. The probable errors in the two cases being of the order ·011 the differences in the contingency coefficients may be regarded as insignificant. Although in three cases out of the four the contingencies of the thumb with the middle, ring and little finger respectively are in ascending order of magnitude, the differences are so small in comparison with the probable errors that no conclusion can be drawn as to the relations between the thumb and the various fingers. We may notice, however, that the rule (*d*) holds good for the thumbs with but two exceptions.

The contingency coefficients given in Tables 14 *b*, 15 *b*, and 16 *b*, are all smaller than the corresponding results of the other series, but a careful study will show that the remarks (*a*) to (*g*) almost invariably apply to these Tables also.

Note. In some preliminary work on this paper I classified the types as follows:—(1) Arches and loops with 1—3 ridges, (2) Loops with 4—10 ridges, (3) Loops with 11—14 ridges, (4) Loops with 15 or more ridges, (5) Whorls, (6) Composites. With this classification the following contingency coefficients were found for corresponding fingers of the two hands:—Thumb ·686, Forefinger ·642, Middle finger ·686, Ring finger ·730, Little finger ·738. These results, which were not corrected for grouping, are seen to agree very closely with those

---

* The method employed is that given in *Biometrika*, Vol. v. Parts I. and II., "On the Probable Error of Mean-Square Contingency," by John Blakeman and Karl Pearson.

of Table 16 *b*, the values being rather larger probably on account of the slightly finer grouping.

10. *Comparison with Results of Previous Work.* It would be well to compare briefly some of my results with those of the two works mentioned on p. 421.

Whiteley and Pearson arrived at the following conclusions:—

(i) The hand is a very highly correlated organ, far more highly correlated than the skull and even somewhat more so than the long bones.

(ii) The parts of the left hand are distinctly more closely correlated than those of the right.

(iii) The order of correlation of the first finger joints is identical for both hands. This order is as follows:—

(*a*) The external fingers have the least correlation and the little finger always less than the index.

(*b*) A finger has always more correlation with a second than with any other finger from which it is separated by the second.

(iv) With corresponding members on both sides the extreme pairs show least correlation, and the pair of middle fingers higher correlation than the pair of ring fingers.

In the paper of Miss Lewenz and Miss Whiteley the chief results which are comparable with those for the finger-prints are the following:—

(v) There is a slight, but we cannot say definitely significant, preponderance in the correlations of the right hand bones over those of the left.

(vi) Dividing the hand into marginal members, i.e. thumb, index and little fingers, and central members, i.e. middle and ring fingers, and the bones into "lower bones," i.e. distal and middle phalanges, and "upper bones," i.e. metacarpal bones and proximal phalanges, the correlations roughly speaking are highest for the upper bones of the central members and become less as we move out from this upper centre towards the lower and marginal parts of the hand. This is true whether we take pairs in lateral or in longitudinal series.

(vii) The highest correlations occur between corresponding bones of the right and left hands.

(viii) Generally there is a "rule of neighbourhood," i.e. any bone is more closely correlated with a second of the same series than with any other from which it is separated by that second.

The above conclusions are to a certain extent mutually corroborative: e.g. (vi) and (iv) are in agreement, and (viii) agrees in substance with (iii *b*). Again (vii) agrees with Table IV, p. 130, of the "First Study," while (iii *a*) is in general supported by Table XXII of the "Second Study." On the other hand (ii) and (v) do not agree. It should be noted, however, that the "First Study" was based on the measurements of the first finger joint only of both hands of 551 women, while for the "Second Study," in which all the finger bones were measured, only 37 to 44 skeleton hands were available. The writers of the latter paper state that in consequence of the comparatively small number of bones measured they look upon that study "as one of suggestion rather than of definite statistical proof," and it is possible that with more adequate data their results might have been somewhat modified and exceptions less numerous.

There appears to be no direct connection between finger bones and the patterns of finger-prints, but it is distinctly interesting to find that some of the most striking relations discovered amongst the former also exist in the latter. In particular, my conclusion (*a*) agrees with (vii), (*c*) with (iii *b*) and with (viii), and (*e*) with (ii) but not with (v).

11. *Concluding Remarks.* The most important conclusions reached in this paper have been summarized on p. 432, and in the Remarks on Tables 8 and 9, pp. 443—445, and on Tables 14—16, p. 449; it scarcely seems necessary to recapitulate them, but a comparison will show an almost perfect agreement although the sets of results have been obtained by entirely different methods.

The essential results of the present paper are that finger-prints are not scattered at random over the fingers; certain types are more or less peculiar to certain fingers, and further the appearance of one type is correlated with the appearance of a second. In this respect certain fingers are more closely related to each other than to any third finger, and the distribution of this relationship is in general similar to what is known of the like distribution of the correlations of the bones of the same fingers.

It has been already stated that the material used is taken entirely from adult males of the lower type of the artisan and labouring classes; it would be of interest to compare the results obtained with those found from the finger-prints of females of the same grade of society, and also when the material is drawn from the professional classes.

Tables I to XX, and XLVI to LV, are of a type which I have not previously met with; novel methods have accordingly been employed in calculating coefficients of contingency and correlation ratios from those Tables. The general investigation of Tables of this type offers an interesting problem, demanding further study.

I am deeply grateful to Professor Pearson for placing at my disposal the necessary material together with a number of books and memoirs bearing on the subject, and for much valuable assistance given during the course of the investigation.

It can scarcely be expected, with such a mass of numerical calculation involved, that the work should be entirely free from inaccuracies, but I trust that no serious errors have escaped detection. The laborious arithmetic has been much lightened by the use of a calculator, for the loan of which my thanks are due to the Government Grant Committee of the Royal Society of London.

The Tables on which the preceding calculations are based are given in the Appendix, pp. 453—478.

APPENDIX.

TABLE I.

Arches and Small Loops, Right.

Arches, *R.*

Small Loops, *R.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	544	26	16	1	1	5	593
1	343	65	16	14	15	—	453
2	256	73	45	18	—	—	392
3	185	87	34	—	—	—	306
4	168	43	—	—	—	—	211
5	45	—	—	—	—	—	45
Totals	1541	294	111	33	16	5	2000

TABLE II.

Arches and Large Loops, Right.

Arches, *R.*

Large Loops, *R.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	286	78	49	23	16	5	457
1	489	114	46	10	0	—	659
2	400	66	12	0	—	—	478
3	245	31	4	—	—	—	280
4	103	5	—	—	—	—	108
5	18	—	—	—	—	—	18
Totals	1541	294	111	33	16	5	2000

TABLE III.

Arches and Whorls, Right.

Arches, *R.*

Whorls, *R.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	512	215	86	27	16	5	861
1	406	61	24	6	0	—	497
2	275	16	1	0	—	—	292
3	168	2	0	—	—	—	170
4	130	0	—	—	—	—	130
5	50	—	—	—	—	—	50
Totals	1541	294	111	33	16	5	2000

TABLE IV.

Arches and Composites, Right.

Arches, *R.*

Composites, *R.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	1042	232	96	33	15	5	1423
1	376	52	13	0	1	—	442
2	107	9	2	0	—	—	118
3	13	1	0	—	—	—	14
4	3	0	—	—	—	—	3
5	0	—	—	—	—	—	0
Totals	1541	294	111	33	16	5	2000

TABLE V.

Small Loops and Large Loops, Right.

Small Loops, *R*.

Large Loops, *R*.	0	1	2	3	4	5	Totals
0	104	69	67	81	91	45	457
1	144	99	142	154	120	—	659
2	143	141	123	71	—	—	478
3	119	101	60	—	—	—	280
4	65	43	—	—	—	—	108
5	18	—	—	—	—	—	18
Totals	593	453	392	306	211	45	2000

TABLE VI.

Small Loops and Whorls, Right.

Small Loops, *R*.

Whorls, *R*.	0	1	2	3	4	5	Totals
0	78	144	204	211	179	45	861
1	106	153	126	80	32	—	497
2	130	92	55	15	—	—	292
3	125	38	7	—	—	—	170
4	104	26	—	—	—	—	130
5	50	—	—	—	—	—	50
Totals	593	453	392	306	211	45	2000

TABLE VII.

Small Loops and Composites, Right.

Small Loops, *R*.

Composites, *R*.	0	1	2	3	4	5	Totals
0	353	294	279	257	195	45	1423
1	165	121	93	47	16	—	442
2	62	34	20	2	—	—	118
3	10	4	0	—	—	—	14
4	3	0	—	—	—	—	3
5	0	—	—	—	—	—	0
Totals	593	453	392	306	211	45	2000

TABLE VIII.

Large Loops and Whorls, Right.

Large Loops, *R*.

Whorls, *R*.	0	1	2	3	4	5	Totals
0	197	289	162	133	62	18	861
1	86	130	148	87	46	—	497
2	47	84	101	60	—	—	292
3	27	76	67	—	—	—	170
4	50	80	—	—	—	—	130
5	50	—	—	—	—	—	50
Totals	457	659	478	280	108	18	2000

TABLE IX.

Large Loops and Composites, Right.

Large Loops, *R.*

		0	*1*	*2*	*3*	*4*	*5*	Totals
Composites, *R.*	*0*	317	471	318	205	94	18	1423
	1	88	153	130	57	14	—	442
	2	41	32	27	18	—	—	118
	3	9	2	3	—	—	—	14
	4	2	1	—	—	—	—	3
	5	0	—	—	—	—	—	0
	Totals	457	659	478	280	108	18	2000

TABLE X.

Whorls and Composites, Right.

Whorls, *R.*

		0	*1*	*2*	*3*	*4*	*5*	Totals
Composites, *R.*	*0*	648	331	180	108	106	50	1423
	1	160	126	84	48	24	—	442
	2	49	31	24	14	—	—	118
	3	3	7	4	—	—	—	14
	4	1	2	—	—	—	—	3
	5	0	—	—	—	—	—	0
	Totals	861	497	292	170	130	50	2000

TABLE XI.

Arches and Small Loops, Left.

Arches, *L.*

		0	*1*	*2*	*3*	*4*	*5*	Totals
Small Loops, *L.*	*0*	463	22	6	0	1	4	496
	1	328	44	23	11	19	—	425
	2	248	59	28	31	—	—	366
	3	189	81	45	—	—	—	315
	4	199	84	—	—	—	—	283
	5	115	—	—	—	—	—	115
	Totals	1542	290	102	42	20	4	2000

TABLE XII.

Arches and Large Loops, Left.

Arches, *L.*

		0	*1*	*2*	*3*	*4*	*5*	Totals
Large Loops, *L.*	*0*	286	115	56	34	20	4	515
	1	437	101	31	8	0	—	577
	2	361	49	13	0	—	—	423
	3	289	20	2	—	—	—	311
	4	136	5	—	—	—	—	141
	5	33	—	—	—	—	—	33
	Totals	1542	290	102	42	20	4	2000

TABLE XIII.

Arches and Whorls, Left.

Arches, *L.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	766	233	89	41	19	4	1152
1	350	44	12	1	1	—	408
2	193	9	1	0	—	—	203
3	120	4	0	—	—	—	124
4	92	0	—	—	—	—	92
5	21	—	—	—	—	—	21
Totals	1542	290	102	42	20	4	2000

TABLE XIV.

Arches and Composites, Left.

Arches, *L.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	1048	237	87	40	20	4	1436
1	375	42	15	2	0	—	434
2	94	9	0	0	—	—	103
3	20	2	0	—	—	—	22
4	4	0	—	—	—	—	4
5	1	—	—	—	—	—	1
Totals	1542	290	102	42	20	4	2000

TABLE XV.

Small Loops and Large Loops, Left.

Small Loops, *L.*

Large Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	82	48	63	83	124	115	515
1	121	73	94	130	159	—	577
2	97	116	108	102	—	—	423
3	100	110	101	—	—	—	311
4	63	78	—	—	—	—	141
5	33	—	—	—	—	—	33
Totals	496	425	366	315	283	115	2000

TABLE XVI.

Small Loops and Whorls, Left.

Small Loops, *L.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	96	183	248	251	259	115	1152
1	103	139	85	57	24	—	408
2	105	63	28	7	—	—	203
3	88	31	5	—	—	—	124
4	83	9	—	—	—	—	92
5	21	—	—	—	—	—	21
Totals	496	425	366	315	283	115	2000

TABLE XVII.

Small Loops and Composites, Left.

Small Loops, *L.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	234	292	264	264	267	115	1436
1	176	110	84	48	16	—	434
2	64	19	17	3	—	—	103
3	17	4	1	—	—	—	22
4	4	0	—	—	—	—	4
5	1	—	—	—	—	—	1
Totals	496	425	366	315	283	115	2000

TABLE XVIII.

Large Loops and Whorls, Left.

Large Loops, *L.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	338	315	195	165	106	33	1152
1	63	94	114	102	35	—	408
2	22	58	79	44	—	—	203
3	26	63	35	—	—	—	124
4	45	47	—	—	—	—	92
5	21	—	—	—	—	—	21
Totals	515	577	423	311	141	33	2000

TABLE XIX.

Large Loops and Composites, Left.

Large Loops, *L.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	380	399	276	230	118	33	1436
1	102	124	119	66	23	—	434
2	23	41	24	15	—	—	103
3	7	11	4	—	—	—	22
4	2	2	—	—	—	—	4
5	1	—	—	—	—	—	1
Totals	515	577	423	311	141	33	2000

TABLE XX.

Whorls and Composites, Left.

Whorls, *L.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	924	249	120	66	56	21	1436
1	173	120	59	46	36	—	434
2	44	26	21	12	—	—	103
3	8	11	3	—	—	—	22
4	2	2	—	—	—	—	4
5	1	—	—	—	—	—	1
Totals	1152	408	203	124	92	21	2000

TABLE XXI.

Arches, Right, and Arches, Left.

Arches, *R.*

Arches, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	1369	145	24	3	1	0	1542
1	146	97	40	7	0	0	290
2	24	45	26	5	2	0	102
3	2	6	16	12	6	0	42
4	0	1	5	6	5	3	20
5	0	0	0	0	2	2	4
Totals	1541	294	111	33	16	5	2000

TABLE XXII.

Arches, Right, and Small Loops, Left.

Arches, *R.*

Small Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	473	14	5	0	2	2	496
1	342	47	17	10	6	3	425
2	249	68	29	15	5	0	366
3	209	74	28	2	2	0	315
4	187	66	26	4	0	0	283
5	81	25	6	2	1	0	115
Totals	1541	294	111	33	16	5	2000

TABLE XXIII.

Arches, Right, and Large Loops, Left.

Arches, *R.*

Large Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	304	95	68	28	15	5	515
1	440	104	28	4	1	0	577
2	363	50	9	1	0	0	423
3	269	36	6	0	0	0	311
4	133	8	0	0	0	0	141
5	32	1	0	0	0	0	33
Totals	1541	294	111	33	16	5	2000

TABLE XXIV.

Arches, Right, and Whorls, Left.

Arches, *R.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	760	243	97	31	16	5	1152
1	354	42	11	1	0	0	408
2	193	8	2	0	0	0	203
3	121	1	1	1	0	0	124
4	92	0	0	0	0	0	92
5	21	0	0	0	0	0	21
Totals	1541	294	111	33	16	5	2000

TABLE XXV.

Arches, Right, and Composites, Left.

		Arches, *R.*						
		0	*1*	*2*	*3*	*4*	*5*	Totals
Composites, *L.*	*0*	1048	244	95	28	16	5	1436
	1	373	44	13	4	0	0	434
	2	95	5	2	1	0	0	103
	3	20	1	1	0	0	0	22
	4	4	0	0	0	0	0	4
	5	1	0	0	0	0	0	1
	Totals	1541	294	111	33	16	5	2000

TABLE XXVI.

Small Loops, Right, and Arches, Left.

		Small Loops, *R.*						
		0	*1*	*2*	*3*	*4*	*5*	Totals
Arches, *L.*	*0*	558	369	279	177	129	30	1542
	1	25	49	74	74	57	11	290
	2	4	15	18	40	21	4	102
	3	1	13	14	11	3	0	42
	4	3	5	7	4	1	0	20
	5	2	2	0	0	0	0	4
	Totals	593	453	392	306	211	45	2000

TABLE XXVII.

Small Loops, Right, and Small Loops, Left.

		Small Loops, *R.*						
		0	*1*	*2*	*3*	*4*	*5*	Totals
Small Loops, *L.*	*0*	373	97	21	5	0	0	496
	1	148	149	88	35	4	1	425
	2	53	118	103	65	26	1	366
	3	14	56	98	86	55	6	315
	4	4	30	63	90	78	18	283
	5	1	3	19	25	48	19	115
	Totals	593	453	392	306	211	45	2000

TABLE XXVIII.

Small Loops, Right, and Large Loops, Left.

		Small Loops, *R.*						
		0	*1*	*2*	*3*	*4*	*5*	Totals
Large Loops, *L.*	*0*	92	76	89	109	115	34	515
	1	154	103	125	121	65	9	577
	2	137	117	90	52	25	2	423
	3	117	107	60	21	6	0	311
	4	70	42	27	2	0	0	141
	5	23	8	1	1	0	0	33
	Totals	593	453	392	306	211	45	2000

TABLE XXIX.

Small Loops, Right, and Whorls, Left.

Small Loops, *R.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	159	217	286	249	198	43	1152
1	130	143	79	43	12	1	408
2	112	58	19	13	1	0	203
3	92	25	6	1	0	0	124
4	79	10	2	0	0	1	92
5	21	0	0	0	0	0	21
Totals	593	453	392	306	211	45	2000

TABLE XXX.

Small Loops, Right, and Composites, Left.

Small Loops, *R.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	323	311	299	267	191	45	1436
1	194	109	82	32	17	0	434
2	60	22	11	7	3	0	103
3	12	10	0	0	0	0	22
4	4	0	0	0	0	0	4
5	0	1	0	0	0	0	1
Totals	593	453	392	306	211	45	2000

TABLE XXXI.

Large Loops, Right, and Arches, Left.

Large Loops, *R.*

Arches, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	290	483	394	255	103	17	1542
1	69	123	70	22	5	1	290
2	51	36	12	3	0	0	102
3	24	16	2	0	0	0	42
4	19	1	0	0	0	0	20
5	4	0	0	0	0	0	4
Totals	457	659	478	280	108	18	2000

TABLE XXXII.

Large Loops, Right, and Small Loops, Left.

Large Loops, *R.*

Small Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	112	134	139	73	32	6	496
1	68	110	110	89	41	7	425
2	76	114	91	52	29	4	366
3	65	128	81	38	3	0	315
4	90	121	45	23	3	1	283
5	46	52	12	5	0	0	115
Totals	457	659	478	280	108	18	2000

TABLE XXXIII.

Large Loops, Right, and Large Loops, Left.

Large Loops, *R.*

Large Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	264	178	51	19	3	0	515
1	121	256	135	59	5	1	577
2	53	135	149	67	18	1	423
3	17	66	101	77	41	9	311
4	2	20	36	47	32	4	141
5	0	4	6	11	9	3	33
Totals	457	659	478	280	108	18	2000

TABLE XXXIV.

Large Loops, Right, and Whorls, Left.

Large Loops, *R.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	262	399	238	163	75	15	1152
1	69	114	125	73	24	3	408
2	35	66	66	31	5	0	203
3	37	41	34	9	3	0	124
4	40	37	11	3	1	0	92
5	14	2	4	1	0	0	21
Totals	457	659	478	280	108	18	2000

TABLE XXXV.

Large Loops, Right, and Composites, Left.

Large Loops, *R.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	331	478	338	195	81	13	1436
1	90	140	107	70	23	4	434
2	27	34	24	13	4	1	103
3	6	5	9	2	0	0	22
4	2	2	0	0	0	0	4
5	1	0	0	0	0	0	1
Totals	457	659	478	280	108	18	2000

TABLE XXXVI.

Whorls, Right, and Arches, Left.

Whorls, *R.*

Arches, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	526	405	262	169	130	50	1542
1	193	69	27	1	0	0	290
2	80	19	3	0	0	0	102
3	38	4	0	0	0	0	42
4	20	0	0	0	0	0	20
5	4	0	0	0	0	0	4
Totals	861	497	292	170	130	50	2000

TABLE XXXVII.

Whorls, Right, and Small Loops, Left.

Whorls, *R.*

Small Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	59	90	97	108	94	48	496
1	137	120	95	44	27	2	425
2	170	126	50	12	8	0	366
3	194	80	35	5	1	0	315
4	202	65	15	1	0	0	283
5	99	16	0	0	0	0	115
Totals	861	497	292	170	130	50	2000

TABLE XXXVIII.

Whorls, Right, and Large Loops, Left.

Whorls, *R.*

Large Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	311	83	36	29	28	28	515
1	241	134	80	46	58	18	577
2	126	128	87	49	32	1	423
3	112	93	63	34	7	2	311
4	58	47	21	11	3	1	141
5	13	12	5	1	2	0	33
Totals	861	497	292	170	130	50	2000

TABLE XXXIX.

Whorls, Right, and Whorls, Left.

Whorls, *R.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	768	281	77	17	9	0	1152
1	79	160	110	41	15	3	408
2	11	42	71	45	32	2	203
3	2	10	25	36	40	11	124
4	1	4	7	27	32	21	92
5	0	0	2	4	2	13	21
Totals	861	497	292	170	130	50	2000

TABLE XL.

Whorls, Right, and Composites, Left.

Whorls, *R.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	744	331	185	93	55	28	1436
1	98	132	78	58	52	16	434
2	16	28	19	16	19	5	103
3	3	5	7	2	4	1	22
4	0	1	2	1	0	0	4
5	0	0	1	0	0	0	1
Totals	861	497	292	170	130	50	2000

TABLE XLI.

Composites, Right, and Arches, Left.

Composites, *R.*

Arches, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	1042	378	107	12	3	0	1542
1	233	46	9	2	0	0	290
2	84	16	2	0	0	0	102
3	40	2	0	0	0	0	42
4	20	0	0	0	0	0	20
5	4	0	0	0	0	0	4
Totals	1423	442	118	14	3	0	2000

TABLE XLII.

Composites, Right, and Small Loops, Left.

Composites, *R.*

Small Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	299	135	52	8	2	0	496
1	294	100	26	5	0	0	425
2	262	82	20	1	1	0	366
3	237	71	7	0	0	0	315
4	233	38	12	0	0	0	283
5	98	16	1	0	0	0	115
Totals	1423	442	118	14	3	0	2000

TABLE XLIII.

Composites, Right, and Large Loops, Left.

Composites, *R.*

Large Loops, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	416	80	17	2	0	0	515
1	416	121	36	3	1	0	577
2	278	111	27	6	1	0	423
3	210	79	20	1	1	0	311
4	86	37	16	2	0	0	141
5	17	14	2	0	0	0	33
Totals	1423	442	118	14	3	0	2000

TABLE XLIV.

Composites, Right, and Whorls, Left.

Composites, *R.*

Whorls, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	878	219	50	3	2	0	1152
1	245	116	40	6	1	0	408
2	129	55	17	2	0	0	203
3	89	28	5	2	0	0	124
4	64	21	6	1	0	0	92
5	18	3	0	0	0	0	21
Totals	1423	442	118	14	3	0	2000

TABLE XLV.

Composites, Right, and Composites, Left.

Composites, *R.*

Composites, *L.*	*0*	*1*	*2*	*3*	*4*	*5*	Totals
0	1100	276	53	6	1	0	1436
1	258	130	43	3	0	0	434
2	57	28	14	3	1	0	103
3	8	8	4	2	0	0	22
4	0	0	3	0	1	0	4
5	0	0	1	0	0	0	1
Totals	1423	442	118	14	3	0	2000

TABLE XLVI.

Arches and Small Loops, Both Hands.

Arches.

Small Loops.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
0	361	7	1	1	1	0	0	0	0	0	2	373
1	205	19	7	6	3	0	0	0	0	5	—	245
2	168	23	12	7	1	0	5	2	5	—	—	223
3	139	36	13	9	8	6	4	10	—	—	—	225
4	105	40	25	10	2	6	10	—	—	—	—	198
5	102	33	25	18	10	10	—	—	—	—	—	198
6	88	37	21	18	15	—	—	—	—	—	—	179
7	74	47	23	21	—	—	—	—	—	—	—	165
8	63	28	18	—	—	—	—	—	—	—	—	109
9	45	21	—	—	—	—	—	—	—	—	—	66
10	19	—	—	—	—	—	—	—	—	—	—	19
Totals	1369	291	145	90	40	22	19	12	5	5	2	2000

TABLE XLVII.

Arches and Large Loops, Both Hands.

Arches.

Large Loops.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
0	109	32	24	32	19	13	12	11	5	5	2	264
1	164	55	34	23	11	6	5	1	0	0	—	299
2	230	70	33	18	4	3	2	0	0	—	—	360
3	225	41	27	9	4	0	0	0	—	—	—	306
4	212	45	16	5	1	0	0	—	—	—	—	279
5	160	22	9	2	0	0	—	—	—	—	—	193
6	119	13	2	1	1	—	—	—	—	—	—	136
7	86	9	0	0	—	—	—	—	—	—	—	95
8	49	3	0	—	—	—	—	—	—	—	—	52
9	12	1	—	—	—	—	—	—	—	—	—	13
10	3	—	—	—	—	—	—	—	—	—	—	3
Totals	1369	291	145	90	40	22	19	12	5	5	2	2000

TABLE XLVIII.

Arches and Whorls, Both Hands.

Arches.

Whorls.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
0	351	160	100	66	30	21	16	12	5	5	2	768
1	240	69	25	16	7	1	2	0	0	0	—	360
2	191	38	11	5	2	0	1	0	0	—	—	248
3	146	14	8	2	1	0	0	0	—	—	—	171
4	124	6	1	1	0	0	0	—	—	—	—	132
5	86	3	0	0	0	0	—	—	—	—	—	89
6	77	1	0	0	0	—	—	—	—	—	—	78
7	71	0	0	0	—	—	—	—	—	—	—	71
8	47	0	0	—	—	—	—	—	—	—	—	47
9	23	0	—	—	—	—	—	—	—	—	—	23
10	13	—	—	—	—	—	—	—	—	—	—	13
Totals	1369	291	145	90	40	22	19	12	5	5	2	2000

TABLE XLIX.

Arches and Composites, Both Hands.

Arches.

Composites.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
0	650	193	101	62	33	18	18	11	5	5	2	1098
1	406	66	34	19	6	3	1	1	0	0	—	536
2	202	22	7	7	1	1	0	0	0	—	—	240
3	74	7	3	1	0	0	0	0	—	—	—	85
4	22	3	0	1	0	0	0	—	—	—	—	26
5	7	0	0	0	0	0	—	—	—	—	—	7
6	6	0	0	0	0	—	—	—	—	—	—	6
7	1	0	0	0	—	—	—	—	—	—	—	1
8	1	0	0	—	—	—	—	—	—	—	—	1
9	0	0	—	—	—	—	—	—	—	—	—	0
10	0	—	—	—	—	—	—	—	—	—	—	0
Totals	1369	291	145	90	40	22	19	12	5	5	2	2000

TABLE L.

Small Loops and Large Loops, Both Hands.

Small Loops.

Large Loops.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
0	44	13	10	18	17	20	31	31	31	30	19	264
1	44	19	13	19	14	30	37	45	42	36	—	299
2	60	24	24	30	35	42	54	55	36	—	—	360
3	55	31	25	39	43	47	32	34	—	—	—	306
4	49	34	41	44	46	40	25	—	—	—	—	279
5	35	39	38	33	29	19	—	—	—	—	—	193
6	34	32	34	22	14	—	—	—	—	—	—	136
7	24	29	22	20	—	—	—	—	—	—	—	95
8	16	20	16	—	—	—	—	—	—	—	—	52
9	9	4	—	—	—	—	—	—	—	—	—	13
10	3	—	—	—	—	—	—	—	—	—	—	3
Totals	373	245	223	225	198	198	179	165	109	66	19	2000

TABLE LI.

Small Loops and Whorls, Both Hands.

Small Loops.

		0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
Whorls.	*0*	23	34	52	68	81	105	111	126	87	62	19	768
	1	19	43	39	55	59	50	42	30	19	4	—	360
	2	29	46	44	47	29	26	17	7	3	—	—	248
	3	32	34	38	30	15	14	6	2	—	—	—	171
	4	46	27	25	16	12	3	3	—	—	—	—	132
	5	47	21	14	5	2	0	—	—	—	—	—	89
	6	52	17	6	3	0	—	—	—	—	—	—	78
	7	50	16	4	1	—	—	—	—	—	—	—	71
	8	39	7	1	—	—	—	—	—	—	—	—	47
	9	23	0	—	—	—	—	—	—	—	—	—	23
	10	13	—	—	—	—	—	—	—	—	—	—	13
Totals		373	245	223	225	198	198	179	165	109	66	19	2000

TABLE LII.

Small Loops and Composites, Both Hands.

Small Loops.

		0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
Composites.	*0*	114	92	113	115	111	121	121	136	95	61	19	1098
	1	128	70	71	64	58	60	41	26	13	5	—	536
	2	78	46	23	37	21	15	16	3	1	—	—	240
	3	32	24	13	6	8	1	1	0	—	—	—	85
	4	13	7	2	3	0	1	0	—	—	—	—	26
	5	2	4	1	0	0	0	—	—	—	—	—	7
	6	5	1	0	0	0	—	—	—	—	—	—	6
	7	0	1	0	0	—	—	—	—	—	—	—	1
	8	1	0	0	—	—	—	—	—	—	—	—	1
	9	0	0	—	—	—	—	—	—	—	—	—	0
	10	0	—	—	—	—	—	—	—	—	—	—	0
Totals		373	245	223	225	198	198	179	165	109	66	19	2000

TABLE LIII.

Large Loops and Whorls, Both Hands.

Large Loops.

		0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
Whorls.	*0*	156	138	133	91	73	51	43	42	29	9	3	768
	1	24	45	54	63	56	39	30	29	16	4	—	360
	2	12	24	36	35	50	35	30	19	7	—	—	248
	3	12	16	30	21	35	35	17	5	—	—	—	171
	4	8	7	20	32	26	23	16	—	—	—	—	132
	5	3	8	20	20	28	10	—	—	—	—	—	89
	6	4	9	24	30	11	—	—	—	—	—	—	78
	7	10	18	29	14	—	—	—	—	—	—	—	71
	8	11	22	14	—	—	—	—	—	—	—	—	47
	9	11	12	—	—	—	—	—	—	—	—	—	23
	10	13	—	—	—	—	—	—	—	—	—	—	13
Totals		264	299	360	306	279	193	136	95	52	13	3	2000

TABLE LIV.

Large Loops and Composites, Both Hands.

Large Loops.

Composites.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
0	180	174	188	144	130	104	79	51	36	9	3	1098
1	39	76	100	96	94	50	32	32	13	4	—	536
2	28	29	45	40	42	30	16	7	3	—	—	240
3	10	11	19	17	8	8	7	5	—	—	—	85
4	3	6	5	6	3	1	2	—	—	—	—	26
5	1	1	2	2	1	0	—	—	—	—	—	7
6	2	1	1	1	1	—	—	—	—	—	—	6
7	1	0	0	0	—	—	—	—	—	—	—	1
8	0	1	0	—	—	—	—	—	—	—	—	1
9	0	0	—	—	—	—	—	—	—	—	—	0
10	0	—	—	—	—	—	—	—	—	—	—	0
Totals	264	299	360	306	279	193	136	95	52	13	3	2000

TABLE LV.

Whorls and Composites, Both Hands.

Whorls.

Composites.	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Totals
0	555	172	109	71	68	26	23	27	22	12	13	1098
1	148	114	71	47	30	40	34	26	15	11	—	536
2	42	52	41	33	22	12	16	12	10	—	—	240
3	17	15	15	12	8	8	4	6	—	—	—	85
4	4	4	7	5	2	3	1	—	—	—	—	26
5	1	1	3	1	1	0	—	—	—	—	—	7
6	1	1	1	2	1	—	—	—	—	—	—	6
7	0	0	1	0	—	—	—	—	—	—	—	1
8	0	1	0	—	—	—	—	—	—	—	—	1
9	0	0	—	—	—	—	—	—	—	—	—	0
10	0	—	—	—	—	—	—	—	—	—	—	0
Totals	768	360	248	171	132	89	78	71	47	23	13	2000

TABLE LVI.

Right Thumb and Index.

Right Thumb.

Right Index.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	29	97	148	50	28	352
SL	12	125	320	139	58	654
LL	2	27	149	125	36	339
W	1	26	144	260	50	481
C	2	15	54	75	28	174
Totals	46	290	815	649	200	2000

TABLE LVII.

Right Thumb and Middle Finger.

Right Thumb.

Right Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	25	66	75	31	15	212
SL	18	174	428	215	86	921
LL	1	26	223	208	58	516
W	2	15	69	160	28	274
C	0	9	20	35	13	77
Totals	46	290	815	649	200	2000

TABLE LVIII.

Right Thumb and Ring Finger.

Right Thumb.

Right Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	14	24	17	7	1	63
SL	17	129	248	64	31	489
LL	5	60	248	166	64	543
W	8	55	229	355	82	729
C	2	22	73	57	22	176
Totals	46	290	815	649	200	2000

TABLE LIX.

Right Thumb and Little Finger.

Right Thumb.

Right Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	8	14	5	3	1	31
SL	30	198	414	162	66	870
LL	5	61	300	304	94	764
W	1	10	58	137	22	228
C	2	7	38	43	17	107
Totals	46	290	815	649	200	2000

TABLE LX.

Right Index and Middle Finger.

Right Index.

Right Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	127	69	9	5	2	212
SL	187	453	119	104	58	921
LL	30	108	144	172	62	516
W	4	16	48	178	28	274
C	4	8	19	22	24	77
Totals	352	654	339	481	174	2000

TABLE LXI.

Right Index and Ring Finger.

Right Index.

Right Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	45	15	3	0	0	63
SL	161	274	17	21	16	489
LL	84	189	127	92	51	543
W	36	130	149	325	89	729
C	26	46	43	43	18	176
Totals	352	654	339	481	174	2000

TABLE LXII.

Right Index and Little Finger.

Right Index.

Right Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	23	8	0	0	0	31
SL	222	402	97	106	43	870
LL	95	199	181	199	90	764
W	4	26	38	136	24	228
C	8	19	23	40	17	107
Totals	352	654	339	481	174	2000

TABLE LXIII.

Right Middle and Ring Fingers.

Right Middle Finger.

Right Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	49	14	0	0	0	63
SL	104	336	43	2	4	489
LL	32	298	181	16	16	543
W	16	189	240	240	44	729
C	11	84	52	16	13	176
Totals	212	921	516	274	77	2000

TABLE LXIV.

Right Middle and Little Fingers.

Right Middle Finger.

Right Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	20	11	0	0	0	31
SL	147	535	129	39	20	870
LL	37	298	280	114	35	764
W	0	47	74	92	15	228
C	8	30	33	29	7	107
Totals	212	921	516	274	77	2000

TABLE LXV.

Right Ring and Little Fingers.

Right Ring Finger.

Right Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	16	13	1	1	0	31
SL	41	400	212	155	62	870
LL	4	67	313	297	83	764
W	0	6	11	199	12	228
C	2	3	6	77	19	107
Totals	63	489	543	729	176	2000

TABLE LXVI.

Left Thumb and Index.

Left Thumb.

Left Index.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	47	133	78	25	30	313
SL	32	313	355	68	65	833
LL	5	40	143	47	47	282
W	3	46	136	166	86	437
C	4	15	55	35	26	135
Totals	91	547	767	341	254	2000

TABLE LXVII.

Left Thumb and Middle Finger.

Left Thumb.

Left Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	31	96	51	17	20	215
SL	46	323	354	85	79	887
LL	7	86	267	113	83	556
W	6	31	62	97	44	240
C	1	11	33	29	28	102
Totals	91	547	767	341	254	2000

TABLE LXVIII.

Left Thumb and Ring Finger.

Left Thumb.

Left Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	13	38	7	5	3	66
SL	45	264	198	38	38	583
LL	21	152	342	110	87	712
W	5	60	172	158	96	491
C	7	33	48	30	30	148
Totals	91	547	767	341	254	2000

TABLE LXIX.

Left Thumb and Little Finger.

Left Thumb.

Left Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	13	18	2	1	1	35
SL	62	380	377	81	59	959
LL	12	123	314	180	139	768
W	3	13	44	53	37	150
C	1	13	30	26	18	88
Totals	91	547	767	341	254	2000

TABLE LXX.

Left Index and Middle Finger.

Left Index.

Left Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	117	84	9	2	3	215
SL	166	525	80	79	37	887
LL	19	187	157	139	54	556
W	6	25	21	171	17	240
C	5	12	15	46	24	102
Totals	313	833	282	437	135	2000

TABLE LXXI.

Left Index and Ring Finger.

Left Index.

Left Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	50	14	1	0	1	66
SL	154	355	31	22	21	583
LL	72	326	141	130	43	712
W	19	93	85	253	41	491
C	18	45	24	32	29	148
Totals	313	833	282	437	135	2000

TABLE LXXII.

Left Index and Little Finger.

Left Index.

Left Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	25	8	1	1	0	35
SL	215	526	87	88	43	959
LL	64	253	164	215	72	768
W	6	26	19	85	14	150
C	3	20	11	48	6	88
Totals	313	833	282	437	135	2000

TABLE LXXIII.

Left Middle and Ring Fingers.

Left Middle Finger.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Ring Finger.	*A*	52	13	1	0	0	66
	SL	119	421	35	2	6	583
	LL	35	311	308	38	20	712
	W	7	88	160	176	60	491
	C	2	54	52	24	16	148
	Totals	215	887	556	240	102	2000

TABLE LXXIV.

Left Middle and Little Fingers.

Left Middle Finger.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Little Finger.	*A*	23	11	0	1	0	35
	SL	149	578	174	43	15	959
	LL	39	256	306	110	57	768
	W	2	25	46	58	19	150
	C	2	17	30	28	11	88
	Totals	215	887	556	240	102	2000

TABLE LXXV.

Left Ring and Little Fingers.

Left Ring Finger.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Little Finger.	*A*	17	13	4	0	1	35
	SL	48	474	294	99	44	959
	LL	1	92	390	212	73	768
	W	0	3	12	120	15	150
	C	0	1	12	60	15	88
	Totals	66	583	712	491	148	2000

TABLE LXXVI.

Right Thumb and Left Thumb.

Right Thumb.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Thumb.	*A*	31	44	10	2	4	91
	SL	13	204	246	53	31	547
	LL	0	30	468	180	89	767
	W	1	3	44	270	23	341
	C	1	9	47	144	53	254
	Totals	46	290	815	649	200	2000

TABLE LXXVII.

Right Thumb and Left Index.

Right Thumb.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Index.	*A*	27	94	123	41	28	313
	SL	16	150	420	181	66	833
	LL	1	17	115	105	44	282
	W	1	17	115	264	40	437
	C	1	12	42	58	22	135
	Totals	46	290	815	649	200	2000

TABLE LXXVIII.

Right Thumb and Left Middle Finger.

Right Thumb.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Middle Finger.	*A*	23	60	86	32	14	215
	SL	17	167	414	210	79	887
	LL	0	40	232	214	70	556
	W	5	18	57	141	19	240
	C	1	5	26	52	18	102
	Totals	46	290	815	649	200	2000

TABLE LXXIX.

Right Thumb and Left Ring Finger.

Right Thumb.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Ring Finger.	*A*	8	32	19	6	1	66
	SL	23	145	276	98	41	583
	LL	7	69	319	229	88	712
	W	3	29	142	261	56	491
	C	5	15	59	55	14	148
	Totals	46	290	815	649	200	2000

TABLE LXXX.

Right Thumb and Left Little Finger.

Right Thumb.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Little Finger.	*A*	10	17	6	2	0	35
	SL	29	207	468	181	74	959
	LL	2	54	275	341	96	768
	W	3	8	40	80	19	150
	C	2	4	26	45	11	88
	Totals	46	290	815	649	200	2000

TABLE LXXXI.

Right Index and Left Thumb.

Right Index.

Left Thumb.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	47	29	10	2	3	91
SL	154	241	61	65	26	547
LL	101	266	158	166	76	767
W	21	58	69	151	42	341
C	29	60	41	97	27	254
Totals	352	654	339	481	174	2000

TABLE LXXXII.

Right Index and Left Index.

Right Index.

Left Index.	*A*	SL_r	SL_u	LL_r	LL_u	*W*	*C*	Totals
A	185	42	58	4	10	9	5	313
SL_r	65	103	75	23	9	11	9	295
SL_u	83	88	187	29	46	59	46	538
LL_r	4	11	4	29	2	25	13	88
LL_u	7	14	23	16	64	47	23	194
W	2	11	13	49	24	287	51	437
C	6	13	12	24	10	43	27	135
Totals	352	282	372	174	165	481	174	2000

TABLE LXXXIII.

Right Index and Left Middle Finger.

Right Index.

Left Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	108	87	9	7	4	215
SL	198	425	109	103	52	887
LL	36	106	168	170	76	556
W	4	21	34	154	27	240
C	6	15	19	47	15	102
Totals	352	654	339	481	174	2000

TABLE LXXXIV.

Right Index and Left Ring Finger.

Right Index.

Left Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	49	14	3	0	0	66
SL	167	315	40	40	21	583
LL	102	218	176	144	72	712
W	20	69	83	262	57	491
C	14	38	37	35	24	148
Totals	352	654	339	481	174	2000

TABLE LXXXV.

Right Index and Left Little Finger.

Right Index.

Left Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	24	10	0	0	1	35
SL	236	435	116	122	50	959
LL	81	183	176	224	104	768
W	5	21	30	86	8	150
C	6	5	17	49	11	88
Totals	352	654	339	481	174	2000

TABLE LXXXVI.

Right Middle Finger and Left Thumb.

Right Middle Finger.

Left Thumb.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	36	43	6	4	2	91
SL	97	311	93	32	14	547
LL	47	374	235	90	21	767
W	11	105	111	95	19	341
C	21	88	71	53	21	254
Totals	212	921	516	274	77	2000

TABLE LXXXVII.

Right Middle Finger and Left Index.

Right Middle Finger.

Left Index.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	110	176	21	1	5	313
SL	86	529	177	24	17	833
LL	7	95	125	42	13	282
W	5	80	144	180	28	437
C	4	41	49	27	14	135
Totals	212	921	516	274	77	2000

TABLE LXXXVIII.

Right Middle and Left Middle Fingers.

Right Middle Finger.

Left Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	115	94	4	0	2	215
SL	84	635	129	25	14	887
LL	8	152	295	70	31	556
W	3	20	65	140	12	240
C	2	20	23	39	18	102
Totals	212	921	516	274	77	2000

TABLE LXXXIX.

Right Middle and Left Ring Fingers.

Right Middle Finger.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Ring Finger.	*A*	49	16	1	0	0	66
	SL	110	398	59	8	8	583
	LL	39	346	241	63	23	712
	W	9	102	165	184	31	491
	C	5	59	50	19	15	148
	Totals	212	921	516	274	77	2000

TABLE XC.

Right Middle and Left Little Fingers.

Right Middle Finger.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Little Finger.	*A*	22	13	0	0	0	35
	SL	148	571	168	53	19	959
	LL	35	289	260	137	47	768
	W	4	30	57	56	3	150
	C	3	18	31	28	8	88
	Totals	212	921	516	274	77	2000

TABLE XCI.

Right Ring Finger and Left Thumb.

Right Ring Finger.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Thumb.	*A*	15	38	13	17	8	91
	SL	34	231	112	121	49	547
	LL	7	167	258	262	73	767
	W	3	28	88	199	23	341
	C	4	25	72	130	23	254
	Totals	63	489	543	729	176	2000

TABLE XCII.

Right Ring Finger and Left Index.

Right Ring Finger.

		A	*SL*	*LL*	*W*	*C*	Totals
Left Index.	*A*	41	141	71	39	21	313
	SL	20	309	244	183	77	833
	LL	1	13	112	129	27	282
	W	0	15	79	307	36	437
	C	1	11	37	71	15	135
	Totals	63	489	543	729	176	2000

TABLE XCIII.

Right Ring and Left Middle Fingers.

Right Ring Finger.

Left Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	47	122	28	13	5	215
SL	15	336	274	179	83	887
LL	1	24	204	266	61	556
W	0	5	26	194	15	240
C	0	2	11	77	12	102
Totals	63	489	543	729	176	2000

TABLE XCIV.

Right Ring and Left Ring Fingers.

Right Ring Finger.

Left Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	37	26	2	0	1	66
SL	20	363	114	45	41	583
LL	4	81	367	183	77	712
W	1	9	30	423	28	491
C	1	10	30	78	29	148
Totals	63	489	543	729	176	2000

TABLE XCV.

Right Ring and Left Little Fingers.

Right Ring Finger.

Left Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	17	15	3	0	0	35
SL	42	399	250	194	74	959
LL	4	68	276	339	81	768
W	0	5	9	125	11	150
C	0	2	5	71	10	88
Totals	63	489	543	729	176	2000

TABLE XCVI.

Right Little Finger and Left Thumb.

Right Little Finger.

Left Thumb.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	8	60	13	5	5	91
SL	16	345	144	24	18	547
LL	2	320	330	81	34	767
W	3	77	165	71	25	341
C	2	68	112	47	25	254
Totals	31	870	764	228	107	2000

TABLE XCVII.

Right Little Finger and Left Index.

Right Little Finger.

Left Index.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	22	204	76	6	5	313
SL	8	477	269	44	35	833
LL	0	76	155	34	17	282
W	1	73	198	128	37	437
C	0	40	66	16	13	135
Totals	31	870	764	228	107	2000

TABLE XCVIII.

Right Little and Left Middle Fingers.

Right Little Finger.

Left Middle Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	17	158	35	2	3	215
SL	12	507	291	42	35	887
LL	0	152	301	71	32	556
W	2	39	89	84	26	240
C	0	14	48	29	11	102
Totals	31	870	764	228	107	2000

TABLE XCIX.

Right Little and Left Ring Fingers.

Right Little Finger.

Left Ring Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	12	50		0	1	66
SL	15	440	10	9	13	583
LL	2	255	398	30	27	712
W	1	89	178	173	50	491
C	1	36	79	16	16	148
Totals	31	870	764	228	107	2000

TABLE C.

Right Little and Left Little Fingers.

Right Little Finger.

Left Little Finger.	*A*	*SL*	*LL*	*W*	*C*	Totals
A	18	17	0	0	0	35
SL	11	721	176	21	30	959
LL	1	115	543	74	35	768
W	1	13	22	99	15	150
C	0	4	23	34	27	88
Totals	31	870	764	228	107	2000

ON THE PROBLEM OF SEXING OSTEOMETRIC MATERIAL

By KARL PEARSON, F.R.S.

It is well known that anthropometric, particularly craniometric measurements give frequency series, which for moderate sized populations follow closely the normal or Laplace-Gaussian distribution. Measurements of stature, cubit head-length, cephalic index, etc., etc., obey with sufficient accuracy for most purposes of science the normal law. This statement may with a high degree of certainty be extended to practically almost all measurements on the adult skeleton. But a new difficulty arises in dealing with the parts of the skeleton: the sexing of the several bones of the human body is by no means certain, and this is especially the case when we come to deal—not with the cranium or the pelvis but with the long bones. In order to get over this difficulty, and to find the constants for each sex, it occurred to me some years back when the sexing of the long bones had presented this problem very forcibly to workers in my laboratory, that the method of my first contribution to the mathematical theory of evolution* might be applied. Namely, we might take the unsexed material and assume it to consist of a compound of male and female data, the frequency curve for each of these being normal; the two components might then be found in the manner of the paper just referred to. The method was especially likely to be successful, when the series was otherwise homogeneous, the numbers large and the character dealt with substantially differentiated sexually. Of course the method does not give the sex of each individual bone, but I have shown in another memoir†, how four to six characters thus resolved form a basis for determining the probable sex of each bone, and this with an accuracy which is very probably as great as, or even greater than, anatomical appreciation unbased on a system of numerical measurement.

One of the few objections to the method is the labour involved in the process. While the analysis required in the application of the method is not so severe that it has not been applied in a large number of cases by workers in the Biometric

---

* *Phil. Trans.* Vol. 185, A, pp. 71—110.

† To appear in the next number of this Journal.

Laboratory, it is still considerably beyond the powers of most of the present workers in anthropometry, and probably no anatomist of the present day has the mathematical knowledge requisite for the solution of the reducing nonic, or the arithmetical patience required for the calculation of its coefficients. It has occurred to me, however, that the work might be considerably shortened by the following considerations. The bones usually dealt with are those found in ancient cemeteries, in plague pits, clearance pits or crypts. It is probable, though by no means certain, that *adult* female bones in such cases would be rather more numerous than male. On the other hand being somewhat smaller they are asserted by some writers as likely to be more frequently broken, and they certainly may more readily escape preservation or measurement. If we take these two causes as counteracting each other, we may assume as a *first* approximation that the numbers of male and female bones will be equal. In the next place it is a result of much anthropometric experience that male and female variations, i.e. their standard deviations, are closely alike. These again we can take equal to a *first* approximation. Accordingly, to this first approximation, our osteometric series may be considered to consist of two equal normal components with different means. Let the mean of the unsexed material be M, and let the actual means of the sexed components be m_1, m_2, their standard deviations be σ_1, σ_2, and their total frequencies n_1 and n_2, where the subscript 1 refers, say, to the males, and 2 to the females. Then m_1, m_2, σ_1, σ_2, n_1 and n_2 are the quantities we desire to discover. Let the moment-coefficients of the total material be, in the usual notation, μ_2, μ_3, μ_4, μ_5 and let $N\,(=n_1+n_2)$ be the total unsexed population. We shall write as customary $\beta_1=\mu_3^2/\mu_2^3$, $\beta_2=\mu_4/\mu_2^2$, $\beta_3=\mu_3\mu_5/\mu_2^4$. Then, if our hypothesis be correct and the material consist very nearly of two equal normal distributions, β_1 and β_3 ought to be very small, while β_2 will be large in relation to them.

It is convenient also to write:

$$\zeta_1=\tfrac{1}{2}(3-\beta_2),\qquad \zeta_2=10\beta_1-\beta_3\quad\text{(i)},$$

$$m_1=M+\gamma_1,\qquad m_2=M+\gamma_2\quad\text{(ii)},$$

$$q_2=\gamma_1\gamma_2/\mu_2,\qquad q_1=(\gamma_1+\gamma_2)/\sqrt{\mu_2}\quad\text{(iii)},$$

$$q_3=q_1q_2\quad\text{(iv)}.$$

Then the fundamental nonic may be written:

$$q_2^9-7\zeta_1q_2^7+\tfrac{3}{2}\beta_1q_2^6-3(\zeta_2-5\zeta_1^2)q_2^5-\left(37\beta_1\zeta_1+\frac{3}{4}\frac{\zeta_2^2}{\beta_1}\right)q_2^4$$
$$+3(4\beta_1^2-3\zeta_1\zeta_2-3\zeta_1^3)q_2^3+3(\beta_1\zeta_2-\tfrac{7}{2}\beta_1\zeta_1^2)q_2^2+8\beta_1^2\zeta_1q_2-\beta_1^3=0\ldots\text{(v)}.$$

Further:
$$q_3=\frac{\sqrt{\beta_1}\left\{\beta_1-6\zeta_1q_2-\dfrac{3}{2}\dfrac{\zeta_2}{\beta_1}q_2^2-4q_2^3\right\}}{2\beta_1-3\zeta_1q_2+q_2^3}\quad\text{(vi)},$$

where the sign of $\sqrt{\beta_1}$ is determined by that of μ_3.

Again
$$n_1=\frac{-\gamma_2}{\gamma_1-\gamma_2}N,\qquad n_2=\frac{\gamma_1}{\gamma_1-\gamma_2}N\quad\text{(vii)}.$$

Lastly:
$$\left.\begin{aligned}\sigma_1^2 &= \mu_2(1+q_2) - \tfrac{1}{3}\mu_3/\gamma_2 - \tfrac{1}{3}\sqrt{\mu_2}\gamma_1\\ \sigma_2^2 &= \mu_2(1+q_2) - \tfrac{1}{3}\mu_3/\gamma_1 - \tfrac{1}{3}\sqrt{\mu_2}\gamma_2\end{aligned}\right\}\ldots\ldots\ldots\ldots\text{(viii)}.$$

Equations (ii), (iii), (iv), (v), (vi), (vii) and (viii) form the complete solution of the problem when we make no approximations whatever*.

If, however, $\beta_1 = \beta_3 = 0$, then, the two components being equal, we have†:

$$\left.\begin{aligned}n_1 &= \quad n_2 = \tfrac{1}{2}N\\ \gamma_1 &= -\gamma_2 = \sqrt{\mu_2}\,\zeta_1^{\frac{1}{4}}\\ \sigma_1 &= \quad \sigma_2 = \sqrt{\mu_2}\,\{1-\sqrt{\zeta_1}\}^{\frac{1}{2}}\end{aligned}\right\}\ldots\ldots\ldots\ldots\text{(ix)}.$$

It will be seen that it is needful in order that the solution may be real that ζ_1 should be positive or $\beta_2 < 3$, i.e. the total frequency should be platykurtic. Now let us suppose that the values given by (ix) are a first approximation and that we need a second approximation in which the two normal curves will be unequal in frequency, mean and standard deviation. Write:

$$n = \tfrac{1}{2}N,\quad \gamma = \sqrt{\mu_2}\,\zeta_1^{\frac{1}{4}},\quad \sigma = \sqrt{\mu_2}\,(1-\sqrt{\zeta_1})^{\frac{1}{2}}\ldots\ldots\ldots\ldots\text{(ix)}^{\text{bis}},$$

and suppose:

$$\begin{aligned}n_1 &= n + \delta n_1, & n_2 &= \quad n + \delta n_2,\\ \gamma_1 &= \gamma + \delta\gamma_1, & \gamma_2 &= -\gamma + \delta\gamma_2,\\ \sigma_1 &= \sigma + \delta\sigma_1, & \sigma_2 &= \quad \sigma + \delta\sigma_2,\end{aligned}$$

where the differentials represent small quantities of which the squares and products may be neglected to a second approximation.

Our equations are‡:

$$\begin{aligned}&n_1 + n_2 = N,\\ &n_1\gamma_1 + n_2\gamma_2 = 0,\\ &n_1(\gamma_1^2+\sigma_1^2) + n_2(\gamma_2^2+\sigma_2^2) = N\mu_2,\\ &n_1(\gamma_1^3 + 3\gamma_1\sigma_1^2) + n_2(\gamma_2^3 + 3\gamma_2\sigma_2^2) = N\mu_3,\\ &n_1(\gamma_1^4 + 6\gamma_1^2\sigma_1^2 + 3\sigma_1^4) + n_2(\gamma_2^4 + 6\gamma_2^2\sigma_2^2 + 3\sigma_2^4) = N\mu_4,\\ &n_1(\gamma_1^5 + 10\gamma_1^3\sigma_1^2 + 15\gamma_1\sigma_1^4) + n_2(\gamma_2^5 + 10\gamma_2^3\sigma_2^2 + 15\gamma_2\sigma_2^4) = N\mu_5.\end{aligned}$$

We now differentiate these and after differentiation put

$$n_1 = n_2 = n,\qquad \gamma_1 = -\gamma_2 = \gamma,\qquad \sigma_1 = \sigma_2 = \sigma.$$

Hence we find:

$$\delta n_1 = -\delta n_2 \ldots\ldots\ldots\ldots\text{(x)},$$

$$n(\delta\gamma_1 + \delta\gamma_2) + 2\gamma\delta n_1 = 0 \ldots\ldots\ldots\ldots\text{(xi)},$$

$$2n\gamma(\delta\gamma_1 - \delta\gamma_2) + 2n\sigma(\delta\sigma_1 + \delta\sigma_2) = 0 \ldots\ldots\ldots\ldots\text{(xii)},$$

$$3n(\delta\gamma_1 + \delta\gamma_2)(\gamma^2+\sigma^2) + 2\delta n_1\gamma(\gamma^2 + 3\sigma^2) + 6n\sigma\gamma(\delta\sigma_1 - \delta\sigma_2) = N\mu_3 \ldots\ldots\text{(xiii)},$$

$$\gamma(\gamma^2 + 3\sigma^2)(\delta\gamma_1 - \delta\gamma_2) + 3\sigma(\gamma^2+\sigma^2)(\delta\sigma_1 + \delta\sigma_2) = 0 \ldots\ldots\ldots\ldots\text{(xiv)},$$

$$\begin{aligned}n(\delta\gamma_1 + \delta\gamma_2)(5\gamma^4 + 30\gamma^2\sigma^2 + 15\sigma^4) + 2\delta n_1\gamma(\gamma^4 + 10\gamma^2\sigma^2 + 15\sigma^4)&\\ + n(\delta\sigma_1 - \delta\sigma_2)\,20\gamma\sigma(\gamma^2 + 3\sigma^2) = N\mu_5 &\ldots\ldots\ldots\ldots\text{(xv)},\end{aligned}$$

---

* They are, in a somewhat better form, those originally given by me in *Phil. Trans.* Vol. 185, A, 1894, pp. 71—110; see Equations (14), (15), (18), (19), (27) and (29) of that memoir.

† *Loc. cit.* footnote, p. 91.

‡ *Loc. cit.* p. 82.

where it must be remembered that the differential terms are introduced solely to account for the asymmetry as represented by μ_3 and μ_5, assumed to be zero to a first approximation.

But (xii) and (xiv) show us that we must have:

$$\delta\gamma_1 = \delta\gamma_2, \qquad \delta\sigma_1 = -\delta\sigma_2.$$

Hence from (xi):

$$\delta n_1 = -n\delta\gamma_1/\gamma \quad \text{(xvi)}.$$

(xiii) now becomes:

$$2\gamma^2\delta\gamma_1 + 6\sigma\gamma\delta\sigma_1 = \mu_3,$$

and (xv):

$$4\gamma^2\delta\gamma_1(\gamma^2 + 5\sigma^2) + 20\gamma\sigma(\gamma^2 + 3\sigma^2)\,\delta\sigma_1 = \mu_5.$$

Whence solving we find:

$$\delta\gamma_1 = \delta\gamma_2 = \tfrac{5}{4}\left(1 + 3\frac{\sigma^2}{\gamma^2}\right)\frac{\mu_3}{\gamma^2} - \frac{3}{8}\frac{\mu_5}{\gamma^4} \quad \text{(xvii)},$$

$$\delta\sigma_1 = -\delta\sigma_2 = -\tfrac{1}{4}\left(1 + 5\frac{\sigma^2}{\gamma^2}\right)\frac{\mu_3}{\sigma\gamma} + \frac{1}{8}\frac{\mu_5}{\sigma\gamma^3} \quad \text{(xviii)},$$

$$\delta n_1 = -\delta n_2 = -\frac{n\delta\gamma_1}{\gamma} \quad \text{(xix)}.$$

These form together with $(\text{ix})^{\text{bis}}$ the complete solution of the problem.

The following example illustrates the procedure: 541 measurements were made of the bicondylar width of English femora, right and left, male and female being mixed. The frequency below resulted.

Frequency Distribution of 541 *Femora for Bicondylar Width.*

mm.	Frequency	mm.	Frequency	mm.	Frequency
61	1	71	23	81	28
62	1	72	33·5	82	23
63	1·5	73	25	83	19
64	5	74	22	84	17·5
65	13·5	75	36	85	19·5
66	14	76	25·5	86	16·5
67	15·5	77	29·5	87	7·5
68	22	78	32·5	88	3
69	31	79	19·5	89	3·5
70	19	80	33	90	0·5

The constants of this distribution were:

$$M = 75{\cdot}8152,$$
$$\mu_2 = 37{\cdot}692{,}112, \qquad \mu_3 = -\ 2{\cdot}587{,}693,$$
$$\mu_4 = 3020{\cdot}893{,}695, \qquad \mu_5 = -\ 83{\cdot}260{,}992.$$

Hence we deduce:

$$\beta_1 = {\cdot}000{,}125{,}047, \qquad \beta_3 = {\cdot}000{,}106{,}750,$$
$$\beta_2 = 2{\cdot}126{,}349, \qquad \zeta_1 = {\cdot}436{,}8255,$$
$$\zeta_2 = {\cdot}001{,}143{,}72.$$

Clearly β_1 and β_3 are so small that the distribution fulfils our condition of being very closely symmetrical. The nonic, equation (v) above, is:

$$\begin{aligned} q_2^9 - 3{\cdot}057{,}789 q_2^7 + {\cdot}000{,}18757 q_2^6 + 2{\cdot}858{,}817 q_2^5 \\ - {\cdot}009{,}867 q_2^4 - {\cdot}754{,}678 q_2^3 - {\cdot}002{,}5011 q_2^2 \\ + {\cdot}000{,}000{,}0546 q_2 - {\cdot}000{,}000{,}000{,}005 = 0, \end{aligned}$$

the last two terms being written down to many figures to show their inappreciableness. The root required is:

$$q_2 = -{\cdot}65679,$$

which by (vi) leads to:

$$\gamma^2 + {\cdot}558{,}050\gamma - 24{\cdot}755{,}802 = 0,$$

and provides the solution:

	Females.	*Males.*
Mean:	70·547 mm.	80·526 mm.
Total Frequency:	255·4	285·6
Standard Deviation:	3·4842 mm.	3·6944 mm.
Modal Ordinate*:	29·24	30·84

............(A).

We have now to inquire how far the same result would be reached, if we had supposed as a first approximation equal Gaussian components and then proceeded to determine a second approximation by aid of (xvii) to (xix).

Equations (ix) give us:

$$\begin{aligned} n_1 &= n_2 = n = 270{\cdot}5, \\ \gamma_1 &= -\gamma_2 = \gamma = 4{\cdot}9912, \\ \sigma_1 &= \sigma_2 = 3{\cdot}5750. \end{aligned}$$

Thus to a first approximation:

	Females.	*Males.*
Mean:	70·824 mm.	80·806 mm.
Total Frequency:	270·5	270·5
Standard Deviation:	3·5750 mm.	3·5750 mm.
Modal Ordinate:	30·19	30·19

............(B).

(B), statistically speaking, is so close to (A) that it gives every confidence of a second approximation practically reproducing (A).

We find:

$$\frac{\mu_3}{\gamma^3} = -{\cdot}020{,}8112, \qquad \frac{\mu_5}{\gamma^5} = -{\cdot}026{,}0386,$$

$$\frac{\sigma^2}{\gamma^2} = {\cdot}513{,}2871, \qquad \frac{\gamma^2}{\sigma^2} = 1{\cdot}948{,}228.$$

---

* $y_0 = \frac{n}{\sqrt{2\pi}\sigma}$ of the normal curve.

Hence by (xvii) to (xix):

$$\begin{aligned}\delta\gamma_1 &= \delta\gamma_2 = -\gamma \times \cdot 056{,}308 = -\cdot 2810,\\ \delta\sigma_1 &= -\delta\sigma_2 = \sigma \times \cdot 029{,}809 = \cdot 1066,\\ \delta n_1 &= -\delta n_2 = +n \times \cdot 056{,}308 = 15\cdot 231.\end{aligned}$$

It will be seen from these results that:

$$\frac{\delta\gamma_1}{\gamma} = -\cdot 0563, \qquad \frac{\delta\sigma_1}{\sigma} = \cdot 0298, \qquad \frac{\delta n_1}{n} = \cdot 0563$$

may be considered fairly small quantities, and that they justify our assumption. We have accordingly:

	Females.	*Males.*	
Mean:	70·543 mm.	80·525 mm.	............(C).
Total Frequency:	255·27	285·73	
Standard Deviation:	3·4684 mm.	3·6816 mm.	
Modal Ordinate:	29·36	30·96	

It is clear that the solutions (C) and (A) are for all practical purposes identical. Thus the short method is justified in the problem of sexing osteometric material. An improper extension of the method to material in which the sexes occur in very unequal groups may be guarded against by simply observing whether β_1 and β_3 are very small quantities.

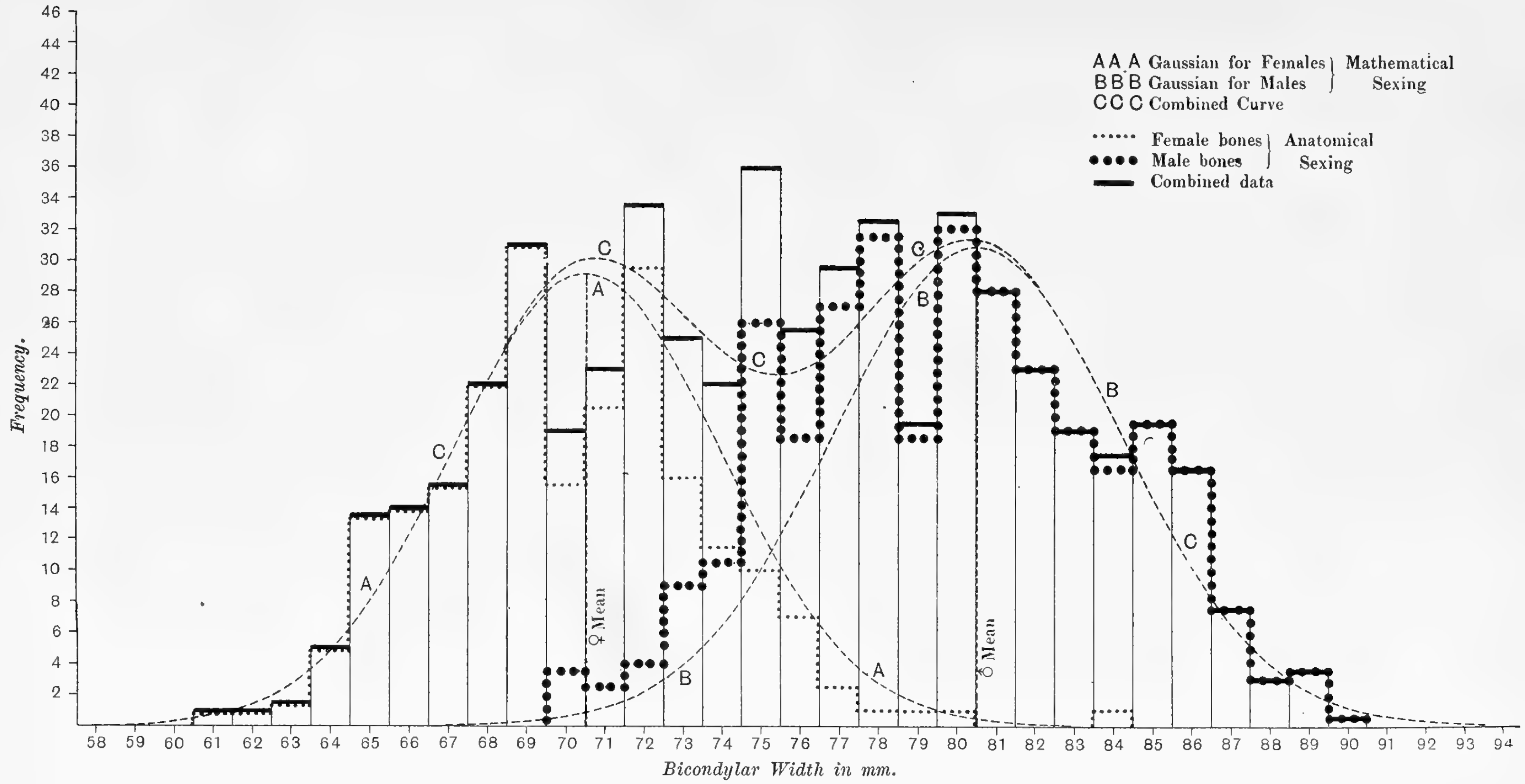
In conclusion it may be desirable to compare the values of these sex-constants as found mathematically with sexing by anatomical appreciation. I owe an anatomical sexing of the same bones to my colleague, Dr Derry.

The following values of the constants resulted:

	Females.	*Males.*	
Mean:	70·098 mm.	79·764 mm.	............(D).
Total Frequency:	221	320	
Standard Deviation:	3·5148 mm.	4·1254 mm.	
Modal Ordinate:	24·55	30·95	

It will be seen that the mathematically deduced constants are not widely divergent from those obtained anatomically, but the accordance if fair is not ideal. The accompanying diagram exhibits the differences in the frequency distributions found by the two methods of sexing. The chief difference lies in the transfer by the anatomist of the larger female bones of the mathematical sexing to the male group. I do not propose to discuss here the relative advantages of the two methods, but would draw attention to a few points of interest:

(i) The solution (D) makes no appeal to measurement in the sexing, it is based purely on an anatomical appreciation. It would therefore be subject to



London 17th century Femora. Bicondylar Width. Comparison of Mathematical and Anatomical Sexing.

personal equation, depending on the features upon which the experience of the individual anatomist leads him to lay most stress. The solution (C) is unique, that is to say, given the same data, all statisticians would reach the same values, of course apart from errors in arithmetic or from the number of decimal places retained in the working. It eliminates the factor of personal equation.

Frequency Distributions of Bicondylar Width in Male and Female Femora sexed by Anatomical Appreciation.

mm.	♀	♂	mm.	♀	♂	mm.	♀	♂
61	1	—	71	20·5	2·5	81	—	28
62	1	—	72	29·5	4	82	—	23
63	1·5	—	73	16	9	83	—	19
64	5	—	74	11·5	10·5	84	1	16·5
65	13·5	—	75	10	26	85	—	19·5
66	14	—	76	7	18·5	86	—	16·5
67	15·5	—	77	2·5	27	87	—	7·5
68	22	—	78	1	31·5	88	—	3
69	31	—	79	1	18·5	89	—	3·5
70	15·5	3·5	80	1	32	90	—	0·5

(ii) (C) would, however, be influenced by the fact that our material is not perfectly homogeneous except for sex; because (*a*) there is a mixture of right and left bones, and, to judge by the anatomical sexing, this may involve a difference of ·7 to ·9 mm. in the means and ·08 to ·24 mm. in the standard deviations; this would add to the heterogeneity, (*b*) our bones may be due to somewhat mixed classes and possibly mixed periods, (*c*) the bicondylar width is liable to be injured by rough treatment of the bone, and this injury will most affect the weaker, and therefore probably the younger, bones. These bones might then be treated as female, a classification which most anatomical sexing also favours. While the total number of these London femora is nearly 800, the bicondylar width could only be measured in 541 cases. This selection will not necessarily be random as to size or sex, and may modify our constants found mathematically from the distribution. On the other hand it would affect also the anatomical appreciation of sex, but only in as far as it was based on the size of the condyles.

(iii) We know from very considerable sexed data that the variation of man and woman is very nearly the same. The coefficients of variation measured in the usual way, i.e. by 100 standard deviation divided by mean, gave:

Mathematical Sexing.	*Anatomical Sexing.*
♀ 4·92 ♂ 4·57	♀ 5·01 ♂ 5·17
$\Delta = \cdot 35$	$\Delta = - \cdot 16$

There was thus closer sexual accord from the anatomical method. But when the same anatomical sexing was applied to the character of the head of the femur in the vertical plane, I found for right bones:

♀ 5·05 ♂ 6·37 $\Delta = -1·32$,

and for left bones:

♀ 4·91 ♂ 6·10 $\Delta = -1·19$,

differences far greater than occur in the mathematical sexing from the bicondylar widths. Accordingly no great stress can be laid on inequalities in the coefficients of variation deduced from either process of sexing.

It would appear to me that we have reached on the whole a reasonable biometric method of sexing. To what extent it can replace the sexing by anatomical appreciation must be left to the future. But it is clear that when anatomists themselves prefer to that appreciation an appeal to a single character, e.g. to the measurement of the femoral head, and only settle by anatomical appreciation the sex of femora with diameters between 45 and 47 mm., then they do not show much confidence in their own method of sexing. An interesting experiment could be made if some 400 to 500 sexed bones were available, and then, without knowledge of the real sex, two or three anatomists and a statistician were to be asked independently to determine the mean and variability of two or three characters of the bones of each sex in this material.

I have cordially to acknowledge the help of my colleague Mr E. Soper in the determination of equations (xvii)—(xix) and in their solution (C) in the numerical case for which I had reached the solution (A); also the labour of my colleague Miss H. Gertrude Jones in the preparation of the diagram which contrasts graphically the mathematical and anatomical solutions of the problem.

FURTHER EVIDENCE OF NATURAL SELECTION IN MAN.

By ETHEL M. ELDERTON, Galton Research Fellow,
AND KARL PEARSON, F.R.S.

(1) The second author of the present paper writing in 1894 a commentary on the statement that "no man, as far as we know, has ever seen natural selection at work," remarked: "Every man who has lived through a hard winter, every man who has examined a mortality table, every man who has studied the history of nations has probably seen natural selection at work*." The emphasis is here to be laid on the word "probably," because the seeing depends on the power and validity of the scientific means adopted to analyse the observed facts. In a paper communicated by the same author to the Royal Society in June 1912†, it was shown from the Registrar-General's series of ten yearly life-tables that when allowance was made for change of environment in the course of the fifty years a very high association existed between the deaths in the first year of life and the deaths in childhood (1 to 5 years). This association was such that if the infantile deathrate *increased* by 10°/₀ the child deathrate *decreased* by 5·3°/₀ in males, while in females the *fall* in the child deathrate was almost 1°/₀ for every *rise* of 1°/₀ in the infantile deathrate. The method of investigating by life-tables could not be extended beyond 1900, because the life-tables for the next ten years (1901–1910) were not then out, and indeed have only just appeared (December 1914). While the infantile deathrate as shown from the life-tables had risen from 1871–1900, the child deathrate had fallen for the same period. During the next decade 1900–1910 both deathrates have fallen together; such a secular change does not in any way modify the argument of the paper, which lies in the statement that whether two deathrates rise together or rise and fall simultaneously we can draw no inferences at all, *until they have been corrected for secular change.* Most economic, demographic and physical variates are changing continuously with time, and no comparison of time graphs or calculation of correlations will demonstrate of necessity anything but spurious association, until

---

* *The Chances of Death and other Studies in Evolution*, Vol. I. p. 166.

† "The Intensity of Natural Selection in Man." *R. S. Proc.* B. Vol. 85, pp. 469—476.

the time factor has been eliminated. It is the deviations from the continuous curves of secular change which may turn out on careful analysis to be truly indicative of causal relationship between the variates under consideration.

The first attempt to get rid of secular change by a method of differences was made by Miss F. E. Cave in 1904 in a paper on barometric correlations*, and shortly afterwards Mr R. H. Hooker published a paper dealing with the same point†. Both these authors used only first differences and gave no general theory of the method. Quite recently "Student" has published a paper‡ giving the fundamental formulae, and indicating how by taking successive differences of two variates and correlating them, we free ourselves from the time or locality influence, and approach the true and probably causal relationship between them. When the correlation of the differences becomes steady, then we have reached the actual correlation of the variates corrected for the time factor, *provided an assumption is made which we shall discuss at greater length below:* see footnote, p. 495. Meanwhile Dr Anderson of Petrograd has been working on the subject, and in a most valuable memoir§ he has added to "Student's" results a number of new theorems; for example, the probable errors of the successive difference correlations when they become steady, and the relations which should be fulfilled between the squares of the standard deviations of successive differences, when the series has become steady. We have thus a double means of ascertaining whether the desired object—the elimination of the time-factor—has been approximately achieved. A third additional test will be indicated in this paper.

This new statistical process has been termed the *Variate Difference Correlation Method*||, and there is small doubt that it is the most important contribution to the apparatus of statistical research which has been made for a number of years past. Its field of application to physical problems alone seems inexhaustible. We are no longer limited to the method of partial correlation, nor compelled to seek for factors which rendered constant will remove the changing influence of environment. In the present case, that of the influence of infantile mortality on child mortality, Pearson endeavoured to eliminate the influence of continual environmental improvement by making the expectation of life at six years constant¶. Snow achieved the same object by correlating the deathrates of one sex for a constant deathrate of the other**. In both these cases substantial evidence of Natural Selection was obtained from the mortality tables. The object of the present paper is to demonstrate by the still more complete elimination of the

---

* *R. S. Proc.* Vol. LXXIV. pp. 407 *et seq.*

† *Royal Statistical Society Journal*, Vol. LXVIII. pp. 396 *et seq.* 1905.

‡ *Biometrika*, Vol. x. pp. 179, 180.

§ *Ibid.* pp. 269—279.

|| Pearson and Cave: "Numerical Illustrations of the Variate Difference Correlation Method." *Biometrika*, Vol. x. pp. 340—355.

¶ *R. S. Proc.* B. Vol. 85, p. 472.

** "The Intensity of Natural Selection in Man." *Drapers' Company Research Memoirs*, Dulau & Co., 1911.

time factor involved in the variate difference correlation method that a selective deathrate plays even in highly civilised states a marked part in the natural history of man.

(2) The material dealt with in this investigation consists of the Registrar-General's returns for births in England and Wales and of deaths in the first five years of life from 1859 to 1908 with the addition of as many years before 1859 as were requisite to make our highest differences fifty in number, and with the addition of as many years after 1908 as were requisite for following up the births of that year to the fifth year of life. Thus actually our data extended from 1850 to 1912. The reason for this procedure lies in the desirability of using a constant population, and not reducing by one a relatively small number like 50 on each differencing. As a result of this process we had to modify Dr Anderson's values for the probable errors for the steady values of the difference correlations because in our case the size of the population does not change as we proceed to higher differences*. The second cause which requires extension of the data is a very important one, and must be illustrated numerically. Consider the table:

Deaths of those born in a given year.

Year	Female Births	*0—1*	*1—2*	*2—3*	*3—4*	*4—5*
1908	478,410	63,594	—	—	—	—
1909	—	—	14,146	—	—	—
1910	—	—	—	5,020	—	—
1911	—	—	—	—	3,449	—
1912	—	—	—	—	—	2,341

Now the deaths of infants 0—1 in 1908 are not necessarily of infants all born in 1908, but the total deaths 63,594 must represent closely the deaths in the 478,410 infants born in that year. Disregarding immigration and emigration, this gives a deathrate per 1000 of 107·495 and leaves 414,816 children alive. Of this group 14,146 may be taken to die in the second year of life, giving a deathrate of 31·990 per mille. There remain 400,670 children who reach the third year of life in 1910, of whom 5,020 die, giving a deathrate of 11·939, and 395,650 survivors. These survivors are followed into 1911 and 1912 in the same manner, and thus we obtain approximately the deathrate up to the fifth year of the male children born in 1908. We thus in bulk follow the same group of children through the first five years of life. Tables I and II give the deathrates for males and females respectively under the heading of the birth year of each group. These deathrates have been taken to three decimals places for the purpose of determining the higher differences correctly to one decimal place. The successive differences of

---

* All the probable errors of the *difference* correlations given in this memoir are these modified Andersonian values, i.e. they are the probable errors on the assumption that the difference correlations have reached steady values.

TABLE I. *Deathrates in each Year of Life for groups born in the Year of 1st column.*

Males.

	0—1	*1—2*	*2—3*	*3—4*	*4—5*
1850	159·781	63·935	34·092	22·138	19·021
1	168·706	64·977	33·882	26·657	18·209
2	173·324	63·533	40·936	24·316	16·022
3	174·808	74·802	35·464	22·298	16·720
4	170·890	60·598	30·740	21·688	21·329
5	169·151	59·696	33·004	29·546	19·780
6	156·756	64·317	39·551	25·594	13·751
7	168·486	67·928	36·777	19·471	13·923
8	172·591	68·712	30·566	19·857	16·268
9	167·106	58·887	31·650	22·325	21·962
1860	162·642	70·401	35·297	30·083	21·325
1	167·634	65·785	41·390	27·654	16·236
2	156·684	73·848	37·326	22·013	14·982
3	163·183	67·537	32·774	21·088	12·552
4	166·309	66·462	34·408	17·660	15·991
5	174·356	68·369	28·278	21·473	17·015
6	173·659	60·603	32·159	22·751	18·105
7	166·905	63·790	32·731	23·220	16·085
8	168·064	62·988	32·357	20·185	12·821
9	169·022	65·716	30·186	17·450	11·624
1870	174·287	62·401	26·760	16·344	16·052
1	171·840	59·853	25·503	21·635	14·727
2	162·321	55·267	30·754	18·731	12·655
3	163·676	61·415	28·007	17·275	12·502
4	164·976	60·316	25·232	16·028	13·499
5	173·145	58·422	25·400	18·848	13·187
6	160·415	56·088	27·992	17·044	13·263
7	149·627	63·344	26·104	17·171	11·738
8	166·266	58·104	27·853	15·280	13·049
9	149·754	66·188	22·245	16·741	12·115
1880	167·313	48·181	25·909	16·036	11·920
1	142·532	59·164	23·333	15·261	10·430
2	153·154	54·365	24·026	14·362	9·416
3	151·184	58·292	23·015	13·823	11·042
4	160·381	53·952	22·077	14·792	9·533
5	151·175	57·960	23·052	13·615	10·073
6	163·081	55·611	21·009	13·974	10·569
7	158·243	50·713	22·169	14·394	9·894
8	150·177	56·882	22·958	13·536	10·150
9	157·476	56·279	22·338	13·750	10·647
1890	164·757	59·098	22·604	14·551	9·623
1	163·761	54·255	20·821	12·615	8·757
2	162·112	52·776	19·442	12·497	10·604
3	173·333	47·035	20·343	14·121	8·546
4	149·633	55·787	21·271	11·750	8·439
5	176·280	51·404	19·535	11·912	8·973
6	160·989	50·293	18·742	11·802	9·366
7	170·291	50·986	19·124	12·396	8·608
8	175·183	48·227	19·380	11·496	8·657
9	176·606	49·837	17·094	11·565	7·165
1900	168·685	44·728	17·400	9·759	7·553
1	165·617	42·597	15·570	10·241	7·135
2	146·791	40·581	16·572	9·587	6·853
3	144·567	45·517	15·368	9·260	6·949
4	158·684	38·921	15·234	10·194	6·458
5	141·193	39·326	15·702	8·893	6·886
6	144·819	38·037	14·222	8·858	5·655
7	130·259	36·615	15·159	7·643	6·230
1908	132·928	34·102	12·529	8·717	5·969

TABLE II. *Deathrates in each Year of Life for groups born in the Year of 1st column.*

Females.

	0—1	*1—2*	*2—3*	*3—4*	*4—5*
1850	130·477	62·235	34·147	22·625	19·278
1	138·306	62·106	33·992	27·087	16·762
2	142·185	61·812	39·787	23·805	16·148
3	144·270	71·939	35·186	22·686	16·948
4	141·521	59·024	30·863	22·226	21·907
5	137·278	57·140	34·057	29·375	20·501
6	129·877	61·902	39·758	26·146	14·305
7	142·203	65·853	36·712	19·990	14·391
8	143·595	64·513	29·716	20·796	17·089
9	138·477	55·535	32·024	24·485	21·454
1860	131·914	66·902	36·060	29·941	20·765
1	137·339	61·860	41·243	27·727	16·287
2	127·203	70·313	36·543	21·905	15·398
3	133·321	63·438	32·637	21·623	12·702
4	138·232	63·395	34·846	18·217	15·547
5	145·388	66·401	28·484	21·437	17·145
6	144·879	58·180	32·392	23·086	17·536
7	137·712	61·641	33·243	23·081	15·653
8	141·756	58·624	31·892	20·060	12·524
9	141·450	60·720	31·004	18·108	11·732
1870	144·596	59·845	26·855	16·235	15·196
1	143·353	56·379	25·062	21·493	13·903
2	136·453	52·652	30·213	18·454	12·571
3	134·079	57·739	27·549	16·950	11·421
4	135·939	57·795	25·498	16·031	13·253
5	142·730	54·032	24·563	18·848	12·833
6	131·718	51·057	27·976	16·685	12·651
7	121·936	59·535	25·135	17·441	11·214
8	137·958	52·463	27·646	15·186	12·491
9	120·643	62·324	21·792	16·911	11·880
1880	137·691	45·620	25·319	15·556	11·335
1	117·221	55·413	22·443	15·359	10·396
2	127·627	50·057	23·539	14·178	9·586
3	122·766	54·048	22·948	13·643	10·606
4	132·701	49·967	21·259	14·913	9·401
5	123·667	53·415	22·740	13·251	10·030
6	134·814	51·103	19·985	13·800	10·295
7	130·689	46·402	21·538	14·594	10·012
8	122·312	53·369	22·456	13·926	10·146
9	129·143	53·307	21·783	14·170	10·569
1890	135·839	54·850	21·607	14·731	9·478
1	132·763	51·567	20·841	12·646	8·900
2	132·414	49·387	19·174	12·708	10·370
3	143·346	44·511	19·842	14·293	8·643
4	123·522	52·267	21·499	12·183	8·300
5	144·326	49·339	18·643	11·817	8·499
6	133·535	46·691	18·439	12·184	9·282
7	140·755	47·998	18·109	12·739	8·642
8	145·001	44·832	18·436	11·490	8·865
9	148·000	45·928	16·954	11·724	7·164
1900	139·148	41·901	16·902	9·712	7·611
1	136·346	39·527	15·156	10·606	7·184
2	118·479	37·064	16·168	9·891	6·792
3	118·004	42·270	14·774	9·315	7·615
4	131·477	36·598	14·136	9·793	6·453
5	114·641	37·084	15·066	8·710	6·902
6	119·668	36·006	13·552	9·168	5·513
7	104·487	33·904	13·789	7·493	6·071
1908	107·495	31·990	11·939	8·546	5·939

these deathrates up to the sixth and, in a few cases, to the tenth were then formed. In our notation m_r is the deathrate in the rth year of life, i.e. from $r-1$ to r years of age, and $\delta_s m_r$ is the sth difference of this deathrate. As we have five deathrates for each sex this involves 10 means, 10 standard deviations and 20 correlation coefficients, but as we have used six successive differences these numbers must be multiplied by seven. The calculation of these differences and of upwards of 150 correlation coefficients has meant very strenuous labour. It must, indeed, be admitted that the application of the variate difference correlation method is not, even with small populations, a light task, but the change from the high positive to low negative and then to high negative values of the correlation is of extraordinary interest, and indicates the stages by which the associations are freed from the spurious influence of the time-factor.

(3) All our correlations are given in Table III (p. 497), but it is desirable to discuss in detail certain groups of them. We take first the correlations of the deathrates in *successive* years. They are:

	Male.	*Female.*
$r_{m_1 m_2}$	$+ \cdot398 \pm \cdot080$	$+ \cdot390 \pm \cdot081$
$r_{m_2 m_3}$	$+ \cdot859 \pm \cdot025$	$+ \cdot864 \pm \cdot024$
$r_{m_3 m_4}$	$+ \cdot924 \pm \cdot014$	$+ \cdot928 \pm \cdot013$
$r_{m_4 m_5}$	$+ \cdot911 \pm \cdot016$	$+ \cdot917 \pm \cdot015$

All these are positive, all are significant and, the first excepted, are very high correlations. There is no significant difference between male and female. The least important is the relation between deaths in infancy and deaths in the first year of childhood. We have in these correlation coefficients the numerical expression of what is obvious in Tables I and II, i.e. as the deathrate in any year of age falls so does the deathrate of the same group in the following year. It is this fact which has led to the erroneous idea that natural selection plays no part in man. The fact, however, simply expresses the continuous change of environment which has been in progress since 1860. During the half-century improved economic conditions, bettered sanitation, and developed medical care have lowered the deathrate at each age*. It is therefore impossible to deduce any argument as to natural selection in man from these correlations until we have removed this continuous influence of the time-factor. This is achieved by the variate difference correlation method. In every case a preliminary examination of Tables I and II shows that the correlation of the first differences of the deathrates of *successive* years is *negative*, and as we take higher and higher differences the intensity of this negative correlation increases, until with the sixth differences it reaches to the

---

* As we have already remarked the infantile deathrate showed little of this improvement till 1905. It was about this same year that the *absolute* number of births in England and Wales began to decline, so that while the population has increased by something like $3\frac{1}{2}$ millions, that population produces about 76,000 fewer babies annually.

very substantial value of about – ·7. In other words a rise in the deathrate of one year of life means a fall in the deathrate of the following year of a most marked kind. While with the sixth differences we are approaching fairly closely steady values it may be doubted whether we have reached them in any case but that of $r_{\delta_6 m_4 . \delta_6 m_5}$. The following are the sixth difference correlations in the case of the deathrates of successive years:

	Male.	*Female.*
$r_{\delta_6 m_1 . \delta_6 m_2}$	– ·688 ± ·090	– ·719 ± ·081
$r_{\delta_6 m_2 . \delta_6 m_3}$	– ·673 ± ·092	– ·660 ± ·095
$r_{\delta_6 m_3 . \delta_6 m_4}$	– ·703 ± ·085	– ·731 ± ·078
$r_{\delta_6 m_4 . \delta_6 m_5}$	– ·695 ± ·087	– ·736 ± ·077

Again the male and female results are in excellent agreement, and we grasp the startling manner in which the new method reverses a judgment based on relations which have been deduced without any regard to secular change.

(4) The question naturally arises: How far are these the "steady" values of the difference correlations measuring the organic relation apart from the time-factor of the deathrates in different years of infancy and childhood?

There are three fundamental tests: (i) The correlation coefficients of successive differences should have ceased to be markedly rising or falling. Table III (p. 497) shows that this is approximately but not absolutely the case, but we have reached a stage in which any further changes are certainly of the order of the probable errors and thus of little significance. The unsteadiness as will be indicated later in better tests is greatest in the differences of the deathrates in the first and second years of life. Here the correlations were taken to the seventh and eighth differences and gave:

	Male.	*Female.*
$r_{\delta_7 m_1 . \delta_7 m_2}$	– ·696 ± ·090	– ·729 ± ·082
$r_{\delta_8 m_1 . \delta_8 m_2}$	– ·692 ± ·094	– ·731 ± ·084

which appear to have reached practical steadiness. Actually the final correlations must be somewhat greater than those obtained from the sixth differences. To push the process further, however, would be of small advantage because higher differences involve introducing earlier data, and the birthrate data before 1855 become more and more unreliable. Again in the extremely high differences, the additional year required for an additional difference if not appertaining to relatively smooth data may in itself, when we have only a small total frequency of 50, produce a certain amount of unsteadiness.

(ii) We may consider the mean values of the differences.

If our first variable be taken* as $x = \phi_1(t) + X$, where X is the intrinsic value of x as apart from the time change, then mean $\delta_{r+1}x$ after steadiness has set in is

Values of σ_X deduced from Sixth Differences for 1st 25, for 2nd 25, and for all 50 years.

	(m_1)		(m_2)		(m_3)		(m_4)		(m_5)	
	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
1st 25 years ...	7·32	6·94	5·51	5·61	2·09	2·30	1·52	1·67	1·05	0·91
All 50 years ...	8·61	7·83	4·71	4·63	1·59	1·77	1·17	1·28	0·86	0·78
2nd 25 years ...	9·70	8·61	3·73	3·37	0·83	0·98	0·66	0·68	0·63	0·62

These values are less steady than we had originally hoped for. Clearly the variability of the X portion of the infantile deathrate has grown greater, and that of the four child deathrates has grown sensibly smaller with the time. The fundamental hypothesis of the variate difference method is therefore only approximately true for this material. We have made some investigations on the assumption that $x = \phi_1(t) + (a + bt)X$, but the values of a and b obtained were by no means satisfactory. We have in hand a further investigation of the problem by the method, originally suggested by one of us, before the difference method was started; namely to subtract from x the value obtained by the best fitting parabola of the sth order in the time and so to reach the actual values of X. The relation of these to the time can then be found with some degree of accuracy. To the male deathrates of the second and fourth years of life we applied parabolae of the third order in the time, and obtained excellent fits; we then subtracted the ordinates of these parabolae from the deathrates and correlated the remainders, d_2 and d_4 say. We found $r_{d_2 d_4} = +{\cdot}312 \pm {\cdot}088$, a value corresponding more nearly with $r_{\delta_5 m_2\, \delta_5 m_4}$ than $r_{\delta_3 m_2\, \delta_3 m_4}$, and indicating that we might more rapidly approach final values by this method than by that of variate differences. But the fitting of high order parabolae is very laborious; at the same time the graphs give excellent tests of the accuracy of the work, and we obtain the actual values of what we have termed X and Y, as represented by d_2 and d_4. We then correlated the *numerical* value of d_2 with the time and found $r_{d_2 t} = -{\cdot}284 \pm {\cdot}089$. It is clear that with correlations of this order with the time, $r_{d_2 d_4}$ would not be modified by the extent of its probable error if we found the partial correlation ${}_t r_{d_2 d_4}$, or corrected the correlation of d_2 and d_4 for the time. There is another point, however, which justifies us in disregarding this variation of X and Y with the time as of secondary importance. The correlation of X with the time is *positive* in the first year's mortality and *negative* in the following four years; thus while it would certainly tend to give a negative value to r_{XY} for the 1st and 2nd years of life, it would tend to give a positive value to the correlation for all successive pairs of years beyond the 1st and 2nd. Now all such successive pairs of years have high negative values, which are therefore minimum values, but these values are all in excellent agreement—roughly equal to $-{\cdot}7$—with that found for the 1st and 2nd years of life. We therefore concluded that the influence of the time on the deviations from the secular curve of change, although very sensible, is of no substantial importance for the correlations.

---

* One of the bases of the variate difference correlation method lies in the assumption that the intrinsic variation is superposed on a secular change of a continuous character; the causes which determined the intrinsic variation X are supposed to be sensibly independent of the time for the period under consideration. We conceive the secular change as given by a parabola, say, of the sth order, but the deviations from this curve are supposed in magnitude and sense to be independent of the time, i.e. due to chance causes which are the same in 1850 as in 1900. This assumption is an important one and must lead to our seeking relatively short periods consistent with a numerical frequency sufficient for significance. It can be roughly tested, of course, by considering σ_X as found from, say, the first and second halves of our observations. In our own case we found:

equal to mean $\delta_{r+1}X$, and this (taking, as we have done, 'backward' differences) is given (the C's being the usual binomial coefficients) by

$$\frac{-(X - {}_rC_1\,X_{-1} + {}_rC_2\,X_{-2} - \dots) + (X_n - {}_rC_1\,X_{n-1} + {}_rC_2\,X_{n-2} - \dots)}{n}.$$

Now if we remember that the X's have chance values uncorrelated with each other then we shall have for the squared standard deviation of the mean $\delta_{r+1}X$,

$$\sigma^2_{\text{mean }\delta_{r+1}X} = \frac{2\sigma_X^2\,(1 + {}_rC_1^2 + {}_rC_2^2 + \dots + {}_rC_r^2)}{n^2},$$

$$= 2\,\frac{|\underline{2r}}{|\underline{r}\,|\underline{r}}\,\frac{\sigma_X^2}{n^2}.$$

Or, the probable error of the mean $(r+1)$th difference after the steady values have been reached

$$= \cdot 67449\sqrt{2\,\frac{|\underline{2r}}{|\underline{r}\,|\underline{r}}}\,.\,\frac{\sigma_X}{n}.$$

At first sight this appears of no value, because σ_X is unknown, but Dr Anderson has given $\sigma_{\delta_{r+1}x}$ in terms of σ_X when steady values have been reached[*], i.e.

$$\sigma^2_{\delta_{r+1}x} = \sigma^2_{\delta_{r+1}X} = \frac{|\underline{2(r+1)}}{|\underline{r+1}\,|\underline{r+1}}\,\sigma^2_X.$$

From this we deduce the probable error of a mean rth difference to be

$$\cdot 67449\sqrt{\frac{r}{2r-1}}\,\frac{\sigma_{\delta_r x}}{n},$$

when we assume steadiness reached.

The values of the means of the differences with their probable errors on the assumption of steadiness are given in Table IV, and the ratio of the means to their differences in Table V.

It will be seen that the positive and negative signs are not scattered quite as much at random as we might have hoped and that this is especially the case in the infantile mortality differences[†]. If we take all the ratios of the means to their probable errors except the first difference, we find their average value 1·16; it should be of course 1·18. Of these ratios 33 are positive and 25 negative. If we omit the ratios for the first year of life, we find 24 negative and 20 positive, while the mean value $= \cdot 98$ as against 1·18, the theoretical ratio of the mean to its probable error. It is obvious that the infantile mortality differences are those which are anomalous. Otherwise the mean differences vary fairly satisfactorily

---

[*] *Biometrika*, Vol. x. p. 272.

[†] It may be noted that at the beginning of the period we have the disturbing influence of war and at the end of the period wholly changed conditions due to a great limitation of births. The means depend on differences of mortality under these conditions.

TABLE III. *Correlations of Deathrates and Differences of Deathrates.*

		First Year (m_1)		Second Year (m_2)		Third Year (m_3)		Fourth Year (m_4)		Fifth Year (m_5)	
		♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
First Year (m_1)	m	—	—	+ ·398 ± ·080	+ ·390 ± ·081	+ ·401 ± ·080	+ ·404 ± ·080	+ ·436 ± ·077	+ ·427 ± ·078	+ ·494 ± ·072	+ ·328 ± ·085
	$\delta_1 m$	—	—	− ·531 ± ·084	− ·545 ± ·082	− ·008 ± ·116	− ·042 ± ·116	+ ·166 ± ·113	+ ·151 ± ·114	+ ·015 ± ·116	− ·003 ± ·116
	$\delta_2 m$	—	—	− ·608 ± ·083	− ·629 ± ·086	+ ·081 ± ·131	+ ·065 ± ·132	+ ·215 ± ·126	+ ·128 ± ·130	− ·086 ± ·131	− ·188 ± ·128
	$\delta_3 m$	—	—	− ·645 ± ·084	− ·667 ± ·080	+ ·151 ± ·141	+ ·109 ± ·142	+ ·238 ± ·136	+ ·118 ± ·142	− ·165 ± ·140	− ·254 ± ·135
	$\delta_4 m$	—	—	− ·664 ± ·086	− ·689 ± ·081	+ ·192 ± ·153	+ ·149 ± ·153	+ ·216 ± ·153	+ ·081 ± ·152	− ·188 ± ·153	− ·291 ± ·140
	$\delta_5 m$	—	—	− ·679 ± ·087	− ·705 ± ·081	+ ·211 ± ·154	+ ·175 ± ·156	+ ·202 ± ·155	+ ·053 ± ·161	− ·216 ± ·154	− ·285 ± ·148
	$\delta_6 m$	—	—	− ·688 ± ·090	− ·719 ± ·081	+ ·227 ± ·159	+ ·200 ± ·161	+ ·205 ± ·161	+ ·035 ± ·168	− ·181 ± ·630	− ·251 ± ·158
	$\delta_7 m$	—	—	− ·696 ± ·090	− ·729 ± ·082	—	—	—	—	—	—
	$\delta_8 m$	—	—	− ·692 ± ·094	− ·731 ± ·084	—	—	—	—	—	—
Second Year (m_2)	m	—	—	—	—	+ ·859 ± ·025	+ ·864 ± ·024	+ ·827 ± ·030	+ ·844 ± ·030	+ ·788 ± ·036	+ ·812 ± ·032
	$\delta_1 m$	—	—	—	—	− ·330 ± ·104	− ·360 ± ·101	+ ·028 ± ·116	+ ·077 ± ·116	− ·143 ± ·114	− ·042 ± ·116
	$\delta_2 m$	—	—	—	—	− ·506 ± ·098	− ·515 ± ·097	+ ·140 ± ·130	+ ·196 ± ·127	− ·073 ± ·132	+ ·129 ± ·130
	$\delta_3 m$	—	—	—	—	− ·587 ± ·094	− ·567 ± ·098	+ ·219 ± ·137	+ ·257 ± ·135	− ·008 ± ·144	+ ·069 ± ·140
	$\delta_4 m$	—	—	—	—	− ·636 ± ·091	− ·614 ± ·096	+ ·282 ± ·141	+ ·304 ± ·139	− ·022 ± ·153	+ ·085 ± ·152
	$\delta_5 m$	—	—	—	—	− ·660 ± ·091	− ·641 ± ·095	+ ·319 ± ·145	+ ·346 ± ·142	− ·015 ± ·161	+ ·082 ± ·160
	$\delta_6 m$	—	—	—	—	− ·673 ± ·092	− ·660 ± ·095	+ ·339 ± ·149	+ ·377 ± ·144	− ·030 ± ·168	+ ·072 ± ·167
Third Year (m_3)	m	—	—	—	—	—	—	+ ·924 ± ·014	+ ·928 ± ·013	+ ·862 ± ·024	+ ·881 ± ·021
	$\delta_1 m$	—	—	—	—	—	—	− ·043 ± ·116	− ·094 ± ·115	− ·098 ± ·115	− ·101 ± ·115
	$\delta_2 m$	—	—	—	—	—	—	− ·316 ± ·119	− ·372 ± ·114	+ ·064 ± ·132	+ ·094 ± ·131
	$\delta_3 m$	—	—	—	—	—	—	− ·529 ± ·104	− ·568 ± ·098	+ ·215 ± ·137	+ ·200 ± ·138
	$\delta_4 m$	—	—	—	—	—	—	− ·618 ± ·095	− ·650 ± ·089	+ ·332 ± ·136	+ ·310 ± ·139
	$\delta_5 m$	—	—	—	—	—	—	− ·666 ± ·090	− ·659 ± ·091	+ ·381 ± ·138	+ ·356 ± ·141
	$\delta_6 m$	—	—	—	—	—	—	− ·703 ± ·085	− ·731 ± ·078	+ ·397 ± ·142	+ ·393 ± ·142
Fourth Year (m_4)	m	—	—	—	—	—	—	—	—	+ ·911 ± ·016	+ ·917 ± ·015
	$\delta_1 m$	—	—	—	—	—	—	—	—	+ ·089 ± ·116	− ·015 ± ·116
	$\delta_2 m$	—	—	—	—	—	—	—	—	− ·272 ± ·123	− ·207 ± ·127
	$\delta_3 m$	—	—	—	—	—	—	—	—	− ·480 ± ·111	− ·567 ± ·098
	$\delta_4 m$	—	—	—	—	—	—	—	—	− ·624 ± ·094	− ·683 ± ·082
	$\delta_5 m$	—	—	—	—	—	—	—	—	− ·705 ± ·081	− ·741 ± ·073
	$\delta_6 m$	—	—	—	—	—	—	—	—	− ·695 ± ·087	− ·736 ± ·077

TABLE IV. *Means and their Probable Errors.*

Variate	Year: 0—1 (m_1)		Year: 1—2 (m_2)		Year: 2—3 (m_3)		Year: 3—4 (m_4)		Year: 4—5 (m_5)	
	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
Actual Deathrate, m	160·67 ± 1·06	132·30 ± ·96	55·15 ± ·88	51·74 ± ·86	24·01 ± ·63	23·63 ± ·65	15·45 ± ·47	15·48 ± ·47	11·27 ± ·36	11·10 ± ·35
1st Difference, $\delta_1 m$ ...	− ·793 ± ·145	− ·722 ± ·133	− ·692 ± ·074	− ·650 ± ·074	− ·361 ± ·033	− ·356 ± ·035	− ·223 ± ·029	− ·245 ± ·029	− ·206 ± ·023	− ·223 ± ·021
2nd " $\delta_2 m$ ...	− ·029 ± ·213	+ ·032 ± ·195	− ·066 ± ·111	− ·011 ± ·110	+ ·072 ± ·045	+ ·103 ± ·049	+ ·014 ± ·035	+ ·005 ± ·036	− ·052 ± ·027	− ·057 ± ·022
3rd " $\delta_3 m$ ...	+ ·497 ± ·380	+ ·582 ± ·348	+ ·035 ± ·202	+ ·110 ± ·199	− ·003 ± ·076	+ ·037 ± ·083	− ·084 ± ·057	− ·085 ± ·060	− ·060 ± ·044	− ·066 ± ·037
4th " $\delta_4 m$ ...	+ 1·343 ± ·712	+ 1·381 ± ·651	+ ·017 ± ·381	+ ·114 ± ·377	− ·237 ± ·135	− ·173 ± ·150	− ·104 ± ·101	− ·101 ± ·110	+ ·028 ± ·074	+ ·021 ± ·066
5th " $\delta_5 m$ ...	+ 2·825 ± 1·357	+ 2·680 ± 1·239	− ·145 ± ·734	− ·046 ± ·724	− ·597 ± ·253	− ·482 ± ·281	+ ·137 ± ·186	+ ·122 ± ·204	+ ·172 ± ·137	+ ·166 ± ·123
6th " $\delta_6 m$ ...	+ 5·618 ± 2·606	+ 5·056 ± 2·372	− ·325 ± 1·427	− ·273 ± 1·404	− ·953 ± ·482	− ·771 ± ·535	+ ·855 ± ·356	+ ·766 ± ·387	+ ·231 ± ·261	+ ·212 ± ·237
7th " $\delta_7 m$ ...	+ 10·833 ± 5·019	+ 9·461 ± 4·549	− ·119 ± 2·784	− ·225 ± 2·735	—	—	—	—	—	—
8th " $\delta_8 m$ ...	+ 19·620 ± 9·704	+ 16·838 ± 8·776	+ 1·273 ± 5·457	+ ·735 ± 5·346	—	—	—	—	—	—

For the sth year $\delta_1 m = m_{s+1} - m_s$ and the differences of higher orders follow the same rule.

TABLE V. *Ratio of Means of Differences to their Probable Errors.*

Variate	Year: 0—1 (m_1)		Year: 1—2 (m_2)		Year: 2—3 (m_3)		Year: 3—4 (m_4)		Year: 4—5 (m_5)	
	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
1st Difference, $\delta_1 m$ ...	− 5·47	− 5·42	− 9·35	− 8·78	− 10·94	− 10·17	− 7·69	− 8·45	− 8·96	− 10·62
2nd " $\delta_2 m$ ...	− ·14	+ ·16	− ·59	− ·10	+ 1·60	+ 2·10	+ ·40	+ ·14	− 1·93	− 2·59
3rd " $\delta_3 m$ ...	+ 1·31	+ 1·67	+ ·17	+ ·55	− ·04	+ ·45	− 1·47	− 1·42	− 1·36	− 1·78
4th " $\delta_4 m$ ...	+ 1·89	+ 2·12	+ ·04	+ ·30	− 1·76	− 1·15	− 1·03	− ·92	+ ·38	+ ·32
5th " $\delta_5 m$ ...	+ 2·08	+ 2·16	− ·20	− ·64	− 2·36	− 1·72	+ ·74	+ ·60	+ 1·26	+ 1·35
6th " $\delta_6 m$ ...	+ 2·16	+ 2·13	− ·23	− ·19	− 1·98	− 1·44	+ 2·40	+ 1·98	+ ·89	+ ·89
7th " $\delta_7 m$ ...	+ 2·16	+ 2·09	− ·43	− ·82	—	—	—	—	—	—
8th " $\delta_8 m$ ...	+ 2·02	+ 1·92	+ ·23	+ ·14	—	—	—	—	—	—

round zero in the required manner. The interest of this test is that we see that the bulk of the time effect has been removed even when we reach the second difference, a result confirmed by the fact that the correlation of the deathrates' second differences is in every case already substantially negative.

(iii) A third set of tests are those which are based on the standard deviations of the differences. In the first place if we assume steadiness to have set in, we can calculate σ_X, the intrinsic standard deviation from the known value of $\sigma_{\delta_s x}$, by means of Dr Anderson's formula cited above (p. 496). Table VI gives the intrinsic values of σ_X, i.e. σ_X as deduced from the variability of the differences. It will be at once observed that for the third difference the mortality ratios of the third, fourth and fifth years of life reach steady standard deviations. In the case of the first year of life it is not till the eighth difference that this result is reached, while in the case of the second year, it can hardly be said to have been obtained with the ninth difference. A distinction should be noted here of which the exact physical significance is not obvious to us. In the second, third and fourth years the intrinsic standard deviations *fall* to steady values, but in the first and second years they *rise* towards those values and these are just the cases where steady values are not absolutely reached.

TABLE VI.

Intrinsic Standard Deviations (σ_X).

Order of Difference	Year: 0—1 (m_1)		Year: 1—2 (m_2)		Year: 2—3 (m_3)		Year: 3—4 (m_4)		Year: 4—5 (m_5)	
	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
1st	7·62	6·96	3·90	3·86	1·75	1·83	1·53	1·52	1·23	1·09
2nd	7·89	7·22	4·13	4·09	1·67	1·81	1·29	1·34	1·00	·83
3rd	8·14	7·45	4·32	4·28	1·62	1·78	1·21	1·29	·93	·80
4th	8·34	7·63	4·47	4·42	1·59	1·76	1·18	1·29	·87	·77
5th	8·50	7·76	4·60	4·54	1·58	1·76	1·17	1·28	·86	·77
6th	8·61	7·83	4·71	4·63	1·59	1·77	1·17	1·28	·86	·78
7th	8·66	7·84	4·80	4·72	—	—	—	—	—	—
8th	8·68	7·85	4·88	4·78	—	—	—	—	—	—
9th	—	—	4·97	4·82	—	—	—	—	—	—

(iv) There is another test for the standard deviations of the differences deduced by Cave and Pearson from the Andersonian results and used by them in their memoir on Italian Index Values*, namely as steadiness is approached the ratio of the squares of standard deviation of successive differences should approach closer and closer to 4, the exact value being

$$\frac{\sigma^2_{\delta_s m}}{\sigma^2_{\delta_{s-1} m}} = 4 - \frac{2}{s}.$$

---

* *Biometrika*, Vol. x. p. 346.

Table VII shows how rapidly the system approximates to the theoretical values in the case of the higher differences.

On the basis of all the tests we have applied we may, we think, conclude that by the sixth difference we have reached values for the correlation of deathrates in successive years which are in all probability close to the organic or intrinsic values. Only in the first and second years of life is steadiness not absolutely reached, but for practical purposes but little change can be anticipated in the correlation coefficients.

TABLE VII.

Ratio of Squared Standard Deviations.

s	m_1		m_2		m_3		m_4		m_5		Mean	Mean	Theory
	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	
1	·949	·956	·354	·369	·142	·144	·194	·192	·211	·181	·370	·368	2
2	3·199	3·221	3·374	3·384	2·731	2·934	2·127	2·317	1·996	1·728	2·685	2·717	3
3	3·547	3·552	3·638	3·633	3·133	3·240	2·944	3·086	2·896	3·109	3·232	3·324	3·333
4	3·676	3·673	3·754	3·741	3·363	3·428	3·341	3·504	3·054	3·227	3·438	3·515	3·500
5	3·738	3·723	3·811	3·793	3·591	3·604	3·509	3·562	3·504	3·641	3·631	3·665	3·600
6	3·756	3·734	3·848	3·828	3·690	3·683	3·708	3·648	3·691	3·758	3·739	3·730	3·667

(5) We can look at the association of deathrates in successive years from another standpoint. We can ask if there be an increase of 10 points in the deathrate for a given year, what increase or decrease will there be of deathrate in the same group in the following year?

In Table VIII below the second column gives the spurious change which is apparent in the crude data, the third column gives the real organic change which is discovered when the time-factor is removed.

TABLE VIII.

Association of Deathrates without and with Annulment of Time-factor.

Result of an increase of 10 deaths per mille in one year of life on the deaths per mille in the next year.

Increase of 10 in Deathrate of	Disregarding Time-factor		Annulling Time-factor	
	♂	♀	♂	♀
1st Year on 2nd Year ...	Increase 3·3	Increase 3·8	Decrease 3·7	Decrease 4·3
2nd Year on 3rd Year ...	Increase 6·1	Increase 6·6	Decrease 2·2	Decrease 2·5
3rd Year on 4th Year ...	Increase 6·9	Increase 6·7	Decrease 5·2	Decrease 5·3
4th Year on 5th Year ...	Increase 7·0	Increase 6·8	Decrease 5·1	Decrease 4·5

It is easy to see how those who contented themselves with crude deathrates, making no allowance for the betterment of deathrates with the time, interpreted a higher deathrate in one year to mean a higher deathrate in the next year of life, and so questioned whether natural selection applied to civilised man. As a matter of fact we see that the true organic relationship of deathrates is much more probably summed up in the statement that a decrease or an increase of deathrate in one year of infancy or childhood is in each case followed by an increase or a decrease in the deathrate of the survivors of the same group in the following year. Disregarding the time-factor we have a result quite incompatible with natural selection; annulling the time-factor, we have a result not only compatible with natural selection, but very difficult of any other interpretation than that of a selective deathrate, i.e. a heavy mortality means a selection of the weaker members, and the exposure to risk in the following year of a selected or stronger population, which has accordingly a lesser deathrate.

(6) We now turn to the problem of how far this influence extends, or probably it would be better to phrase it: how far this influence can be traced. It is not only that the age group we follow does not absolutely consist of the same individuals but even with those members that are the same there is very often change of environment due not to time but to a change of locality or of economic condition affecting individuals. Added to this there is a continuous immigration and emigration. But beyond these causes weakening the association, there is another difficulty of great importance arising from what has happened in the intervening years. We wish to find out how an increase of deathrate in the sth year of life affects the deathrate in the $(s+2)$th year of life, but the events in the $(s+1)$th year will largely dominate and, perhaps, screen the results we are seeking. Such problems are always arising in statistical research. For example, a child may resemble its grandfather simply because both grandfather and child are like the child's father. We know that the problem is answered statistically by inquiring what is the relation between a character in the child and the grandparent for a constant value of the character in the parent. In precisely the same manner we must in the present problem inquire: What is the correlation between the deathrates in the sth and $(s+2)$th year of life for constant deathrate in the $(s+1)$th year of life?

TABLE IX.

Influence of Natural Selection at Interval of Two Years.

Partial Correlation of	For constant	♂	♀
$\delta_6 m_1$ and $\delta_6 m_3$ ...	$\delta_6 m_2$	$-\cdot4307$	$-\cdot5242$
$\delta_6 m_2$ and $\delta_6 m_4$ ...	$\delta_6 m_3$	$-\cdot2555$	$-\cdot2058$
$\delta_6 m_3$ and $\delta_6 m_5$ ...	$\delta_6 m_4$	$-\cdot1798$	$-\cdot3129$

We shall of course work with the sixth difference correlations in order to free ourselves substantially from the time-factor.

Here again the judgment based on the partial correlation of the crude deathrates is in all six cases reversed. For every one of the partial coefficients of crude deathrates shows that for intervening year with a constant deathrate, an increase of deathrate in the earlier year means an increase, not a decrease in the later year. Actually an increase in the one year is shown in Table X in all cases to be followed by a decrease at two years' interval.

TABLE X.

Influence of Natural Selection at Interval of Two Years.

Result of an increase of 10 deaths per mille in the second following year.

Increase of 10 in Deathrate of	For constant deathrate in	♂	♀
1st Year on that of 3rd Year	2nd Year	Decrease ·81	Decrease 1·28
2nd Year on that of 4th Year	3rd Year	Decrease ·61	Decrease ·52
3rd Year on that of 5th Year	4th Year	Decrease ·99	Decrease 1·4

It will be seen that these values are appreciable although far less important than the decreases produced in a following year by an increase in the immediately preceding year. Thus we judge that a selection of the weakly children in one year is largely influential on the deathrate of the immediately following year, and diminishes, as we might anticipate, with increase of time.

Some objection might, however, be taken to the sixth difference correlations, when we consider deathrates of the same group two years apart. They are

	Male.	*Female.*
$r_{\delta_6 m_1 . \delta_6 m_3}$	$+ \cdot 227 \pm \cdot 159$	$+ \cdot 200 \pm \cdot 161$
$r_{\delta_6 m_2 . \delta_6 m_4}$	$+ \cdot 339 \pm \cdot 149$	$+ \cdot 377 \pm \cdot 144$
$r_{\delta_6 m_3 . \delta_6 m_5}$	$+ \cdot 397 \pm \cdot 142$	$+ \cdot 393 \pm \cdot 142$

It will be seen that while they are all of the same sign and fairly accordant for both sexes the probable errors are becoming very substantial relative to the coefficients. We have indeed too limited a range of years.

(7) If now we take out the correlation coefficients of the sixth differences for three years' interval, and again for four years' interval we find great irregularities.

	Male.	*Female.*
$r_{\delta_6 m_1 . \delta_6 m_4}$	$+ \cdot 205 \pm \cdot 161$	$+ \cdot 035 \pm \cdot 168$
$r_{\delta_6 m_2 . \delta_6 m_5}$	$- \cdot 030 \pm \cdot 168$	$+ \cdot 072 \pm \cdot 167$
$r_{\delta_6 m_1 . \delta_6 m_5}$	$- \cdot 181 \pm \cdot 163$	$- \cdot 251 \pm \cdot 158$

The correlations now do not agree in sign, they are insignificant having regard to their probable errors, and there is no close correspondence for the two sexes.

We should need a far longer period than 50 years to determine certainly even the signs of these correlations, and their real magnitudes would require still ampler data. It would appear impossible to assert on the basis of the above values of the correlations at three and four years' intervals more than the insignificance of the associations between deathrates of the same groups at intervals of more than two years*. In other words the effect of intense selection appears to be exhausted after an interval of two years. The word "appears" is used purposely because there must be some spurious weakening of the effect due to our not being able to follow absolutely the same individuals.

(8) We have further studied to some extent the relationship between the male and female deathrates. There is almost perfect correlation between male and female deathrates in any given year of life after we annul the time-factor. Thus, if we represent female deathrates by m', we have as illustrations:

$$r_{\delta_6 m_1 \,.\, \delta_6 m_1'} = +\cdot 9905,$$
$$r_{\delta_6 m_2 \,.\, \delta_6 m_2'} = +\cdot 9880,$$
$$r_{\delta_6 m_3 \,.\, \delta_6 m_3'} = +\cdot 9687,$$
$$r_{\delta_6 m_4 \,.\, \delta_6 m_4'} = +\cdot 9800.$$

Of course the sole significance of these values lies in the fact that years of stress, whether due to climatic or epidemic causes, affect equally infants or children of both sexes of the same age. But these very high values in our opinion cast considerable doubt on the partial correlations derived from them. We have in fact

$$_3r_{12} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{N}{D},$$

and if we suppose r_{12} and r_{23} nearly equal, then if r_{13} be of the above high value N will be extremely small, but D is also, owing to the presence of the factor $\sqrt{1 - r_{13}^2}$, very small. Thus $_3r_{12}$ although it may be very considerable is the ratio

---

* Actually the partial correlations of the sixth differences at three years' interval based on the above values are:

Correlation of	For constant	♂	♀
$\delta_6 m_1$ and $\delta_6 m_4$	$\delta_6 m_2$ and $\delta_6 m_3$	+·526	+·181
$\delta_6 m_2$ and $\delta_6 m_5$	$\delta_6 m_3$ and $\delta_6 m_4$	+·251	+·485

These are certainly all positive, but they are irregular as between the sexes and probably quite unreliable for the reasons already given. Should a more extended experience show that there is a real if slight *positive* correlation between deathrates at three years' interval, while there is considerable *negative* correlation at one and two years' intervals, we should be compelled to discuss whether there may not be something periodic in the nature of the heavy and light deathrates of infancy and childhood. We have been unable to trace any sign of such periodicity either in the deathrates or in the graphs drawn, but we do not believe that a very short periodicity would be eliminated by the variate difference method using any moderate number of differences. We cannot on this point accept Dr Anderson's view. See *Biometrika*, Vol. x. p. 279.

of two small quantities and any disturbing cause which but slightly modifies the value of either r_{12} or r_{23} may even change the sign of N and so swing ${}_3r_{12}$ over from a considerable positive to a considerable negative value*.

We can consider the correlations between the female deathrate in one year and the male deathrate in a second year, supposing of course time influence annulled. We have

$$\begin{aligned} r_{\delta_6 m_1 . \delta_6 m_2'} &= -\cdot 6674 & (r_{\delta_6 m_1 . \delta_6 m_2} &= -\cdot 6879), \\ r_{\delta_6 m_1' . \delta_6 m_2} &= -\cdot 7337 & (r_{\delta_6 m_1' . \delta_6 m_2'} &= -\cdot 7188), \\ r_{\delta_6 m_3 . \delta_6 m_4'} &= -\cdot 7313 & (r_{\delta_6 m_3 . \delta_6 m_4} &= -\cdot 7032), \\ r_{\delta_6 m_3' . \delta_6 m_4} &= -\cdot 7278 & (r_{\delta_6 m_3' . \delta_6 m_4'} &= -\cdot 7313). \end{aligned}$$

Thus we see that the same remarkably high negative correlations exist between the male and female deathrates of successive years of groups born in the same year as exist between male and male or female and female deathrates within the same group in successive years. In fact in two out of the four correlations the *cross* relationships are higher than the direct, although the differences are scarcely significant. Here again there is nothing noteworthy, considering the very high correlations just noted to exist between the male and female deathrates of groups born in the same year. We can, however, endeavour to correct such values by finding the relationship between the deathrate in females in the first year of life and males born in the same year in their second year of life for a constant deathrate of males in the first year of life. Or still more stringently between the deathrates of females in the first year of life with males in the second year of life for constant male deathrate in the first year of life and constant female deathrate in the second year of life. We should anticipate that such values would come out small or insignificant, if our interpretation of the high negative correlations between deathrates of the same group in successive years of life be a correct one, i.e. that the high deathrate leaves a stronger population. For a heavy deathrate in the females of one year should not leave a stronger population of males for the following year *after correction by partial correlation.*

We obtained the following correlations:

$$\begin{aligned} {}_{\delta_6 m_1} r_{\delta_6 m_1' . \delta_6 m_2} &= -\cdot 5240 \pm \cdot 0692, \\ {}_{\delta_6 m_1'} r_{\delta_6 m_1 . \delta_6 m_2'} &= +\cdot 4665 \pm \cdot 0746. \end{aligned}$$

---

* The reader must note that we say a "disturbing cause"; it is not the mere result of random sampling affecting N. The probable error of $N = r_{12} - r_{13} r_{23}$ for a sample of size n is given by

$$\cdot 67449 \sigma_N = \cdot 67449 \frac{1}{\sqrt{n}} \{D^2 - N^2 [2(1 - r_{13}^2) + 2(1 - r_{23}^2) + 1 - r_{12}^2 - 3]\}^{\frac{1}{2}},$$

and is thus quite easy to calculate. We have tested it on a number of cases of partial correlations worked out for this paper and find that if $\cdot 67449\sigma_N$ is of the same order as N, then $\cdot 67449\sigma_{{}_3r_{12}}$ is of much the same order as ${}_3r_{12}$. In other words, if N is so small relative to its probable error that it might easily have a reversed sign, then ${}_3r_{12}$ is insignificant as compared to its probable error also. For example, $N = \cdot 0446$ and $D = \cdot 0956$ leads to ${}_3r_{12} = \cdot 4665$ with a probable error of $\cdot 0746$. ${}_3r_{12}$ is accordingly considerable and significant, but the probable error of N is only $\cdot 0105$, and we can hardly suppose the sign of ${}_3r_{12}$ due to a random sampling variation in the sign of N.

These values were so startling and so contradictory, that we proceeded to eighth differences with the results:

$$_{\delta_8 m_1}r_{\delta_8 m_1' \,.\, \delta_8 m_2} = -\cdot 6013 \pm \cdot 0609$$

$$_{\delta_8 m_1'}r_{\delta_8 m_1 \,.\, \delta_8 m_2'} = +\cdot 5481 \pm \cdot 0667,$$

which emphasised as well as confirmed the previous results.

Now it seems absurd to suppose that the deaths of female infants in one year can organically influence the deaths of males of the same group in the next year, or male infants the deaths of females in the successive year. But the extraordinary feature of these results is that while a high deathrate of female infants *lessens* the deathrate of males in the second year of life of the same group, a high deathrate of male infants increases the deathrate of females in the second year of life of the same group.

In order to throw further light on the matter we investigated male and female deathrate correlations in the third and fourth years of life. We found

$$_{\delta_6 m_3}r_{\delta_6 m_3' \,.\, \delta_6 m_4} = -\cdot 2640 \pm \cdot 0887,$$

$$_{\delta_6 m_3'}r_{\delta_6 m_3 \,.\, \delta_6 m_4'} = -\cdot 0082 \pm \cdot 0954.$$

The second is practically zero, the first of no importance having regard to the high values of the correlation of deathrates of groups of the same sex in the third and fourth years of life (♂: − ·703 ± ·085; ♀: − ·731 ± ·078). Had we come to these values at first we should have been content, but the cross relation between the infant deaths of one sex and the deaths in the second year of life of the opposite sex was undoubtedly puzzling.

We then proceeded to still further limit our conditions by determining the partial correlation between female infants in one year and males in the second year of life of the same birth-year when the deathrates of the males in the first year of life and of the females in the second were both constant. We obtained

$$_{\delta_6 m_1 \,.\, \delta_6 m_2'}r_{\delta_6 m_1' \,.\, \delta_6 m_2} = +\cdot 1632 \pm \cdot 0928,$$

$$_{\delta_6 m_1' \,.\, \delta_6 m_2}r_{\delta_6 m_1 \,.\, \delta_6 m_2'} = +\cdot 2997 \pm \cdot 0868.$$

Having regard to their probable errors these are of a quite different and negligible significance when compared with the values of

$$_{\delta_6 m_1}r_{\delta_6 m_1' \,.\, \delta_6 m_2} \quad \text{and} \quad _{\delta_6 m_1'}r_{\delta_6 m_1 \,.\, \delta_6 m_2'}$$

given above.

It is worth while noting that

$$_{\delta_6 m_2'}r_{\delta_6 m_1' \,.\, \delta_6 m_2} = -\cdot 2188 \pm \cdot 0908,$$

$$_{\delta_6 m_2}r_{\delta_6 m_1 \,.\, \delta_6 m_2'} = +\cdot 1088 \pm \cdot 0943$$

also give values of no practical importance. Or, to annul the spurious influence of infantile deaths of one sex, A, on deaths in the second year of sex, B, of the same group, it is more effective to render constant the deaths of A in the second year of life than of B in the first year of life.

In the light of this result we have found the correlations between deathrates of sex A in the third and sex B in the fourth year of life, for constant deathrate of sex A in the fourth year of life.

We have

$$_{\delta_6 m_4{}'}r_{\delta_6 m_3{}' \,.\, \delta_6 m_4} = -\cdot0818 \pm \cdot0948,$$

$$_{\delta_6 m_4}r_{\delta_6 m_3 \,.\, \delta_6 m_4{}'} = -\cdot1477 \pm \cdot0933.$$

Both of these may be taken as zero, having regard to their probable errors.

Thus on the whole, while the relation between the deathrate of a group of one sex in one year and the deathrate of the remainder in the following year of life appears after the annulment of the time-factor to be very considerable and negative, there does not appear to be any organic relation between the deathrate of sex A in one year and sex B in the following year, if we proceed by the method of partial correlation. But at the same time we believe that this method must be used with very considerable caution, and that to avoid erroneous conclusions the whole problem must be investigated from a variety of standpoints in cases like the present where one of the three total correlations is extremely high. The numerator N ranges in the cases we have been discussing from about ·01 to ·05 and with a small total frequency like 50, any disturbing cause—apart from random variation—may have marked influence*.

(9) The conclusion which we have formed is that in the present problem of natural selection it is probably better to annul the environmental factor by the variate difference method rather than to proceed by the method of partial correlation as we have hitherto done.

By the former method we have shown that for both sexes a heavy deathrate in one year of life means a markedly lower deathrate in the same group in the following year of life, and that this extends in a lessened degree to the year following that, but is not by the present method easy to trace further. It is difficult to believe that this important fact can be due to any other source than the influence of natural selection, i.e. a heavy mortality leaves behind it a stronger population. Nature is not concerned with the moral or the immoral, which are standards of human conduct, and the duty of the naturalist is to point out what goes on in Nature. There can now scarcely be a doubt that even in highly organised human communities the deathrate is selective, and physical fitness is the criterion for survival. To assert the existence of this selection and measure its intensity must be distinguished from advocacy of a high infant mortality as a factor of racial efficiency. This reminder is the more needful as there are not wanting those who assert that demonstrating the existence of natural selection in man is identical with decrying all efforts to reduce the infantile deathrate.

We have to acknowledge the great assistance we have received from our colleague Miss Beatrice M. Cave in the laborious arithmetical work of this paper.

---

* If $F=N/D$, where N and D are both small, but F finite, then $\delta F/F=\delta N/N-\delta D/D$ and small disturbances produce great results in F.

FREQUENCY DISTRIBUTION OF THE VALUES OF THE CORRELATION COEFFICIENT IN SAMPLES FROM AN INDEFINITELY LARGE POPULATION.

BY R. A. FISHER.

1. My attention was drawn to the problem of the frequency distribution of the correlation coefficient by an article published by Mr H. E. Soper* in 1913. Seeing that the problem might be attacked by means of geometrical ideas, which I had previously found helpful in the consideration of samples, I have examined the two articles by "Student†," upon which Mr Soper's more elaborate work was based, with a view to checking and verifying the conclusions there attained.

"Student," if I do not mistake his intention, desiring primarily to obtain a just estimate of the accuracy to be ascribed to the mean of a small sample, found it necessary to allow for the fact that the mean square error of such a sample is not generally equal to the standard deviation of the normal population from which it is drawn. He was led, in fact, to study the frequency distribution of the mean square error. He calculated algebraically the first four moments of this frequency curve, both about the zero point, and about its mean, observed a simple law to connect the successive moments, and discovered a frequency curve, which fitted his moments, and gave the required law.

Thus if $x_1, x_2, \dots x_n$ are the members of a sample,

$$n\bar{x} = x_1 + x_2 + \dots + x_n,$$

and

$$n\mu^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2,$$

the frequency with which the mean square error lies in the range $d\mu$ is proportional to

$$\mu^{n-2} e^{-\frac{n\mu^2}{2\sigma^2}} d\mu.$$

This result, although arrived at by empirical methods, was established almost beyond reasonable doubt in the first of "Student's" papers. It is, however, of interest to notice that the form establishes itself instantly, when the distribution of the sample is viewed geometrically.

---

* *Biometrika*, Vol. IX. p. 91.

† *Ibid.* Vol. VI. pp. 1 and 302.

In the second of these two papers the more difficult problem of the frequency distribution of the correlation coefficient is attempted. For samples of 2 the frequency distribution between the only two possible values -1 and $+1$ was determined by Sheppard's theorem to be in the ratio $\frac{\pi}{2}+\sin^{-1}\rho : \frac{\pi}{2}-\sin^{-1}\rho$, where ρ is the correlation of the population. Besides this theoretical result, "Student" appeals only to experimental data. From these he derives an empirical form for the distribution when $\rho=0$, and makes several valuable suggestions. It has been the greatest pleasure and interest to myself to observe with what accuracy "Student's" insight has led him to the right conclusions. The form when $\rho=0$ is absolutely correct, and as a further instance I may quote the remark* "I have dealt with the cases of samples of 2 at some length, because it is possible that this limiting value of the distribution, with its mean of $\frac{2}{\pi}\sin^{-1}\rho$ and its second moment coefficient of $1-\left(\frac{2}{\pi}\sin^{-1}\rho\right)^2$, may furnish a clue to the distribution when n is greater than 2." As a matter of fact it is just these quantities with which we shall be concerned.

To Mr Soper's laborious and intricate paper I cannot hope to do justice. I have been able to establish the substantial accuracy and value of his approximations. It is one of the advantages of approaching a problem from opposite standpoints that Mr Soper's forms are most accurate for those larger values of n, where the exact formulae become most complicated.

2. The problem of the frequency distribution of the correlation coefficient r, derived from a sample of n pairs, taken at random from an infinite population, may be solved, when that population can be represented by a normal surface, with the aid of certain very general conceptions derived from the geometry of n dimensional space. In this paper the general form will first be demonstrated, and for a few important cases some of the successive moments will be derived. Incidentally it will be of interest to compare the exact form with Mr Soper's approximation, and with reference to the experimental data supplied by "Student."

If the frequency distribution of the population be specified by the form

$$df=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\,e^{-\frac{1}{1-\rho^2}\left\{\frac{(x-m_1)^2}{2\sigma_1^2}-\frac{2\rho\,(x-m_1)\,(y-m_2)}{2\sigma_1\sigma_2}+\frac{(y-m_2)^2}{2\sigma_2^2}\right\}}\,dx\,dy,$$

where df is the chance that any observation should fall into the range $dx\,dy$, then the chance that n pairs should fall within their specified elements is

$$\frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n}\,e^{-\frac{1}{1-\rho^2}\sum_1^n\left\{\frac{(x-m_1)^2}{2\sigma_1^2}-\frac{2\rho\,(x-m_1)\,(y-m_2)}{2\sigma_1\sigma_2}+\frac{(y-m_2)^2}{2\sigma_2^2}\right\}}\,dx_1dy_1\ldots dx_ndy_n\ldots\text{(I)},$$

and this we interpret as a simple density distribution in $2n$ dimensions.

---

* *Biometrika*, Vol. VI. p. 304.

For the variables x and y it is now necessary to substitute the statistical derivatives determined by the equations

$$n\bar{x}=\sum_1^n(x), \qquad n\bar{y}=\sum_1^n(y),$$
$$n\mu_1^2=\sum_1^n(x-\bar{x})^2, \quad n\mu_2^2=\sum_1^n(y-\bar{y})^2,$$
$$nr\mu_1\mu_2=\sum_1^n(x-\bar{x})(y-\bar{y}),$$

and it is evident that the only difficulty lies in the expression of an element of volume in $2n$ dimensional space in terms of these derivatives.

The five quantities above defined have, in fact, an exceedingly beautiful interpretation in generalised space, which we may now examine.

3. Considering first the space of n dimensions in which the variations of x are represented, the mean and mean square error of n observations are determined by the relations of P, the point representing the n observations, to the line

$$x_1=x_2=x_3=\ldots=x_n,$$

for the perpendicular PM drawn from P upon this line will lie in the region

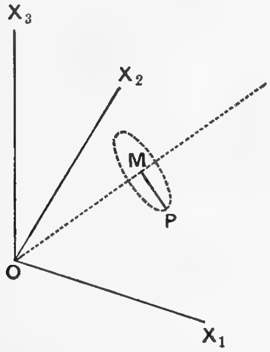
$$x_1+x_2+\ldots+x_n=n\bar{x},$$

and will meet it at the point M, where

$$x_1=\bar{x}, \quad x_2=\bar{x}, \quad \ldots x_n=\bar{x};$$

further, since, $PM^2=(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2$,

the length of PM is $\mu_1\sqrt{n}$.



An element of volume in this n dimensional space may now without difficulty be specified in terms of $\bar{x}$ and μ_1; for, given $\bar{x}$ and μ_1, P must lie on a sphere in $n-1$ dimensions, lying at right angles to the line OM, and the element of volume is

$$C\mu_1^{n-2}d\mu_1d\bar{x},$$

where C is some constant, which need not be determined.

The point in $2n$ dimensional space which is represented by the n pairs of observations must be such that its projection on the n dimensional space, in which x is represented, lies upon a certain sphere of radius $\mu_1\sqrt{n}$, and on the space in which y is represented, upon another sphere of radius $\mu_2\sqrt{n}$, and now, when we come to the interpretation of r, we must observe that to each point on the first sphere there corresponds a certain point on the second sphere, to which it bears the relation

$$\frac{x_1-\bar{x}}{y_1-\bar{y}}=\frac{x_2-\bar{x}}{y_2-\bar{y}}=\ldots=\frac{x_n-\bar{x}}{y_n-\bar{y}}.$$

In general this relation does not hold for the n pairs of observations, and the two projections will not fall at corresponding points on the two spheres. If now one of the spheres be turned round so as to occupy the same space as the other, and so that the lines upon which x_1 and y_1, and the other pairs of coordinates, are measured, coincide, then corresponding points will lie on the same radii, and the correlation coefficient r measures the cosine of the angle between the radii to the two points specified by the observations.

Taking one of the projections as fixed at any point on the sphere of radius μ_2, the region for which r lies in the range dr, is a zone, on the other sphere in $n-1$ dimensions, of radius $\mu_1\sqrt{n}\sqrt{1-r^2}$, and of width $\mu_1\sqrt{n}\,dr/\sqrt{1-r^2}$, and therefore having a volume proportional to $\mu_1^{n-2}(1-r^2)^{\frac{n-4}{2}}\,dr$.

4. We may now turn to the direct simplification of the expression (I), at each stage discarding any factors which do not involve r.

$$e^{-\frac{1}{1-\rho^2}\sum_1^n\left\{\frac{(x-m_1)^2}{2\sigma_1^2}-\frac{2\rho(x-m_1)(y-m_2)}{2\sigma_1\sigma_2}+\frac{(y-m_2)^2}{2\sigma_2^2}\right\}}dx_1dy_1dx_2dy_2\ldots dx_ndy_n$$

may be reduced to

$$e^{-\frac{n}{1-\rho^2}\left\{\frac{(\bar{x}-m_1)^2+\mu_1^2}{2\sigma_1^2}-\frac{2\rho\{r\mu_1\mu_2+(\bar{x}-m_1)(\bar{y}-m_2)\}}{2\sigma_1\sigma_2}+\frac{(\bar{y}-m_2)^2+\mu_2^2}{2\sigma_2^2}\right\}}$$
$$d\bar{x}\,d\bar{y}\,\mu_1^{n-2}d\mu_1\,\mu_2^{n-2}d\mu_2\,(1-r^2)^{\frac{n-4}{2}}\,dr,$$

or to
$$e^{-\frac{n}{1-\rho^2}\left\{\frac{\mu_1^2}{2\sigma_1^2}-\frac{2\rho r\mu_1\mu_2}{2\sigma_1\sigma_2}+\frac{\mu_2^2}{2\sigma_2^2}\right\}}\mu_1^{n-2}\mu_2^{n-2}(1-r^2)^{\frac{n-4}{2}}\,d\mu_1d\mu_2dr.$$

In order to integrate this expression from 0 to ∞, with respect to μ_1 and μ_2, let

$$\zeta=\frac{\mu_1\mu_2}{\sigma_1\sigma_2},\qquad e^z=\frac{\mu_1\sigma_2}{\mu_2\sigma_1},$$

and we have

$$\int_{-\infty}^{\infty}dz\int_0^{\infty}\zeta^{n-2}d\zeta\,.\,e^{-\frac{n}{1-\rho^2}(\cosh z-\rho r)\zeta}\,.\,(1-r^2)^{\frac{n-4}{2}}\,dr,$$

or
$$\int_0^{\infty}\frac{dz}{(\cosh z-\rho r)^{n-1}}\,.\,(1-r^2)^{\frac{n-4}{2}}\,dr,$$

which, on substituting $\cos\theta$ for $-\rho r$, may be expressed in terms of a Legendre function in the form

$$(i \operatorname{cosec}\theta)^{n-1}\, Q_{n-2}(i\cot\theta)\,.\,(1-r^2)^{\frac{n-4}{2}}\, dr \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{II}).$$

Again
$$\int_0^\infty \frac{dz}{\cosh z + \cos\theta} = \frac{\theta}{\sin\theta},$$

so that
$$\int_0^\infty \frac{dz}{(\cosh z + \cos\theta)^{n-1}} = \frac{1}{\underline{|n-2}} \left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-2} \frac{\theta}{\sin\theta},$$

and since this is a function of ρr only, we may express the frequency distribution by the convenient expression

$$(1-r^2)^{\frac{n-4}{2}} \frac{\partial^{n-2}}{\partial r^{n-2}} \left(\frac{\theta}{\sin\theta}\right) dr.$$

Professor Pearson has shown that this last result can be obtained directly from Sheppard's theorem* that

$$\frac{1}{2\pi\Sigma_1\Sigma_2\sqrt{1-R^2}} \int_0^\infty \int_0^\infty e^{-\frac{1}{2(1-R^2)}\left(\frac{\mu_1^2}{\Sigma_1^2} - \frac{2R\mu_1\mu_2}{\Sigma_1\Sigma_2} + \frac{\mu_2^2}{\Sigma_2^2}\right)} d\mu_1\, d\mu_2 = \frac{\cos^{-1}(-R)}{2\pi};$$

making the substitutions

$$\frac{1}{(1-R^2)\Sigma_1^2} = \frac{n}{(1-\rho^2)\sigma_1^2},$$

$$\frac{1}{(1-R^2)\Sigma_2^2} = \frac{n}{(1-\rho^2)\sigma_2^2},$$

$$\frac{R}{(1-R^2)\Sigma_1\Sigma_2} = \frac{nr\rho}{(1-\rho^2)\sigma_1\sigma_2},$$

which give
$$R = \rho r$$

and
$$\cos^{-1}(-R) = \theta,$$

we obtain

$$\frac{n}{\sigma_1\sigma_2(1-\rho^2)} \int_0^\infty \int_0^\infty e^{-\frac{n}{2(1-\rho^2)}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{2\rho r\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_2^2}{\sigma_2^2}\right)} d\mu_1\, d\mu_2 = \frac{\theta}{\sin\theta}$$

and hence differentiating $(n-2)$ times with respect to r, the required expression is obtained.

5. The form which we have now obtained may be applied without difficulty to all small even values of n, and in such cases is peculiarly suitable for the calculation of moments.

When $n = 2$ the ordinate of the curve, with abscissa r, is

$$\frac{\theta}{(1-r^2)\sin\theta},$$

which becomes hyperbolic in the neighbourhoods of -1 and $+1$. The value

---

* *Phil. Trans.* Vol. 192, A, p. 141.

of r is, therefore, as we know, either -1 or $+1$, and the proportion, in which these occur, depends upon ρ. The ratio of the infinite areas included with the asymptotes of the above curve is

$$\frac{\cos^{-1}\rho}{\cos^{-1}(-\rho)},$$

so that the mean value of a number of observations is $\dfrac{\sin^{-1}\rho}{\frac{\pi}{2}}$.

When $n = 4$ there is still no approach to normality, the curve takes the form

$$\frac{1}{\sin^3\theta}(\theta - 3\cot\theta + 3\theta\cot^2\theta),$$

which, when r is positive, increases regularly from its value of $\frac{4}{15}$ when $\theta = 0$, to infinity, to which it approaches as θ approaches π. Unless ρ is actually equal to 1, in which case r is also 1 of necessity, the curve has finite ordinates at both extremes. For calculating the number of values which should fall within any given range, the integral, $\dfrac{1}{\sin^2\theta}(1 - \theta\cot\theta)$, may be directly tabulated, as has been done in forming the accompanying table of "Student's" observations, and the corresponding expectations. The values given by Mr Soper's formula are apposed for comparison.

Table for comparison with p. 114, *Biometrika, Vol. IX.*

r	Calculated frequency m	Observed	Difference e	$\frac{e^2}{m}$	H.E. Soper's approximation	Difference e	$\frac{e^2}{m}$
·905—1	202·1	175·5	−15·0	·69	230·3	−17·2	·90
·805—·905	124·9	136·5			98·9		
·705—·805	88·7	84	− 3·8	·09	72·1	+20·3	3·18
·605—	65·1	66			57·6		
·505—	49·9	55	+12·3	1·73	48·0	+11·8	1·58
·405—	37·8	45			40·2		
·305—	30·6	24·5	− 6·4	·74	34·3	−15·0	3·52
·205—	24·8	24·5			29·7		
·105—	20·5	19	−11·6	3·58	25·6	−21·6	9·80
·005—	17·1	7			22·0		
$\bar{1}$·905—	14·5	22	+ 7·1	1·87	18·8	− ·8	0·02
$\bar{1}$·805—	12·4	12			16·0		
$\bar{1}$·705—	10·7	13	− 4·0	·80	13·5	− 8·7	3·06
$\bar{1}$·605—	9·3	3			11·2		
$\bar{1}$·505—	8·1	12	+12·7	10·54	9·0	+12·1	9·21
$\bar{1}$·405—	7·2	16			6·9		
$\bar{1}$·305—	6·3	7	+ 5·1	2·19	5·1	+ 8·6	8·80
$\bar{1}$·205—	5·6	10			3·3		
$\bar{1}$·105—	5·1	4	+ 3·6	1·38	1·9	+10·5	44·10
$\bar{1}$—$\bar{1}$·105	4·3	9			·6		
—	—	745	—	23·61	—	—	84·17

6. The direct process of integration by parts applied to such expressions as

$$\int_{-1}^{+1} (1-r^2)^{\frac{n-4}{2}} \frac{\partial^{n-1}}{\partial r^{n-1}} \frac{\theta^2}{2}\, dr \quad \text{and} \quad \int_{-1}^{+1} (1-r^2)^{\frac{n-4}{2}}\, r \frac{\partial^{n-1}}{\partial r^{n-1}} \frac{\theta^2}{2}\, dr,$$

when n is even, merely introduces the sums and differences of the terms $\frac{\partial^p}{\partial r^p} \frac{\theta^2}{2}$ at the extremes, where r is -1 or $+1$, with coefficients which are, in any particular case, easily calculable.

Thus, n being 6,

$$\int_{-1}^{+1} (1-r^2) \frac{\partial^5}{\partial r^5} \frac{\theta^2}{2}\, dr = \left[(1-r^2)\frac{\partial^4}{\partial r^4}\frac{\theta^2}{2}\right]_{-1}^{+1} + \left[2r \frac{\partial^3}{\partial r^3}\frac{\theta^2}{2}\right]_{-1}^{+1} - \left[2\frac{\partial^2}{\partial r^2}\frac{\theta^2}{2}\right]_{-1}^{+1}$$

$$= 2 \times \text{the sum of the extreme values of } \frac{\rho^3}{\sin^3\theta}(\theta - 3\cot\theta + 3\theta\cot^2\theta)$$

$$- 2 \times \text{the difference of the extreme values of } \frac{\rho^2}{\sin^2\theta}(1-\theta\cot\theta).$$

If $\rho = \sin\alpha$, so that the extreme values of θ are $\frac{\pi}{2} - \alpha$ and $\frac{\pi}{2} + \alpha$, the sums and differences may readily be expressed in terms of α, and the first few may here be tabulated: the table has been carried back as far as is necessary for the calculation of the fourth moment.

	sum	difference
$\frac{\sin^2\theta}{4\rho^2}\left\{\frac{7+2\theta^2}{4} - 3\theta\cot\theta - \frac{7-6\theta^2}{4}\cot^2\theta\right\}$	—	$\frac{\pi\cot^2\theta}{4}(a + 3\tan a + 3a\tan^2 a)$
$\frac{\sin\theta}{\rho}\left\{\theta + \left(1 - \frac{\theta^2}{2}\right)\cot\theta\right\}$	$\pi\cot a\,(1 + a\tan a)$	$\cot a\left\{2a - 2\tan a + \left(\frac{\pi^2}{4} + a^2\right)\tan a\right\}$
$\frac{\theta^2}{2}$	$\frac{\pi^2}{4} + a^2$	πa
$\frac{\rho}{\sin\theta} \cdot \theta$	$\pi\tan a$	$2a\tan a$
$\frac{\rho^2}{\sin^2\theta}(1 - \theta\cot\theta)$	$2\tan^2 a\,(1 + a\tan a)$	$\pi\tan^3 a$
$\frac{\rho^3}{\sin^3 a}(\theta - 3\cot\theta + 3\theta\cot^2\theta)$	$\pi\tan^2 a\,(1 + 3\tan^2 a)$	$2\tan^3 a\,(a + 3\tan a + 3a\tan^2 a)$
$\frac{\rho^4}{\sin^4 a}(4 - 9\theta\cot\theta + 15\cot^2\theta - 15\theta\cot^3\theta)$	$2\tan^4 a(4 + 9a\tan a + 15\tan^2 a + 15a\tan^3 a)$	$\pi\tan^4 a\,(9\tan a + 15\tan^3 a)$

There are here two natural series, which appear alternately as sums and differences; the simpler, which may be expressed in the form

$$\frac{\pi}{2}\sin^p\alpha\left(\frac{\partial}{\cos\alpha\,\partial\alpha}\right)^p \alpha,$$

is essentially a series of Legendre functions of the first kind; and may be expressed as

$$\frac{\pi}{2}\cdot\tan^p\alpha\,\frac{\underline{|p-1}}{i^{p-1}}\,P_{p-1}(i\tan\alpha);$$

and it is these only which occur in the evaluation of the even moments.

7. It is, however, desirable to obtain general expressions for these integrals in terms of n and ρ, and to evaluate them when n is odd.

For this purpose let us introduce a quantity ϕ, such that

$$\cos\phi=\cos\theta-k,$$

then, when k is sufficiently small, we may expand ϕ^2 by Taylor's theorem, so that

$$\frac{\phi^2}{2}=\frac{\theta^2}{2}+k\frac{\partial}{\sin\theta\partial\theta}\frac{\theta^2}{2}+\frac{k^2}{\underline{|2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2}+\dots.$$

Now let

$$k=\rho h\sqrt{1-r^2},$$

then

$$\frac{\phi^2}{2}=\frac{\theta^2}{2}+\rho h\sqrt{1-r^2}\,\frac{\partial}{\sin\theta\partial\theta}\frac{\theta^2}{2}+\frac{\rho^2h^2(1-r^2)}{\underline{|2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2}+\dots,$$

and differentiating twice with respect to h

$$\rho^2(1-r^2)\left(\frac{\partial}{\sin\phi\partial\phi}\right)^2\frac{\phi^2}{2}=\rho^2(1-r^2)\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2}+h\rho^3(1-r^2)^{\frac{3}{2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^3\frac{\theta^2}{2}+\dots,$$

whence, dividing by $(1-r^2)^{\frac{3}{2}}$, we obtain

$$\frac{\rho^2}{\sqrt{1-r^2}}\left(\frac{\partial}{\sin\phi\partial\phi}\right)^2\frac{\phi^2}{2}=\frac{\rho^2}{(1-r^2)^{\frac{1}{2}}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2}+h\rho^3\left(\frac{\partial}{\sin\theta\partial\theta}\right)^3\frac{\theta^2}{2}$$
$$+\frac{\rho^4h^2}{\underline{|2}}(1-r^2)^{\frac{1}{2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^4\frac{\theta^2}{2}+\dots,$$

so that

$$\int_{-1}^{+1}r^p(1-r^2)^{\frac{n-4}{2}}\frac{\partial^{n-1}}{\partial r^{n-1}}\frac{\theta^2}{2}\,dr$$

may be obtained by multiplying by $\underline{|n-3}$ the coefficient of h^{n-3} in

$$\rho^2\int_{-1}^{+1}\frac{r^p\,dr}{\sqrt{1-r^2}}\cdot\frac{1-\phi\cot\phi}{\sin^2\phi},$$

when $\cos\phi=\cos\theta-\rho h\sqrt{1-r^2}=-\rho(r+h\sqrt{1-r^2})$.

Our object might equally be achieved by the evaluation of the integral

$$\rho\int_{-1}^{+1}\frac{r^p\,dr}{1-r^2}\left(\frac{\phi}{\sin\phi}-\frac{\theta}{\sin\theta}\right).$$

The quantity ϕ is determined by the equation

$$\cos\phi=\cos\theta-\rho h\sqrt{1-r^2},$$

that is

$$\cos\phi=-\rho(r+h\sqrt{1-r^2}).$$

If now
$$r = \sin\beta,$$
$$h = \tan\epsilon,$$
then
$$\cos\theta = -\rho\sin\beta,$$
$$\cos\phi = -\rho\sqrt{1+h^2}\sin(\beta+\epsilon) = -\rho\sqrt{1+h^2}\sin\beta',$$
and as r passes from -1 to $+1$,

β passes from $-\frac{\pi}{2}$ to $+\frac{\pi}{2}$,

θ from $\frac{\pi}{2}-\alpha$ to $\frac{\pi}{2}+\alpha$,

β' from $-\frac{\pi}{2}+\epsilon$ to $\frac{\pi}{2}$ and thence to $\frac{\pi}{2}+\epsilon$,

and ϕ from $\frac{\pi}{2}-\alpha$ to $\frac{\pi}{2}+\alpha'$ and thence back to $\frac{\pi}{2}+\alpha$,

where $\sin\alpha' = \rho\sqrt{1+h^2}$, ϕ oscillates in the same manner as θ, with a somewhat greater amplitude, and slightly in advance in respect of phase.

The expression
$$\rho^2\int_{-1}^{+1}\frac{1-\phi\cot\phi}{\sin^2\phi}\,\frac{dr}{\sqrt{1-r^2}}$$
may now be reduced to
$$\rho^2\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}}\frac{1-\phi\cot\phi}{\sin^2\phi}\,d\beta = \rho^2\int_{-\frac{\pi}{2}+\epsilon}^{\frac{\pi}{2}+\epsilon}\left(\frac{1}{1-\sin^2\alpha'\sin^2\beta'}+\frac{\phi\sin\alpha'\sin\beta'}{(1-\sin^2\alpha'\sin^2\beta')^{\frac{3}{2}}}\right)d\beta'$$
$$=\rho^2\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}}\frac{d\beta'}{1-\sin^2\alpha'\sin^2\beta'}+\pi\rho^2\int_{+\frac{\pi}{2}}^{+\frac{\pi}{2}+\epsilon}\frac{\sin\alpha'\sin\beta'\,d\beta'}{(1-\sin^2\alpha'\sin^2\beta')^{\frac{3}{2}}}$$
$$+\rho^2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{(\phi)\sin\alpha'\sin\beta'\,d\beta'}{(1-\sin^2\alpha'\sin^2\beta')^{\frac{3}{2}}}$$
$$=\frac{\rho^2\pi}{\cos\alpha'}+\frac{\pi\rho^2\sin\alpha'}{\cos^2\alpha'}\left(\frac{\sin\epsilon}{\cos\alpha}\right)+\frac{\pi\rho^2}{\cos^2\alpha'}(1-\cos\alpha')$$
$$=\frac{\rho^2\pi}{\cos^2\alpha'}\left(1+\frac{\sin\alpha\tan\epsilon}{\cos\alpha}\right),$$
but
$$\cos^2\alpha' = 1-\rho^2(1+h^2) = \cos^2\alpha-\sin^2\alpha\tan^2\epsilon,$$
so that
$$\rho^2\int_{-1}^{+1}\frac{1-\phi\cot\phi}{\sin^2\phi}\,\frac{dr}{\sqrt{1-r^2}} = \frac{\pi\tan^2\alpha}{1-h\tan\alpha}.$$

From this evaluation we deduce the general form
$$\int_{-1}^{+1}(1-r^2)^{\frac{n-4}{2}}\frac{\partial^{n-1}}{\partial r^{n-1}}\frac{\theta^2}{2}\,dr = \underline{|n-3}\,\pi\tan^{n-1}\alpha \;\ldots\ldots\ldots\ldots\ldots\text{(III)}.$$

The absolute frequency df, with which r falls in the range dr, is therefore

$$\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\,\underline{|n-3}}(1-r^2)^{\frac{n-4}{2}}\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-2}\frac{\theta}{\sin\theta}\,dr.$$

8. I do not see how to integrate the other expressions of the type

$$\rho^2\int_{-1}^{+1}\frac{1-\phi\cot\phi}{\sin^2\phi}\,\frac{r^p\,dr}{\sqrt{1-r^2}},$$

although a form could probably be obtained when p is even. The general expression for the second moment may, however, be deduced by means of a reduction formula.

By a process of integration by parts it appears that, if we write

$$\int_{-1}^{+1}(1-r^2)^{\frac{n-4}{2}}\,r^p\,\frac{\partial^{n-1}}{\partial r^{n-1}}\,\frac{\theta^2}{2}\,dr=I_{n.p},$$

then
$$I_{n+2.2}=I_{n+2.0}+nI_{n.0}-n(n-1)I_{n.2},$$

and since
$$I_{4.2}=2\pi\left(\frac{\tan^3\alpha}{2}-\tan\alpha+\alpha\right),$$

we may obtain successively

$$I_{6.2}=24\pi\left(\frac{\tan^5\alpha}{4}-\frac{\tan^3\alpha}{3}+\tan\alpha-\alpha\right),$$

$$I_{8.2}=720\pi\left(\frac{\tan^7\alpha}{6}-\frac{\tan^5\alpha}{5}+\frac{\tan^3\alpha}{3}-\tan\alpha+\alpha\right),$$

and so on, yielding, when n is even, the expression

$$I_{n.2}=I_{n.0}-\pi\,\underline{|n-2}\int_0^\alpha\tan^{n-2}x\,dx,$$

a form which may well hold when n is odd.

The above expressions are useful in tabulating the numerical values of the second moment, $\bar{r}^2+\sigma^2$, of the unit curve, which may easily be calculated in succession for different values of n when $\tan^2\alpha$ is taken to have some simple value.

9. Before leaving this aspect of the subject it is worth while to give a more detailed examination of the mean of the frequency curves of r when $n=4$.

Two formulae are arrived at by Mr Soper, which are equivalent approximations of the second degree

I. $$\bar{r}=\rho\left[1-\frac{1-\rho^2}{2n}\left\{1+\frac{3}{4n}(1+3\rho^2)\right\}\right]=\rho\left[1-\frac{1-\rho^2}{8}\left\{1+\frac{3}{16}(1+3\rho^2)\right\}\right],$$

II. $$\bar{r}=\rho\left[1-\frac{1-\rho^2}{2(n-1)}\left\{1-\frac{1}{4(n-1)}(1-9\rho^2)\right\}\right]=\rho\left[1-\frac{1-\rho^2}{6}\left\{1-\frac{1}{12}(1-9\rho^2)\right\}\right],$$

and these we shall compare with the form

$$\text{III.} \qquad \bar{r} = \frac{2}{\pi}(\alpha + \cot\alpha - \alpha\cot^2\alpha),$$

ρ	·1000	·2000	·3000	·4000	·5000	·6000	·7000	·8000	·9000	·9500
I	·0853	·1710	·2578	·3463	·4377	·5333	·6347	·7443	·8649	·9304
II	·0847	·1697	·2555	·3419	·4310	·5241	·6236	·7330	·8566	·9254
III	·0850	·1704	·2570	·3451	·4360	·5301	·6290	·7357	·8540	·9209

It will be observed that the approximations lie on either side of the exact value over the greater part of the range, and that the error of the first approximation increases up to the value when $\rho = \cdot 9$. The second formula gives the correct value somewhere between ·8 and ·9, and is thereafter too large.

For the particular case	$\rho = \cdot 6608$,
I find (formula III)	$\bar{r} = \cdot 5897$, nearly the maximum difference from ρ,
Mr Soper gives (p. 109) the value	·5933
and the experimental data	·5609.

The two theoretical values are much nearer to each other than either is to the experimental value. On the whole, it is obvious that even in this unfavourable case Mr Soper's formulae possess remarkable accuracy.

10. The use of the correlation coefficient r as independent variable of these frequency curves is in some respects highly unsatisfactory. For high values of r the curve becomes extremely distorted and cramped, and although this very cramping forces the mean $\bar{r}$ to approach ρ, the difference compared with $1-\rho$ becomes inordinately great. Even for high values of n, the distortion in this region becomes extreme, and since at the same time the curve rapidly changes its shape, the values of the mean and standard deviation cease to have any very useful meaning. It would appear essential in order to draw just conclusions from an observed high value of the correlation coefficient, say ·99, that the frequency curves should be reasonably constant in form.

The previous paragraphs suggest that more natural variables for the treatment of our formulae are afforded by the transformations

$$t = \tan\beta = \frac{r}{\sqrt{1-r^2}},$$

$$\tau = \tan\alpha = \frac{\rho}{\sqrt{1-\rho^2}}.$$

The expression for the frequency curve (II)

$$(1-r^2)^{\frac{n-4}{2}} \left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1} \frac{\theta^2}{2}\,dr$$

now becomes

$$\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\,\frac{dt}{(1+t^2)^{\frac{n-1}{2}}}$$

and the range of the curve is extended from $-\infty$ to $+\infty$.

It is interesting that in the important case, $r=0$, the frequency reduces to $\dfrac{dt}{(1+t^2)^{\frac{n-1}{2}}}$ and the curves are identical with those found by "Student" for z, the probability integral of which he has tabulated in his first paper.

11. The moments of these curves are obtained by the evaluation of the expressions

$$\int_{-\infty}^{\infty}\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\,\frac{dt}{(1+t^2)^{\frac{n-1}{2}}},\quad \int_{-\infty}^{\infty}\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\,\frac{t\,dt}{(1+t^2)^{\frac{n-1}{2}}},$$

and so on; of these the first is known already (III) to have the value

$$\frac{\pi\,\lfloor\underline{n-3}}{(1-\rho^2)^{\frac{n-1}{2}}}$$

and the others may be obtained in succession, for

$$I_{n.p}=\int_{-\infty}^{\infty}\frac{\partial^{n-1}}{(\sin\theta\,\partial\theta)^{n-1}}\,\frac{\theta^2}{2}\,\frac{t^p\,dt}{(1+t^2)^{\frac{n-1}{2}}}=\frac{\partial^{n-1}}{\partial\rho^{n-1}}\int_{-\infty}^{\infty}\frac{1}{r^{n-1}}\,\frac{\theta^2}{2}\,\frac{t^p\,dt}{(1+t^2)^{\frac{n-1}{2}}}$$

$$=\frac{\partial^{n-1}}{\partial\rho^{n-1}}\int_{-\infty}^{\infty}\frac{\theta^2}{2}\cdot\frac{dt}{t^{n-1-p}}=\frac{\partial^p}{\partial\rho^p}\,I_{n-p.0},$$

so that the first moment

$$\int_{-\infty}^{\infty}\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\cdot\frac{t\,dt}{(1+t^2)^{\frac{n-1}{2}}}=\frac{\partial}{\partial\rho}\cdot\frac{\pi\,\lfloor\underline{n-4}}{(1-\rho^2)^{\frac{n-2}{2}}}=\frac{\pi\,\lfloor\underline{n-4}\,(n-2)\,\rho}{(1-\rho^2)^{\frac{n-4}{2}}};$$

hence

$$\bar{t}=\frac{n-2}{n-3}\,\frac{\rho}{\sqrt{1-\rho^2}}=\frac{n-2}{n-3}\,\tau.$$

The mean, therefore, is greater than the true value τ by a constant fraction of its value. And this fraction decreases in the simplest possible manner as n increases.

In the same way, we may evaluate the second moment,

$$\bar{t}^2+\sigma^2=\frac{1}{n-4}\,\{1+(n-1)\,\tau^2\}$$

and

$$\sigma^2=\frac{1}{n-4}\left\{1+\tau^2+\frac{(n-2)}{(n-3)^2}\,\tau^2\right\};$$

the third moment

$$\sqrt{\beta_1}\,\sigma^3=\frac{(n-2)\,\tau}{(n-3)(n-4)(n-5)}\left\{3\,(1+\tau^2)+\frac{2\tau^2(n-1)}{(n-3)^2}\right\},$$

and the fourth moment

$$\beta_2\sigma^4 = \frac{3}{(n-4)(n-6)}\left\{(1+\tau^2)^2 + \frac{6(n-2)\tau^2}{(n-3)(n-5)}(1+\tau^2) + \frac{6(n-2)(3n^2-11n+12)\tau^4}{(n-3)^4(n-5)}\right\}.$$

For high values of n, all but the first terms tend to vanish; β_1 tends to vary as ρ^2, and β_2 tends to become independent of ρ. In effect for high values of τ, where ρ^2 is nearly equal to unity, the form of the curve is nearly constant, but the skewness measured by β_1 decreases to zero at the origin, and changes its sense, when τ and ρ change their sign.

Tables are appended for inspection rather than for reference which show the nature and extent of these changes in the form of the curves.

Table of σ^2.

$\tau^2 =$	*·01*	*·03*	*·10*	*·30*	*1·00*	*3·00*	*10·00*	*30·00*	*100·00*
$n =$									
8	·2531	·2593	·2810	·3430	·5600	1·140	3·350	9·550	31·250
13	·1123	·1148	·1234	·1481	·2344	·4811	1·344	3·811	12·444
18	·07219	·07372	·07908	·09438	·1479	·3010	·8365	2·367	7·722
23	·05319	·05429	·05817	·06925	·1080	·2188	·6066	1·714	5·592
33	·03484	·03555	·03805	·04518	·7015	·1415	·3912	1·105	3·602
43	·02590	·02643	·02827	·03353	·05194	·1045	·2886	·8146	2·655
53	·02062	·02103	·02249	·02666	·04123	·08288	·2287	·6451	2·103

Table of β_1.

$\tau^2 =$	*·01*	*·03*	*·10*	*·30*	*1·00*	*3·00*	*10·00*	*30·00*	*100·00*	∞
$n =$										
8	·05685	·1662	·5076	1·230	2·450	3·788	3·965	4·153	4·184	4·252
13	·01517	·04776	·1376	·3400	·7058	1·018	1·205	1·271	1·296	1·3065
18	·008399	·02463	·07645	·1914	·4016	·5857	·6990	·7395	·7546	·7619
23	·005757	·01691	·05247	·1317	·3016	·4093	·4910	·5208	·5314	·5361
33	·003518	·01035	·03214	·08100	·1731	·2559	·3031	·3260	·3334	·3366
43	·002530	·007435	·02315	·05841	·1251	·1858	·2237	·2376	·2429	·2452
53	·001973	·005798	·01807	·04562	·09800	·1458	·1757	·1868	·1910	·1928

Table of β_2.

$\tau^2 =$	*00*	*·01*	*·03*	*·10*	*·30*	*1·00*	*3·00*	*10·00*	*30·00*	*100·00*	∞
$n =$											
8	6·0000	6·1137	6·3179	7·0179	8·4767	10·9668	12·9652	14·1116	14·5024	14·6508	14·7159
13	3·8571	3·8802	3·9248	4·0663	4·3770	4·9397	5·4240	5·7147	5·8186	5·8578	5·8750
18	3·5000	3·5121	3·5356	3·6104	3·7937	4·0828	4·3532	4·5186	4·5783	4·6009	4·6109
23	3·3529	3·3612	3·3768	3·4271	3·5556	3·7486	3·9356	4·0511	4·0930	4·1089	4·1159
33	3·2222	3·2271	3·2365	3·2667	3·3343	3·4619	3·5773	3·6493	3·6756	3·6856	3·6899
43	3·1622	3·1656	3·1723	3·1938	3·2422	3·3261	3·4172	3·4692	3·4886	3·4958	3·4991
53	3·1277	3·1303	3·1356	3·1522	3·1898	3·2640	3·3281	3·3676	3·3826	3·3883	3·3909

12. The fact that the mean value $\bar{r}$ of the observed correlation coefficient is numerically less than ρ might have been interpreted as meaning that given a single observed value r, the true value of the correlation coefficient of the population from which the sample is drawn is likely to be greater than r. This reasoning is altogether fallacious. The mean $\bar{r}$ is not an intrinsic feature of the frequency distribution. It depends upon the choice of the particular variable r in terms of which the frequency distribution is represented. When we use t as variable, the situation is reversed. Whereas in using r we cramp all the high values of the correlation into the small space in the neighbourhood of $r=1$, producing a frequency curve which trails out in the negative direction and so tending to reduce the value of the mean, by using t, we spread out the region of high values, producing asymmetry in the opposite sense, and obtain a value $\bar{t}$ which is greater than τ. The mean might, in fact, be brought to any chosen point, by stretching and compressing different parts of the scale in the required manner. For the interpretation of a single observation the relation between $\bar{t}$ and τ is in no way superior to that between $\bar{r}$ and ρ. The variable t has been chosen primarily in order to give stability of form to the frequency curves in different parts of the scale. It is in addition a variable to which the analysis naturally leads us, and which enables the mean and moments to be readily calculated, and so a comparison to be made with the standard Pearson curves, but it is not, with these advantages, in a unique position. In some respects the function, $\log\tan\frac{1}{2}\left(\alpha+\frac{\pi}{2}\right)$, is its superior as independent variable.

I have given elsewhere* a criterion, independent of scaling, suitable for obtaining the relation between an observed correlation of a sample and the most probable value of the correlation of the whole population. Since the chance of any observation falling in the range dr is proportional to

$$(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\,dr$$

for variations of ρ, we must find that value of ρ for which this quantity is a maximum, and thereby obtain the equation

$$\frac{\partial}{\partial\rho}\left\{(1-\rho^2)^{\frac{n-1}{2}}\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\right\}=0.$$

Since
$$\int_0^\infty\frac{dx}{(\cosh x+\cos\theta)^{n-1}}=\frac{1}{\underline{|n-1}}\left(\frac{\partial}{\sin\theta\,\partial\theta}\right)^{n-1}\frac{\theta^2}{2}$$

we have
$$\int_0^\infty\frac{\partial}{\partial\rho}\left\{(1-\rho^2)^{\frac{n-1}{2}}\frac{dx}{(\cosh x+\cos\theta)^{n-1}}\right\}=0,$$

---

* R. A. Fisher, "On an absolute criterion for fitting frequency curves," *Messenger of Mathematics*, February, 1912.

which leads by a process of simplification to the equation

$$\int_0^\infty \frac{dx}{(\cosh x - \rho r)^n}(r - \rho \cosh x) = 0.$$

Since $\cosh x$ is always greater than ρr, the factor in the numerator, $r - \rho \cosh x$, must change sign in the range of integration. We therefore see that r is greater than ρ. Further an approximate solution may be obtained for large values of n. The integrand is negligible save when x is very small, and we may write

$$1 + \frac{x^2}{2} \text{ for } \cosh x$$

and

$$(1 - \rho r)^n e^{\frac{nx^2}{2(1-\rho r)}} \text{ for } (\cosh x - \rho r)^n.$$

Then

$$r\int_0^\infty e^{-\frac{nx^2}{2(1-\rho r)}} dx = \rho \int_0^\infty \left(1 + \frac{x^2}{2}\right) e^{-\frac{nx^2}{2(1-\rho r)}} dx,$$

and in consequence, as a first approximation,

$$r = \rho\left(1 + \frac{1 - r^2}{2n}\right).$$

The corresponding relation between t and τ is evidently

$$t = \tau\left(1 + \frac{1}{2n}\right).$$

It is now apparent that the most likely value of the correlation will in general be less than that observed, but the difference will be only half of that suggested by the mean, $\bar{t}$.

It might plausibly be urged that in the choice of an independent variable we should aim at making the relation between the mean and the true value approach the above equation, or rather that to which the above is an approximation, or that we should aim at reducing the asymmetry of the curves, or at approximate constancy of the standard deviation. In these respects the function

$$\log \tan \tfrac{1}{2}\left(\alpha + \frac{\pi}{2}\right) \text{ that is, } \tanh^{-1}\rho$$

is not a little attractive, but so far as I have examined it, it does not tend to simplify the analysis, and approaches relative constancy at the expense of the constancy proportionate to the variable, which the expressions in τ exhibit*.

---

* [It may be worth noting that Mr Fisher's t is the ϕ-square root mean square contingency—of the more usual notation, and is the expression used in determining the probability that correlated material has been obtained by random sampling from uncorrelated material. Ed.]

ON THE DISTRIBUTION OF THE STANDARD DEVIATIONS OF SMALL SAMPLES: APPENDIX I. TO PAPERS BY "STUDENT" AND R. A. FISHER.

(EDITORIAL.)

CONSIDER the population distributed according to the law

$$y = \frac{N}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \quad \text{.................................(i)},$$

and let a sample of n represented by the variate values $x_1, x_2 \ldots x_n$ be taken from it. Then the probability δP that this sample will lie between

$$x_1 \text{ and } x_1 + \delta x_1,\ x_2 \text{ and } x_2 + \delta x_2 \ldots x_n \text{ and } x_n + \delta x_n$$

is

$$\delta P = \frac{N^n}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2}\frac{S(x_s-m)^2}{\sigma^2}} \delta x_1 \delta x_2 \ldots \delta x_n$$

$$= \text{const.} \times e^{-\frac{1}{2}\frac{S(x_s-\bar{x})^2}{\sigma^2} - \frac{1}{2}\frac{n(\bar{x}-m)^2}{\sigma^2}} \delta x_1 \delta x_2 \ldots \delta x_n \quad \text{...........(ii)},$$

where $\bar{x} = \frac{1}{n} S(x_s)$. If $\Sigma^2 = \frac{1}{n} S(x_s - \bar{x})^2$ we may write:

$$\delta P = \text{const.} \times e^{-\frac{1}{2}\left(\frac{n\Sigma^2}{\sigma^2} + \frac{n(\bar{x}-m)^2}{\sigma^2}\right)} \delta x_1 \delta x_2 \ldots \delta x_n \quad \text{..............(iii)}.$$

Changing as Mr Fisher does (see p. 510 above) to $\bar{x}$ and Σ as coordinates we have:

$$\delta P = \text{const.} \times e^{-\frac{1}{2}\left(\frac{n\Sigma^2}{\sigma^2} + \frac{n(\bar{x}-m)^2}{\sigma^2}\right)} \Sigma^{n-2} \delta\bar{x}\, \delta\Sigma.$$

We see at once from this* that the law of distribution of samples of means is the normal curve

$$y = y_0 e^{-\frac{1}{2}\frac{n(\bar{x}-m)^2}{\sigma^2}} \quad \text{.................................(iv)}$$

---

* Of course the form reached above shows that for normal distributions there is no correlation between deviations in the mean and in the standard deviation of samples, a familiar fact.

with mean $\bar{x} = m$, the mean of the population, and with standard deviation $= \sigma/\sqrt{n}$, a well-known result.

On the other hand the distribution of samples of standard deviations is

$$y = y_0 \Sigma^{n-2} e^{-\frac{1}{2}\frac{n\Sigma^2}{\sigma^2}} \text{ ....(v).}$$

This curve was first reached by "Student" as a highly probable result following from the relations he had obtained from the moments of Σ^2*. Mr Fisher's work thus enables us to justify "Student's" assumption. "Student" has discussed at some length the distribution curve for Σ. He has obtained the values of the moment coefficients μ_2, μ_3 and μ_4 and the general expressions for the means when n is even and odd. The whole problem is of such importance that it seems worth reconsidering, and providing tables showing the approach of the distribution curve to normality as n rises from 4 to 100.

The following investigation largely repeats work given by "Student," but it expresses the values for μ_3, μ_4, and β_1 and β_2 in a different form†. We shall not use approximate expressions for the constants, for the order of terms in $1/n$ depends so largely on the relative magnitude of their coefficients, that such expressions become unreliable for values of n under 100.

Clearly (v) is a skew curve with range limited at one end, $\Sigma = 0$, and not at the other, $\Sigma = \infty$. See Figure p. 524.

We shall write the standard deviation of Σ, σ_Σ, and the moments of the frequency about the end of the range O as M_1', M_2', etc., while the moment-coefficients about Q will be as usual $\mu_1 (= 0)$, μ_2, etc. Obviously $\mu_2 = \sigma_\Sigma^2$. It is desirable to ascertain $\tilde{\Sigma}$, $\bar{\Sigma}$, σ_Σ and the skewness as well as β_1 and β_2 for the distribution. We do this to show the rapidity of change to a normal distribution. It is well, however, to notice *a priori* that for n large the distribution does become normal.

---

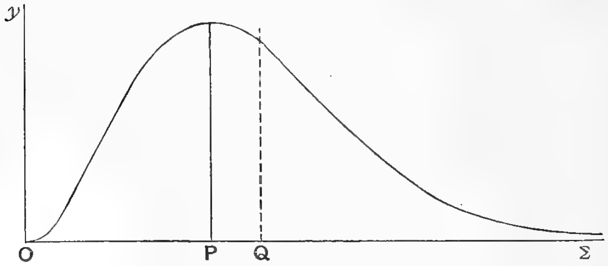
* "Student's" approximate values for β_1 and β_2 (*loc. cit.* p. 10) are, we fear, erroneous. He gives $D^2 = n - \frac{3}{2} + \frac{1}{8n}$, but it is needful to have a further term in $\frac{1}{n^2}$ in order to obtain β_1 and β_2 correctly to the *second* approximation in $\frac{1}{n}$. If this further term be p/n^2, then:

$$\beta_1 = \frac{1}{2n}\left(1 + \frac{64p - 3}{4n}\right), \text{ as against "Student's" } \frac{1}{2n}\left(1 - \frac{3}{4n}\right),$$

$$\beta_2 \quad 3 + \frac{16p}{n}, \quad \text{,, \quad ,, \quad ,,} \quad 3\left(1 - \frac{1}{4n^2}\right).$$

An examination of our table (p. 529) shows that "Student's" corrections are not of the right sign to agree with the facts, and that further no constant value of p would give good results even for fairly high values of n, i.e. it is probable that the term in $\frac{1}{n^3}$ in D^2 is of equal importance with that in $\frac{1}{n^2}$.

† "The Probable Error of a Mean," *Biometrika*, Vol. VI. pp. 1—25, more especially pp. 4, 6, and 8 to 10.



$OP = \text{mode} = \check{\Sigma}, \qquad OQ = \text{mean} = \bar{\Sigma}.$

To obtain this approximation to (v) let us assume $\Sigma = \check{\Sigma} + \epsilon$, and suppose ϵ small. Expanding $\log y$ we find:

$$\log y = \log y_0 + (n-2)\log \check{\Sigma} - \tfrac{1}{2}n(\check{\Sigma}/\sigma)^2 + \frac{(n-2)}{\check{\Sigma}}\left(1 - \frac{n}{n-2}\frac{\check{\Sigma}^2}{\sigma^2}\right)\epsilon$$
$$- \frac{n\epsilon^2}{2\sigma^2}\left(1 + \frac{n-2}{n}\frac{\sigma^2}{\check{\Sigma}^2}\right) + \text{terms in } \epsilon^3.$$

Hence since $\check{\Sigma}$ is at our choice we will take it so that

$$\check{\Sigma} = \sqrt{\frac{n-2}{n}}\,\sigma \qquad \text{(vi)}$$

and thus:

$$y = y_0 \check{\Sigma}^{n-2} e^{-\frac{1}{2}\frac{n\check{\Sigma}^2}{\sigma^2}} e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma^2/(2n)}} + \text{etc.}$$

Or, if ϵ be small compared with σ, the distribution is the normal curve:

$$y = y_0' e^{-\frac{1}{2}\frac{\epsilon^2}{\sigma^2/(2n)}} \qquad \text{(vii)},$$

with mean at $\check{\Sigma} = \sqrt{\frac{n-2}{n}}\,\sigma$ and standard deviation $\sigma/\sqrt{2n}$. If n as usual be considerable, this agrees with the ordinary result, i.e. $\bar{\Sigma} = \sigma$ and $\sigma_{\Sigma} = \sigma/\sqrt{2n}$, the distribution being treated as normal.

We will now deal with the full result (v). We have:

$$M_p' = \int_0^\infty y\Sigma^p\,d\Sigma = y_0\int_0^\infty \Sigma^{n+p-2} e^{-\frac{1}{2}\frac{n\Sigma^2}{\sigma^2}}\,d\Sigma \qquad \text{(viii)},$$

and clearly M_p' depends on a knowledge of

$$L_q = \int_0^\infty u^q e^{-\frac{1}{2}u^2}\,du \qquad \text{(ix)},$$

for we have:

$$M_p' = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n+p-1} L_{n+p-2}.$$

Integrating by parts we find:

$$L_q = (q-1)\,L_{q-2}$$
$$= (q-1)(q-3)\ldots\ldots \begin{cases} \ldots 3\,.\,1\,L_0, \text{ if } q \text{ be even} \\ \ldots 4\,.\,2\,L_1, \text{ if } q \text{ be odd} \end{cases} \quad \ldots\ldots(\text{x}).$$

Now $$L_0 = \int_0^\infty e^{-\frac{1}{2}u^2}du = \sqrt{\frac{\pi}{2}},$$

and $$L_1 = \int_0^\infty u e^{-\frac{1}{2}u^2}du = 1,$$

thus M_p' is determined, and will depend on whether $n+p$ be even or odd.

But $$M_2' = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n+1} L_n = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n+1}(n-1)\,L_{n-2},$$

$$M_0' = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n-1} L_{n-2}.$$

Hence $$\mu_2' = M_2'/M_0' = \sigma^2\,\frac{n-1}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{xi}).$$

To find the modal value $\tilde{\Sigma}$ we must differentiate (v), and we have

$$(n-2)\,\tilde{\Sigma}^{n-3}e^{-\frac{1}{2}\frac{n\tilde{\Sigma}^2}{\sigma^2}} - \frac{n\tilde{\Sigma}}{\sigma^2}\,\tilde{\Sigma}^{n-2}e^{-\frac{1}{2}\frac{n\tilde{\Sigma}^2}{\sigma^2}} = 0,$$

which gives

$$\tilde{\Sigma} = \sqrt{\frac{n-2}{n}}\,\sigma \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{xii}),$$

a result in agreement with the mean $\breve{\Sigma}$ of the approximate solution (vi), as we should anticipate.

It now remains to find M_1' and M_0' absolutely

$$M_1' = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n} L_{n-1},$$

$$M_0' = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n-1} L_{n-2}.$$

Suppose n even, then

$$\left.\begin{aligned} L_{n-1} &= (n-2)(n-4)\ldots 2\times 1 \\ L_{n-2} &= (n-3)(n-5)\ldots 1\times\sqrt{\frac{\pi}{2}} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots(\text{xiii}).$$

Hence for n even

$$\bar{\Sigma} = \mu_1' = M_1'/M_0' = \frac{\sigma}{\sqrt{n}}\,\frac{n-2}{n-3}\,\frac{n-4}{n-5}\ldots\frac{2}{1}\sqrt{\frac{2}{\pi}} \quad \ldots\ldots\ldots\ldots(\text{xiv}\,\textsc{a}).$$

Again for n odd

$$L_{n-1} = (n-2)(n-4)\ldots 1\times\sqrt{\frac{\pi}{2}},$$

$$L_{n-2} = (n-3)(n-5)\ldots 2\times 1,$$

and hence

$$\bar{\Sigma} = \frac{\sigma}{\sqrt{n}}\,\frac{(n-2)(n-4)\ldots 1}{(n-3)(n-5)\ldots 2}\sqrt{\frac{\pi}{2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(\text{xiv}\,\textsc{b}).$$

These accurate values of $\overline{\Sigma}$, the mean standard deviation of samples, were first given by "Student" (*loc. cit.* p. 8). Now by Wallis' Theorem

$$\sqrt{\frac{\pi}{2}(2n+1)} = \frac{\text{product of even numbers up to } 2n}{\text{product of odd numbers up to } 2n-1}.$$

Thus (xiv A) for n large tends to become

$$\overline{\Sigma} = \frac{\sigma}{\sqrt{n}}\sqrt{n-1} = \sigma\sqrt{\frac{n-1}{n}},$$

and (xiv B)

$$\overline{\Sigma} = \frac{\sigma}{\sqrt{n}}\,\frac{n-2}{\sqrt{n-2}} = \sigma\sqrt{\frac{n-2}{n}}.$$

These values, however, really only suffice to show the approach of $\overline{\Sigma}$ to σ, as they depend on the neglect of terms of the order $\frac{1}{n}$ as compared to 1, and we should get absurd results for $\sigma_{\Sigma}{}^2$ by subtracting the square of the above values of $\overline{\Sigma}$ from μ_2' in (xi). All they really tell us is that for n large $\overline{\Sigma} = \sigma$, but they give no true, approximation in $\frac{1}{n}$*.

If we use Stirling's Theorem up to the third term†, i.e.

$$x! = \sqrt{2\pi x}\, x^x e^{-x}\left(1 + \frac{1}{12x} + \frac{1}{288x^2}\right),$$

we obtain

$$\overline{\Sigma} = \sigma\left(1 - \frac{3}{4n} - \frac{7}{32n^2}\right) \quad \text{..........................(xv)},$$

$$\sigma_{\Sigma}{}^2 = \frac{\sigma^2}{2n}\left(1 - \frac{1}{4n}\right) \quad \text{..................................(xvi)},$$

$$\mu_3 = \frac{\sigma^3}{4n^2}, \qquad \mu_4 = \frac{3\sigma^4}{4n^2}.$$

But we should be compelled to introduce the term $-\frac{139}{51840x^3}$ into Stirling's expression to reach the second terms in μ_3 and μ_4. As we have indicated (p. 523, ftn.), such a term, even if used, will not lead to profitable results. It is better to work with the full formulae. It is desirable to find the full third and fourth moment coefficients in order to determine β_1 and β_2 and so measure as n increases the rapidity of approach to the normal curve.

---

* "Student" has used an extension of Wallis' Theorem, which will suffice for certain constants only.

† We can write (xiv A)

$$\overline{\Sigma} = \frac{\sigma}{\sqrt{n}}\,\frac{\left(2^{\frac{1}{2}n-1}\,\lfloor\underline{\tfrac{1}{2}n-1}\right)^2}{\lfloor\underline{n-2}}\sqrt{\frac{2}{\pi}}, \quad \text{......................................(xvii)},$$

and (xiv B)

$$\overline{\Sigma} = \frac{\sigma}{\sqrt{n}}\,\frac{(n-2)\,\lfloor\underline{n-3}}{\left(2^{\frac{1}{2}(n-3)}\,\lfloor\underline{\tfrac{1}{2}(n-3)}\right)^2}\sqrt{\frac{\pi}{2}}, \quad \text{..................................(xviii)},$$

and then apply Stirling's Theorem.

We have:

$$M_4' = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n+3} L_{n+2} = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n+3}(n+1)\,L_n$$

$$= \frac{\sigma^2}{n}(n+1)\,M_2'.$$

Hence $$\mu_4' = M_4'/M_0 = \frac{\sigma^2}{n}(n+1)\,\mu_2' = \frac{\sigma^4(n^2-1)}{n^2} \quad \text{.................(xix)},$$

$$M_3' = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n+2} L_{n+1} = y_0\left(\frac{\sigma}{\sqrt{n}}\right)^{n+2} nL_{n-1}$$

$$= \frac{\sigma^2}{n}\,nM_1'.$$

Hence $$\mu_3' = M_3'/M_0' = \sigma^2\mu_1' \quad \text{...........................(xx)}.$$

Transferring to mean:

$$\mu_3 = \mu_3' - 2\mu_2\mu_1' - \mu_2'\mu_1'$$

$$= \sigma^2\mu_1'\left(1 - 2\sigma_\Sigma^2/\sigma^2 - \frac{n-1}{n}\right)$$

$$= \frac{\sigma^2\bar{\Sigma}}{n}\left(1 - \frac{\sigma_\Sigma^2}{\sigma^2/2n}\right) \quad \text{.................................(xxi)}.$$

Thus μ_3 will grow small, not only owing to the factor $\frac{1}{n}$, but because σ_Σ^2 tends to equal $\sigma^2/2n$ as n increases.

Now $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\sigma^4\bar{\Sigma}^2}{n^2}\,\frac{\left(1 - \frac{\sigma_\Sigma^2}{\sigma^2/2n}\right)^2}{\sigma_\Sigma^6}.$$

Or $$\beta_1 = 8n\left(\frac{\bar{\Sigma}}{\sigma}\right)^2\left(1 - \frac{\sigma_\Sigma^2}{\sigma^2/2n}\right)^2\Big/\left(\frac{\sigma_\Sigma^2}{\sigma^2/2n}\right)^3 \text{...................(xxii)}.$$

Here $\bar{\Sigma}/\sigma$ is of the form $1 - \frac{\chi_1}{n}$, and

$$\sigma_\Sigma^2 = \frac{\sigma^2}{2n}\left(1 - \frac{\chi_2}{n}\right),$$

and thus β_1 tends as n increases to take the form $8\chi_2^2/n$, but as χ_2 may be a considerable numerical coefficient $8\chi_2^2$ may be commensurable with n till n is very considerable.

We next turn to μ_4, and shall endeavour to express it in terms of $\mu_2 = \sigma_\Sigma^2$.

Since $$\mu_2' = \mu_2 + \bar{\Sigma}^2 = \sigma^2\left(1 - \frac{1}{n}\right)$$

by (xi), we have

$$\bar{\Sigma}^2 = \sigma^2\left(1 - \frac{1}{n}\right) - \mu_2.$$

Further by (xx)

$$\mu_3'\mu_1' = \sigma^2\mu_1'^2 = \sigma^4\left(1-\frac{1}{n}\right) - \sigma^2\mu_2.$$

Thus

$$\begin{aligned}\mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4\\ &= \sigma^4\left\{1-\frac{1}{n^2} - 4\left(1-\frac{1}{n}\right) + 4\frac{\mu_2}{\sigma^2} + 6\left(1-\frac{1}{n}\right)\left(1-\frac{1}{n}-\frac{\mu_2}{\sigma^2}\right)\right.\\ &\qquad\qquad\left. - 3\left(1-\frac{1}{n}\right)^2 + \frac{6\mu_2}{\sigma^2}\left(1-\frac{1}{n}\right) - 3\frac{\mu_2^2}{\sigma^4}\right\}\\ &= \sigma^4\left\{\frac{5}{4n^2} - \frac{1}{2n}\left(4-\frac{3}{n}\right)\left(1-\frac{\sigma_\Sigma^2}{\sigma^2/2n}\right) - \frac{3}{4n^2}\left(1-\frac{\sigma_\Sigma^2}{\sigma^2/2n}\right)^2\right\}\ldots\ldots\ldots\end{aligned}$$

Hence

$$\beta_2 = \frac{1}{\{\sigma_\Sigma^2/(\sigma^2/2n)\}^2}\left\{5 - 2n\left(4-\frac{3}{n}\right)\left(1-\frac{\sigma_\Sigma^2}{\sigma^2/2n}\right) - 3\left(1-\frac{\sigma_\Sigma^2}{\sigma^2/2n}\right)^2\right\} \quad\ldots\text{(xxiii)}.$$

Our results for μ_3 and μ_4, although expressed in other notation, are in accordance with "Student's" (*loc. cit.* p. 9), so also are our results (xv) and (xvi) although reached by a different method of approximation. We do not agree with his approximate values for μ_3, μ_4 or β_1 and β_2.

The calculations to find $\overline{\Sigma}/\sigma$, $\sigma_\Sigma/(\sigma/\sqrt{2n})$, β_1 and β_2 presented some trouble. In order to be correct to the four figures of decimals in the tabled results, tables of ten-figure logarithms had to be used in the logarithmic part of the work. Formulae (xvii) and (xviii) of the ftn. p. 526 were adopted, using Degen's Tables of the Logarithms of Factorials. M_2' was calculated to nine figures, and even then, as n became large, the determination of the antilogarithms presented considerable difficulty. Further the powers of $1 - \sigma_\Sigma^2/(\sigma^2/2n)$ gave rise to trouble. The numerical work was undertaken by Ethel M. Elderton and Beatrice M. Cave, to whom very hearty thanks are due. We think the results may be depended on to the figures tabulated.

It will be seen that by the time $n = 50$ the mode is as close to the mean as we should expect to find in any random sample of normal material; the average mean $\overline{\Sigma}$ is only $1{\cdot}5\,{}^\circ\!/_\circ$ from the usually adopted value σ, and the average standard deviation σ_Σ only $0{\cdot}3\,{}^\circ\!/_\circ$ from its customary value $\sigma/\sqrt{2n}$. Further β_1 and β_2 are ·0105 and 3·0003 respectively, or for all practical purposes have reached their normal values. We think it must be concluded that for samples of 50 the usual theory of the probable error of the standard deviation holds satisfactorily, and that to apply it for the case of $n = 25$ would not lead to any error which would be of importance in the majority of statistical problems.

On the other hand, if a small sample, $n < 20$ say, of a population be taken, the value of the standard deviation found from it will be usually *less* than the standard deviation of the true population. If we take the most probable value, $\tilde{\Sigma}$, as that which has most likely been observed, then the result should be divided by the number in the column entitled mode $\tilde{\Sigma}/\sigma$ to obtain the most reasonable value for σ. For example, if Σ be observed, and $n = 20$, then the most reasonable value to give σ is $\Sigma/{\cdot}9487$.

The paper by Mr Fisher and the accompanying table more or less complete the work on the distribution of standard-deviations outlined by "Student" in 1908.

Table of Values of the Constants of the Frequency Distribution of the Standard Deviations of Samples drawn at random from a Normal Population.

Size of Sample n	Mode $\tilde{\Sigma}/\sigma$	Mean $\bar{\Sigma}/\sigma$	Standard Deviation		Measures of Deviation from Normality		
			σ_Σ/σ	$\sigma_\Sigma/(\sigma/\sqrt{2n})$	Skewness	β_1	β_2
4	·7071	·7979	·3367	·9524	·2696	·2359	3·1082
5	·7746	·8407	·3052	·9651	·2168	·1646	3·0593
6	·8165	·8686	·2808	·9725	·1857	·1255	3·0370
7	·8452	·8882	·2612	·9774	·1648	·1011	3·0251
8	·8660	·9027	·2452	·9808	·1495	·0845	3·0181
9	·8819	·9139	·2318	·9834	·1378	·0725	3·0136
10	·8944	·9227	·2203	·9853	·1285	·0634	3·0106
11	·9045	·9300	·2104	·9868	·1209	·0564	3·0085
12	·9129	·9359	·2017	·9881	·1144	·0507	3·0070
13	·9199	·9410	·1940	·9891	·1088	·0461	3·0059
14	·9258	·9453	·1871	·9900	·1041	·0422	3·0049
15	·9309	·9490	·1809	·9907	·0998	·0390	3·0042
16	·9354	·9523	·1752	·9914	·0961	·0362	3·0036
17	·9393	·9551	·1701	·9919	·0927	·0337	3·0032
18	·9428	·9576	·1654	·9924	·0897	·0316	3·0028
19	·9459	·9599	·1611	·9928	·0869	·0297	3·0025
20	·9487	·9619	·1570	·9932	·0844	·0281	3·0022
25	·9592	·9696	·1407	·9948	·0745	·0219	3·0014
30	·9661	·9748	·1285	·9956	·0674	·0180	3·0009
35	·9710	·9784	·1191	·9963	·0620	·0153	3·0007
40	·9747	·9811	·1114	·9967	·0577	·0132	3·0005
45	·9775	·9832	·1051	·9977	·0541	·0117	3·0004
50	·9798	·9849	·0997	·9974	·0512	·0105	3·0003
55	·9816	·9863	·0951	·9977	·0488	·0095	3·0003
60	·9832	·9874	·0911	·9979	·0467	·0087	3·0002
65	·9845	·9884	·0875	·9980	·0447	·0080	3·0002
70	·9856	·9892	·0844	·9982	·0430	·0074	3·0002
75	·9866	·9900	·0815	·9983	·0415	·0069	3·0001
80	·9874	·9906	·0789	·9984	·0402	·0064	3·0001
85	·9882	·9911	·0766	·9985	·0389	·0060	3·0001
90	·9888	·9916	·0744	·9986	·0378	·0057	3·0001
95	·9894	·9921	·0725	·9987	·0367	·0054	3·0001
100	·9899	·9925	·0706	·9987	·0358	·0051	3·0000

TUBERCULOSIS AND SEGREGATION.

BY ALICE LEE, D.SC.

(1) IN his book *The Prevention of Tuberculosis* (London: Methuen, no date on the issue we have used) Dr A. Newsholme has examined the influence of segregation on Tuberculosis. This is the topic of Chapter XXXV. In the opening of this chapter, he writes:

> The exact measure of institutional segregation of phthisis is the ratio stating how many of the total days of sickness (number of patients and number of days of sickness) are passed in institutions. This ratio and the equivalents for it which have to be used in practice may for convenience be called the *segregation ratio*. The need for equivalents for the ratio as stated above arises from the fact that we are dealing with actual recorded experience, and the material has to be taken from the records as they happen to exist. (p. 266.)

After noting the incompleteness of the records, Dr Newsholme continues:

> It becomes necessary therefore to select other figures which vary approximately with the total days of tuberculous sickness and the total days of tuberculous sickness passed in institutions. (p. 266.)

We shall discuss below what "indirect measures of segregation" Dr Newsholme selects, but he gives the following most proper caution with regard to them:

> In using these indirect measures of institutional treatment of tuberculosis and of its prevalence it must be remembered that they are indirect and approximate. Thus, for instance, figures for institutional treatment usually give the number of cases and not days of treatment, and while they tell how many people were segregated in institutions do not show the average duration, still less the quality of the treatment. Any of these indirect forms of segregation ratio has therefore to be verified wherever possible by the application to the same community and period of one or more other forms of the ratio, and checked wherever practicable by a special examination of sample constituent communities whose figures are included in the total. (p. 268.)

Dr Newsholme in the course of his chapter gives a number of very high correlations between the phthisis deathrate and the indirect forms of the segregation ratio he has selected, and he interprets these as well as a long series of graphs as demonstrating that institutional segregation has been a most important factor in the diminution of the phthisis deathrate. Now any two variates which are changing continuously with the time—say, the consumption of bananas per head of the population and the fall in the birthrate—will exhibit high correlation and will show graphically very high association, if plotted to appropriate scales and on a common time basis. Until the time factor has been removed, either by partial correlation or otherwise, it would be most unwise to interpret such cases as providing any causal relationship.

It seemed accordingly worth while to reinvestigate Dr Newsholme's problems with the aid of a rather more adequate statistical apparatus.

(2) We must frankly confess at the outset that we have had great difficulty in following Dr Newsholme's description of the methods he has adopted to measure the amount of segregation. His charts do not seem always in accordance with his tables, and both are occasionally out of agreement with his definitions. As he does not give the raw data on which his correlations are based, but only condensed versions of them in his tables and graphs, it is impossible to test his conclusions without returning to the original sources, which are not always stated, and when we have found them and our results differ, we are unable to say whether the difference is due to failure in his or in our arithmetic, or to divergences between his and our records.

Dr Newsholme uses in all some six measures of the segregation ratio, four intentionally and two apparently by inadvertence.

Let $P =$ total population of a given area, $\phi =$ the total number of annual deaths from phthisis. Then ϕ/P multiplied by 10,000 or 100,000, as the case may be, gives the crude deathrate from phthisis. Let D_i be the deaths from *all* causes which occur in institutions and D the total deaths in the same area, then $100D_i/D$ is Dr Newsholme's first approximation to the segregation ratio*. On p. 270 he gives two tables which show in (*a*) England and Wales as a whole, (*b*) in London, that, while in the course of forty years $1000\phi/P$ has practically halved, $100D_i/D$ has practically doubled. The data, Dr Newsholme tells us, show "not only a very close correspondence between the increase of total institutional segregation measured by the ratio in question and the decrease of phthisis, but an even more striking similarity in the ratio at which these changes have occurred" (p. 271). This is illustrated by a graph on p. 271, in which the *logarithms* of the phthisis deathrate are plotted to time against the logarithms of the indices of institutional deaths to all deaths†. We do not know why Dr Newsholme has chosen this method of representation; it certainly, with his choice of scales, makes the two curves roughly parallel, but this does not demonstrate the "similarity in the ratios at which these changes have occurred." For, if the actual values be plotted to the time, the curve of phthisis deathrate is *convex* and the institutional deathrate *concave* to the time axis, in other words while the rate of one is increasing,

---

* The assumption made appears to be that for the period in question D_i is proportional to the institutional deaths from phthisis,—a very big assumption.

† The logarithms of the ratios of institutional deaths to all deaths appear to be either wrongly plotted or wrongly calculated.

the rate of the other is decreasing during the period in question,—always on the supposition that we plot the results as Dr Newsholme has done with *reversed* directions of increasing scales for the two indices. He states that "the experience is summarised in the high correlation coefficients of ·91 for England and Wales (1878—1903) and ·90 for London (1866—1904)" (p. 271). The correlations found from his actual tables do not appear to agree with these, being, for example, − ·93 for England and Wales with the *negative* sign as we should anticipate; but as Dr Newsholme does not give the same years for his correlation coefficients as in his tables, he may have worked out his coefficients for individual years. It is impossible to test the matter, as neither the figures nor their source are provided.

If, however, we take his Tables LXII and LXIII, and apply the variate difference method* to Dr Newsholme's data as they stand in his book, which are all the data available, we find

Correlation of Phthisis Deathrate and Ratio of Deaths in Institutions to Total Deaths.

England and Wales:	Third Differences	− ·174 ± ·293,
London:	Second Differences	− ·094 ± ·252.

In other words the data show no significant relationship between this measure of segregation and the phthisis deathrate, when the time-factor is annulled, even with the early differences. It is impossible to press the matter further because the data are far too sparse for difference treatment, but the results, such as they are, are sufficient to indicate that Dr Newsholme's high correlations are solely due to the fact that both variates are continuously changing with the time†.

(3) As a second measure of segregation Dr Newsholme takes $100\phi_i/\phi$ and $1000\phi/P$ is then correlated with this, ϕ_i being the deaths from phthisis in institutions. On p. 275 Dr Newsholme gives very meagre data for Brighton, Sheffield and Salford in groups of years, six pairs of values for Sheffield, five for Brighton and four for Salford. It is thus impossible to test these for annulment of the time-factor, and no references are given to the sources of the original data. On p. 276 we read:

> Coefficients of correlation summarising this correspondence for long series of single years work out at ·67 for Salford from 1884 to 1904 and ·80 for Sheffield from 1876 to 1905‡.

If the arithmetical values be correct, they should certainly have negative signs, but even then they would not demonstrate anything but the increasing use of institutions and the decreasing prevalence of phthisis during the years in question.

---

* *Biometrika*, Vol. x. pp. 179, 341.

† These values might be modified if we could go to higher differences, but this is impossible on the very limited data which Dr Newsholme provides. On these data all we can state is that no evidence of organic relationship between the variates, such as is asserted by Dr Newsholme to exist, can be demonstrated.

‡ There is no statement as to why Brighton has been omitted.

There is, however, a much graver criticism to be made of Dr Newsholme's method in this measure of segregation. He proposes to correlate

$$100\phi_i/\phi \quad \text{and} \quad 1000\phi/P,$$

and interprets the high correlations as a sign of the value of segregation in reducing the phthisis deathrate. We have not his data to test his conclusions by, but we can compare them against certain results for 38 years in (i) England and Wales, (ii) Scotland, and (iii) Ireland. Here they are:

Correlation of Phthisis Deathrate and Ratio of Institutional Phthisical to all Phthisical Deaths.

Years	District	Correlation
1866—1903	Scotland	$-\cdot 9815 \pm \cdot 0040$
1866—1903	England and Wales	$-\cdot 9750 \pm \cdot 0054$
1866—1903	Ireland	$-\cdot 8720 \pm \cdot 0262$
1876—1905	Sheffield	$-\cdot 80 \pm \cdot 0443$
1884—1904	Salford	$-\cdot 67 \pm \cdot 0811$

The reader may imagine in this table a confirmation of Dr Newsholme's results, for the larger material gives higher values of the correlations. On the contrary, these correlations have been obtained by taking as the measure of segregation the ratio

$$100 \times \frac{\textit{Mean}\text{ Institutional deaths per annum from phthisis 1866–1903}}{\text{Annual Total deaths from phthisis}}.$$

Now it is clear that this index never *varies* with the increasing percentage of institutional deaths from phthisis. Yet all the correlations are greater than Dr Newsholme's! We have little doubt that he would get higher values than he has done, if he replaced the actual institutional deaths per annum by the constant mean value. In other words the results reached by him are of no significance, for we get higher correlations by putting a single fictitious value for the annual institutional deathrate.

The real source of his result is not the strong influence of segregation on phthisis, but the spurious correlation introduced by using the phthisis deaths, ϕ, in the numerator of one variate, $1000\phi/P$, and in the denominator of the other, $100\phi_i/\phi$. Thus no scientific results of value can be found from Dr Newsholme's second measure of segregation.

In discussing this second measure of segregation, Dr Newsholme lays great stress on the part played by asylums for the insane in segregating the tuberculous. He notes that the percentage of lunatics treated privately with relatives and others was 18·4 in 1859 and fell to 5·5 in 1902, thus marking increasing segregation during the period of fall in the phthisis deathrate. He states (p. 274) that: "the deathrate from tuberculosis in borough and county asylums in 1901 was 15·8 per cent. of the inmates, and over ten times as great as in the general population." Now

Dr Newsholme's figure appears to be quoted from the *56th Annual Report of the Commissioners in Lunacy*, and in this case it should read 15·8 per 1000 and not per 100, and although Dr Newsholme appears to have made a similar slip in dealing with the deathrate in the general population, he seems to be comparing deaths from all forms of tuberculosis among the insane—some of which have possibly a direct relation to their insanity—with deaths from phthisis alone in the general population. Further he has made no allowance for the very marked difference between the age distributions of the two groups he is comparing. The difference is so great that a phthisis deathrate of 1·46 per 1000 in the general male population is equivalent to one of 2·41 per 1000 among the insane population of males. Even if the corrected deathrate among the insane for phthisis were ten times its magnitude among the sane, we fail to understand what Dr Newsholme means when he asserts that: "the segregation of each tuberculous lunatic has been equivalent to the withdrawal of ten ordinary tuberculous persons" (p. 274). Because tuberculosis among lunatics is ten times as frequent—judging by deaths, and accepting for the purpose of argument Dr Newsholme's figures—why should the isolation of one tuberculous lunatic be equivalent to the withdrawal of ten sane tuberculous persons? That must suppose a tuberculous lunatic capable of spreading ten times the infection of a tuberculous but sane individual. All Dr Newsholme could say would be that from the standpoint of segregation it is ten times more desirable to segregate *any* lunatic, than *any* sane person, for the former is ten times as likely to die of tuberculosis. Dr Newsholme brings no evidence to show that the individual tuberculous lunatic is ten times as dangerous as the individual tuberculous sane person. As a matter of fact we still need very careful investigation of the relation of lunacy to tuberculosis, not only having regard to some forms of tuberculosis as possible sources of feeble-mindedness, if not of insanity, but also having regard to whether the old idea of asylum segregation as a possible cause of the spread of tuberculosis among lunatics is wholly erroneous, and we might further examine whether the new idea that the majority of tuberculous lunatics were tuberculous on admission is in its turn wholly sound*. In the present state of our knowledge we think the assertion that the increased segregation of lunatics has substantial relation to the decrease in the phthisis deathrate is quite unproven.

(4) Dr Newsholme's third approximation to the segregation ratio is the index $100p_i/p_\tau$ where p_i is the number of paupers in institutions and p_τ is the total number of paupers, indoor and outdoor. Unfortunately Dr Newsholme's usage does not agree with his definition. The index he appears to use is generally $100p_\tau/p_i$, and the values of this are given in the last column of Table LXV (p. 277) and Table LXVII (p. 279). In Table LXVI (on p. 277), however, the 100 factor is dropped and p_i/p_τ again used in the heading to the central column,

---

* Many lunatics enter and re-enter asylums, it does not follow because they died of tuberculosis and were tuberculous on last admission that their tuberculosis was there on first admission.

although the figures in that column appear to refer to $100p_\tau/p_i$. Below this table occur the words:

> This experience for the entire series of individual years is expressed by a coefficient of correlation of – ·94 between segregation measured by the fraction of pauper population treated in institutions and the phthisis deathrate. (p. 277.)

The correlation to support Dr Newsholme's views should be negative if

$$100p_i/p_\tau$$

has been used, and positive if $100p_\tau/p_i$ has been used. But as many of his other correlations are given with the wrong sign, it is difficult to discover what measure of segregation he actually has used. To add to the confusion the index actually plotted is $\log p_\tau/p_i$, and not $100p_i/p_\tau$, which is what Dr Newsholme defines as his index. We have accordingly in our analysis of the figures, to be given later, used both indices $100p_i/p_\tau$ and $100p_\tau/p_i$.

It is very difficult to appreciate how the ratio $100p_i/p_\tau$ can effectively measure the segregation ratio—it is indeed impossible to agree with Dr Newsholme's view that any of his indices "measure with approximate accuracy the ratio which states how many of total days of tuberculous sickness are passed in institutions."

The policy of compelling as many paupers as possible to go into the workhouse was directly adopted with a view to diminishing the total pauperism as well as abuses connected with outdoor relief, and that policy is the source of increase in the index $100p_i/p_\tau$. Had Dr Newsholme examined his own Tables LXV, LXVII and LXIX carefully, he would have seen that the percentage of indoor paupers on the general population has remained *almost constant* for the period in question, while the total paupers per cent. of the general population in England with Wales and in Scotland have decreased. If the same relative number of paupers are segregated now as formerly, how can this segregation have diminished the chances of infection in the community? We can hardly assume that all paupers are tuberculous, or markedly so relatively to other men, so that the reduction of the number of outside paupers by indoor segregation is equivalent practically to a reduction *pro tanto* (note the extraordinarily high correlations!) of the number of tuberculous in the community. If so, then the reduction of the tuberculous deathrate would be due not to the segregation, but to the large decrease in the total pauperism relative to the population of this country. The correlation, as we shall demonstrate, is not between the segregation of paupers and the phthisis deathrate, but between the diminution of total pauperism and the phthisis deathrate. We shall investigate how far this relationship between total pauperism and the phthisis deathrate is "organic," i.e. continues after the annulment of the time-factor, or is purely due to the fact that both pauperism and phthisis have diminished during the forty-year period under consideration.

It was this third definition of a segregation ratio in conjunction with the fourth segregation ratio to be considered later that led us to realise that the whole problem must be dealt with afresh, and the modern methods of partial correlation and variate difference correlation applied to its various aspects. We have taken the period used by Dr Newsholme, 1866–1903 inclusive, and have used the figures for each individual year thus obtaining 38 entries, which are few indeed, but the best we can probably do with data of this kind, and therefore directly comparable with Dr Newsholme's results, for he seems to have used individual years for his correlations although he does not always say so (cf. pp. 271 and 280), and notwithstanding that his tables are all given for five-year periods.

The population numbers for England and Wales (Table A) were taken from the Registrar-General's *Annual Report* for 1909, and the phthisis deaths from the *Reports* for 1866–1903; the average of each five years' period agrees with Dr Newsholme's values for phthisis, but the values for indoor and for total paupers do not quite agree with his. Dr Newsholme was therefore written to and asked whence he obtained his numbers. He was kind enough to reply, but said that he was unable to refer at the moment to the original tables, but that undoubtedly the data were the statistics given in the *Annual Reports* of the Registrars-General for England, Scotland and Ireland. We then examined the Local Government Board returns and found that Dr Newsholme apparently had used the pauper returns for the January quarter of each year. We kept therefore to the Registrar-General's *Report*, as the numbers there given are based on the Local Government Board's returns for the whole year, which are a fairer measure of pauperism than those for the January quarter alone.

For Scotland, our numbers (Table A) agree with Dr Newsholme's for both phthisis and indoor paupers, except when we take the first five-year period (1866–70), where they differ slightly. In the case of total paupers for the periods 1866–70, 1881–85, and 1896–1900 our figures do not agree*. We cannot find any reason for these divergences except a slip in his or our arithmetic, or the possibility that a wrong number of *outside* paupers has been taken by one or other of us. We do not think the differences in the values are such as to invalidate a comparison of results.

In Ireland the only serious discrepancy in our values is in the total number of paupers for the period 1876–80.

These discrepancies, however, emphasise the very necessary rules for statistical treatment: (i) that the ultimate raw data should be published with every inquiry, and (ii) it should be stated exactly where they are taken from, and how they have been treated.

Table A gives our raw data, Table B our deathrates and indices based thereon. We have correlated the phthisis deathrate taken as $10^5\phi/P$ with $100p_i/p_\tau$ and

---

* We are unable to compare his and our data for individual years, because Dr Newsholme has only published his data for five-year periods.

$100p_\tau/p_i$. Taking first England and Wales, and calling these three indices respectively I_ϕ, I_i and I_τ, we find:

$$\text{Correlation of } I_\phi \text{ and } I_i = -\cdot 9664 \pm \cdot 0072,$$

$$I_\phi \text{ and } I_\tau = +\cdot 9298 \pm \cdot 0148.$$

Dr Newsholme gives $-\cdot 94$ as the coefficient of correlation "between segregation measured by the fraction of pauper population treated in institutions and the phthisis deathrate" (p. 277). Having regard to his confusion of I_i and I_τ and his frequent interchange of the signs of correlation coefficients, we can only say our results confirm his high numerical value, but not his actual figure.

But does this actual figure mean that there is any real relationship between segregation and the phthisis deathrate? To test this, we replaced the index I_i by $I_{\bar{i}}$, where

$$I_{\bar{i}} = 100\,\frac{\textit{Mean}\text{ number of indoor paupers per }10^5\text{ for the population, 1866–1903}}{10^5 \times p_\tau/P}$$

$$= 100\left(\frac{\bar{p}_i}{P}\right)\Big/\left(\frac{p_\tau}{P}\right).$$

In this index the relative number of indoor paupers is assumed to remain *absolutely constant.* We found:

$$\text{Correlation for England and Wales of } I_\phi \text{ and } I_{\bar{i}} = -\cdot 9459 \pm \cdot 0115,$$

that is to say we get substantially the same value, a value higher than Dr Newsholme's, by putting the number of indoor paupers relative to the general population *constant* throughout the period. It is very difficult, in the face of such a result, to suppose that segregation of paupers has anything whatever to do with the diminution of the phthisis deathrate. It is clearly due to a negative correlation of a high magnitude between $\dfrac{1}{p_\tau/P}$ and ϕ/P, or to a positive correlation between $\dfrac{p_\tau}{P}$ and $\dfrac{\phi}{P}$, i.e. to a correlation between a high *total* pauper rate and a high phthisis deathrate. Dr Newsholme's result merely reduces to the statement that total pauperism in England and Wales has diminished contemporaneously with phthisis. If the result has nothing to do with segregation, can we assert that the reduction of phthisis is causally related to the reduction in total pauperism?

Overlooking for a moment a new objection to be raised later, let us apply the variate difference method to the correlation of ϕ/P with $100p_i/p_\tau$ and $100p_\tau/p_i$ in the cases of England with Wales, of Scotland, and of Ireland; also to the correlation of ϕ/P with the index $100\,(\overline{p_i/P})/(p_\tau/P)$ in the case of England with Wales. The following are the results:

TABLE I.

	England with Wales			Scotland		Ireland	
Correlation of with	$10^5\phi/P$ I_i	$10^5\phi/P$ I_i	$10^5\phi/P$ I_τ	$10^5\phi/P$ I_i	$10^5\phi/P$ I_τ	$10^5\phi/P$ I_i	$10^5\phi/P$ I_τ
Crude Indices	− ·946 ± ·012	− ·966 ± ·007	+ ·930 ± ·015	− ·952 ± ·010	+ ·920 ± ·017	− ·881 ± ·024	+ ·893 ± ·022
Δ_1	+ ·090 ± ·134	− ·258 ± ·126	+ ·340 ± ·120	− ·265 ± ·126	+ ·250 ± ·127	− ·280 ± ·125	+ ·235 ± ·128
Δ_2	− ·201 ± ·149	− ·461 ± ·123	+ ·542 ± ·110	− ·240 ± ·147	+ ·182 ± ·151	− ·264 ± ·145	+ ·180 ± ·151
Δ_3	− ·335 ± ·153	− ·508 ± ·127	+ ·567 ± ·116	− ·205 ± ·164	+ ·086 ± ·170	− ·226 ± ·163	+ ·162 ± ·167
Δ_4	− ·407 ± ·155	− ·518 ± ·136	+ ·547 ± ·130	− ·186 ± ·179	+ ·024 ± ·185	− ·182 ± ·179	+ ·133 ± ·182
Δ_5	− ·475 ± ·153	− ·528 ± ·143	+ ·529 ± ·142	− ·182 ± ·191	− ·003 ± ·198	− ·145 ± ·194	+ ·108 ± ·195
Δ_6	− ·538 ± ·149	− ·543 ± ·147	+ ·583 ± ·151	—	—	− ·112 ± ·206	+ ·081 ± ·208
Δ_7	− ·584 ± ·145	− ·562 ± ·150	+ ·539 ± ·156	—	—	—	+ ·044 ± ·219
Δ_8	− ·614 ± ·143	− ·587 ± ·151	+ ·557 ± ·159	—	—	—	− ·004 ± ·230

It will be seen from this table that whether we use the index I_i or its inverse I_τ, we get practically the same results—naturally with changed sign. But the results themselves are of extraordinary interest. For both Scotland and Ireland, when we proceed to annul the time-factor by correlating successive differences, we find that the high correlations interpreted by Dr Newsholme as marking a relation between pauper segregation and phthisis deathrate entirely disappear or become less than their probable errors. There is thus no organic relation between these variates as measured by the above indices. In the case of England and Wales, however, while there is a reduction on annulment of the time-factor to roughly two-thirds of the high value noted by Dr Newsholme, this value does not tend to disappear with increasing differences. Thus in England with Wales, as apart from the remainder of Great Britain, there would at first sight appear to be an organic relation between segregation of paupers and the phthisis deathrate. But our first column under the England with Wales section shows that if we fix the percentage in the general population of these indoor paupers and then annul the time-factor, we reach a *slightly higher* value of this apparent organic relation. It has therefore nothing to do with segregation. Thus Dr Newsholme's interpretation of his original high correlations appears in every case fallacious.

There are two methods of testing this result, i.e. the absence of organic relationship between indoor pauperism and phthisis. Suppose we correlate the crude numbers of phthisis deaths per annum and of indoor paupers per annum, the resulting coefficient will have very small logical value because both these variates are continuously changing with the time*. But now suppose we annul

---

* It is noteworthy that the England with Wales and the Scotland correlation coefficients for these crude variates are high and negative, but for Ireland the coefficient is moderate and positive. Thus the factors at work must be totally different in the two Islands. Since indoor paupers relative to the population have remained singularly constant the increase of phthisis deaths must have been much slower than the population increase in Great Britain, but somewhat faster in Ireland.

the time-factor by correlating the differences of these variates, then we shall free ourselves from the influence of the time-variate, and in doing this we shall also free ourselves practically from the influence of change of population, which is a time change.

The following table resulted from this investigation.

TABLE II.

Correlation of Crude Phthisis Deaths (ϕ) *and Indoor Paupers* (p_i).

Variates	England with Wales	Scotland	Ireland
Crude	− ·934 ± ·014	− ·718 ± ·053	+ ·457 ± ·086
Δ_1	− ·376 ± ·116	− ·206 ± ·130	− ·092 ± ·134
Δ_2	− ·302 ± ·141	− ·219 ± ·148	− ·103 ± ·154
Δ_3	− ·213 ± ·164	− ·180 ± ·166	− 143 ± ·168
Δ_4	− ·100 ± ·183	− ·157 ± ·181	− ·147 ± ·181
Δ_5	− ·016 ± ·198	− ·158 ± ·193	− ·140 ± ·194

It will be seen that for all three countries, whether we start with the positive correlation of the Irish or the negative correlation of the English and Scottish returns, there is no remaining significant correlation after annulment of the time-factor between indoor pauperism and phthisis.

A second method of verifying our conclusions is to find the partial correlation between indoor pauperism and phthisis deaths for a constant value of the total population and a constant value of total pauperism. We thus ask the question whether with a constant population and a constant amount of total pauperism, an increase of indoor pauperism would organically affect the number of deaths from phthisis. By making the population and the total pauperism constant we are largely producing an annulment of the time-factor and ascertaining whether a change in the number of indoor paupers due to causes other than temporal influences the number of deaths from phthisis.

The system of correlation coefficients given in Table III, p. 540, was determined:

Here the values of ${}_{Pp_\tau}r_{p_i\phi}$ for England with Wales and for Scotland confirm the conclusions we have reached by other methods, i.e. there is no significant relationship at all between phthisis and *indoor* pauperism. The value for Ireland is, perhaps, significant, but having regard to its smallness (− ·3 ± ·1) and the size of its probable error, no one can lay real stress on it, in opposition to the results of the other two countries. In general the coefficients for the Irish data appear very anomalous, and certainly divergent from those for Great Britain.

Thus our investigation of the relation between indoor pauperism and phthisis appears to be entirely opposed to Dr Newsholme's conclusions. We find the segregation of paupers to have no substantial influence on deaths from phthisis. The one outstanding point at present, the relation between p_τ/P and ϕ/P after annulment of the time-factor (see our p. 538), has no bearing on the segregation problem of Dr Newsholme.

TABLE III.

Total and Partial Correlation Coefficients of Crude Numbers of Indoor and Total Paupers (p_i and p_T), Total Population (P), and Phthisis Deaths (ϕ).

	Coefficients	England with Wales	Scotland	Ireland
Total Coefficients	$r_{p_i\phi}$	− ·934 ± ·014	− ·718 ± ·053	+ ·457 ± ·087
	r_{p_iP}	+ ·955 ± ·010	+ ·831 ± ·034	+ ·763 ± ·046
	$r_{\phi P}$	− ·950 ± ·011	− ·896 ± ·022	+ ·479 ± ·084
	$r_{p_ip_T}$	− ·544 ± ·077	− ·528 ± ·079	− 251 ± ·103
	$r_{\phi p_T}$	+ ·577 ± ·073	+ ·780 ± ·043	+ ·070 ± ·109
	r_{p_TP}	− ·674 ± ·060	− ·805 ± ·038	− ·684 ± ·058
Partial Coefficients	${}_Pr_{p_i\phi}$	− ·287 ± ·100	+ ·111 ± ·108	+ ·162 ± ·107
	${}_{p_T}r_{p_i\phi}$	− ·905 ± ·020	− ·575 ± ·073	+ ·492 ± ·083
	${}_{Pp_T}r_{p_i\phi}$	− ·189 ± ·106	− ·017 ± ·109	− ·305 ± ·099

To approach nearer to the meaning of the relation between total pauperism and phthisis we determined the correlation between p_T and ϕ for constant P, and found

$$ {}_Pr_{p_T\phi} = - \cdot 277 \pm \cdot 101, $$

which is barely significant having regard to its probable error.

Now after elimination of the time-factor, we found for the correlation of ϕ/P and I_i at the eighth difference $-\cdot 614 \pm \cdot 143$, but this is the same as the correlation of $\dfrac{1}{p_T/P}$ and ϕ/P. Hence the correlation of p_T/P and ϕ/P must be very significant, *positive* and of the order ·6. Now if p_T and ϕ *after the removal of the time-factor* were practically independent of each other, there would be a high positive correlation between p_T/P and ϕ/P, due to the fact that P when it takes—after annulment of the time-factor—any random deviation appears in both variates' denominators. In other words, we are inclined to believe that the high negative correlation between ϕ/P and I_i is solely due to spurious correlation arising from the nature of the indices used.

To throw still more light on the matter we have investigated the correlation between the total number of paupers and the total number of deaths from phthisis when the time-factor approaches annulment. It will be seen from the table below that for both Scotland and Ireland there is finally no relationship at all between phthisis deaths and total pauperism. On the other hand, England with Wales is tending to a value at least approaching to the crude correlation. We have therefore this noteworthy result: England with Wales starts with a considerable value and concludes with an equally great value, Scotland starts with a high value and ends with a zero value, Ireland starts with an insignificant value and ends with an insignificant value. If pauperism were causative of phthisis, it is hard to believe that this would not manifest itself in the Scottish and Irish returns; these negative any such hypothesis. It would appear that there are essential differences in the treatment of pauperism in the three countries. I suggest, but I cannot demonstrate the view, that phthisis itself leads to pauperism in England, i.e. that the relatives of the phthisical breadwinner more often are allowed to become paupers in England than in the sister countries. In other words, that the only organic relationship between pauperism and phthisis we have been able to discover *may* be due to a relatively harsh treatment in England of the dependents of the phthisical.

TABLE IV.

Relation between Total Pauperism (p_T) and Deaths from Phthisis (ϕ).

	England with Wales	Scotland	Ireland
Crude values	+ ·577 ± ·073	+ ·780 ± ·043	+ ·070 ± ·109
Δ_1	− ·095 ± ·134	+ ·025 ± ·135	+ ·164 ± ·132
Δ_2	+ ·174 ± ·151	+ ·025 ± ·156	+ ·144 ± ·152
Δ_3	+ ·286 ± ·158	+ ·012 ± ·172	+ ·131 ± ·169
Δ_4	+ ·347 ± ·163	+ ·033 ± ·185	+ ·110 ± ·183
Δ_5	+ ·413 ± ·164	+ ·027 ± ·198	+ ·090 ± ·196

To show how effectively the variate difference correlation method removes time influence, we may note that we correlated total population (P) with total pauperism (p_T) and total phthisis deaths (ϕ) with total population by this method, with a view to ascertaining whether the relation between p_T and ϕ would be modified, if we determined it for constant population.

The following results were reached:

TABLE IV^{bis}. *England with Wales.*

	Total Population and Total Pauperism (P and p_T)	Total Population and Phthisis Deaths (P and ϕ)
Crude values	− ·674 ± ·060	− ·950 ± ·011
Δ_1	+ ·457 ± ·107	− ·039 ± ·135
Δ_2	− ·016 ± ·156	− ·205 ± ·149
Δ_3	− ·022 ± ·171	− ·089 ± ·170
$_4$	− ·031 ± ·185	+ ·002 ± ·185

Thus we see that apart from the time-factor there is no relation whatever between either pauperism or phthisis and population. In the relation between total pauperism and phthisis deaths, no further correction for population is needful than that obtained by the annulment of the time-factor as in Table IV. Table IV *bis* shows us that neither pauperism nor phthisis is organically related to population, although we might well have anticipated that greater density of population would influence pauperism and provide greater chances of infection, and so of deaths, in the case of phthisis.

(5) We now come to Dr Newsholme's fourth and last measure of segregation. It is "the ratio in which the number of paupers treated in workhouses and workhouse infirmaries stand to the total number of deaths in the community" (p. 276). In our notation this is p_i/ϕ, or as an index $100p_i/\phi$. But in the figures actually given in Table LXV (p. 277), and headed Segregation Ratio, Dr Newsholme appears to be using $100\phi/p_i$. The same remark applies to Tables LXVIII and LXIX (pp. 280—281). Thus it is difficult to be certain of what Dr Newsholme intends to be taken as his fourth measure of segregation. In our discussion below we have used both $100\phi/p_i$ and $100p_i/\phi$ to provide for both contingencies and to check our results.

Unfortunately Dr Newsholme makes little attempt to justify either his third or fourth ratio as an approximate measure of segregation. It will be remembered that he has defined the true method of measuring segregation to consist in forming the ratio "stating how many of the total days of sickness (number of patients and number of days of sickness) are passed in institutions" (p. 267). In this fourth index of segregation he replaces phthisical patients in institutions by *indoor paupers*, and total of phthisical patients by total deaths from phthisis, dropping any question of the number of days of sickness. At the very least this seems to involve two assumptions, (*a*) either that all indoor paupers are phthisical or that for the period in question the proportion of indoor paupers who are phthisical has remained constant, (*b*) that for the period in question the number of deaths from phthisis has remained a constant fraction of the total number of cases of phthisis. It is difficult to see how, without such assumptions, such figures can "measure with approximate accuracy the ratio which states how many of the total days of tuberculous sickness are passed in institutions" (p. 267).

Yet in another paragraph Dr Newsholme quotes with apparent approval the statement of Mr Fleming, who speaks of the "great change in the character of workhouse inmates during recent years....The able-bodied inmates are gone and the sick inmates have come" (p. 273). Such a statement is absolutely inconsistent with the assumption (*a*) above.

To justify (*b*) we must assert that for the last fifty years of the nineteenth century there has been no change in efficiency of treatment in the case of tuberculosis, for without this we cannot assume that deaths from phthisis are even an approximate measure of the number of cases (p. 267). The fact that the reduction in the phthisis deathrate has been substantially different for the different age groups, and is especially marked in the case of children, seems to indicate that recovery, at least from puerile phthisis, is more frequent now than formerly. However, not to spend more time on these assumptions—which, it appears to us that Dr Newsholme has by no means justified—let us examine whither this fourth method of approximately measuring segregation leads us. Table V gives the necessary coefficients.

TABLE V.

Correlation and Difference Correlations of $10^5\phi/P$ *and* $100p_i/\phi$ *or* $100\phi/p_i$.

Variate $10^5\phi/P$ with	England with Wales		Scotland		Ireland	
	$100p_i/\phi$	$100\phi/p_i$	$100p_i/\phi$	$100\phi/p_i$	$100p_i/\phi$	$100\phi/p_i$
Crude	– ·760 ± ·046	+ ·976 ± ·005	– ·861 ± ·028	+ ·944 ± ·012	– ·712 ± ·054	+ ·666 ± ·061
Δ_1	– ·868 ± ·033	+ ·848 ± ·038	– ·755 ± ·058	+ ·772 ± ·055	– ·819 ± ·045	+ ·707 ± ·068
Δ_2	– ·879 ± ·035	+ ·875 ± ·037	– ·824 ± ·050	+ ·834 ± ·047	– ·922 ± ·023	+ ·755 ± ·067
Δ_3	– ·895 ± ·034	+ ·874 ± ·041	– ·809 ± ·059	+ ·824 ± ·055	– ·954 ± ·015	+ ·791 ± ·064
Δ_4	– ·895 ± ·037	+ ·860 ± ·048	– ·811 ± ·064	+ ·805 ± ·065	– ·964 ± ·013	+ ·805 ± ·065
Δ_5	– ·898 ± ·038	+ ·847 ± ·056	– ·786 ± ·076	+ ·788 ± ·075	– ·970 ± ·012	+ ·831 ± ·061
Δ_6	– ·907 ± ·037	+ ·850 ± ·058	– ·788 ± ·079	+ ·794 ± ·077	– ·973 ± ·011	+ ·848 ± ·059
Δ_7	– ·917 ± ·035	+ ·835 ± ·067	– ·792 ± ·082	+ ·791 ± ·082	—	+ ·857 ± ·056*

Now this table at any rate demonstrates a very high correlation between ϕ/P and p_i/ϕ, while the previous table for Dr Newsholme's third approximate segregation ratio led in the case of England with Wales to the value – ·587, and in the case of Scotland and Ireland to negligibly small values! Dr Newsholme himself writes: "Any of these indirect forms of segregation ratio has therefore to be verified wherever possible by the application to the same community and period of one or more other forms of the ratio, and checked where practicable by a special examination of sample constituent communities whose figures are included in the total. This has been done so far as the information obtainable allowed. It will be seen that the results obtained by applying different ratios to the experience of the same country and period are usually, though not invariably, in good agreement" (p. 268).

What is quite clear from the above results is that, while in the case of Dr Newsholme's two chief measures of segregation, there is very sensible difference in the case of England with Wales, there is an absolute discordance in the cases of both Scotland and Ireland. Accordingly on the basis of his own axiom, that we must check our results by application of one or more other forms of the ratio,

---

* This correlation continues to rise until it reaches ·929 with the thirteenth difference, but with such high differences the "population" is so reduced that the method ceases really to be reliable.

we are bound to reject these ratios as even approximate measures of segregation*.

But it would not be satisfactory to leave the matter here and not provide some explanation of why this fourth segregation ratio, both before and after the annulment of the time-factor, leads to such high correlations. Luckily the matter is capable of a perfectly straightforward and obvious explanation, which would have been anticipated had Dr Newsholme had in mind the danger of "spurious correlation."

What he is correlating are essentially ϕ/P and p_i/ϕ. The latter may be written $(p_i/P)/(\phi/P)$. Now p_i/P is practically constant during the period in question. Hence Dr Newsholme is correlating ϕ/P with $1/(\phi/P)$, or a variate with its reciprocal. In other words we may anticipate something very closely approaching perfect correlation. The deviation from such correlation arises from the fact that p_i/P is not absolutely steady, although its variations are very probably nearly random. The assertion therefore that this fourth measure of segregation assists in demonstrating the close relation between the fall in the phthisis deathrate and institutional segregation is based on a fallacy which entirely overlooks "spurious correlation."

It will be seen therefore that not one of Dr Newsholme's methods of reaching an approximate measure of the segregation is satisfactory, and they lead to contradictory and inconclusive results. Whether there is any really substantial relation between the prevalence of phthisis and institutional segregation we do not yet know. All we can say is that Dr Newsholme has entirely failed to demonstrate it, if it actually exists.

(6) Before concluding this paper it may be of interest to judge how far it justifies the application of the method of variate difference correlation to such problems as are here dealt with.

In the first place, the correlations of successive differences should approach steady values. This is generally—as the reader can judge by examining Tables I, II, IV and V—but not invariably, the case. The test cannot, however, be completed, as the method ought not to be pressed to such high differences that the order of the difference is a large percentage of the original "population."

We doubt whether it is advisable to carry differences beyond the 8th in a population of 38. 20 °/₀ to 25 °/₀ reduction in the population is as much surely as it is safe to allow where the original population is so small in number. It is true that a population of 38 itself is capable of exciting the derision of trained

---

* Under the circumstances it is, perhaps, unnecessary to draw attention to Dr Newsholme's statement that "the specific result of pauper segregation must have been lower in Ireland than in England or Scotland" (p. 282). Free of the time-factor the correlations of phthisis deathrate and Dr Newsholme's fourth segregation ratio are *higher* in Ireland than in England or Scotland. This criticism as well as Dr Newsholme's original remark are of no importance, because the fourth segregation ratio correlation is entirely spurious.

statisticians, and ought never to be used where hard work can produce larger numbers. But in annual returns, as has been indicated by others, a period of 30 to 50 years is often the maximum attainable, and we must take what we can get. In the present case the probable errors of the difference correlations—based on the Andersonian formulae for steady conditions—show us that we can form fairly legitimate conclusions from the results reached.

A second test that we have applied is the approach to the theoretical values in the function $\sigma^2_{\delta_m x}/\sigma^2_{\delta_{m-1} x}$ where $\delta_m x$ is the mth difference of the variate x.

The following table shows that there is a reasonable approach to these theoretical values in the calculated standard deviations of the differences, and suffices to justify the application of the variate difference method within the limits of practical statistics. We have continued the differences beyond the values used in some of the correlation results to indicate the sort of irregularities which may be expected to occur when using high differences in small populations. Terminal irregularities then begin to affect the uniform rise of $\sigma^2_{\delta_m x}/\sigma^2_{\delta_{m-1} x}$.

TABLE VI. *Values of* $\sigma^2_{\delta_m x}/\sigma^2_{\delta_{m-1} x}$ *and their approach to* $4 - \frac{2}{m}$.

m	Theoretical Series	Phthisis and Death-rate $10^5\phi/P$			Third Segregation Ratio $100p_i/p_\tau$			Fourth Segregation Ratio $100p_i/\phi$			Third Segregation Ratio Reciprocal: $100p_\tau/p_i$			Fourth Segregation Ratio Reciprocal: $100\phi/p_i$		
		England and Wales	Scotland	Ireland	England and Wales	Scotland	Ireland	England and Wales	Scotland	Ireland	England and Wales	Scotland	Ireland	England and Wales	Scotland	Ireland
1	2·000	·027	·056	·478	·019	·030	·036	·046	·127	·637	·016	·038	·072	·076	·158	·078
2	3·000	2·317	2·225	2·999	1·605	1·781	2·009	1·630	1·812	2·220	1·193	1·412	2·559	1·934	1·689	2·355
3	3·333	2·928	2·806	3·308	2·791	3·111	3·071	2·700	2·805	3·054	2·493	2·715	3·247	2·878	2·506	2·625
4	3·500	3·135	3·202	3·422	3·180	3·621	3·307	3·142	3·306	3·283	2·864	3·329	3·471	3·081	3·090	2·840
5	3·600	3·547	3·450	3·502	3·431	3·743	3·456	3·331	3·558	3·392	3·215	3·418	3·597	3·222	3·398	2·922
6	3·667	3·574	3·575	3·532	3·606	—	3·560	3·441	3·698	3·469	3·584	—	3·666	3·411	3·654	3·132
7	3·714	3·729	3·630	3·564	3·697	—	—	3·588	3·791	—	3·780	—	3·688	3·595	3·688	3·272
8	3·750	3·802	—	3·595	3·729	—	—	—	—	—	3·843	—	—	—	—	3·335
9	3·778	3·840	—	3·685	3·770	—	—	—	—	—	3·865	—	—	—	—	3·498
10	3·800	3·864	—	3·764	3·820	—	—	—	—	—	3·869	—	—	—	—	3·625
11	3·818	3·881	—	3·823	3·892	—	—	—	—	—	3·893	—	—	—	—	3·726
12	3·833	3·877	—	3·818	3·932	—	—	—	—	—	3·907	—	—	—	—	3·740
13	3·846	3·843	—	3·747	3·930	—	—	—	—	—	3·970	—	—	—	—	3·668
14	3·857	3·779	—	3·659	3·903	—	—	—	—	—	3·827	—	—	—	—	3·604

TABLE A.

	England with Wales.				Scotland.				Ireland.			
Year	Number of Deaths from Phthisis	Population	Number of Indoor Paupers	Number of Total Paupers	Number of Deaths from Phthisis	Population	Number of Indoor Paupers	Number of Total Paupers	Number of Deaths from Phthisis	Population	Number of Indoor Paupers	Number of Total Paupers
1866	55714	21409684	130599	867092	8809	3215129	7070	126042	10119	5522942	50510	62019
1867	55042	21677525	138736	921295	8827	3245098	7929	128361	10383	5486509	53147	67950
1868	51423	21948713	149251	956964	8778	3275350	8794	130446	9440	5465914	54484	71195
1869	52270	22223299	148896	964491	8827	3305885	8346	130003	9983	5449094	52118	70093
1870	54231	22501316	148509	978807	9443	3336707	7928	126239	9989	5418512	49115	70183
1871	53376	22788594	143568	948006	9366	3368921	7749	122909	10427	5398179	46094	70303
1872	52589	23096495	138579	852261	8989	3404798	7402	116356	10430	5372890	46053	73107
1873	51355	23408556	138110	793550	8465	3441056	7898	109634	10341	5327938	47391	77602
1874	49379	23724834	137604	747332	8068	3477704	7769	103688	9416	5298979	47699	77931
1875	52943	24045385	134654	714148	8892	3514744	7673	98868	10240	5278629	45809	76050
1876	51775	24370267	138911	670182	8703	3552183	7586	95682	10092	5277544	43936	75405
1877	51353	24699539	148197	666888	8580	3590022	8046	93380	10433	5286380	44888	78053
1878	52856	25033259	155206	683408	8672	3628268	8763	91795	10406	5282246	48406	84202
1879	51272	25371489	164915	735773	8467	3665443	9140	95138	10619	5265625	52063	91284
1880	48201	25714288	172286	730286	8190	3705994	9296	95089	11124	5202648	54610	111624
1881	47541	26046142	177114	734501	7822	3742564	9040	93958	10027	5145770	53038	114568
1882	48715	26334942	176444	716310	8069	3770657	8964	90624	10258	5101018	51262	109804
1883	50053	26626949	175738	710940	8434	3798961	8722	88216	10775	5023811	50385	111106
1884	49325	26922192	177686	699977	7923	3827478	9041	85608	10583	4974561	48241	106210
1885	48175	27220706	179657	707738	7905	3856307*	9007	86362	10709	4938588	47009	105640
1886	47872	27522532	181649	732099	7926	3885155	9495	88070	10694	4905895	46749	124205
1887	44935	27827706	183663	729086	7359	3914318	9095	86984	10329	4857119	46442	112049
1888	44248	28136258	185699	731543	6941	3943701	8863	86394	9812	4801312	45638	109413
1889	44738	28448239	182069	711206	7152	3973305	8471	84766	9951	4757385	44046	106336
1890	48366	28763673	178335	684575	7726	4003132	8182	82454	10176	4717959	43009	105090
1891	46515	29085819	174515	663157	7427	4036245	8160	80454	10038	4680376	41744	104272
1892	43323	29421392	179470	667866	6853	4078910	8527	80014	10048	4633808	41582	101316
1893	43632	29760842	190469	690452	7007	4122029	8824	81002	9869	4607462	41836	99737
1894	41641	30104201	195677	701428	7248	4165606	9212	82320	9626	4589260	42263	99519
1895	42490	30451528	200980	736927	7830	4209645	9083	84317	9768	4559936	41719	99247
1896	40251	30802858	200219	717707	7062	4254153	9569	86063	9051	4542061	41990	96386
1897	41642	31158245	199413	716640	7395	4299132	9908	87221	9748	4529917	42849	97545
1898	41335	31517725	204865	721756	7299	4344589	10032	86941	9678	4518478	44186	110798
1899	42408	31881365	204041	691826	7332	4390530	9965	84969	9480	4502401	43627	101040
1900	42987	32249187	199945	677233	7365	4436958	9868	85850	10076	4468501	41882	99925
1901	41224	32621263	205514	691571	6860	4483880	10306	85432	9549	4445630	42386	99956
1902	40671	32997626	211185	702849	6616	4531299	10865	86999	9400	4432274	43132	101053
1903	40132	33378338	220297	717634	6686	4579223	11711	88823	9559	4413658	43523	100490

* 3856207 in *Annual Reports* before 1908.

TABLE B.

	England with Wales.				*Scotland.*				*Ireland.*			
Year	Number of Deaths from Phthisis per 10^5 of the Population	Number of Deaths from Phthisis per 10^2 Indoor Paupers	Number of Indoor Paupers per 10^2 Deaths from Phthisis	Number of Indoor Paupers per 10^2 Total Paupers	Number of Deaths from Phthisis per 10^5 of the Population	Number of Deaths from Phthisis per 10^2 Indoor Paupers	Number of Indoor Paupers per 10^2 Deaths from Phthisis	Number of Indoor Paupers per 10^2 Total Paupers	Number of Deaths from Phthisis per 10^5 of the Population	Number of Deaths from Phthisis per 10^2 Indoor Paupers	Number of Indoor Paupers per 10^2 Deaths from Phthisis	Number of Indoor Paupers per 10^2 Total Paupers
1866	260	43	234	15·1	274	125	80	5·6	183	20	500	81
1867	254	40	252	15·1	272	111	90	6·2	189	20	513	78
1868	234	34	290	15·6	268	100	100	6·7	173	17	576	77
1869	235	35	285	15·4	267	106	94	6·4	183	19	522	74
1870	241	37	274	15·2	283	119	84	6·3	184	20	492	70
1871	234	37	269	15·1	278	121	83	6·3	193	23	442	66
1872	227	38	264	16·3	264	122	82	6·4	194	23	442	63
1873	219	37	269	17·4	246	107	93	7·2	194	22	458	61
1874	208	34	279	18·4	232	104	96	7·5	178	20	506	61
1875	220	39	254	18·9	253	116	86	7·7	194	22	447	60
1876	212	37	268	20·7	245	114	87	7·9	191	23	436	58
1877	208	35	289	22·2	239	107	94	8·6	197	23	431	58
1878	211	34	294	22·7	239	99	101	9·6	194	21	472	57
1879	202	31	322	22·4	231	93	108	9·6	198	20	499	57
1880	187	28	357	23·6	221	88	114	9·8	214	20	491	49
1881	183	27	373	24·1	209	86	116	9·6	194	19	531	46
1882	185	28	362	24·6	214	90	111	9·9	201	20	500	47
1883	188	29	351	24·7	222	97	104	9·9	215	21	467	45
1884	183	28	360	25·4	207	88	114	10·5	213	22	455	45
1885	177	27	373	25·4	205	88	114	10·4	218	23	437	45
1886	174	26	379	24·8	204	84	120	10·8	219	23	435	38
1887	162	25	409	25·2	188	81	123	10·4	214	22	447	41
1888	157	24	420	25·4	176	78	128	10·3	205	22	464	42
1889	157	25	407	25·6	180	85	118	10·0	209	23	443	41
1890	168	27	369	26·1	193	95	106	9·9	216	24	423	41
1891	160	27	375	26·3	184	91	110	10·1	214	24	416	40
1892	147	24	414	26·9	168	80	124	10·7	217	24	414	41
1893	147	23	437	27·6	170	79	126	10·9	214	24	424	42
1894	138	21	470	27·9	174	79	127	11·2	209	23	439	42
1895	140	21	473	27·3	186	86	116	10·8	214	23	427	42
1896	130	20	497	27·9	166	74	136	11·1	199	22	465	44
1897	134	21	479	27·8	172	75	134	11·3	215	23	440	44
1898	131	20	496	28·4	168	73	138	11·5	214	22	457	40
1899	133	21	481	29·5	167	74	136	11·7	211	22	460	43
1900	133	22	465	29·5	166	75	134	11·5	226	24	416	42
1901	126	20	499	29·7	153	67	150	12·1	215	23	444	42
1902	123	19	519	30·0	146	61	164	12·5	212	22	459	43
1903	120	18	549	30·7	146	57	175	13·2	217	22	455	43

THE INFLUENCE OF ISOLATION ON THE DIPHTHERIA ATTACK- AND DEATH-RATES.

BY ETHEL M. ELDERTON, Galton Research Fellow
AND KARL PEARSON, F.R.S.

(1) *Introductory.* The problem of the advantages of isolation, not only in the case of diphtheria but of other diseases of an infectious character, is likely, owing to modern views as to "carriers" and other sources of transmission, to be much discussed in the near future. It is therefore well to consider what may be learnt from the statistics available. The questions which naturally arise are of the following kind:

(i) In districts with a maximum of isolation is there a minimum of incidence?

(ii) In districts with a maximum of isolation is there a minimum deathrate from the disease isolated?

There cannot be the slightest doubt that, if these two questions were answered in the affirmative and we could show that the incidence was markedly less and the deathrate significantly smaller in districts where isolation was most stringently carried out, then these results would be advanced as a strong argument in favour of isolation.

To the trained statistician, however, no conclusion based upon such results without much further analysis would have any validity. To illustrate this point, let us consider the hypothetical case that medical or popular opinion in a given town has been persistently in favour of increasing the isolation-rate, and further suppose that in this district improved economic conditions have increased the immunity, or bettered sanitation lowered the incidence, while at the same time new methods of treatment have lowered the deathrate of the disease; it will be clear that in considering the statistical results over a course of years we should find a high isolation-rate *negatively* correlated with both the incidence- and the death-rates. Thus if we considered this correlational as a causal nexus, we should be raising an apparently strong argument in favour of a maximum of isolation, which would be based on the statistical fallacy, that when two quantities are both changing continuously with the time, this must of itself denote a causal relation.

In precisely the same way a *positive* correlation between the isolation rate and the attack- or death-rates by no means justifies us in asserting that isolation is worse than non-effective. It is conceivable that in the period or the district under consideration with an increasing isolation-rate there might be decreased immunity in the population, greater virulence of the disease, or even a limit to the available isolation accommodation, so that in the case of attacks of an epidemic nature the isolation *rate* would not increase proportionately to the cases, or indeed might even diminish*. Further, if apart from the changes in a single district, we consider a great variety of districts, it may chance that the greatest isolation-rate occurs in those districts where the disease has been found most prevalent, because it appeared the most obvious remedy, and thus a greater attack- or death-rate would be no real measure of the futility of high isolation.

If, however, it should turn out that on the whole the higher isolation rate is associated with the higher attack-rate or the higher death-rate then it will be clear (i) that there is ground for demanding a closer investigation as to the advantages of isolation, and (ii) that we may be overlooking the real method, or at least one or more important factors, of the transmission of the disease. It is conceivable that isolation of all cases during attack may be of far less importance than isolation of certain special cases for a shorter or longer period well subsequent to the attack, and after they would normally have resumed their ordinary avocations†.

The main problems which arise are accordingly these:

(i) Have isolation-, attack- and death-rates changed continuously with the time, and are the apparent correlations really suggestive of causal relationships?

(ii) Are associations between isolation-rate and attack- and death-rates really spurious arising from the fact that where the attack- and death-rates have been severe there the remedy which appeared nearest to hand was more isolation?

---

* For example, if there were only 100 hospital beds available, and out of 100 cases 50 were sent to hospital, the isolation-rate would be 50 °/₀; but if in the next year there were 300 cases and all the beds were used, the isolation-rate would be only 33 °/₀. Thus limited accommodation may tend to produce a *negative* correlation between isolation-rate and attack-rate, so that a *positive* correlation between these two rates may be of more importance than its apparent significance. It is extremely probable that some of the falls in isolation-rates are really due to an increase of incidence, so that the same percentage of cases cannot be met by the available hospital accommodation.

† It is, on the hypothesis of natural selection, a plausible view that the parasites—including under this term all disease organisms—which ultimately survive must tend to become innocuous to their hosts, and thus the decreasing virulence of certain diseases may be accounted for. The organism is destroyed owing to the death of the host or its own death at his recovery, or it has been modified by selection so as to become innocuous to its host relative to his immunity. But immunity is a matter of personal equation, and thus the function of the "carrier" in preserving and spreading a conceivably less nocuous form of the organism becomes clearer. We are not unaware of the view that the organism remains the same, but that the immunity is increased owing to "practise" of the leucocytes, but such a view requires the assumption of inheritance of acquired characters to explain reduced disease virulence, and further compels us to assume two types of immunity, the one which destroys the organism, and the other which without modifying it, establishes so to speak a mutual *modus vivendi*.

(iii) Are the districts which have adopted most isolation really urban districts where isolation was easiest to adopt and where possibly economic or social conditions favoured the spread of the disease or, in the case of the death-rate, the disease encountered a less resistant population?

(iv) What evidence is there to show that the districts which have rapidly increased their isolation-rates have subsequently lower attack- or death-rates?

If no one of these problems can be fully answered,—even in the case of a single disease—with the data at present available, at least light can be thrown on the lines which their solution in the future must take; and further something can be done to prevent hasty generalisation and excessive dogmatism as to the advantages or disadvantages of the isolation system. It can never be too strongly insisted upon, because it is so often forgotten, that preventive medicine is essentially an *experimental* science, and that in nine cases out of ten the efficiency of any line of action can only be adequately tested by statistics and by statistics collected after the expenditure of many thousand pounds, possibly spread over a long period of years, in carrying out this line of action*.

(2) *Material.* In endeavouring to throw some light on the above problems we have fortunately received data of very considerable value from Dr E. H. Snell, the Medical Officer of Health for the City of Coventry. He obtained for a period of nine years, 1904–1912 inclusive, for about eighty towns or districts of large population but of very varying local conditions, (i) the annual number of diphtheria cases, (ii) the number removed to hospital, (iii) the number of deaths. We have added to this material the estimated population of the town or district, and further certain data as to the economic and social conditions. Unfortunately there is no existing adequate measure of the general sanitary condition of individual towns, although the construction of a general sanitary "index number" would be of remarkable value in many forms of inquiry. We took as our measures of social condition:

(*a*) Death-rate of infants under a year.

(*b*) Amount of overcrowding, that is to say the percentage of the population in private families living more than two in a room.

(*c*) Density of population, i.e. the number of persons to the acre.

---

* Assert that it is most desirable to test the effect of sanatoria and of tuberculin in cases of tuberculosis, but do not dogmatically proclaim them as "cures" for phthisis, until statistics have been collected in sufficient amount and have been adequately and dispassionately examined to prove or disprove your statements. Insist on compulsory inoculation for enteric in the case of all recruits, but do not make it optional and then publish letters in the newspapers giving perfectly idle statistics, or go round to the camps giving popular lantern lectures to the recruits showing the gravestones of uninoculated persons, the portraits of persons dying of enteric, or much enlarged pictures of bacilli! If you think it experimentally worth doing, inoculate; but don't bring inoculation about by emphasising the dread of pain or the fear of death, both of which it is the first essential for a soldier wholly to disregard.

(*d*) Economic prosperity as measured by the number of indoor and outdoor servants of both sexes per 100 private families.

Our data are based on the census of 1911 as providing more ample information on these points. It will we think be admitted that the list of towns dealt with provides a very fair sample of the urban populations of this country. It ranges from manufacturing towns* like Preston, Rochdale and Bolton, mining and iron towns like Rhondda, Wigan and Middlesbrough, sea-ports like Hull, Liverpool and Southampton, to county towns like York and Reading, watering places like Brighton and Blackpool, suburban districts like Acton and Hornsey, and residential towns like Oxford or Bath. We ought from such a list to be able to throw some light on the relation of isolation to incidence under a variety of social conditions, if indeed these latter are factors in the problem at all.

(3) *What are the crude correlations between Isolation-Rate, Attack-Rate and Mortality-Rate?* The isolation-rate (I) has been measured as the average per cent. of cases removed to hospital during a five or four year period. We have two such periods, the earlier period 1904–1908 and the later period 1909–1912. The attack-rate has been measured per 1000 of the population, *uncorrected for age distribution.* Since diphtheria is largely a disease of infancy and childhood this neglect of the age correction—the reduction to a standard population—may seem serious. But in the first place we had not the age incidence in the individual districts, and in the second place we satisfied ourselves that such correction, if it could have been made, would not substantially modify any argument we have based on our data. For we calculated the attack-rate (A') on the population under 15 years of age, as well as the attack-rate (A) on the total population of the districts. We found the correlation between the two methods of measuring the attack-rate was $+\cdot972$, which indicates how close is the relation between the two methods of measuring the attack-rate and how little influence small variations in the proportion of less immune persons in the population due to age differences could have on the results†.

The *attack-rate* (A) has been measured as the number of cases per 1000 of the population. The *mortality-rate* has been measured in two different ways; first as the population mortality, the death-rate in the ordinary sense (M) or the deaths per 1000 of the population; and secondly the *case death-rate* or the mortality (m) per 100 attacked. We now give the crude correlations between I and A. They are:

First Period: 1904–1908, $r_{IA} = +\cdot427 \pm \cdot063$,
Second Period: 1909–1912, $r_{IA} = +\cdot290 \pm \cdot069$.

---

* See table, p. 567, for 76 of the 80 towns, the four others with full data only for the second period being Reading, Stoke, Dewsbury and Edmonton.

† The formula giving the juvenile attack-rate A' in terms of the crude attack-rate A is:

$$A' = 1\cdot3094\,A + \cdot0164$$

with a probable error of $\pm\cdot1369$.

Thirty-three towns were selected at random out of the 80, and gave the following results for A' calculated from A and A' as observed. The theoretical mean error $= \cdot162$; the mean error of the defects

Thus both periods show significant if not very large correlation*. The difference ($\cdot 137 \pm \cdot 093$) between the coefficients for the two periods is, however, probably not significant. Thus in towns with greater isolation-rate there is certainly a higher attack-rate, and equally certainly no argument can be based on the crude figures to prove that the more the isolation the less prevalent is diphtheria. We will now turn to the death-rate M, and we find:

$$\begin{array}{lll} \text{First Period:} & 1904\text{–}1908, & r_{IM} = +\cdot 153 \pm \cdot 075, \\ \text{Second Period:} & 1909\text{–}1912, & r_{IM} = -\cdot 012 \pm \cdot 075. \end{array}$$

In the first period isolation was associated with a higher diphtheria death-rate, in the second period with a lower diphtheria death-rate, but neither are of any real significance. Thus all we can conclude from the crude figures is that they show no evidence that isolation has reduced the general death-rate from diphtheria.

We next take the case death-rate (m) and we have for the two periods:

$$\begin{array}{lll} \text{First Period:} & 1904\text{–}1908, & r_{Im} = -\cdot 509 \pm \cdot 057, \\ & & r_{Am} = -\cdot 527 \pm \cdot 056, \\ \text{Second Period:} & 1909\text{–}1912, & r_{Im} = -\cdot 534 \pm \cdot 054, \\ & & r_{Am} = -\cdot 495 \pm \cdot 057. \end{array}$$

---

is − ·153 and of the excesses + ·134; this shows very fair accordance, 17 deviations being positive and 16 negative. The greatest deviations occur in Hornsey, Bath and Brighton, where residential neighbourhoods show fewer children, and in Edmonton, Walthamstow and Rhondda where there are probably excess of children. On the whole the general order is very well maintained, and the general attack-rate closely fixes the juvenile attack-rate. In any further collection of material, it would of course be well to have the age-distribution of cases.

	Observed	Calculated	Δ		Observed	Calculated	Δ
Derby ...	4·42	4·25	− ·17	Edmonton ...	1·36	1·80	+ ·44
Southampton ...	2·67	2·70	+ ·03	Bath ...	1·18	·87	− ·31
Hornsey ...	2·48	1·86	− ·82	Newport ...	1·14	1·25	+ ·11
Bristol ...	2·33	2·23	− ·10	Rhondda ...	1·10	1·34	+ ·24
Reading ...	2·13	1·95	− ·18	Bury ...	1·08	·91	− ·17
Nottingham ...	2·09	1·99	− ·10	Rotherham ...	1·06	1·18	+ ·12
Salford ...	2·04	2·16	+ ·12	Dewsbury ...	1·01	1·01	− ·00
Ilford ...	1·98	2·09	+ ·11	Blackburn ...	·99	·91	− ·08
Brighton ...	1·94	1·68	− ·26	Manchester ...	·94	·97	+ ·03
Stockton ...	1·90	2·12	+ ·22	Oxford ...	·89	·79	− ·10
Ipswich ...	1·87	1·89	+ ·02	Bolton ...	·86	·85	− ·01
Grimsby ...	1·85	1·90	+ ·05	Rochdale ...	·82	·75	− ·07
Walthamstow ...	1·85	2·10	+ ·25	Northampton ...	·78	·76	− ·02
Coventry ...	1·84	1·93	+ ·09	Barnsley ...	·75	·88	+ ·13
Plymouth ...	1·61	1·56	− ·05	Wigan ...	·58	·64	+ ·06
Wakefield ...	1·40	1·39	− ·01	W. Bromwich ...	·45	·53	+ ·08
Smethwick ...	1·38	1·55	+ ·17				

* We endeavoured to see whether the correlation of isolation- and attack-rates would be modified if we took the attack-rate on children under 15 years. This made little difference, r being raised only from $+\cdot 290 \pm \cdot 069$ to $+\cdot 315 \pm \cdot 068$.

According to these correlations, when or where the isolation-rate is high, the case mortality is low. Further when or where the attack-rate is high the case mortality is low. Now we know that:

$$\begin{aligned} I &= 100 \quad \times \text{isolated cases} \div \text{all cases}, \\ A &= 1000 \times \text{all cases} \div \text{population}, \\ m &= 100 \quad \times \text{deaths} \div \text{all cases}. \end{aligned}$$

Hence if we selected the number attacked at random and chose the deaths to be simply some number less than this, we should expect to find a considerable *negative* correlation between A and m; and as we actually do find such a correlation, we cannot be certain that the actually observed values of r_{Am} are not due to "spurious correlation." If they were "organic" we should interpret them to mean that a widespread epidemic (A large) was a less virulent epidemic (m small). On the other hand the spurious correlations of I and m would be *positive* in value, while the actual correlations are *negative*. Thus it would seem that while a high isolation-rate is associated with high attack-rate, it must be "organically" associated with a lessened case mortality. In other words while isolation does not, on the crude figures, appear to lessen the frequency of disease, it does appear to lessen the mortality among the attacked. This result appeared to be of such very great importance, if thoroughly established, that we determined to inquire into it further. It seemed reasonable to believe that the bulk of persons attacked might have better care in a hospital than in their own homes and thus isolation indirectly lessen the ill effects of the disease.

We accordingly endeavoured to approach the problem from a somewhat different standpoint: Given two districts with the same total number of persons attacked (a), will that district with more isolated (i), have fewer or more deaths (d)? The answer to this question depends on whether the partial correlation coefficient of total isolated cases with total deaths for constant number attacked is negative or positive. We found:

Correlations	First Period 1904–1908	Second Period 1909–1912
r_{id} = Isolated Cases and Deaths ...	$+{\cdot}860 \pm {\cdot}020$	$+{\cdot}867 \pm {\cdot}019$
r_{ia} = Isolated Cases and Attacked ...	$+{\cdot}937 \pm {\cdot}010$	$+{\cdot}968 \pm {\cdot}005$
r_{ad} = Attacked and Deaths	$+{\cdot}907 \pm {\cdot}014$	$+{\cdot}918 \pm {\cdot}012$
${}_ar_{id}$ = Isolated Cases and Deaths for constant number of Attacked	$+{\cdot}066 \pm {\cdot}077$	$-{\cdot}220 \pm {\cdot}072$

Thus in the first period for a given number of attacked more isolation was associated with more deaths, and in the second period for a given number of attacked, with fewer deaths; but in both periods, having regard to the probable errors, we cannot assert any real significance, or be reasonably certain that where there is more isolation, there recovery is more likely to occur.

We shall see later that the correlation between I and m for constant total number of attacks is not the same thing as the correlation of the total isolated and total deaths for constant total number of attacks. And this divergence, often in a marked degree, of partial correlations for rates and for absolute numbers is not unfamiliar to those who have had to deal with disease statistics. In the present case it renders still more obscure any argument drawn in favour of isolation from apparently lesser case mortality.

(4) *On the degree to which "spurious correlation" may be influencing the attack- and death-rates.* It seemed desirable if possible to throw further light on this point and accordingly we correlated attack- and death-rates with the total population. It will be remembered that:

$$A = 1000 \times \text{cases} \div \text{population},$$
$$M = 1000 \times \text{deaths} \div \text{population},$$

and accordingly if A and M be correlated with the population P, we might anticipate that if cases and deaths had no relation to population, there would be a high negative correlation arising from A and M both varying inversely as P. We were comforted by finding practically insignificant *positive* correlations. Thus:

Correlations	First Period 1904–1908	Second Period 1909–1912
r_{PA} = Population and Attack-rate	+ ·137 ± ·075	+ ·054 ± ·075
r_{PM} = Population and Death-rate	+ ·131 ± ·076	+ ·116 ± ·074
r_{PI} = Population and Isolation-rate	+ ·152 ± ·075	+ ·102 ± ·075

The last correlation cofficients show us that there is very little relation between the size of a population and the amount of isolation practised. Further these isolation correlations in which there is no obvious source of spurious correlation are as significant as those of population with attack- and death-rates where the possibility of "spurious correlation" is manifest. We conclude accordingly that risk is more uniformly distributed over population than we had anticipated, and that the correlations between the three rates I, A and M are really open to "organic" interpretation.

The next point which arises for discussion is whether the presence of the total number attacked (a) in the rates I and m can produce spurious correlation. If so we should anticipate that the absolute number a would be *negatively* correlated with both isolation and case mortality rates. We found:

Correlations	First Period 1904–1908	Second Period 1909–1912
r_{aI} = Total attacked and Isolation-rate	+ ·264 ± ·072	+ ·226 ± ·072
r_{am} = Total attacked and Case-Mortality	− ·251 ± ·072	− ·203 ± ·072

The first set of these coefficients are not even negative and therefore cannot be due to "spurious correlation," although such correlation may have reduced their organic values. They admit, however, of an easy interpretation, namely that: where the number attacked has been large the isolation has been more practised. The second set of coefficients might be due to spurious correlation, but they again admit of a simple interpretation as apart from "spurious correlation," namely that: when the attacks are numerous the deaths are relatively few, because a wide-spread epidemic means a mild epidemic. All four coefficients are significant, and pair and pair they are quite consistent but in no case are they of any marked importance. They enable us, however, to correlate the isolation-rate and the case-mortality for a constant total attacked, i.e. to find the partial correlation ${}_ar_{Im}$. We have the following results:

		First Period 1904–1908	Second Period 1909–1912
${}_ar_{Im} = {}_ar_{\frac{i}{a}\cdot\frac{d}{a}} =$	Correlation of Isolation-rate and Case mortality for constant number attacked	$-\cdot474 \pm \cdot056$	$-\cdot512 \pm \cdot057$

while we have already found:

${}_ar_{id} =$	Correlation of number isolated with number of deaths for constant number attacked	$+\cdot066 \pm \cdot077$	$-\cdot220 \pm \cdot072$

Correlation of Total Numbers Isolated and Total Registered Deaths.

Total Numbers Isolated.

Deaths Registered.	*0—250*	*250—500*	*500—750*	*750—1000*	*1000—1250*	*1250—1500*	*1500—1750*	*1750—2000*	*2000—2250*	*2250—2500*	*2500—2750*	*2750—3000*	*3000—3250*	Totals
0— 75	64	27	2	—	1	—	—	—	—	—	—	—	—	94
75—150	16	9	3	5	2	—	1	—	—	—	—	—	—	36
150—225	1	1	—	2	2	2	1	—	—	—	—	—	—	9
225—300	—	—	1	—	2	1	—	—	—	—	—	—	—	4
300—375	—	—	—	—	1	2	1	1	1	—	—	—	—	6
375—450	—	—	—	—	2	—	—	1	—	—	1	—	1	5
450—525	—	—	—	—	—	—	1	—	—	—	—	—	—	1
525—600	—	—	—	—	—	1	—	—	—	—	—	—	—	1
600—675	—	—	—	—	—	—	—	—	—	—	—	—	—	0
675—750	—	—	—	—	—	—	—	—	—	—	—	—	—	0
750—825	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Totals	81	37	6	7	10	6	4	2	1	0	1	0	2	157
Means	54·2	59·8	112·5	133·9	232·5	312·5	382·5 at 2125 isolated							103·4

It will therefore be clear that removing the variation in number attacked has made only slight reductions in the values of the correlation coefficients between isolation-rate and case-mortality. The discrepancy between the absolute numbers' and the rates' correlations is not to be accounted for by "spurious correlation" involved in the use of total numbers attacked in both rates. It must therefore be due to: (i) lack of linearity in certain of the regressions, (ii) high values in the coefficients of variation in certain of the quantities under discussion, or to a combination of these causes. With the small size of the populations under discussion it is by no means easy to test the true linearity of the regressions, even if we do what appears legitimate in this case, namely pool our data for the earlier and later periods. Our actual correlations have all been found without grouping by the direct product-moment formula, but we give on pp. 556 and 558 two grouped tables to illustrate the difficulties which arise in analysis. Our first table is for the total numbers isolated and the total deaths registered. It will be seen at once that the marginal distributions are intensely skew, crowding up into the corner of few deaths and few cases isolated, so that they appear to asymptote to the zero values of the coordinates. Further, Diagram I shows that the regression curve of deaths on total number isolated is, if just sensibly, still not markedly skew. Turning to the actual numbers given by this table we have the following series of constants:

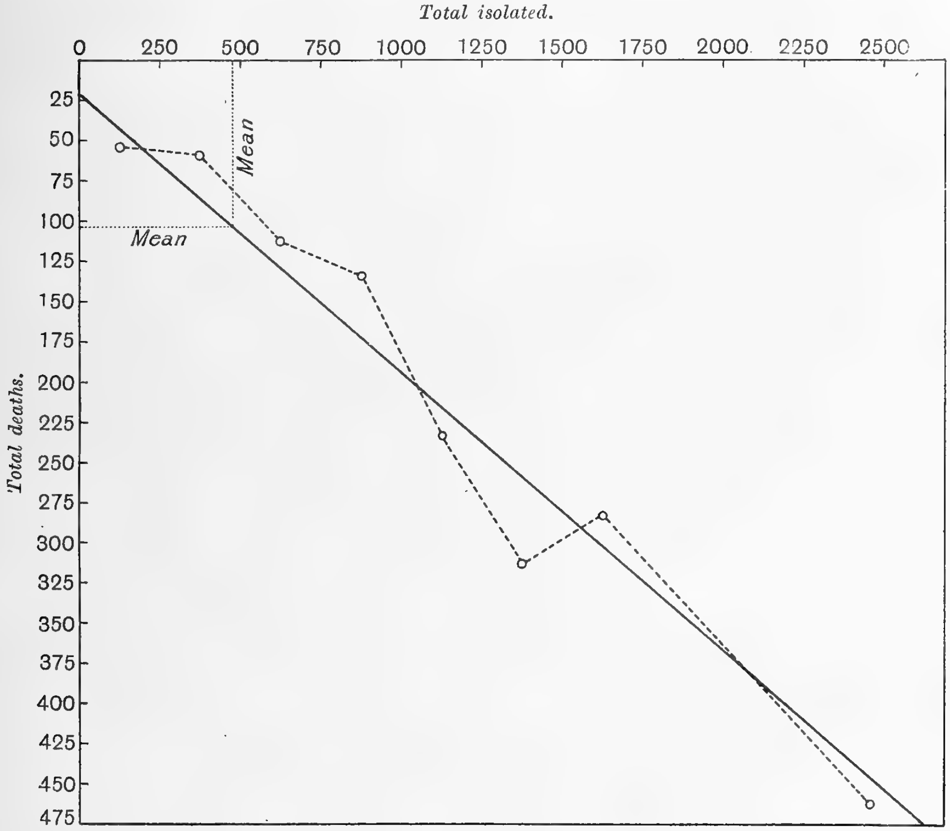


DIAGRAM I.

	Numbers Isolated (i)	Registered Deaths (d)
Mean	$\bar{\imath} = 475{\cdot}33$	$\bar{d} = 103{\cdot}42$
Standard Deviation	$\sigma_i = 571{\cdot}25$	$\sigma_d = 118{\cdot}61$
Coefficient of Variation ... ... (=S.D./Mean)	$v_i = 1{\cdot}20$	$v_d = 1{\cdot}15$
Correlation Coefficient and Ratio	$r_{id} = {\cdot}8348^* \pm {\cdot}0163$	$\eta_{di} = {\cdot}8564$ †

* Agrees reasonably well with the *non-grouped* values for the two *separate* periods.

† Found by taking means of all 13 column-arrays.

Clearly these results are of much interest; they show that the difference of η for deaths on isolation over r is not as great numerically as, perhaps, the graph suggests, but they indicate the markedly high values for the coefficients of variation. Now it is quite straightforward algebra to prove that

$$ {}_a r_{i/a,\, d/a} = {}_a r_{i,\, d}, $$

provided we may neglect terms of the square and product order in v_i and v_d compared with unity, and this is perfectly legitimate when these coefficients of variation are, as is usual in anthropometric measurements, quite small quantities. But in the present case these quantities are greater than unity and their squares are not negligible as compared with unity, thus we need not be surprised at the marked inequality of ${}_a r_{i/a,\, d/a}$* and ${}_a r_{i,\, d}$† found above. The values of the former show a marked relation between the case mortality and the isolation-rate, and the values of the latter indicate no appreciable betterment in the deaths due to increased isolation. Before we consider which of these coefficients gives us in the present case the better result as a guide to practical conduct, let us examine the correlation table for isolation-rate and case-mortality for the same 157 observations.

Correlation of Isolation-Rate I and Case-Mortality m.

Isolation-Rate.

Case-Mortality.	*0—10*	*10—20*	*20—30*	*30—40*	*40—50*	*50—60*	*60—70*	*70—80*	*80—90*	Totals
4— 8	1	—	—	—	1	7	6	5	8	28
8—12	2	—	3	3	4	8	11	9	5	45
12—16	5	4	3	8	3	8	6	6	1	44
16—20	2	—	2	1	2	3	3	1	2	16
20—24	9	—	2	1	2	—	2	1	—	17
24—28	3	—	1	2	—	—	—	—	—	6
28—32	1	—	—	—	—	—	—	—	—	1
Totals	23	4	11	15	12	26	28	22	16	157
Means	19·04	14·00	16·18	15·60	14·00	11·08	11·71	11·09	9·25	13·26

The following constants were found for this table:

	Isolation-Rate	Case Mortality
Mean	$\bar{I} = 49{\cdot}52$	$\bar{m} = 13{\cdot}26$
Standard Deviation	$\sigma_I = 25{\cdot}52$	$\sigma_m = 5{\cdot}58$
Coefficient of Variation	$v_I = 0{\cdot}52$	$v_m = 0{\cdot}42$
Correlation Coefficient and Ratio	$r_{Im} = -{\cdot}5291 \pm {\cdot}038$	$\eta_{mI} = {\cdot}5546$

The graph of the regression of case mortality on isolation-rate shows small evidence of skew regression (see Diagram II), and this is again confirmed by the difference between r_{Im} and η_{mI} being fairly small. The marginal frequency distributions show, however, considerable skewness, and that for the isolation-rate is lumped up at the end where there is no isolation: more than half the numerator of η_{mI} being contributed by the towns with little or no isolation. It is desirable to consider these towns further. They have an attack-rate of ·76, which is sensibly

---

* This is the ${}_a r_{Im}$ of our p. 556.

† The values are given on our p. 554.

less than the mean attack-rate (1·30), but they have a case-mortality of 19·04 as against the average case-mortality of 13·26; the 17 towns* with no isolation at all give a case-mortality of 19·4. It would thus appear that the towns with little or no isolation are those with a lower average attack-rate, but with rare exceptions their case-mortality is high.

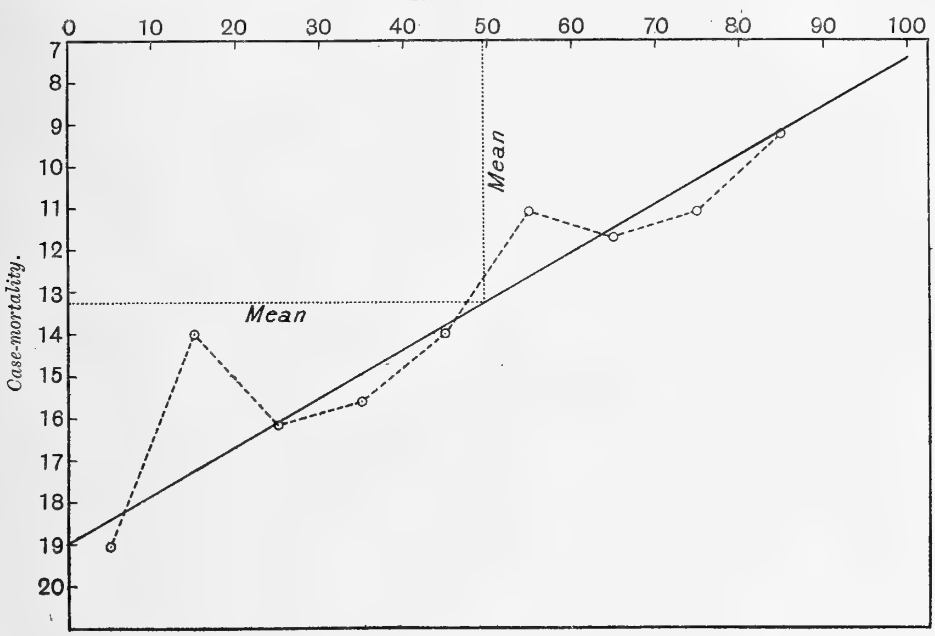


DIAGRAM II.

To test the influence of these towns with little or no isolation, we have removed the column 0—10 isolation-rate group and recalculated r_{Im} and η_{mI}; we find

$$r_{Im} = -\cdot 4120 \pm \cdot 0484, \quad \eta_{mI} = \cdot 4810.$$

Thus while the correlations are somewhat reduced by excluding the towns with little or no isolation there is still in the towns which do isolate a very sensible relation between the degree of isolation and the case-mortality, and this relation exhibits rather more skewness.

We may sum up as follows: The relation between greater isolation and a lessened case-mortality appears to be a real one. We have shown that it is hardly due to spurious correlation, as this would have produced a positive correlation and further no great changes are made when we correct for inequality in the numbers

---

* South Shields (1st and 2nd Periods), Sunderland (1st and 2nd Periods), Barrow (1st Period), Preston (1st Period), Wigan (1st Period), Smethwick (1st and 2nd Periods), Walsall (1st and 2nd Periods), West Bromwich (1st and 2nd Periods), Coventry (1st and 2nd Periods), Barnsley (1st and 2nd Periods). Of these towns West Bromwich in the 1st period had the highest case-mortality recorded of any of our 80 towns, while Smethwick in both periods, and Coventry and Barnsley in the 2nd period with no isolation had case-mortalities below the general average.

attacked. The regression is roughly linear and only very partially due to the high case-mortality in towns with no isolation. It is probable that where there is a large amount of isolation, the care of patients falls largely into the hands of a few men with a more extensive experience of the disease, and that this reduces the case-mortality.

Against this may be set the fact that the correlations between the absolute numbers of deaths and of cases isolated for constant numbers attacked are insignificant. The divergence between the two methods of approaching the problem is, however, explicable because the coefficients of variation of the absolute numbers are greater than unity, and the identity of the correlations reached by the two methods depends on the neglect of the squares and products of the coefficients of variation compared to unity. It may be asked: Why in this case we prefer the partial correlation found from the rates to that found from the absolute numbers? We reply: Because the partial correlation coefficient for the absolute numbers depends on very high total correlations, and if these correlations be, as we have shown, non-linear, then the partial correlation coefficient not only loses its full meaning, but may, as experience has shown us, easily change its sign as well as its magnitude. We would suggest that in a minor sense total mortalities and total isolations are bound to give "restricted tables," for deaths and isolated cases are perforce less than the numbers attacked, and that in such "restricted" tables, there is a general tendency to skew correlation and to a spurious factor*. On the other hand it is true that case-mortalities and isolation-rates cannot exceed $100\,^\circ/_\circ$ or fall short of $0\,^\circ/_\circ$, but these limits are the same for every array and do not vary from array to array as in the previous case. On the whole we think it safe to say that isolation is associated with greater prevalence of the disease and with a lessened case-mortality.

(5) *Is there any significant Relation between Isolation-Rate and General Diphtheria Death-Rate?* We have seen (p. 553) that insignificant correlations exist between I and M, and it is difficult to understand how a spurious factor could have modified this result. In the first place the small values of r_{PI} and r_{PM} on p. 555 show us that the value of ${}_Pr_{IM}$ is sensibly the same as r_{IM}; thus, for a constant population there is no sensible association between diphtheria mortality and isolation. But now let us ask whether for a constant attack-rate, isolation does not lessen general diphtheria mortality. We have:

Correlation	First Period	Second Period
r_{IM} = Isolation-rate and Death-rate ...	+·1532	−·0119
r_{IA} = Isolation-rate and Attack-rate ...	+·4268	+·2905
r_{MA} = Death-rate and Attack-rate ...	+·6772	+·6879
Hence		
${}_Ar_{IM}$ = Isolation-rate and Death-rate for constant Attack-rate	−·204 ± ·074	−·305 ± ·068

---

* See especially the illustrations of such "restricted" tables and their regression lines in a paper by Waite on Finger-Prints: *Biometrika*, Vol. x, pp. 421—478.

Both of these values may be considered significant and negative, and hence when the attack-rate is constant there is a sensible, if not very close relationship between increased isolation and reduced general mortality from diphtheria.

This confirms the view already reached that while isolation is associated with higher attack-rate its effect is to lessen the number of deaths whether they be reckoned as case-mortality or general population death-rate.

(6) *What is the meaning of the Association between Isolation and increasing prevalency of Diphtheria?* The analysis of this problem is more complicated. The obvious answer of those who advocate increasing isolation would be that it has been adopted in those districts where the disease is most prevalent, and this of course may turn out to be correct. But we may ask in turn upon what statistics they depend to demonstrate their view that isolation lessens the prevalence of the disease and is therefore advantageous, if our data demonstrate that where there is more isolation, there there is more diphtheria? It can only be by an analysis of no simple character that it is possible to deduce from such data that the practice of isolation has *lessened* the amount of the disease.

There is, however, a preliminary problem to be dealt with. The isolation-rate has been increasing very sensibly from 1904 to 1912, the attack-rate has lessened although very slightly, the case-mortality has lessened and the mortality on the population is considerably less. These facts are exhibited in the following table:

Variate	Symbol	Means		Standard Deviations	
		1904–1908	1909–1912	1904–1908	1909–1912
Attack-rate per 1000 population	A	1·33	1·28	·657	·639
Isolation-rate per 100 attacked	I	42·4	55·7	25·52	25·18
Mortality per 1000 population	M	·174	·138	·080	·061
Mortality per 100 attacked	m	14·6	12·1	5·72	5·01

Now it may well be, since the attack-rate has changed so little, that in the towns with increasing attack-rate there has been increasing isolation, both quantities changing with the time, but having no causal relation the one to the other. It is of some interest therefore to consider the type of districts in which isolation is most practised. In the first place we ask if any known bad social conditions are associated with prevalence of diphtheria. We took as our measure of sanitary conditions (i) the infant death-rate, or the deaths of children under one year per 1000 births, (ii) overcrowding, or the percentage of the population in private families with more than two in a room. We found the following results:

Variates Correlated	First Period 1904–1908	Second Period 1909–1912
Attack-rate and Infant Death-rate	$-·206 \pm ·074$	$-·206 \pm ·072$
Attack-rate and Overcrowding ...	$-·153 \pm ·075$	$-·136 \pm ·074$

These are not very considerable, but they are consistent, and indicate, as far as they go, that the incidence of diphtheria is not dependent upon such measures as the above of unfavourable sanitary conditions.

If we now turn to the correlation between the mortality-rate on the population and these measures of unfavourable sanitary conditions we find:

Variates Correlated	First Period 1904–1908	Second Period 1909–1912
Death-rate from Diphtheria and Infant Death-rate	+ ·081 ± ·076	+ ·118 ± ·074
Death-rate from Diphtheria and Overcrowding ...	+ ·061 ± ·079	+ ·004 ± ·075

All these are indeed positive, but they are of no significance and if they were significant would be so small as to be of no importance. The first indeed might have been anticipated to show a higher value, for a certain number of deaths from diphtheria must be deaths of infants. We can only conclude that as far as these measures of unsanitary conditions are concerned they do not in any way determine the diphtheria death-rate.

We now turn to the isolation-rate and find:

Variates Correlated	First Period 1904–1908	Second Period 1909–1912
Isolation-rate and Infant Deathrate ...	− ·414 ± ·064	− ·375 ± ·065
Isolation-rate and Overcrowding ...	− ·236 ± ·073	− ·235 ± ·071

These are significant although not very large and we conclude that most isolation is practised in those districts which have the lowest infant deathrate and the least overcrowding; the correlations are sensible if not very large. In other words the towns with better health conditions have adopted more extensively the practice of isolating diphtheria cases.

It seemed further of interest to determine: (i) whether diphtheria and isolation were more or less associated with urban conditions, and we took for this purpose the number of persons per acre, and (ii) whether the well-to-do character of the district, as measured by the number of domestic servants, indoor and outdoor, male and female per 100 private families, has any influence on the incidence of mortality from, or the isolation of diphtheria. We found:

Variates Correlated	First Period 1904–1908	Second Period 1909–1912
Persons per Acre and Attack-rate ...	+ ·165 ± ·075	+ ·043 ± ·075
Persons per Acre and Death-rate ...	+ ·169 ± ·075	+ ·115 ± ·074
Persons per Acre and Isolation-rate ...	+ ·073 ± ·076	+ ·053 ± ·075

Not one of these correlations is of any importance, if indeed any of them can be considered significant. It is thus clear that the intensity of urban conditions has very little to do with the prevalence of diphtheria, for if anything the suburban conditions have the lesser death-rate; clearly isolation has no sensible relation to number of persons per acre.

Turning now to our measure of the prosperity of the district, we find that it has no influence on the attack-rate, that it sensibly, but not very intensely affects the mortality, the higher death-rate occurring in the poorer districts, and that isolation is associated quite significantly with the prosperity of the district, i.e. the more well-to-do the district the more isolation is practised*.

Variates Correlated	First Period 1904–1908	Second Period 1909–1912
Number of Domestic Servants and Attack-rate ...	$+\cdot095 \pm \cdot076$	$+\cdot024 \pm \cdot075$
Number of Domestic Servants and Death-rate ...	$-\cdot219 \pm \cdot073$	$-\cdot308 \pm \cdot068$
Number of Domestic Servants and Isolation-rate	$+\cdot437 \pm \cdot062$	$+\cdot363 \pm \cdot065$

We conclude therefore that the more prosperous and generally healthier districts are associated with fuller isolation, and that the more prosperous, but not necessarily the more healthy districts, have the less diphtheria death-rate. On the other hand the incidence of the disease seems independent of the prosperity or density of population of the district and to be somewhat greater in those towns where the sanitary conditions as judged by infant death-rate and overcrowding are better†.

Thus as far as our measures go, we must conclude that diphtheria is not to be considered as a disease of markedly urban districts, of overcrowded or of insanitary districts‡. It would appear that the more prosperous and healthy districts have the greater isolation and that these are subject to somewhat the greater incidence.

---

* Of course this may largely mean that the more prosperous towns introduce isolation to remove the supposed danger of infection when servants of the families of the well-to-do are attacked.

† In order to ascertain whether the variates persons per acre (p_a) and overcrowding (O) were merely measures of the size of the town population (P) we correlated P with p_a and with O and found:

$$r_{Pp_a} = +\cdot404 \pm \cdot064 \text{ (1904–8)}, = +\cdot402 \pm \cdot063 \text{ (1909–12)},$$
$$r_{PO} = +\cdot091 \pm \cdot076 \text{ (1904–8)}, = +\cdot074 \pm \cdot075 \text{ (1909–12)}.$$

Thus overcrowding has no relation to the size of the town, the larger towns do not show more overcrowding. There is, however, a considerable association of persons per acre with total population, the larger towns having more persons per acre without exhibiting any more significant overcrowding. Making the population constant we find:

First Period 1904–1908		Second Period 1909–1912	
Total Correlation	Partial Correlation	Total Correlation	Partial Correlation
$r_{Ap_a} = +\cdot165 \pm \cdot075$ $r_{AO} = -\cdot153 \pm \cdot075$	${}_pr_{Ap_a} = +\cdot122 \pm \cdot076$ ${}_pr_{AO} = -\cdot167 \pm \cdot075$	$r_{Ap_a} = +\cdot043 \pm \cdot075$ $r_{AO} = -\cdot136 \pm \cdot074$	${}_pr_{Ap_a} = +\cdot023 \pm \cdot075$ ${}_pr_{AO} = -\cdot140 \pm \cdot074$

Thus correcting for population only makes the relation between persons per acre and incidence still more insignificant, while the relation between incidence and *less* overcrowding becomes slightly greater, without rising to any real importance.

‡ This result must not offhand be extended to subdistricts of our towns, it is an inter-urban and not intra-urban statement.

It will be seen at once that this conclusion opens up new problems: (i) Is the greater isolation the outcome of greater incidence, the only remedy suggestable for greater incidence being a more complete isolation? (ii) Is the greater incidence in some manner a result of the greater isolation, and does it really tell against isolation as a remedy against the spread of diphtheria? The association of greater isolation with local prosperity would then be merely a measure of the economic capacity of the district for carrying out the accepted sanitary code. (iii) If (ii) is to be answered in the negative, then is there any factor in prosperity which makes for a greater diphtheritic incidence? The final answers to these problems can probably not be given on the basis of the present data. The correlations under discussion although significant are not of such a marked character as to provide more than provisional statements, or indeed more than suggestions for further inquiry and tabulation.

(7) *Does greater Isolation follow increasing Incidence, or greater Incidence follow increasing Isolation?* The problem is a much more subtle one than appears at first sight. What we have established is that those towns with the higher isolation-rate have the higher attack-rate. It does not follow from this that the individual town which increases its isolation-rate will increase its attack-rate. To determine whether this is so we took as our variates: increase in isolation-rate ($\dot{I}$) between the periods 1904–8 and 1909–12*, and the similar increase $\dot{A}$ of the attack-rate. We found:

$$r_{\dot{A}\dot{I}} = + \cdot 256 \pm \cdot 072,$$

a value probably significant, although not quite so large as that found for the inter-urban relation:

$$r_{AI} = + \cdot 427 \pm \cdot 063 \text{ (for 1904–8)},$$
$$= + \cdot 290 \pm \cdot 069 \text{ (for 1909–12)}.$$

We can, we think, therefore conclude that the towns which increase their isolation-rate are those with increasing attack-rate, just as the towns with higher isolation-rate are those with higher attack-rate.

But this does not answer the question as to which is "the cart" and which "the horse"! Does the increased attack-rate precede or follow the increased isolation-rate? To answer this question we divided our material into three periods each of three years, let us say T_1, T_2 and T_3. Then the attack-rate increase between T_2 and T_1 was correlated with the isolation-rate in T_3, and the isolation-rate increase between T_2 and T_1 with the attack rate in T_3. In other words we asked whether towns with most rapid increase of attack-rate in the

---

* That is the total number treated in hospital × 100 and divided by the total number of attacks was taken for the first period and for the second period, and their difference (second period—first period) was treated as *increase* in isolation-rate. In the same way the sum of the totals attacked for the years of the first period × 1000 and divided by the sum of the calculated intercensal populations for the same years was treated as the attack-rate, and the difference of second and first period values taken as the increase in the attack-rate.

early periods had most isolation in the later period, and whether the towns with most rapid increase of isolation-rate in the earlier periods had most incidence in the later period. We found:

$$r_{\dot{A}_{2-1}I_3} = -\cdot 004 \pm \cdot 077,$$
$$r_{\dot{I}_{2-1}A_3} = +\cdot 085 \pm \cdot 077.$$

Thus there is no significant relation whatever between either increase of attack-rate or increase of isolation-rate in the first periods, and the isolation or the incidence in the following period.

As criticism of this result it might, perhaps, be suggested that the correlation of $\dot{A}_{2-1}$ and I_3 will be influenced by what has been the course of I in the periods T_1 and T_2 and the nature of A in T_3; we have accordingly, in order to test this, made the isolation-rate constant in the first two periods and the attack-rate constant in the third period and find

$$_{I_{1+2}A_3}r_{\dot{A}_{2-1}I_3} = +\cdot 147 \pm \cdot 076.$$

This is still of no real significance, although the sign appears to indicate that where the isolation-rate has been constant then increasing attack-rate in the earlier period is followed by very slightly more isolation in the third period, even if the attack-rate in that period be itself constant.

Similarly we determined:

$$_{A_{1+2}I_3}r_{\dot{I}_{2-1}A_3} = +\cdot 077 \pm \cdot 077.$$

This coefficient shows that towns which have increased their isolation-rate during a period of constant attack are not liable to sensibly heavier attack in the following period.

It would thus seem that our first two problems are both to be answered in the negative. Towns which increase their isolation are not those which in the following period have most incidence, nor are those which have increasing incidence markedly those of most isolation in the following period. Attack and isolation appear to have no causative relation, and the association we have found between more isolation and more incidence seems to be contemporaneous rather than successional. We are, it seems, compelled to search for something in the environment, which favours incidence and at the same time isolation. The only common factor that we have been able to reach at present is the prosperity and general healthy condition of the town. Under these circumstances there appears to be economic possibility of greater isolation, but why should there be greater incidence? Is it possible that in the more prosperous towns there is greater consumption of some easily contaminated commodities, which may act as carriers of the disease, or more concourse of those of susceptible ages at places of public amusement or instruction?

(8) *Test of the "organic" nature of the correlation of Isolation- and Attack-rates by the method of Variate Differences.* If the suggestion made at the end of the last section be correct we should anticipate that by the use of the method

of variate differences we should free ourselves from the influence of the time factor, if attack-rate and isolation-rate increase simultaneously in the more prosperous towns, but without organic association. We have nine years' returns, but the epidemic nature of diphtheria in many cases does not give one great confidence in applying the method of variate differences to *individual* years. We considered that it would not be wise to deal with smaller intervals than three-year periods, and should have preferred had the data been available to work with five-year intervals. As it is, we cannot with three-year intervals for each town go beyond the *second* differences. We have accordingly 228 isolation-rates and attack-rates obtained from 76 towns for each of three three-year periods, 152 first differences, and 76 second differences. We may symbolise them as I' and A', $\delta_1 I'$ and $\delta_1 A'$ and $\delta_2 I'$ and $\delta_2 A'$. We found the following results:

$$r_{I'A'} \quad = + \cdot 332 \pm \cdot 040,$$

$$r_{\delta_1 I' \delta_1 A'} = + \cdot 236 \pm \cdot 052,$$

$$r_{\delta_2 I' \delta_2 A'} = + \cdot 159 \pm \cdot 075.$$

The first of these results compares reasonably with the previous results for the first and second periods on p. 552, i.e.

$$1904\text{–}1908: \qquad r_{IA} = + \cdot 427 \pm \cdot 063,$$

$$1909\text{–}1912: \qquad r_{IA} = + \cdot 290 \pm \cdot 069,$$

with a mean value of + ·358. And this is the more true because the values of r_{IA} were found by the product moment method without grouping, while $r_{I'A'}$ was obtained from grouping in a correlation table*.

Now the above values bring out very markedly that when we endeavour to remove the influence of the time factor and to obtain a purely organic relationship between I and A, we more than halve the correlation between them by proceeding to the second difference only. If we might suppose that a hyperbola would give the asymptotic value of $r_{\delta_s I' \delta_s A'}$ from the above three known correlations we should have

$$r_{\delta_s I' \delta_s A'} = \frac{7 \cdot 084}{8 \cdot 105 + s} - \cdot 542,$$

which indicates, although no stress can be laid on actual numbers, that at about the fifth difference $r_{\delta_s I' \delta_s A'}$ would tend to become negative. All we think it possible to say would be that if the time factor be eliminated there is very little *positive* organic association between high isolation-rate and high attack-rate to be cleared up,—certainly not more than is indicated by the correlation on p. 565:

$${}_{I_{1+2}A_3} r_{\dot{A}_{2-1} I_3} = + \cdot 147 \pm \cdot 076,$$

---

* It may be noted that $r_{\delta_1 A' \delta_1 I'}$ was also found from a correlation table, but $r_{\delta_2 I' \delta_2 A'}$ as having only 76 cases by product moment without grouping.

which seems to suggest that other things being constant increasing incidence is to some slight extent followed—probably as the only suggested remedy—by the higher isolation rates*.

(9) *Can any other Factors be determined which measure the Relation between urban conditions and the Incidence of Diphtheria?* It is worth while from this standpoint to place the towns with which we have dealt in the order of incidence, each town being credited with the mean of the three attack-rates for each of three-year periods. Now an examination of the four columns of this table shows that, with the exception of Oxford—which has a child incidence (·89 as compared to ·70) considerably above the population incidence owing to relatively few children—the towns with the least diphtheria are the Midland, and particularly the Northern manufacturing towns. These constitute practically the whole of the first column of 19 towns. The last column contains the big ports and certain suburban metropolitan districts, indeed all these for which we have data except Plymouth, Devonport and Tottenham fall into the second half of the

Seventy-six Towns in order of their Diphtheria Incidence Rates 1904–1912.

1	West Bromwich	(·40)	20	Rotherham	(·91)	39	Birkenhead	(1·20)	58	Brighton	(1·61)
2	Northampton	(·45)	21	South Shields	(·92)	40	Rhondda	(1·24)	59	Stockton	(1·64)
3	Wigan	(·48)	22	Preston	(·93)	41	Smethwick	(1·25)	60	Grimsby	(1·66)
4	Walsall	(·49)	23	Wallasey	(·94)	42	Barrow	(1·25)	61	Leyton	(1·73)
5	Stockport	(·53)	24	Bath	(·95)	43	Newport	(1·25)	62	West Ham	(1·75)
6	Oldham	(·59)	25	Bootle	(·96)	44	Wimbledon	(1·30)	63	Salford	(1·78)
7	Bolton	(·59)	26	York	(·99)	45	Great Yarmouth	(1·31)	64	Nottingham	(1·79)
8	Oxford	(·70)	27	Blackpool	(·99)	46	Southend-on-Sea	(1·32)	65	St Helens	(1·80)
9	Barnsley	(·71)	28	Tynemouth	(1·00)	47	Birmingham	(1·32)	66	Walthamstow	(1·83)
10	Southport	(·72)	29	Tottenham	(1·02)	48	Gillingham	(1·34)	67	Ilford	(1·87)
11	Rochdale	(·73)	30	Halifax	(1·03)	49	Ipswich	(1·36)	68	Southampton	(1·95)
12	Leicester	(·76)	31	Sheffield	(1·03)	50	Liverpool	(1·37)	69	Cardiff	(2·03)
13	Manchester	(·79)	32	Plymouth	(1·07)	51	Hornsey	(1·39)	70	Enfield	(2·11)
14	Bury	(·80)	33	Coventry	(1·11)	52	Darlington	(1·39)	71	Hull	(2·16)
15	Blackburn	(·83)	34	Warrington	(1·11)	53	Acton	(1·43)	72	Bristol	(2·20)
16	Wolverhampton	(·86)	35	Devonport	(1·13)	54	Newcastle	(1·45)	73	Croydon	(2·41)
17	Burnley	(·86)	36	Sunderland	(1·14)	55	Burton on Trent	(1·48)	74	Portsmouth	(2·51)
18	Huddersfield	(·89)	37	Bournemouth	(1·17)	56	Bradford	(1·56)	75	Derby	(3·45)
19	Wakefield	(·91)	38	Middlesbrough	(1·20)	57	Willesden	(1·56)	76	Lincoln	(3·65)

---

* It is perhaps worth while putting on record the additional statistical constants obtained in deducing the above correlations, as they are probably fairly reliable values and should be compared with the two period constants on p. 561:

$\bar{A}'$ = Mean Attack-rate 1·26; Standard Deviation, Attack-rate ·655

$\bar{I}'$ = Mean Isolation-rate 47·75; Standard Deviation, Isolation-rate 26·341

$\overline{\delta_1 A'}$ = Mean Increase in Attack-rate − ·086; Standard Deviation of change in Attack-rate ·648

$\overline{\delta_1 I'}$ = Mean Increase in Isolation-rate 9·03; Standard Deviation of Increase in Isolation-rate 1·05

Thus while most towns have been sensibly increasing their amount of isolation by 17 °/₀ to 18 °/₀ of its mean value, the decrease in the attack-rate has only been 6 °/₀ to 7 °/₀ of the mean incidence, and the correlations show that this decrease has not occurred in the towns with marked increase of isolation.

table, and there can be little doubt that on the whole sea-port conditions and the big new neighbourhoods round London favour, while manufacturing conditions restrict, the incidence of diphtheria. We have not data, however, available upon which we could test water and milk supply, or extent of consumption of milk and fish in these towns. The results for Derby and Lincoln are remarkable, but they are high for all three periods, and this notwithstanding the rapid increase of isolation in those towns.

At first sight it seemed to us that the towns in the first column were markedly those in which there had been a greatly restricted birthrate*, while those in the last column were towns of greater fertility. Taking the births per 100 married women from 15 to 45 (B) we found:

$$r_{AB} = + \cdot 013 \pm \cdot 075.$$

Thus there is no association between incidence and the well-to-do character of a town as estimated by a low birthrate.

Again having regard to the character of the towns in our first column, it occurred to us to test the incidence in relation to the employment of males in manufacturing processes involving smoke. We took out of the 1911 census the percentage (S) of males over 10 years of age, who fell under a rough test of smoke-producing occupations, namely IX. 1, X. 1–2, 5–8, XIV. 1, XV. and XVIII. 1–6 of the Registrar-General's classification, and we found:

$$r_{AS} = - \cdot 180 \pm \cdot 073.$$

This is possibly significant and would undoubtedly be emphasised had we included as a factor the women engaged in textile industries. There seems therefore some slight reason to suppose that the conditions favourable to smoke production are unfavourable to the spread of diphtheria.

If the data could be procured, it would be worth while considering water and milk supply and the extent of fish consumption in the towns we have dealt with. If these were found to be of little influence, the road would certainly be clearer for dealing with the chronic diphtheritic human carrier as the chief source of the spread of the diphtheria bacillus.

(10) *Conclusions.*

(*a*) No influence of greater isolation in reducing the attack-rate from diphtheria is discoverable. In fact there is a sensible, if not large, *positive* association between the isolation-rate and attack-rate.

(*b*) The case mortality is somewhat less where there is more isolation. This may very probably be accounted for by more cases coming under specialised medical care.

---

* We had partially in view here also the possibility that restricted birthrate meant employment of women and thus less breast-feeding and greater use of milk, so that cross-currents might be at work.

(*c*) The attack-rate appears to be greater in the more prosperous towns and in towns of somewhat better sanitary conditions. We have not found the prevalence of diphtheria associated with overcrowding or with the conditions leading to high infant mortality.

(*d*) While a low birthrate, taken either as a measure of prosperity, or as a measure of the employment of women and so of the prevalence of hand feeding, appears to have no significance for the attack-rate of diphtheria, smoke-producing manufactures are probably unfavourable to the prevalence of the disease, which appears to attach itself in the main to the large ports and metropolitan suburban districts.

(*e*) The association between the attack- and isolation-rates observed is not very significant, and while it might, to a very small extent, be due to increased isolation following or accompanying increased attack, it is more probably an association due to the more prosperous towns practising more isolation, and also to there being some element in prosperity which assists the spread of the disease.

Generally all the correlations are of a low order; they contain, however, nothing to support the theory that isolation markedly limits the incidence of diphtheria; the disease itself does not appear where overcrowding is greatest nor where the population is most dense; on the other hand isolation is most practised in those towns where domestic servants are most common and which may be supposed to be most prosperous. The chief argument for isolation—*which can be drawn from the present data*—is a lessened case-mortality, but such mortality might be obtained in all probability by specialised medical service as apart from isolation.

MISCELLANEA.

I. On the Probable Error of a Coefficient of Mean Square Contingency.

By KARL PEARSON, F.R.S.

LET the sampled population be considered as to two variates and be represented by the total M and the cell-frequency m_{pq} for the pth row and qth column cell. Further let the vertical marginal frequencies be given by $m_{:q}$ and the horizontal marginal frequencies by $m_{p:}$, so that

$$m_{1q}+m_{2q}+\ldots+m_{pq}+\ldots=m_{:q},$$

$$m_{p1}+m_{p2}+\ldots+m_{pq}+\ldots=m_{p:}.$$

Let the corresponding quantities for the sample be N, n_{pq}, $n_{:q}$ and $n_{p:}$.

Then we know that the mean square contingency ϕ^2 is given by

$$\phi^2=S\left\{\frac{\left(N\dfrac{m_{:q}}{M}\dfrac{m_{p:}}{M}-n_{pq}\right)^2}{N\times N\dfrac{m_{:q}}{M}\dfrac{m_{p:}}{M}}\right\} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(i)},$$

summed for every cell.

Now in the great bulk of statistical phenomena we do not know more of the sampled population than is given by the sample, and thus to determine ϕ^2 we must put $\dfrac{m_{:q}}{M}$ and $\dfrac{m_{p:}}{M}$ equal to the most probable values known to us*, namely, $\dfrac{n_{:q}}{N}$ and $\dfrac{n_{p:}}{N}$. Doing this we obtain the usual value for the mean square contingency

$$\phi^2=S\left\{\left(\frac{n_{:q}n_{p:}}{N}-n_{pq}\right)^2\Big/n_{:q}n_{p:}\right\} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(ii)}.$$

Starting from (ii) Blakeman and Pearson have found† the probable error of the mean square contingency. The process is admittedly very laborious and although it has now been used fairly often, it must be confessed that its chief value is to obtain appreciation of the probable errors of contingency coefficients in general, rather than in any usefulness in recording significant differences between long series of individual coefficients.

But it has not been sufficiently recognised that the probable error thus found is that of the *approximate* value of the mean square contingency (ii) and not that of its true value (i). It is indeed the probable error of the expression actually used, but it is not the probable error of the true value as given by (i). The latter is easy to find and deserves consideration. Let us write

$$Nm_{:q}m_{p:}/M^2=\mu_{pq},$$

then

$$\phi_t^2=\frac{1}{N}S\left\{\frac{(\mu_{pq}-n_{pq})^2}{\mu_{pq}}\right\}=-1+\frac{1}{N}S\left\{\frac{n^2_{pq}}{\mu_{pq}}\right\},$$

---

* Pearson, *Philosophical Magazine*, Vol. L. 1900, pp. 164 *et seq.*

† *Biometrika*, Vol. v. pp. 191 *et seq.*

where we shall use ϕ_t^2 and ϕ_a^2 for the true and approximate values of the mean square contingency. Thus

$$1+\phi_t^2=\frac{1}{N}S\left(\frac{n^2_{pq}}{\mu_{pq}}\right)\quad\text{..(iii).}$$

Now for a sample of constant size μ_{pq} is constant and therefore representing small deviations by differentials

$$\delta\phi_t^2=\frac{2}{N}S\left(\frac{n_{pq}\,\delta n_{pq}}{\mu_{pq}}\right).$$

Square, add for all samples and divide by the number of such samples and we have

$$\sigma^2{}_{\phi_t^2}=\frac{4}{N^2}S\left(\frac{n^2{}_{pq}}{\mu^2{}_{pq}}\sigma^2{}_{n_{pq}}\right)+\frac{8}{N^2}\Sigma\left(\frac{n_{pq}n_{p'q'}}{\mu_{pq}\mu_{p'q'}}\times\sigma_{n_{pq}}\sigma_{n_{p'q'}}r_{n_{pq}n_{p'q'}}\right),$$

where $\sigma_{n_{pq}}$ is the standard deviation of n_{pq} and $r_{n_{pq}n_{p'q'}}$ is the correlation of deviations in n_{pq} and $n_{p'q'}$; S is a summation for every cell and Σ for every pair of cells.

But it is well known* that

$$\sigma^2{}_{n_{pq}}=N\frac{m_{pq}}{M}\left(1-\frac{m_{pq}}{M}\right)\chi,$$

$$\sigma_{n_{pq}}\sigma_{n_{p'q'}}r_{n_{pq}n_{p'q'}}=-N\frac{m_{pq}}{M}\frac{m_{p'q'}}{M}\chi,$$

where χ is the factor $1-(N-1)/(M-1)$, usually put unity, since M is as a rule large compared with N, and which will be here put unity for the remainder of the work.

Hence $$\sigma^2{}_{\phi_t^2}=\frac{4}{N}S\left(\frac{n^2{}_{pq}m_{pq}}{M\mu^2{}_{pq}}\right)-\frac{4}{N}S\left(\frac{n^2{}_{pq}m^2{}_{pq}}{M^2\mu^2{}_{pq}}\right)-\frac{8}{N}\Sigma\left(\frac{n_{pq}n_{p'q'}}{\mu_{pq}\mu_{p'q'}}\frac{m_{pq}m_{p'q'}}{M^2}\right)$$

$$=\frac{4}{N}S\left(\frac{n^2{}_{pq}m_{pq}}{M\mu^2{}_{pq}}\right)-\frac{4}{N}\left\{S\left(\frac{n_{pq}m_{pq}}{M\mu_{pq}}\right)\right\}^2\quad\text{..(iv).}$$

This is the standard deviation of the true value of the mean square contingency, and in most cases will be of no service, for we do not know the true values of m_{pq} and μ_{pq}.

If we put these equal to the values obtained from the actual sample under consideration we obtain the approximate value of the standard deviation of the true mean square contingency, which we may represent by the symbol $(\sigma_{\phi_t^2})_a$ and compare with what Blakeman and Pearson found, i.e. $(\sigma_{\phi_a^2})_t$. Thus our alternatives are

$$\phi_a^2\pm{\cdot}67449\,(\sigma_{\phi_t^2})_a,$$

and $$\phi_a^2\pm{\cdot}67449\,(\sigma_{\phi_a^2})_t.$$

The real thing is $$\phi_t^2\pm{\cdot}67449\sigma_{\phi_t^2}.$$

Shall we obtain a better insight into the variation of this by taking the approximate values of both ϕ_t^2 and $\sigma_{\phi_t^2}$, or by taking the probable error of ϕ_a^2? The problem is a subtle one, and, perhaps, only to be solved by experiment, not by theory. Of course when we take numerous samples and calculate ϕ_a^2, then $\sigma_{\phi_a^2}$ will measure their variability. But this is not what we seek. We use ϕ_a^2 as an approximation to ϕ_t^2, and it is the variability of the true value that we want. Are we not right in choosing $(\sigma_{\phi_t^2})_a$ as its best value? In short would not—on the average of a great number of samples—$(\sigma_{\phi_t^2})_a$ give us a closer result to $\sigma_{\phi_t^2}$ than $(\sigma_{\phi_a^2})_t$?

Returning now to equation (iv) and putting in the observed values for m_{pq}, μ_{pq} we have

$$(\sigma_{\phi_t^2})_a=\frac{4}{N}\left[S\left\{\frac{Nn^3{}_{pq}}{(n_{:q}n_{p:})^2}\right\}-\left\{S\left(\frac{n^2{}_{pq}}{n_{:q}n_{p:}}\right)\right\}^2\right]\quad\text{...........................(v).}$$

---

* The values here given are the true values before we approximate by putting $m_{pq}/M=n_{pq}/N$, etc.

Or, after some reductions

$$(\sigma_{\phi_t^2})_a=\frac{2}{\sqrt{N}}\{\psi_a^3+\phi_a^2-\phi_a^4\}^{\frac{1}{2}} \quad \text{.................................(vi)},$$

where

$$\psi_a^3=S\left\{\frac{N\left(n_{pq}-\frac{n_{:q}n_{p:}}{N}\right)^3}{(n_{:q}n_{p:})^2}\right\} \quad \text{.................................(vii)}$$

and

$$\phi_a^2=S\left\{\frac{\left(n_{pq}-\frac{n_{:q}n_{p:}}{N}\right)^2}{n_{:q}n_{p:}}\right\} \quad \text{....................................(viii)}.$$

Again we have

$$\sigma_{\phi_t^2}=2\phi_t\sigma_{\phi_t};$$

$$\therefore\quad (\sigma_{\phi_t})_a=\frac{1}{\sqrt{N}}\left\{\frac{\psi_a^3}{\phi_a^2}+1-\phi_a^2\right\}^{\frac{1}{2}} \text{..............................(ix)}.$$

Now what we usually need is the probable error of the contingency coefficient

$$C_2=\sqrt{\phi^2/(1+\phi^2)}.$$

But

$$\sigma_{C_2}=\sigma_\phi(1-C_2^2)^{\frac{3}{2}}=\sigma_\phi/(1+\phi^2)^{\frac{3}{2}}.$$

Thus the probable error of the coefficient of mean square contingency

$$\cdot67449\times\sigma_{C_2}=\frac{\cdot67449}{\sqrt{N}}\left\{\frac{\psi_a^3/\phi_a^2+1-\phi_a^2}{(1+\phi_a^2)^3}\right\}^{\frac{1}{2}} \quad \text{..........................(x)}.$$

This expression is much simpler than that for the probable error of the actually used value as given by Blakeman and Pearson*. It is not, however, asserted that it possesses greater theoretical validity. Those authors illustrate their formula by calculating the probable error of the contingency coefficient in the case of the association of handwriting and general intelligence in 1801 schoolgirls. They find

$$C=\cdot2957\pm\cdot0192.$$

In the course of their work they deduce

$$\phi_a^2=\cdot09580,$$
$$(\sigma_{\phi_a})_t=\cdot03268,$$
$$\psi_a^3=\cdot14865.$$

Using these values we have from (ix)

$$(\sigma_{\phi_t})_a=\frac{1}{\sqrt{1801}}\left\{\frac{\cdot14865}{\cdot09580}+\cdot90420\right\}^{\frac{1}{2}}=\cdot03693.$$

It is clear therefore that $(\sigma_{\phi_t})_a$ does not differ very substantially from $(\sigma_{\phi_a})_t$. Calculating from (x) the probable error of C_2, we find it $=\cdot0217$, while the Blakeman-Pearson process gave ·0192. The two values only differ by ·0025, which is unlikely to be of importance in the case of most inferences in practical statistics.

Beyond the knowledge of ϕ_a^2 only ψ_a^3 is required by the present process.

This may be written

$$\psi_a^3=\frac{1}{N}S\left\{\frac{\left(n_{pq}-\frac{n_{:q}n_{p:}}{N}\right)^2}{\frac{n_{:q}n_{p:}}{N}}\times\frac{n_{pq}-\frac{n_{:q}n_{p:}}{N}}{\frac{n_{:q}n_{p:}}{N}}\right\}.$$

In finding the mean square contingency ϕ_a^2, however, the three expressions

$$\frac{\left(n_{pq}-\frac{n_{:q}n_{p:}}{N}\right)^2}{\frac{n_{:q}n_{p:}}{N}},\quad n_{pq}-\frac{n_{:q}n_{p:}}{N}\quad\text{and}\quad\frac{n_{:q}n_{p:}}{N}$$

---

* *loc. cit.* p. 194.

must have been written down for each cell and thus $\psi_a{}^3$ can be readily calculated. We can also treat $\psi_a{}^3$ as given by

$$\psi_a{}^3 = S\left\{\frac{Nn^3{}_{pq}}{(n_{:q}n_{p:})^2}\right\} - 1 - 3\phi_a{}^2,$$

but the cubing of the often rather large cell frequencies is troublesome, just as it is rather more troublesome to calculate

$$\phi_a{}^2 = S\left(\frac{n^2{}_{pq}}{n_{:q}n_{p:}}\right) - 1$$

than

$$\phi_a{}^2 = \frac{1}{N} S\left\{\frac{\left(n_{pq} - \frac{n_{:q}n_{p:}}{N}\right)^2}{\frac{n_{:q}n_{p:}}{N}}\right\},$$

owing to the largeness of the squares in the former expression.

II. Measurements of Medieval English Femora.

In a forthcoming memoir on the English Long Bones there will be a good deal to be said about the conclusions reached by Dr Parsons in his recent paper on the Rothwell femora. Meanwhile he has started an attack on the Biometric School in a Journal whose columns are not open to adequate reply,—i.e. to a reply of not greater length than the published attack—from members of that school. In his communication he suggested that I was unacquainted with the condition of affairs at Rothwell, and behind this charge tried to escape any answers to the essential questions I asked him, and thus those questions still remain unanswered.

The communication I made ran as follows:

My informant who I hope is trustworthy speaks of (i) "the great mass of bones beneath the church at Rothwell" and (ii) of "the great collection of human bones beneath the old parish church at Rothwell"; further (iii) "there are probably some 5000 or 6000 individuals represented in the vault at Rothwell, either altogether or in part"; and again (iv) "The stack varies in height and breadth, but is nowhere as high or broad as that at Hythe, although it is much longer. I know that at Hythe there are the remains of rather over 4000 people,....... I think that this collection contains more than this, partly because the stack is so much longer, partly because the bones are so much more decomposed and have therefore settled more."

Manouvrier after much *piecing and mending* while only able to measure the lengths of about 16 femora from the *neolithic* burial places of Montigny and Esbly, was yet able to determine the pilastric index of 127, and the platymeric index in 127 bones, that is to say in *eight* times as many bones as those for which he could obtain the maximum length. And had he dealt fully with the head and neck and the popliteal region, the multiplying factor would probably have been ten. Had piecing, mending and a maximum of care in handling been used, I can hardly believe that what Manouvrier achieved at Montigny was not possible for Dr Parsons at Rothwell.

Dr Parsons writes: "If the remains of femurs, whether they are fit or unfit for measurement, are counted it will be found that females are quite as numerous as males though measurable male femurs from their stronger build are *less liable to break in being extricated from the pile of bones*, and so there are more of them available for measurement." The italics are mine.

Much depends on the method of 'extrication,' and if the capacity of a bone to stand a hole being drilled in it with a bradawl be part of the necessary fitness for measurement then the number might undoubtedly be limited. But trusting to what I know has been achieved by the

French, I feel convinced that if Dr Parsons could measure 277 femoral heads where the femoral length was measurable, he could easily have measured 2000 heads in all and thus have ascertained, definitely, whether his Rothwell series is unique in showing a significant depression in frequency between 45 and 47 mm. Further he could on such material by dealing with numbers 8 to 10 times those he has provided have given definite answers to many of the problems concerning platymery and the pilastric and popliteal indices, which other observers have been vainly trying to solve on far less adequate and in many cases far more fragmentary material.

I would note that Dr Parsons gives no reply at all to my question of why he used Dwight's measurements as a criterion of sex when they referred to bones *with the cartilages attached*, because without this reply his careful attention to 'other points' when the head fell between 45 and 47 mm. seems one-sided, and of no value in sexing the collection as a whole. He further gives no reply whatever to my question of why it is the male end, not the *dwarf* end, of his female distribution which is lacking, if absence of females be due to breakage.

I would also state (i) that I have not sexed the Rothwell bones and therefore cannot say how far I should or should not agree with Dr Parsons. Dr Lee using the best available mathematical process found 145 ♀s and 133 ♂s, while Dr Parsons has 103 ♀s and 174 ♂s. How this shows any agreement I fail to perceive; (ii) that I have made no assertion about the bones being of the 13th and 14th centuries. I merely headed my letter with Dr Parsons' heading "Measurements of Medieval English Femora," and asked why, if Dr Parsons holds these bones to be such, he considers them *without* cartilages comparable with the mixed results of a modern American dissecting room *plus* the cartilages.

K. P.

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS

Journal of Anatomy and Physiology

CONDUCTED BY

SIR WILLIAM TURNER, K.C.B. ARTHUR THOMSON, University of Oxford
ALEX. MACALISTER, University of Cambridge ARTHUR KEITH, Royal College of Surgeons
ARTHUR ROBINSON, University of Edinburgh

VOL. XLIX

ANNUAL SUBSCRIPTION 21/- POST FREE

CONTENTS OF PART III.—APRIL 1915

James C. Brash, M.A., M.B., B.Sc. Vertebral Column with six and a half Cervical and thirteen True Thoracic Vertebræ, with Associated Abnormalities of the Cervical Spinal Cord and Nerves. Douglas G. Reid, M.B., Ch.B. Edin., M.A. Trin. Coll. Camb. Appendices Epiploicæ and Pericolic Membranes. R. J. M'Connell, M.B. The Topographical and Applied Anatomy of the Gall-Bladder and Bile-Ducts. Thomas Walmsley, M.B. Observations on Certain Structural Details of the Neck of the Femur. Thomas Walmsley, M.B. The Neck of the Femur as a Static Problem. M. F. Lucas, M.B., B.S. (1) Two Cases of Cervical Ribs. (2) An Anomalous Arrangement of the Vagi. Anatomical Notes.

LONDON: CHARLES GRIFFIN AND COMPANY, Ltd., Exeter Street, Strand

---

JOURNAL OF THE ROYAL ANTHROPOLOGICAL INSTITUTE

Vol. XLIV. July—December, 1914

Contents:—

The Cheddar Man: A Skeleton of Late Palaeolithic Date. (With Plates XXIII—XXV.) By C. G. Seligman and F. G. Parsons. Cave Exploration at Gibraltar in 1912. By W. L. H. Duckworth, M.D., Sc.D., Cambridge. The Stability of Caste and Tribal Groups in India. By W. Crooke. The Orientation of the Dead in Indonesia. By W. J. Perry. Lois de Croissance. By Dr Paul Godin. The Prehistoric Pottery of the Canary Islands and its Makers. (With Plates XXVI—XXXV.) By Hon. John Abercromby, LL.D., F.S.A. Scot. Magic and Witchcraft on the Chōtā-Nāgpūr Plateau.—A Study in the Philosophy of Primitive Life. By Sarat Chandra Roy, M.A., B.L., Ranchi. Les Touareg du Sud. (With Plates XXXVI, XXXVII.) By Fr. de Zeltner. Flint-finds in connection with Sand. By Reginald A. Smith, F.S.A. Some Recent Work on Later Quaternary Geology and Anthropology, with its bearing on the question of "Pre-Boulder-Clay Man". By A. Irving, D.Sc., B.A. Notes on the Hypogeum at Hal-Saflieni, Malta. By Rev. H. J. Dukinfield Astley, M.A., Litt.D., F.R.Hist.S. Hasan and the Princess. By G. W. Murray. The Experimental Investigation of Flint Fracture and its application to Problems of Human Implements. (With Plates XXXVIII—XL.) By S. Hazzledine Warren, F.G.S. Miscellanea.

214 PAGES, WITH 18 PLATES AND MANY ILLUSTRATIONS IN THE TEXT.

Also Index and Table of Contents to Vol. XLIV.

PRICE 15*s.* NET

LONDON: THE ROYAL ANTHROPOLOGICAL INSTITUTE, 50, Great Russell Street, W.C.
General Agent: FRANCIS EDWARDS, 83a, High Street, Marylebone, W.
OR THROUGH ANY BOOKSELLER.

MAN

A MONTHLY RECORD OF ANTHROPOLOGICAL SCIENCE

Published under the direction of the Royal Anthropological Institute of Great Britain and Ireland. Each number of **MAN** consists of at least 16 Imp. 8vo. pages, with illustrations in the text together with one full-page plate; and includes Original Articles, Notes, and Correspondence; Reviews and Summaries; Reports of Meetings; and Descriptive Notices of the Acquisitions of Museums and Private Collections.

Price, 1*s.* Monthly or 10*s.* per Annum prepaid.

TO BE OBTAINED FROM THE
ROYAL ANTHROPOLOGICAL INSTITUTE, 50, Great Russell Street, W.C.
AND THROUGH ALL BOOKSELLERS

CONTENTS

(All Rights reserved)

The publication of a paper in *Biometrika* marks that in the Editor's opinion it contains either in method or material something of interest to biometricians. But the Editor desires it to be distinctly understood that such publication does not mark assent to the arguments used or to the conclusions drawn in the paper.

Biometrika appears about four times a year. A volume containing about 500 pages, with plates and tables, is issued annually.

Papers for publication and books and offprints for notice should be sent to Professor Karl Pearson, University College, London. It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted elsewhere without leave of the Editor. It is very desirable that a copy of all measurements made, not necessarily for publication, should accompany each manuscript. In all cases the papers themselves should contain not only the calculated constants, but the distributions from which they have been deduced. Diagrams and drawings should be sent in a state suitable for direct photographic reproduction, and if on decimal paper it should be *blue* ruled, and the lettering only pencilled.

Papers will be accepted in German, French or Italian. In the first case the manuscript should be in Roman not German characters.

Contributors receive 25 copies of their papers free. Fifty additional copies may be had on payment of 7/- per sheet of eight pages, or part of a sheet of eight pages, with an extra charge for Plates; these should be ordered when the final proof is returned.

The subscription price, payable in advance, is 30*s.* *net* per volume (post free); single numbers 10*s.* *net.* Volumes I—X (1902—15) complete, 30*s.* *net* per volume. Bound in Buckram 34/6 *net* per volume. Index to Volumes I to V, 2*s.* *net.* Subscriptions may be sent to C. F. Clay, Cambridge University Press, Fetter Lane, London, E.C., either direct or through any bookseller, and communications respecting advertisements should also be addressed to C. F. Clay.

Till further notice, new subscribers to *Biometrika* may obtain Vols. I—X together for £11 *net*—or bound in Buckram for £13 *net.*

The Cambridge University Press has appointed the University of Chicago Press Agents for the sale of *Biometrika* in the United States of America, and has authorised them to charge the following prices:—$7.50 net per volume; single parts $2.50 net each.

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS.