Vol. XIV. Parts I and II July, 1922

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

ISSUED BY THE BIOMETRIC LABORATORY
UNIVERSITY COLLEGE, LONDON
AND PRINTED BY THE
UNIVERSITY PRESS, CAMBRIDGE

Price Thirty Shillings net

[*Issued* 26 *July*, 1922]

JOURNAL OF ANATOMY

(ORIGINALLY THE JOURNAL OF ANATOMY AND PHYSIOLOGY)

CONDUCTED, ON BEHALF OF THE ANATOMICAL SOCIETY OF GREAT BRITAIN AND IRELAND, BY

Professor EDWARD FAWCETT, University of Bristol
Professor J. P. HILL, University of London
Professor ARTHUR ROBINSON, University of Edinburgh
Professor E. BARCLAY-SMITH, University of London
Professor Sir ARTHUR KEITH, Royal College of Surgeons

LINCOLN'S-INN-FIELDS, LONDON, W.C. 2

ANNUAL SUBSCRIPTION 40/- POST FREE

VOL. LVI

CONTENTS OF PARTS III AND IV.—APRIL AND JULY, 1922. PRICE **20/-** NET

CAMBRIDGE UNIVERSITY PRESS
Fetter Lane, London, E.C. 4: C. F. Clay, Manager

VOLUME XIV JULY, 1922 NOS. 1 AND 2

BIOMETRIKA

THE STANDARD DEVIATIONS OF FRATERNAL AND PARENTAL CORRELATION COEFFICIENTS.

BY KIRSTINE SMITH, D.Sc., Lond.

CONTENTS.

INTRODUCTION.

No attempts have been made as far as I know to calculate special formulae for the standard deviations of fraternal and parental correlation coefficients. The usual formula for the standard deviation of a correlation coefficient* which is deduced on the supposition that the values of the same variable are mutually uncorrelated is generally used also for this case, although it is only correct for a

* *Vide* Pearson and Filon: *Phil. Trans.* Vol. 191 A, p. 229, 1898.

fraternal correlation coefficient calculated from only two siblings of each family and for a parental correlation coefficient when only one offspring value from each family enters into the calculation. When the material of observation, as is usually the case in investigations of inheritance in higher mammals, consists of families of varying size, and correlation tables are used in which the same weight is given to each observed pair of siblings or pair of parent and offspring, without regard to the size of the family, a rational treatment of the probable error is excluded at the outset. With material in hand which makes it possible to examine numerous siblings, it is most reasonable to confine the investigation to a constant number of offspring from each family. In this case the deduction of formulae for the standard deviations of the two correlation coefficients does not present special difficulties, and this problem will be solved here.

We shall suppose that each group of q siblings belongs to the same litter or that from other reasons their order of birth is indifferent. Then each pair of siblings or each pair of parent and offspring ought to take a like part in the calculation, and q siblings give rise to $\frac{1}{2}q\,(q-1)$ pair of brothers and q pair of parent and offspring which all of them are entered in the calculation.

The fraternal correlation can thus be calculated either from a correlation table which is made symmetrical so that it contains $q\,(q-1)$ entries from each fraternity or by the formula quoted p. 10 which gives an identical result.

I. Fraternal Correlation.

Although this investigation aims especially at fraternal correlation it concerns of course other calculations of correlation in which the material consists of classes of equal size inside which the individuals are mutually correlated, all of them forming like parts. In the following we shall therefore name a group of siblings a *class*.

Suppose we have a material consisting of q individuals from each of n classes inside which the individuals are correlated while individuals from different classes are uncorrelated. We can then consider such a material as one of many possible samples of the same nature and size drawn from a population consisting of classes of individuals correlated as mentioned. It is therefore possible to face the problem of finding the law of errors for the *mean value*, the *standard deviation* of the character concerned and further for the *correlation coefficient* inside a class, supposing that these are all calculated from a sample like the one now considered.

Let the sample be $y_1, y_2, y_3 \ldots y_{nq}$ with mean value $\overline{y}$ and standard deviation σ. No special notation will be introduced for individuals of the same class, but summation of products is indicated by Σ when all factors of the product belong to the same class, and by S when factors of the same product belong to two or more classes. The summations always extend to all n classes.

(*a*) *The Mean Value.*

For the sample in hand we have

$$\bar{y} = \frac{1}{nq} \Sigma (y_1).$$

The mean value of $\bar{y}$ for a great number of samples coincides, according to the suppositions, with the mean of the population and this we choose for the zero point of y. The squared standard deviation of $\bar{y}$ is therefore found simply by squaring the expression above, summing for all the samples imagined and taking the mean value of the result. We thus find

$$\sigma_{\bar{y}}^2 = \frac{1}{n^2 q^2} \{\overline{\Sigma (y_1^2)} + 2\overline{\Sigma (y_1 y_2)} + 2\overline{S (y_1 y_2)}\},$$

where a bar above a summation indicates that the mean value has to be taken of the sums for all samples, i.e. for the population. Let the standard deviation of the population be s and the correlation coefficient for individuals of the same class r, we then have

$$\overline{\Sigma (y_1^2)} = nqs^2$$

and

$$\overline{\Sigma (y_1 y_2)} = \tfrac{1}{2} nq (q-1) rs^2.$$

As individuals of different classes are uncorrelated $\overline{\Sigma (y_1 y_2)}$ is equal to 0, and accordingly we find

$$\sigma_{\bar{y}}^2 = \frac{s^2}{nq} \{1 + (q-1) r\} * \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1).$$

This contains s and r for the population, which are, as a rule, only known from the sample in hand. It will be seen in the following, what is the approximation obtained by putting s and r equal to the values found from the sample.

(*b*) *The Mean Value of* σ^2*—the* presumptive *Standard Deviation.*

For our sample we find

$$\sigma^2 = \frac{1}{nq} \Sigma (y_1^2) - \bar{y}^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2).$$

By taking the mean of σ^2 for a great number of samples we find from this, remembering that the mean of $\bar{y}^2$ equals $\sigma_{\bar{y}}^2$,

$$\overline{\sigma^2} = s^2 \left(1 - \frac{1 + (q-1) r}{nq}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3).$$

When we take the value found for σ^2 as an approximation to $\overline{\sigma^2}$, we find accordingly the *presumptive* value of the standard deviation of the population by the formula

$${}_p\sigma^2 = s^2 = \sigma^2 \frac{nq}{nq - \{1 + (q-1) r\}},$$

which for $r = 0$ or $q = 1$ takes the form known for uncorrelated observations.

* *Vide Comptes-Rendus des Trav. du Lab. Carlsberg*, Vol. XIV. No. 11, 1921, Copenhagen, p. 32.

For the S.D. of $\bar{y}$ we find by introducing s in (1)

$$\sigma_{\bar{y}}^2 = \sigma^2 \frac{1+(q-1)r}{nq-\{1+(q-1)r\}}.$$

(c) *The Standard Deviation of* σ^2.

The S.D. of the σ^2 of our sample is found from $\sigma_{\sigma^2}^2 = \overline{\sigma^4} - (\overline{\sigma^2})^2$, where the latter term is already known. From (2) we find for the calculation of σ^4

$$n^2q^2\sigma^2 = (nq-1)\Sigma(y_1^2) - 2\Sigma(y_1y_2) - 2S(y_1y_2) \quad \ldots\ldots\ldots\ldots\ldots\ldots(4),$$

and from this

$$n^4q^4\overline{\sigma^4} = (nq-1)^2\overline{(\Sigma(y_1^2))^2} + 4\overline{(\Sigma(y_1y_2))^2} + 4\overline{(S(y_1y_2))^2} +$$
$$- 4(nq-1)\overline{\Sigma(y_1^2)\Sigma(y_1y_2)} - 4(nq-1)\overline{\Sigma(y_1^2)S(y_1y_2)} + 8\overline{\Sigma(y_1y_2)S(y_1y_2)} \ldots(5).$$

For the calculation of the mean values contained in this equation, the six products of product sums must be examined. We find

$$\left.\begin{aligned}
(\Sigma(y_1^2))^2 &= \Sigma(y_1^4) + 2\Sigma(y_1^2y_2^2) + 2S(y_1^2y_2^2) \\
(\Sigma(y_1y_2))^2 &= \Sigma(y_1^2y_2^2) + 2\Sigma(y_1^2y_2y_3) + 6\Sigma(y_1y_2y_3y_4) + 2S_s(y_1y_2y_3y_4)^* \\
\Sigma(y_1^2)\Sigma(y_1y_2) &= \Sigma(y_1^3y_2) + \Sigma(y_1^2y_2y_3) + S_s(y_1^2y_2y_3)
\end{aligned}\right\} \ldots(6).$$

When the multiplication of products containing the factor $S(y_1y_2)$ is carried out, it is clear that we need not consider such sums of products where the product contains a factor which is uncorrelated with all the other factors of the product, because the mean values of such product sums are 0. In the products $\Sigma(y_1^2)S(y_1y_2)$ and $\Sigma(y_1y_2)S(y_1y_2)$ all the sums of products are of this kind, the factors being distributed either in two classes of which one contains 3 and the other 1 factor or in three classes with respectively 2, 1 and 1 in each.

We therefore find

$$\left.\begin{aligned}
(S(y_1y_2))^2 &= S(y_1^2y_2^2) + 2S_s(y_1^2y_2y_3) + 4S_{s\ s}(y_1y_2y_3y_4) + \alpha_1 \\
\Sigma(y_1^2)S(y_1y_2) &= \alpha_2 \\
\Sigma(y_1y_2)S(y_1y_2) &= \alpha_3,
\end{aligned}\right\} \ldots\ldots\ldots(7),$$

where the mean values of the α's for the population are 0.

Let us denote the product moment corresponding to $y_1^m y_2^n y_3^p y_4^q$ by β_{mnpq} if all factors belong to the same class and in the opposite case let us insert 'd' or 's' as denoting different or same class.

* In the sums S all factors of a product are supposed to belong to different classes except those which are denoted by an 's' inserted between them, as belonging to the same class.

We find then

$$\left.\begin{aligned}
\overline{\Sigma(y_1^4)} &= nq\beta_4 \\
\overline{\Sigma(y_1^3 y_2)} &= nq(q-1)\beta_{31} \\
\overline{\Sigma(y_1^2 y_2^2)} &= \tfrac{1}{2}nq(q-1)\beta_{22} \\
\overline{\Sigma(y_1^2 y_2 y_3)} &= \tfrac{1}{2}nq(q-1)(q-2)\beta_{211} \\
\overline{\Sigma(y_1 y_2 y_3 y_4)} &= \tfrac{1}{24}nq(q-1)(q-2)(q-3)\beta_{1111} \\
\overline{S(y_1^2 y_2^2)} &= \tfrac{1}{2}n(n-1)q^2\beta_{\substack{2\,2\\d}} \\
\overline{S(y_1^2 y_2 y_3)} &= \tfrac{1}{2}n(n-1)q^2(q-1)\beta_{\substack{2\,1\,1\\d\,s}} \\
\overline{S(y_1 y_2 y_3 y_4)} &= \tfrac{1}{8}n(n-1)q^2(q-1)^2\beta_{\substack{1\,1\,1\,1\\s\,d\,s}}
\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (8).$$

Till now no suppositions have been made as to the law of distribution of the y's, but *in the following calculation we shall suppose that the distribution is normal and the correlation between individuals of the same class normal.*

For the general case of normal correlation between n variables the product moments have been determined by Sverker Bergström *. Taking the standard deviations as units of the variable and denoting the correlation coefficients by $r_{12}, r_{23}\ldots$, where for instance r_{23} means the correlation coefficient between the 2nd and 3rd variable of a product moment β'_{mnpq}, he finds the following formulae for the product moments of the 4th order :

$$\left.\begin{aligned}
\beta_4' &= 3 \\
\beta'_{31} &= \beta'_{13} = 3r_{12} \\
\beta'_{22} &= 2r_{12}^2 + 1 \\
\beta'_{211} &= 2r_{12}r_{13} + r_{23} \\
\beta'_{1111} &= r_{12}r_{34} + r_{13}r_{24} + r_{14}r_{23}
\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (9).$$

Substituting our special values for the correlation coefficient we find

$$\left.\begin{aligned}
\beta_4 &= 3s^4 \\
\beta_{31} &= 3rs^4 \\
\beta_{22} &= (2r^2+1)s^4 \\
\beta_{211} &= r(1+2r)s^4 \\
\beta_{1111} &= 3r^2s^4 \\
\beta_{\substack{2\,2\\d}} &= s^4 \\
\beta_{\substack{2\,1\,1\\d\,s}} &= rs^4 \\
\beta_{\substack{1\,1\,1\,1\\s\,d\,s}} &= r^2s^4
\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (10).$$

and further

We are now by means of (8) and (10) in a position to evaluate the mean values of the products put down under (6) and (7).

* *Vide* S. Bergström: *Biometrika*, Vol. XII. 1918, p. 177.

We find

$$\left.\begin{aligned}
\overline{(\Sigma(y_1^2))^2} &= nq\{nq+2+2(q-1)r^2\}s^4 \\
\overline{(\Sigma(y_1y_2))^2} &= \tfrac{1}{2}nq(q-1)\{1+2(q-2)r+[\tfrac{1}{2}nq(q-1)+q^2-3q+3]r^2\}s^4 \\
\overline{\Sigma(y_1^2)\Sigma(y_1y_2)} &= \tfrac{1}{2}nq(q-1)\{nq+4+2(q-2)r\}rs^4 \\
\overline{(S(y_1y_2))^2} &= \tfrac{1}{2}n(n-1)q^2\{1+(q-1)r\}^2s^4 \\
\overline{\Sigma(y_1^2)S(y_1y_2)} &= 0 \quad \text{and} \quad \overline{\Sigma(y_1y_2)S(y_1y_2)} = 0
\end{aligned}\right\} \quad (11).$$

The calculation of $\overline{\sigma^4}$ may now be continued. We find, by substituting the above mean values in (5),

$$n^2q^2\overline{\sigma^4} = s^4\{n^2q^2-1-2(nq+1)(q-1)r+(q-1)[2nq-(q-1)]r^2\}.$$

From (3) is found

$$n^2q^2(\overline{\sigma^2})^2 = s^4\{n^2q^2-2nq+1-2(nq-1)(q-1)r+(q-1)^2r^2\},$$

and accordingly

$$\sigma^2_{\sigma^2} = \overline{\sigma^4}-(\overline{\sigma^2})^2 = \frac{2s^4}{n^2q^2}\{nq-1-2(q-1)r+(q-1)(nq-q+1)r^2\},$$

or arranged according to powers of nq

$$\sigma^2_{\sigma^2} = \frac{2s^4}{nq}\left\{1+(q-1)r^2-\frac{1}{nq}[1+r(q-1)]^2\right\} \quad \ldots\ldots\ldots\ldots\ldots (12).$$

This formula for the S.D. of the squared standard deviations is thus exact, supposing that the correlation be normal.

For great values of n or rather of $\dfrac{nq}{1+(q-1)r^2}$ we may consider the S.D. of σ^2 a differential, so that

$$\sigma^2 = \overline{\sigma^2}+\delta\sigma^2 = \overline{\sigma^2}+2\sigma\delta\sigma.$$

From $\delta\sigma^2 = 2\sigma\delta\sigma$ we find by squaring and taking mean value for a great number of samples,

$$\sigma^2_{\sigma^2} = 4\sigma^2\sigma_\sigma^2,$$

and by substituting the value of $\sigma^2_{\sigma^2}$, omitting the last term,

$$\sigma_\sigma^2 = \frac{s^2}{2nq}\{1+(q-1)r^2\},$$

or, as with the accuracy obtainable we have

$$s^2 = \sigma^2,$$

it follows that:

$$\sigma_\sigma^2 = \frac{\sigma^2}{2nq}\{1+(q-1)r^2\}.$$

We notice when comparing this formula with (1) that only for $r=1$ and $r=0$ does the rule

$$\sigma_\sigma^2 = \tfrac{1}{2}\sigma_y^2$$

hold good.

The fraternal correlation coefficient ρ for the present sample is, when all the $\frac{1}{2}q(q-1)$ pairs of siblings are used for the calculation, defined by

$$\rho = \frac{\Pi}{\sigma^2},$$

where

$$\Pi = \frac{2}{nq(q-1)} \Sigma(y_1 y_2) - \overline{y^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13).$$

To determine the S.D. of ρ one requires in addition to $\sigma^2_{\sigma^2}$, the S.D. of Π and the product moment for Π and σ^2.

(*d*) *Mean Value and Standard Deviation of the Product Moment* Π.

Taking mean value of (13) for a great number of samples we find as

$$\Sigma(\overline{y_1 y_2}) = \tfrac{1}{2} nq(q-1) r s^2 \text{ and } (\overline{\bar{y}^2}) = \sigma_{\bar{y}}^2,$$

$$\overline{\Pi} = s^2 \left\{ r - \frac{1}{nq} [1 + (q-1) r] \right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14).$$

For calculating the mean value of Π^2 (13) may be written

$$n^2 q^2 (q-1) \Pi = -(q-1) \Sigma(y_1^2) + 2(nq - q + 1) \Sigma(y_1 y_2) - 2(q-1) S(y_1 y_2) \ldots (15),$$

from which follows

$$n^4 q^4 (q-1)^2 \overline{\Pi^2} = (q-1)^2 \overline{(\Sigma(y_1^2))^2} + 4(nq - q + 1)^2 \overline{(\Sigma(y_1 y_2))^2} +$$
$$- 4(q-1)(nq - q + 1) \overline{\Sigma(y_1^2) \Sigma(y_1 y_2)} + 4(q-1)^2 \overline{(S(y_1 y_2))^2},$$

the mean values of the two products being 0 according to (11). Substituting the rest of the values from (11) we find

$$(q-1) n^2 q^2 \overline{\Pi^2} = s^4 \{ 2nq - (q-1) + 2r[nq(q-3) - (q-1)^2]$$
$$+ r^2 [n^2 q^2 (q-1) - 2nq(q-2) - (q-1)^3] \} \ldots\ldots (16),$$

and by squaring (14) is found

$$(q-1) n^2 q^2 (\overline{\Pi})^2 = s^4 \{ q - 1 - 2r[nq(q-1) - (q-1)^2]$$
$$+ r^2 [n^2 q^2 (q-1) - 2nq(q-1)^2 + (q-1)^3] \}.$$

By subtraction of this equation from (16) we arrive at

$$\sigma^2_\Pi = \overline{\Pi^2} - (\overline{\Pi})^2 = \frac{2s^4}{nq(q-1)} \left\{ 1 - \frac{q-1}{nq} + 2r \left[q - 2 - \frac{(q-1)^2}{nq} \right] \right.$$
$$\left. + r^2 \left[q^2 - 3q + 3 - \frac{(q-1)^3}{nq} \right] \right\},$$

or arranged according to nq

$$\sigma^2_\Pi = \frac{2s^4}{nq(q-1)} \left\{ 1 + 2r(q-2) + r^2(q^2 - 3q + 3) - \frac{q-1}{nq} [1 + r(q-1)]^2 \right\},$$

which may also be written

$$\sigma^2_\Pi = \frac{2s^4}{nq(q-1)} \left\{ [1 + r(q-2)]^2 + r^2(q-1) - \frac{q-1}{nq} [1 + r(q-1)]^2 \right\} \ldots (17).$$

(*e*) *The Product Moment,* $\Pi_{\Pi\sigma^2}$, *of* Π *and* σ^2.

By multiplication of (4) and (15) and taking mean value for a great number of samples we find for the mean value of the product $\Pi\sigma^2$

$$n^4q^4(q-1)\,\overline{\Pi\sigma^2} = -(nq-1)(q-1)\,\overline{(\Sigma(y_1^2))^2} - 4(nq-q+1)\,\overline{(\Sigma(y_1y_2))^2}$$
$$+ 2\{n^2q^2 - nq^2 + 2(q-1)\}\,\overline{\Sigma(y_1^2)\,\Sigma(y_1y_2)} + 4(q-1)\,\overline{(S(y_1y_2))^2},$$

the mean values of the two products being zero according to (11). Introducing the rest of the mean values from (11), we have

$$n^2q^2\,\overline{\Pi\sigma^2} = s^4\{-nq - 1 + r[n^2q^2 - nq(q-4) - 2(q-1)] + r^2[nq(q-3) - (q-1)^2]\}.$$

From (3) and (14) is found

$$n^2q^2\,\overline{\Pi}\,.\,\overline{\sigma^2} = s^4\{-nq + 1 + r[n^2q^2 - nq^2 + 2(q-1)] + r^2[-nq(q-1) + (q-1)^2]\}.$$

As
$$\Pi_{\Pi\sigma^2} = \overline{\Pi\sigma^2} - \overline{\Pi}\,.\,\overline{\sigma^2},$$

it follows from the two foregoing equations that

$$\Pi_{\Pi\sigma^2} = \frac{2s^4}{n^2q^2}\{-1 + 2r[nq - (q-1)] + r^2[nq(q-2) - (q-1)^2]\},$$

or
$$\Pi_{\Pi\sigma^2} = \frac{2s^4}{nq}\left\{r[2 + (q-2)r] - \frac{1}{nq}[1 + (q-1)r]^2\right\} \quad \text{.....(18).}$$

(*f*) *The Standard Deviation of the Fraternal Correlation Coefficient.*

If the sample is great in proportion to $(q-1)r$ the errors of Π and σ^2 can be treated as differentials and we have for the correlation coefficient calculated from a sample

$$\rho = \frac{\overline{\Pi} + \delta\Pi}{\overline{\sigma^2} + \delta\sigma^2} = \frac{\overline{\Pi}}{\overline{\sigma^2}} + \frac{1}{\overline{\sigma^2}}\delta\Pi - \frac{\overline{\Pi}}{(\overline{\sigma^2})^2}\delta\sigma^2,$$

and
$$\bar{\rho} = \frac{\overline{\Pi}}{\overline{\sigma^2}} = \frac{r - \frac{1}{nq}\{1 + (q-1)r\}}{1 - \frac{1}{nq}\{1 + (q-1)r\}},$$

and therefore neglecting the term containing $\frac{1}{nq}$ which according to these suppositions cannot be evaluated

$$\bar{\rho} = r.$$

From $\delta\rho = \frac{1}{\overline{\sigma^2}}\left\{\delta\Pi - \frac{\overline{\Pi}}{\overline{\sigma^2}}\delta\sigma^2\right\}$ we find by squaring and forming mean value

$$\sigma^2_\rho = \frac{1}{(\overline{\sigma^2})^2}\left\{\sigma^2_\Pi + \left(\frac{\overline{\Pi}}{\overline{\sigma^2}}\right)^2\sigma^2_{\sigma^2} - 2\frac{\overline{\Pi}}{\overline{\sigma^2}}\Pi_{\Pi\sigma^2}\right\}.$$

When the values from (3), (12), (14), (17) and (18) are introduced in this formula and the terms containing the higher power of $\frac{1}{nq}$ are neglected, we get

$$\sigma_\rho^2 = \frac{2}{nq(q-1)}\{[1 + r(q-2)]^2 + r^2(q-1)\} + \frac{2r^2}{nq}\{1 + (q-1)r^2\} - \frac{4r^2}{nq}\{2 + (q-2)r\},$$

from which is found

$$\sigma_\rho^2 = \frac{2}{nq(q-1)}\{1 + r(q-2) - r^2(q-1)\}^2$$

and

$$\sigma_\rho = \sqrt{\frac{2}{nq(q-1)}}(1-r)\{1 + (q-1)r\} \quad \text{.....(19).}$$

For $q = 2$ this formula coincides with the usual formula for the standard deviation of a correlation coefficient calculated from two series of values of two variables corresponding in pairs, the values of each series being mutually uncorrelated.

(*g*) *Numerical Evaluation of the Formula for the* S.D. *of a Fraternal Correlation Coefficient.*

The number, N, of observed pairs of observations being equal to $\frac{1}{2}nq(q-1)$ the formula (19) may also be written

$$\sigma_\rho = \frac{1}{\sqrt{N}}(1-r)\{1 + (q-1)r\}.$$

Comparing materials of observations with different number of siblings q, we see that for the calculation of fraternal correlation information of each available pair of siblings has a value inversely proportional to $\{1 + (q-1)r\}^2$. The ratio $v_q = \left(\frac{1+r}{1+(q-1)r}\right)^2$ serves as a measure for the value which must be attributed to information of an observed pair among q siblings, supposed that all of the $\frac{1}{2}nq(q-1)$ pair of siblings are used for the calculation, and supposed that the value of information of a pair of siblings for $q = 2$ is put equal to 1. On the other hand $\frac{1}{v_q}$ indicates the ratio between the numbers of pairs of siblings which are required for obtaining the same accuracy in the correlation coefficient in the case of q and in the case of two siblings from each family. Table I gives the numerical values of v for different values of r and q.

TABLE I.

$$v_q = \left(\frac{1+r}{1+(q-1)r}\right)^2.$$

q	$r=0{\cdot}1$	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
2	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000
3	·840	·735	·660	·605	·563	·529	·502	·479	·460
4	·716	·563	·468	·405	·360	·327	·301	·280	·264
5	·617	·444	·349	·290	·250	·221	·200	·184	·171
6	·538	·360	·270	·218	·184	·160	·143	·130	·119
7	·473	·298	·216	·170	·141	·121	·107	·096	·088
8	·419	·250	·176	·136	·111	·095	·083	·074	·068
9	·373	·213	·146	·111	·090	·076	·066	·059	·054
10	·335	·184	·123	·093	·074	·063	·054	·048	·044

We notice that for values of r somewhat greater than 0·5, such as are usually found for mammals, v_3 has already decreased to about $\frac{1}{2}$ and v_4 to about $\frac{1}{3}$. By giving the same weight to each pair of siblings when forming fraternal correlation tables from a material consisting of fraternities of different size, we therefore fail very largely to pay due regard to the observations. With material under consideration, as for example anthropometric data, which according to its nature consists of small groups of siblings of varying number, and which is not so numerous that we can afford to omit observations from the calculation to make q constant for all fraternities, the rational proceeding must be to sort the material according to the number of siblings and calculate the correlation coefficient of each group separately.

It is then possible to effect considerable saving of time and labour in the investigation of correlation by avoiding the forming of fraternal correlation tables and using instead the formula

$$r = \frac{1}{q-1}\left(q\,\frac{\sigma_q^{\,2}}{\sigma^2} - 1\right)^*,$$

where σ_q is the directly calculated S.D. for mean values of fraternities. The results found by the formula are identical with those of the defining formula, so that the only objection to this method of calculation is the lack of opportunity to examine the shape of the regression curve.

From the correlation coefficients found for different values of q†, it is finally possible with knowledge of their S.D.'s to calculate a mean value of the fraternal correlation coefficient and its S.D.

In investigations of inheritance with animals with numerous offspring, where a great number of siblings are available, we have to face the problem of deciding what number of siblings it is profitable to employ for the investigation.

We shall state provisionally the problem as follows: with which value of q do we, provided the number of examined offspring individuals (nq) be fixed, obtain the most accurately determined fraternal correlation coefficient? Or in other words for which value of q is

$$\frac{1}{q-1}\{1 + r\,(q-1)\}^2 \text{ a minimum?}$$

* *Vide* K. Smith, *Comptes-Rendus des Trav. du Lab. Carlsb.* Vol. XIV. No. 11, 1921, p. 8, where the formula is deduced for the special case $q = 10$.

† In the memoir quoted it is shewn (p. 29) that the above formula may also be written

$$r = 1 - \frac{q}{q-1}\,\frac{\sigma^2_{f,q}}{\sigma^2},$$

$\sigma^2_{f,q}$ being the squared S.D. inside fraternities of q siblings and being calculated as a mean of such values obtained from each of the n fraternities. We may here instead of $\sigma_{f,q}$ introduce the presumptive S.D. inside a fraternity ${}_p\sigma_f$ that is the S.D. we expect to find in fraternities consisting of a great number of siblings. The relation is

$${}_p\sigma_f{}^2 = \frac{q}{q-1}\,\sigma^2_{f,q},$$

so that we find

$$r = 1 - \frac{{}_p\sigma_f{}^2}{\sigma^2},$$

which shews that the value of r arrived at must be expected to be independent of q.

The condition of minimum is

$$q = 1 + \frac{1}{r}.$$

Corresponding to the values $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$ and $\frac{3}{5}$ for r the values of q are 5, 4, 3 and $\frac{8}{3}$.

In examining the question of the most profitable number of siblings, attention must also be paid to the determination of the parental correlation and the question will therefore be further discussed in the following section. Besides it cannot be left out of consideration that, as a rule, it will be easier to examine the same number of individuals distributed among a smaller than among a greater number of fraternities. When regard only is had to fraternal correlation, the values of q obtained above must therefore be considered the minimum values.

For a more detailed illustration of the variation of the S.D. of the fraternal correlation coefficient with the number of siblings Table II has been calculated. The table gives the values of the S.D. for 1000 observations distributed among from 500 to 100 fraternities, the sizes of which therefore vary from 2 to 10.

TABLE II.

The Standard Deviation of a Fraternal Correlation Coefficient calculated from 1000 *observed Individuals.*

q	$r=\frac{1}{4}$	$r=\frac{1}{3}$	$r=\frac{1}{2}$	$r=\frac{3}{5}$
2	·0419	·0398	·0335	·0286
3	·0356	·0351	·0316	·0278
4	·0339	·0344	·0323	·0289
5	·0335	·0348	·0335	·0304
6	·0337	·0356	·0350	·0320
7	·0342	·0365	·0365	·0336
8	·0349	·0376	·0380	·0352
9	·0356	·0387	·0395	·0367
10	·0363	·0398	·0410	·0382

The table does not show a rapid increase of the S.D. when the number of siblings increases beyond the most profitable number found above. But a comparison of the values for $q=5$ and for $q=10$ still shows that the latter are respectively 8°/₀, 14°/₀, 22°/₀ and 25°/₀ greater than the former, so that when there are 10 siblings in each fraternity respectively 18°/₀, 31°/₀, 50°/₀ and 58°/₀ more individuals are required to obtain the same accuracy than when there are only 5 siblings from each family.

(*h*) *Application of the Formula to previous Calculations of Correlation.*

In an investigation* concerning the characters, *number of vertebrae* ('Vert.'), *number of rays in the pectoral fins* ('Pd.' and 'Ps.') and *number of pigment spots* ('Pigm.') in *Zoarces viviparus* from the station Nakkehage in Isefjord, Denmark,

* K. Smith, *Comptes-Rendus des Trav. du Lab. Carlsberg*, Vol. XIV. No. 11, 1921.

the fraternal correlation coefficient was calculated for 6 (for pigment spot only 5) samples from different years consisting of fraternities of 10 siblings. In this case the probable error of the fraternal correlation coefficient is according to (19)

$$\text{P.E.}\,(r) = \frac{0{\cdot}67449}{\sqrt{45n}}\,(1-r)\,(1+9r).$$

Table III gives for each sample the values of n, r and P.E. (r), as well as r for all the samples each weighted according to the S.D.

TABLE III.

Fraternal Correlation.

Year when sample taken	Vert.		Pd.		Pigm.	
	n	$r \pm$ P.E.	n	$r \pm$ P.E.	n	$r \pm$ P.E.
1914	138	0·4590 ± ·0238	132	0·3169 ± ·0231	—	—
1915	168	0·4693 ± ·0215	174	0·4196 ± ·0211	75	0·3175 ± ·0306
1916	123	0·5108 ± ·0248	122	0·3985 ± ·0251	87	0·3418 ± ·0289
1917	177	0·4715 ± ·0209	176	0·3634 ± ·0206	127	0·4112 ± ·0247
1918	153	0·4801 ± ·0225	156	0·3329 ± ·0215	113	0·3074 ± ·0247
1919	98	0·4066 ± ·0281	98	0·2893 ± ·0260	86	0·3722 ± ·0296
From total samples	—	0·4689 ± ·0095	—	0·3564 ± ·0092	—	0·3517 ± ·0122

For the mean values of r probable errors have previously been calculated based on the 6 or 5 values found. These probable errors had for

	Vert.	Pd.	and	Pigm. respectively
the values	0·0094	0·0137	and	0·0128,

which for Vert. and Pigm. agree extremely well with the theoretical values now found, while for Pd. the error had been estimated somewhat too great.

II. Parental Correlation.

For investigation of parental correlation we have a sample consisting as above of nq offspring values $y_1, y_2, y_3 \ldots\ldots y_{nq}$ distributed in n classes with q in each, and in addition, containing for each class an observed parental value x. We aim at finding the correlation between x and y's of the same class.

Let the parental correlation be r_p and the S.D. for x's s' in the population which we may imagine that the sample represents, and let us choose the mean value of the population as zero point for x.

The parental correlation coefficient is from the sample determined by

$$\rho_p = \frac{\Pi_{xy}}{\sigma\sigma'},$$

where σ' is the S.D. of x calculated from the sample, and Π_{xy} is the product moment for x and y determined by

$$\Pi_{xy} = \frac{1}{nq} \Sigma (x_1 y_1) - \bar{x}\,\bar{y} \quad \ldots\ldots(20).$$

As in the previous section Σ denotes a sum of products each of which consists of factors from the same class. In the sums S each product contains factors from at least two classes, and when two factors belong to the same class it is indicated by an 's' inserted between them.

For evaluation of the standard deviation of ρ_p the S.D. of Π_{xy}, σ and σ' are required, as well as the product moments for each pair of these three functions.

(*a*) *Mean Value and Standard Deviation of the Product Moment* Π_{xy}.

The equation (20) may also be written

$$\Pi_{xy} = \frac{n-1}{n^2 q} \Sigma (x_1 y_1) - \frac{1}{n^2 q} S (x_1 y_1) \quad \ldots\ldots(21).$$

By taking the mean value for a great number of samples we therefore find

$$\overline{\Pi}_{xy} = \frac{n-1}{n} r_p s s' \quad \ldots\ldots(22).$$

From (21) we find by squaring and taking mean value

$$n^4 q^2 \overline{\Pi^2_{xy}} = (n-1)^2 \overline{(\Sigma (x_1 y_1))^2} + \overline{(S (x_1 y_1))^2} - 2 (n-1) \overline{\Sigma (x_1 y_1)\, S (x_1 y_1)} \ldots(23).$$

Together with the determination of the mean values occurring here, we shall determine the other mean values of products required for the evaluation of σ_{ρ_p}. They are such as arise from multiplication of $\Sigma (x_1 y_1)$ and $S (x_1 y_1)$ with each of the two groups $\Sigma (y_1^2)$, $\Sigma (y_1 y_2)$, $S (y_1 y_2)$ and $\Sigma (x_1^2)$, $S (x_1 x_2)$ and also those which contain a factor of each of the two latter groups. As in the foregoing section, we need, however, not consider products of a Σ and an S, because such products may be developed into sums of products all containing a factor uncorrelated with all the other factors of the product, from which it follows that the mean value for a great number of samples is zero for each of these sums of products. It remains to determine the following products :

$$\left.\begin{aligned}
(\Sigma (x_1 y_1))^2 &= \Sigma (x_1^2 y_1^2) + 2\Sigma (x_1^2 y_1 y_2) + 2S (x_1 \underset{s}{y_1} x_2 \underset{s}{y_2}) \\
(S (x_1 y_1))^2 &= S (x_1^2 y_1^2) + 2S (x_1^2 \underset{s}{y_1\ y_2}) + 2S (x_1 \underset{s}{y_1} x_2 \underset{s}{y_2}) + \epsilon_1 \\
\Sigma (x_1 y_1)\, \Sigma (y_1^2) &= \Sigma (x_1 y_1^3) + \Sigma (x_1 y_1^2 y_2) + S (x_1 \underset{s}{y_1} y_2^2) \\
\Sigma (x_1 y_1)\, \Sigma (y_1 y_2) &= \Sigma (x_1 y_1^2 y_2) + 3\Sigma (x_1 y_1 y_2 y_3) + S (x_1 \underset{s}{y_1} y_2 \underset{s}{y_3}) \\
S (x_1 y_1)\, S (y_1 y_2) &= S (x_1 \underset{s}{y_1} y_2^2) + 2S (x_1 \underset{s}{y_1} y_2 \underset{s}{y_3}) + \epsilon_2 \\
\Sigma (x_1 y_1)\, \Sigma (x_1^2) &= \Sigma (x_1^3 y_1) + S (x_1^2 \underset{s}{x_2}\ y_1) \\
S (x_1 y_1)\, S (x_1 x_2) &= S (x_1^2 \underset{s}{x_2}\ y_1) + \epsilon_3 \\
\Sigma (x_1^2)\, \Sigma (y_1^2) &= \Sigma (x_1^2 y_1^2) + S (x_1^2 y_1^2) \\
\Sigma (x_1^2)\, \Sigma (y_1 y_2) &= \Sigma (x_1^2 y_1 y_2) + S (x_1^2 \underset{s}{y_1\ y_2}) \\
S (x_1 x_2)\, S (y_1 y_2) &= S (x_1 \underset{s}{y_1} x_2 \underset{s}{y_2}) + \epsilon_4
\end{aligned}\right\} \quad \ldots\ldots(24).$$

ϵ_1, ϵ_2, ϵ_3 and ϵ_4 are sums, the means of which are 0. The product moments are, as in the previous section, denoted by β and the indices concerning "x's" are placed in front of β, for instance $\frac{1}{nq}\overline{\Sigma(x^2y^2)}$ is denoted by ${}_2\beta_2$. We thus find for the mean values of the sums occurring in (24):

$$\left.\begin{aligned}
\overline{\Sigma(x_1^3y_1)} &= nq\,{}_3\beta_1\\
\overline{\Sigma(x_1^2y_1^2)} &= nq\,{}_2\beta_2\\
\overline{\Sigma(x_1^2y_1y_2)} &= \tfrac{1}{2}nq(q-1)\,{}_2\beta_{11}\\
\overline{\Sigma(x_1y_1^3)} &= nq\,{}_1\beta_3\\
\overline{\Sigma(x_1y_1^2y_2)} &= nq(q-1)\,{}_1\beta_{21}\\
\overline{\Sigma(x_1y_1y_2y_3)} &= \tfrac{1}{6}nq(q-1)(q-2)\,{}_1\beta_{111}\\
\overline{\underset{s}{S}(x_1^2x_2y_1)} &= n(n-1)\,q\,{}_{2\ 1}\beta_{1}\\
&\qquad\qquad {}_{d\ s}\\
\overline{S(x_1^2y_1^2)} &= n(n-1)\,q\,{}_2\underset{d}{\beta_2}\\
\overline{\underset{s}{S}(x_1^2y_1y_2)} &= \tfrac{1}{2}n(n-1)\,q(q-1)\,{}_2\beta_{1\ 1}\\
&\qquad\qquad {}_{d\ s}\\
\overline{\underset{s}{S}(x_1y_1\underset{s}{x_2}y_2)} &= \tfrac{1}{2}n(n-1)\,q^2\,{}_{1\ 1}\beta_{1\ 1}{}^{*}\\
&\qquad\qquad {}_{d\ s\ d}\\
\overline{\underset{s}{S}(x_1y_1y_2^2)} &= n(n-1)\,q^2\,{}_1\beta_{1\ 2}\\
&\qquad\qquad {}_{s\ d}\\
\overline{\underset{s}{S}(x_1y_1\underset{s}{y_2}y_3)} &= \tfrac{1}{2}n(n-1)\,q^2(q-1)\,{}_1\beta_{1\ 1\ 1}\\
&\qquad\qquad {}_{s\ d\ s}
\end{aligned}\right\}\ \ldots\ldots\ldots\ldots\ldots\ldots(25).$$

From Bergström's formulae (9) we find, when introducing r_p, r and 0 for the correlation coefficients and remembering that in his formulae s and s' are taken as units for y and x:

$$\left.\begin{aligned}
{}_3\beta_1 &= 3r_p s'^3 s\\
{}_2\beta_2 &= (2r_p{}^2+1)\,s'^2s^2\\
{}_2\beta_{11} &= (2r_p{}^2+r)\,s'^2s^2\\
{}_1\beta_3 &= 3r_p\,s's^3\\
{}_1\beta_{21} &= r_p(1+2r)\,s's^3\\
{}_1\beta_{111} &= 3rr_p s's^3\\
{}_{2\ 1}\beta_{1} &= r_p s'^3 s\\
{}_{d\ s}\qquad & \\
{}_2\underset{d}{\beta_2} &= s'^2s^2\\
{}_2\beta_{1\ 1} &= rs'^2s^2\\
{}_{d\ s}\qquad & \\
{}_{1\ 1}\beta_{1\ 1}{}^{*} &= r_p{}^2s'^2s^2\\
{}_{d\ s\ d}\qquad & \\
{}_1\beta_{1\ 2} &= r_p s's^3\\
{}_{s\ d}\qquad & \\
{}_1\beta_{1\ 1\ 1} &= rr_p s's^3\\
{}_{s\ d\ s}\qquad &
\end{aligned}\right\}\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (26).$$

* In this single case the notation fails, as it ought to be indicated that the first x and the last y belong to the same class.

Applying (25) and (26) we find for the mean values of the products under (24) the following values:

$$\left.\begin{aligned}
\overline{(\Sigma(x_1y_1))^2} &= nq\{q(n+1)r_p^2 + 1 + (q-1)r\}s'^2s^2 \\
\overline{(S(x_1y_1))^2} &= nq(n-1)\{qr_p^2 + 1 + (q-1)r\}s'^2s^2 \\
\overline{\Sigma(x_1y_1)\Sigma(y_1^2)} &= nq\{nq + 2 + 2(q-1)r\}r_p s's^3 \\
\overline{\Sigma(x_1y_1)\Sigma(y_1y_2)} &= \tfrac{1}{2}nq(q-1)\{2 + (nq+2q-2)r\}r_p s's^3 \\
\overline{S(x_1y_1)S(y_1y_2)} &= n(n-1)q^2\{1 + (q-1)r\}r_p s's^3 \\
\overline{\Sigma(x_1y_1)\Sigma(x_1^2)} &= nq(n+2)r_p s'^3s \\
\overline{S(x_1y_1)S(x_1x_2)} &= n(n-1)qr_p s'^3s \\
\overline{\Sigma(x_1^2)\Sigma(y_1^2)} &= nq(n+2r_p^2)s'^2s^2 \\
\overline{\Sigma(x_1^2)\Sigma(y_1y_2)} &= \tfrac{1}{2}nq(q-1)\{2r_p^2 + nr\}s'^2s^2 \\
\overline{S(x_1x_2)S(y_1y_2)} &= \tfrac{1}{2}n(n-1)q^2r_p^2s'^2s^2
\end{aligned}\right\} \quad \ldots\ldots(27).$$

We may now continue the calculation of $\overline{\Pi^2_{xy}}$. Introducing the mean values in (23), we get

$$n^2q\,\overline{\Pi^2_{xy}} = (n-1)\{nqr_p^2 + 1 + (q-1)r\}s'^2s^2.$$

From (22) we find

$$n^2q\,(\overline{\Pi}_{xy})^2 = (n-1)^2qr_p^2s'^2s^2,$$

and when this equation is subtracted from the foregoing

$$\sigma^2_{\Pi_{xy}} = \overline{\Pi^2_{xy}} - (\overline{\Pi}_{xy})^2 = \frac{n-1}{n^2q}\{qr_p^2 + 1 + r(q-1)\}s'^2s^2 \quad \ldots\ldots\ldots(28).$$

(*b*) *The Product Moment,* $\Pi_{\Pi_{xy},\,\sigma^2}$, *of* Π_{xy} *and* σ^2.

Multiplication of (4) and (21) gives

$$n^4q^3\Pi_{xy}\cdot\sigma^2 = (nq-1)(n-1)\Sigma(y_1^2)\Sigma(x_1y_1) - 2(n-1)\Sigma(x_1y_1)\Sigma(y_1y_2) + 2S(x_1y_1)S(y_1y_2) + \gamma_1,$$

where γ_1 consists of terms $S \times \Sigma$, the mean values of which are zero.

Taking the mean value and applying (27) we therefore find

$$n^3q^2\,\overline{\Pi_{xy}\cdot\sigma^2} = (n-1)r_p\{nq(nq+1) + rnq(q-1)\}s's^3.$$

For $\overline{\Pi_{xy}}\cdot\overline{\sigma^2}$ we get from (3) and (22)

$$n^3q^2\,\overline{\Pi_{xy}}\cdot\overline{\sigma^2} = (n-1)r_p\{nq(nq-1) - rnq(q-1)\}s's^3,$$

and accordingly from the two latter equations

$$\Pi_{\Pi_{xy},\,\sigma^2} = \overline{\Pi_{xy}\cdot\sigma^2} - \overline{\Pi_{xy}}\cdot\overline{\sigma^2} = \frac{2(n-1)}{n^2q}r_p\{1 + r(q-1)\}s's^3 \ldots\ldots \quad (29).$$

(*c*) *The Product Moment,* Π_{Π_{xy},σ'^2}, *of* Π_{xy} *and* σ'^2.

For σ'^2 we obtain, from the formulae (4), (3) and (12), which concern σ^2, by substituting x for y and putting q equal to 1:

$$\sigma'^2 = \frac{n-1}{n^2}\Sigma(x_1^2) - \frac{2}{n^2}S(x_1x_2) \quad \text{.....(30),}$$

$$\overline{\sigma'^2} = \frac{n-1}{n}s'^2 \quad \text{.....(31),}$$

and

$$\sigma^2_{\sigma'^2} = \frac{2(n-1)}{n^2}s'^4 \quad \text{.....(32).}$$

By multiplication of (21) and (30) we get

$$n^4q\,\Pi_{xy}\,.\,\sigma'^2 = (n-1)^2\Sigma(x_1y_1)\Sigma(x_1^2) + 2S(x_1y_1)S(x_1x_2) + \gamma_2,$$

for which the mean value by application of (27) is found to be

$$n^2\,\overline{\Pi_{xy}\,.\,\sigma'^2} = (n^2-1)\,r_p s'^3 s.$$

From (22) and (31) follows

$$n^2\,\overline{\Pi_{xy}}\,.\,\overline{\sigma'^2} = (n-1)^2\,r_p s'^3 s,$$

so that

$$\Pi_{\Pi_{xy},\sigma'^2} = \overline{\Pi_{xy}\,.\,\sigma'^2} - \overline{\Pi_{xy}}\,.\,\overline{\sigma'^2} = \frac{2(n-1)}{n^2}\,r_p s'^3 s \quad \text{.....(33).}$$

(*d*) *The Product Moment,* Π_{σ^2,σ'^2}, *of* σ^2 and σ'^2.

For the product $\sigma^2\sigma'^2$ we find by multiplying (4) and (30):

$$n^4q^2\sigma^2\sigma'^2 = (nq-1)(n-1)\Sigma(x_1^2)\Sigma(y_1^2) - 2(n-1)\Sigma(x_1^2)\Sigma(y_1y_2) + 4S(x_1x_2)S(y_1y_2) + \gamma_3.$$

The mean value of γ_3 is zero, and therefore by taking the mean value and using (27) we get

$$n^2q\,\overline{\sigma^2\sigma'^2} = (n-1)\{nq - 1 + 2qr_p^2 - (q-1)r\}\,s'^2s^2,$$

and when from this is subtracted

$$n^2q\,\overline{\sigma^2}\,\overline{\sigma'^2} = (n-1)\{nq - 1 - (q-1)r\}\,s'^2s^2,$$

we arrive at

$$\Pi_{\sigma^2\sigma'^2} = \frac{2(n-1)}{n^2}\,r_p^2s'^2s^2 \quad \text{.....(34).}$$

(*e*) *The Standard Deviation of the Parental Correlation Coefficient.*

For the logarithm of the parental correlation coefficient calculated from our sample we have

$$\log\rho_p = \log\Pi_{xy} - \tfrac{1}{2}\log\sigma^2 - \tfrac{1}{2}\log\sigma'^2.$$

For great values of n, which allow us to treat the deviations of σ'^2, σ^2 and Π_{xy} from their mean values as differentials, it follows from the above equation by differentiation that

$$\frac{\delta\rho_p}{\rho_p} = \frac{\delta\Pi_{xy}}{\Pi_{xy}} - \tfrac{1}{2}\frac{\delta\sigma^2}{\sigma^2} - \tfrac{1}{2}\frac{\delta\sigma'^2}{\sigma'^2} \quad \text{.....(35).}$$

With the accuracy here employed, which excludes the determination of terms containing the higher power of $\frac{1}{n}$, we have

$$\rho_p = \overline{\rho_p} = \frac{\overline{\Pi}_{xy}}{\sqrt{\overline{\sigma^2} \cdot \overline{\sigma'^2}}} = r_p .$$

From (35) we find by squaring and taking the mean value for a great number of samples

$$\sigma^2_{\rho_p} = r_p^2 \left\{ \frac{\sigma^2_{\Pi_{xy}}}{(\overline{\Pi}_{xy})^2} + \tfrac{1}{4} \frac{\sigma^2_{\sigma^2}}{(\overline{\sigma^2})^2} + \tfrac{1}{4} \frac{\sigma^2_{\sigma'^2}}{(\overline{\sigma'^2})^2} - \frac{\Pi_{\Pi_{xy}\sigma^2}}{\overline{\Pi}_{xy}\overline{\sigma^2}} - \frac{\Pi_{\Pi_{xy}\sigma'^2}}{\overline{\Pi}_{xy} \cdot \overline{\sigma'^2}} + \tfrac{1}{2} \frac{\Pi_{\sigma^2\sigma'^2}}{\overline{\sigma^2}\,\overline{\sigma'^2}} \right\},$$

which by introducing the values from (3), (12), (22), (28), (29), (31), (32), (33) and (34) and neglecting the term containing the higher power of $\frac{1}{n}$ leads to

$$\sigma^2_{\rho_p} = \frac{1}{nq} \{1 + r(q-1) + q r_p^2\} + \frac{r_p^2}{2nq} \{1 + (q-1) r^2\} + \frac{r_p^2}{2n} - \frac{2 r_p^2}{nq} \{1 + (q-1) r\} - \frac{2 r_p^2}{n} + \frac{r_p^4}{n},$$

or

$$\sigma^2_{\rho_p} = \frac{1}{nq} \left\{ 1 + (q-1) r - \frac{r_p^2}{2} [q + 3 + (q-1) r (4 - r)] + q r_p^4 \right\},$$

which may be written

$$\sigma^2_{\rho_p} = \frac{(1 - r_p^2)^2}{n} - \frac{q-1}{nq} (1 - r) \left\{ 1 - r_p^2 \frac{3 - r}{2} \right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (36).$$

The first term is the usual expression obtained for $q = 1$. From this, for $q > 1$, one must subtract a term which for given values of r and r_p increases with q.

(*f*) *The Standard Deviation of the Slopes of the Regression Curves.*

We shall finally add the formulae of the S.D. of the slope of the regression curves for the calculation of which we have all the material ready. The regression coefficients are determined by

$$\alpha_p = \frac{\Pi_{xy}}{\sigma'^2} \text{ and } \alpha_a = \frac{\Pi_{xy}}{\sigma^2}.$$

By differentiation, squaring and taking mean value, we find

$$\sigma^2_{\alpha_a} = \alpha_a^2 \left\{ \frac{\sigma^2_{\Pi_{xy}}}{(\overline{\Pi}_{xy})^2} + \frac{\sigma^2_{\sigma^2}}{(\overline{\sigma^2})^2} - 2 \frac{\Pi_{\Pi_{xy}\sigma^2}}{\overline{\sigma^2}\,\overline{\Pi}_{xy}} \right\},$$

and a corresponding equation for $\sigma^2_{\alpha_p}$.

From these we find by introducing the S.D. and product moments

$$\sigma^2_{\alpha_p} = \frac{s^2}{nqs'^2} \{1 + r(q-1) - q r_p^2\} *,$$

and

$$\sigma^2_{\alpha_a} = \frac{s'^2}{nqs^2} \{1 + r(q-1) - q r_p^2 + 2(q-1) r_p^2 (1 - r)^2\}.$$

* *Vide* K. Smith, *l.c.*, pp. 6, 7, where the same formula is deduced in a different form, containing σ_q instead of r. The two expressions are easily seen to be identical when the term $\frac{1}{n}$ is neglected.

(*g*) *Numerical Evaluation of the Formula for the* S.D. *of a Parental Correlation Coefficient.*

We shall first examine how valuable a material consisting of n groups of q siblings with corresponding parental values is compared with nq pairs of values from different families. Denoting the S.D.'s of ρ_p calculated from the two materials by $\sigma_{q\rho_p}$ and $\sigma_{1\rho_p}$ we find by applying (36)

$$v_q' = \frac{\sigma_1^2\rho_p}{\sigma_q^2\rho_p} = \frac{(1-r_p^2)^2}{q(1-r_p^2)^2-(q-1)(1-r)\left\{1-r_p^2\frac{3-r}{2}\right\}} \quad \ldots\ldots\ldots(37).$$

This ratio indicates the value of an observed pair, when the parental value also occurs combined with $(q-1)$ other offspring values, in proportion to the value of an observed pair when the parental value only occurs once in the calculation.

The numerical values of (37) are, for values of r_p and r, fairly well representative of the values met with in investigations of inheritance given in Table IV.

TABLE IV.

$v_q' = \sigma^2{}_{1\rho_p} : \sigma^2{}_{q\rho_p}$.

q	$r_p = \cdot3$ $r = \cdot4$	$r_p = \cdot4$ $r = \cdot5$	$r_p = \cdot5$ $r = \cdot6$
1	1·000	1·000	1·000
2	·735	·698	·666
3	·581	·536	·499
4	·481	·435	·399
5	·410	·366	·332
6	·357	·316	·285
7	·316	·278	·249
8	·284	·248	·221
9	·258	·224	·199
10	·236	·204	·181

It appears that entering into the same parental correlation table families with numbers of offspring varying from, for example, 1 to 5 the same weight is given to pairs of observations which according to Table IV ought to vary in weight from 1 to $\frac{1}{3}$.

It is therefore a more rational proceeding to sort the families according to the number of offspring and deal with each group separately. The work may then be shortened by calculating the correlation coefficient between the parental value and the mean for the offspring from which the parental correlation for individuals is obtained by multiplying with $\frac{\sigma_q}{\sigma}$, σ_q being as above (see I(*g*)) the S.D. for means of fraternities of q individuals. It is then possible to calculate the correlation coefficient with S.D. for each group of families and finally calculate a mean value for the correlation coefficient.

In investigations of inheritance with animals with numerous offspring it is as a rule easier to provide information of a given number of individuals among a small number of families than to examine the same number of individuals if they belong to a larger number of families. The labour required is therefore not proportional to the number of individuals and it must be estimated for the individual materials whether the encumbrance of dealing with a relatively large number of families is duly compensated for by the reduction of the number of individuals hereby permissible.

It does not seem at the outset probable, but it may be possible, that, even in cases in which parent and offspring are equally easily available for investigation, a shortening of labour, that is, a diminution of the total number of observed individuals, may be obtainable by examining several offspring individuals of each family. We will therefore examine for which value of q, $\sigma^2_{\rho_p}$ is a minimum when $n(q+1)$ is put equal to a constant k. We find the condition

$$(1-r_p^2)^2 - \left(1+\frac{1}{q^2}\right)(1-r)\left\{1-r_p^2\frac{3-r}{2}\right\} = 0,$$

from which follows

$$q^2 = \frac{(1-r)\left\{1-r_p^2\frac{3-r}{2}\right\}}{(1-r_p^2)^2 - (1-r)\left\{1-r_p^2\frac{3-r}{2}\right\}}.$$

To obtain a survey we introduce a few sets of values for r_p and r for which we give the result in Table V.

TABLE V.

r_p	r	q
0·20	0·25	1·8
0·30	0·40	1·3
0·50	0·60	1·0

It will be seen, that for sufficiently small values of r and r_p it is profitable to examine several siblings of each family in those cases where the examination of an offspring individual requires the same labour as that of a parent.

As a guide for the choice of the number of offspring in the more frequently occurring case when it is easier to provide data of offspring than of parent, we give in Table VI for some values of r_p and r the number of observations which, for varying values of q, yield the same accuracy in the parental correlation coefficient as 1000 parents with 1000 offspring.

It appears from the table that while the number of offspring increases evenly with increasing q the number of parents decreases more and more slowly, so that the compensation obtained in this way for the increased total number of offspring

tends to be very small for increasing q. Already by increasing q from 5 to 6 we find, for $r_p = \cdot 3$ and $r = \cdot 4$, that to outweigh the augmentation of 360 in the number of offspring, we only get a diminution of 21 in the number of parents.

TABLE VI.

Number of Parental and Offspring Individuals which for varying q yield the same Accuracy to ρ_p.

	$r_p = \cdot 3$ $r = \cdot 4$ $\sigma_{\rho_p} = \cdot 0288$		$r_p = \cdot 4$ $r = \cdot 5$ $\sigma_{\rho_p} = \cdot 0266$		$r_p = \cdot 5$ $r = \cdot 6$ $\sigma_{\rho_p} = \cdot 0237$	
q	Number of Parents	Number of Offspring	Number of Parents	Number of Offspring	Number of Parents	Number of Offspring
1	1000	1000	1000	1000	1000	1000
2	680	1360	717	1433	751	1502
3	573	1720	622	1866	668	2004
4	520	2081	575	2299	627	2507
5	488	2441	546	2732	602	3009
6	467	2801	528	3166	585	3511
7	452	3161	514	3599	573	4013
8	440	3522	504	4032	565	4516
9	431	3882	496	4465	558	5018
10	424	4242	490	4898	552	5520

For fraternal correlation we have found (see Table II) that the most profitable number of offspring was 3—4 for the values of r now considered, and that a somewhat greater number was not substantially opposed to economy of work. Whether the number ought to be increased beyond 3—4 or confined to even fewer offspring individuals from each family depends in each investigation upon the relative difficulty of observing parents and offspring.

(*h*) *Application of the Formula to previous Calculations of Correlation.*

For the investigation of *Zoarces viviparus* mentioned in the previous section, in which 10 offspring individuals were examined for each mother, we have according to (36) the following formula for the probable error of the maternal correlation coefficient:

$$\text{P.E.}\,(r_p) = \frac{0{\cdot}67449}{\sqrt{n}} \left\{ (1 - r_p{}^2)^2 - \frac{9}{10}(1 - r) \left[1 - r_p{}^2 \frac{3 - r}{2} \right] \right\}^{\frac{1}{2}}.$$

In Table VII are found the values of r_p for the three characters examined as well as their probable errors calculated from this formula. Giving each of these values of r_p its due weight we have calculated a mean value and its probable error.

TABLE VII.

Maternal Correlation.

Year when sample taken	Vert.	Pd.	Pigm.
	$r_p \pm$ P.E.	$r_p \pm$ P.E.	$r_p \pm$ P.E.
1914	0·3513 ± ·0343	0·2409 ± ·0332	—
1915	0·4375 ± ·0281	0·3215 ± ·0303	0·3762 ± ·0381
1916	0·4139 ± ·0355	0·2116 ± ·0387	0·3622 ± ·0373
1917	0·3775 ± ·0298	0·2824 ± ·0293	0·3722 ± ·0332
1918	0·4382 ± ·0298	0·2928 ± ·0298	0·3710 ± ·0308
1919	0·3674 ± ·0378	0·1851 ± ·0387	0·3380 ± ·0398
From total samples	0·4021 ± ·0131	0·2654 ± ·0133	0·3654 ± ·0158

It appears that these probable errors agree extremely well with those originally calculated* on the basis of the 5 or 6 values of the correlation coefficient obtained from 5 or 6 samples.

Summary.

In the first section we dealt with fraternal correlation and a formula was deduced for the standard deviation of the fraternal correlation coefficient for the case when the material of observation consists of equal numbers of offspring from each family and when each available pair of siblings is introduced into the calculation. The formula is calculated on the supposition of normal distribution and normal fraternal correlation.

It is shewn by means of the formula that forming fraternal correlation tables for fraternities of different numbers and giving each pair of observations the same weight we disturb very highly the distribution of weight which the observations must claim according to their nature. We find further from the formula that when the number of observed offspring from each family may be freely chosen the best determination of fraternal correlation from a given number of observations is obtained by taking $\left(1 + \frac{1}{r}\right)$ offspring individuals from each family (r = frater. corr. coeff.).

In the second section we deduce, also supposing normal distribution and normal correlation, the S.D. of the parental correlation coefficient calculated from a material comprising equal numbers of offspring from each family. The formula shews that forming parental correlation tables of a material consisting of families of different sizes we also in an unfortunate manner disturb the due distribution of weight among the pairs of observation. It is shewn that if observations of

* *Vide l.c.*, p. 24, Table 6.

parents are as easily produced as those of offspring it is, for determination of parental correlation, only for small values of the corr. coeffs., for instance $r_p < \frac{1}{4}$ and $r < \frac{1}{3}$, profitable to include more than one offspring individual from each family in the calculation. For the case more frequently occurring, when the observation of parents represents more labour or greater cost than that of offspring, we have for certain values of r_p and r and varying sizes of fraternities calculated such numbers of parents and of offspring which yield the same accuracy to the parental correlation as 1000 parents with corresponding 1000 offspring. Table VI shews that when the number of siblings exceeds 4—5, there is not much gained by increasing it.

Considering both fraternal and parental correlation we may therefore generally conclude that an essential increase in the number of offspring beyond $1 + \frac{1}{r}$, i.e. in practice 3—4, is only then to be recommended, when it causes a relatively insignificant increase in labour.

This research has been occasioned by the investigations of inheritance carried out by the Carlsberg Laboratorium Kobenhavn and I am much indebted to Dr. Johs. Schmidt for the interest he has taken in my work.

ON THE VARIATIONS IN PERSONAL EQUATION AND THE CORRELATION OF SUCCESSIVE JUDGMENTS.

By EGON S. PEARSON, Trinity College, Cambridge.

CONTENTS.

I. Introduction.

Starting from Bessel's discovery, in the early part of the last century, of the existence of a definite relative personal equation for two observers recording transits by the eye and ear method, there has been a continuous discussion among astronomers on the errors which such personal equations may introduce, and on the methods of eliminating them or correcting for them*. In such discussions it has been the usual practice to take the yearly mean personal equation, whether relative or absolute, of different observers and to use this mean personal equation as the basis of any correction to be applied to observations made in that year. From a comparison of the yearly means it is admitted that there may be gradual secular changes in personal equation, but it is found that for experienced observers there is usually very little variation. In text-books on Practical Astronomy brief mention of the subject is usually made, and the conclusion drawn is that for an observer in normal health, the personal equation in any one type of observation will remain sensibly constant for "short periods" of time; an exact definition of the words "short period" is not and clearly cannot be attempted†. It is further assumed that variations from the personal equation are due to accidental errors and may be taken as randomly distributed in accordance with the Gaussian Law. With the recent introduction of photography and mechanical methods of record, the interest of the astronomer in the subject has to some extent diminished, but there are many fields of scientific observation where the human element cannot be eliminated, and in the modern researches of the psychologist we find a study is made of problems of this type for their own interest and for the light which they may throw on the working of the human machine.

One very important aspect of the problem of personal equation, and of particular import to the astronomer, was discussed in detail in a paper entitled "On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation," published in the *Phil. Trans.* (Vol. 198 A, p. 235). In this case various series of experiments were carried out simultaneously by three observers under identical conditions and it was found that there was a marked correlation between the variations in absolute personal equation of the different observers. This in itself was sufficient to show that the judgments of any one observer were not randomly distributed about his mean personal equation. The purpose of the present paper is to discuss the variations in judgment of *one* observer, and to inquire how far the evidence of four or five experiments suggests that the theory of personal equation and of errors of judgment, as usually accepted, requires modification.

The subject is a large one, and much beyond the scope of a single paper; but by making careful inquiries of this type with the help of statistical methods, it

* For example, *Monthly Notices*, Vol. XL. 1880, pp. 75, 165, 302 (Discussion of Greenwich Observations of the Moon); *Monthly Notices*, Vol. XLIV. 1884, pp. 1 and 39 (Greenwich Observations of the Sun); *Monthly Notices*, Vol. LVII. 1897, p. 504 (General Discussion of relative personal Equations).

† For example, in Campbell's *Elements of Practical Astronomy*, 1899, p. 157; Young's *General Astronomy*, Revised Edn. § 114, and Chauvenet's *Spherical and Practical Astronomy*, 4th Edn. II. p. 189.

may be possible to construct a more generalised theory of errors of judgment than that which has hitherto been adopted, and although the practical corrections which such a theory will impose may not be large, yet a more detailed knowledge of the nature of the variations and perhaps some insight into the psychological and physiological factors which underlie them, will give the observer a clearer idea of the precautions to be taken to avoid error and a greater justification for confidence in his results.

II. GENERALISED THEORY OF PERSONAL EQUATION.

Before proceeding to the reduction of the Experiments which have been carried out, I will consider whether it is not possible to make a very general, and yet simple, analysis of personal equation. Let us suppose that we have a large number, N, of observations, which have been made in separate groups, or at what may be termed separate *sessions*. For the astronomer, a session will be a night's work; for the physicist or psychologist, one continuous set of readings or observations. Any particular observation y may be designated (1) by τ, a function of the time when it was recorded, measured from some fixed epoch, or (2) by the number of the session in which it was made, and t, the time of record measured from the commencement of that session. E.g. an observation made in the pth session may be written either as y_τ or ${}_p y_t$. We will suppose that the secular change can be represented by the function $\phi(\tau)$, but in addition to this change there may be another of a different type which may be termed the *sessional change*, and will be represented by the function $f_p(t)$. The fundamental difference between a secular and sessional change is this: if there is a break of some hours or perhaps days between two series of observations, the sessional change of the first series will have no influence on the judgments of the second series, while the secular change will continue from series to series. The sessional change is thus peculiar to its own session or series of observations, although it is very possible that the same type of change may be repeated in session after session; it may be a change resulting simply from fatigue or perhaps from more complex causes. Figure 3 (p. 46) provides a good illustration of secular and sessional changes; the centres of the small circles represent the mean values of twenty different series of observations, and it will be seen that the general tendency is for a drop in mean judgment from left to right of the diagram; this is the secular change. The sessional changes are represented by the continuous lines drawn through the centres of the circles, and the slope of these lines is on the whole seen to be very constant throughout the twenty series. In this case the secular and sessional changes are acting in the same direction, but they may well act in opposite directions.

We have thus seen that an observation y may be expressed in the form

$$y = \phi(\tau) + f_p(t) + Y_t \quad \text{.....(i),}$$

where Y_t is the residual after the removal of secular and sessional changes. The duration of the session is likely to be so short compared with the period over which the secular change is measured, that τ may be taken as practically constant

for any one session, and $\phi(\tau_p)$ may be described as the secular term in the observations of the pth session. It remains therefore to consider the function $f_p(t)$. Supposing that there were n observations made in a session, it would of course be possible to fit an $(n-1)$th order parabola on which all the observations would lie, so that the values of Y_t would all be zero, but such a curve would be entirely useless. If the observations are made at finite intervals so that we can imagine that one may be interpolated between two others, owing to the mass of random errors to which each judgment is subject, we should not for a moment expect that the interpolated error would lie on, or even close to the $(n-1)$th order parabola. A curve of far lower order would probably give a much better fit. If the sessional change is a sign of some physiological change of state which is affecting the observer's judgment, it is natural to suppose that it can be represented fairly closely by some simple curve—a low order parabola if not a straight line, or perhaps, if periodic, a sine curve. Suppose that in a practical case, a first or second order parabola has been fitted to the observations of a session; then it will be easy to test whether the residuals Y_t follow a Gaussian distribution; a simple practically sufficient, if not theoretically sufficient test would be to find whether

$$\left.\begin{aligned} &\sum_{t=1}^{n}(Y_t)=0, \quad \sum_{t=1}^{n}(Y_t^3)=0 \text{.....(ii)} \\ &\beta_2=\frac{\mu_4}{(\mu_2)^2}=\frac{\Sigma(Y_t^4)}{\{\Sigma(Y_t^2)\}^2}=3 \text{.....(iii)} \end{aligned}\right\} \text{approximately.}$$

But there is a further possibility; it may be found that although the relations (ii) and (iii) hold approximately, the Y_t's are not randomly distributed in time, and that there is in fact a correlation between the successive values of Y_t, so that

$$r_{Y_t;\, Y_{t+k}}=\frac{\sum_{t=1}^{n}(Y_t Y_{t+k})}{\sqrt{\sum_{t=1}^{n}(Y_t^2)\sum_{t=1}^{n}(Y_{t+k}^2)}}\neq 0$$

for perhaps several positive integral values of k from 1 upwards.

To emphasise the importance of the different terms in the relation

$${}_py_t=\phi(\tau_p)+f_p(t)+Y_t \text{.....(i) } bis,$$

let us take the case of an astronomer who makes a number of observations, often at many days' interval. He will take a mean

$$\bar{y}=\text{mean}\,\phi(\tau_p)+\text{mean}\,f_p(t),$$

but he must not suppose that the quantities

$${}_py_t-\bar{y}=\phi(\tau_p)-\text{mean}\,\phi(\tau_p)+f_p(t)-\text{mean}\,f_p(t)+Y_t$$

follow a Gaussian distribution. It will be only a part of the expression that does so, the Y_t's, and it is possible that even these may not be true.

Further it is clear that successive values of ${}_py_t-\bar{y}$ will not be independent; correlation will arise from the inclusion of both the secular and sessional terms,

and perhaps too from a relationship between the successive Y_t's. There may be no large scale sessional change, and it may be possible to correct for a secular change in personal equation, but even then the mean of a small number "m" of successive observations, subject to its probable error $\cdot 6745\sqrt{\frac{1}{m}}\sigma_m$, will not be a satisfactory approximation to the true value of the quantity observed, if these "m" observations are correlated. Suppose for example that the points in Figure 14 (p. 76) represent a series of successive observations which have been corrected for any secular change in personal equation; the linear sessional change is small and has been represented by the continuous straight line, while the dotted straight line represents the mean value of the 63 observations. Yet many sets of 10 consecutive observations could be taken, the difference between the mean of which and that of the whole 63 would be far greater than would be anticipated from the value of the probable error calculated from the expression above. This is because the observations are not randomly distributed in time.

In addition to secular and sessional changes in the value of an estimation, there may be similar changes in the standard deviation; the judgments may become more erratic or less so. A sessional change giving an increase in standard deviation would suggest the effect of fatigue; and secular change decreasing the standard deviation might be the indication of increased accuracy with experience. An example of secular change in personal equation and standard deviation is illustrated in the diagram on p. 84; the details of this will be discussed more fully in the reduction of Experiment *D*, but it is here sufficient to say that the central curve represents the smoothed personal equation, while the distance between any point on this curve and either of the outer curves gives the smoothed standard deviation at that point or period in the series of observations. It will be seen that the standard deviation increases in the later observations.

It would be out of place at this point to enter further into the details of variation in personal equation and correlation of judgments, but I think that enough has been said to indicate the general lines of enquiry. In choosing the experiments which will be described in the following sections, the aim has been to select those in which there was likely to be considerable variation in judgment, and where consequently the secular and sessional changes, if present, would be clearly recognizable and the correlation of successive judgments easy to measure. It was also important that the errors in measurement should be small compared with the variations in judgment.

It may of course be urged that the experiments should have been carried out by an observer who was unaware of the lines of enquiry and therefore not liable to bias of any form, but this was not practicable, and in fact none of the reductions had been completed nor the general theory developed before all the experiments had been carried out, and I do not think that the observations could have been affected by any conscious or unconscious prejudice.

III. The Experiments.

The present paper is based on the reduction of the following Experiments:

A. Estimation of the value of a Third, or Trisection Experiment.

B. Estimation of the value of a Half, or Bisection Experiment.

C. Estimation of Time, by counting of Ten Seconds.

D. Estimation of Ten Seconds without intermediate counting.

E. Some repeated measurements of fine structure in a Stellar Spectrum, with a Zeiss Comparator.

The first four Experiments were carried out by the writer in accordance with a uniform scheme; each Experiment was divided into 20 series of 63 observations, making 1260 observations in all. Only one series (or 63 observations) was done at a sitting to avoid as far as possible the effect of fatigue; in the case of Experiments *A* and *B* the sequence of the series was much broken, spreading over some weeks, but *C* and *D* were carried out within four consecutive days. The dates of the series are given with the detailed discussion of the observations below.

(*a*) *Experiments A and B.*

Figure 1 is a copy of one of the printed forms used for these experiments; the longer line was used for *A*; distance between inner edges of bounding marks 7·53 inches; the shorter line was used for *B*; distance between inner edges of bounding marks 5·94 inches.

The lines were on the same form simply for convenience in printing, etc. and that not used was concealed while the observation on the other was being made; a fresh line was used for each of the 1260 observations. In carrying out a series a pile of 63 forms was placed on a table slightly tilted up towards the observer, and straight in front of him, with a good light coming from the left-hand side, the pencil being in his right hand. He then made a short pencil stroke across the line at the point which he estimated was one-third way along the line from the left-hand end (Experiment *A*), or at the point which he considered to bisect the line (Experiment *B*). He then turned the form over, face downwards at his side, and proceeded to deal with the next form in the same manner, continuing until the 63 were finished*. The pencil stroke was made after a rapid eye estimate, the aim being to record the first impression of third or half formed upon seeing the fresh line, and to avoid hesitation; the average time taken in going through a series of 63 observations was 5 minutes 40 seconds for Trisection, 5 minutes 22 seconds for Bisection, or 5·4 seconds and 5·1 seconds respectively between judgments.

To avoid bias, it would have been desirable to complete all the observations of an experiment before commencing the measurement of any of the series, but

* Actually in Experiments *A* and *B* 70 forms were marked in each series; the first 7 were to enable the observer to "get his eye in," and the measures of them were not used at all in the reduction.

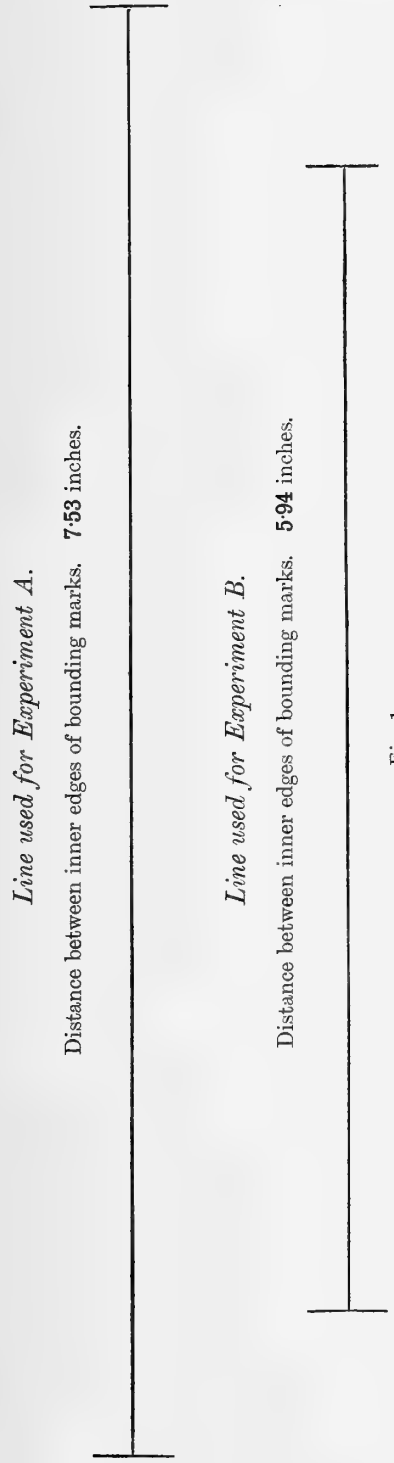


Fig. 1.

from considerations of time and as all the forms were not printed at the commencement this was not done. In some cases therefore a series was measured directly after it had been marked, and if the observer happened to remember that its estimates were considerably too large or too small, his judgment would almost certainly be influenced when marking the next succeeding series; the correlation of judgments within this second series would hardly be altered, but any natural secular change which had been occurring from series to series might be broken*.

The measures of the observations were made with a ruler divided to fiftieth's of an inch, so that readings could be taken to one hundredth of an inch with fair accuracy.

(*b*) *Experiments C and D.*

These two experiments were carried out with the help of a chronograph. The instrument was run by clockwork, and had a paper tape on which records could be made independently by two pens worked by small electromagnets. One pen was put in circuit with a second's pendulum, a platinum pointer at the end of which made contact at each swing through the vertical position by cutting through a bead of mercury, the other pen was connected with a tapping key. The rate of the driving clock was not quite uniform, and the pendulum second-marks on the tape were therefore necessary in reckoning the intervals between the marks made by the other pen, corresponding to taps of the key. As the estimate in both experiments was one of 10 seconds, it was found that except for a few cases in Experiment *D*†, the true value of the time interval between the taps could be represented with sufficient accuracy by the factor e/p, where,

* See p. 49, remark in Table I, regarding Series IX and X.

† In Experiment *D*, some of the estimates had values nearer 20 seconds than 10 seconds, and here half the distance on the tape between the nearest corresponding 20 seconds was taken for p.

e was the distance measured on the tape between consecutive marks of the key.

p the length on the tape of the nearest corresponding 10 seconds recorded by the pendulum pen.

Had the pendulum been beating exactly one second, $10 \times \frac{e}{p}$ seconds would have been the true length of the estimate; actually the period as found by comparison for a long run with a watch was,

before Experiments *C* and *D* (6th December) 1·020 seconds
after „ „ (16th „) 1·019 „ ,

so that the length of estimate with sufficient accuracy is $10{\cdot}2 \times \frac{e}{p}$ seconds. It is the factor $\frac{e}{p}$ that will be used throughout the reductions.

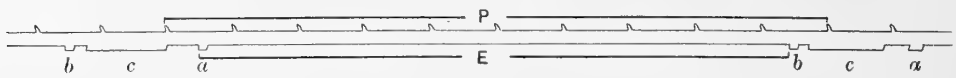


Fig. 2. Shows a small piece of tape, and the points from which the measurements were made.

If the amplitude of the pendulum was rather small, it was sometimes noticeable that the intervals between the second marks were alternately longer and shorter; this was due either to slight deformation in the shape of the mercury bead or (what is really the same thing) from the centre of the bead not having been placed exactly under the equilibrium position of the platinum pointer. But in taking for measurement the even number of 10 seconds, such errors would be inappreciable.

In both experiments the beginning and end of the estimate were recorded by sharp taps on the key (at *a* and *b* respectively in Figure 2); a long drawn tap (*c* in figure) then followed to make a break before the next estimate was recorded. The interval between the *b* tap of one observation and the *a* tap of the following varied from $1\frac{1}{2}$ to $2\frac{1}{2}$ seconds. This method of record soon became quite automatic, and very few mistaps occurred.

The measurements on the tape were made from the sharp beginnings of the marks, which correspond to the making of the electric contact at the beginning of the tap on the key.

In Experiment *C* the counting was "sotto voce," the first tap being made on the count "nought," the last on "ten"; in order that the counts might be quite uniform the word "sen" was used instead of the two-syllabled "seven." The counting was usually done in step to a slight beat of the thumb on the key (not hard enough, of course, to make contact), and it was fairly easy to keep the attention concentrated during the counts. In Experiment *D* there was no counting and it was far harder to keep one's mind fixed; in fact the mental effort required was quite noticeable, and I found that a greater interval of rest was required

between each series than for *C*. It is mainly by reference to the passing of external events, to changes the duration of which we can infer from previous experience, that we estimate any but the shortest intervals of time. In the counting experiment, the second-intervals between each of the 10 counts which made up the observation were comparatively short, and the beating of the thumb or fingers became almost mechanical; the interval of course varied but was not subject to violent fluctuations. But while most people are able to estimate a second interval with fair accuracy, it would need very much practice to estimate a 10 second interval, and in my case I found it quite impossible to concentrate attention for 10 seconds, solely on the passing of time. I soon found myself imagining that I saw the seconds' hand of a watch, passing usually from the position where 60 is marked on the dial to the 10; but it was not another case of counting, for I did not note the passing of each individual second mark, only having a vague idea of the position of the 5 second division line. If I tried to think of nothing, my thoughts probably wandered on to other subjects, until I came up with a start, and realising that I had very little idea of how long before I had pressed the key to start the observation, pressed it to finish, with the greatest uncertainty. To keep attention fixed, it appeared that I must try to record the stages of the passage of 10 seconds, and this I was doing vaguely on the imaginary clock face, but I must say that the seconds' hand was very refractory, at times appearing to stop or even move backwards, and was often so slow that I had to close the observation before it reached the 10 second mark.

I have given the above description at some length in order to shew that there was an essential difference between Experiments *C* and *D*, which is borne out by the figures of the reduction given later in this paper. The observer with the key sat in a separate room where the beats of the chronograph could not be heard. Experiment *D* was actually carried out in the week previous to *C*; before starting, a few trials at estimating 10 seconds had been made with a watch, but these were not repeated after the commencement. Again, some 10 second counts were made with a watch before starting on *C*, but no comparison with a watch or clock was made during the course of the experiment. The measuring up of *C* and *D* was left until both experiments were completed, so that the chance of some bias to the judgment, which occurred in the case of *A* and *B* was avoided.

(*c*) *Experiment E.*

This consists of nine series of readings made with a Zeiss Comparator at the Solar Physics Observatory, Cambridge, on photographic plates of the spectrum of Nova Aquilae III. The readings were taken in the first place in order to calculate the Probable Errors of the measurements of certain types of structure featuring in the broad emission bands, and each series consists of readings taken from 51 consecutive settings on a particular marking, either a maximum or the edge of a maximum. Although the number of readings is not sufficient for any great weight to be attached to the results, they are, I think, of sufficient interest to be included. In the instrument used, the plate to be measured is fixed to a slide,

which is moved horizontally in a greased slot by pressure with the hand; the measurer looks through one eyepiece and pushes the slide until the feature on the plate of which he is wishing to measure the position, comes under a cross wire in the focus of the eyepiece; then looking through a second eyepiece at the scale attached to the slide, he takes the reading, the last two figures of which are read from a graduated wheel attached to a micrometer screw-head. In making a measurement there are therefore two adjustments:

(1) The setting of the marking in the plate under the cross wire in the first eyepiece.

(2) The shifting of two very close parallel wires by a micrometer screw in the second eyepiece, until a line of division on the scale appears to lie exactly in the centre between them.

Far the greater source of error arises from the first setting, particularly if the marking on the plate is not clear cut. In taking a series of measurements, the observer should always move the slide from the same direction—that is he should always push it or always pull it, until he thinks that the marking is bisected or "edged" by the cross wire, and then he should stop; if he obviously overshoots the mark he should start again, and not hesitatingly move the slide backwards and forwards in search of what he thinks may be the best setting. By shifting the slide into position from the same direction, the measures may be all subject to a fairly constant personal equation due to "over push" or "under push," "over pull" or "under pull" of the slide, but this effect may be eliminated by reversing the plate in the instrument, making a fresh series of measures, and taking the mean of the two. In this particular set of readings the slide was always "pulled" into its final position.

(*d*) It is hoped that the results of some further experiments of a different type in estimating length which were kindly undertaken for me by Mr E. A. Milne of Trinity College, and Mr L. J. Comrie of St John's College, Cambridge, will be included in a future paper.

IV. Terminology.

Experiments *A*, *B*, *C* and *D* were arranged in accordance with a uniform scheme, each Experiment being divided into 20 "series" consisting of 63 observations. The series will be designated by the Roman numerals I, II...XX in the order in which they were carried out, and the 63 observations* in a series by the letters

$$y_1, y_2, \dots y_t \dots y_{63}.$$

In dealing with each Experiment one of the first objects will be to ascertain whether there is any correlation between successive judgments, and the manner in which this correlation, if existent, falls off as the interval between the judgments correlated is increased. To obtain these coefficients of correlation it is necessary

* The first 7 observations, see footnote, p. 28, being always disregarded.

to divide the observations of each series into "groups," and thus we have the 50 observations

$y_1, y_2, \ldots y_{50}$ form Group 1 with mean d_1 and standard deviation σ_1,

$y_2, y_3, \ldots y_{51}$ „ „ 2 „ „ d_2 „ „ „ σ_2,

.....

$y_k, y_{k+1}, \ldots y_{50+k-1}$ „ „ k „ „ d_k „ „ „ σ_k,

.....

$y_{14}, y_{15}, \ldots y_{63}$ „ „ 14 „ „ d_{14} „ „ „ σ_{14}.

By "the correlation of successive judgments at intervals of one," I shall understand the correlation of the 50 observations of Group 1 of a series with the 50 corresponding observations of Group 2 of that series; this will be expressed as ρ_1. Similarly "the correlation of successive judgments at intervals of k," or ρ_k, is the correlation of the corresponding observations in Groups 1 and $k+1$. In fact ρ_k is given by

$$\rho_k = \frac{\frac{1}{50}\sum_{t=1}^{50} y_t y_{t+k} - d_1 d_{k+1}}{\sigma_1 \cdot \sigma_{k+1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(iv)}.$$

When these constants are to be referred to some particular series, say the pth, the prefix p will be placed before them, e.g. ${}_p\sigma_1$, ${}_p\sigma_k$, ${}_p\rho_k$, etc.

A comparison of the d's, σ's and ρ's of the different series will be instructive, but as each of these constants has been calculated from 50 observations only, to obtain quantities with smaller probable errors we must combine the observations of the 20 series. Thus we shall obtain

$$D_k = \frac{1}{mn}\sum_m \sum_{t=1}^{n} y_{t+k-1} = \frac{1}{m}\sum_m (d_k) \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(v)},$$

where $n = 50$, the number in a group,

$m = 20$, the number of series,

and $\sum_m$ indicates summation for all 20 series.

$$P_k = \frac{1}{mn}\sum_m \sum_{t=1}^{n} y_t y_{t+k} - D_1 D_{k+1}$$

$$= \frac{1}{m}\sum_m (\rho_k \sigma_1 \sigma_{k+1} + d_1 d_{k+1}) - D_1 D_{k+1}.$$

$$P_k = \frac{1}{m}\sum_m (\rho_k \sigma_1 \sigma_{k+1}) + \frac{1}{m}\sum_m (D_1 - d_1)(D_{k+1} - d_{k+1}) \text{ in view of (v)} \quad \ldots\text{(vi)}.$$

Putting $k = 0$, in (vi) we have as the square of the standard deviation

$$\left.\begin{aligned} S_1^2 &= \frac{1}{m}\sum_m (\sigma_1^2) + \frac{1}{m}\sum_m (D_1 - d_1)^2 \\ \text{and similarly} \quad S^2_{k+1} &= \frac{1}{m}\sum_m (\sigma^2_{k+1}) + \frac{1}{m}\sum_m (D_{k+1} - d_{k+1})^2 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\text{(vii)},$$

and finally the coefficient of correlation $\mathbf{R}_k$ is given by

$$\mathbf{R}_k = \frac{P_k}{S_1 S_{k+1}} \quad \text{.....(viii).}$$

D_k and S_k are the mean and standard deviation of the combined observations—1000 in all—of the 20 Groups k, while $\mathbf{R}_k$ is the correlation between the 1000 observations in the 20 Groups 1 and the corresponding 1000 observations in the 20 Groups $k+1$, where it must be remembered that owing to the break between each series the 50th observation in Series I is correlated with the $(50+k)$th observation in that series, and not with the kth in Series II, etc.

It will be seen from the equations (vi) and (viii) that it is possible for $\mathbf{R}_k$ to have a large value even though the coefficients of correlation of successive judgments for the separate series are negligible. For though $\sum_m (\rho_k \sigma_1 \sigma_{k+1})$ may be zero for $k \geqslant p$, let us say, where p may perhaps be 3 or 4, it is clear that the coefficients for the combined series, $\mathbf{R}_k$, will not vanish as k increases unless

$$L_k = \frac{\sum_m (D_1 - d_1)(D_{k+1} - d_{k+1})}{S_1 S_{k+1}} \to 0.$$

In fact if L_k (and therefore $\mathbf{R}_k$) does not vanish for values of k for which the ρ_k's of the individual series vanish, this is a sign of the existence of a secular change running through the series; the means of the separate series differ significantly from the mean of the combined 1000 observations, that is to say they differ significantly from each other. Now it is important to obtain a measure of the correlation of successive judgments, when freed from this secular term. First I define S_k' by the relation

$$S_k' = \sqrt{\frac{1}{m} \sum_m (\sigma_k^2)} \quad \text{.....(ix),}$$

($m = 20$, $\sum_m$ indicating summation for the 20 series); it is the standard deviation of the 1000 observations in the combined Groups k after the secular change has been removed. Then $\mathbf{R}_k'$ is given by

$$\mathbf{R}_k' = \frac{\frac{1}{m} \sum_m (\rho_k \sigma_1 \sigma_{k+1})}{S_1' S_{k+1}'} \quad \text{.....(x),}$$

this is the correlation of successive judgments freed from secular change; before correlating the observations we are in fact fitting the series means together, by subtracting ${}_1d_1 - D_1$ from the observations of the 1st Group of Series I, ${}_1d_2 - D_2$ from the 2nd Group and so on, and again subtracting ${}_1d_{k+1} - D_{k+1}$ from the observations of the $(k+1)$th Group of Series I, etc.

Again it may be desirable to examine the residuals after a sessional change has been removed from the observations of each series, in addition to the general secular term. Suppose that an observation in the pth Series can be expressed in the form introduced on page 25

$${}_p y_t = \phi(\tau_p) + f_p(t) + {}_p Y_t \quad \text{.....(i) bis,}$$

where $\phi(\tau_p)$ represents the secular term which we take as constant for all the observations of the pth Series, and $f_p(t)$ gives the sessional change, then S_1'' will be the standard deviation of the 1000 residuals in the twenty 1st Groups, S_k'' of the 1000 residuals in the twenty kth Groups, etc., so that

$$S_k'' = \sqrt{\frac{1}{mn} \sum_{m} \sum_{t=1}^{n} (Y^2_{t+k-1})} \quad \text{.....(xi),}$$

the mean of the residuals being zero, and $m = 20$, $n = 50$ again; while the correlation of the successive residuals at intervals of k, after the removal of secular and sessional terms, or $\mathbf{R}_k''$ will be given by

$$\mathbf{R}_k'' = \frac{\frac{1}{mn} \sum_{m} \sum_{t=1}^{n} (Y_t Y_{t+k})}{S_1'' . S''_{k+1}} \quad \text{.....(xii).}$$

TABLE OF CONSTANTS.

In the following table definitions are given of the most important of the constants referred to in the preceding section and of others to be introduced in the sequel.

1. The kth Group of the pth Series consists of the 50 observations

$${}_pY_k,\ {}_pY_{k+1},\ \ldots\ldots\ {}_pY_{k+50-1}.$$

As each Series consists of 63 observations, there are 14 Groups in each of the 20 Series,

n will often be used for 50, the number of observations in a Group,

m „ „ 20 „ „ Series.

2. *The crude Observations.*

(a) For the pth Series.

$\bar{d}$ = mean of the whole 63 observations.

${}_pd_k$ = mean of observations in kth Group.

${}_p\sigma_k$ = standard deviation of observations in kth Group.

${}_p\rho_k$ = coefficient of correlation between corresponding observations of Groups 1 and $k+1$, i.e. between ${}_pY_1$ and ${}_pY_{k+1}$, ${}_pY_2$ and ${}_pY_{k+2}$, etc.

${}_p\sigma_\delta$ = standard deviation of the first forward differences of the observations in Group 1, i.e. of ${}_pY_2 - {}_pY_1$, ${}_pY_3 - {}_pY_2 \ldots {}_pY_{51} - {}_pY_{50}$.

${}_pb$ = slope of the straight line $y - {}_pd_1 = {}_pb\left(t - \frac{n+1}{2}\right)$ which fits "best" the 50 observations ${}_pY_1$, ${}_pY_2$, ... ${}_pY_t$, ... ${}_pY_n$ of Group 1.

${}_p\sigma_k'$ = standard deviation of residuals left after the ordinates of this "best" fitting straight line have been subtracted from the observations of Group k.

${}_p\rho_k'$ = coefficient of correlation between these residuals of Group 1 and Group $k+1$.

In the reduction of the results of the experiments, unless it is necessary to specify a particular series, the prefix p before these constants will usually be omitted for brevity.

(*b*) For the combined 20 series.

$\bar{D}$ = mean of the whole 1260 (= 20 × 63) observations of an experiment.

D_k = mean of the 1000 observations in the combined kth Groups of the 20 series.

S_k = standard deviation of the 1000 observations in the combined kth Group of the 20 series.

$\mathbf{R}_k$ = coefficient of correlation between the 1000 observations in the 1st Groups and the 1000 corresponding observations in the $k+1$th Groups.

${}_sR_k$ = coefficient of correlation between the 1000 sth forward differences of the observations in the 1st Groups and the corresponding differences of the observations in the $k+1$th Groups.

S_δ = standard deviation of the 1000 first forward differences of the observations in the 1st Groups.

3. *The Observations freed from the Secular Change.*

The "secular term" in the observation ${}_py_t$ considered as a member of the kth Group is ${}_pd_k$. Thus the mean of the 1000 observations in the kth Groups each freed from its secular term will be zero.

S_k' = standard deviation of the 1000 observations (freed from secular term) in the kth Groups.

$\mathbf{R}_k'$ = coefficient of correlation between the 1000 observations in the 1st Groups and the 1000 corresponding observations in the $k+1$th Groups (all freed from secular term).

4. *The Observations freed from both Secular and Sessional Change.*

$y = f_p(t)$ is the curve representing the sessional change in the pth Series, so that $f_p(t)$ is the "sessional term" in ${}_py_t$, the tth observation in the pth Series.

${}_pY_t$ = the residual left after removing the secular and sessional terms from ${}_py_t$.

S_k'' = standard deviation of the 1000 Y's in the kth Groups.

$\mathbf{R}_k''$ = coefficient of correlation between the 1000 Y's in the 1st Groups and the corresponding 1000 Y's in the $k+1$th Groups.

${}_p\alpha_t$ = the part of ${}_pY_t$ representing the actual estimate which the observer wishes to record.

${}_p\beta_t$ = the part of ${}_pY_t$ representing a complex of accidental errors superimposed on ${}_p\alpha_t$ in the process of record.

G_k = standard deviation of the sessional terms in the 1000 observations of the kth Groups.

F_k = 1st order product moment coefficient about the mean of these sessional terms in the 1st Groups and the corresponding terms in the $k+1$th Groups.

V. On Methods of Reduction.

(*a*) *Variate Difference Correlation.*

It will become evident in the detailed discussion of the results of the experiments, that a considerable part of the correlation of the successive judgments is due to a secular change with time, occurring from series to series, and in the case of the Trisections, to a sessional change as well occurring within the series; I therefore propose to consider at this point how far the Variate Difference Correlation Method is applicable in this type of problem, and to do this will approach the matter from a slightly more general point of view than that of "Student" in *Biometrika*, Vol. x. p. 179.

Suppose that x and y are the two variables to be correlated, with corresponding values

$$x_1,\ x_2,\ \dots\ x_t,\ \dots\ x_v\ \dots,$$

$$y_1,\ y_2,\ \dots\ y_t,\ \dots\ y_v\ \dots,$$

and that we may express x_t and y_t in the form

$$x_t = F_1(t) + X_t,$$

$$y_t = F_2(t) + Y_t,$$

where $F_1(t)$ and $F_2(t)$ are polynomials of degree n in t, the unit of t being the interval of time or space between the successive values of the variates, which is supposed equal and constant; X_t and Y_t are independent of the secular or sessional change represented by F_1 and F_2.

Let us now obtain a general expression for

(1) $r_{\Delta_n x_t;\ \Delta_n y_t}$ or ${}_nR$, the correlation of the nth forward differences of x_t and y_t.

(2) $r_{\Delta_n X_t;\ \Delta_n Y_t}$ or ${}_nR'$ „ „ „ X_t „ Y_t.

Now

$$\Delta_n x_t = (1-\epsilon)^n x_{n+t} = x_{n+t} - n x_{n+t-1} \dots (-1)^s \frac{n!}{s!\,(n-s)!} x_{n+t-s} \dots (-1)^n x_t \dots \text{(xiii)},$$

where the operator ϵ is defined by $\epsilon^s x_t = x_{t-s}$, etc.

Further we must assume that

(*a*) $\sum_{t=1}^{v} x_{t+h}$ = constant for all values of h small compared with v,

$= 0$, by suitable choice of origin,

$\sum_{t=1}^{v} y_{t+h} = 0$,

from which it follows that $\sum_{t=1}^{v} \Delta_n x_{t+h} = 0 = \sum_{t=1}^{v} \Delta_n y_{t+h}$,

(*b*) $\sum_{t=1}^{v} (x^2_{t+h})$ = constant $= v\sigma_x^2$ for all values of h small compared with v,

$\sum_{t=1}^{v} (y^2_{t+h}) = v\sigma_y^2$ „ „ „ „ „

(c) $\sum_{t=1}^{v}(x_{t+h}\,x_{t+h+k}) = v \times {}_x\rho_k \sigma_x^2$ for all values of h small compared with v,

$\sum_{t=1}^{v}(y_{t+h}\,y_{t+h+k}) = v \times {}_y\rho_k \sigma_y^2$ " " " "

$\sum_{t=1}^{v}(x_{t+h+k}\,y_{t+h}) = v \times {}_{xy}\rho_k \sigma_x \sigma_y$ " " " "

Similar relations will hold for the residuals X and Y.

Then a little consideration shews that the sum of the coefficients of the products of the x's and y's whose indices differ by p in the expression

$$\Delta_n x_t\,\Delta_n y_t \text{ or } (1-\epsilon)^n x_{n+t}\,(1-\epsilon)^n y_{n+t}$$

is the coefficient of $v \times {}_{xy}\rho_p \sigma_x \sigma_y$ in the product moment

$$\sum_{t=1}^{v} \Delta_n x_t \,.\, \Delta_n y_t\,;$$ call this coefficient a_p.

Now $\left.\begin{matrix}\epsilon^r \text{ operating on } x_{n+t} \text{ gives } x_{n+t-r}\\ \epsilon^{r'} \quad " \quad " \quad y_{n+t} \quad " \quad y_{n+t-r'}\end{matrix}\right\}$,

and if $(n+t-r)-(n+t-r')=p$, then $r'-r=p$; hence a_p is the sum of the coefficients of the products $\epsilon_1^{\,r}\,\epsilon_2^{\,r'}$ in the expansion of $(1-\epsilon_1)^n(1-\epsilon_2)^n$ for which $r-r'=p$, or the coefficient of ϵ^p in

$$\left(1-\frac{1}{\epsilon}\right)^n (1-\epsilon)^n,$$

or of ϵ^{n+p} in $$(-1)^n (1-\epsilon)^{2n},$$

so that $$a_p = (-1)^p \frac{2n\,!}{(n+p)\,!\,(n-p)\,!}.$$

Hence finally writing $j = n+p$ we have

$$\frac{1}{v}\sum_{t=1}^{v}\Delta_n x_t\,\Delta_n y_t = \sigma_x \sigma_y \sum_{j=0}^{2n}(-1)^{n+j}\frac{2n\,!}{(2n-j)\,!\,j\,!}\,{}_{xy}\rho_{j-n} \quad \text{.....(xiv)},$$

where negative values of the subscript of ρ imply that the subscript of x is less than that of y; e.g. ${}_{xy}\rho_{-p}$ is the correlation between x_t and y_{t+p}.

Similarly for the standard deviations of the nth differences

$$\frac{1}{v}\sum_{t=1}^{v}(\Delta_n x_t)^2 = \sigma_x^2 \sum_{j=0}^{2n}(-1)^{n+j}\frac{2n\,!}{(2n-j)\,!\,j\,!}\,{}_x\rho_{j-n} \quad \text{.....(xv)},$$

$$\frac{1}{v}\sum_{t=1}^{v}(\Delta_n y_t)^2 = \sigma_y^2 \sum_{j=0}^{2n}(-1)^{n+j}\frac{2n\,!}{(2n-j)\,!\,j\,!}\,{}_y\rho_{j-n} \quad \text{.....(xvi)},$$

and for the correlation between the differences

$${}_nR = \frac{\sum_{j=0}^{2n}(-1)^{n+j}\frac{2n\,!}{(2n-j)\,!\,j\,!}\,{}_{xy}\rho_{j-n}}{\sqrt{\left\{\sum_{j=0}^{2n}(-1)^{n+j}\frac{2n\,!}{(2n-j)\,!\,j\,!}\,{}_x\rho_{j-n}\right\}\left\{\sum_{j=0}^{2n}(-1)^{n+j}\frac{2n\,!}{(2n-j)\,!\,j\,!}\,{}_y\rho_{j-n}\right\}}} \quad \text{...(xvii)}.$$

The correlation of the nth forward differences of the residuals X_t and Y_t or ${}_nR'$ will equal an exactly similar expression to the last, in which ${}_{XY}\rho$, ${}_X\rho$ and ${}_Y\rho$ are

substituted for ${}_{xy}\rho\ {}_{x}\rho$ and ${}_{y}\rho$. But as $F_1(t)$ and $F_2(t)$ are polynomials of degree n in t, we know that

$$\left.\begin{aligned}\Delta_n x_t &= \Delta_n X_t + \text{constant}\\ \Delta_n y_t &= \Delta_n Y_t + \text{constant}\end{aligned}\right\},$$

and therefore

$${}_nR = r_{\Delta_n x_t;\ \Delta_n y_t} = r_{\Delta_n X_t;\ \Delta_n Y_t} = {}_nR',$$

that is to say we may equate ${}_nR$ to an expression similar to that on the right hand side of (xvii) above, except that the correlation coefficients of the residuals, namely: ${}_{XY}\rho$, ${}_{X}\rho$ and ${}_{Y}\rho$ are to be substituted for ${}_{xy}\rho$, ${}_{x}\rho$ and ${}_{y}\rho$.

Now in the usual problem to which the Variate Difference Method is applied it is assumed that after taking a sufficient number of differences we shall approach a state in which the corresponding values of X_t and Y_t, the residuals left after the ordinates of an nth order parabola have been subtracted from x_t and y_t, are mutually at random in time or space; or that

$${}_{XY}\rho_p = 0, \quad {}_{X}\rho_p = 0, \quad {}_{Y}\rho_p = 0,$$

for all values of p other than zero, and that

$${}_{X}\rho_0 = 1 = {}_{Y}\rho_0, \qquad {}_{XY}\rho_0 = r_{XY},$$

i.e. the correlation between X_t and Y_t. Upon this assumption it follows at once from the modified form of (xvii) that

$${}_nR = {}_{XY}\rho_0 \quad \text{or} \quad r_{\Delta_n x_t\, \Delta_n y_t} = r_{XY},$$

the fundamental relation of the original Variate Difference Correlation Method.

Let us now turn to the particular type of problem in which we wish to correlate *the successive values of the same variate.* If we are correlating the values at intervals of k, we shall have as corresponding variables, not x_t and y_t but y_t and y_{t+k} so that

$$\left.\begin{aligned}&{}_{xy}\rho_{j-n} \text{ becomes } \rho_{j+k-n} \text{ and } {}_{XY}\rho_{j-n} \text{ may be written } \rho_{j+k-n}{}^{(n)}\\ &{}_{x}\rho_{j-n} \quad \text{,,} \quad \rho_{j-n} \quad \text{,,} \quad {}_{X}\rho_{j-n} \quad \text{,,} \quad \text{,,} \quad \rho_{j-n}{}^{(n)}\\ &{}_{y}\rho_{j-n} \quad \text{,,} \quad \rho_{j-n} \quad \text{,,} \quad {}_{Y}\rho_{j-n} \quad \text{,,} \quad \text{,,} \quad \rho_{j-n}{}^{(n)}\end{aligned}\right\},$$

where as in the notation of page 35 ρ_p is the correlation of successive values of the variate at intervals of p, and $\rho_p{}^{(n)}$ the correlation of successive residuals (at intervals of p) which are left after the subtraction of the ordinates of an nth order parabola representing the secular change. Hence we have from equation (xvii) that ${}_nR_k$, or the correlation between the nth forward differences of y_t and y_{t+k} is given by

$${}_nR_k = \frac{\sum_{j=0}^{2n} (-1)^{n+j} \dfrac{2n!}{(2n-j)!\,j!}\, \rho_{k+j-n}}{\sum_{j=0}^{2n} (-1)^{n+j} \dfrac{2n!}{(2n-j)!\,j!}\, \rho_{j-n}} \qquad \text{.....(xviii),}$$

$$= \frac{\sum_{j=0}^{2n} (-1)^{n+j} \dfrac{2n!}{(2n-j)!\,j!}\, \rho_{k+j-n}{}^{(n)}}{\sum_{j=0}^{2n} (-1)^{n+j} \dfrac{2n!}{(2n-j)!\,j!}\, \rho_{j-n}{}^{(n)}} \qquad \text{.....(xix),}$$

where negative values of the subscript of ρ and $\rho^{(n)}$ are to be treated as positive: e.g. if $k=1$, $n=5$, $j=1$, then $\rho_{k+j-n}=\rho_{-3}=\rho_3$.

We are again supposing that this secular change can be represented by $y=f(t)$, a polynomial of degree n in t, but we cannot expect that after removing a parabola of even 5th or 6th order*, the residuals $Y_1, Y_2, \ldots Y_t, \ldots Y_v$ will be mutually at random in time or space; if we anticipate correlation between Y_t and Y_{t+k}, we must also be prepared for correlation between Y_t and Y_{t+k-1}, and in any case the correlation between Y_t and Y_t or ρ'_{k+j-n} where $j=n-k$, will be unity. Hence we cannot make the assumptions of the first problem (that ${}_{XY}\rho_p=0$, etc.), in fact

$$r_{\Delta_n Y_t \Delta_n Y_{t+k}} \text{ is not equal to } r_{Y_t Y_{t+k}}.$$

Now consider the use which may be made of equations (xviii) and (xix). If the values of the ρ_p's have been calculated from the crude values of the variate, the quickest method of finding the correlations of differences ${}_nR_k$ is not by direct calculation but by putting these known values of the ρ_p's into the right hand side of (xviii). Then using (xix) we have a number of equations connecting the $\rho_p{}^{(n)}$'s, and the question that at once arises is whether there are sufficient equations to determine these coefficients? It will be seen at once that there cannot be; if we are proceeding to nth differences, we can obtain q equations by putting $k=1, 2, \ldots q$, but these will contain coefficients $\rho_1{}^{(n)}$, to $\rho_{n+q}{}^{(n)}$; in fact n more equations are required. By using the appropriate equations for the Product Moments and for the Standard Deviation of nth differences corresponding to (xiv), (xv) and (xvi) we could obtain one further equation, but at the same time we introduce one further unknown, the standard deviation of the residuals.

That these equations will be indeterminate, can be seen from another standpoint; the nth difference correlation equations (xviii) and (xix) will be satisfied not only by the ρ_p's and $\rho_p{}^{(n)}$'s as defined above, but by the correlation of the residuals left after the ordinates of a parabola of any order less than n, have been subtracted from the crude observations. Nor can further equations for the $\rho_p{}^{(n)}$'s, be obtained by proceeding to $n+1$, or higher differences; the further relations obtained will not be independent, for example

$${}_{n+1}R_1=\frac{-1+2\,{}_nR_1-{}_nR_2}{2\,(1-{}_nR_1)} \text{ etc.}$$

The possible application of these difference correlation equations is considered in the next section.

(*b*) *The Application of the Results of the preceding Section.*

Although the correlation of differences does not appear to provide a general method for obtaining the correlation of successive values of a variate after secular changes have been removed, the equations (xviii) and (xix) will be found of considerable assistance in certain cases.

* The figures will probably not warrant the taking of differences of much higher orders than 5th or 6th.

The results of the analysis given in the three illustrative problems below will be used in obtaining the values of various constants in the reduction of the experiments in the later sections. It seemed desirable to collect the algebra together in this way, but in reading this paper the reader may find it more convenient to pass on and refer back to the theory when occasion arises for the numerical application of the results.

Problem 1. In this and the following illustrations of the method of the preceding section, the notation of Section IV for the correlation of judgment will be used.

I shall suppose that we have m series of observations through the course of which there is some form of secular change; the means of the different series, or the values of ${}_pd$, varying considerably. The coefficients of correlation for the combined series, $\mathbf{R}_1$, $\mathbf{R}_2$, ... $\mathbf{R}_k$ ... $\mathbf{R}_s$ have been calculated, and also the single coefficient $\mathbf{R}_1'$, the correlation of the successive values of the observations (at intervals of 1) after the series means have been fitted together—i.e. after removal of secular change.

It is clear that $\Delta_1 y_t = \Delta_1 Y_t'$, where $y_t = d_1 + Y_t'$, within any one series, and

$$\sum_m \sum_{t=1}^{n} (\Delta_1 y_t \cdot \Delta_1 y_{t+k}) = \sum_m \sum_{t=1}^{n} (\Delta_1 Y'_t \cdot \Delta_1 Y'_{t+k}) \text{ etc.},$$

where $\sum_m$ again stands for summation for the m series, so that the 1st difference correlation equations (xviii) and (xix) are applicable, and become

$${}_1R_1 = \frac{-1 + 2\mathbf{R}_1 - \mathbf{R}_2}{2(1 - \mathbf{R}_1)} \quad {}_1R_k = \frac{-\mathbf{R}_{k-1} + 2\mathbf{R}_k - \mathbf{R}_{k+1}}{2(1 - \mathbf{R}_1)} \qquad k = 2, \text{ to } s-1 \quad \ldots(\text{xx}),$$

$$= \frac{-1 + 2\mathbf{R}_1' - \mathbf{R}_2'}{2(1 - \mathbf{R}_1')} \quad = \frac{-\mathbf{R}'_{k-1} + 2\mathbf{R}_k' - \mathbf{R}'_{k+1}}{2(1 - \mathbf{R}_1')} \qquad k = 2, \text{ to } s-1 \quad \ldots(\text{xxi}).$$

From (xx) we get the values of ${}_1R_k, k = 1, 2 \ldots s-1$, and using these and value of $\mathbf{R}_1'$ already supposed to be known, the $s-1$ equations (xxi) will give the $s-1$ unknowns $\mathbf{R}_2', \ldots \mathbf{R}_s'$.

The accuracy of this method will of course depend on the errors involved in the assumptions (*a*), (*b*), and (*c*) of page 37 above.

Problem 2. To obtain the coefficients of correlation of the successive residuals left after the ordinates of the "best" fitting straight lines have been subtracted from each of m series of observations, that is, after the removal of a linear sessional change as well as a secular change. In the notation of p. 35 these coefficients may therefore be written

$$\mathbf{R}_1'', \ \mathbf{R}_2'' \ldots \mathbf{R}_k'' \ldots.$$

In the first place let us obtain the constants of the straight line "best" fitting the 50 observations of Group 1 of a series; this can be done by the method of Least Squares.

If for any series the equation to the line is

$$y = d + b\left(t - \frac{n+1}{2}\right) \qquad (n = 50 \text{ as before}) \quad \ldots\ldots\ldots\ldots(\text{xxii}),$$

where the tth observation is

$$y_t = d + b\left(t - \frac{n+1}{2}\right) + Y_t,$$

we have that

$$K = \sum_{t=1}^{n} Y_t^2$$

$$= \sum_{t=1}^{n} \left\{y_t - d - b\left(t - \frac{n+1}{2}\right)\right\}^2, \text{ is to be a minimum,}$$

therefore

$$\frac{\partial K}{\partial d} = 0 \text{ and } \frac{\partial K}{\partial b} = 0,$$

or

$$\sum_{t=1}^{n} Y_t = 0 \text{ whence } \sum_{t=1}^{n} y_t = nd,$$

and

$$\sum_{t=1}^{n} \left\{y_t - d - b\left(t - \frac{n+1}{2}\right)\right\}\left(t - \frac{n+1}{2}\right) = 0$$

giving

$$\sum_{t=1}^{n} \left\{y_t\left(t - \frac{n+1}{2}\right)\right\} = b \sum_{t=1}^{n} \{t^2 - (n+1)\,t + \tfrac{1}{4}(n+1)^2\}.$$

Or, the first order product moment coefficient about the mean of y_t and t

$$\bar{p}_{11} = b\,\frac{(n^2-1)}{12},$$

giving for the constants of the best fitting line

$$d = d_1 = \frac{1}{n}\sum_{t=1}^{n} y_t$$

$$b = \frac{12}{(n^2-1)}\,\bar{p}_{11}.$$

The next step is to obtain the correlation of the successive residuals left after the ordinates of this line have been subtracted from the observations.

We shall have that

$$n\sigma_1\sigma_2\rho_1 = \sum_{t=1}^{n} \left\{d + b\left(t - \frac{n+1}{2}\right) + Y_t\right\}\left\{d + b\left(t+1 - \frac{n+1}{2}\right) + Y_{t+1}\right\}$$

$$- nd\left(d + \frac{y_{n+1} - y_1}{n}\right)$$

$$= d\sum_{t=1}^{n}(y_t + y_{t+1}) - nd^2 + b\sum_{t=1}^{n}\left\{\left(t+1-\frac{n+1}{2}-1\right)(y_{t+1}-d)\right.$$

$$\left. + \left(t - \frac{n+1}{2} + 1\right)(y_t - d)\right\} - b^2\sum_{t=1}^{n}\left(t - \frac{n+1}{2}\right)\left(t - \frac{n-1}{2}\right)$$

$$+ \sum_{t=1}^{n} Y_t\,Y_{t+1} - nd^2 - d\,(y_{n+1} - y_1)$$

$$= \sum_{t=1}^{n} Y_t\,Y_{t+1} + b\left\{2n\bar{p}_{11} + \frac{n+1}{2}(y_{n+1} - d) + \frac{n-1}{2}(y_1 - d) - y_{n+1} + y_1\right\}$$

$$- b^2\sum_{t=1}^{n}\left\{t^2 - nt + \frac{n^2-1}{4}\right\}$$

$$= \sum_{t=1}^{n} Y_t\,Y_{t+1} + b\left\{2n\bar{p}_{11} - nd + \frac{n+1}{2}\,y_1 + \frac{n-1}{2}\,y_{n+1}\right\} - b^2\frac{n\,(n^2-1)}{12}$$

$$= \sum_{t=1}^{n} Y_t\,Y_{t+1} + \frac{b^2}{12}\,n\,(n^2-1) + b\left\{\frac{n+1}{2}\,y_1 + \frac{n-1}{2}\,y_{n+1} - nd\right\},$$

and if ρ_1' be the correlation of the successive residuals and σ_1' and σ_2' the corresponding standard deviations in Groups 1 and 2, we have finally

$$\rho_1'\sigma_1'\sigma_2' = \rho_1\sigma_1\sigma_2 - \frac{b^2}{12}(n^2-1) - \frac{b}{2n}\{(n+1)y_1 + (n-1)y_{n+1} - 2nd\} \quad \ldots(\text{xxiii}).$$

Similarly we have

$$\begin{aligned} n\sigma_1^2 &= \sum_{t=1}^{n}\left\{d + b\left(t - \frac{n+1}{2}\right) + Y_t\right\}^2 - nd^2 \\ &= 2d\sum_{1}^{n}(y_t) - 2nd^2 + 2b\sum_{t=1}^{n}\left\{\left(t - \frac{n+1}{2}\right)(y_t - d)\right\} - b^2\sum_{t=1}^{n}\left(t - \frac{n+1}{2}\right)^2 + \sum_{t=1}^{n}Y_t^2 \\ &= \sum_{t=1}^{n}Y_t^2 + 2bn\bar{p}_{11} - \frac{b^2}{12}n(n^2-1) \\ &= \sum_{t=1}^{n}Y_t^2 + \frac{b^2}{12}n(n^2-1), \end{aligned}$$

whence it follows that

$$\sigma_1'^2 = \sigma_1^2 - \frac{b^2}{12}(n^2-1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{xxiv}).$$

And again,

$$\begin{aligned} n\sigma_2^2 &= \sum_{t=1}^{n}\left\{d + b\left(t + 1 - \frac{n+1}{2}\right) + Y_{t+1}\right\}^2 - n\left(d + b + \frac{Y_{n+1} - Y_1}{n}\right)^2 \\ &= 2d\sum_{t=1}^{n}y_{t+1} - 2nd^2 + 2b\sum_{t=1}^{n}\left\{\left(t + 1 - \frac{n+1}{2}\right)(y_{t+1} - d)\right\} - b^2\sum_{t=1}^{n}\left(t - \frac{n-1}{2}\right)^2 \\ &\quad + \sum_{t=1}^{n}Y^2{}_{t+1} - n\left(\frac{Y_{n+1} - Y_1}{n}\right)^2 - nb^2 - 2nbd - 2(b+d)(y_{n+1} - y_1 - nb) \\ &= n\sigma_2'^2 + 2bn\bar{p}_{11} + nb^2 - b^2\sum_{t=1}^{n}\left(t^2 - (n-1)t + \frac{(n-1)^2}{4}\right) \\ &\qquad\qquad + 2b\left(\frac{n+1}{2}y_1 + \frac{n-1}{2}y_{n+1} - nd\right) \\ &= n\sigma_2'^2 + \frac{b^2}{12}n(n^2-1) + b\{(n+1)y_1 + (n-1)y_{n+1} - 2nd\} \end{aligned}$$

$$\sigma_2'^2 = \sigma_2^2 - \frac{b^2}{12}(n^2-1) - \frac{b}{n}\{(n+1)y_1 + (n-1)y_{n+1} - 2nd\} \quad \ldots\ldots\ldots\ldots\ldots(\text{xxv}).$$

If the values of ρ_1' have been calculated by this means for each of the m series, we shall have for the combined series,

$$\mathbf{R}_1'' = \frac{\sum_m(\rho_1'\sigma_1'\sigma_2')}{\sqrt{\sum_m(\sigma_1'^2)\sum_m(\sigma_2'^2)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{xxvi}),$$

a modified form of equation (xii).

As we are subtracting the ordinates of a *different* straight line from each series, a modification of the first-difference equations may be necessary. The

1st order product moment coefficient, for the m combined series*, of successive first differences at intervals of k is given by

$$
\begin{aligned}
{}_1P_k &= \frac{1}{mn}\sum_m\sum_{t=1}^{n}(y_t - y_{t+1})(y_{t+k} - y_{t+k+1}) - \frac{1}{m^2}\left\{\sum_m \frac{y_1 - y_{n+1}}{n}\right\}\left\{\sum_m \frac{y_{k+1} - y_{k+n+1}}{n}\right\} \\
&= \frac{1}{mn}\sum_m\sum_{t=1}^{n}(Y_t - Y_{t+1} - b)(Y_{t+k} - Y_{t+k+1} - b) \\
&\qquad - \frac{1}{m^2}\left\{\sum_m\left(-b + \frac{Y_1 - Y_{n+1}}{n}\right)\right\}\left\{\sum_m\left(-b + \frac{Y_{k+1} - Y_{k+n+1}}{n}\right)\right\} \\
&= \frac{1}{mn}\sum_m\sum_{t=1}^{n}(Y_t - Y_{t+1})(Y_{t+k} - Y_{t+k+1}) - \left\{\sum_m \frac{Y_1 - Y_{n+1}}{nm}\right\}\left\{\sum_m \frac{Y_{k+1} - Y_{k+n+1}}{nm}\right\} \\
&\qquad - \frac{1}{m}\sum_m\left(b\frac{y_1 - y_{n+1} + y_{k+1} - y_{n+k+1}}{n}\right) + \left\{\sum_m \frac{b}{m}\right\}\left\{\sum_m \frac{y_1 - y_{n+1} + y_{k+1} - y_{n+k+1}}{mn}\right\} \\
&\qquad - \sum_m \frac{b^2}{m} + \left\{\sum_m \frac{b}{m}\right\}^2.
\end{aligned}
$$

Or finally,

$$
{}_1P_k = (-\mathbf{R}_{k-1}'' + 2\mathbf{R}_k'' - \mathbf{R}_{k+1}'')\,S''^2 - Q_k - \overline{b^2} \quad \text{.....(xxvii)},
$$

making the assumptions (a), (b), and (c) of p. 37, and where

1. $Q_k = \dfrac{1}{m}\sum_m\left(b\dfrac{y_1 - y_{n+1} + y_{k+1} - y_{n+k+1}}{n}\right) - \left\{\sum_m \dfrac{b}{m}\right\}\left\{\sum_m \dfrac{y_1 - y_{n+1} + y_{k+1} - y_{n+k+1}}{mn}\right\}$.

2. $\sqrt{\overline{b^2}}$ is the standard deviation of the b's.

There will be similar corrected expressions for the *standard deviations* of the combined first differences.

If we are justified in neglecting terms of the order of $Q_k + \overline{b^2}$, we may use the first difference equations,

$$
\left.
\begin{aligned}
{}_1R_k &= \frac{-\mathbf{R}_{k-1} + 2\mathbf{R}_k - \mathbf{R}_{k+1}}{2(1 - \mathbf{R}_1)} \\
&= \frac{-\mathbf{R}_{k-1}'' + 2\mathbf{R}_k'' - \mathbf{R}_{k+1}''}{2(1 - \mathbf{R}_1'')}, \quad k = 1, 2 \ldots s-1
\end{aligned}
\right\} \text{.....(xxviii)},
$$

where, as in Problem 1, the known $\mathbf{R}_k$'s will give the ${}_1R_k$'s, and it will only be necessary to calculate directly the one quantity $\mathbf{R}_1''$, in order to obtain

$$
\mathbf{R}_2'',\ \mathbf{R}_3'' \ldots \mathbf{R}_s''.
$$

Problem 3. In the last illustration it may happen that while $Q_k + \overline{b^2}$ is so small as to cause only a negligible error in the value of $\mathbf{R}_2''$ found from

$$
{}_1R_1 = \frac{-1 + 2\mathbf{R}_1'' - \mathbf{R}_2''}{2(1 - \mathbf{R}_1'')},
$$

* ${}_pb$ is the slope of best fitting line in the pth Series.

the cumulative effect of this error may be considerable in the value found for $\mathbf{R}_s''$ ($s = 12$, say). If then we take second differences

$$
\begin{aligned}
{}_2P_k &= \frac{1}{mn} \sum_m \sum_{t=1}^{n} (y_t - 2y_{t+1} + y_{t+2})(y_{t+k} - 2y_{t+k+1} + y_{t+k+2}) \\
&\quad - \frac{1}{m^2} \left\{ \sum_m \frac{y_1 - y_2 - y_{n+1} + y_{n+2}}{n} \right\} \left\{ \sum_m \frac{y_{k+1} - y_{k+2} - y_{k+n+1} + y_{k+n+2}}{n} \right\} \\
&= \frac{1}{mn} \sum_m \sum_{t=1}^{n} (Y_t - 2Y_{t+1} + Y_{t+2})(Y_{t+k} - 2Y_{t+k+1} + Y_{t+k+2}) \\
&\quad - \frac{1}{m^2} \left\{ \sum_m \frac{Y_1 - Y_2 - Y_{n+1} + Y_{n+2}}{n} \right\} \left\{ \sum_m \frac{Y_{k+1} - Y_{k+2} - Y_{k+n+1} + Y_{k+n+2}}{n} \right\} \\
&= (\mathbf{R}_{k-2}'' - 4\mathbf{R}_{k-1}'' + 6\mathbf{R}_k'' - 4\mathbf{R}_{k+1}'' + \mathbf{R}_{k+2}'')\, S''^2,
\end{aligned}
$$

and is independent of the differing values of the b's.

The appropriate equations are in fact of type,

$$
{}_2R_k = \frac{\mathbf{R}_{k-2}'' - 4\mathbf{R}_{k-1}'' + 6\mathbf{R}_k'' - 4\mathbf{R}_{k+1}'' + \mathbf{R}_{k+2}''}{2\,(3 - 4\mathbf{R}_1'' + \mathbf{R}_2'')} \quad \text{.....(xxix),}
$$

for $k = 1, 2, 3 \ldots s-2$, where $\mathbf{R}_{-1}'' = \mathbf{R}_1''$ etc. and $\mathbf{R}_0'' = 1$. Then using the known value of $\mathbf{R}_1''$, and that of $\mathbf{R}_2''$, found as in Problem 2 from the first difference equation, these $s-2$ equations will give the $s-2$ unknowns $\mathbf{R}_3'' \ldots \mathbf{R}_s''$.

It is clear that similar methods could be applied in the case of sessional changes of higher order, but I have taken the algebra in these three Problems, as the results will be used in the reduction of the experiments later on. The general explanation and equations may have appeared long, but the actual calculation in any particular case of such quantities as ${}_1R_1, {}_1R_2, \ldots {}_1R_k$, or ${}_2R_1, \ldots {}_2R_k$, and then of $\mathbf{R}_2', \ldots \mathbf{R}_k'$, and $\mathbf{R}_2'', \ldots \mathbf{R}_k''$, is exceedingly simple, and far shorter than a direct calculation from the crude figures would be. In two cases the correlations were calculated both by the difference correlation method and directly without approximation, and the agreement of the former results with the latter established confidence in this method of approximation.

VI. Experiment *A* (Trisection). Reduction of Observations.

(*a*) *The individual Series.*

The observations of this Experiment have been reduced in more detail than in the other cases; the values of ρ_k, $k = 1, 2, \ldots 13$, were found separately for each series, and these and the values of d and σ—the means and standard deviations of the Groups—are given in Tables I, II and III. Several points of interest will be noted; in the first place the observations have a marked tendency to decrease (i.e. for the estimate of a third to become smaller) both in the course of a series (as is seen by the general decrease of d_k as k increases) and also in passing from the earlier to the later series. These are examples of what have been termed Sessional and Secular Changes. These changes are illustrated in Figure 3 where the centres of the circles give the values of d_1 for each Series, the length of the dotted lines from either side of these points representing the standard deviations σ_1, and the

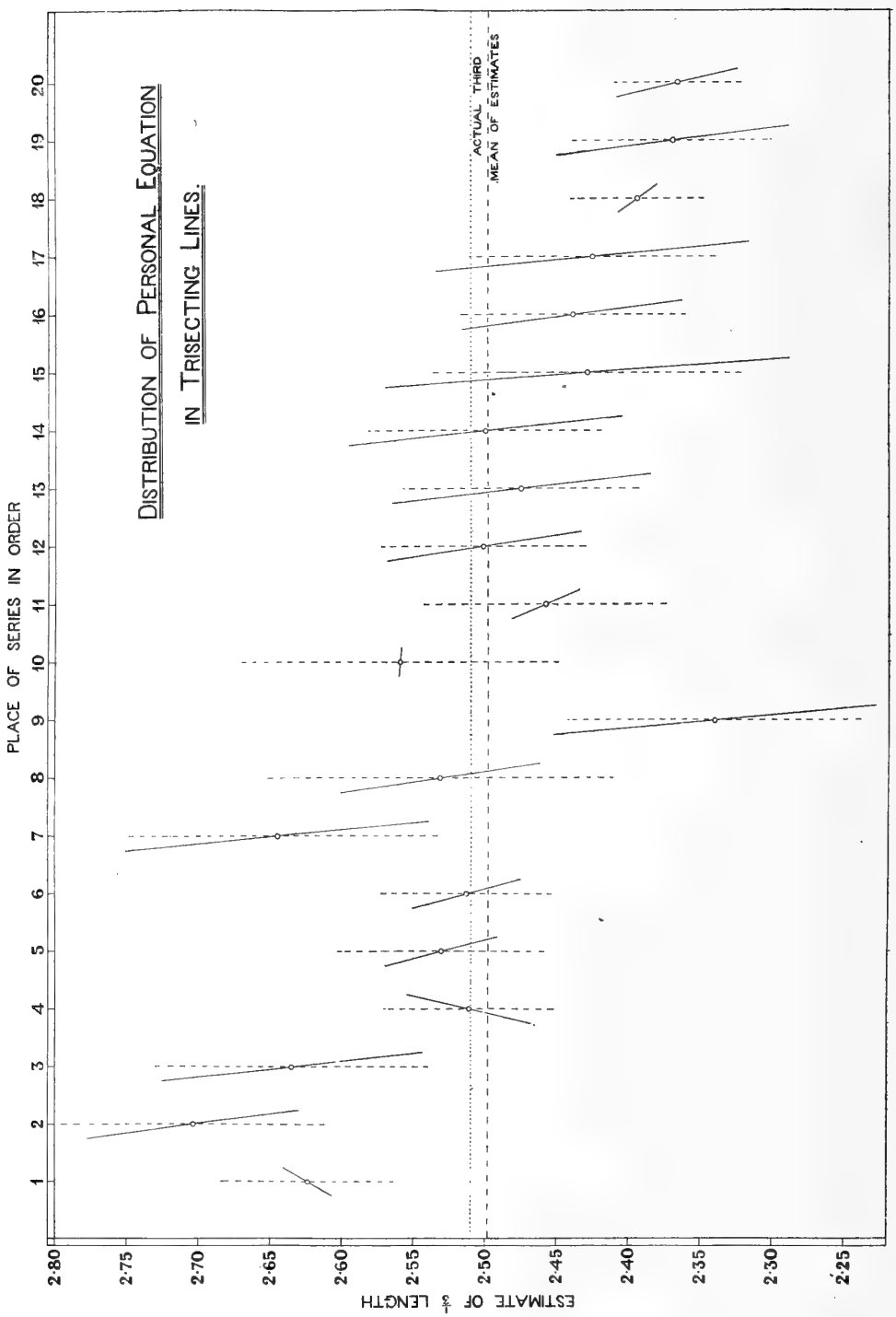

DISTRIBUTION OF PERSONAL EQUATION
IN TRISECTING LINES.
ACTUAL THIRD
MEAN OF ESTIMATES
PLACE OF SERIES IN ORDER
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
ESTIMATE OF ⅓ LENGTH
2·80 2·75 2·70 2·65 2·60 2·55 2·50 2·45 2·40 2·35 2·30 2·25

Fig. 3.

continuous lines through the points representing the "best" fitting straight lines for the 50 observations of Group 1; the slopes of these last lines, or constants ${}_pb$, have been calculated by the Least Square method as in Problem 2, p. 41, and their values are given in the 3rd column of Table IV.

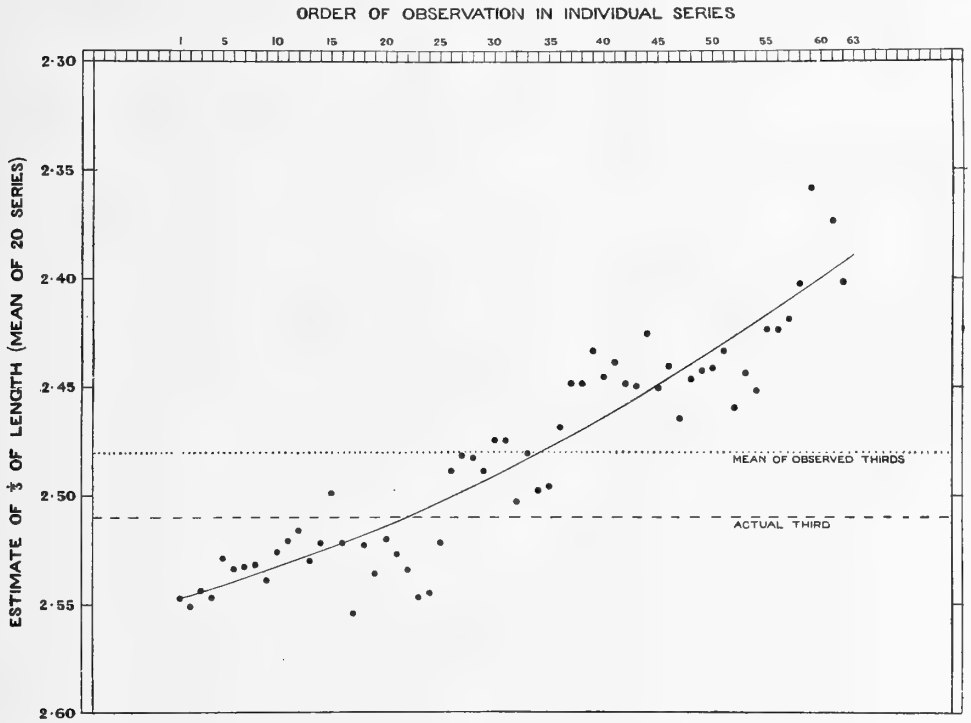
Another way of examining the sessional change, and of obtaining a typical representation of it, is to calculate the average values for the 20 series of y_t the tth observation in a series; thus

$$\bar{y}_t = \frac{1}{m}\sum_m y_t = \frac{1}{m}\sum_m (\bar{d} + Y_t)$$
$$= \bar{D} + \frac{1}{m}\sum_m Y_t,$$

where ${}_p\bar{d}$ stands for the mean of the pth series (63 observations) as opposed to ${}_pd_k$, the mean of a particular Group k of that series.

The values of Y_t represent the sessional variation in any series about the mean of that series or session of observations, and the sequence $\bar{y}_t - \bar{D}$, $t = 1, 2, 3, \ldots 63$, will clearly represent the mean sessional change. The values of $\bar{y}_t$ are given at the end of Table II and have been plotted in Figure 4, where they have been fitted with the second order parabola (calculated by least squares)

$$y = \cdot 486 + \cdot 00255t - \cdot 0000189t^2 \ldots\ldots\ldots\ldots\ldots\ldots (xxx).$$



TRISECTION EXPERIMENT. MEAN SESSIONAL CHANGE.

Fig. 4.

Figures 3 and 4 together show very clearly the marked sessional change; while the former shows that except in a few series, notably Series I, IV and X, the regression is remarkably constant in its value, the latter indicates that the sessional change is better represented by a parabola than by a straight line.

The sessional change can also be represented numerically with the help of the correlation ratio of y_t upon t. If we are dealing with the observations freed from the secular change, that is after the removal of the means ${}_p\bar{d}$ from the 63 observations of the pth series we have η_{y_t} given by

$$\eta_{y_t}^2 = \frac{\sum_{t=1}^{63} (\bar{y}_t - \bar{D})^2}{63S'^2}, \text{ where } S'^2 = \frac{1}{1260} \sum_m \sum_{t=1}^{63} ({}_p y_t - {}_p\bar{d})^2,$$

or S' is the standard deviation of the whole 1260 observations after the removal of the secular term*. Then the ratio of the mean square distance of every observation from the regression line or line of means $\bar{y}_t$, to the standard deviation of the observations is

$$\frac{\sqrt{\frac{1}{1260} \sum_{t=1}^{63} \sum_m (y_t - \bar{y}_t)^2}}{S'} = \sqrt{1 - \eta_{y_t}^2} \quad \text{.....(xxxi),}$$

where $\sum_m$ indicates summation for the 20 series.

This is a measure of the closeness of fit of the observations in a series to the mean sessional change as represented by the values $\bar{y}_t$; the larger η_{y_t} and therefore the smaller $\sqrt{1 - \eta_{y_t}^2}$ is, the more nearly does a sessional change of the same form recur in series after series. A comparison of the values of $\sqrt{1 - \eta_{y_t}^2}$ for the different experiments will show the relative significance of their mean sessional changes.

In the present case the value of η_{y_t} is found to be ·579 ± ·013, while

$$\sqrt{1 - \eta_{y_t}^2} = \cdot 815.$$

It would be an interesting problem to obtain the correlation of the successive residuals left after the ordinates of the "best" fitting parabola for each series had been subtracted from the observations of that series; but although this has not been done, a fair idea of the degree to which the correlation of the successive judgments in the individual series is due to the sessional change can be obtained by removing the "best" fitting straight lines from each series. The values calculated for the ${}_pb$'s have been referred to above, and using these and the equations (xxii)—(xxiv) of pp. 41—43, the values of σ_1' and ρ_1', or the standard deviations and correlations of successive observations freed from the linear sessional changes, have been calculated and are given in the 4th and 6th columns of Table IV. The ρ_1''s are all less than the corresponding ρ_1's, except in Series X where they are

* Actually it is only the values of the Group Standard Deviations S_1', S_2' ... S_{14}' which have been calculated; they are not all equal (as shown in Table V) owing to the sessional change in standard deviation, but an approximation to S' sufficiently accurate for the purpose will be given by taking

$$S'^2 = \tfrac{1}{14} \{S_1'^2 + S_2'^2 + \ldots + S_{14}'^2\}.$$

TABLE I. Trisection.

Coefficients of Correlation for Separate Series, Dates, Times and Remarks.

Series	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}	ρ_{11}	ρ_{12}	ρ_{13}	Date (1920)	Remarks	Time taken
I	+·3008	+·2435	−·0416	+·1813	+·0189	+·0021	+·0301	+·0041	+·1183	+·1075	+·0334	−·0938	−·0237	a.m. 7 May	—	—
II	+·5485	+·5134	+·4577	+·4283	+·4391	+·3557	+·3651	+·2673	+·1842	+·1716	+·2197	+·2756	+·2265	a.m. 9 „	—	—
III	+·5560	+·5065	+·3653	+·3981	+·3778	+·2284	+·2642	+·2944	+·4221	+·3317	+·3953	+·4397	+·5145	a.m. 16 „	—	—
IV	−·0460	+·2168	+·0399	−·0942	+·0948	−·0273	+·2952	+·0783	−·0510	+·0682	+·0505	−·1315	+·0307	a.m. 17 „	After measuring plates for 1½—2 hrs at Observatory	—
V	+·3234	+·2309	+·2022	+·1691	+·0485	+·0963	+·1570	−·0047	−·0776	−·0442	−·1614	−·0936	−·1219	a.m. 22 „	—	—
VI	+·3390	+·4403	+·1749	+·1472	+·2066	−·0580	+·1486	+·0733	+·0663	+·1874	+·1848	+·4314	+·3999	a.m. 23 „	—	—
VII	+·6457	+·4532	+·3681	+·2193	+·2849	+·2142	+·1076	+·0509	+·0186	+·0746	+·1200	+·1348	+·2051	a.m. 11 July		6m 45s
VIII	+·6089	+·5938	+·3917	+·3340	+·3702	+·3829	+·2660	+·4145	+·3246	+·3802	+·4091	+·3602	+·3013	p.m. 11 July	Tired and not very fit	7 0
IX	+·7075	+·4058	+·3596	+·3458	+·3359	+·4014	+·3555	+·3506	+·3637	+·3376	+·2206	+·1252	+·1420	a.m. 12 „		5 45
X	+·7151	+·6392	+·3890	+·1505	+·0283	−·0735	+·0467	+·0615	+·1381	+·2571	+·2252	+·2707	+·0433	p.m. 12 „	Was aware of underestimation in IX	6 15
XI	+·7381	+·4814	+·3095	+·1658	+·0993	+·0295	−·1086	−·2638	−·3844	−·3775	−·4303	−·4428	−·4122	a.m. 13 „	—	5 30
XII	+·6360	+·5137	+·4434	+·4119	+·2917	+·2638	+·2298	+·0611	+·0482	+·1065	+·0569	−·1401	−·0515	p.m. 13 „	—	6 15
XIII	+·6897	+·5563	+·3618	+·3391	+·3109	+·2961	+·3247	+·3010	+·3204	+·2917	+·2201	+·1660	+·0862	a.m. 14 „	—	6 0
XIV	+·7965	+·6875	+·6018	+·4370	+·2968	+·2458	+·2182	+·1141	+·1273	+·1751	+·2697	+·3373	+·4034	p.m. 14 „	Eyes tired from much reading	6 15
XV	+·8568	+·8256	+·7196	+·6029	+·5277	+·3665	+·2920	+·1950	+·1249	+·1016	+·1219	+·1821	+·2741	p.m. 19 „	Plate measuring in morning at — [Observatory	5 45
XVI	+·7412	+·5754	+·4567	+·3267	+·2352	+·2133	+·0812	+·0544	+·1275	+·1142	−·0066	−·0560	−·0397	a.m. 20 „		6 0
XVII	+·6556	+·6168	+·4732	+·5055	+·4488	+·5434	+·5380	+·5920	+·3595	+·4370	+·2971	+·4283	+·2851	a.m. 21 „	—	6 15
XVIII	+·3144	+·0829	−·0376	−·1662	−·2014	−·1357	+·0463	+·0697	+·0665	−·0834	−·0554	+·0658	+·1440	a.m. 22 „	—	6 30
XIX	+·7219	+·5919	+·6033	+·5623	+·5148	+·4560	+·4145	+·4380	+·3376	+·1793	+·1527	+·1000	−·0227	a.m. 24 „	—	6 15
XX	+·5072	+·2875	+·2672	+·2536	+·1140	+·1257	+·2581	+·2481	+·2465	+·3125	+·3233	+·2468	+·2321	a.m. 27 „	—	6 45

Probable Errors of Coefficient of Correlation calculated from 50 Pairs of the Variates.

Value of ρ ...	·80	·70	·60	·50	·40	·30	·20	·10
Value of P.E.	± ·0343	± ·0486	± ·0610	± ·0715	± ·0801	± ·0868	± ·0916	± ·0944

Mean time taken for a series of 70 observations (including the 7 preliminary trials*) 6m 14s

Mean interval between records of judgment ... ... 5s·4

* See p. 28 footnote.

TABLE II. Trisection.

Means of Series Groups (from origin 2·5000 inches).

Series	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11	Group 12	Group 13	Group 14
I	+·1238	+·1192	+·1208	+·1176	+·1184	+·1188	+·1222	+·1234	+·1226	+·1216	+·1206	+·1202	+·1164	+·1122
II	+·2036	+·2032	+·1996	+·1930	+·1882	+·1826	+·1754	+·1718	+·1678	+·1648	+·1598	+·1542	+·1516	+·1484
III	+·1350	+·1350	+·1312	+·1280	+·1268	+·1230	+·1210	+·1212	+·1168	+·1106	+·1022	+·0962	+·0936	+·0924
IV	+·0114	+·0124	+·0128	+·0112	+·0126	+·0170	+·0206	+·0210	+·0206	+·0178	+·0178	+·0156	+·0130	+·0140
V	+·0308	+·0298	+·0292	+·0284	+·0262	+·0256	+·0232	+·0184	+·0138	+·0078	+·0036	−·0042	−·0100	−·0152
VI	+·0132	+·0110	+·0078	+·0074	+·0040	+·0012	·0000	−·0030	−·0062	−·0122	−·0168	−·0206	−·0254	−·0298
VII	+·1448	+·1390	+·1374	+·1378	+·1372	+·1384	+·1392	+·1362	+·1346	+·1292	+·1274	+·1230	+·1220	+·1214
VIII	+·0314	+·0270	+·0234	+·0210	+·0184	+·0120	+·0086	+·0040	+·0028	−·0038	−·0054	−·0078	−·0106	−·0132
IX	−·1596	−·1630	−·1662	−·1708	−·1730	−·1744	−·1760	−·1760	−·1780	−·1786	−·1804	−·1806	−·1816	−·1814
X	+·0590	+·0582	+·0580	+·0624	+·0680	+·0710	+·0716	+·0736	+·0732	+·0722	+·0652	+·0610	+·0534	+·0488
XI	−·0418	−·0414	−·0384	−·0368	−·0364	−·0386	−·0424	−·0470	−·0464	−·0470	−·0456	−·0428	−·0408	−·0404
XII	+·0014	+·0004	+·0004	−·0004	−·0036	−·0040	−·0070	−·0114	−·0130	−·0140	−·0146	−·0182	−·0186	−·0192
XIII	−·0248	−·0264	−·0266	−·0258	−·0264	−·0284	−·0322	−·0356	−·0384	−·0412	−·0452	−·0490	−·0510	−·0520
XIV	·0000	−·0022	−·0052	−·0078	−·0126	−·0178	−·0226	−·0274	−·0334	−·0398	−·0462	−·0486	−·0522	−·0568
XV	−·0710	−·0760	−·0814	−·0858	−·0920	−·0968	−·1032	−·1080	−·1126	−·1184	−·1216	−·1224	−·1244	−·1254
XVI	−·0610	−·0632	−·0648	−·0690	−·0732	−·0770	−·0824	−·0870	−·0910	−·0946	−·0982	−·1016	−·1022	−·1026
XVII	−·0746	−·0808	−·0856	−·0892	−·0936	−·0976	−·1004	−·1010	−·1044	−·1106	−·1142	−·1170	−·1184	−·1218
XVIII	−·1056	−·1086	−·1122	−·1142	−·1132	−·1144	−·1144	−·1146	−·1176	−·1194	−·1206	−·1222	−·1246	−·1272
XIX	−·1300	−·1332	−·1362	−·1412	−·1460	−·1510	−·1538	−·1548	−·1564	−·1608	−·1638	−·1684	−·1696	−·1726
XX	−·1334	−·1330	−·1330	−·1348	−·1370	−·1386	−·1404	−·1424	−·1450	−·1456	−·1476	−·1496	−·1496	−·1512
Means (*D*)	−·00237	−·00463	−·00645	−·00845	−·01036	−·01245	−·01465	−·01693	−·01951	−·02310	−·02618	−·02914	−·03145	−·03358

Mean Values of the "t^{th}" Observation of each Series.

t	$\bar{y}_t$	t	$\bar{y}_t$	t	$\bar{y}_t$
1	2·547	22	2·534	43	2·450
2	·551	23	·547	44	·426
3	·544	24	·545	45	·451
4	·547	25	·522	46	·441
5	·529	26	·489	47	·465
6	·534	27	·482	48	·447
7	·533	28	·483	49	·443
8	·532	29	·489	50	·442
9	·539	30	·475	51	·434
10	·526	31	·475	52	·460
11	·521	32	·503	53	·444
12	·516	33	·481	54	·452
13	·530	34	·498	55	·424
14	·522	35	·496	56	·424
15	·499	36	·469	57	·419
16	·522	37	·449	58	·403
17	·554	38	·449	59	·359
18	·523	39	·434	60	·371
19	·536	40	·446	61	·374
20	·520	41	·439	62	·402
21	2·527	42	2·449	63	2·423

Mean of all observations = 2·4803

TABLE III. TRISECTION.

Standard Deviations of Series Groups (in inches).

Series	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11	Group 12	Group 13	Group 14
I	·06093	·06575	·06599	·06739	·06682	·06672	·06382	·06263	·06343	·06495	·06565	·06635	·06926	·07217
II	·09184	·09197	·09510	·09992	·10373	·11079	·11767	·11960	·12158	·12161	·12354	·12654	·12813	·12962
III	·09561	·09561	·09474	·09506	·09427	·09586	·09548	·09526	·09840	·10851	·11789	·12381	·12526	·12385
IV	·05902	·05857	·05828	·06024	·05935	·05900	·05908	·05941	·05948	·06084	·06084	·06509	·06856	·06726
V	·07210	·07204	·07191	·07167	·07113	·07156	·07530	·07661	·07962	·08117	·08376	·08602	·08920	·08963
VI	·05921	·05984	·06087	·06104	·06274	·06590	·06699	·06935	·07230	·07718	·08184	·08260	·08358	·08629
VII	·11154	·11190	·11043	·11095	·11036	·11104	·11198	·10962	·10823	·10605	·10457	·10181	·10178	·10182
VIII	·12060	·12592	·13248	·13538	·13719	·13882	·13871	·14170	·14216	·14421	·14526	·14515	·14478	·14430
IX	·10233	·10443	·10490	·10777	·10804	·10793	·10711	·10711	·10528	·10486	·10479	·10474	·10372	·10402
X	·11141	·11226	·11229	·10950	·10768	·10445	·10360	·10258	·10326	·10400	·11265	·11697	·11872	·11543
XI	·08435	·08450	·08814	·09026	·09068	·09086	·09229	·09181	·09176	·09194	·09227	·09239	·09112	·09086
XII	·07105	·07071	·07071	·07082	·07028	·06977	·06978	·06920	·06914	·06888	·06795	·06755	·06732	·06693
XIII	·08244	·08116	·08098	·08154	·08153	·08257	·08278	·08446	·08422	·08489	·08521	·08240	·08269	·08219
XIV	·08141	·07961	·07697	·07658	·07781	·07747	·07807	·08197	·08267	·08744	·09428	·09510	·09659	·09756
XV	·10711	·10387	·09996	·09621	·09334	·08997	·08332	·08005	·07560	·07401	·07226	·07212	·07341	·07395
XVI	·07819	·07511	·07395	·06914	·06565	·06084	·05767	·05182	·05341	·05587	·05799	·06168	·06162	·06154
XVII	·08594	·08501	·07836	·07756	·07576	·07651	·07507	·07511	·07292	·07604	·07351	·07097	·07012	·07036
XVIII	·04644	·04821	·05247	·05273	·05365	·05379	·05379	·05390	·05655	·05863	·06065	·06217	·06087	·06040
XIX	·06920	·06866	·06997	·06678	·06579	·06552	·06325	·06198	·06157	·05949	·05706	·05696	·05550	·05351
XX	·04443	·04446	·04446	·04531	·04277	·04285	·04350	·04343	·04291	·04319	·04379	·04477	·04477	·04403
S_k'	·08450	·08478	·08499	·08524	·08502	·08533	·08535	·08545	·08576	·08722	·08908	·09010	·09072	·09065

The values of S_k at the bottom of the Table have been taken from Table VI below.

equal, but though there is in general a considerable reduction, it is clear that neither a linear sessional change nor a parabolic one of the form represented by equation (xxx) account for the greater part of the correlation of successive judgments.

The coefficients ρ_1 and also ρ_1' vary considerably from series to series, but there is no very marked progressive secular change. On the whole both ρ_1 and ρ_1' are large when the standard deviation is large, and a measure of this correspondence will be given by the correlation of ρ and σ. This can be obtained most readily, and with sufficient accuracy for the purpose, from the correlation of the ranks of these variates, by the method referred to in *Biometrika*, Vol. x. p. 416*.

The results are

$$\text{correlation between } \rho_1 \text{ and } \sigma_1 + \cdot 52 \pm \cdot 11 = r_{\rho\sigma},$$
$$\text{„ \quad „ \quad } \rho_1' \text{ and } \sigma_1' + \cdot 60 \pm \cdot 10 = r'_{\rho\sigma}.$$

The difference is not significant, and we may draw the conclusion which could not have been assumed *a priori*, that the correlation of successive judgments is larger when the variations in judgment are larger, and that this relationship does not appear to be reduced when the large linear sessional change has been removed. Large values of σ might have implied erratic observation and small relation between

TABLE IV. *Constants of Individual Series* (*Trisection*).

(The definition of these constants is given on p. 35.)

1	2	3	4	5	6	7	8
Series	d_1	b	ρ_1'	ρ_1	σ_1'	σ_1	σ_δ
I	2·6238	+ ·000673	+ ·2925	+ ·3008	·06015	·06093	·0721
II	·7036	− ·002964	+ ·4149	+ ·5485	·08125	·09182	·0873
III	·6350	− ·003626	+ ·3643	+ ·5560	·08001	·09561	·0901
IV	·5114	+ ·001718	− ·2521	− ·0460	·05356	·05902	·0854
V	·5309	− ·001555	+ ·2520	+ ·3234	·06853	·07210	·0839
VI	·5132	− ·001529	+ ·2270	+ ·3390	·05495	·05921	·0681
VII	·6448	− ·004244	+ ·4918	+ ·6457	·09322	·11154	·0939
VIII	·5314	− ·002788	+ ·5478	+ ·6089	·11369	·12060	·1067
IX	·3404	− ·004477	+ ·4979	+ ·7075	·07935	·10233	·0783
X	·5590	− ·000036	+ ·7151	+ ·7151	·11141	·11141	·0841
XI	·4582	− ·000972	+ ·7320	+ ·7381	·08317	·08435	·0610
XII	·5014	− ·002720	+ ·4851	+ ·6360	·05923	·07105	·0606
XIII	·4752	− ·003594	+ ·5101	+ ·6897	·06409	·08244	·0649
XIV	·5000	− ·003818	+ ·6433	+ ·7965	·05993	·08141	·0519
XV	·4290	− ·005588	+ ·6810	+ ·8568	·07051	·10711	·0573
XVI	·4390	− ·003071	+ ·6408	+ ·7412	·06441	·07819	·0562
XVII	·4254	− ·004369	+ ·2569	+ ·6556	·05840	·08594	·0713
XVIII	·3944	− ·000580	+ ·2870	+ ·3144	·04568	·04644	·0544
XIX	·3700	− ·003236	+ ·4935	+ ·7219	·05107	·06920	·0516
XX	2·3666	− ·001725	+ ·2850	+ ·5072	·03680	·04443	·0441

Mean value of $b = -\cdot 002425$.

* The theory is based on the hypothesis that the variates follow a normal distribution, and though this may not be strictly true for the ρ_1's and σ_1's the method probably gives a sufficiently accurate approximation to the value of their correlation.

successive judgments, and at the same time high correlation might have been expected to result in small variation. The significance of this will be discussed in the concluding sections of this paper.

In Table III giving the σ's, it will be seen that in general the standard deviations increase in the later groups; though this may be due in part to the parabolic form of the sessional change, with its tendency to an increasing drop towards the end of the series, it is possible that it also indicates a fatigue effect setting in, and causing the later observations in a series to be more erratic; the same phenomenon appears in the Bisection Experiment where there is no appreciable sessional change within the series. It may in fact be looked on as a sessional change in the standard deviation.

At the end of Table I are given the dates on which the different series were carried out, remarks noted at the time as to the condition of the observer, and, for the last 14 series, the time taken to mark off the 70 forms*. It will be seen that there was a large gap between the times of carrying out the first six and the last fourteen series, and this interval of nearly two months broke the continuity of the secular change in the means of the series. In Figure 5 the means of Group 1 (or

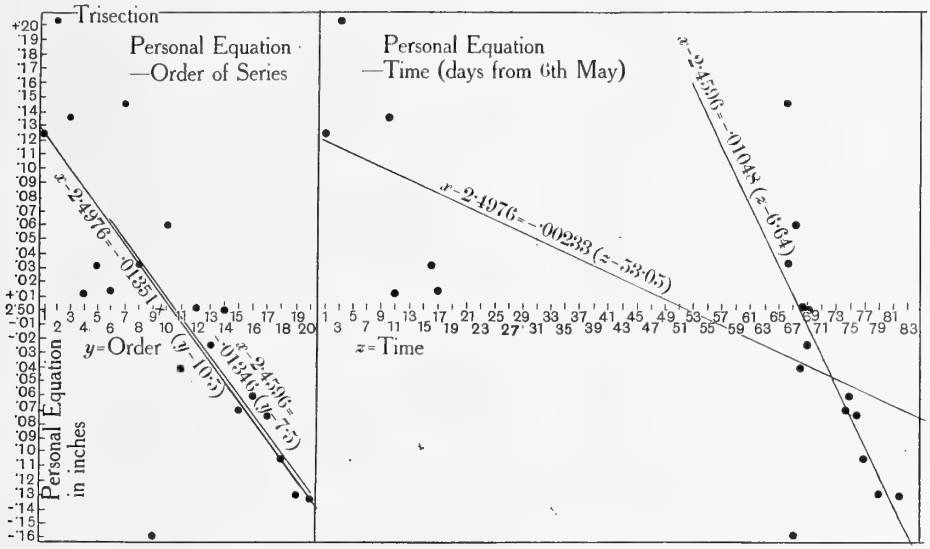


Fig. 5.

the d_1's of each series) have been plotted firstly with the order of series and secondly with the date of series.

If x is the personal equation, or mean value of the observations of Group 1 of a series measured in inches,

y the number or order of a Series,

z the number of days between the 6th May and the date on which the series was carried out,

* Reference to the 7 trial forms first marked, in addition to the 63 of the Series proper, is made on p. 28 in footnote.

we have for the regression straight line of x on y

$$x - 2{\cdot}4976 = -{\cdot}01351\,(y - 10{\cdot}5) \quad \text{.....(xxxii),}$$

and for the regression of x on z

$$x - 2{\cdot}4976 = -{\cdot}00233\,(z - 53{\cdot}05) \quad \text{... .....(xxxiii),}$$

and these lines have been drawn in the diagram.

The corresponding coefficients of correlation between (1) personal equation and order, (2) personal equation and time, and (3) time and order, a meaningless coefficient but required to find the partial correlations, are

$$(1)\ r_{xy} = -{\cdot}800 \pm {\cdot}054,$$
$$(2)\ r_{xz} = -{\cdot}692 \pm {\cdot}079,$$
$$(3)\ r_{yz} = +{\cdot}882 \pm {\cdot}033,$$

and the partial correlations are

$$r_{xy.z} = -{\cdot}559 \pm {\cdot}104,$$
$$r_{xz.y} = +{\cdot}049 \pm {\cdot}150.$$

But the interval between the May and July series was so large, that the series should perhaps be considered as forming two groups, one of six and the other of fourteen. Taking the last fourteen series, we have the regression lines

$$x - 2{\cdot}4596 = -{\cdot}01346\,(y - 7{\cdot}5),$$

the Series VII being given the order 1, VIII, 2 etc., and

$$x - 2{\cdot}4596 = -{\cdot}01048\,(z - 6{\cdot}64),$$

z being the days between 10th July and date of Series. These lines have also been drawn on the diagrams.

The correlations are

$$(1)\ r_{xy} = -{\cdot}674 \pm {\cdot}098,$$
$$(2)\ r_{xz} = -{\cdot}673 \pm {\cdot}099,$$
$$(3)\ r_{yz} = +{\cdot}956 \pm {\cdot}016,$$

giving partial correlations

$$r_{xy.z} = -{\cdot}143 \pm {\cdot}177,$$
$$r_{xz.y} = -{\cdot}138 \pm {\cdot}177.$$

The point of interest is this: there is a secular change in personal equation from series to series; is this change more closely related to the number of series or sessions that have gone before (that is, almost, to the experience gained), or is it due to some change with time in the observer's outlook? Suppose that it was arranged to carry out observations on a number of different days with varying intervals of time between them, and that on each day a certain number of different series of observations or sessions were undertaken at regular intervals of perhaps an hour or less; any series could then be classified as the pth series of the qth day. Then $r_{xy.z}$ (the partial correlation of personal equation and order, time being kept constant) would give a measure of the relationship between change in personal equation and order of series in any one day. This will not necessarily be the same

as the sessional change, for it has been supposed that this latter occurs only during the course of a sitting, and is broken by the interval of rest in between. On the other hand if we take all the pth series of the various days, then $r_{xz.y}$ (the partial correlation of personal equation and time, order being constant) gives the relation between change in personal equation and time, taken over a long period.

The long break in the middle of the Trisection Experiment takes away any real significance from the difference between $r_{xy.z}$ (− ·559) and $r_{xz.y}$ (+ ·049) for the twenty series, and in the case of the last fourteen series these coefficients are equal (− ·143 and − ·138), because the intervals between the series were nearly uniform. In the Timing Experiments, C and D, the arrangement of the series in groups on consecutive days leads to considerably more interesting results*.

A comparative measure of the consistency of the consecutive judgments in the different series, is the standard deviation of first differences, or

$$\sigma_\delta = \sqrt{\frac{\sum_{t=1}^{n} (y_{t+1} - y_t)^2}{n}} = \sigma_1 \sqrt{2(1 - \rho_1)}$$

approximately. The values of this expression are given in the 8th column of Table IV.

Now suppose we compare the constants in Table IV, the dates and remarks at the end of Table I and the diagrammatic representation of seven of the series given in Figure 6. The first series to be remarked on is IV; most of the series were carried out at the beginning of the morning before any other work, and it is possibly the fact that IV was done soon after a spell of measuring spectrograms with a Zeiss comparator that explains the exceptional values of ρ_1 and ρ_1', namely $\rho_1 = -\cdot 0460$, $\rho_1' = -\cdot 2521$. The σ_δ, or standard deviation of 1st differences, is no higher than for the other series done at about the same time (in May), and the σ_1 is lower. The first graph in Figure 6 gives the diagram of this series; the rapid fluctuations in judgment about a very steady, if slowly changing, personal equation may perhaps have some physiological significance. In the second and third graphs of Figure 6 are represented two of the four Series VII—X which were done when the observer was not very fit; they have large values for σ_1, and the σ_δ's are large compared with those of the ten series which follow, showing that the judgments were rather erratic; the correlation is however high. In VIII there is a great jump between the 44th and 45th judgments, from 2·22 to 2·66, and the gradual drop down, which follows, to 2·20 (for the 52nd judgment) is a good example of a way in which successive judgments are correlated. In Series XI (not represented among the graphs) there appears to be a periodic variation, for the correlation falls steadily from $\rho_1 = +\cdot 7381$ to $\rho_{12} = -\cdot 4428$.

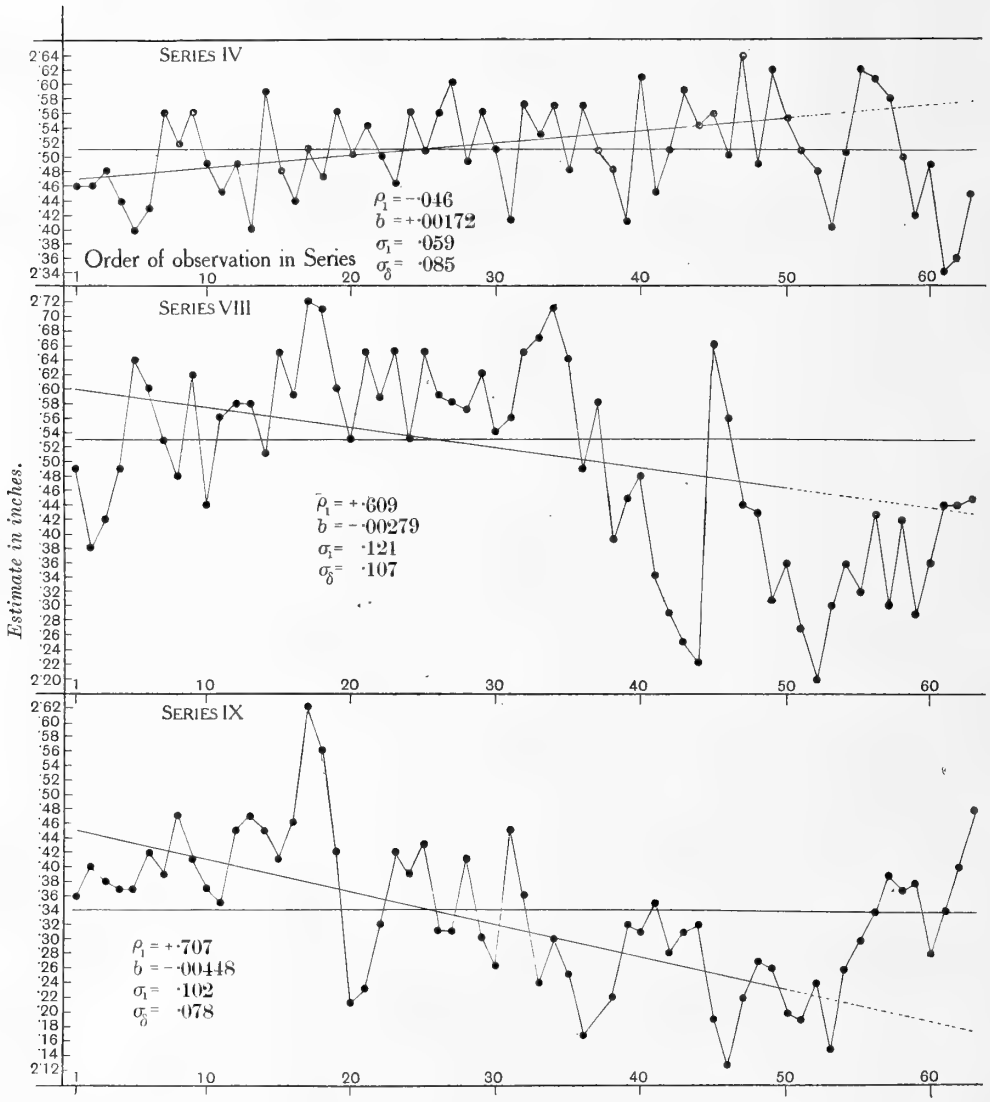
XIII, XV and XVI are typical highly correlated series with large sessional changes; the σ_δ's as well as the σ_1's are considerably smaller than in the series VII—X. In examining the fourth to sixth graphs we notice what may be called the large scale correlated variations superimposed upon the linear sessional change;

* See pp. 70, 75 and 83, and below.

it is due to these variations that the values of ρ_1' remain so large, and it is their absence that makes the correlation in IV so low. The last graph (Series XX) is that for which the σ_1 and σ_δ are least, and yet ρ_1 is quite high (+ ·507).

As a last instance of these points, we may compare the constants* for XV and XVIII:

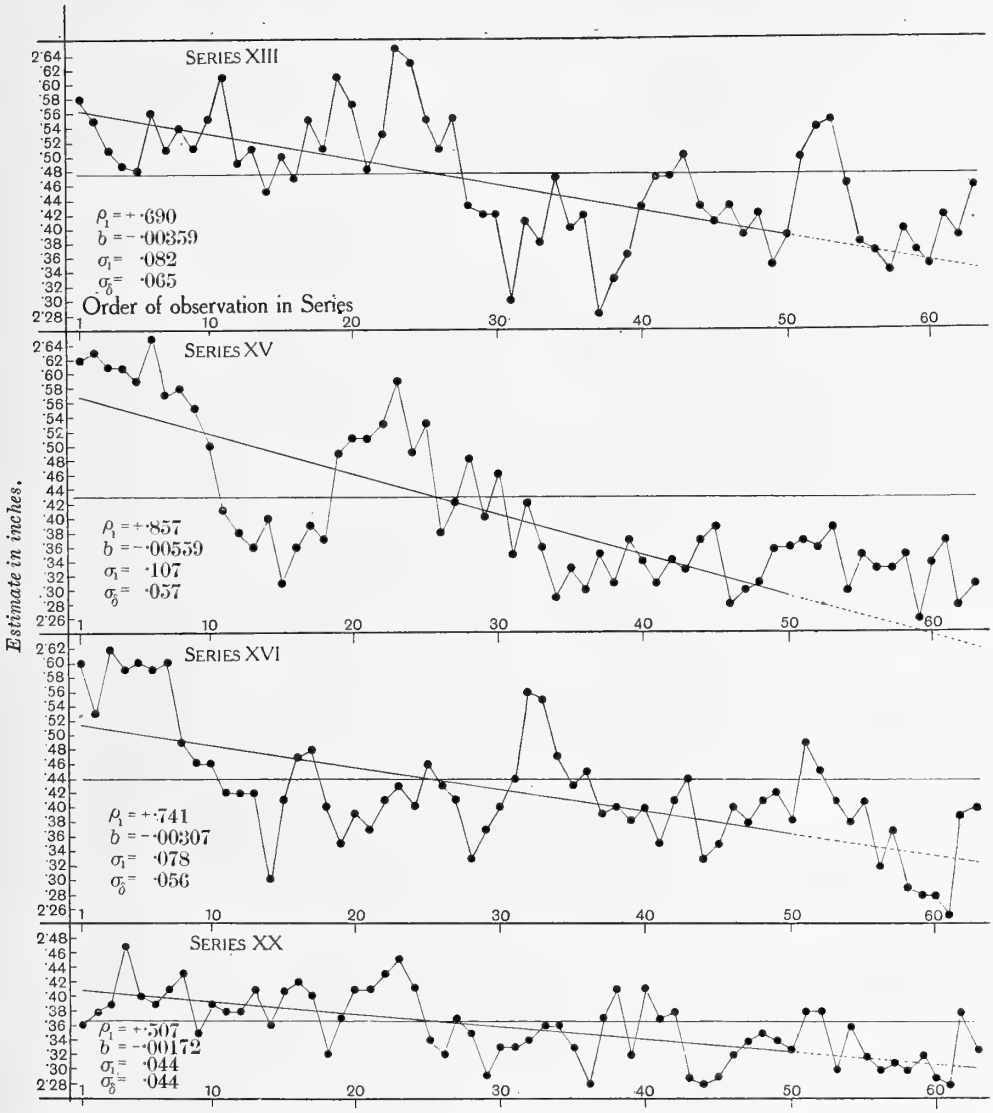
	b	ρ_1'	ρ_1	σ_1	σ_δ
XV	− ·005588	+ ·6810	+ ·8568	·10711	·0573
XVIII	− ·000580	+ ·2870	+ ·3144	·04644	·0544



The horizontal line intersecting each graph gives the mean of the first 50 observations in that series.

Fig. 6. Trisection. Diagrams representing variations in judgment.

* For definitions of these constants the table on page 35 may be referred to.



The horizontal line intersecting each graph gives the mean of the first 50 observations in that series.

Fig. 6. Trisection. Diagrams representing variations in judgment (continued).

XV has a large linear sessional change, but superimposed on this there must be considerable correlated variations, for the removal of the best fitting straight line only reduces ρ_1 (·8568) to ρ_1' (·6810). XVIII has variations altogether on a smaller scale; the correlation of successive judgments is low and it is barely affected by the removal of the linear sessional change. And yet though the σ_1 for XV is more than twice as great as for XVIII, the σ_δ's, or measures of the average jump in estimation from judgment to judgment, are practically identical; the importance of this constant σ_δ as a measure of variability of judgment is discussed on p. 69 below.

TABLE V.

Constants of Combined Series (Trisection).

	$k=1$	2	3	4	5	6	7	8	9	10	11	12	13	14
D_k	2·50													2·50
	− ·00237	− ·00463	− ·00645	− ·00845	− ·01036	− ·01245	− ·01465	− ·01693	− ·01951	− ·02310	− ·02618	− ·02914	− ·03145	− ·03358
S_k	·12892	·12926	·12961	·13003	·13042	·13103	·13146	·13165	·13195	·13270	·13347	·13354	·13338	·13310
	± ·00194	± ·00195	± ·00195	± ·00196	± ·00197	± ·00198	± ·00198	± ·00199	± ·00199	± ·00200	± ·00201	± ·00201	± ·00201	± ·00201
$\mathbf{R}_k$	+ ·8385	·7946	·7361	·7059	·6904	·6694	·6642	·6498	·6333	·6346	·6232	·6215	·6151	
	± ·0063	± ·0079	± ·0098	± ·0107	± ·0112	± ·0118	± ·0119	± ·0123	± ·0128	± ·0127	± ·0130	± ·0131	± ·0133	
S_k'	·08450	·08478	·08499	·08524	·08502	·08533	·08535	·08545	·08576	·08722	·08908	·09010	·09072	·09065
	± ·00127	± ·00128	± ·00128	± ·00129	± ·00128	± ·00129	± ·00129	± ·00129	± ·00129	± ·00132	± ·00134	± ·00136	± ·00137	± ·00137
$\mathbf{R}_k'$ (1)	+ ·6246	·5231	·3878	·3155	·2812	·2324	·2218	·1910	·1647	·1826	·1681	·1719	·1603	
	± ·0130	± ·0155	± ·0181	± ·0192	± ·0196	± ·0202	± ·0203	± ·0206	± ·0208	± ·0206	± ·0207	± ·0207	± ·0208	
$\mathbf{R}_k'$ (2)	+ ·6246	·5226	·3867	·3166	·2807	·2320	·2200	·1866	·1483	·1514	·1250	·1211	·1062	
${}_1R_k$	− ·3641	+ ·0452	− ·0876	− ·0455	+ ·0170	− ·0489	+ ·0285	+ ·0065	− ·0551	+ ·0393	− ·0300	+ ·0146		
${}_2R_k$	− ·6500	+ ·1987	− ·0641	− ·0075	+ ·0471	− ·0525	+ ·0364	+ ·0145	− ·0572	+ ·0600	− ·0418			

TABLE VI.

Coefficients of Correlation of Successive Observations freed from Linear Sessional Change.

	$k=1$	2	3	4	5	6	7	8	9	10	11	12	13
$\mathbf{R}_k''$ from 1st Difference Equations	+ ·4892	+ ·3504	+ ·1654	+ ·0699	+ ·0209	− ·0455	− ·0619	− ·1074	− ·1595	− ·1553	− ·1912	− ·1965	− ·2167
	± ·0162	± ·0187	± ·0207	± ·0212	± ·0213	± ·0213	± ·0212	± ·0211	± ·0208	± ·0208	± ·0206	± ·0205	± ·0203
$\mathbf{R}_k''$ from 2nd Difference Equations	—	—	+ ·1655	+ ·0703	+ ·0218	− ·0437	− ·0589	− ·1026	− ·1522	− ·1447	− ·1765	− ·1770	− ·1917

(*b*) *The Combination of Series.*

Having discussed the reduction of the individual series, I will proceed to consider the results of combining the 20 series. The formulae (v) to (viii) on pp. 33 and 34 give the values of D_k, S_k, and $\mathbf{R}_k$ which are tabled below. Remembering that D_1 and S_1 are the mean and standard deviation of the combined observations $y_1, y_2 \ldots y_{50}$ from each of the 20 series, D_2 and S_2 the mean and standard deviation of the combined observations $y_2, y_3 \ldots y_{51}$, and finally D_{14} and S_{14} of $y_{14}, y_{15} \ldots y_{63}$, we see that the progressive decrease in D_k as k increases indicates the shortening in the estimate of a third during the course of a sitting, while the increase in S_k may perhaps be partly due to increasing variability of judgment due to fatigue. The values of $\mathbf{R}_k$ are large, but this is to be expected owing to the large changes in personal equation from series to series; in fact for $k = 13$ it will be found that the limiting expression L_k of page 34 gives

$$L_{13} = +\cdot 5435, \text{ while } \mathbf{R}_{13} = \cdot 6151.$$

The reason for this difference between L_{13} and $\mathbf{R}_{13}$ is that $\sum_m (\rho_{13}\sigma_1\sigma_{14})$, and therefore $\mathbf{R}_{13}'$, does not vanish. The next step is to obtain the values of S_k' and $\mathbf{R}_k'$, or the standard deviations and correlations of successive judgments after the secular change has been removed. They are found from Equations (ix) and (x) of p. 34 and are given in Table V (5th and 6th rows).

There is here an opportunity of testing the accuracy of the Difference Correlation method discussed in Section V (*b*); the case is that of Problem 1, page 41, the values of $\mathbf{R}_1, \mathbf{R}_2 \ldots \mathbf{R}_{13}$ are known and give the correlation of 1st differences, ${}_1R_1, {}_1R_2 \ldots {}_1R_{12}$; these together with the coefficients of correlation of 2nd differences to be used later, are given at the bottom of Table V. Then using the value + ·6246 for $\mathbf{R}_1'$, we get the values of the 12 quantities $\mathbf{R}_2' \ldots \mathbf{R}_{13}'$, which have been inserted in the 7th row of Table V. It will be seen that the values obtained by this approximate method agree well with the others, the differences being within the probable error of the $\mathbf{R}_k'$'s up to and including $\mathbf{R}_9'$; beyond this the approximate values become rapidly too small, the error, from the form of Equations (xx) and (xxi), being clearly cumulative. This failure is certainly largely due to the fact that the errors involved in the assumptions (*a*), (*b*) and (*c*) of p. 37 are not negligible when the later groups enter into the correlation, for we have already seen that both D_k and S_k change steadily with k.

The values of S_k' and $\mathbf{R}_k'$ in Table V correspond to the average values of the standard deviations and correlations of successive judgments in the individual series, i.e. of the σ's and ρ's given in Tables III and I. Owing to the sessional change which occurs during the course of nearly all the series, $\mathbf{R}_k'$ does not vanish as k increases, but appears to approach a limiting value in the neighbourhood of + ·16. By obtaining for the separate series, the coefficient of correlation, ρ_1', of the successive observations at intervals of one, freed from the *linear* sessional change, a step has been made towards the further reduction of the problem. $\mathbf{R}_1''$, the

coefficient for the combined series corresponding to ρ_1' of the individual series, is given by

$$\mathbf{R}_1'' = \frac{\sum_m (\rho_1' \sigma_1' \sigma_2')}{\sqrt{\left\{\sum_m (\sigma_1'^2)\right\}\left\{\sum_m (\sigma_2'^2)\right\}}} \quad \text{..... (xxvi) bis,}$$

and taking ρ_1', σ_1' and σ_2' calculated from Equations (xxiii), (xxiv) and (xxv) we find that

$$\mathbf{R}_1'' = +\cdot 48922 \pm \cdot 01623.$$

Then $\mathbf{R}_2''$, $\mathbf{R}_3''$... $\mathbf{R}_{13}''$ can be found by the method of Problem 2, p. 41; or again using the value of $\mathbf{R}_2''$ found from the first difference equations, we may proceed to second differences as in Problem 3, and so obtain $\mathbf{R}_3''$... $\mathbf{R}_{12}''$. In this particular case there is no need to use the second difference equations*, but the values of the $\mathbf{R}_k''$'s have been worked out by both methods, as numerical examples of the theory of Sections V (*a*) and (*b*). A comparison of the values given in Table VI shows that there is no significant difference between the results of the two methods†, and the agreement found earlier in this section between the values of $\mathbf{R}_k'$ calculated directly and those found from the difference equations, warrants confidence in the results for $\mathbf{R}_k''$. Although the negative values of $\mathbf{R}_k''$ are probably too large for the higher values of k (just as the later positive values of $\mathbf{R}_k'$ in Table V, row 7, were too small), there is no doubt, I think, that the correlation of the successive observations freed from the linear sessional change, does become negative at $k = 5$, 6 or 7 and remain negative for the higher values of k. A word of qualification is necessary; the linear sessional change to be removed has been represented by the line "best" fitting the *first* 50 *observations* of each series, and a glance at Figure 4 shows that the mean values of the later observations in the series of 63 would lie well off this line because of the parabolic form of the sessional change; the negative values found for $\mathbf{R}_8''$, $\mathbf{R}_9''$, etc. may probably be largely accounted for by this fact. A more satisfactory approximation to the correlation of successive judgments freed from sessional change will be obtained in Section XI below.

As $\sigma_\delta = \sigma_1 \sqrt{2(1-\rho_1)}$, referred to on p. 55 above, gives the standard deviation of the first differences of consecutive judgments in a single series, we shall have as a corresponding measure for the combined twenty series

$$S_\delta = S_1' \sqrt{2(1-\mathbf{R}_1')}.$$

For the Trisection Experiment

$$S_\delta = \cdot 0732.$$

* To get an idea of the order of the terms Q_k and $\overline{b^2}$ which are being neglected, the values were calculated for two values of k, with the result

$$\left.\begin{aligned} k=1,\ Q_k + \overline{b^2} &= -\cdot 000001064 \\ k=9,\ Q_k + \overline{b^2} &= +\cdot 000000192 \end{aligned}\right\}.$$

† The probable errors in the Table have been calculated from the usual formula $\epsilon = \pm \cdot 6745 \frac{1-R^2}{\sqrt{N}}$, and do not cover the errors arising from the method of approximation.

(*c*) *On the possible Effect of shifting the Head during the course of a Series.*

It was suggested to me that the correlation of successive judgments in this and in the Bisection Experiment might be due to periodic shifting of the head from side to side during the course of a series, some parallax effect of the two eyes making corresponding variations in the estimation of a third (or a half) of the line on the form. Now such an explanation might account for part of the correlation in these two experiments, but it could not explain the regular secular and sessional changes in the Trisection, except by the highly improbable hypothesis that the observer's head leant over increasingly to one side during the course of a sitting, and that he started with it more on one side in the later series than in the earlier ones. But beyond this, the fact that correlation is found also in the timing experiments suggests that it is of deeper and more complex origin. It is likely to arise from many unknown causes affecting the environment and condition of the observer, and if one of them is a relative shifting of the eyes, it is of interest, for it will enter into many kinds of observations, where the observer who takes the readings is not looking through a fixed eyepiece.

To test the effect of a relative shift between head and paper, 42 of the forms were taken, and trisected in the usual way, but for alternate groups of seven the paper was shifted 4 inches relatively from side to side. The measures of the estimates and their means are in Table VII. The three sets of seven under the heading I, were made with the forms in one position, the three sets under II with the forms shifted 4 inches to the right. The difference is noticeable at once; readings I are smaller than II, and at the same time the curious effect of sessional change is appearing, the later readings of I and again of II, being on the whole smaller than the earlier ones. Now in carrying out the observations of the Trisection and Bisection Series, the body and head were kept as steady as possible, and it is unlikely that frequent shifts as large as 4 inches could have occurred; further the differences between the means of readings I and II, are much smaller than the actual variations in judgment shown in the diagrams of Figure 6.

But as a further test, a series of 63 forms were marked off, with the head fixed mechanically; the results are given in Table VIII with the usual notation. The correlations are not as high as many of those in Table I, but they are comparable with those of Series I, V, VI, XVIII. The sessional change is also indicated by the decrease in d_k as k increases*. Without carrying through a good number of series with fixed head, no useful comparison can be made, but I think that the evidence of this one series is sufficient to justify the assertion that a shifting of the head from side to side cannot account for the greater part of the correlation of successive judgments.

(*d*) *Summary.*

First considering the individual series, it was noticed that there was a secular change in Personal Equation with time—i.e. the means decreased in passing from

* The value of σ_δ or ·074 may be compared with that of S_δ for the ordinary series of the Experiment *A*, which was ·0732.

TABLE VII.

Experiments on Shift of Head.

Order of Observations	I	Order of Observations	II	Order of Observations	I	Order of Observations	II	Order of Observations	I	Order of Observations	II
1	2·39	8	2·48	15	2·45	22	2·49	29	2·36	36	2·45
2	2·39	9	2·50	16	2·43	23	2·49	30	2·41	37	2·38
3	2·39	10	2·57	17	2·39	24	2·55	31	2·38	38	2·40
4	2·44 Mean	11	2·47 Mean	18	2·46 Mean	25	2·48 Mean	32	2·42 Mean	39	2·39 Mean
5	2·45 2·416	12	2·44 2·500	19	2·40 2·414	26	2·47 2·483	33	2·37 2·377	40	2·46 2·410
6	2·44	13	2·53	20	2·39	27	2·44	34	2·35	41	2·37
7	2·41	14	2·51	21	2·38	28	2·46	35	2·35	42	2·42

TABLE VIII.

	$k=1$	2	3	4	5	6	7	8	9	10	11	12	13	14
d_k	2·5194	2·5188	2·5160	2·5144	2·5136	2·5096	2·5050	2·5048	2·5040	2·5052	2·5064	2·5036	2·5022	2·5008
σ_k	·07004	·06965	·07034	·06955	·06948	·07071	·06546	·06558	·06588	·06598	·06678	·06904	·06652	·06602
ρ_k	+ ·4438	+ ·4084	+ ·2856	+ ·2121	+ ·0719	+ ·0107	+ ·0298	− ·0017	+ ·1521	+ ·1287	+ ·0430	+ ·1097	+ 0619	—

$\sigma_\delta = \sigma_1 \sqrt{2\,(1-\rho_1)} = ·074.$

the earlier to the later series; in addition there was a remarkably constant sessional change within each series, this change being again a decrease from the earlier to the later observations. There was something in these changes almost analogous to an elastic strain, during the course of a series the estimation of a third drops, in the interval between the series there is a recovery, but not a complete recovery, for the first judgments in the succeeding series start at a little lower level than the first, but well above the last judgments in the series before; this slight "permanent deformation" caused by the "strain" represented in the sessional change, results in the secular fall. The figure below gives an ideal representation of this.

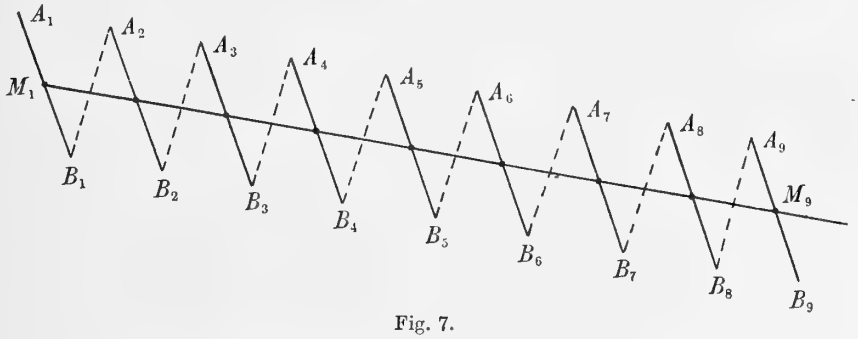


Fig. 7.

A_1B_1, A_2B_2, ... sessional change in Series 1, 2 ... etc.
B_1A_2, B_2A_3, ... "recovery" during interval between Series 1 and 2, 2 and 3 etc.
M_1M_9 the resulting secular change.

Then combining the twenty series, in order to get more reliable results, the coefficients of correlation of successive judgments, $\mathbf{R}_k$, were obtained; owing to the secular and sessional changes these coefficients had very high values and as k increased, apparently tended to a limit at about $+\cdot60$. By fitting the means of the series together, the secular change was eliminated, and a series of coefficients $\mathbf{R}_k'$ obtained, which represented the average value of the correlation in a series; owing to the sessional change the $\mathbf{R}_k''$s did not appear to tend to zero as k increased but to a limit at $+\cdot16$ or $+\cdot15$. The correlation of successive values of the residuals, left after subtracting the ordinates of the straight line "best" fitting the first 50 observations of each series from the observations of that series, gave a set of coefficients $\mathbf{R}_k''$, which fell off very rapidly and became negative when k equalled 6 or 7; the large negative values of the coefficients for the high values of k were probably due in part to the method of approximation used, and also to the fact that the straight line fitting the first 50 observations in a series did not represent satisfactorily the sessional change.

The values of $\mathbf{R}_k'$ calculated (up to $k = 13$), gave no evidence of any tendency to periodicity in this coefficient, although there was evidence of this occurring in some of the individual series; periodicity in $\mathbf{R}_k'$ would indicate marked variations of roughly the same period occurring at any rate in a large number of the series.

It will be shown in a later section that the values of $\mathbf{R}_k'$, $k = 1 \ldots 13$, can be fitted very closely by a curve of the type $y = p + qr^k$, where p, q, and r are constants.

Finally it was shown that the correlation of successive judgments could not be due to a shifting of the head during the course of observation, although this might perhaps be one of many contributory causes.

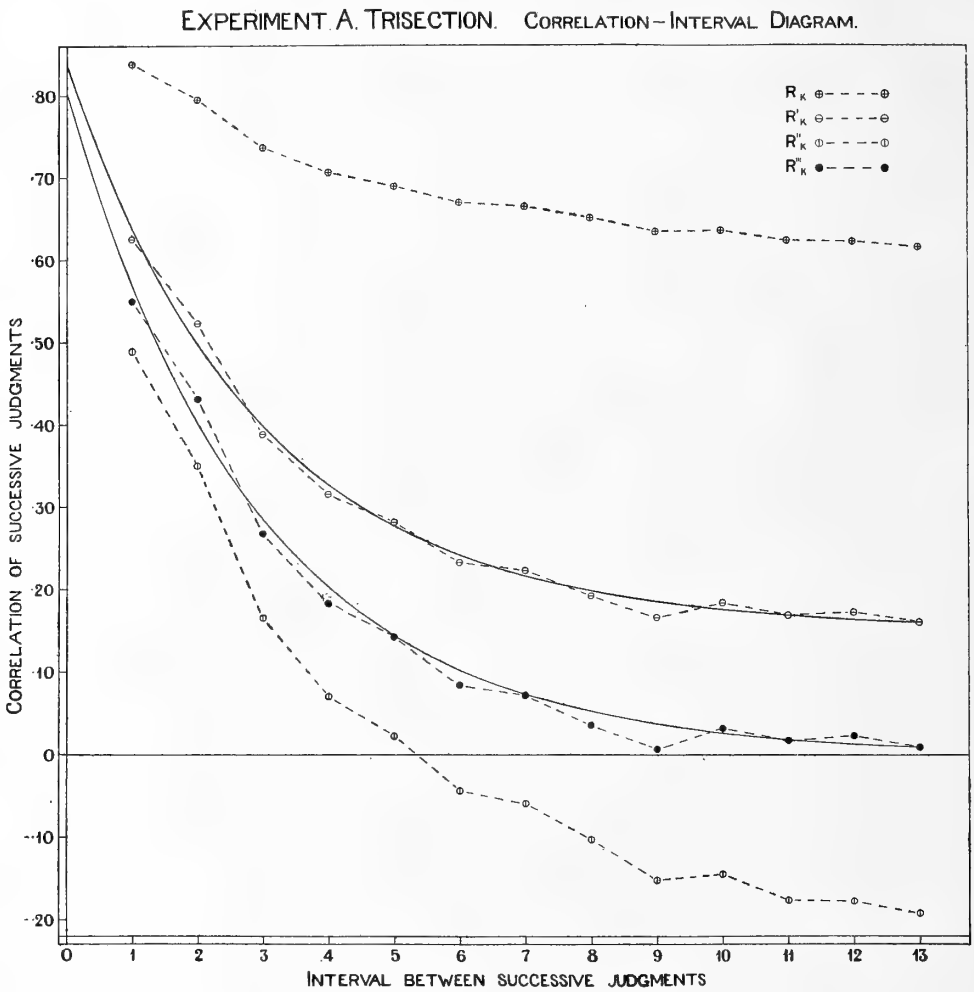


Fig. 8.

In Figure 8 the values of $\mathbf{R}_k$, $\mathbf{R}_k'$, and $\mathbf{R}_k''$ (for linear sessional change) are plotted to k; the theoretical curves of the Equations (xlix) and (lvi) shown in the Figure will be discussed in Section XI, and also the points referred to as R_k'''.

VII. EXPERIMENT *B*, BISECTION. REDUCTION OF OBSERVATIONS.

(*a*) *The individual Series.*

In this Experiment the coefficients of correlation of successive judgments for the individual series were not all worked out, but only the values of ρ_1; these are tabled with σ_1 and d_1 in Table IX. The values of d_k for each series, $k = 1 \ldots 14$ are also given in Table X. It will be seen that there are not the same marked secular or sessional changes as characterised the Trisection Series. In Figure 9 the means of Groups 1 of each Series—or the d_1's—have been plotted to "order" and to "time," and again if

x is the personal equation or mean,

y the number of order of series,

z the number of days between 13th June and the date on which the series was carried out,

we have for the regression lines

$$x - 2{\cdot}8793 = -{\cdot}0010131\,(y - 10{\cdot}5) \quad \ldots\ldots\ldots\ldots\ldots\ldots(\text{xxxiv}),$$

$$x - 2{\cdot}8793 = -{\cdot}0005359\,(z - 32{\cdot}10) \quad \ldots\ldots\ldots\ldots\ldots\ldots(\text{xxxv}).$$

These lines have been drawn in the diagrams; the coefficients of correlation are

$$r_{xz} = -{\cdot}337 \pm {\cdot}134, \qquad r_{xy} = -{\cdot}156 \pm {\cdot}147, \qquad r_{yz} = +{\cdot}945 \pm {\cdot}016,$$

giving partial correlations

$$r_{xy.z} = +{\cdot}529 \pm {\cdot}109, \qquad r_{xz.y} = -{\cdot}589 \pm {\cdot}099.$$

TABLE IX. *Constants of Bisection Experiments.*

Series	d_1	σ_1	ρ_1	$\sigma_\delta = \sigma_1\sqrt{2(1-\rho_1)}$	Dates (1920) and time at start	Time taken for series
I	2·8648	·04997	+·4942	·0503	11 a.m. / 2.45 p.m. } 13 June	6m 0s
II	·8624	·05461	+·2609	·0664		5 20
III	·9262	·03821	+·0823	·0518	p.m. 15 „	5 45
IV	·8642	·04690	+·4107	·0509	10 a.m. / 3 p.m. } 29 „	5 30
V	·8290	·05158	+·5870	·0469		5 45
VI	·9114	·04609	+·5768	·0424	a.m. 30 „	5 45
VII	·9178	·04415	+·2993	·0523	10 a.m. / 12.15 p.m. } 1 July	5 45
VIII	·9218	·04766	+·4360	·0506		6 0
IX	·8724	·04384	+·1389	·0575	10.30 a.m. / 6.30 p.m. } 2 „	5 30
X	·8990	·04579	+·1018	·0614		6 15
XI	·9238	·03617	−·0423	·0522	9.30 a.m. / 6.45 p.m. } 6 „	6 30
XII	·9298	·04810	+·5089	·0477		6 45
XIII	·8806	·04407	+·2769	·0530	a.m. 7 „	5 45
XIV	·8312	·04955	+·4445	·0522	11 a.m. / 2.30 p.m. } 1 August	6 45
XV	·8242	·03606	+·3334	·0416		6 15
XVI	·7976	·04135	+·3190	·0483	p.m 16 „	Not recorded
XVII	·8566	·03739	+·5531	·0353	a.m. 17 „	
XVIII	·8808	·03497	+·2776	·0420	p.m. 18 „	
XIX	·8890	·03986	+·5407	·0382	p.m. 19 „	
XX	2·9030	·02610	+·1404	·0342	p.m. 20 „	

Probable Errors of Coefficients of Correlation calculated from 50 pairs of the variates.

ρ	P. E.
·80	± ·0343
·70	± ·0486
·60	± ·0610
·50	± ·0715
·40	± ·0801
·30	± ·0868
·20	± ·0916
·10	± ·0944
·00	± ·0945

Mean time taken for a series of 70 observations (including the 7 preliminary trials*) 5m 58s

Mean interval between records of judgment 5·11s

* See p. 28, footnote.

TABLE X. Bisection.

Table of Means of Groups (in inches).

Series	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11	Group 12	Group 13	Group 14
I	2·8648	2·8646	2·8656	2·8640	2·8652	2·8652	2·8644	2·8648	2·8650	2·8648	2·8654	2·8658	2·8674	2·8668
II	·8624	·8656	·8676	·8692	·8704	·8710	·8736	·8756	·8774	·8782	·8788	·8772	·8774	·8770
III	·9262	·9260	·9256	·9238	·9228	·9220	·9232	·9234	·9236	·9242	·9244	·9234	·9224	·9226
IV	·8642	·8666	·8684	·8684	·8704	·8718	·8754	·8786	·8824	·8834	·8856	·8868	·8886	·8890
V	·8290	·8304	·8310	·8320	·8334	·8344	·8344	·8348	·8342	·8362	·8376	·8378	·8380	·8376
VI	·9114	·9136	·9154	·9172	·9196	·9212	·9214	·9210	·9218	·9196	·9202	·9218	·9212	·9206
VII	·9178	·9192	·9196	·9200	·9196	·9228	·9228	·9230	·9236	·9214	·9210	·9218	·9194	·9180
VIII	·9218	·9230	·9256	·9260	·9264	·9260	·9220	·9206	·9206	·9188	·9170	·9172	·9154	·9132
IX	·8724	·8736	·8704	·8688	·8686	·8690	·8682	·8680	·8688	·8676	·8676	·8660	·8628	·8610
X	·8990	·8986	·9004	·9030	·9028	·9034	·9024	·9024	·9010	·9018	·9016	·9028	·9024	·9002
XI	·9238	·9234	·9264	·9290	·9286	·9278	·9264	·9262	·9252	·9252	·9286	·9290	·9302	·9298
XII	·9298	·9300	·9318	·9316	·9374	·9398	·9416	·9426	·9412	·9400	·9398	·9392	·9404	·9414
XIII	·8806	·8804	·8816	·8798	·8796	·8808	·8818	·8840	·8876	·8876	·8902	·8916	·8950	·8964
XIV	·8312	·8322	·8340	·8350	·8362	·8392	·8394	·8404	·8406	·8406	·8416	·8426	·8458	·8478
XV	·8242	·8254	·8260	·8254	·8266	·8286	·8294	·8304	·8306	·8324	·8342	·8364	·8386	·8408
XVI	·7976	·7978	·7974	·7970	·7966	·7976	·7996	·8006	·8020	·8044	·8060	·8068	·8086	·8094
XVII	·8566	·8548	·8522	·8512	·8514	·8518	·8514	·8514	·8518	·8520	·8536	·8544	·8528	·8530
XVIII	·8808	·8790	·8784	·8770	·8740	·8720	·8714	·8694	·8682	·8676	·8686	·8674	·8664	·8664
XIX	·8890	·8886	·8864	·8828	·8816	·8802	·8796	·8778	·8778	·8778	·8786	·8788	·8786	·8784
XX	2·9030	·9024	·9018	·9022	·9014	·9026	·9028	·9030	·9036	·9050	·9034	·9032	·9048	2·9068
Means	2·87928	2·87976	2·88028	2·88017	2·88063	2·88136	2·88156	2·88190	2·88235	2·88243	2·88319	2·88350	2·88381	2·88381

The dates on which the series were carried out—the z's—are given at the end of Table IX; the distribution was more satisfactory than that of the Trisections, and the significance of these two partial correlations will be referred to shortly.

The variation in the means of the series is much smaller than in the case of the Trisections; we have here a range from 2·93 to 2·80 ins. while in the other, from 2·70 to 2·34 ins.; in both cases the secular change is in the direction which lessens the measures, i.e. the marks on the forms in the later series were on the whole further to the observer's left hand than in the earlier series. Nor does experience appear to increase accuracy, for the true position of the half is at 2·97 inches (and of the third at 2·51 inches).

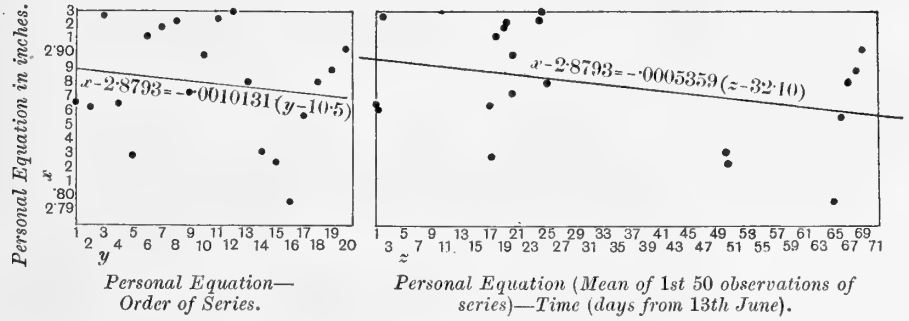


Fig. 9. Bisection. Means of Groups 1 of each series plotted with Order of Series and Date of Series.

Next considering the sessional change, the values of $\bar{y}_t$ (defined on p. 47) have been plotted in Figure 10; the straight line "best" fitting these points is

$$\bar{y}_t - 2{\cdot}8816 = +\,{\cdot}0003534\,(t-32) \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xli)},$$

where t is the order of observation in a series, and the coefficient of correlation between $\bar{y}_t$ and t is $+\,{\cdot}5294 \pm {\cdot}0137$*.

Using the relations of page 48, it is found that

$$\eta_{\bar{y}_t} = {\cdot}271 \pm {\cdot}018, \quad \sqrt{1-\eta^2_{\bar{y}_t}} = {\cdot}963,$$

and on comparing this latter value with that for the Trisections (·815) we see that in the present case the mean sessional change is of less significance.

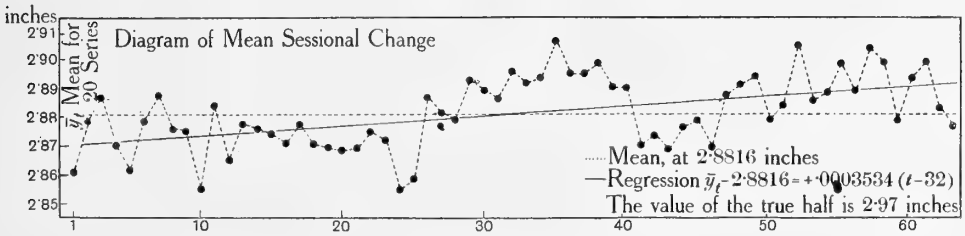
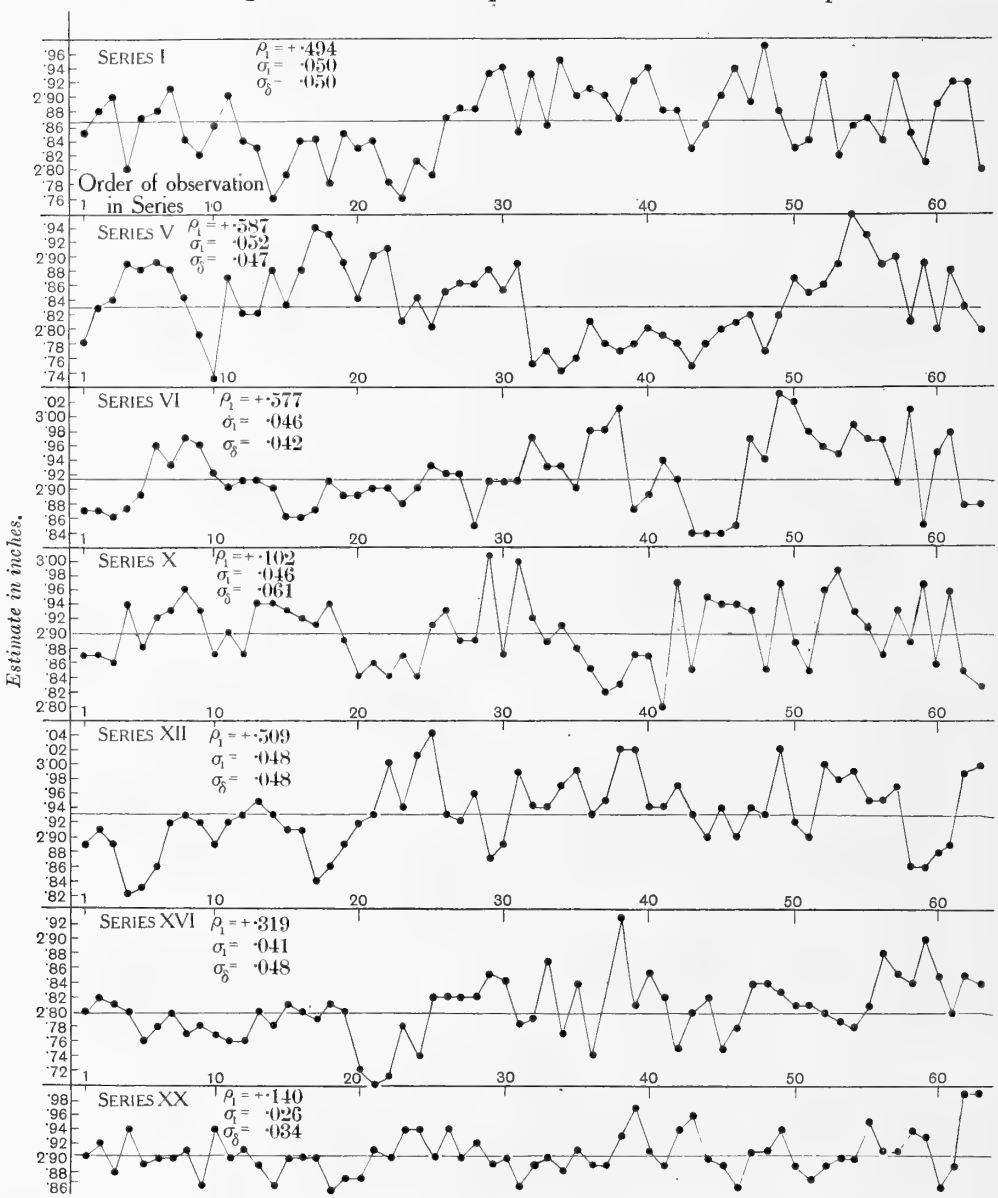


Fig. 10. Bisection. t, Order of Observation in Series.

It will be noticed in looking at Figure 10 that the points $(t, \bar{y}_t)$ appear to be subject to a fairly consistent periodic variation about the regression line, the

* This correlation between the mean tth observation ($\bar{y}_t$) and t must be distinguished from the correlation between the tth observation (y_t) and t, which is $+\,{\cdot}143$, and as it should be, less than η.

complete period covering from 20 to 22 observations. Without a detailed analysis of the separate series, it is not possible to say whether there is a period of this order underlying the variations in judgment in all series, or whether this periodicity in $\bar{y}_t$ results from large variations in one or two series; the diagrams of seven of the series, in Figure 11 do not certainly suggest any marked periodic variation, and it is possible that the drop at about the 44th and the peak near the



The horizontal line intersecting each graph gives the mean of the first 50 observations in that series.

Fig. 11. Bisections. Diagrams representing variations in judgment.

55th observations in Series V and VI, would go far to account for the similar features in the $\bar{y}_t$ diagram, the "y" scale of which is four times greater than that in Figure 11.

Using the method of Correlation of Ranks*, the correlation between σ_1 and ρ_1 has been calculated for the 20 Series; the result is

$$r_{\sigma_1\rho_1} = + \cdot 420 \pm \cdot 124.$$

Another coefficient which may be calculated, is that of the correlation between σ_δ, or the standard deviation of first differences of consecutive judgments, and ρ_1; using the same method as for $r_{\sigma_1\rho_1}$, it is found that

$$r_{\sigma_\delta\rho_1} = - \cdot 416 \pm \cdot 125 \text{ and again } r_{\sigma_\delta\sigma_1} = + \cdot 465 \pm \cdot 118.$$

Now ρ_1, σ_1 and σ_δ are not three independent quantities, as they are connected by the relation

$$\sigma_\delta = \sqrt{\sum_{t=1}^{n} \frac{(y_{t+1} - y_t)^2}{n}} = \sigma_1 \sqrt{2(1-\rho_1)},$$

and it is open to question, which two are the most fundamental. In the ordinary theory of the Combination of Observations, where it is assumed that ρ_1 is zero, it is natural to consider σ_1 (or σ) as a fundamental constant, the measure of the accuracy of judgment; σ_δ appears to have no special significance and merely equals $\sqrt{2}\sigma$. If however there is a correlation of successive judgments, σ loses its importance; if we take a small number, p, of successive observations and calculate their standard deviation, s_p, we can no longer say that s_p, subject to its probable error $\pm \cdot 6745 \frac{s_p}{\sqrt{2p}}$, will be equal to σ, the standard deviation of a long series of judgments. On the other hand there is every reason to expect that the σ_δ found from a few observations will give a fair approximation to the σ_δ found from a large number. σ is dependent to a high degree on the sessional change; for example it has been shown† that if this change can be represented by a straight line of the form $y = bt$, then σ', or the standard deviation of the observations freed from this change is given by

$$\sigma'^2 = \sigma^2 - \frac{b^2}{12}(n^2 - 1).$$

It is true that σ_δ is dependent to some extent on the sessional change, but far less so; for instance in the case of the linear sessional change, σ_δ', the standard deviation of the first differences of the successive residuals left after the removal of the line, is given approximately by the relation

$$\sigma_\delta'^2 = \sigma_\delta^2 - b^2.$$

And for any form of sessional change which is likely to occur in experiments of the type we are considering, the correction to the difference between two successive observations necessary to get the corresponding difference between the

* p. 52 and footnote.

† Section V (*b*) p. 43.

residuals after the removal of the sessional term, will be very small indeed compared with the standard deviation of this difference, or σ_δ. It is therefore suggested that in the combination of correlated observations, σ_δ, the average value of the jump in estimation between two successive judgments, is of more fundamental importance than σ. As an example, consider the diagrams of the observations of Series X and Series XX in Figure 11; the correlation, ρ_1, is very low in both cases, but it is suggested that the physiological significance of the difference in type between the two, lies in the fact that σ_δ for Series X is nearly twice as large as σ_δ for Series XX, rather than in the difference in the σ_1's. Or again in the diagrams of the Trisection Experiment, Figure 6, I would emphasise the same point in a comparison of the difference between the two highly correlated Series VIII and XVI.

Now returning to the coefficients of partial correlation

$$r_{xy.z} = +\cdot529 \pm \cdot109, \quad r_{xz.y} = -\cdot589 \pm \cdot099.$$

With the interpretation suggested on p. 54 for these coefficients, we are led to a rather suggestive conclusion. If we are dealing with a number of series carried out at equal intervals of time in the course of one, or even perhaps two days, but effectively at one epoch when comparison is made with the long range of nearly 70 days covered by the Bisection Series, then the correlation between x and y is positive, or the pencil mark in the later series tends to be made further to the observer's right than in the earlier series; this change is in the same direction as the sessional change within a series. There is indeed a curious coincidence, on which of course no stress must be laid,

$$r_{xy.z} = +\cdot529 \pm \cdot109, \quad r_{\bar{y}t.t} = +\cdot5294 \pm \cdot0137.$$

That is to say the correlation between the mean of a series and the order of that series when a number of series are done in close succession, is of the same sign and magnitude as the correlation between the mean tth observation and its order, t, in the series. But if we are dealing with all the pth series of sets which have been carried out on different days with varying and perhaps many days' interval between, then the coefficient $r_{xz.y}$ is negative, or the bisection-marks on the later days have on the whole a tendency to move to the left of the observer; this is in the direction of the secular change.

The conclusion which it seems possible to draw is this; if a number of series are done at very short intervals, the interval of rest between the series will not be sufficient to break the effect of the sessional change; but if a considerable interval elapses between the carrying out of the series, then the sessional change in one series has no influence on the judgments in the succeeding series, but a quite distinct secular change may be noticeable. In the Bisection Experiment both secular and sessional changes are very small, but they are acting in opposite directions. If these two changes are due to different physiological factors, it seems possible that it is the fact that they are acting in opposite directions in the Bisection Experiment which causes them to be of so much smaller magnitude than in the Trisection Experiment, where they were acting in the same direction.

(*b*) *The Combination of the Series.*

For the combined series, the coefficients of correlation of successive judgments $\mathbf{R}_k$ for $k = 1, 2 \ldots 13$ were calculated from 13 correlation tables each based on the 1000 combined observations; the results for D_k, S_k and $\mathbf{R}_k$ are tabled below (Table XI). The effect of the slight sessional change is noticeable in the increasing values of D_k.

Using the values of D_k, S_k and $\mathbf{R}_k$ and of ${}_pd_k$ from Table X, Equations (vi), (vii) and (viii) give $\underset{m}{\Sigma}(\rho_k\sigma_1\sigma_{k+1})$ and $\underset{m}{\Sigma}(\sigma_k{}^2)$ for $k = 1, 2 \ldots 14$. Equations (ix) and (x) then give the values of S_k' and $\mathbf{R}_k'$ contained in the 5th and 6th rows of Table XI. The value of $\mathbf{R}_1'$ found by this method should be compared with that found with the help of the ρ_1's, σ_1's and σ_2's of the individual series, namely

$$\mathbf{R}_1' = \frac{\underset{m}{\Sigma}(\rho_1\sigma_1\sigma_2)}{\sqrt{\underset{m}{\Sigma}(\sigma_1{}^2)\underset{m}{\Sigma}(\sigma_2{}^2)}} = +\cdot3578 \pm \cdot0186 \ldots\ldots\ldots\ldots\ldots\text{(x) bis.}$$

The difference which is well within the probable error arises from the fact that $\mathbf{R}_1$ has been found by grouping the observations in a correlation table, while the ρ_1's, σ_1's and σ_2's were found by direct multiplication of the crude values of the observations.

Another method of obtaining the $\mathbf{R}_k''$s is from the first difference correlation equations, or the method of Problem 1, p. 41; the results are given in the 7th row of Table XI, while the constants ${}_1R_k$, the coefficients of correlation of successive first differences required in the solution, are in the 8th row of the Table. Comparing the values of $\mathbf{R}_k'$ found by the two methods, we find good agreement up to $k = 6$, but beyond this point the $\mathbf{R}_k''$s of the second and approximative method assume much too large negative values*. It is however evident from the results of the first method that $\mathbf{R}_k'$ does become negative, and as it could not remain negative indefinitely as k increased, there seems here to be another indication that a periodic variation exists among the judgments at any rate in a certain number of the series. For a complete period covering from 20 to 22 observations suggested by the $\bar{y}_t$ diagram, $\mathbf{R}_k'$ should have a minimum value at $\mathbf{R}_{10}'$ or $\mathbf{R}_{11}'$; the figures suggest that the minimum occurs somewhat earlier, at about $\mathbf{R}_8'$, but the probable errors for these small coefficients are very large. When time is available it would be interesting to examine further the significance of this periodicity.

The points $(\mathbf{R}_k, k)$ and $(\mathbf{R}_k', k)$ have been plotted in Figure 12.

It will be noticed that the S_k''s in the later groups are larger than in the earlier, this suggesting again as in the case of the Trisections, that the observations become slightly more erratic towards the end of a series.

* This result tends to confirm the suggestion made on p. 60 that the difference correlation method gave too large negative values for $\mathbf{R}_k''$ in the Trisection Experiment.

TABLE XI.

Constants of Combined Series (Bisection).

l		$k=1$	2	3	4	5	6	7	8	9	10	11	12	13	14
2	D_k	2·87936	2·87982	2·88038	2·88034	2·88068	2·88136	2·88154	2·88190	2·88236	2·88236	2·88314	2·88338	2·88366	2·88368
3	S_k	·05727 ± ·00086	·05706 ± ·00086	·05748 ± ·00087	·05770 ± ·00087	·05783 ± ·00087	·05770 ± ·00087	·05789 ± ·00087	·05814 ± ·00088	·05840 ± ·00088	·05792 ± ·00087	·05773 ± ·00087	·05794 ± ·00087	·05813 ± ·00088	·05850 ± ·00088
4	$\mathbf{R}_k$	+ ·6299 ± ·0129	·5599 ± ·0146	·5322 ± ·0153	·4755 ± ·0165	·4470 ± ·0171	·4243 ± ·0175	·3905 ± ·0181	·3727 ± ·0184	·3756 ± ·0183	·3817 ± ·0182	·3865 ± ·0181	·3621 ± ·0185	·3518 ± ·0187	
5	S_k'	·04364 ± ·00066	·04348 ± ·00066	·04360 ± ·00066	·04355 ± ·00066	·04373 ± ·00066	·04375 ± ·00066	·04437 ± ·00067	·04490 ± ·00068	·04541 ± ·00068	·04548 ± ·00069	·04566 ± ·00069	·04607 ± ·00069	·04658 ± ·00070	·04727 ± ·00071
6	$\mathbf{R}_k'$ 1st method	+ ·3519 ± ·0187	+ ·2306 ± ·0202	+ ·1828 ± ·0206	+ ·0849 ± ·0212	+ ·0389 ± ·0213	+ ·0119 ± ·0213	− ·0344 ± ·0213	− ·0542 ± ·0213	− ·0383 ± ·0213	− ·0204 ± ·0213	− ·0051 ± ·0213	− ·0302 ± ·0213	− ·0339 ± ·0213	
7	$\mathbf{R}_k'$ 2nd method	+ ·3578 ± ·0186	+ ·2364	+ ·1883	+ ·0900	+ ·0404	+ ·0009	− ·0577	− ·0886	− ·0835	− ·0730	− ·0647	− ·1070	− ·1250	
8	${}_1R_k$	− ·4055	− ·0571	+ ·0390	− ·0379	− ·0079	+ ·0149	− ·0215	− ·0280	− ·0042	+ ·0017	+ ·0394	− ·0189		

$$S_\delta = S_1' \sqrt{2\,(1-\mathbf{R}_1')} = \cdot 04946.$$

(*c*) *Comparison with Experiment A.*

The difference between the results of the two experiments is probably due to the fact that the estimation of a half is so much easier than the estimation of a third. The variations in the latter observations are all on a larger scale than in the former; the secular and sessional changes are very much greater, and if we compare the values of the fundamental constants, we find:

	S_1'	S_δ	$\mathbf{R}_1'$
Trisection	·0845	·0732	+ ·6246 ± ·0130
Bisection	·0436	·0495	+ ·3519 ± ·0187

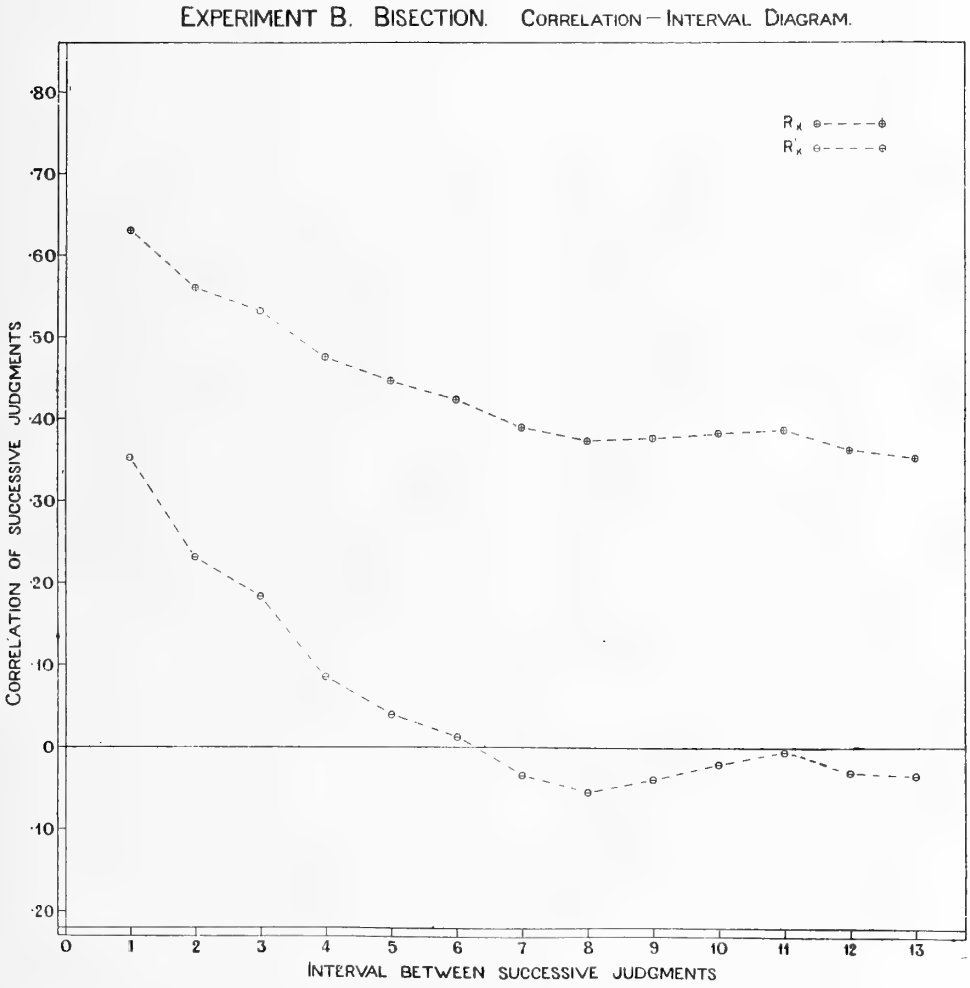


Fig. 12.

and even after the removal of the greater part of the sessional change—(the best fitting straight lines)—the coefficient $\mathbf{R}_1''$ for the Trisections is + ·4892, or greater than $\mathbf{R}_1'$ for the Bisections. The ratio of the values of S_δ, or roughly 3 to 2, is a

measure of the relative uncertainty of the observer in making his estimate in the two different Experiments.

There is some evidence for a slight periodicity in the judgments in the Bisection Series; if there is any period in the Trisections it must cover at least 26 observations, for there is no indication of a significant increase in the values of $\mathbf{R}_k$ as far as calculated, i.e. up to $\mathbf{R}_{13}$.

VIII. Experiment *C*. Counting of 10 Seconds. Reduction of Observations.

(*a*) *The Individual Series.*

The values of d_1, σ_1 and ρ_1 for each of the 20 series are given in Table XII as well as the hour and date; the means (d_1) have been plotted to the order of series in Figure 13.

If x is the mean in the factor e/p for a series,

y the order of series,

z the time in hours and fractions of an hour between 2.0 p.m. on December 13, and the commencement of series

we have for the regression lines,

$$x - \cdot 9186 = - \cdot 006056\,(y - 10\cdot 5) \quad \text{.....(xxxvi),}$$

$$x - \cdot 9186 = - \cdot 001552\,(z - 38\cdot 24) \quad \text{.....(xxxvii),}$$

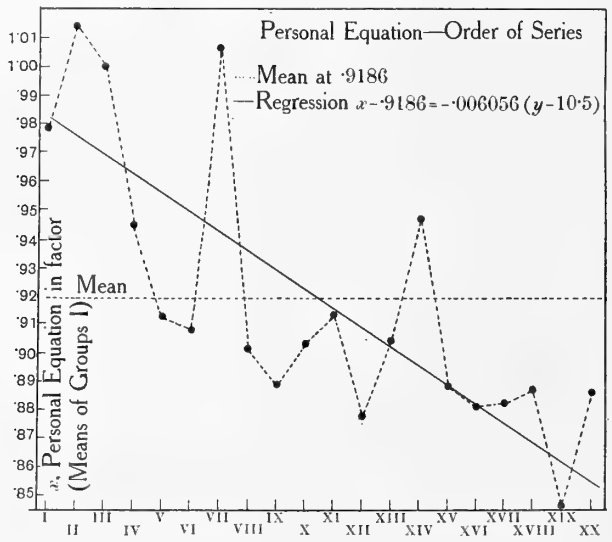


Fig. 13. 10 Second Counting. y, Order or Number of Series.

of which the first is represented in Figure 13.

The coefficients of correlation are,

$$r_{xy} = - \cdot 754 \pm \cdot 065, \quad r_{xz} = - \cdot 775 \pm \cdot 060, \quad r_{yz} = + \cdot 977 \pm \cdot 007,$$

giving partial correlation coefficients

$$r_{xy.z} = +\cdot022 \pm \cdot151, \quad r_{xz.y} = -\cdot271 \pm \cdot140.$$

The secular change corresponds to a gradual decrease in estimate throughout the course of the experiment; the value of the factor e/p for a true 10 second estimate would be $\frac{10\cdot0}{10\cdot2} = \cdot98$, and this was closely approached by the means of the first three series, which were carried out on the first day, shortly after trial counts had been made with a watch. No further check with a watch was made during the succeeding days, and the length of estimation decreased and finally appears to have reached a fairly steady value at about ·88. The mean for the 20 Series was ·9186, or a count of 9·37 seconds.

TABLE XII.

Constants of Individual Series (Counting Seconds).

Series	d_1	σ_1	ρ_1	$\sigma_\delta = \sigma_1\sqrt{2(1-\rho_1)}$	Date (1920)	Time at Start
I	·9786	·04030	+·5283	·0391	13 December	2.30 p.m.
II	1·0140	·04331	+·4988	·0434	„	3.15 p.m.
III	·9998	·03844	+·0625	·0526	„	3.45 p.m.
IV	·9446	·03732	+·4027	·0408	14 December	10.15 a.m.
V	·9128	·03394	+·4378	·0360	„	11.20 a.m.
VI	·9090	·03015	+·5437	·0288	„	12.0 noon
VII	1·0070	·03981	+·3819	·0443	„	2.30 p.m.
VIII	·9012	·02488	+·4550	·0260	„	3.5 p.m.
IX	·8886	·03934	+·4326	·0419	„	3.35 p.m.
X	·9030	·02851	+·5439	·0272	15 December	10.0 a.m.
XI	·9130	·02982	+·5326	·0288	„	10.35 a.m.
XII	·8774	·01852	+·2850	·0221	„	11.10 a.m.
XIII	·9046	·02402	+·4894	·0243	„	11.50 a.m.
XIV	·9464	·02903	+·5085	·0288	„	2.30 p.m.
XV	·8880	·04162	+·7589	·0289	„	3.5 p.m.
XVI	·8812	·04947	+·8549	·0266	16 December	10.0 a.m.
XVII	·8828	·03945	+·6566	·0327	„	10.30 a.m.
XVIII	·8872	·02750	+·5406	·0264	„	11.5 a.m.
XIX	·8468	·02486	+·1266	·0329	„	11.35 a.m.
XX	·8864	·03345	+·6369	·0285	„	12.10 p.m.

Probable Errors of Coefficients of Correlation calculated from 50 pairs of the variates.

ρ	P.E.
·80	± ·0343
·70	± ·0486
·60	± ·0610
·50	± ·0715
·40	± ·0801
·30	± ·0868
·20	± ·0916
·10	± ·0944
·00	± ·0954

With the interpretation of p. 54, the insignificant value of the coefficient $r_{xy.z}$, suggests that for a number of series done in quick succession, there will be no change in personal equation; we shall therefore not expect to find any large general sessional change in the series. The diagram of mean sessional change is given in Figure 14, where $\overline{y}_t$ is plotted to t.

The equation of the straight line best fitting the points is

$$\overline{y}_t - \cdot919 = +\cdot0000731\,(t-32) \quad \text{.....(xxxviii),}$$

and has been drawn in Figure 14.

Using the relations and interpretation of page 48, it is found that

$$\eta_{y_t} = \cdot212 \pm \cdot018 \text{ and } \sqrt{1-\eta_{y_t}^2} = \cdot977,$$

so that the mean sessional change is of even less significance than for the Bisections. In fact it is clear from the diagram that the regression line (xxxviii) very nearly coincides with the line of mean judgment, $y = \cdot 919$.

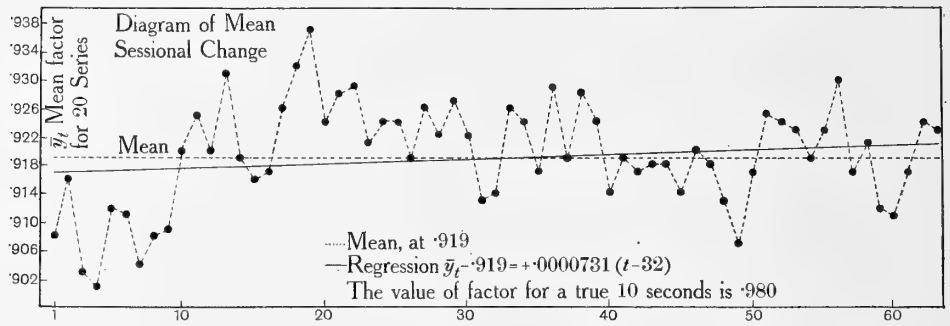


Fig. 14. 10 Second Counting. t, Order of Observation in Series.

The σ_δ's have been found for all the individual series, and using the values of S_1' and $\mathbf{R}_1'$ given below, we have for the combined series

$$S_\delta = S_1' \sqrt{2\,(1 - \mathbf{R}_1')} = \cdot 0338.$$

The method of correlation of Ranks gives

$$r_{\sigma_1 \rho_1} = + \cdot 329 \pm \cdot 135,$$

showing again that large variation is associated with high correlation.

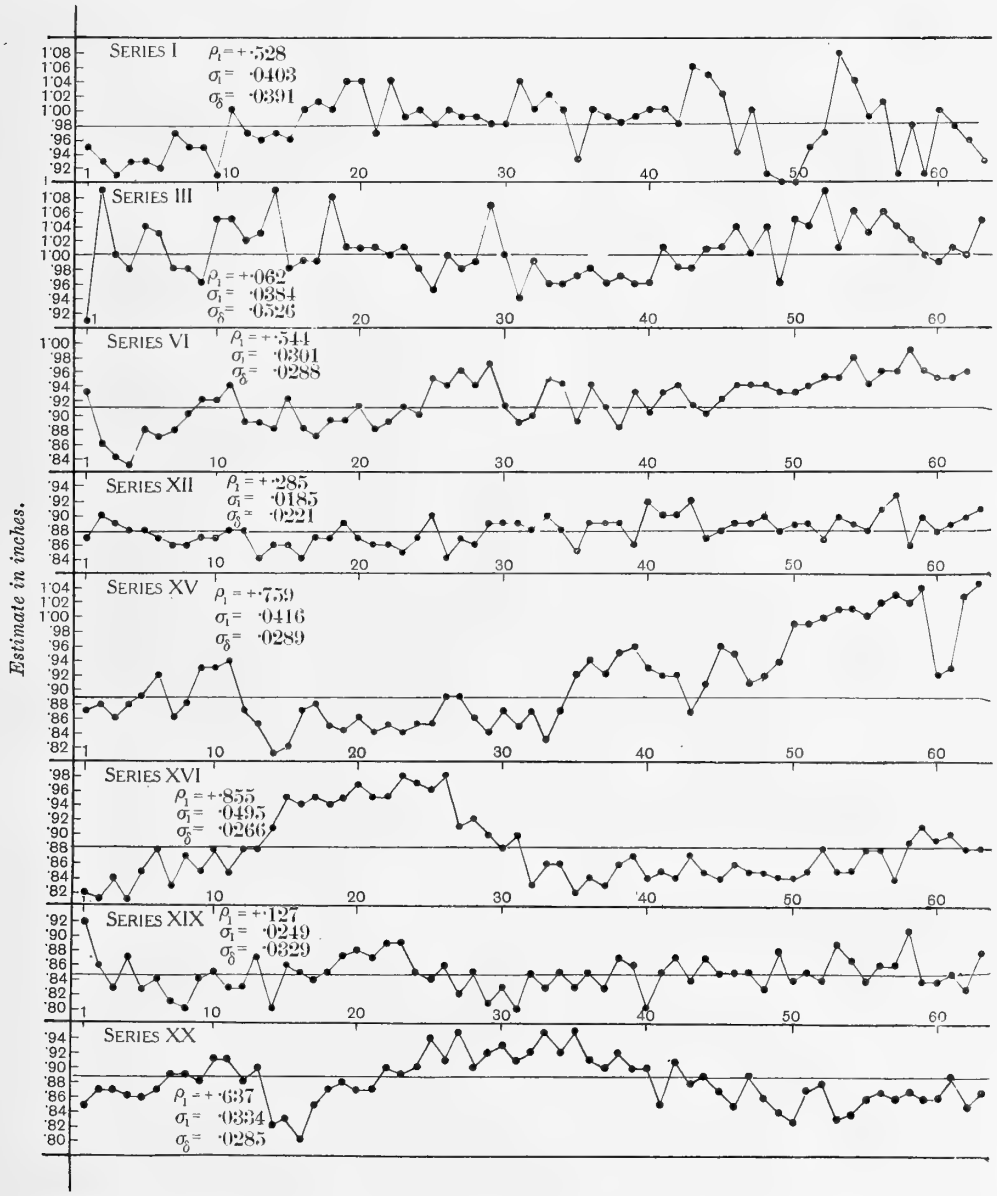
In Figure 15 are given eight representative series graphs which provide a good illustration of the variations in judgment. In the first two graphs (I and III), σ_δ is large and there are many sudden fluctuations, but in the later series σ_δ is lower and very constant in value. What may be described as the smoothness in the change of judgment is in some cases particularly noticeable; for example in the stretch between

$$\left.\begin{array}{l} y_{44} \text{ and } y_{53}, \text{ Series VI} \\ y_{2} \text{ and } y_{12}, \text{ Series XII} \\ y_{47} \text{ and } y_{59}, \text{ Series XV} \end{array}\right\}.$$

In making comparison with the similar diagrams for Trisections and Bisections allowance must be made for the differences in scale, but I think it is clear that this "smoothness" or gradual variation is a special feature of the 10 second counting; there is for instance no diagram of Trisections or Bisections which can compare with that of Series XVI of the counting, for high correlation combined with very gradual variation. But such a result is not unexpected, if the procedure of the experiment with the continuous counting be remembered.

A further point of interest is to examine how far a sudden "break" or discontinuity in the length of estimate influences the succeeding judgments. Among the 1000 observations forming the Groups 1 of the 20 series there are 61 "breaks" or differences between successive judgments of ·07 or over (in terms of the factor e/p),

i.e. of over twice S_δ, the standard deviation of first differences. In the diagram of Figure 16, ten observations are represented; the break between y_{t-1} and y_t or $y_t \sim y_{t-1}$, is supposed to be equal to, or greater than, ·07. If this large break influences the succeeding judgments, it is to be expected that the differences $y_{t+1} \sim y_t$, $y_{t+2} \sim y_t$, ... etc. will be smaller on the average than the differences $y_t \sim y_{t-1}$, $y_t \sim y_{t-2}$, ... etc.



The horizontal line intersecting each graph gives the mean of the first 50 observations in that series.

Fig. 15. 10 Second Counting. Diagrams representing variations in judgment.

In the first row of Table XIII are given the standard deviations of these differences taken from the 61 breaks; now in 14 of these cases there is what may be called a "double break," that is, after making one large variation to y_t, the judgment returns approximately to its previous state, both $y_t \sim y_{t-1}$ and $y_t \sim y_{t+1}$ being greater than or equal to ·07. While such cases may represent true variations in judgment, it is very possible that they result from some accidental error, a slowness in pressing the tapping key or in catching up the counting at the com-

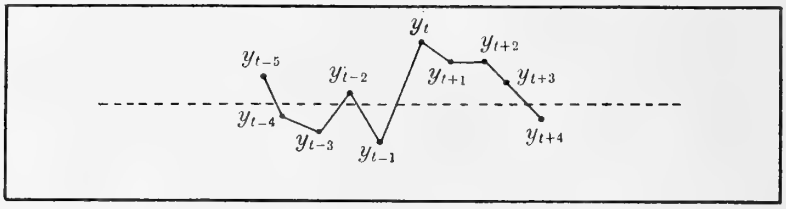


Fig. 16. Experiment *C*. Effect of a large break in judgment.

mencement of the observation, which was realised at the time and was not due to a real change in estimate. In the second row of the Table, therefore, are given the standard deviations taken from the 33 sets where there was no double break, and, in the third row, the standard deviations of 1st differences (taken from the whole 1000 judgments)

$$\text{between } y_t \text{ and } y_{t+1} = S_\delta = S_1' \sqrt{2\,(1 - \mathbf{R}_1')},$$

$$\text{„ } \quad y_t \text{ and } y_{t+2} \quad = S_1' \sqrt{2\,(1 - \mathbf{R}_2')},$$

$$\text{„ } \quad y_t \text{ and } y_{t+3} \quad = S_1' \sqrt{2\,(1 - \mathbf{R}_3')}, \text{ etc.}$$

TABLE XIII.

Standard Deviations of Differences between Judgment after "Break", y_t, and the Judgments y_{t-6} to y_{t+6}.

No. of Row	Number of Judgments	Previous Judgments					Succeeding Judgments				
		y_{t-6}	y_{t-4}	y_{t-3}	y_{t-2}	y_{t-1}	y_{t+1}	y_{t+2}	y_{t+3}	y_{t+4}	y_{t+6}
1	From 61 sets	·0692	·0647	·0624	·0624	·0851	·0476	·0541	·0553	·0549	·0623
		± ·0042	± ·0040	± ·0038	± ·0038	± ·0052	± ·0029	± ·0033	± ·0034	± ·0034	± ·0038
2	From 33 sets	·0757	·0682	·0636	·0667	·0810	·0346	·0421	·0508	·0483	·0635
		± ·0063	± ·0057	± ·0053	± ·0055	± ·0067	± ·0029	± ·0035	± ·0042	± ·0040	± ·0053
3	From total 1000	·0442	·0416	·0401	·0373	·0338	·0338	·0373	·0401	·0416	·0442

The probable errors are calculated from the usual expression, $\pm\,·6745\,\sigma/\sqrt{2n}$.

If we consider the values of these standard deviations together with their probable errors, we may say definitely that the effect of a large break or discontinuity in judgment is quite significant, and that the influence appears to last for at least four or five judgments. It cannot of course be decided whether the

breaks were caused by some chance external factor, or were due to a conscious change in estimate made by the observer on deciding, whether rightly or wrongly, that his second count was too short or too long*.

It will be noticed that the standard deviations of differences between these special pairs of judgments are in all cases greater than the corresponding standard deviations from the total 1000 judgments; this is to be expected, for the judgments y_t from which all the differences are taken are not a random selection of 61 (or 33) judgments, but include many of the most erratic and therefore those furthest from the mean.

(*b*) *The Combination of the Series.*

In combining the twenty series, D_k, S_k and $\mathbf{R}_k$ were calculated from the thirteen correlation tables of the judgments, and the values of these constants are given in Table XIV below. A glance at any one of the correlation tables showed that the 1000 judgments in any group did not follow a normal distribution, and in order to get a measure of this, the coefficient of skewness for the 1000 judgments in the combined Groups 1 (i.e. for the judgments $y_1, y_2 \ldots y_{50}$ of the twenty series) was calculated from the expression

$$\text{Skewness} = \frac{\sqrt{\beta_1}\,(\beta_2 + 3)}{2\,(5\beta_2 - 6\beta_1 - 9)},$$

where β_1 and β_2 are the fundamental ratios of the moments about the mean given by

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

The result was as follows:

$$\beta_1 = \cdot 2726, \quad \beta_2 = 2 \cdot 9739, \quad \text{Skewness} = \cdot 3684 \pm \cdot 0339,$$

showing a very significant degree of skewness, and the frequency follows a Type I curve of limited range.

The distribution of these 1000 observations made within a period of four consecutive days, gives but another example of the frequent inapplicability of the Normal Error Law.

Using the values of ρ_1, σ_1 and σ_2, $\mathbf{R}_1'$ is obtained from

$$\mathbf{R}_1' = \frac{\sum_m (\rho_1 \sigma_1 \sigma_2)}{\sqrt{\sum_m (\sigma_1^2) \sum_m (\sigma_2^2)}} = + \cdot 5200 \pm \cdot 0156 \ldots\ldots\ldots\ldots\ldots\text{(x) bis,}$$

and the remaining values of $\mathbf{R}_k'$, $k = 2, \ldots 13$, by the approximate method of Problem 1, p. 41. Perhaps the chief source of error in the method is variation in S_k, which has been assumed constant; in this experiment the range of S_k is only 1·8 % compared with 3·6 % for the Trisections and 2·5 % for the Bisections, and the results which are contained in the 6th row of Table XIV may be regarded, therefore, with reasonable confidence. As before, for the higher values of k, $\mathbf{R}_k'$ may be

* Eleven definite interruptions in the ordinary routine of counting, due to a mistap on the key or a miscount of the 10 seconds, were recorded at the time of observation, but only three of these resulted in breaks of judgment $\geq \cdot 07$, the limiting value taken in the above investigation.

TABLE XIV.

Constants of Combined Series (Counting Seconds).

		$k=1$	2	3	4	5	6	7	8	9	10	11	12	13	14
2	D_k	·9192	·9196	·9198	·9202	·9206	·9208	·9211	·9214	·9217	·9217	·9216	·9215	·9215	·9214
3	S_k	·05759 ± ·00087	·05747 ± ·00087	·05741 ± ·00087	·05762 ± ·00087	·05789 ± ·00087	·05810 ± ·00088	·05837 ± ·00088	·05845 ± ·00088	·05833 ± ·00088	·05847 ± ·00088	·05842 ± ·00088	·05839 ± ·00088	·05844 ± ·00088	·05812 ± ·00088
4	$\mathbf{R}_k$	+ ·8317 ± ·0066	·7955 ± ·0078	·7644 ± ·0089	·7456 ± ·0095	·7286 ± ·0100	·7129 ± ·0105	·6929 ± ·0111	·6885 ± ·0112	·6712 ± ·0117	·6619 ± ·0120	·6586 ± ·0121	·6365 ± ·0127	·6277 ± ·0129	
5	S_k'* (1)	·03455	·03431												
	S_k' (2)	·03426	—	—	—	—	—	—	—	—	—	—	—	—	·03444
6	$\mathbf{R}_k'$	+ ·5200 ± ·0156	+ ·4168 ± ·0176	+ ·3281 ± ·0190	+ ·2746 ± ·0197	+ ·2263 ± ·0202	+ ·1818 ± ·0206	+ ·1248 ± ·0210	+ ·1128 ± ·0211	+ ·0636 ± ·0212	+ ·0374 ± ·0213	+ ·0282 ± ·0213	− ·0348 ± ·0213	− ·0598 ± ·0212	
7	$\mathbf{R}_k'$ from equation (lvii)	+ ·528	+ ·418	+ ·331	+ ·262	+ ·208	+ ·164	+ ·130	+ ·103	+ ·082	+ ·065	+ ·051	+ ·040	+ ·032	
8	Difference ...	− ·008	− ·001	− ·003	+ ·013	+ ·018	+ ·018	− ·005	+ ·010	− ·018	− ·028	− ·023			
9	${}_1R_k$	− ·3925	− ·0150	− ·0367	− ·0054	− ·0040	+ ·0130	− ·0468	+ ·0386	− ·0239	− ·0177	+ ·0560	− ·0396		

* S_k' (1) was obtained from the relation $S_k' = \sqrt{\frac{1}{m}\Sigma_m \sigma_k^2}$, the values of σ_k being obtained without grouping.

S_k' (2) was obtained from the grouped data of the correlation tables. Its probable error is ± ·00052.

a little too low, and as a test of the amount of cumulative error which may be affecting $\mathbf{R}_{13}'$, I have worked out this constant directly from the relations

$$\left\{\begin{aligned} \mathbf{R}_{13} &= \frac{\mathbf{R}_{13}' S_1' S_{14}' + \frac{1}{m}\sum_m (D_1 - d_1)(D_{14} - d_{14})}{S_1 \times S_{14}}, \\ S_1^2 &= S_1'^2 + \frac{1}{m}\sum_m (D_1 - d_1)^2, \quad S_{14}^2 = S_{14}'^2 + \frac{1}{m}\sum_m (D_{14} - d_{14})^2, \end{aligned}\right.$$

EXPERIMENT C. 10 SECOND COUNTING. CORRELATION-INTERVAL DIAGRAM.

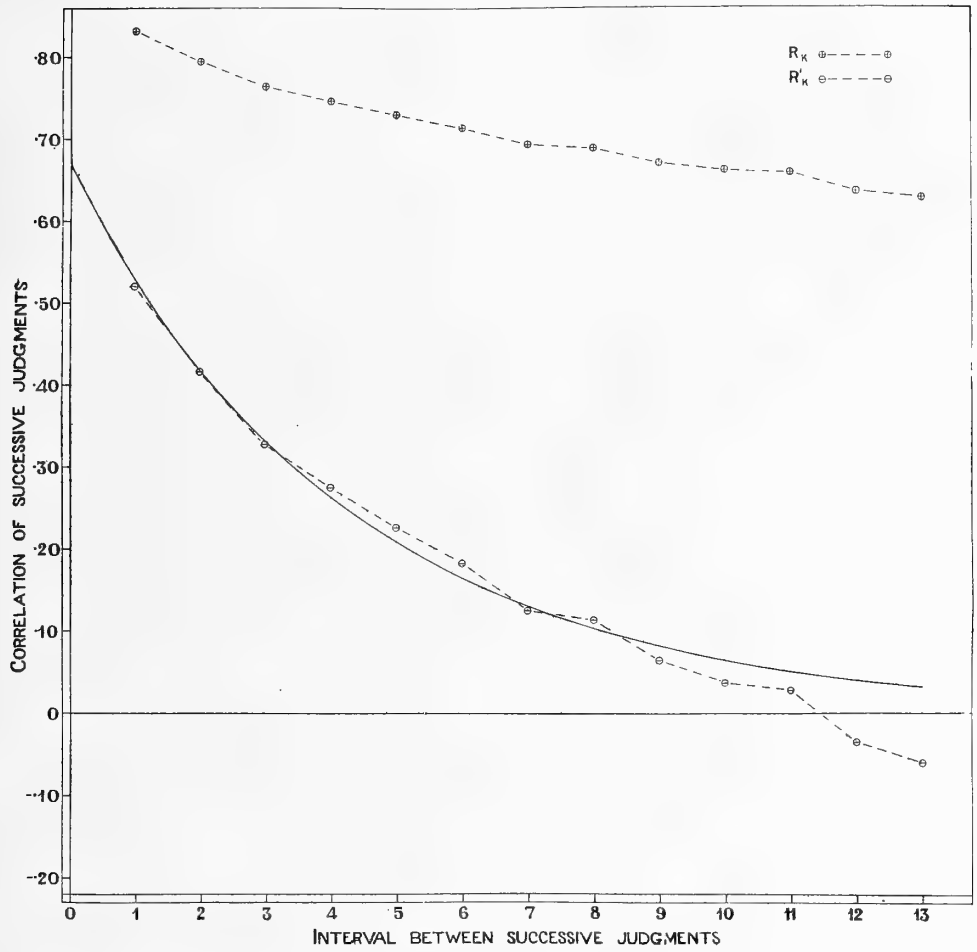


Fig. 17.

with the following results:

$$\mathbf{R}_{13}' = -\cdot 0124 \pm \cdot 0213, \quad L_{13} = +\cdot 632\,*,$$
$$S_1' = \cdot 03426, \qquad S_{14}' = \cdot 03444.$$

$\mathbf{R}_{13}'$ and presumably $\mathbf{R}_{12}'$ are not therefore significantly negative, and it seems probable that $\mathbf{R}_k'$ tends to zero as k increases, without oscillating about that value. The points $(k, \mathbf{R}_k)$ and $(k, \mathbf{R}_k')$ are plotted in Figure 17; the theoretical curve drawn in the diagram will be referred to in Section XI.

* L_k, or the limit to which $\mathbf{R}_k$ approaches as $\mathbf{R}_k'$ tends to zero is discussed on p. 34.

IX. Experiment *D*. Estimation of 10 seconds. Reduction of Observations.

(*a*) *The Individual Series.*

In Table XV are given the values of d (the mean of the 63 observations of a series, *not* those of Group 1 only), and of σ_1 and ρ_1 for the individual series; the low values of ρ_1 will be noted at once, and also the high values of σ_1 compared with those in the Counting Experiment. In Figure 19 below the means have been plotted to order of series, and if

x is the mean in the factor e/p,

y the order of series,

z the time in hours and fractions of an hour between 10 a.m. on December 7th, and the commencement of series,

TABLE XV.

Constants of Individual Series (*Estimate of Seconds*).

Series	d	σ_1	ρ_1	Time of Start	Date (1920)
I	1·151	·1217	+·1518 ± ·0932	10.45 a.m.	
II	1·111	·1254	+·2332 ± ·0902	11.30 a.m.	
III	1·109	·1330	−·0249 ± ·0953	12.10 p.m.	7th December
IV	1·052	·1393	+·1803 ± ·0923	2.0 p.m.	
V	·973	·1292	+·2632 ± ·0888	3.0 p.m.	
VI	1·119	·1349	+·1300 ± ·0938	10.15 a.m.	
VII	1·011	·1312	+·3673 ± ·0825	11.0 a.m.	
VIII	1·073	·1318	+·1631 ± ·0929	2.0 p.m.	8th December
IX	1·003	·1108	+·1976 ± ·0917	2.30 p.m.	
X	1·089	·0989	+·0380 ± ·0953	3.15 p.m.	
XI	1·204	·1519	+·3405 ± ·0843	10.0 a.m.	
XII	1·204	·1467	+·1415 ± ·0935	11.0 a.m	
XIII	1·091	·1166	+·3241 ± ·0854	12.0 midday	9th December
XIV	1·036	·1059	+·0566 ± ·0951	2.0 p.m.	
XV	1·132	·1884	+·4814 ± ·0733	3.15 p.m.	
XVI	1·170	·1500	+·1036 ± ·0944	10.0 a.m.	
XVII	1·421	·1520	−·0834 ± ·0947	11.0 a.m.	
XVIII	1·300	·1591	+·2314 ± ·0903	12.0 midday	10th December
XIX	1·243	·1708	+·2260 ± ·0905	2.0 p.m.	
XX	1·170	·1833	+·1659 ± ·0928	2.45 p.m.	

Correlation between ρ_1 and σ_1, $r_{\sigma_1 \rho_1} = +·176 \pm ·146$ (calculated from correlation of ranks).

we have for the regression lines

$$x - 1·1333 = +·01018\,(y - 10·5) \quad \text{..... (xxxix),}$$

$$x - 1·1333 = +·002493\,(z - 38·62) \quad \text{.....(xl).}$$

The coefficients of correlation are

$$r_{xz} = +·638 \pm ·089, \quad r_{xy} = +·562 \pm ·103, \quad r_{yz} = +·983 \pm ·005,$$

giving partial correlation coefficients

$$r_{xz.y} = +\cdot570 \pm \cdot102, \quad r_{xy.z} = -\cdot470 \pm \cdot118.$$

These latter coefficients suggest that the secular change for observations spread over a number of days will be a lengthening in estimation, but that, if a number of series are done in rapid succession, the tendency will be for a shortening; in fact we should expect the sessional change to be in the opposite direction to the secular, as for the Bisections.

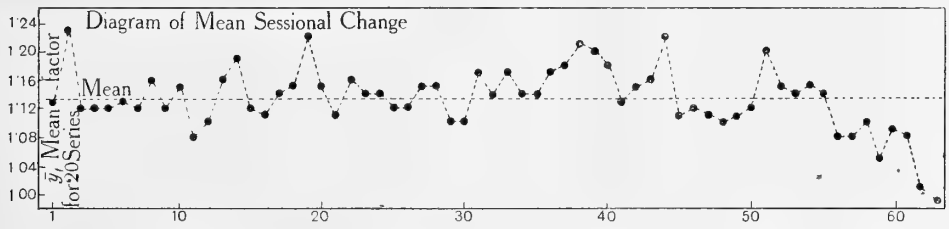


Fig. 18. 10 Second Estimation. t, Order of Observation in Series.

The values of $\bar{y}_t$ have been plotted in Figure 18; the best fitting line has not been calculated, but it would certainly correspond very closely with the mean, $y = 1{\cdot}1333$. There is in fact apparently no mean sessional change, though the drop in the last eight values of $\bar{y}_t$ may be significant, and a mark of the tendency suggested by the negative value of $r_{xy.z}$.

In Figure 19 the centres of the small circles represent the positions of the means of the 63 observations of each series; these points have been fitted with the cubic

$$x = 1{\cdot}093971 + {\cdot}022116\,(y-10{\cdot}5) + {\cdot}001174\,(y-10{\cdot}5)^2 - {\cdot}0002002\,(y-10{\cdot}5)^3 \ldots \text{(xli)},$$

which is the middle of the three curves. There is evidence of a slight secular change, the length of the estimation *increasing* towards the end of the experiment. If however it is remembered that the 20 series were carried out in 4 days, it will be seen that there is in general a *decrease* in estimation in the course of the 5 series done in any one day. It is this daily drop that the coefficient $r_{xy.z} (= -\cdot470)$ is picking out. Now in addition to the secular change in personal equation, the figures in Table XV suggest that there is also a secular change in standard deviation. The vertical lines on each side of the series-means in Figure 19 equal in length the corresponding standard deviations, or σ_1's. These values of σ_1 have been fitted with the cubic

$$x' = {\cdot}129006 + {\cdot}001072\,(y-10{\cdot}5) + {\cdot}000302\,(y-10{\cdot}5)^2 + {\cdot}0000214\,(y-10{\cdot}5)^3 \ldots \text{(xlii)},$$

and the other two curves in the diagram have ordinates equal to $x + x'$ and $x - x'$, so that the distance between the central curve and either of the outer curves, gives the smoothed value of the standard deviation at the point. The diagram provides a generalised representation of a secular change in personal equation and standard deviation.

The factor for a true 10 second interval would be $\frac{10{\cdot}0}{10{\cdot}2} = {\cdot}98$, and was most nearly approached by the means of Series V, VII and IX, while in the case of XVII the mean estimation nearly reached the high value of 15 seconds.

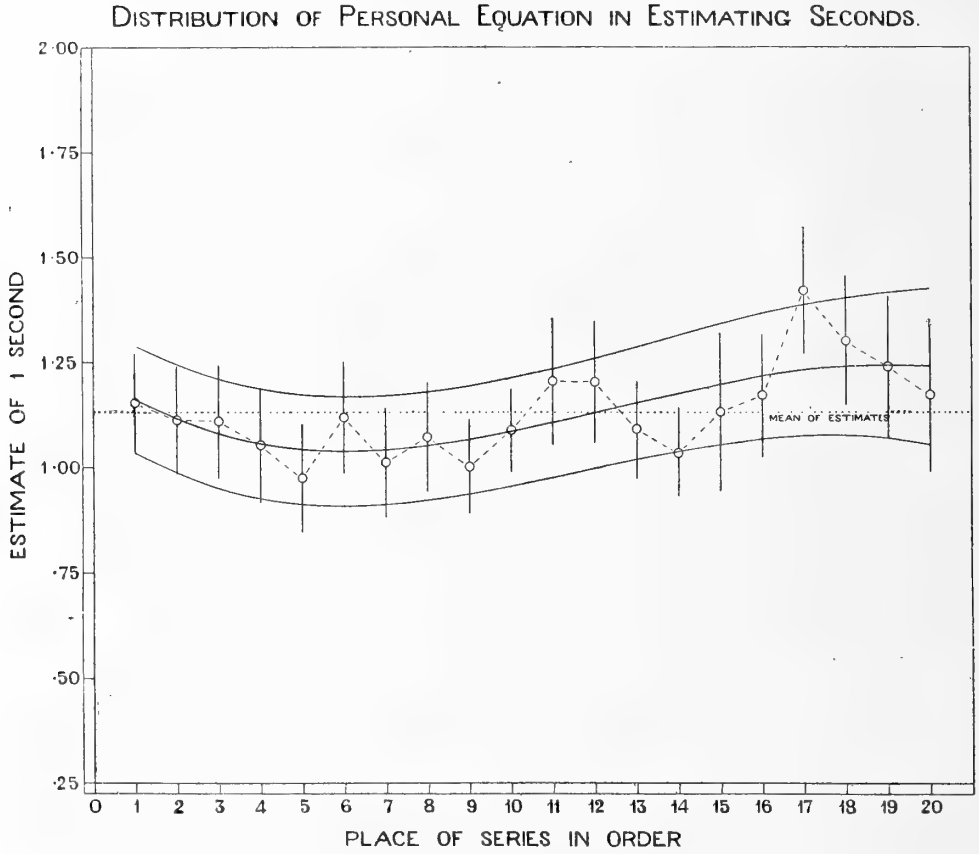


Fig. 19.

(*b*) *The Combination of the Series.*

In combining the twenty series, D_k, S_k and $\mathbf{R}_k$ were calculated from the thirteen correlation tables of the observations of the combined series. Using the correlations and standard deviations of the separate series, $\mathbf{R}_1'$ is obtained from

$$\mathbf{R}_1' = \frac{\sum_m (\rho_1 \sigma_1 \sigma_2)}{\sqrt{\sum_m (\sigma_1^2) \sum_m (\sigma_2^2)}} = +{\cdot}19841 \pm {\cdot}02049 \ldots\ldots\ldots\ldots \text{(x) bis,}$$

and $$S_1' = {\cdot}14101, \quad S_2' = {\cdot}14056.$$

Then using this value of $\mathbf{R}_1'$, and the first difference correlation equations (Problem 1, p. 41), $\mathbf{R}_k'$ can be calculated for $k = 2, \ldots 12$. The values of these quantities are given in the Table XVI below.

The fall in $\mathbf{R}_k$ is small, and although there is considerable irregular variation from $\mathbf{R}_7$ onwards, it appears that $\mathbf{R}_k$ will not vanish as k increases, but approach a constant value in the neighbourhood of + ·35. This can be tested; we have from the equations (vi) to (x)

$$\mathbf{R}_k = \frac{\mathbf{R}_k' S_1' S'_{k+1} + \frac{1}{m}\sum_m (D_1 - d_1)(D_{k+1} - d_{k+1})}{\sqrt{\left\{S_1'^2 + \frac{1}{m}\sum_m (D_1 - d_1)^2\right\}\left\{S'^2_{k+1} + \frac{1}{m}\sum_m (D_{k+1} - d_{k+1})^2\right\}}} \quad \ldots\ldots \text{(xliii)},$$

TABLE XVI.

Constants of Combined Series (Estimating Seconds).

	$k=1$	2	3	4	5	6	7	8	9	10	11	12	13
D_k	1·1421	1·1440	1·1415	1·1413	1·1421	1·1423	1·1416	1·1402	1·1395	1·1391	1·1382	1·1378	1·1351
S_k	·1749 ± ·0026	·1749 ± ·0026	·1759 ± ·0027	·1761 ± ·0027	·1764 ± ·0027	·1760 ± ·0027	·1754 ± ·0026	·1757 ± ·0026	·1763 ± ·0027	·1756 ± ·0026	·1760 ± ·0027	·1767 ± ·0027	·1773 ± ·0027
$\mathbf{R}_k$	+ ·4825 ± ·0164	·4269 ± ·0174	·3965 ± ·0180	·3913 ± ·0181	·3764 ± ·0183	·3755 ± ·0183	·3983 ± ·0180	·4045 ± ·0178	·3488 ± ·0187	·3691 ± ·0184	·3524 ± ·0187	·3831 ± ·0182	
S_k'	·14101 ± ·00213	·14056 ± ·00212											
$\mathbf{R}_k'$	+ ·19841 ± ·02049	+ ·1123 ± ·0211	+ ·0652 ± ·0212	+ ·0570 ± ·0212	+ ·0338 ± ·0213	+ ·0332 ± ·0213	+ ·0693 ± ·0212	+ ·0798 ± ·0212	− ·0056 ± ·0213	+ ·0267 ± ·0213	+ ·0017 ± ·0213	+ ·0501 ± ·0213	
${}_1R_k$	− ·4463	− ·0243	− ·0243	+ ·0094	− ·0135	− ·0229	+ ·0160	+ ·0598	− ·0734	+ ·0357	− ·0458		

$$S_\delta = S_1' \sqrt{2(1 - \mathbf{R}_1')} = \cdot 1785.$$

and as the sessional change for the series is very small, we may make the approximation

$$\sum_m (D_1 - d_1)(D_{k+1} - d_{k+1}) = \sum_m (D_1 - d_1)^2 = \sum_m (D_{k+1} - d_{k+1})^2 \text{ for all values of } k,$$

and in view of the constancy of S_k,

$$S_1' = S'_{k+1} \text{ for all values of } k.$$

Then on the assumption that there is no significant periodic variation in the observations,

$$\mathbf{R}_k' \to 0 \text{ as } k \text{ increases,}$$

and from (xliii)

$$\mathbf{R}_k \to \frac{\frac{1}{m}\sum_m (D_1 - d_1)^2}{S_1'^2 + \frac{1}{m}\sum_m (D_1 - d_1)^2} = + \cdot 354.$$

The correlations $\mathbf{R}_k'$ become rapidly insignificant; the values tabulated are of course subject to the errors of the method of approximation, but as in the case of the 10 second Counting Experiment, these should not be large owing to the constancy of S_k*. The points $(k, \mathbf{R}_k)$ and $(k, \mathbf{R}_k')$ are plotted in Figure 20; the two curves there drawn will be referred to in Section XI below.

* The difference between S_1 and S_{13} is one of 1·4 °/₀ only.

(c) *Comparison of Experiments C and D.*

It has been found that in both the Counting and the Estimating Experiments there is evidence of a secular change in personal equation, and that in both cases the tendency is for the estimates to depart further from the true value of 10 seconds in the later series; in the Counting Seconds there is a decrease, in the Estimating Seconds an increase in length of estimate. There is also very little evidence of regular sessional change in either experiment.

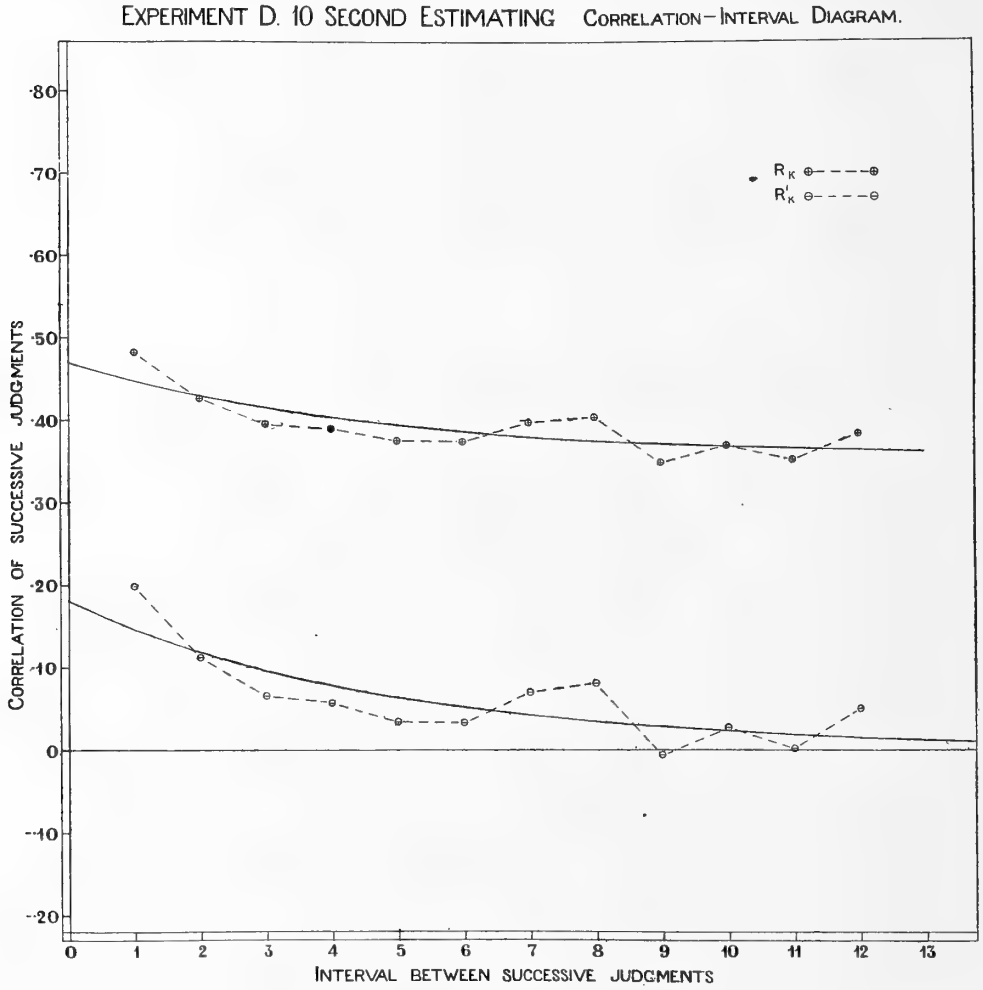


Fig. 20.

Beyond this the similarity ceases; it is only necessary to compare the values of the chief constants (defined on p. 36),

	S_1'	S_δ	$\mathbf{R}_1'$
Counting	·0343	·0338	$+ ·5200 \pm ·0156$
Estimating	·1410	·1785	$+ ·1984 \pm ·0205$

The variations in judgment in the Estimating Experiment are very large compared with those in the other, and at the same time there is low correlation between successive judgments, so that the observations will be found to be scattered far more nearly in accordance with the Normal Error Law than in the three preceding experiments. In the case of the Counting Experiment, the skew distribution of the 1000 observations has already been referred to. But for one or two exceptions (as III and XIX) the individual series in the Counting conform more closely to a general type than in the Trisections or Bisections, and this results in the very smooth values of the constants $\mathbf{R}_k$ and $\mathbf{R}_k'$.

X. EXPERIMENT *E*. PLATE MEASUREMENTS WITH ZEISS COMPARATOR.

The values of ρ_1 only have been calculated; these, with σ_1 and a brief description of the nature of the marking measured, are given in the Table XVII; σ_1 is in millimetres. Series I—VIII involved settings of both slide and micrometer, IX of micrometer only.

No great weight can be attached to the result of one series of 50 readings on a marking, but it is justifiable to draw certain conclusions from the results of the eight series. In the first place, there appears to be a significant correlation between the successive measures of the edge of a band (I and II), but in measuring the centres, i.e. in bisecting a bright maximum with the cross wire, there is on the whole no correlation. This perhaps might be expected; the edges of bands or maxima in photographic spectra are not quite sharply cut, so that some uncertainty must exist in the observer's mind as to where the real edge should be taken to be; his opinion on this point may vary throughout the course of the sitting, and consequently correlation will be found between the successive readings. On the other

TABLE XVII.

Series	ρ_1	σ_1	Description of marking	
I	$+ \cdot 384 \pm \cdot 081$	·0016	Sharp edge of bright band	Edge
II	$+ \cdot 467 \pm \cdot 075$	·0015	Slightly vaguer edge than I	Edge
III	$+ \cdot 117 \pm \cdot 094$	·0007	Clear and narrow maximum	Centre
IV	$+ \cdot 090 \pm \cdot 095$	·0012	,, ,, ,,	Centre
V	$+ \cdot 021 \pm \cdot 095$	·0016	,, ,, ,,	Centre
VI	$- \cdot 001 \pm \cdot 095$	·0019	Broad and obscure ,,	Centre
VII	$- \cdot 050 \pm \cdot 095$	·0022	,, ,, soft ,,	Centre
VIII	$+ \cdot 227 \pm \cdot 091$	·0041	,, ,, ,, ,,	Centre
IX	$+ \cdot 288 \pm \cdot 087$	·0004	Micrometer screw settings only	

hand, in the bisection of a narrow maximum, there will be little doubt as to the position of the centre; the real estimate of the observer will vary but slightly, and the variations in the reading will be due mainly to failure in breaking off the push

or pull of the slide at the right moment. It is possible that unconscious "over-pulls" or "under-pulls" may go in runs together, but the measures seem to show that this is not the case, and that the correlation of successive judgments is due rather to correlated changes of mental estimate than to those of a more physical character. If it were more difficult to bisect a maximum, if there were greater opportunity for variation, it is probable that there would be a correlation of successive judgments, and this is perhaps illustrated by the case of Series VIII, which has the largest standard deviation (·0041) and also a correlation ($\rho_1 = + \cdot 227 \pm \cdot 091$) possibly significant.

The result of IX suggests that there is a correlation between successive settings of the micrometer wires in the second eyepiece; this correlation would of course enter into the results of I—VIII, but the standard deviation of IX (·0004) is so small that the effect will be insignificant where the variations in slide settings are large.

As a matter of practical application these results serve to emphasise the importance of the routine of measurement usually adopted; if, for example, it is proposed to take four readings of each of a number of markings on a plate, the four readings should not be made in succession, but all the markings should be measured once, and then perhaps a short interval taken before the second measuring is made, and so on. This method should eliminate the error in the mean of several measurements of a marking, which may arise from a correlation of successive judgments, as well as errors due to change in temperature of instrument or plate, etc.

XI. Analysis of the correlation between successive judgments.

(*a*) *The Theory of correlated Estimates and accidental Errors.*

It has been seen that in the case of the Bisection and Timing Experiments when the secular term was removed the coefficients of correlation of the successive judgments, or the constants $\mathbf{R}_k'$, diminished to approximately zero values as k, the interval between the judgments correlated, was increased. In the Trisection Experiment, owing to the marked sessional change which was repeated in practically all the series, $\mathbf{R}_k'$ appeared to approach a value of $+ \cdot 16$ and not zero as k was increased; the sessional change in this case appeared to be of parabolic rather than linear form, and it seemed possible that if the ordinates of the "best" fitting parabola of each series were removed from the observations, the coefficients of correlation of the residuals, or the $\mathbf{R}_k'''$s, would tend to zero as k increased, as in the case of the other three experiments in which there was no large sessional change. The points representing the values of $\mathbf{R}_k'$ which have been plotted in Figures 8, 12, 17 and 20 appear on the whole to lie so nearly on a smooth curve, that it is of no little interest to inquire whether we can obtain equations to such curves based on some definite theory of the physiological factors underlying the variations in an observer's judgment.

In the first place we have seen that neither a secular change in personal equation—the variation in series means—nor a simple sessional change such as that represented by the straight line or by a second order parabola considered in the Trisection Experiment, will account for the whole of the correlation of successive judgments. We must therefore conclude that quite apart from the large scale variations in judgment which are due to the more gradual changes of state in the observer resulting, perhaps, from experience or fatigue, there is a definite relationship between the small scale variations in judgment; if judgment y_t is greater than the average of the five or six preceding judgments, then we shall on the whole expect that y_{t+1}, the next judgment, will also be greater. I propose therefore to consider what results will follow from the assumption that y_t has a correlation r with y_{t-1} and y_{t+1}, but that for y_{t+1} or y_{t-1} constant it has no partial correlation with y_{t-2} and y_{t+2} or judgments at greater intervals. In other words we will suppose that the observer's estimation at any moment is only influenced by the preceding estimation, and only through this, and not directly, by the earlier estimations.

Let us take the successive judgments y_t, y_{t+1}, $y_{t+2} \ldots y_{t+k} \ldots$ and suppose that the total correlation between y_t and y_{t+k} is ρ_k, where $k = 1, 2, 3, \ldots$, and $\rho_1 = r$. If there is no partial correlation between y_t and y_{t+2}, y_{t+1} being constant we must have

$$\rho_2 - \rho_1^2 = 0 \text{ or } \rho_2 = r^2.$$

In the same way if there is no partial correlation between y_t and y_{t+3} when y_{t+1} (or y_{t+2}) is constant,

$$\rho_3 - \rho_1\rho_2 = 0 \text{ or } \rho_3 = r^3,$$

and in general we find that

$$\rho_k = r^k \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(xliv)}.$$

In reaching this simple result there is a point however that has been overlooked; it has been assumed that there is some physiological or psychological significance in the correlation of an estimate of a quantity and in the preceding estimate, but it must be remembered that the value which the observer records may not be exactly that which he wished to record, or in other words he may be unable to record his true estimate. Thus in bisecting a line it is likely that the pencil point will not strike the paper exactly at the spot intended, or in counting 10 seconds the tapping of the key may not be exactly synchronised with the beginning or end of the count, and there may be many other little external influences of which the observer is unaware, which will all combine to form what may be termed an accidental error superimposed upon the true correlated estimation. Let us examine how the relation (xliv) will be modified by introducing the idea of these accidental and uncorrelated errors; we must suppose that the observer's recorded judgment y_t is made up of two parts, α_t his actual estimate at the moment of record and β_t some complex of accidental errors affecting his record. Then

$$y_t = \alpha_t + \beta_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(xlv)}.$$

Now if we assume that the accidental errors β_t are as like to be positive as negative, and that they will not be correlated in any manner among themselves

nor with the fundamental part of the judgment α_t, we shall have the following approximate relations

$$\left.\begin{array}{lll} \sum_1^N \beta_{t+k} = 0 \text{ for } k = 1, 2, 3, \ldots, & \text{where } N \text{ is large compared with } k & \\ \sum_1^N \beta_t \beta_{t+k} = 0 \quad ,, \quad ,, & & \\ \sum_1^N \beta_{t+k} \alpha_{t+k'} = 0 \quad ,, \quad ,, & \text{where } k \text{ and } k' \text{ take any of the values } 1, 2, 3, \ldots \text{etc.} & \end{array}\right\} \quad \ldots\ldots\text{(xlvi)}.$$

But the correlation between successive values of the y's at intervals of k is

$$\rho_k = \frac{\sum_{t=1}^N (\alpha_t + \beta_t)(\alpha_{t+k} + \beta_{t+k}) - N \sum_{t=1}^N \frac{\alpha_t + \beta_t}{N} \sum_{t=1}^N \frac{\alpha_{t+k} + \beta_{t+k}}{N}}{\sqrt{\left\{\sum_{t=1}^N (\alpha_t + \beta_t)^2 - N\left(\sum_{t=1}^N \frac{\alpha_t + \beta_t}{N}\right)^2\right\}\left\{\sum_{t=1}^N (\alpha_{t+k} + \beta_{t+k})^2 - N\left(\sum_{t=1}^N \frac{\alpha_{t+k} + \beta_{t+k}}{N}\right)^2\right\}}}$$

$$= \frac{\sum_{t=1}^N (\alpha_t \alpha_{t+k}) - N \sum_{t=1}^N \frac{\alpha_t}{N} \sum_{t=1}^N \frac{\alpha_{t+k}}{N}}{\sqrt{\left\{\sum_{t=1}^N \alpha_t^2 - N\left(\sum_{t=1}^N \frac{\alpha_t}{N}\right)^2 + \sum_{t=1}^N \beta_t^2\right\}\left\{\sum_{t=1}^N \alpha_{t+k}^2 - N\left(\sum_{t=1}^N \frac{\alpha_{t+k}}{N}\right)^2 + \sum_{t=1}^N \beta_{t+k}^2\right\}}}$$

in view of the relations (xlvi)

$$= \frac{[\alpha_t \alpha_{t+k}]}{\sqrt{(\bar{\alpha}_1^2 + \bar{\beta}_1^2)(\bar{\alpha}_{k+1}^2 + \bar{\beta}_{k+1}^2)}},$$

where $[\alpha_t \alpha_{t+k}]$ is the first order product moment coefficient referred to mean of the successive α's at intervals of k,

and $\sqrt{\bar{\alpha}_k^2}$ is the standard deviation of $\alpha_k, \alpha_{k+1}, \ldots \alpha_{k+N}$,

$\sqrt{\bar{\beta}_k^2}$,, ,, ,, $\beta_k, \beta_{k+1}, \ldots \beta_{k+N}$,

and $\sqrt{\bar{\alpha}_k^2 + \bar{\beta}_k^2}$,, ,, ,, $y_k, y_{k+1}, \ldots y_{k+N}$.

Now unless there is a steady sessional change in the α's, we may assume that for large values of N

$$\bar{\alpha}_1^2 = \bar{\alpha}_2^2 = \ldots = \bar{\alpha}_k^2 = \ldots = \bar{\alpha}^2, \text{ say,}$$

and similarly unless the accidental errors are steadily increasing or decreasing in magnitude

$$\bar{\beta}_1^2 = \bar{\beta}_2^2 = \ldots = \bar{\beta}^2,$$

and we have $$\rho_k = \frac{[\alpha_t \alpha_{t+k}]}{\bar{\alpha}^2 + \bar{\beta}^2} = \frac{\bar{\alpha}^2}{\bar{\alpha}^2 + \bar{\beta}^2} \cdot \frac{[\alpha_t \alpha_{t+k}]}{\bar{\alpha}^2} = \frac{\bar{\alpha}^2}{\bar{\alpha}^2 + \bar{\beta}^2} \cdot r_{\alpha_t, \alpha_{t+k}}.$$

But on the assumption made above of zero partial correlation between two estimates which are not consecutive, we have found that $r_{\alpha_t, \alpha_{t+k}}$, the correlation between the observer's real estimates at intervals of k, can be expressed in the form r^k, and therefore

$$\rho_k = \frac{\bar{\alpha}^2}{\bar{\alpha}^2 + \bar{\beta}^2} r^k = q r^k \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xlvii)},$$

where q is a constant not depending on the interval k. With this expression for the correlation we shall of course find an apparent partial correlation between the judgments at intervals greater than one; for example the partial correlation between y_t and y_{t+2}, y_{t+1} being constant, is $\frac{q(1-q)r^2}{1-q^2r^2}$, and does not vanish unless $q=1$. According to the theory suggested this is however a spurious correlation due solely to the presence of the accidental errors.

The next problem is to inquire how far a relation of the type of (xlvii) will fit the correlation coefficients which have been calculated for the Experiments *A*, *B*, *C* and *D*. In the first place, in order to get as smooth values for the coefficients as possible we must combine the 20 series, which we may do if we remove the secular change as represented by the variation in the series means; this step is clearly necessary for we are considering the relationship between judgments made in close proximity and are not concerned for the moment with the variation in personal equation from day to day. We must therefore deal with the coefficients of correlation $\mathbf{R}_k'$ and endeavour to fit a curve $z=qr^x$ through the points $x=k$, $z=\mathbf{R}_k'$. I will consider the different experiments in turn.

(*b*) *Application of Theory to results of Experiments.*

Experiment A.

The curve represented by $z=qr^x$ is asymptotic to the x axis (as $r<1$), so that if it is to fit the points $(k, \mathbf{R}_k')$ it is necessary that $\mathbf{R}_k'$ should tend to zero as k increases. But the values of $\mathbf{R}_k'$ given in Table V, p. 58, appear to tend as k increases, to a limiting value between $+\cdot16$ and $+\cdot15$ rather than to zero. I think that this results from the marked sessional changes which have been represented in mean form by a second order parabola (see Equation (xxx) and Figure 4), and that if there is a physiological significance in the distinction between the sessional change and the residual variations of the observations when freed from this change, it will be of interest to find out how the coefficients of correlation of these successive residuals—what have been termed the $\mathbf{R}_k''$'s—fall off as the interval or k is increased. Should it be found that the $\mathbf{R}_k''$'s follow the law

$$\mathbf{R}_k'' = qr^k,$$

the argument in favour of distinguishing the sessional change from the residual variations will be strengthened.

It was found that the values of $\mathbf{R}_k'$ given in Table V could be fitted closely by a curve of the form

$$z = p + qr^k \quad \text{.....(xlviii),}$$

where p, q and r are constants.

A rough trial gave the following approximate values:

$$p_0 = \cdot157, \quad q_0 = \cdot69, \quad r_0 = \cdot73.$$

Now if $z = f(p, q, r)$

$$= f(p_0, q_0, r_0) + \delta p \frac{\partial f_0}{\partial p_0} + \delta q \frac{\partial f_0}{\partial q_0} + \delta r \frac{\partial f_0}{\partial r_0} \quad \text{to first order,}$$

$$= p_0 + q_0 r_0^k + \delta p + r_0^k \delta q + k q_0 r_0^{k-1} \delta r,$$

we have as equations of condition for a least square solution

$$\delta p + r_0^k \delta q + k q_0 r_0^{k-1} \delta r = \mathbf{R}_k' - p_0 - q_0 r_0^k, \text{ for } k = 1, 2, \dots 13.$$

Using the values of p_0, q_0 and r_0 given above, the corrections δp, δq and δr were calculated and gave finally as the best fitting numerical equation,

$$\mathbf{R}_k' = \cdot 1524 + \cdot 6817 (\cdot 7105)^k \text{ ....(xlix).}$$

TABLE XVIII.

Values of the $\mathbf{R}_k$*'s for Trisection Experiments.*

1	2	3	4	5	6	7	8
k	$\mathbf{R}_k'$ (direct calculation)	$\mathbf{R}_k'$ (from equation (xlix))	Difference Col. 2—Col. 3	Probable Error of $\mathbf{R}_k'$	Values obtained from (lii) on assumption of constancy of G_k		$\mathbf{R}_k''$ (from equation (lvi))
					S_k''	$\mathbf{R}_k''$	
0	—	+ ·834	—	—	—	—	+ ·804
1	+ ·625	·637	− ·012	± ·013	·0773	+ ·550	·571
2	·523	·497	+ ·026	± ·016	·0776	·431	·406
3	·388	·397	− ·009	± ·018	·0778	·268	·288
4	·315	·326	− ·011	± ·019	·0781	·183	·205
5	·281	·276	+ ·005	± ·020	·0778	·142	·146
6	·232	·240	− ·008	± ·020	·0782	·084	·103
7	·222	·215	+ ·007	± ·020	·0782	·071	·074
8	·191	·197	− ·006	± ·021	·0783	·035	·052
9	·165	·184	− ·019	± ·021	·0787	·006	·037
10	·183	·175	+ ·008	± ·021	·0802	·031	·026
11	·168	·168	·000	± ·021	·0823	·017	·019
12	·172	·164	+ ·008	± ·021	·0834	·023	·013
13	+ ·160	+ ·160	·000	± ·021	·0840	+ ·009	+ ·009
14	—	—	—	—	·0840	—	—

In the second column of Table XVIII are given the values of $\mathbf{R}_k'$ taken from Table V and in the fifth column their probable errors; the values of $\mathbf{R}_k'$ given by equation (xlix) are in the third column, and in the fourth are the differences col. 2—col. 3. It will be seen that the fit is a good one, the difference being only greater than the probable error in the case of $\mathbf{R}_2'$. The points (k, $\mathbf{R}_k'$) and the curve of (xlix) are shown in Figure 8 (p. 64).

The problem before us is therefore this; can we explain the constant p in equation (xlviii) in terms of the sessional changes? We have seen that the mean sessional change for the 20 series can be represented by a parabola of the second order, but we must allow for a different change in each series. Let us suppose that

$$y = f_p(t)$$

will represent the sessional change in the pth Series after the secular term represented by the series mean has been removed, so that instead of equation (xlv) of p. 89, we have

$$y_t' = f_p(t) + \alpha_t + \beta_t = f_p(t) + Y_t \text{ ....(l),}$$

where $Y_t = \alpha_t + \beta_t$.

Then if $\sum_m$ indicates summation for the m (or 20) series, $n = 50$, the number of observations in each group of a series, and k takes any of the group numbers 1, 2, ... 14, since $y = f_p(t)$ will be the "best" fitting curve of its type $\sum_{t=1}^{n} Y_{t+k-1} = 0$ approximately, and on combining the m series

$$\sum_m \sum_{t=1}^{n} (Y_{t+k-1}) = 0.$$

Again we have no reason to suppose that there will be any correlation between the sessional term $f_p(t)$ and the residual Y_t, so that

$$\sum_m \sum_{t=1}^{n} \{Y_{t+k-1} f_p(t+k'-1)\} = 0,$$

for all values of k and k' between 1 and 14.

As y_t' is freed from the secular term, using the relations above we have that

$$\mathbf{R}_k' = \frac{\sum_m \sum_{t=1}^{n} \{(f_p(t)+Y_t)(f_p(t+k)+Y_{t+k})\} - mn \sum_m \sum_{t=1}^{n} \left\{\frac{f_p(t)}{mn}\right\} \sum_m \sum_{t=1}^{n} \left\{\frac{f_p(t+k)}{mn}\right\}}{\sqrt{\left[\sum_m \sum_{t=1}^{n} (f_p(t)+Y_t)^2 - mn \left\{\sum_m \sum_{t=1}^{n} \frac{f_p(t)}{mn}\right\}^2\right] \left[\sum_m \sum_{t=1}^{n} (f_p(t+k)+Y_{t+k})^2 - mn \left\{\sum_m \sum_{t=1}^{n} \frac{f_p(t+k)}{mn}\right\}^2\right]}} \quad \text{.....(li),}$$

$$= \frac{\mathbf{R}_k'' S_1'' S_{k+1}'' + F_k}{\sqrt{(S_1''^2 + G_1^2)(S_{k+1}''^2 + G_{k+1}^2)}},$$

where $\mathbf{R}_k''$ is the coefficient of correlation between Y_t and Y_{t+k}, S_1'' and S_{k+1}'' are the standard deviations of the Y's of Groups 1 and $k+1$ (see (xi) and (xii) on page 35), and

$$\left.\begin{aligned} F_k &= \frac{1}{mn} \sum_m \sum_{t=1}^{n} f_p(t) f_p(t+k) - \left\{\sum_m \sum_{t=1}^{n} \frac{f_p(t)}{mn}\right\} \left\{\sum_m \sum_{t=1}^{n} \frac{f_p(t+k)}{mn}\right\} \\ G_k^2 &= \frac{1}{mn} \sum_m \sum_{t=1}^{n} \{f_p(t+k-1)\}^2 - \left\{\sum_m \sum_{t=1}^{n} \frac{f_p(t+k-1)}{mn}\right\}^2 \end{aligned}\right\}.$$

It will be seen that G_k is the standard deviation of the ordinates of the curves representing the sessional changes, $y = f_p(t)$, which correspond to the observations in the kth groups, while $\frac{F_k}{G_1 G_{k+1}}$ is the correlation of these successive ordinates at intervals of k. If the sessional changes were linear this correlation would be unity, and a little consideration will show that if the sessional change in each series can be represented by a curve of gradual bend, the correlation will not be far from this value. For example in the case of the parabola (Equation (xxx), p. 47) which was fitted to the *mean* sessional change and is drawn in Figure 4, it is found that

$$\frac{F_{13}}{G_1 G_{14}} = +\cdot 994.$$

We shall therefore make no great error in assuming that

$$F_k = G_1 G_{k+1},$$

and it follows that the relation for $\mathbf{R}_k'$ can be expressed in the form

$$\mathbf{R}_k' = \frac{1}{\sqrt{\left(1+\frac{S_1''^2}{G_1''^2}\right)\left(1+\frac{S_{k+1}''^2}{G_{k+1}^2}\right)}} + \frac{1}{\sqrt{\left(1+\frac{G_1^2}{S_1''^2}\right)\left(1+\frac{G_{k+1}^2}{S_{k+1}''^2}\right)}}\mathbf{R}_k'' \quad \text{.....(lii)}$$
$$= p_k + l_k \mathbf{R}_k'',$$

which must be compared with the relation

$$\mathbf{R}_k' = p + q \,.\, r^k \quad \text{.....(xlviii) bis,}$$

where $p = \cdot 1524, \quad q = \cdot 6817, \quad r = \cdot 7105,$

that has been found empirically to fit the actual values of $\mathbf{R}_k'$.

If the expressions p_k and l_k were constant for $k = 1, 2 \ldots 14$ an interpretation of (lii) would be at once suggested. Namely that $\mathbf{R}_k''$, the coefficient of correlation of the successive residuals Y_t and Y_{t+k} left after the removal of the secular and sessional changes is expressible in the form

$$\mathbf{R}_k'' = q' r^k \text{.....(liii),}$$

that is to say, making allowance for the presence of accidental errors, the law of relationship between the successive estimates suggested on p. 90 above, holds good. Now without finding the curve which represents the sessional change in each series we do not know the values of S_k'' and G_k. We have however that

$$S_k''^2 + G_k^2 = S_k'^2 \text{.....(liv),}$$

where S_k' is the standard deviation of the observations in the kth groups after the removal of the secular term. The values of S_k' are given in Table V, p. 58; they are seen to increase as k increases and therefore p_k and l_k can only be constant for all values of k if

$$\frac{S_1''^2}{G_1^2} = \frac{S_2''^2}{G_2^2} = \ldots = \frac{S_{14}''^2}{G_{14}^2} \text{.....(lv).}$$

That the relations (lv) should hold approximately is not at all improbable; for with a sessional change of the parabolic form of the curve (xxx) illustrated in Figure 4, the standard deviations of the ordinates in the later groups will increase owing to the increasing drop of the curve towards the end of the series while S_k'' may increase with k owing to greater variation towards the end of a session.

In fact for this particular mean series with its sessional curve represented by (xxx) it is found that

$$G_1 = \cdot 0336 \text{ ins.}, \quad G_{14} = \cdot 0406 \text{ ins.},$$

while $$S_1'' = \cdot 0165 \text{ ins.}, \quad S_{14}'' = \cdot 0201 \text{ ins.},$$

that is to say, the variations superimposed upon the main sessional change (the distances of the points plotted in Figure 4 from the parabola) become greater towards the end of the series when the observer's judgment perhaps became more erratic as he grew tired. These values give $\frac{S_1''}{G_1} = \cdot 49$, $\frac{S_{14}''}{G_{14}} = \cdot 50$ suggesting that the relations (lv) do hold very closely. What we find therefore in this typical mean

series represented by Figure 4 may well be expected to hold approximately in the individual series.

If then p_k is constant for $k = 1, 2, \dots 14$ and equals p, we find readily from equations (lii) and (lv) that

$$l_k = 1 - p_k = 1 - p,$$

and hence $$(1-p)\, R_k'' = qr^k \quad \text{or} \quad \mathbf{R}_k'' = \frac{q}{1-p}\, r^k.$$

Making use of the numerical values $p = \cdot 1524$, $q = \cdot 6817$, $r = \cdot 7105$ we obtain finally

$$\mathbf{R}_k'' = \cdot 8043\, (\cdot 7105)^k \quad \text{......(lvi)},$$

as the theoretical expression for the correlation of the successive residuals after the observations have been freed from secular and sessional change. This curve is the lower of the two curves drawn in Figure 8. The points which are there plotted about this curve are the points $(k, \mathbf{R}_k'')$* obtained from equation (lii)

(*a*) On the assumption that $G_1^2 = G_2^2 = \dots = G_{14}^2 =$ constant,

(*b*) $$\frac{1}{\sqrt{\left(1 + \frac{S_1''^2}{G_1^2}\right)\left(1 + \frac{S_{14}''^2}{G_{14}^2}\right)}} = p_{14} = \cdot 1524,$$

(*c*) Making use of equations (liv) and the tabled values of S_k'.

The close fit of the curve to these points shows that the manner in which the values of $\mathbf{R}_k''$ fall off as k increases is not much affected by the different assumptions regarding the relations of the S_k'''s and the G_k's made in the two cases †.

Experiment B.

Reference has been made on p. 71 to evidence for a slight periodicity in the observations of this Experiment, which gives rise to small but apparently significant negative values to $\mathbf{R}_k'$, for $k \geqslant 7$. Further investigation might enable a correction for this periodicity to be made, but at present it is not possible to express $\mathbf{R}_k''$ with exactness in the form

$$\mathbf{R}_k' = qr^k.$$

For the purpose of comparison with the other experiments we can however obtain values of q and r which will give a rough fit for the first few values of $\mathbf{R}_k'$. Thus if we take

$$r = \cdot 72, \quad q = \cdot 47 \ddagger,$$

we get the values

$$\mathbf{R}_1' = \cdot 34, \quad \mathbf{R}_2' = \cdot 24, \quad \mathbf{R}_3' = \cdot 18, \quad \mathbf{R}_4' = \cdot 13,$$

which agree roughly with the actual values given in Table XI, namely

$$\mathbf{R}_1' = \cdot 352, \quad \mathbf{R}_2' = \cdot 231, \quad \mathbf{R}_3' = \cdot 183, \quad \mathbf{R}_4' = \cdot 085.$$

* In Figure 8 these points have been indicated by $\mathbf{R}_k'''$ to distinguish them from the correlation coefficients of residuals after removal of linear sessional change, there denoted by $\mathbf{R}_k''$.

† The values of R_k'' calculated from equation (lvi) and of R_k'' and S_k'' calculated on the assumption of the constancy of G_k are given in the 8th, 7th and 6th columns of Table XVIII.

‡ $r = \cdot 72$ is the value of the mean of the ratios $\frac{R_2'}{R_1'}$, $\frac{R_3'}{R_2'}$ and $\frac{R_4'}{R_3'}$, and using this value for r, q was taken as $\cdot 47$ by rough trial.

Experiment C.

At the end of the section dealing with the reduction of the observations for this experiment the conclusion reached was that $\mathbf{R}_{12}'$ and $\mathbf{R}_{13}'$ were not significantly negative; no difficulty therefore arises in fitting a curve of the form $y = qr^k$ to the values of $\mathbf{R}_k'$ given in the 6th row of Table XIV, p. 80. This was effected by the method of least squares, with the result

$$\mathbf{R}_k' = \cdot 6673 \times (\cdot 7917)^k \quad \text{.....(lvii).}$$

In the 7th row of Table XIV are given the values of $\mathbf{R}_k'$ calculated from this equation, and in the 8th row the differences

$$(\mathbf{R}_k' \text{ from observations}) - (\mathbf{R}_k' \text{ from curve}).$$

If these differences are compared with the probable errors of $\mathbf{R}_k'$, it will be seen that the fit is very satisfactory, for the later calculated values of $\mathbf{R}_k'$ are in any case uncertain; $\mathbf{R}_{12}'$ and $\mathbf{R}_{13}'$ were indeed not used in the least square solution as they were known to have too high negative values.

Experiment D.

On p. 85 it was suggested that $\mathbf{R}_k$ would approach the value $+ \cdot 354$ as k increased. In this case a curve of the form

$$\mathbf{R}_k = \cdot 354 + qr^k,$$

was fitted to the calculated values of $\mathbf{R}_k$. The fitting was carried out by moments. Making $\mathbf{R}_k - \cdot 354 = z$, we have

$$\left.\begin{aligned} \sum_1^s (z) &= qr \frac{1 - r^s}{1 - r} = N, \text{ say, where } s \text{ is the number of ordinates, or } 12 \\ \sum_1^s (zk) &= q(r + 2r^2 + 3r^3 + \ldots + sr^s) = N \times \mu_1' \end{aligned}\right\},$$

whence

$$\mu_1' = \frac{1}{1 - r} - \frac{sr^s}{1 - r^s} \quad \text{.....(lviii),}$$

and is the distance of the mean from an origin at unit distance from the first ordinate qr,

$$N = qr \frac{1 - r^s}{1 - r} \quad \text{.....(lix).}$$

The constants μ_1' and N are known; solving (lviii) by approximation we have r, and then (lix) gives q.

The values are

$$\left.\begin{aligned} q &= \cdot 1153 \\ r &= \cdot 8121 \end{aligned}\right\},$$

and finally,

$$\mathbf{R}_k = \cdot 354 + \cdot 1153\,(\cdot 8121)^k \quad \text{.....(lx).}$$

Then using the approximate relation

$$\mathbf{R}_k' = (\mathbf{R}_k - \cdot 354) \times \frac{S_1'^2 + \frac{1}{m} \sum_m (D_1 - d_1)^2}{S_1'^2}$$

—which is a modified form of Equation (xliii)—we obtain for $\mathbf{R}_k'$ the equation

$$\mathbf{R}_k' = \cdot 1785\,(\cdot 8121)^k \quad \text{.....(lxi).}$$

Both of the curves, represented by equation (lx) and (lxi), have been drawn in Figure 20, and show a satisfactory fit, if the roughness of the data is taken into account.

The results of the Trisection and of the Ten-second Counting Experiments, and as far as the rough form of the data will allow, of the Ten-second Estimating Experiment, suggest therefore that there is some foundation for the theory of relationship between successive estimates put forward at the beginning of the present Section. To reach the expression qr^k for the correlation of successive judgments at intervals of k, it has been necessary in all cases to remove the secular change, and in one case a sessional change as well, but if these changes correspond in themselves to some definite mental or physical processes which can be separated in some degree from the causes underlying the residual variations, then we are justified in inquiring into the significance of the constants q and r. It has been suggested that

$$q = \frac{\bar{\alpha}^2}{\bar{\alpha}^2 + \bar{\beta}^2} \quad \text{.....(lxii),}$$

so that q is dependent on the ratio between the correlated and the uncorrelated parts of the observer's judgment, that is between what I have considered as the true estimate and the accidental errors superimposed in the process of record. Now using (lxii) and the relation*

$$\sqrt{\bar{\alpha}^2 + \bar{\beta}^2} = S' \quad \text{.....(lxiii),}$$

(or S'' for the Trisections where it has been necessary to allow for a sessional change), we find that

$$\sqrt{\bar{\alpha}^2} = \sqrt{q}\, S', \quad \sqrt{\bar{\beta}^2} = \sqrt{(1-q)}\, S' \quad \text{.....(lxiv),}$$

and the values calculated in this way for $\sqrt{\bar{\alpha}^2}$ and $\sqrt{\bar{\beta}^2}$ are given in Table XIX.

TABLE XIX.

Experiment	q	S'	$\sqrt{\bar{\alpha}^2}$	$\sqrt{\bar{\beta}^2}$	r
Trisection	·80	·080 (=S'') in inches	·071	·036	·71
Bisection (approximate only) ...	·47	·045 in inches	·031	·033	·72
Ten-second Counting	·67	·034 in factor	·028	·020	·79
Ten-second Estimating ...	·18	·141 in factor	·060	·128	·81

If the Trisection and Bisection results are compared it will be seen that the standard deviations of the accidental errors ($\sqrt{\bar{\beta}^2}$) are nearly the same but that there is a large difference between the measures of the variations of the true

* It will be seen that owing to a sessional change in standard deviation, S_k'' for the Trisections (Table XVIII) and S_k' for the Bisections (Table XI) increase with k. To obtain an approximate value for the standard deviation of the whole 1200 observations as opposed to that for the 1000 observations of any particular Group k, I have used in equations (lxiii) and (lxiv) S' (or S'') given by

$$S'^2 = \tfrac{1}{14}\,(S_1'^2 + S_2'^2 + S_3'^2 + \ldots + S_{14}'^2).$$

estimates ($\sqrt{\bar{\alpha}^2}$). This is a result which we should anticipate, for the method of recording the estimate was the same in each experiment, and accidental errors of the same magnitude would occur in both cases; on the other hand the observer was faced with a more difficult problem in estimating a third than in estimating a half, and this is shown by the greater variability of his estimate in the former case (·07 against ·03)*. For the Timing experiments, we find no correspondence between the $\sqrt{\bar{\beta}^2}$'s; the great difference between the counting of ten seconds and the attempted concentration of mind on the passing of an unbroken ten second interval has been emphasized in the description of the experiments above, and a correspondence was hardly to be expected. The standard deviations are in terms of the factors e/p and must be multiplied by 10·2 if required in seconds.

If now we turn to the values of r given in the last column of Table XIX, it will be seen that they lie near together, and although that for the Bisections is not an exact measure, there is a suggestion of close agreement between the r's in the pairs of similar experiments, for we have estimations of length with ·71 and ·72, and estimations of time with ·79 and ·81. This coefficient is a measure of the rate at which the correlation of successive judgments falls off or the influence of previous estimates vanishes from the observer's mind: on the theory of zero partial correlation it is simply the coefficient of correlation between a true estimate freed from accidental errors and the preceding estimate.

On any theory r would seem to be a fundamental constant not varying greatly for different types of observations, but perhaps varying considerably for different observers. The fact that it is so nearly the same for experiments with a five second interval between observations (Trisection and Bisection) and for others with an interval of ten seconds or more (Counting and Estimating) shows that the correlation of successive judgments is a function not only of the *time interval* between two judgments but also of the *number of intervening judgments*. For if it were purely a function of the time interval we should expect to find a greater difference between the values of r found for experiments with a five second interval and a ten second interval. Indeed if the experiments were exactly the same but for difference in interval, $\mathbf{R}_1'$ for that with ten seconds would equal $\mathbf{R}_2'$ for that with five seconds. Further experiments of the same type in which the interval between the recording of judgments was varied would undoubtedly throw much light on this point.

XII. Prediction.

If the values of "m" successive judgments are known and there is no correlation between them, the "most probable" value of the $(m+1)$th judgment, that is the most reasonable guess at its value that can be made, is the mean of the "m" judgments. If however the successive judgments are correlated, then it is possible to predict the value of the $(m+1)$th with much greater expectation of accuracy.

* This may be compared with the ratio of 3 to 2 given on p. 73 from a comparison of the S_δ's before making any allowance for the accidental errors.

In the Experiments *B*, *C* and *D* it has been found that the correlation between judgments at intervals of k, made in the same session, can be expressed approximately in the form

$$\mathbf{R}_k' = qr^k \quad \text{.....(lxv),}$$

while for Experiment *A*, owing to the large sessional change, the expression was

$$\mathbf{R}_k' = p + qr^k \quad \text{.....(lxvi).}$$

The decrease of correlation in geometrical progression expressed by (lxv) follows precisely the law of ancestral heredity, for which the multiple regression equations required for prediction have already been worked out*. It is not therefore proposed to go further into the problem in the present Paper, nor to inquire whether the general multiple regression equations would reduce to as simple a form when the correlation is expressed by equation (lxvi) rather than (lxv).

XIII. SUMMARY AND CONCLUSIONS.

The secular change in personal equation is shown by the variation in the series' means, but it is only in Experiment *A* and perhaps Experiment *C*, where the general trend of the variations is markedly in one direction, that we find that type of change which is usually understood when a secular change is referred to. In the Bisection Experiment *B* the linear secular change is very small and its existence might well not be recognized, and yet the series' means are subject to fluctuations far exceeding those of random sampling. For the probable error of the mean of a series (or of the observations in Group 1) is

$$\pm \cdot 67449 \times \frac{S_1'}{\sqrt{50}} = \pm \cdot 00416,$$

but if we take the distribution consisting of the 20 series means, d_1, we find that the standard deviation is ·037375, giving for the probable error of a mean d_1

$$\pm \cdot 02521,$$

which is more than six times as large as the probable error we have calculated by considering the variations within a series. It is therefore clear that the 50 observations in a series are not random samples of the whole "universe" of observations, as they should be on the Gaussian hypothesis of normal errors.

It is again only in Experiment *A* that there is a fairly consistent sessional change from series to series which an observer might easily recognize and possibly allow for, and yet if we turn to any of the graphs for the Bisection or Seconds-counting which show the variations of judgment within a series (Figures 11 and 15), it will be seen how very often the mean of ten consecutive judgments will give but a poor approximation to the mean of the series; we cannot take the judgments within one series as scattered at random. When dealing with a sample of m

* The Galton-Pearson Law of Ancestral Heredity; the offspring and the mean of the kth grandparents have qr^k for their correlation.

correlated variates, the usual expression for the probable error of the mean is (1) $\pm$ ·67449 $\frac{(1-r^2)}{\sqrt{m}}\sigma_m$, as compared with (2) $\pm$ ·67449 $\frac{\sigma_m}{\sqrt{m}}$, when the variates are not correlated, but owing to the sessional variations to which a large part of the correlation is due, the expression (1) being the smaller, is in the present case a worse measure than (2), of the probable limits of divergence of the mean of the sample from the mean of the series. The graphs of Figures 6, 11 and 15 show that there is a tendency for the judgments to vary in waves, to be first on one side of the mean for the series, and then to change to the other, but with no definite period of variation. It is owing to these large correlated variations which cannot be expressed in any simple sessional term, that the coefficients of correlation, r_{ρ_1,σ_1}, between σ_1 and ρ_1 have been found to have positive values ranging from + ·52 $\pm$ ·11 in Experiment *A* to + ·18 $\pm$ ·15 in *D*, showing that greater variation is associated with higher correlation of successive judgments.

An analysis has suggested that the coefficients of correlation of the crude values of the observations at intervals of k can be expressed in the generalized form

$$\mathbf{R}_k = \frac{S_1''S_{k+1}''\mathbf{R}_k'' + F_k + \frac{1}{m}\sum_m (D_1-d_1)(D_{k+1}-d_{k+1})}{\sqrt{\left\{S_1''^2 + G_1^2 + \frac{1}{m}\sum_m (D_1-d_1)^2\right\}\left\{S_{k+1}''^2 + G_{k+1}^2 + \frac{1}{m}\sum_m (D_{k+1}-d_{k+1})^2\right\}}} \quad \text{.....(lxvii),}$$

where

$\sum_m (D_1-d_1)(D_{k+1}-d_{k+1})$, $\sum_m (D_k-d_k)^2$ etc. are terms representing the secular change,

F_k and G_k are functions of the sessional change, and

$\mathbf{R}_k''$ and S_k'' are the correlation coefficients and standard deviations of the residuals left after secular and sessional changes have been removed.

In two experiments it has been found that $\mathbf{R}_1$ is greater than + ·80, which shows clearly that the estimates have not been distributed randomly in time.

The coefficients $\mathbf{R}_k''$ appear to fall off in geometrical progression, and to be closely represented by expressions of the form qr^k, in which q and r are constant for any experiment; it has been found that the introduction of the quantities F and G in equation (lxvii) in addition to the secular terms, is only necessary if there is a significant sessional change which repeats itself in series after series. Thus in Experiment *C*, where there was no such change, $\mathbf{R}_k$ could be expressed by the relation

$$\mathbf{R}_k = \frac{qr^kS_1'S'_{k+1} + \frac{1}{m}\sum_m (D_1-d_1)(D_{k+1}-d_{k+1})}{\sqrt{\left\{S_1'^2 + \frac{1}{m}\sum_m (D_1-d_1)^2\right\}\left\{S'^2_{k+1} + \frac{1}{m}\sum_m (D_{k+1}-d_{k+1})^2\right\}}} \quad \text{...(lxviii).}$$

A tentative interpretation has been given to the results of this analysis. The observations in Experiment *A* suggested that there was some physiological significance in the distinction between the secular and sessional changes, and this was

confirmed in Experiment *B*, where it was found that there was evidence of a linear sessional change acting in the opposite direction to the secular change. A discussion of the values of the partial correlation coefficients $r_{xy \cdot z}$ (personal equation and order, time constant) and $r_{xz \cdot y}$ (personal equation and time, order constant) suggested that if the interval between the successive series were made very short, it might not be sufficient to break the effect of the sessional change. The correlated variations which have been found to follow the law $\mathbf{R}_k' = qr^k$, have been considered as in some way separate from and superimposed upon the other more steady changes. Starting from the tentative assumption that there is little or no *partial* correlation between the observer's true estimates at intervals greater than one—that is to say that the observer's judgment at any moment is only influenced by the judgment immediately preceding, and only through this and not directly by the earlier judgments—it has been shown that the constant q in the relation

$$\mathbf{R}_k' = qr^k \quad \text{.....(lxv) bis}$$

can be accounted for by the presence of uncorrelated accidental errors which are superimposed on the correlated variations in the observer's true estimate. Without further investigation it would be difficult to distinguish between what may perhaps be termed the physiological and the psychological factors; in the experiments that have been undertaken the variations in recorded judgment depend partly on the movements of the hand, so that the former factors are likely to have played some part as well as the latter. The successive recording motions of the hand may have been correlated as well as the variations in mental estimate.

The importance of the results of course depends on how far they may be considered as typical of any practical series of observations made by the astronomer or the physicist. Experiments were admittedly chosen in which it was expected that the variations in judgment would be large, and for the experienced observer working at the type of observation in which he has had much practice, the errors would no doubt be smaller, but it seems to me likely that the phenomena which have been discussed will be present in the judgments of other observers even if on a smaller scale. Experience and accuracy may be gained by practice, but it does not follow that the correlation between successive judgments will disappear. The secular and sessional changes may be small, but if rough comparisons of only the yearly mean personal equations of different observers are made, the finer changes, which may be of considerable importance in a combination of observations, cannot be recognized. The Law of Normal Errors requires but two constants to describe adequately any series of observations:

(1) the mean,

(2) the standard-deviation,

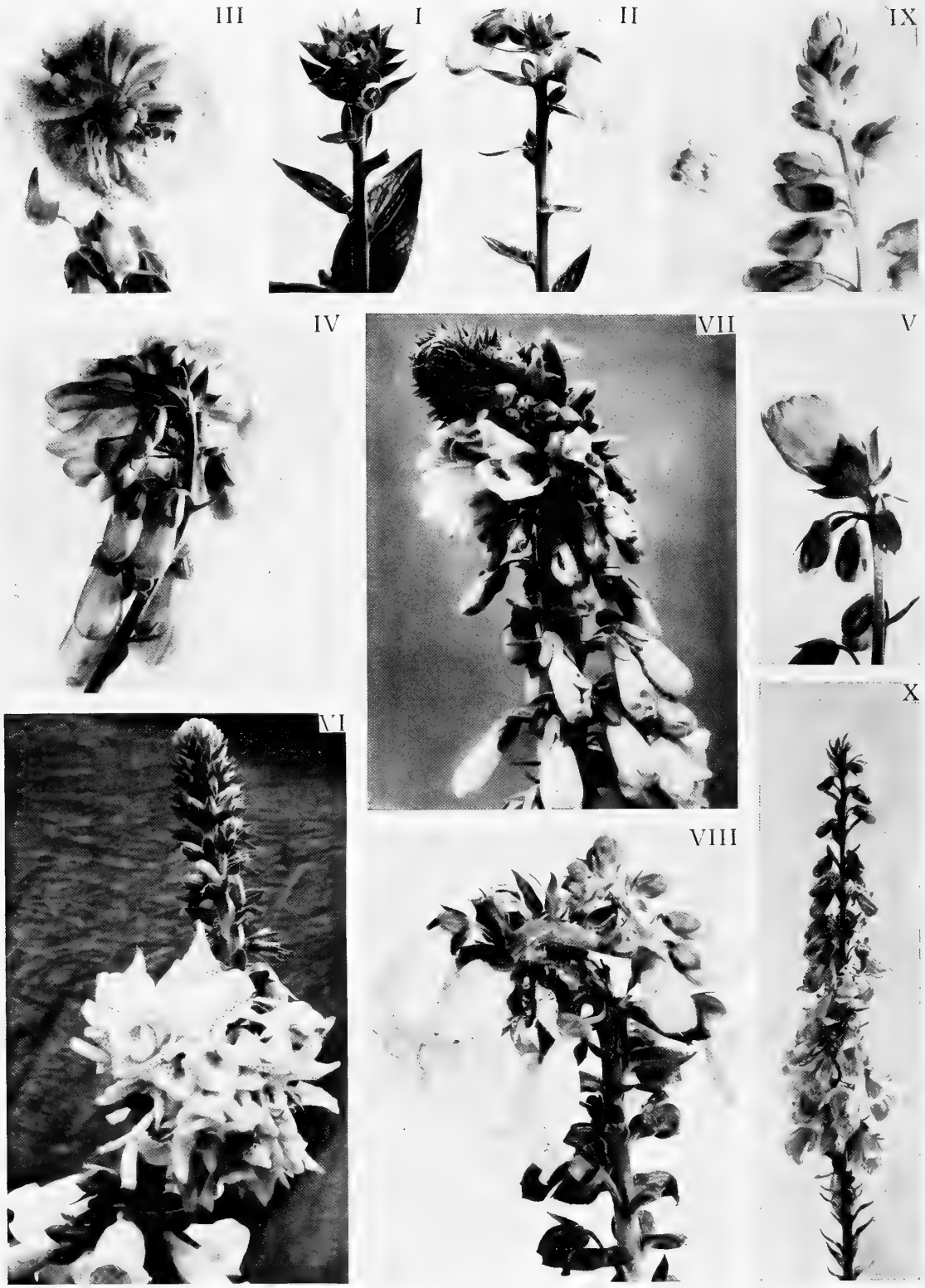
while the introduction of a third may be necessary if a gradual secular change in personal equation is noticed. But the more generalized Theory of Errors discussed in the preceding sections requires more detailed information and a greater number of constants to define the character of an observer's personal equation and variations in judgment. We shall require to know how the personal equation and the standard

deviation vary both within a session and over long periods of time, and if there is any correlation between successive judgments, what is the form of the function ψ, which gives the value of the successive correlation coefficients in the relation

$$\mathbf{R}_k' = \psi\,(k).$$

It is only by a detailed analysis of the observations themselves or of others carried out *ad hoc*, copying them as closely as possible, that full information on these points can be obtained; but if the possible complexities which may be present in the variations of judgment are fully realised, a great deal may be done in practical cases by the arrangement of the observations and the combination of the results, to eliminate the factors whose magnitude is unknown and to correct for others which are more easy to ascertain.

I have heartily to thank Miss I. McLearn for making the diagrams for Figs. 3, 4, 8, 12, 17, 19 and 20, and Miss M. Noel Karn for assistance in some of the computation.



Figs. I—IX. Pelorism of various intensities. Fig. X. Split corolla.

INHERITANCE IN THE FOXGLOVE, AND THE RESULT OF SELECTIVE BREEDING.

By ERNEST WARREN, D.Sc. Lond.

In *Biometrika*, Vol. XI. pp. 302—327, 1917, the author published a preliminary report on the earlier results obtained in the breeding of foxgloves; and the present paper contains some account of the final results of the selection experiments.

In 1914 ten foxglove plants (*Digitalis gloxiniaeflora*), obtained from various sources and of different characteristics, were crossed among themselves and also self-fertilised. In subsequent years, 1915—19, new generations were obtained chiefly by the self-fertilisation of selected parents. The measurement, or when not possible the grading, of certain characters (pelorism, colour, size of flower, spotting of flower, etc.) was undertaken in all the generations in order to determine the effect of selection when selfing alone occurred in an apparently pure race.

1. Pelorism.

Mendelian inheritance occurred in a typical fashion. A peloric plant crossed with a non-peloric plant produced non-peloric offspring. On selfing these, or crossing them together, there resulted on the average one peloric to three non-pelorics.

Of the 10 parent plants two exhibited the peloric condition in a fully developed form, and the rest were non-peloric. The character was very perfectly recessive, and by breeding, it was found that three of the remaining plants were really heterozygous, while all the others were non-peloric and homozygous.

It was soon observed that the peloric condition was by no means a clearly defined and fixed character. Pelorism in the foxglove may be regarded as an abnormal lack of power to produce internodes between the flower-buds, and consequently there may result considerable fusion of such buds with one another.

The maximum stage of pelorism is seen when the main-axis is short and abruptly ceases to grow in height. Only two or three normal flowers may be produced by the axis, and its blunt, sharply truncated end is surrounded by a whorl of bracts or sepals, petals being absent. Sometimes a ring of sessile anthers occurs (Pl. I, figs. I, II).

In typical pelorism the inability to produce internodes affects the terminal portions of all of the flower-axes of a plant, both central and side-axes. A variable number of flower-buds fuse and the corollas unite and may form a large symmetrical cup or saucer of some ornamental value, but the sepals mostly remain

separate (figs. III, IV). When numerous flower-buds fuse a dense rosette may be formed by the petals, and the result is not pleasing. The peloric or crown-flower opens early, often before any of the normal flowers. After the crown-flower has faded, the main-axis usually grows through the centre of it, and may even produce a second crown-flower (fig. VI); but in the case of the side-shoots the axis generally ends in an ovary and no further growth occurs (fig. V).

If the peloric tendency is not so well-marked, the main-axis may be only slightly affected by the suppression of several internodes, and by the partial fusion of flower-buds, at a variable distance above the lowest normal flower of the axis. Sometimes a considerable number of internodes may be unduly shortened, so as to produce excessive crowding of flowers which do not actually fuse (fig. VII), and frequently a strongly marked spiral bending of the axis occurs (fig. VIII).

At other times the suppression of the internodes may occur only high up on the flowering axis close to where it normally ceases to grow (fig. IX).

When the central axis is strongly peloric the side-axes are invariably so, and in all other cases the side-axes exhibit greater pelorism than the main-axis.

Finally, the main-axis may be quite normal and show no peloric tendency, but the side-axes may still be strongly peloric.

The last trace of pelorism in a plant is shown when only one or two of the weaker side-axes exhibit some slight sign of a peloric tendency.

It is unfortunate that it has not been found possible to devise any practical method of measuring the intensity of pelorism, and therefore the plants have been arranged in four grades.

0° grade = no peloric tendency.

1°— 25° grade = those in which the central axis is non-peloric, but the side-axes exhibit some peloric tendency.

26°— 50° grade = main-axis non-peloric, but side-axes may reach full pelorism.

51°— 75° grade = main-axis partially peloric, side-axes fully so.

76°—100° grade = plants ranging to complete pelorism in all axes.

In the generations produced from 1914—19 there were in all 128 fertilisations of different classes of individuals, recessive (peloric), homozygous dominant (non-peloric) and heterozygous dominant (non-peloric) plants, and families were raised. In the table on p. 105 the experimental and theoretical results are compared. The fertilisations of the classes $DD \times DD$, $RR \times RR$, and $DR \times DR$ include both selfing and crossing. The sum totals of the experimental and theoretical results are remarkably close; being, crowned, 1019 experimental and 1013 theoretical; non-crowned, 1169 experimental and 1175 theoretical.

It must be noted here that a plant was recorded as "peloric" or "crowned" if it exhibited the least tendency towards pelorism in any of the axes. Taking all the classes or groups together it may be said that the inheritance of the quality of pelorism is typically Mendelian. The group $RR \times RR$ should include no non-crowned offspring, and the 7 which occurred were obtained by gradual selection.

The group in which the experimental result diverged the most widely from the theoretical result was $DR \times RR$ (heterozygous plants crossed with recessives) and it would be interesting to know whether such is generally the case in Mendelian inheritance.

Gametic Nature of Pairings	Number of Families	Number of Offspring	Number of *Crowned* Offspring		Number of *Non-Crowned* Offspring	
			Experimental	Theoretical	Experimental	Theoretical
$DD \times DD$	16	266	0	0	266	266
$RR \times RR$	43	741	734	741	7	0
$DR \times DR$	38	777	187	194	590	583
$DR \times DD$	5	93	0	0	93	93
$DR \times RR$	12	156	98	78	58	78
$DD \times RR$	14	155	0	0	155	155
Totals	128	2188	1019	1013	1169	1175

The Inheritance of the Degree or Intensity of Pelorism.

If a peloric plant be crossed with a non-peloric homozygous dominant, the offspring are heterozygous and non-peloric, and if these are self-fertilised or crossed together the peloric character re-appears in an apparently unchanged and undiluted condition. If, on the other hand, a strongly peloric plant is crossed with a weakly peloric one the offspring are more or less intermediate, and if the offspring are selfed or fertilised together the intermediate nature of the peloric character tends to be retained.

In the accompanying table *A*, *B*, *C*, *D*, *E* are plants of various gametic constitution. On selfing (*A*) the offspring were all fully peloric. On selfing some 5 offspring, *A*, 2—9, the plants produced were all essentially fully crowned.

On crossing two recessive plants (*A* and *E*) of different peloric intensities (see bottom of table) the offspring tended to be intermediate.

On crossing (*A*) with an ordinary plant (*B*) the offspring were non-peloric and heterozygous. On selfing two of these plants, ($A \times B$) pls. 2 and 7, the offspring were either fully peloric, or non-peloric (heterozygous and homozygous). On selfing two recessives, ($A \times B$) 2, pls. 8 and 9, obtained from ($A \times B$) pl. 2, the offspring were all nearly completely peloric. Thus, there was no clearly marked dilution or apparent contamination by crossing a peloric plant with a non-peloric one. When, however, the same recessive plant (*A*) was crossed with a heterozygous plant (*C*) having in its gametes a weak peloric tendency of about 35° there was much variation in the offspring, and on selfing some of these plants, ($A \times C$) 1, 2, 7, 11, and raising a new generation it was obvious that considerable dilution of the peloric tendency had occurred. On crossing the same plant (*A*) with a heterozygous plant (*D*) having a stronger peloric tendency (75°) in its gametes it was clear that in the next generation raised ($A \times D$) 6, 5, 11 less dilution had taken place than in the former case.

Pelorism—Various Pairings.

Parentage	Peloric Offspring				Non peloric	Offspring (selfed)	Peloric Offspring				Non-peloric
	100°	75°	50°	25°			100°	75°	50°	25°	
A (100° pelorism) Selfed = *RR* × *RR*	33	0	0	0	0	*A* pl. 2 (100° pelorism)	13	1	0	0	0
						A pl. 3 „	3	0	0	0	0
						A pl. 4 „	2	0	0	0	0
						A pl. 6 „	6	0	0	0	0
						A pl. 9 „	1	0	0	0	0
A ♀ (100° pelorism) × *B* ♂ (0° pelorism and homozygous) *A* × *B* = *RR* × *DD*	0	0	0	0	13	(*A* × *B*) pl. 2 (non-peloric and heterozygous)	6	0	0	0	21
						(*A* × *B*) pl. 7 (non-peloric and heterozygous)	5	0	0	0	23
						(*A* × *B*) 2 pl. 8 (100° pelorism)	12	1	0	0	0
						(*A* × *B*) 2 pl. 9 „	5	0	0	0	0
C (heterozygous) Selfed = *DR* × *DR*	1	0	2	3	13						
A ♀ (100° pelorism) × *C* ♂ (non-crowned and heterozygous with, say, 35° pelorism in gamete) *A* × *C* = *RR* × *DR*	4	12	3	1	7	(*A* × *C*) pl. 1 (75° pelorism)	20	4	9	2	0
						(*A* × *C*) pl. 2 (50° pelorism)	6	7	1	1	0
						(*A* × *C*) pl. 7 (heterozygous)	—	—	2	0	10
						(*A* × *C*) pl. 11 „	—	2	3	1	7
D (heterozygous) selfed	1	1	1	0	8						
A ♀ (100° pelorism) × *D* ♂ (non-crowned and heterozygous with, say, 75° pelorism in gamete)	11	2	0	0	10	(*A* × *D*) pl. 6 (100° pelorism)	17	2	2	0	0
						(*A* × *D*) pl. 5 „	21	0	6	0	0
						(*A* × *D*) pl. 11 (75° pelorism)	27	5	1	0	0
A ♀ (100°) × *E* ♂ (50°)	4	1	3	0	0						

In the last generation it will be seen that there was no sharp separation of the plants into two groups attributable to the two grandparental factors. Thus, in the case of (*A* × *C*) pl. 2 (50°) the offspring are not clearly divisible into those of 100° resembling *A*, and those of 35° attributable to *C*; in other words there was no obvious segregation into two degrees of pelorism.

On the factorial and chromosome hypotheses we must suppose that the factor or factors governing the peloric character tend to become mutually changed and intermediate in nature when the male and female chromosomes containing the factors for the two degrees of pelorism lie alongside each other in the zygote.

It will be of interest to obtain a general measure of the strength of inheritance between mid-parent and offspring with respect to the transmission of the degree or intensity of pelorism. For this purpose only recessives were used, involving

30 mid-parents. Employing Prof. Karl Pearson's method the accompanying table gives the correlation surface.

Pelorism—Correlation Table—Recessives. Mid-parent and Offspring.

Offspring. Grade of Pelorism.

Mid-parents. Grade of Pelorism	76°—100°	51°— 75°	26°— 50°	1°— 25°	Totals
1°— 25°	6	2	23	18	49
26°— 50°	58	61	68	11	198
51°— 75°	64	31	15	—	110
76°—100°	143	14	11	5	173
Totals	271	108	117	34	530

The coefficient of correlation, calculated from the table, between mid-parent and offspring is ·52. The result can be regarded as only a very rough approximation, since a satisfactory method of measuring pelorism has yet to be found. The figure obtained is somewhat low, but it would seem to indicate that the inheritance of the degree of pelorism is of the nature of ordinary blended inheritance.

The point of interest to notice is that the union of two peloric plants of different peloric intensities influences the gametes, while the union of a peloric plant with a homozygous non-peloric plant does not very readily affect the purity of the gametes with respect to pelorism.

Pelorism. Effect of Selection in a homogeneous race.

A peloric plant (*C*) with pelorism of about 85° intensity was self-fertilised, and the offspring, 16 in number, were as follows: 7 with 100°, 4 with 75° and 5 with 50° of pelorism.

Parentage (Self-fertilisation)	Crowned Offspring				Not crowned	Parentage (Self-fertilisation)	Crowned Offspring				Not crowned
	100°	75°	50°	25°			100°	75°	50°	25°	
C (85°)	7	4	5	0	0						
C 2 (50°)	7	7	10	0	0	*C* 2, 11 (75°)	6	13	5	0	0
C 7 (50°)	17	11	18	2	0	*C* 2, 2 (50°)	2	16	10	0	0
C 7, 10 (25°) ...	0	2	18	2	5	*C* 2, 8 (50°)	1	13	11	0	0
C 7, 10, 20 (25°)	0	0	1	2	7	*C* 7, 10, 20, 4 (0°)	0	0	0	0	6

Two of these plants of 50° (*C* 2 and 7) were selfed, and the generation raised exhibited a lowered pelorism. The various selections made and the results obtained are shown in the accompanying table. It will be seen that finally on the selfing of plant *C* 7, 10, 20, 4 (0°) only non-peloric offspring were obtained.

2. General Coloration of the Corolla.

As described in the previous report (*loc. cit.*) the intensity of the purple coloration was measured by comparing it with a colour-scale founded on the intensity of colour by transmitted light of varying depths of a standard colour-solution.

Purple and white foxgloves exhibit the ordinary Mendelian relationship, purple being dominant. A confusing aspect of the problem is introduced by the fact that "white" foxgloves are not necessarily entirely white, since they may exhibit a faint purple coloration which on the colour-scale adopted may amount to about 5. On crossing such a plant with an ordinary purple plant segregation occurs when the heterozygous offspring are self-fertilised. Any higher coloration, say 10—15, does not exhibit segregation, but gives a blended inheritance, and such a plant is to be regarded as a very pale purple one and not "white." From certain observations that have been made it is probable that a similar condition occurs in the Blue Agapanthus lily, since some of the "white" plants have flowers faintly tinged with blue. It is quite likely that the phenomenon is general, and it may throw an important light on the physical theory of heredity. Possibly it may be surmised that a factor for a coloration of less than 5 units is unable to blend with, or influence, the factor controlling a higher coloration, in that we have reached the lowest dynamic unit.

Of the ten original plants, five were purple and homozygous, four were purple and heterozygous and one was white or recessive. These were very variously crossed in all manner of ways. In the accompanying table the experimental results are compared with the Mendelian expectation for the different gametic pairings.

General Coloration of Corolla—Breeding Results.

Gametic Nature of Pairings	Number of Families	Number of Offspring	White		Purple	
			Experimental	Expectation	Experimental	Expectation
$DD \times DD$	120	1620	2+3	0	1615	1620
$RR \times RR$	17	336	330	336	6	0
$DR \times DR$	50	785	190	196	595	589
$DR \times DD$	11	103	0	0	103	103
$DR \times RR$	8	76	24	38	52	38
$DD \times RR$	8	87	0	0	87	87
Totals	214	3007	549	570	2458	2437

In the gametic group $DD \times DD$ (homozygous purple × homozygous purple) there were 1620 offspring. These should have been all purple, but there were two white plants which occurred in two deeply coloured families and three white plants which occurred in one pale-coloured family. I do not believe that there was contamination, and it is probable that the two former plants were sports, while the three latter plants were produced by selection.

In the group $RR \times RR$ (white × white) there were 336 offspring, and these should have been all white, but there were six pale-coloured plants. The difficulty in distinguishing a tinged "white" plant from a pale-coloured plant may account for this result, but I favour the view that we are here witnessing the beginning of a coloured race.

The result given by $DR \times DR$ (heterozygous purple × heterozygous purple) is very closely Mendelian. Out of 785 offspring there were 190 white plants while the expectation was 196.

Heterozygous plants crossed with dominants ($DR \times DD$) gave nothing but coloured plants, and this was also the case with dominants crossed with recessives ($DD \times RR$).

The gametic group $DR \times RR$ (heterozygous plants × recessives) gave a result which diverged rather widely from the expectation: there were insufficient whites, there being 24 whites and 52 purples instead of 38 of each. The numbers are somewhat small for drawing conclusions, but it is important to notice that in the character of pelorism it was the same gametic group which diverged the most widely of all the classes from the theoretical expectation. On the chromosome hypothesis it may be conjectured that possibly preferential pairing of the male and female chromosomes may explain the discrepancy.

The Inheritance of the Intensity of Coloration.

On crossing a purple homozygous plant with a white plant the offspring were all heterozygous and all coloured, but the intensity of the coloration was mostly reduced very considerably. On selfing these offspring the next generation yielded some homozygous dominants in which the original colour-intensity of the grandparent was regained; thus, at first sight it appeared that there had been no real dilution of the colour by crossing with the white. This was my first impression from the earlier results, but with more extended experience I found that there was certain evidence that the crossing with the white did have some deleterious action on the intensity of the coloration of the dominant grandchildren, although the coloration which appeared was much greater than a half and half blend with white.

If two homozygous dominants of marked difference in colour-intensity were crossed, the offspring tended to be intermediate. On selfing these offspring the next generation was similarly intermediate, and there was no segregation into the two different intensities of the grandparents. Thus a true blend of the two intensities had taken place.

In the accompanying table the results of some instructive crossings and self-fertilisations are given. In Series I two dominants (*E* and *F*) of different colour-intensities were selfed and the families raised showed that the parents were homozygous. On crossing (*E*) and (*F*) a family of intermediate offspring was obtained.

Colour-Intensity—Various Pairings.

No. of Series	Gametic Constitution	Parentage	Mid-Parental Colour	Colour-Scale—Offspring										Mean of Coloured Offspring
				Coloured plants									"White"	
				125°—139°	110°—124°	95°—109°	80°—94°	65°—79°	50°—64°	35°—49°	20°—34°	5°—19°	0°—4°	
I	*DD* × *DD*	*E* (selfed)	90	—	2	0	5	8	—	—	—	—	—	83
	DD × *DD*	*F* (selfed)	68	—	—	—	—	6	3	—	—	—	—	64
	DD × *DD*	♀ *E* × ♂ *F*	79	—	—	—	4	3	5	—	—	—	—	71
	DD × *DD*	(*E* × *F*) pl. 16 (selfed) ...	82	—	—	2	6	2	—	—	—	—	—	87
	DD × *DD*	(*E* × *F*) pl. 9 (selfed) ...	61	—	—	—	—	1	10	1	—	—	—	56
II	*DD* × *RR*	♀ *E* × ♂ (White) ...	45	—	—	—	—	11	1	4	2	—	—	60
	DR × *DR*	*E* × (White) pl. 18 (selfed)	71	—	—	1	1	5	1	1	1	—	2	66
III	*DD* × *DD*	*B* (selfed)	95	3	0	2	3	6	1	—	—	—	—	102
	DD × *RR*	♀ *B* × ♂ (White) ...	47	—	—	—	1	4	4	1	1	—	—	59
	DR × *DR*	(*B* × White) pl. 1 (selfed)	80	—	—	—	—	1	2	—	—	—	—	64
	DR × *DR*	(*B* × White) pl. 5 (selfed)	32	—	—	—	—	—	—	2	3	—	5	31
IV	*DD* × *DD*	*B* (selfed)	95	3	0	2	3	6	1	—	—	—	—	102
	DR × *DR*	*A* (selfed)	70	—	—	—	4	12	4	1	—	—	7	70
	DD × *DR*	♀ *B* × ♂ *A*	82	2	0	1	4	4	2	—	—	—	—	105
	DD × *DD*	(*B* × *A*) pl. 7 (selfed) ...	130	1	1	3	12	13	—	—	—	—	—	87
	DD × *DD*	(*B* × *A*) pl. 2 (selfed) ...	65	—	—	—	2	13	12	—	—	—	—	67
V	*DD* × *DD*	*B* (selfed)	95	3	0	2	3	6	1	—	—	—	—	102
	DR × *DR*	*C* (selfed)	34	—	—	—	—	—	1	9	2	—	4	40
	DD × *DR*	♀ *B* × ♂ *C*	65	—	—	—	—	2	4	2	—	—	—	53
	DD × *DD*	(*B* × *C*) pl. 8 (selfed) ...	50	—	—	—	—	—	2	1	—	—	—	50
	DD × *DD*	pl. 4	50	—	—	—	—	4	17	3	—	—	—	59
	DD × *DD*	pl. 7	58	—	—	—	—	5	4	1	1	—	—	60
	DD × *DD*	pl. 6	68	—	—	—	—	6	14	9	—	—	—	54
	DD × *DD*	pl. 1	70	—	—	2	4	20	4	—	—	—	—	74

Two of these offspring were selected, (*E* × *F*) pls. 16 and 9, as widely divergent from each other as possible, and selfed. In the families obtained there was no tendency for the occurrence of segregation into the two colour-intensities of (*E*) and (*F*) respectively. There was thus a definite blend, and the means of the two families approached the respective colour-intensities of the two self-fertilised plants.

In Series II the same homozygous dominant plant (*E*), with colour-intensity of 90°, was crossed with a white plant and all the offspring were heterozygous and intermediate. On selfing one of the darker coloured offspring, no. 18, the dominant plants raised tended to be of about the same colour-intensity as the grandparent

(*E*). In Series III a dark-coloured homozygous dominant plant (*B*) was also crossed with a white plant. One of the darkest heterozygous offspring (*B* × White) pl. 1 was selfed and the coloured plants raised tended to be paler than the grandparent, but the family was small.

In Series IV the dark-coloured homozygous plant (*B*) was crossed with a dark heterozygous plant (*A*). From the offspring raised, two were selected and selfed, one very dark and the other moderately dark. The two families included only coloured plants, and consequently the parents may be supposed to have been homozygous. The moderately dark parent (*B* × *A*) pl. 2 failed to produce any offspring as dark as the grandparent (*B*).

In Series V the same plant (*B*) was crossed with a light heterozygous plant (*C*). From the offspring produced five homozygous dominants were selfed, and in the five families raised only two plants reached the colour-intensity of the grandparent (*B*).

On taking all these results together it may be said that there is evidence for the view that crossing a dark race of foxgloves with white plants tends to dull the colour-intensity of homozygous dominants of subsequent generations.

General Coloration—Strength of Inheritance and Effect of Selection.

In 1914 a dark-coloured homozygous plant (B_4 ♀) was crossed with a somewhat pale-coloured heterozygous plant (C_1 ♂) = $DD \times DR$ = III. The offspring would consist theoretically of approximately equal numbers of dominants and heterozygous individuals. The reciprocal cross (C_1 ♀ × B_4♂) was also made = II. Several dominants were selfed and families were raised. Out of these families certain plants were selected and selfed and new families were obtained. This procedure was continued until 1917, and the results are given in the accompanying table. The families of the different years are arranged in ascending order of the colour-intensities of the parents. On comparing the means of the families with the colour-grade of the parents (shown in brackets) it will be at once seen that small variations in the colour-intensity of the parents tended to be transmitted to the offspring. It is obvious that the table exhibits the effect of selection in self-fertilised homozygous generations.

For example we may take the following:

Homozygous plant,	II. 1	had a colour of	70	and a mean of offspring	74
An offspring of above,	II. 1, 4	„	„	74	„	„	„	82
An offspring of above,	II. 1, 4, 17	„	„	110	„	„	„	95
Homozygous plant,	III. 2	„	„	75	„	„	„	66
An offspring of above,	III. 2, 1	„	„	66	„	„	„	55
An offspring of above,	III. 2, 1, 18	„	„	80	„	„	„	85
An offspring of above,	III. 2, 1, 18, 28	„	„	95	„	„	„	100

Reverse selection is shown also:

Homozygous plant,	III. 2	„	„	75	„	„	„	66
An offspring of above,	III. 2, 5	„	„	52	„	„	„	57
An offspring of above,	III. 2, 5, 5	„	„	40	„	„	„	41
An offspring of above,	III. 2, 5, 5, 12	„	„	30	„	„	„	32

Inheritance of Colour-Intensity among Dominants.

Grades of Colour-Scale (Offspring)	*DR* × *DD*	Dominant Generations (Self-fertilisation)																		
	II = $C_1(34) \times B_4(80)$	Parents II 4 (50)	II 6 (68)	II 1 (70)	II 4, 8 (45)	II 4, 6 (49)	II 6, 3 (49)	II 6, 4 (54)	II 4, 2 (61)	II 6, 11 (61)	II 6, 1 (68)	II 4, 12 (69)	II 1, 4 (74)	II 1, 2 (91)	II 1, 1 (120)	II 1, 2, 20 (53)	II 6, 11, 6 (59)	II 1, 4, 3 (64)	II 1, 1, 2 (77)	II 1, 4, 17 (110)	
30— 39	2	—	1	—	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
40— 49	0	3	8	—	15	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
50— 59	4	8	13	1	6	1	15	5	1	1	2	—	—	5	1	3	—	1	—	—	—
60— 69	1	12	4	9	—	7	—	7	7	7	6	3	4	2	2	4	3	3	1	—	—
70— 79	1	1	3	14	—	2	—	—	2	8	1	—	7	4	5	—	1	3	0	—	—
80— 89	—	—	—	1	—	—	—	—	—	—	—	—	6	3	1	—	—	2	6	1	—
90— 99	—	—	—	4	—	—	—	—	—	—	—	—	3	4	—	—	—	—	—	0	—
100—109	—	—	—	1	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	2	—
110—119	—	—	—	—	—	—	—	—	—	—	—	—	2	0	—	—	—	—	—	—	—
120—129	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—
Means	53	59	54	74	51	66	54	62	66	69	64	64	82	78	70	59	65	71	81	95	—

Grades of Colour-Scale (Offspring)	III = $B_4(80) \times C_1(34)$	Parents III 2 (75)	III 2, 5 (52)	III 2, 7 (52)	III 2, 13 (63)	III 2, 1 (66)	III 2, 2 (71)	III 2, 8 (72)	III 2, 5, 5 (40)	III 2, 5, 18 (52)	III 2, 7, 9 (62)	III 2, 5, 10 (72)	III 2, 1, 18 (80)	III 2, 5, 5, 12 (30)	III 2, 5, 10, 17 (37)	III 2, 5, 5, 11 (42)	III 2, 1, 18, 10 (53)	III 2, 5, 10, 22 (70)	III 2, 1, 18, 12 (85)	III 2, 1, 18, 21 (95)	III 2, 1, 18, 28 (95)
20— 29	—	—	—	—	—	—	—	—	—	—	—	—	—	3	—	—	—	—	—	—	—
30— 39	1	—	—	—	—	—	—	—	4	2	—	1	—	7	2	2	2	—	—	—	—
40— 49	0	—	3	3	—	5	—	—	7	8	—	0	—	—	4	5	6	—	—	—	—
50— 59	1	4	9	6	3	5	—	1	—	5	5	2	1	—	1	4	2	1	—	—	—
60— 69	0	17	3	2	1	2	6	11	—	—	4	2	1	—	—	—	—	0	1	3	—
70— 79	2	9	2	—	—	0	11	9	—	—	1	4	2	—	—	—	—	0	1	0	1
80— 89	1	1	—	—	—	1	1	0	—	—	—	2	3	—	—	—	—	2	1	1	4
90— 99	—	—	—	—	—	—	—	1	—	—	—	—	5	—	—	—	—	—	—	2	9
100—109	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	2	4
110—119	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	8
120—129	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	4
130—139	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Means	67	66	57	54	60	55	71	69	41	47	62	68	85	32	43	47	44	74	75	91	100

Thus, starting with a plant of about 70 colour-intensity we arrive by selection of self-fertilised plants at mean family intensities of 100 in one direction and 32 in the reverse direction.

In another series, starting with a homozygous dominant plant of colour-intensity of about 11, I have by selection obtained plants in which the corolla exhibited no general tint. On selfing the pale plant no white plants occurred,

and the offspring were all pale-coloured; but when the intensity was decreased by selection to about 4, the "white" plants showed Mendelian segregation, for the offspring arising from the plants produced from a cross with a dark-coloured plant were sharply divisible into strongly coloured and "white" individuals.

As a further example of selection, I started with a homozygous medium-coloured (48) plant (*G*). This was selfed and a family of 31 coloured plants was raised, there were no whites. Thus, the parent plant may be regarded as homozygous. A plant (*G* 3) in this family, not far removed in colour (55) from the average, was selfed and the resulting family had a mean colour approximating to the colour of the parent. A light-coloured (27) plant (*G* 3, 20) and a dark (81) plant (*G* 3, 13) in this last family were selfed also, and the two families raised tended to resemble their respective parents. In a succeeding generation further progress was obtained in securing a dark race and a pale race. The necessary details are given in the accompanying diagrammatic table. The families printed in heavy type are those leading to a dark race, while those in ordinary type are passing into a light race.

Formation of Light and Dark Races from a Dominant (homozygous) G.

Parents			Offspring—Scale of Colour								
Number		Colour	95—104	85—94	75—84	65—74	55—64	45—54	35—44	25—34	15—24
G (selfed)	...	*48*	—	—	—	1	9	15	6	—	—
G pl. 3	...	*55*	—	—	2	1	0	4	5	3	1
G 3, pl. 20	...	*38*	—	—	—	—	—	—	4	1	—
G 3, pl. 13	...	*81*	—	**2**	**3**	**2**	**2**	**2**	**3**	—	—
G 3, pl. 20	...	*27*	—	—	—	—	—	—	4	5	2
G 3, 13, pl. 2	...	*80*	**2**	**6**	**2**	—	—	—	—	—	—

Correlation Table—Colour-Intensity—Dominants (homozygous). Series II and III.

Parents. Grade of Colour-Intensity	130—139	120—129	110—119	100—109	90— 99	80— 89	70— 79	60— 69	50— 59	40— 49	30— 39	20— 29	Offspring
30— 39	—	—	—	—	—	—	—	—	1	4	9	3	17
40— 49	—	—	—	—	—	—	2	7	26	29	10	—	74
50— 59	—	—	—	—	—	—	4	31	38	23	4	—	100
60— 69	—	—	—	—	—	3	18	37	31	13	1	—	103
70— 79	—	—	2	2	8	19	54	50	9	0	1	—	145
80— 89	—	—	—	2	5	4	3	2	1	—	—	—	17
90— 99	2	5	9	7	15	8	5	5	5	—	—	—	61
100—109	—	—	—	—	—	—	—	—	—	—	—	—	0
110—119	—	—	—	2	0	1	—	—	—	—	—	—	3
120—129	—	—	—	—	—	1	5	2	1	—	—	—	9
Totals	2	5	11	13	28	36	91	134	112	69	25	3	529

In the last table, p. 113, a correlation surface is shown between parents and offspring. It is formed from the series of families given in the table preceding the last, and arising by self-fertilisation.

The constants calculated from the table are: standard deviation of weighted parents 1·7805 units, and of offspring 1·8962 units, coefficient of correlation ·707.

In this table 39 families were involved, as detailed in the previous table. The starting points were four homozygous dominant plants occurring in the two families raised from the reciprocal crosses ($C_1 \times B_4$) and ($B_4 \times C_1$).

3. Brown Spots.

The amount of spotting on the inside of the corolla is not closely correlated to the intensity of the general purple coloration of the flower, for even in white plants the spots may be numerous and of a deep purple colour. In coloured plants the spots were almost always dark purple. As a very rare exception in the coloured plants (4 plants in about 2500) some of the spots were russet brown, and in the case of the larger spots there was a middle area of brown bordered by a margin of purple. In white flowers the spots were fairly frequently brownish-green or brown. In such brown spotted white flowers I could never detect the slightest tinge of purple on the general surface of the corolla, while in purple-spotted white flowers a faint tinge of purple could often be seen. The brown spots of white flowers might not become visible until the flowers were on the point of fading, and in the case of any given white plant it was wholly impossible to affirm that brown spots were, or would be, entirely absent from all of the flowers.

With the exception of the four plants mentioned above there was a sharp discontinuity to the naked eye between purple spots and brown spots, intermediate conditions being absent. The brown colouring matter may be regarded as altered or decomposed anthocyanin. In purple spots a microscopic examination often showed a certain amount of decomposition; but, with the exception of the four plants, the amount was not enough to alter the colour of the spots sufficiently for detection by the naked eye. Thus, the discontinuity lies between a normal small amount of decomposition, and an abnormal entire decomposition. It may be stated that under ordinary circumstances brown or greenish spots (as seen by the naked eye) are linked to a perfectly white corolla, but purple spots occur in both purple and "white" flowers, and an apparently perfectly white corolla may also bear purple spots.

If a brown spotted plant is crossed with a purple spotted one the offspring are all purple spotted and heterozygous. The brown spotted condition is inherited in Mendelian fashion, and is recessive to purple spots.

No special crossings have been made to investigate the matter, and the results which are given below are merely picked out from the records of the numerous families which have been raised for other purposes.

In the accompanying table it is useless to include families in which there was no taint of whiteness, since all the individuals (except 4 plants out of 2500) had purple spots.

Brown Spots—Families White or Some Taint of Whiteness.

Gametic Nature of Pairings	Number of Families	Number of Offspring	Purple Spotted		Brown Spotted	
			Experimental	Mendelian Expectation	Experimental	Mendelian Expectation
$DD \times DD$	13	344	344	344	0	0
$RR \times RR$	11	169	0	0	169	169
$DR \times DR$	13	213	166	160	47	53
$DR \times DD$	15	137	137	137	0	0
$DR \times RR$	1	8	3	4	5	4
$DD \times RR$	6	70	70	70	0	0
Totals	59	941	720	715	221	226

It is obvious from the table that the brown spotted condition exhibits Mendelian inheritance.

4. Inheritance of Certain Sport Abnormalities.

Crenate Margin.—In a homogeneous family of 29 plants there appeared one plant in which the free edge of the mouth of the flower exhibited a well-marked serrated condition. All the flowers of a main-axis of considerable size were similarly affected, and later, lateral flowering axes were formed, and the flowers were also serrate. The character was sufficiently marked to be noticeable at a casual glance of the plant, and since all the numerous flowers were alike in this particular, the character was clearly inherent in the plant, and was not due to a chance environmental disturbance influencing a young growing axis or certain flower-buds. The plant was self-fertilised, and it was confidently expected that the character would reappear in the offspring. Out of a family of some 20 plants 12 flowered and no sign of the peculiar serrated condition could be detected in any one of the plants. Here we have a conspicuous character in a large healthy plant affecting every flower of all the flowering axes, and yet apparently it was incapable of being transmitted to the offspring.

Split Corolla.—In a homogeneous family (XXXIV) of 27 plants there appeared one plant in which in the great majority of the numerous flowers the corolla was symmetrically divided into an upper, a lower and two lateral pieces by four lateral splits extending down to the base of the flower. The plant was a large, healthy one and produced a number of similar lateral axes. At least 90 °/₀ of the flowers were completely split (Pl. I, fig. 10).

In a family (VIII 7) unrelated to the above there were 16 plants, and of these, four plants were similarly affected. In one of these plants practically all

(99 °/₀) of the flowers were entirely split into four pieces, while in the remaining three plants some 50—60 °/₀ of the flowers were split. All the plants were large and vigorous. It was thought that very probably the character would exhibit Mendelian inheritance. The results of crossing and selfing are shown in the accompanying table.

Inheritance of Split Corolla.

Parents															
	Mid-parental degree of Splitting	0	0	94	0	13	18	0	4	99	0	0	45	0	0	0
	Registered Number	XXXIV (selfed)	VIII 7 (selfed)	XXXIV 4 (90 °/₀) × VIII 7, 9 (99 °/₀) = S. J.	S. J. pl. 9 (selfed)	S. J. pl. 18 (selfed)	S. J. pl. 6 (selfed)	S. J. 18 pl. 11 (selfed)	S. J. 18 pl. 4 (selfed)	S. J. 18 pl. 10 (selfed)	II 6, 1 (selfed)	XXXIV (selfed)	II 6, 1 (0 °/₀) × XXXIV 4 (90 °/₀) = R. J. ...	R. J. pl. 9 (selfed)	R. J. pl. 16 (selfed)	R. J. pl. 16 pl. 14 (selfed) ... ...
Offspring. Percentage of Splitting	No Splitting	26	12	8	7	2	3	1	0	0	9	26	12	15	10	10
	1—14	0	0	6	3	4	2	1	2	0	0	0	0	0	0	1
	15—29	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
	30—44	0	0	0	0	1	1	2	2	0	0	0	0	0	0	2
	45—59	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0
	60—74	0	2	0	1	1	3	3	2	1	0	0	0	0	0	0
	75—99	1	1	0	2	5	4	1	1	3	0	1	0	0	0	0

The first mentioned plant (XXXIV 4) with 90 °/₀ of the flowers split was crossed with an unrelated plant with some 99 °/₀ of the flowers split (5th vertical column of table). Of the 17 offspring 8 plants were wholly unsplit, while the remainder exhibited the character in a very greatly weakened condition. Three of these offspring, S. J. nos. 9, 18 and 6 having 0 °/₀, 13 °/₀ and 18 °/₀ of the flowers split respectively, were selfed, and the families raised all contained some plants very conspicuously split, but the character was more marked in the two families raised from parents 18 and 6 which showed some degree of splitting. In a subsequent generation (S. J. 18 pl. 4 and S. J. 18 pl. 10) raised by selfing, the character became very strongly pronounced.

An unrelated non-split plant (II 6, 1) was crossed with the first mentioned plant having at least 90 °/₀ of the flowers split (XXXIV 4). In the family of

12 plants raised none of the plants exhibited splitting. Two of these offspring (R. J. nos. 9, 16) were selfed and no splitting occurred in the two families. Another generation was raised from R. J. 16, plant 14 and some re-appearance of splitting was detected. The table includes all the split plants which have occurred among some 3000 plants which have been under observation.

The results obtained indicate that heredity has some influence, but the data are insufficient for determining the nature of the transmission which does not bear a Mendelian aspect.

Creased Upper Lip.—In a certain plant in the majority of the flowers the upper surface and lip exhibited a conspicuous pucker or crease. This plant was crossed with an unrelated normal plant with no crease. Most of the seedlings were killed by the violent elements, but four plants were raised, and in one, a number of flowers exhibited a crease, which, however, was much less developed than in the paternal parent. The data are scanty, but the hereditary transmission does not seem to be Mendelian.

Spontaneous Appearance of White plants.—Among the numerous homozygous dominant coloured families that have been raised a white plant appeared spontaneously on two occasions in two unrelated families. These plants, of course, bred true, and as there was no evidence of contamination of the seed the plants must be regarded as new sports.

5. INHERITANCE OF SEED-LENGTH.

The mean length of the seed varied considerably in different plants. No discontinuous variation could be detected, and inheritance was of the blended type. Ten seeds were taken at random from one or more capsules of a number of plants of certain series and the means determined. The seeds of a capsule exhibited a moderate amount of variation, but they were monomorphic in varietal crossings, and not dimorphic as was noticed in an interspecific crossing. The distribution was more or less normal. Unfortunately there was very considerable variation in the mean size of the seeds in different capsules of the same plant, and consequently no very accurate determination of the strength of inheritance was possible with this character without an excessive number of measurements. As it was, the investigation entailed the measurement of about 1000 seeds.

A plant, C_1 (mean seed-length 639 units), was crossed with B_4 (mean seed-length 628 units) and a family was raised; $C_1 \times B_4 = \text{II}$. In family II twelve plants were selfed, namely II 1, II 2 ... II 12, the seeds were measured and twelve families were obtained. In family II 1 three plants were selfed and the seed-length determined, namely (II 1) 1, (II 1) 2 and (II 1) 4. The means of the seed-lengths of these three plants were compared with the seed-length of the parent II 1. Similarly, for example, in family II 1, 2 two plants were selfed, namely (II 1, 2) 5 and (II 1, 2) 20, and the means of the seed-lengths of these two plants were compared with the seed-length of the parent II 1, 2. The data are given in the accompanying table.

Mean Seed-length, Parents and Offspring.

Parent (selfed)		Offspring (selfed)		Parent (selfed)		Offspring (selfed)		Parent (selfed)		Offspring (selfed)	
Desig-nation	Mean Seed-length	Desig-nation	Mean Seed-length	Desig-nation	Mean Seed-length	Desig-nation	Mean Seed-length	Desig-nation	Mean Seed-length	Desig-nation	Mean Seed-length
II 1	606	II 1, 1	572	II 4	592	II 4, 8	628	II 9	653	II 9, 3	629
		II 1, 2	668			II 4, 12	598				
		II 1, 4	649								
				II 6	620	II 6, 1	621	II 10	646	II 10, 1	660
II 1, 2	668	II 1, 2, 5	668			II 6, 3	641			II 10, 2	669
		II 1, 2, 20	642			II 6, 4	670			II 10, 5	649
						II 6, 11	695			II 10, 7	660
II 1, 4	649	II 1, 4, 3	655							II 10, 8	713
		II 1, 4, 17	674	II 6, 11	695	II 6, 11, 6	665			II 10, 13	681
								II 10, 1	660	II 10, 1, 18	642
II 2	528	II 2, 1	624	II 7	547	II 7, 1	671				
		II 2, 3	582			II 7, 12	570	II 10, 5	649	II 10, 5, 5	598
		II 2, 5	637			II 7, 14	624			II 10, 5, 10	629
		II 2, 16	566							II 10, 5, 18	649
				II 7, 1	671	II 7, 1, 7	649				
II 3	629	II 3, 1	686					II 10, 7	660	II 10, 7, 9	653
		II 3, 4	686	II 8	620	II 8, 2	629				
		II 3, 15	672			II 8, 3	624	II 11	615	II 11, 8	617
II 4	592	II 4, 2	668	II 9	653	II 9, 2	633	II 12	679	II 12, 9	642
		II 4, 6	657			II 9, 11	620				
						II 9, 10	630				

C_1 (self-pollen) seed-length = 639
B_4 (self-pollen) „ = 628
C_1 (B_4 pollen) „ = 642, these last seeds produced fam. II.

The coefficient of correlation, calculated from the above numbers, between parents (selfed) and offspring (selfed) is ·378. This is low for mid-parental correlation; but as all the generations arose by self-fertilisation we ought to have practically no correlation at all according to the pure-line hypothesis, for the two original parents (C_1 and B_4) were closely similar to each other in the character under investigation.

6. Purple Spotting of the Corolla.

The purple spotting of the lower surface of the corolla-tube and lower lip varied greatly in the original parent plants, and the character was obviously inherited. The amount of spotting had little relationship to the intensity of the general coloration of the corolla, and "white" flowers were sometimes richly spotted with purple.

The percentage area of the lower surface covered with spots was estimated by comparing the flowers with a series of diagrams each covered with a definitely known percentage of spotting. With practice it was found that sufficiently uniform results could be obtained by this method.

In plants which had lost completely the power of producing any purple coloration whatever, the spots were brown and usually small and scanty, and among such plants an almost entire absence of spots of any kind occasionally occurred. We have already seen that with regard to the colour of the spots (brown and purple) Mendelian segregation takes place.

In the inheritance of the *amount* of purple spotting no Mendelian relationship could be detected. The smallest amount of purple spotting met with in coloured foxgloves equalled about 1 %, and the maximum about 70 %. It will be remembered that on crossing a dark purple plant with a plant bearing flowers very faintly tinged with purple (say, colour 4 of standard), definite segregation into "white" and purple plants occurred in the second generation following; but on crossing a plant possessing an abundance of purple spots (say, 50 %) with a plant bearing very few purple spots (say, 2 % or 3 %) no such segregation was found, and the spotting tended to remain intermediate in amount.

In the numerous crosses that have been made for various purposes the condition of the spotting was observed, and it is undoubtedly true that the means of the spotting of the families resulting from the crosses tended on the average to approximate to the spotting of the mid-parent, $\frac{1}{2}(♂ + ♀)$. No difference could be detected between the reciprocal crosses of two plants.

Influence of Selection and Strength of Inheritance in Self-fertilised Generations.

In this connection details of Series II and III may be given (see p. 120). Plant C_1 with 11 % spotting was crossed with pollen of plant B_4 with spotting 48 % = II. Seven of the offspring were selfed and the spotting of the resulting families was determined. Subsequently two other generations were raised by selfing. Plant B_4 was crossed with pollen of C_1 = III. Four of the offspring were selfed and subsequently three other generations were raised by self-fertilisation.

The distributions of the spotting in the families of the different generations are shown in the accompanying table. In each generation the families are arranged in the ascending order of the parental spotting (see the top and middle horizontal lines). A casual inspection indicates at once that the general trend of the family-distributions follows the gradual increase in the spotting of the parents.

As an example of selection we may take:

III 2 (9 %) selfed	produced with others a plant	III 2, 5 (15 %)
III 2, 5 (15 %) selfed	„ „ „	III 2, 5, 10 (22 %)
III 2, 5, 10 (22 %) selfed	„ „ „	III 2, 5, 10, 17 (27 %)
III 2, 5, 10, 17 (27 %) selfed	produced a family with mean spotting of 39 %	

Thus, we have passed from a plant with 9 % spotting to a plant with 27 %, which on selfing produced a family with a mean spotting of 39 %.

With reference to the strength of inheritance two tables are given on p. 121, one for parents and offspring, and one for grandparents and grandchildren. The respective coefficients of correlation are ·560 and ·395. This correlation does not arise by the mixture of two races which have been sorted out by segregation

Purple Spotting—Families from Self-fertilised Parents. Series II and III.

Offspring. Percentage Grade of Spotting	III = B_4 (48) × C_1 (11)		III 1	III 3	III 4	III 2, 1	III 2, 2	III 2, 5	III 2, 7	III 4, y	III 2, 13	III 2, 8	III 1, 11	III 4, x	III 2, 5, 18	III 2, 7, 9	III 2, 5, 5	III 4, y, 30	III 4, x, 27	III 2, 1, 18	III 2, 5, 10	III 4, x, 23	III 4, x, 5	III 4, x, 22	III 2, 5, 12	III 2, 1, 18, 12	III 2, 1, 18, 10	III 2, 1, 18, 28	III 2, 5, 5, 11	III 2, 5, 10, 22	III 2, 5, 10, 17	Means
0— 3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
4— 7	—	—	—	—	—	3	5	2	1	—	—	—	1	—	6	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—
8—11	1	1	2	—	—	1	4	7	3	1	—	8	1	—	1	—	—	—	—	—	—	—	—	—	2	—	3	—	—	—	—
12—15	0	1	3	1	2	2	2	4	5	1	2	0	4	3	7	—	—	—	—	—	—	—	—	—	1	1	6	—	—	—	—
16—19	1	3	7	0	1	4	1	7	1	3	—	2	0	4	2	—	—	—	—	—	—	—	—	—	4	4	5	1	3	—	—
20—23	3	—	9	3	5	1	1	2	—	3	1	1	2	4	—	—	—	—	—	—	—	—	—	—	4	3	6	4	1	1	—
24—27	—	—	10	—	3	—	—	—	—	2	—	—	1	3	—	—	—	—	—	—	—	—	—	—	4	1	3	6	4	0	—
28—31	—	—	7	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	3	1	3	2	1	—
32—35	—	—	3	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	2	4	3	1	0	—
36—39	—	—	4	—	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	0	1	—
40—43	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	1	1	—
44—47	—	—	0	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	0	1	—
48—51	—	—	2	—	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—
Means	18		23	20	27	13	13	14	12	18	16	13	16	23	8	14	11	18	23	19	25	32	28	31	19	25	30	21	29	28	38	39
Parents	Spotting of Parents																														
III 2	9	12																													
III 1	20	23																													
III 3	20	20																													
III 4	20	27																													
III 2, 1	12	13																													
III 2, 2	14	13																													
III 2, 5	15	14																													
III 2, 7	15	12																													
III 4, y	15	18																													
III 2, 13	17	16																													
III 2, 8	19	13																													
III 1, 11	23	16																													
III 4, x	24	23																													
III 2, 5, 18	2	8																													
III 2, 7, 9	6	14																													
III 2, 5, 5	10	11																													
III 4, y, 30	13	18																													
III 4, x, 27	18	23																													
III 2, 1, 18	19	19																													
III 2, 5, 10	22	25																													
III 4, x, 23	25	32																													
III 4, x, 5	27	28																													
III, 4, x, 22	29	31																													
III 2, 5, 5, 12	11	19																													
III 2, 1, 18, 12	16	25																													
III 2, 1, 18, 10	17	30																													
III 2, 1, 18, 28	19	21																													
III 2, 5, 5, 11	19	29																													
III 2, 5, 10, 22	26	38																													
III 2, 5, 10, 17	27	39																													

Offspring. Percentage Grade of Spotting	C_1 (11) selfed		II 3	II 6	II 1	II 8	II 2	II 2	II 7	II 6	II 1, 1	II 6, 11	II 1, 2	II 4	II 1, 2	II 1, 1	II 6, 3	II 6, 1	II 1, 4	II 6, 4	II 4, 6	II 7, 14	II 4, 2	II 7, 12	II 8, 2	II 4, 12	II 4, 8	II 8, 3	II 6, 11, 6	II 1, 2, 20	II 1, 4, 17	II 1, 4, 3	II 1, 1, 2	Means
0— 3	1	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
4— 7	5	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8—11	2	—	3	1	3	1	2	3	3	3	1	1	6	2	6	1	4	6	4	6	1	1	3	2	6	1	2	1	9	1	6	3	1	—
12—15	1	—	7	9	5	0	8	3	0	3	2	6	6	2	6	1	3	1	1	1	3	2	4	2	4	2	1	0	6	1	1	1	1	—
16—19	2	—	5	4	9	0	2	1	1	5	2	4	6	1	1	1	1	1	1	4	2	1	4	2	2	1	1	0	2	1	1	1	1	—
20—23	—	—	3	10	10	0	1	0	1	2	1	2	1	1	2	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	—
24—27	—	—	2	2	2	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
28—31	—	—	1	2	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
32—35	—	—	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
36—39	—	—	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
40—43	—	—	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
44—47	—	—	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
48—51	—	—	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
52—55	—	—	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
56—59	—	—	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	—
Means	10		17	18	18	16	20	16	12	21	18	16	11	21	16	18	21	34	19	24	27	37	20	22	20	32	30	43	16	14	14	16	37	18
Parents	Spotting of Parents																																	
II 3	13	17																																
II 6	14	18																																
II 1	16	18																																
II 8	16	28																																
II 2	17	16																																
II 2	17	20																																
II 4	28	21																																
II 1, 2	10	11																																
II 1, 1	13	18																																
II 6, 11	13	16																																
II 7, 1	14	16																																
II 7, 2	15	16																																
II 6, 3	19	21																																
II 6, 1	20	34																																
II 1, 4	21	19																																
II 6, 4	21	24																																
II 4, 6	22	27																																
II 7, 14	24	37																																
II 4, 2	25	20																																
II 7, 12	25	22																																
II 8, 2	25	20																																
II 4, 12	28	32																																
II 4, 8	30	30																																
II 8, 3	45	43																																
II 6, 11, 6	13	16																																
II 1, 2, 20	17	14																																
II 1, 4, 17	17	14																																
II 1, 4, 3	18	16																																
II 1, 1, 2	21	23																																

during the different self-fertilised generations. Inspection of the tables shows that the distributions of the various families give no indication whatever of the occurrence of segregation into little spotted and much spotted plants. The gradual rise in the degree of spotting of the different parents is followed by a gradual increase in the spotting of the respective families obtained by self-fertilisation. The fact that the correlation between the grandparents and grandchildren is less than that between the parents and offspring is further evidence that the small, apparently fortuitous, variations in spotting occurring among self-fertilised generations are inherited. This result is opposed to the pure-line hypothesis, according to which such small variations are regarded as slightly different expressions of the same identical character which remains unchanged in its essence from one self-fertilised generation to another. If such were the case

Correlation Table—Spotting—Parents and Offspring. Series II and III.

Offspring. Grades of Spotting.

Parents. Grades of Spotting	48—51	44—47	40—43	36—39	32—35	28—31	24—27	20—23	16—19	12—15	8—11	4—7	0—3	Totals
0— 3	—	—	—	—	—	—	—	—	—	1	6	8	—	15
4— 7	—	—	—	—	—	—	—	—	2	7	1	—	—	10
8—11	—	—	—	—	—	—	2	2	18	18	17	10	1	68
12—15	—	—	—	—	—	1	8	28	40	68	19	5	2	171
16—19	—	1	1	5	14	10	28	34	33	49	20	1	—	196
20—23	2	3	5	4	9	16	25	26	25	17	3	1	—	136
24—27	—	2	5	6	3	8	15	14	10	3	1	—	—	67
28—31	—	1	3	3	6	10	9	4	10	2	2	—	—	50
32—35	—	—	—	—	—	—	—	—	—	—	—	—	—	0
36—39	—	—	—	—	—	—	—	—	—	—	—	—	—	0
40—43	—	—	—	—	—	—	—	—	—	—	—	—	—	0
44—47	—	2	—	1	—	—	—	—	—	—	—	—	—	3
Totals	2	9	14	19	32	45	87	108	138	165	69	25	3	716

Correlation Table—Spotting—Grandparents and Grandchildren. Series II and III.

Grandchildren. Grades of Spotting.

Grand-parents. Grades of Spotting	44—47	40—43	36—39	32—35	28—31	24—27	20—23	16—19	12—15	8—11	4—7	0—3	Totals
8—11	—	—	1	4	2	6	8	19	35	19	5	2	101
12—15	1	1	2	5	4	17	21	20	31	10	12	—	124
16—19	2	3	6	12	9	20	22	25	37	15	2	1	154
20—23	2	4	1	1	4	7	9	15	13	3	1	—	60
24—27	—	1	6	4	7	8	0	4	—	—	—	—	30
28—31	2	3	1	4	7	8	6	8	1	—	—	—	40
Totals	7	12	17	30	33	66	66	91	117	47	20	3	509

the small variations would be fluctuating, non-inheritable variations; but the results in the present case are definitely against a supposition of this kind.

It might be urged by some that the result is really due to the existence of genotypes, and that variations within the limits of each genotype are not inheritable. The distributions of the families in the table do not indicate the occurrence of genotypes of any considerable magnitude. If the genotypes are supposed to be very small the practical result would become indistinguishable from the inheritance of continuous variations.

7. Ratio of Breadth to Length of Corolla.

The breadth was measured as the maximum horizontal width across the mouth of the corolla of a fully expanded flower in which the anthers had opened; the length was the maximum distance measured along the mid-adcauline surface with the lower lip stretched out straight in the long axis of the flower. It is convenient to express the ratio in the form, $\frac{\text{Breadth}}{\text{Length}}$ 1000. The mean of the ratios of the four lowest flowers of an axis was taken as the mean of the plant.

The original parent plants varied widely in this ratio, and the families raised by selfing tended to have the same ratio as their parents.

A plant bearing wide flowers was crossed with one having narrow flowers, and the offspring tended to be intermediate. On selfing these offspring the new generation exhibited, of course, considerable variation, but taken as a whole the intermediate condition was retained, and there was clearly no segregation into wide flowers and narrow flowers. Thus, the different degrees of this character blend readily on crossing, and the mode of inheritance is very similar to that of the spotted condition.

The results of a multitude of crossings of plants bearing variously shaped flowers have been carefully determined and tabulated, and there is no question about the general accuracy of the statement made above. In the present place we may confine our attention to the self-fertilised generations of Series II and III (p. 123).

A plant (♀ C_1) with relatively wide flowers (ratio 608) was crossed with a plant (♂ B_4) having relatively narrow flowers (ratio 487). The family (= II) had flowers approximately intermediate. The reciprocal cross = III. The distributions of the families of the various generations raised by selfing are shown in the accompanying table. The families of each generation are given in an ascending order of the ratios of the parents. As in the case of the character of spotting it will be seen that there is a clearly marked tendency for the mean ratios of the families to approximate to the ratios of the respective parents. In none of the families do we find any definite segregation into plants with wide flowers and plants with narrow flowers resembling those of the two progenitors of the series.

Wide and narrow races could be raised by selection using only self-fertilisation.

Thus in family III with a mean ratio of 531 there was a single plant (III 2) with as high a ratio as 575. This was selfed and the mean ratio of the offspring

Ratio of Flower—Families from selfed plants. Series II and III.

Offspring. Grades of Ratios Width / Length ×1000	B_4 (selfed) 487	C_1 (selfed) 608	II = $C_1 \times B_4$	Ratio	475	497	518	529	571	578	580	443	476	482	491	496	499	509	527	529	529	529	540	541	542	553	554	601	487	510	534	567	585
				Parents	II 6	II 3	II 4	II 7	II 1	II 8	II 2	II 6, 4	II 6, 3	II 4, 2	II 4, 6	II 4, 12	II 7, 2	II 1, 2	II 6, 1	II 4, 8	II 7, 1	II 7, 12	II 7, 14	II 1, 4	II 1, 1	II 6, 11	II 8, 3	II 8, 2	II 1, 4, 3	II 1, 4, 17	II 1, 1, 2	II 1, 2, 20	II 6, 11, 6
350—379	—	—	—		1	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
380—409	—	—	—		1	1	—	—	—	—	—	1	—	—	1	—	—	—	0	—	1	—	—	—	—	—	—	—	—	—	—	—	—
410—439	—	—	—		1	1	—	—	—	—	—	2	—	1	1	3	—	—	1	1	0	—	—	2	—	—	—	—	—	—	—	—	—
440—469	2	—	—		2	2	4	—	—	—	2	5	1	5	1	—	1	2	3	1	5	0	—	0	—	—	2	—	—	—	—	—	—
470—499	5	—	2		9	4	10	1	4	—	2	3	3	1	2	—	0	7	3	4	4	4	—	4	5	6	1	3	—	—	—	—	2
500—529	2	2	2		3	7	5	4	5	—	5	—	—	2	6	—	1	1	1	6	1	6	1	4	3	2	0	—	5	—	2	3	1
530—559	1	2	1		7	2	1	3	9	2	1	—	—	1	2	—	—	1	—	3	0	0	—	3	1	3	2	1	1	—	1	4	0
560—589	—	2	3		1	3	2	2	8	0	4	—	—	—	—	—	—	3	—	—	1	2	—	1	—	2	—	1	—	2	3	—	1
590—619	—	1	—		—	—	—	1	1	1	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	—	1	1	—	—
620—649	—	1	—		—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
650—679	—	—	—		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Means	494	561	535		495	506	497	538	546	567	522	451	502	477	493	427	485	511	459	505	477	507	522	505	502	508	522	582	502	577	556	524	511

Offspring. Grades of Ratios Width / Length ×1000	III = $B_4 \times C_1$	Ratio	490	528	535	575	451	518	518	533	549	574	587	588	674	421	434	483	490	505	524	528	551	559	611	432	488	497	531	537	556	569	598
		Parents	III 3	III 4	III 1	III 2	III 4, x	III 4, y	III 2, 8	III 2, 1	III 2, 5	III 2, 2	III 2, 13	III 1, 11	III 2, 7	III 4, x, 27	III 4, x, 5	III 4, x, 22	III 4, x, 23	III 2, 5, 10	III 4, y, 30	III 2, 7, 9	III 2, 1, 18	III 2, 5, 5	III 2, 5, 18	III 2, 5, 10, 22	III 2, 5, 5, 11	III 2, 5, 10, 17	III 2, 1, 18, 12	III 2, 1, 18, 10	III 2, 5, 5, 12	III 2, 1, 18, 21	III 2, 1, 18, 28
350—379	—		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
380—409	—		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—
410—439	—		—	1	1	—	4	—	—	—	—	—	—	—	—	—	4	1	2	2	1	—	—	—	—	—	—	1	—	—	—	—	—	—
440—469	—		1	4	4	—	3	3	1	—	—	—	—	—	—	2	2	4	1	5	0	—	—	1	—	1	3	2	—	—	—	—	—	—
470—499	1		2	1	10	3	5	1	2	—	—	2	—	—	—	1	1	4	3	2	1	—	1	1	—	2	3	1	4	—	—	—	—	—
500—529	2		1	7	10	2	1	4	6	1	3	2	—	—	1	—	1	1	1	1	3	—	0	4	—	—	3	2	5	1	2	—	—	—
530—559	1		—	1	5	7	1	5	8	7	2	8	—	2	0	—	0	—	—	—	5	1	5	3	2	—	2	1	7	4	4	1	2
560—589	1		—	1	3	8	—	—	0	1	6	2	2	2	2	—	1	—	—	—	1	1	3	2	7	—	—	—	3	1	4	3	6
590—619	—		—	1	—	3	—	—	1	0	5	—	0	5	2	—	—	—	—	—	—	3	4	—	4	—	—	—	2	3	—	5	13
620—649	—		—	—	—	6	—	—	—	2	2	—	1	2	3	—	—	—	—	—	—	3	—	—	2	—	—	—	1	1	—	1	7
650—679	—		—	—	—	2	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	2	—	—	—	—	—	—	1	—	—	—	2
680—709	—		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Means	531		481	506	507	574	465	509	526	563	579	538	592	592	614	464	469	474	468	440	522	618	561	526	587	477	499	488	550	573	547	592	606

was 574. In this family there was a plant (III 2, 1) with a ratio of 533 and the mean of offspring = 563. III 2, 1, 18 (ratio 551) produced a family with mean 561, and III 2, 1, 18, 28 (ratio 598) produced a family with a mean of 606.

In the reverse direction, through III 2, III 2, 5, III 2, 5, 10 and III 2, 5, 10, 22 we pass from a parent of ratio 575 to a family having a mean ratio of 477.

With the data given in the preceding table, correlation tables have been prepared for parents and offspring, and grandparents and grandchildren.

Correlation Table—Ratios of Corolla—Parents and Offspring.
Series II and III.

Offspring.

Parents. Grades of Ratios $\frac{B}{L}$ 1000	680—709	650—679	620—649	590—619	560—589	530—559	500—529	470—499	440—469	410—439	380—409	350—379	Totals
410—439	—	—	—	—	1	0	1	4	5	4	—	—	15
440—469	—	—	—	—	—	1	1	8	8	6	1	—	25
470—499	—	—	—	—	4	16	39	35	23	11	3	1	132
500—529	—	2	3	8	15	28	45	40	28	6	2	1	178
530—559	—	1	6	15	29	47	38	32	7	3	—	—	178
560—589	—	2	12	15	30	34	18	13	2	—	—	—	126
590—619	1	2	9	18	14	5	—	—	—	—	—	—	49
620—649	—	—	—	—	—	—	—	—	—	—	—	—	0
650—679	—	2	3	2	2	0	1	—	—	—	—	—	10
Totals	1	9	33	58	95	131	143	132	73	30	6	2	713

Correlation Table—Ratios of Corolla—Grandparents and Grandchildren.
Series II and III.

Grandchildren.

Grandparents. Grades of Ratios $\frac{B}{L}$ 1000	680—709	650—679	620—649	590—619	560—589	530—559	500—529	470—499	440—469	410—439	380—409	350—379	Totals
440—469	—	—	—	—	1	0	3	9	9	7	—	—	29
470—499	—	—	—	—	2	3	10	15	11	3	1	1	46
500—529	—	—	—	—	4	22	36	25	22	12	2	—	123
530—559	1	3	14	38	37	34	24	16	9	2	1	—	179
560—589	—	2	8	10	18	33	21	21	3	2	—	—	118
590—619	—	—	—	—	—	—	—	—	—	—	—	—	0
620—649	—	—	—	—	—	—	—	—	—	—	—	—	0
650—679	—	2	3	3	1	1	—	—	—	—	—	—	10
Totals	1	7	25	51	63	93	94	86	54	26	4	1	505

The coefficients of correlation are ·601 for parents and offspring and ·492 for grandparents and grandchildren. The latter figure is somewhat high; but taking the results altogether they are incompatible with any notion of pure-lines.

8. General Conclusions.

In the various characters that have been dealt with in the crossing of different strains of the garden foxglove we have seen that in pelorism, colour of corolla and colour of spots, the mode of inheritance is Mendelian with reference to the qualities: peloric and non-peloric, purple and white corolla, purple spots and brown spots. If, however, there are any marked differences in the intensities of these qualities, the mode of inheritance of the intensity of the quality was found to be of the blended type.

The other characters examined were quantitative in nature, such as degree of the development of purple spots and the ratio of breadth to length of corolla, and these characters blended completely.

When the intensity of a quality is very slight and approaching zero the difficulty arises as to which category the individual should be referred. When Mendelian inheritance is in evidence the critical point may apparently be determined by the occurrence of segregation. Thus, if a homozygous plant with a very faint tinge of purple (say an intensity of about 4) is crossed with a homozygous strongly coloured plant, segregation occurs in the so-called F_2 generation, and we obtain on the average 1 faintly tinged plant to 3 much more darkly coloured plants. When, however, the pale plant has a somewhat greater intensity (say about 10), the F_2 and subsequent generations are intermediate, and definite segregation does not occur. In accordance with this procedure a plant with flowers having an intensity of general coloration which did not reach 5 of the scale was classed as "white." Without employing such a line of demarcation the results obtained were wholly unintelligible.

From the strict Mendelian standpoint, in the example given above, it would probably be affirmed that the faint tinge of purple on "white" flowers is not really a fractional part of the general purple coloration of coloured plants, but is a distinct character governed by a different factor or set of factors in the chromosomes. To one who has grown the plants this view appears an artificial one. In my previous account I stated that there appeared to be a distinct gap among my plants between "white" plants and coloured plants, and that colorations of about 8—25 of the scale were extremely rare or almost absent, but I have subsequently obtained a number of plants having such intensities of coloration, passing imperceptibly down to absolute whiteness. Consequently it is quite unlikely that the faint tinge of purple on "white" flowers is anything else than the last remnant of a general purple coloration.

It is quite similar in the character of pelorism, but the difficulty in finding a suitable method of measuring this character renders the matter less obvious. Thus, it would appear that if a character is not present beyond a certain minimum or unit quantity it may be unable to blend on crossing with a plant possessing the character in a well-marked degree.

With reference to the characters which blend, the accompanying table summarizes the results obtained for parental correlation. Mid-parents and self-fertilised parents are regarded as comparable.

Character	Number of Offspring	Coefficient of Correlation. Parents and Offspring
Intensity of pelorism (homozygous recessive, mid-parents and self-fertilised parents)	530	·520
Intensity of general purple coloration (homozygous dominants, self-fertilised parents)	529	·707
Seed-length (self-fertilised parents)	46	·378
Spotting (self-fertilised parents)	716	·560
Ratio of Corolla (self-fertilised parents) ...	713	·601

The probable errors of these results are reasonably small and the average coefficient for the 5 characters is ·553 which is not far removed from the average coefficient found by Professor Karl Pearson for a large number of characters in a variety of different organisms.

It must be again emphasized that these results are based on self-fertilised generations of pedigree plants of known gametic constitution, and on Johannesen's theory of pure-lines these parental coefficients should be zero, or at least very small.

The evidence of the present investigation is therefore definitely against any general application of the theory of pure-lines and of genotypes of any appreciable magnitude, and further it indicates that selective breeding within self-fertilised generations of a homogeneous race is capable of modifying that race to a marked degree.

EXPLANATION OF PLATE I.

Figs. 1 and 2.—Pelorism of maximum intensity; grade 100°. Corollas absent, sessile anthers.

Figs. 3 and 4.—Perfect pelorism, grade 100°. Corollas joined along their split edges forming a complete saucer. Stamens with filaments.

Fig. 5.—Peloric flower of side-axis; the axis terminates in an ovary.

Fig. 6.—Pelorism of grade 100°. Numerous flowers fused irregularly forming a rosette, the axis has grown through the crown.

Figs. 7 and 8.—Incomplete pelorism of main axes, grade 75°. A spiral bending often occurs.

Fig. 9.—Faintly defined pelorism. When such occurred on the lateral axes the plant was said to possess a grade of 25°. Side view, and view from above.

Fig. 10.—Flowering axis of a conspicuous sport in which practically all the corollas are completely split longitudinally into four elongated blades. Nature of inheritance obscure.

The photographs were kindly taken by Dr Conrad Akerman.

ON POLYCHORIC COEFFICIENTS OF CORRELATION.

By KARL PEARSON, F.R.S. and EGON S. PEARSON.

(1) One of the difficulties which are constantly recurring in statistical practice is that of the correlation or contingency table in which the two variates are classified in broad categories. We may indeed proceed by the method of mean square contingency and correct for the grouping of both variates by the class index corrections on the assumption that the marginal totals for both variates may be assumed to follow approximately normal distributions. Such a procedure gives reasonable satisfactory results*, provided the marginal totals are not in very unequal groupings and the correlation is not intense (say, ·85 and above). The polychoric table has been discussed by Ritchie-Scott and he has described a method of reaching a polychoric coefficient of correlation from the weighted mean of the possible tetrachoric values†. Such a process is, however, so laborious that it can hardly establish itself in practice. From the theoretical standpoint, however, Ritchie-Scott's paper was of great interest (i) as guiding us by the size of the probable errors to discriminate between the valuable and worthless dichotomies in tetrachoric determinations of the correlation‡, (ii) as providing standard values by which those obtained by other procedures could be directly tested.

We shall endeavour to reach in this paper another form of polychoric coefficient,—that is a correlation coefficient which does use all the information given in a polychoric table,—but which requires less analysis than Ritchie-Scott's weighted mean coefficient. Thus what may be lost in exactness will possibly be repaid by practical efficiency. There is another point also of very considerable illustrative importance; we desire wherever the data are suitable actually to exhibit in the form of a graph the relation between the two variates. This should be possible in the case of a polychoric table, and in the past has frequently been done by approximate methods of more or less validity.

We can indeed take such methods as our present starting point as they will directly indicate to the reader our line of approach.

We start with the hypothesis that the marginal totals of our polychoric table can be represented on a normal scale. This is no great assumption in itself. If a true quantitative scale ever becomes available it can be attached at once and with little trouble to the normal scale. To exhibit a variate on a normal scale makes

* By "reasonably satisfactory results," we mean that in cases which can be directly checked by the product moment method the difference is within the range of practical insignificance as judged by probable error.

† *Biometrika*, Vol. XII. pp. 93—133.

‡ Thus in a 3×3 table it is possible for two of the corner dichotomies, i.e. those unassociated with the diagonal in the sense of the correlation, to have even *negative* weights, so that they should be omitted in finding the mean.

no greater assumption than when we exhibit a pressure-volume curve as a straight line by using a logarithmic scale.

Now let the polychoric table be such that in the population N under discussion, the sth category of the first variate A contains $n_{s\cdot}$ individuals and the s'th category of the second variate B contains $n_{\cdot s'}$ individuals, while the number of individuals who combine in the population N the sth category of A and the s'th category of B is $n_{ss'}$.

Now when we proceed to exhibit the categories of the A-variate on a normal scale, the process will give us two important quantities:

(*a*) We shall have the ratio of abscissa to standard deviation at the dichotomy between each pair of broad categories.

If $n_{1\cdot}$, $n_{2\cdot}$, $n_{3\cdot}$, ... $n_{s\cdot}$, ... be the frequencies of the A-variate for the several categories the values of the ratios of abscissae to standard deviation will be specified as

$$-\infty, \quad h_1, \quad h_2, \quad h_3, \quad h_4, \ldots h_{s-1}, \quad h_s, \ldots +\infty.$$

Here h_{s-1}, h_s are the values on either side of the category $n_{s\cdot}$ and if there be q categories, $n_{\cdot 1}$ is bounded by h_0 or $-\infty$ and h_1, while $n_{\cdot q}$ is bounded by h_{q-1} and h_q or $+\infty$. The lower h's will have negative and the upper positive signs and the greatest care must be taken to see that the proper signs are given to the values of h. Similarly if the frequencies of the various categories of the B-variate be

$$n_{\cdot 1}, \quad n_{\cdot 2}, \quad n_{\cdot 3}, \ldots n_{\cdot s'}, \ldots,$$

the values of the ratios of ordinates to standard deviation will be represented by

$$-\infty, \quad k_1, \quad k_2, \quad k_3, \ldots k_{s'-1}, \quad k_{s'}, \ldots k_{q'}, +\infty,$$

where $k_{s'-1}$ and $k_{s'}$ give the dichotomies on either side of $n_{\cdot s'}$.

We may consider the coordinate at the back of the variate A when represented on a normal scale to be x', the origin being taken at the mean on the normal scale. Hence if the standard deviation be σ_x, we shall find it convenient to write the absolute normal abscissae

$$x' = \sigma_x x, \quad h_s' = \sigma_x h_s.$$

Similarly we take y' for the coordinate at the back of the variate B, measured from the mean, and write:

$$y' = \sigma_y y, \quad k_s' = \sigma_y k_s,$$

where σ_y is the standard deviation of B. Clearly until a quantitative scale has been determined we shall know h, k, x, y but not h', k', x', y', σ_x and σ_y.

(*b*) We shall determine the ratio of abscissa to standard deviation, or the ratio of ordinate to standard deviation of the centroids or means of the groups $n_{s\cdot}$ and $n_{\cdot s'}$.

Let $$H_s = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h_s^2}, \qquad K_{s'} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k_{s'}^2},$$

then the means of the categories $n_{\cdot s}$ and $n_{s'\cdot}$ are determined by

$$\bar{h}_s = (H_{s-1} - H_s) \Big/ \frac{n_{s\cdot}}{N}, \qquad \bar{k}_{s'} = (K_{s-1} - K_s) \Big/ \frac{n_{\cdot s'}}{N} \quad \ldots\ldots\ldots\ldots\text{(i)}$$

respectively. The numerical values of $\bar{h}_s$ and $\bar{k}_{s'}$ can be easily ascertained from the table published recently of ordinates of normal curve to permilles of area*. Care must be taken in every case to give the correct sign to $\bar{h}_s$ and $\bar{k}_{s'}$.

Now if there were no correlation, $\bar{h}_s$ and $\bar{k}_{s'}$ combined would give the mean of the group $n_{ss'}$, and they give a fair approximation to the result if there are numerous categories, that is if the range of the categories be small.

The correlation found from these marginal centroids would then be

$$r_c = S(n_{ss'}\bar{h}_s\bar{k}_{s'})/N \quad \text{.....(ii),}$$

but as Ritchie-Scott has shown† this r_c diverges much more than r_ϕ the mean square contingency value from the true correlation, and considerably more than the tetrachoric or polychoric coefficients do. The reason for this is clear and was pointed out by one of us in 1913‡. Namely $\bar{h}_s$ and $\bar{k}_{s'}$ do not give the coordinates of the mean of $n_{ss'}$. In fact $n_{ss'}\bar{h}_s\bar{k}_{s'}$ is not the contribution of $n_{ss'}$ to the product-moment.

We propose in the present paper to give first the actual contributions of $n_{ss'}$ to the means and product-moments of the two variates and then to apply these results in order to obtain (*a*) a polychoric coefficient, and (*b*) a graph of the relation of the two variates.

The essential assumptions that will be made are the following:

(i) The marginal totals having been reduced to a normal scale, and the correlation being supposed to be r, we shall calculate what the contents of the sth-s'th cell would be on the assumption that the frequency surface is the normal surface represented by the given correlation and the marginal totals reduced to normal scales. We shall further calculate the x-moment, the y-moment and the xy product-moment of the sth-s'th cell on the same hypothesis.

(ii) From these data we shall determine the most suitable value to give to r, so that the actually observed frequencies differ least from those that would be given by such a correlation surface. We shall also obtain a formula for calculating the mean value of y for the array of B-variates, $n_{s\cdot}$ in number, which corresponds to the sth category of A. We shall thus be in a position to plot the regression line of B on A and test at the same time the closeness with which it fits the thus calculated array means, both variates being represented on a normal scale.

We shall write the real coefficient of correlation of the population r, the coefficient as found from a single sth-s'th cell, as $r_{ss'}$, and those found from the $n_{s\cdot}$ and $n_{\cdot s'}$ arrays as $r_{s\cdot}$ and $r_{\cdot s'}$ respectively.

$\bar{h}_{ss'}$, $\bar{k}_{ss'}$ will be the A- and B-variate means of the sth-s'th cell and $\pi_{ss'}$ the product-moment, per unit of the population, of the frequency in the sth-s'th cell about the mean axes as determined from the marginal totals on the normal scale.

* See *Biometrika*, Vol. XIII. pp. 426–8.
† *Biometrika*, Vol. XII. p. 122.
‡ *Biometrika*, Vol. IX. p. 138.

(2) The developments we require involve the use of the tetrachoric functions. The tetrachoric function of the order t is given by*

$$\tau_t = \frac{1}{\sqrt{t!}}\left(-\frac{d}{dx}\right)^{t-1}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \quad \text{.....(iii).}$$

The tetrachoric functions τ_1 to τ_6 are tabled for *positive* values of x in *Tables for Statisticians and Biometricians*† to five decimal places. For *negative* values of x tetrachoric functions of an odd order remain unchanged, but those of an even order must have their sign as given in the tables reversed.

It will frequently be needful to take the difference of the tetrachoric functions at the boundaries of a marginal category. Thus if $\tau_t(h)$ denotes the value of the tetrachoric function for $x = h$, we shall need for the sth marginal total

$$\tau_t(h_s) - \tau_t(h_{s-1}).$$

This difference we shall write, for brevity,

$$\Im_s \tau_t,$$

and in obtaining its numerical value from tables of the tetrachoric functions it is essential to remember that s (or s') is supposed to increase in the positive direction of the axis of x (or y), and that when h (or k) is negative attention must be paid to changing the sign of the tabled value of τ_t, if t be even.

The formula for determining the successive tetrachoric functions for a given value of x is

$$\tau_t = xp_t\tau_{t-1} - q_t\tau_{t-2} \quad \text{.....(iv),}$$

where p_t and q_t are given by the following table:

t	p_t	q_t	t	p_t	q_t
2	·707,1068	·000,0000	14	·267,2612	·889,4990
3	·577,3503	·408,2483	15	·258,1989	·897,0851
4	·500,0000	·577,3503	16	·250,0000	·903,6962
5	·447,2136	·670,8204	17	·242,5356	·909,5085 +
6	·408,2483	·730,2968	18	·235,7023	·914,6592
7	·377,9645	·771,5168	19	·229,4157	·919,2547
8	·353,5534	·801,7838	20	·223,6068	·923,3804
9	·333,3333	·824,9578	21	·218,2179	·927,1051
10	·316,2278	·843,2740	22	·213,2007	·930,4842
11	·301,5113	·858,1163	23	·208,5144	·933,5637
12	·288,6751	·870,3880	24	·204,1241	·936,3819
13	·277,3501	·880,7047	25	·200,0000	·938,9709

Since $\tau_1 = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$, it can be found at once from the tables for the ordinates of the normal curve, and will indeed have been computed at each division in order

* The reasons why the tetrachoric functions are tabled with the factor $1/\sqrt{t!}$ are: (*a*) because this factor greatly simplifies our formulae and (*b*) because a factor of some such order is essential, if we are to have manageable tabulated values. As a matter of fact the factor chosen reduces all tetrachoric functions to numerical values lying between 0 and 1.

† Cambridge University Press, pp. 42—51.

to determine $\bar{h}_{s.}$ and $\bar{h}_{.s'}$. It is then often simpler to work directly with (iv) rather than interpolate into the tabled values of the functions.

In an earlier paper* dealing with the tetrachoric functions one of us has shown that if

$$z = \frac{N}{2\pi\sqrt{(1-r^2)}}\, e^{-\frac{1}{2}\frac{x^2-2rxy+y^2}{1-r^2}}$$

be the equation to a normal correlation surface the variates being measured in the standard deviations as units, then

$$z/N = \tau_1\tau_1' + 2r\tau_2\tau_2' + 3r^2\tau_3\tau_3' + \ldots + (t+1)\, r^t\tau_{t+1}\tau'_{t+1} + \ldots \quad \ldots\ldots(\text{v}),$$

where $\tau_t = \tau_t(x)$ and $\tau_t' = \tau_t(y)$.

Now in order to proceed further it is needful to determine the following integrals:

$$\int_{h_{s-1}}^{h_s} \tau_t dx, \qquad \int_{h_{s-1}}^{h_s} x\tau_t dx.$$

We can determine these by using (iii) after in the second case integrating by parts. We have:

$$\begin{aligned}
\int_{h_{s-1}}^{h_s} \tau_t dx &= \frac{1}{\sqrt{t!}} \int_{h_{s-1}}^{h_s} \left(-\frac{d}{dx}\right)^{t-1} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} dx \\
&= \left[-\frac{1}{\sqrt{t!}} \left(-\frac{d}{dx}\right)^{t-2} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\right]_{h_{s-1}}^{h_s} \\
&= -\frac{1}{\sqrt{t}}\, \vartheta_s \tau_{t-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\text{vi}).
\end{aligned}$$

Again:

$$\begin{aligned}
\int_{h_{s-1}}^{h_s} x\tau_t dx &= \int_{h_{s-1}}^{h_s} \frac{1}{\sqrt{t!}}\, x \left(-\frac{d}{dx}\right)^{t-1} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} dx \\
&= \left[\frac{1}{\sqrt{t!}}(-x)\left(-\frac{d}{dx}\right)^{t-2} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\right]_{h_{s-1}}^{h_s} \\
&\qquad + \frac{1}{\sqrt{t!}} \int_{h_{s-1}}^{h_s} \left(-\frac{d}{dx}\right)^{t-2} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} dx \\
&= -\left[\frac{1}{\sqrt{t}}\, \tau_{t-1}\, x\right]_{h_{s-1}}^{h_s} + \frac{1}{\sqrt{t}} \int_{h_{s-1}}^{h_s} \tau_{t-1} dx \\
&= \left[-\frac{1}{\sqrt{t}}\, \tau_{t-1}\, x - \frac{1}{\sqrt{t(t-1)}}\, \tau_{t-2}\right]_{h_{s-1}}^{h_s} \\
&= -\frac{1}{\sqrt{t}} \left[\tau_{t-1}\, x + \frac{1}{\sqrt{t-1}}\, \tau_{t-2}\right]_{h_{s-1}}^{h_s} \quad \ldots\ldots\ldots\ldots(\text{vi}) \text{ bis.}
\end{aligned}$$

But by (iv):

$$\tau_{t-1}\, x = \frac{\tau_t + q_t\tau_{t-2}}{p_t},$$

where $p_t = 1/\sqrt{t}, \quad q_t = (t-2)/\sqrt{t(t-1)}.$

* *Phil. Trans.* Vol. 195 A, p. 4, Equation (xiv), with a slight change of notation. In that paper, $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} v_n \frac{1}{\sqrt{(n+1)!}}$ is written for τ_{n+1} and $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \frac{w_n}{\sqrt{(n+1)!}}$ for τ'_{n+1}.

Thus: $$\tau_{t-1}x + \frac{1}{\sqrt{t-1}}\tau_{t-2} = \sqrt{t}\,\tau_t + \sqrt{t-1}\,\tau_{t-2}.$$

Accordingly

$$\int_{h_{s-1}}^{h_s} x\tau_t dx = -\frac{1}{\sqrt{t}}\left[\sqrt{t}\,\tau_t + \sqrt{t-1}\,\tau_{t-2}\right]_{h_{s-1}}^{h_s}$$

$$= -\frac{1}{\sqrt{t}}(\sqrt{t}\,\vartheta_s\tau_t + \sqrt{t-1}\,\vartheta_s\tau_{t-2}) \quad \text{.....(vii).}$$

The latter form throws us back on $\vartheta_s\tau_t$ which will have to be calculated to determine the integral in (vi) for the successive values of t and s.

On the other hand a table of

$$T_{t-1} = \sqrt{t}\,\tau_t + \sqrt{t-1}\,\tau_{t-2} \quad \text{.....(viii)}$$

would be a convenient method of determining the integral and tables of T might be easily formed, say up to T_6.

In this case we may write (vii):

$$\int_{h_{s-1}}^{h_s} x\tau_t dx = -\frac{1}{\sqrt{t}}\vartheta_s T_{t-1} \quad \text{.....(ix).}$$

We are now in the position to compute all the requisite integrals we need; if we write $\bar{n}_{ss'}$ for the contents of the sth-s'th cell, then on the supposition that the surface is normal, has correlation r and follows the actual marginal frequencies, we have:

$$\frac{\bar{n}_{ss'}}{N} = \int_{h_{s-1}}^{h_s}\int_{k_{s-1}}^{k_s}\frac{z}{N}\,dxdy$$

$$= \vartheta_s\tau_0\vartheta_{s'}\tau_0' + r\vartheta_s\tau_1\vartheta_{s'}\tau_1' + r^2\vartheta_s\tau_2\vartheta_{s'}\tau_2' + \ldots + r^p\vartheta_s\tau_p\vartheta_{s'}\tau_p' + \ldots \quad \text{(x),}$$

$$\frac{\bar{n}_{ss'}}{N}\bar{h}_{ss'} = \int_{h_{s-1}}^{h_s}\int_{k_{s-1}}^{k_s}\frac{xz\,dxdy}{N} = \vartheta_s T_0\vartheta_{s'}\tau_0' + r\vartheta_s T_1\vartheta_{s'}\tau_1' + r^2\vartheta_s T_2\vartheta_{s'}\tau_2'$$

$$+ \ldots + r^p\vartheta_s T_p\vartheta_{s'}\tau_p' + \ldots \quad \text{.....(xi),}$$

$$\frac{\bar{n}_{ss'}}{N}\bar{k}_{ss'} = \int_{h_{s-1}}^{h_s}\int_{k_{s-1}}^{k_s}\frac{yz\,dxdy}{N} = \vartheta_s\tau_0\vartheta_{s'}T_0' + r\vartheta_s\tau_1\vartheta_{s'}T_1' + r^2\vartheta_s\tau_2\vartheta_{s'}T_2'$$

$$+ \ldots + r^p\vartheta_s\tau_p\vartheta_{s'}T_p' + \ldots \quad \text{.....(xii),}$$

$$\frac{\bar{n}_{ss'}}{N}\pi_{ss'} = \int_{h_{s-1}}^{h_s}\int_{k_{s-1}}^{k_s}\frac{xyz\,dxdy}{N} = \vartheta_s T_0\vartheta_{s'}T_0' + r\vartheta_s T_1\vartheta_{s'}T_1' + r^2\vartheta_s T_2\vartheta_{s'}T_2'$$

$$+ \ldots + r^p\vartheta_s T_p\vartheta_{s'}T_p' + \ldots \quad \text{.....(xiii).}$$

It is desirable to say a few words about the functions τ_0 and T_0 which may at first present difficulties to the reader. $-\tau_0$ clearly stands for the integral

$$\int_{h_{s-1}}^{h_s}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx, \text{ i.e. } \int_{h_{s-1}}^{h_s}\tau_1 dx,$$

and is therefore simply $n_{s\cdot}/N$.

Similarly $-\tau_0' = n_{\cdot s'}/N$.

Next clearly $-\Im_s T_0$ stands for

$$\int_{h_{s-1}}^{h_s} \tau_1 x\,dx = \int_{h_{s-1}}^{h_s} \frac{1}{\sqrt{2\pi}} x e^{-\frac{1}{2}x^2}\,dx$$
$$= -\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right]_{h_{s-1}}^{h_s}$$
$$= -\Im_s \tau_1,$$

or
$$\Im_s T_0 = \Im_s \tau_1,$$

which is precisely the value given by (viii).

Thus (viii) is shown to be correct even for this special case although a form like (vi) bis through which it is reached shows difficulties.

Similarly $\Im_s T_0' = \Im_s \tau_1'$*.

The remainder of the τ's knowing τ_0 and τ_1 come directly from (iv) and the T's are always given by (viii).

Now it is clear that (x) to (xiii) provide a large number of ways of determining r. We might find r, i.e. $r_{ss'}$, from the single cell by writing in (x) $n_{ss'}$ for $\bar{n}_{ss'}$. Or we may find

$$\bar{h}_{s\cdot} = \frac{1}{n_{s\cdot}} \mathop{S}_{s'} (n_{ss'} \bar{h}_{ss'})$$
$$= \frac{N}{n_{s\cdot}} \mathop{S}_{s'} \left(\frac{n_{ss'}}{\bar{n}_{ss'}} \{\Im_s T_0 \Im_{s'} \tau_0' + r \Im_{s'} T_1 \Im_{s'} \tau_1' + \ldots + r^p \Im_s T_p \Im_{s'} \tau_p' + \ldots\}\right) \quad \text{(xiv)},$$

where $\bar{n}_{ss'}$ is given by (x). But $\bar{h}_{s\cdot}$ is the known centroid of the $n_{s\cdot}$ marginal total, and accordingly the above is an equation to find r, i.e. $r_{s'}$, from a given column of the table.

If we use this value of $r_{s\cdot}$ in (x) and (xii) to find $\bar{n}_{ss'}$, and $\bar{k}_{ss'}$, we obtain the theoretical cell frequency and y-mean of the cell as found from a column.

Now sum $\bar{k}_{ss'}$ for every value of s' and we find $\bar{k}_{s\cdot}$ the y mean of a column depending on the data as found from the column, i.e.

$$\bar{k}_{s\cdot} = \frac{N}{n_{s\cdot}} \mathop{S}_{s'} \left(\frac{n_{ss'}}{\bar{n}_{ss'}} \{\Im_s \tau_0 \Im_{s'} T_0' + r \Im_s \tau_1 \Im_{s'} T_1' + \ldots + r^p \Im_s \tau_p \Im_{s'} T_p' + \ldots\}\right) \quad \text{(xv)},$$

where $n_{ss'}$ is the observed cell frequency and $\bar{n}_{ss'}$ the frequency found by (x) when we insert the value of r as found from (xiv). We are thus in a position theoretically to determine on a normal scale the mean of a column from the correlation actually determined from that column. This would be the ideal method of determining the mean of a row or column; but it would involve a great deal of hard work, as with the two regression curves we should need to find r for every row and column by an equation of a high order. Hence in most cases we are likely to content ourselves by finding r for the whole table and then use this value in (x) to determine $\bar{n}_{ss'}$ and in (xv) to find the mean of the array. $\bar{k}_{s\cdot}$ plotted to the known $\bar{h}_{s\cdot}$ on the normal scale will give the regression curve.

* We can thus take $T_0 = \tau_1$ and $T_0' = \tau_1'$.

The question now arises as to the manner in which we can find r for the whole table most effectively.

Clearly we might assume the product-moment components from (xiii) and sum for all cells. We should have

$$\sum_{s,s'}\left(\frac{n_{ss'}\pi_{ss'}}{N}\right)=r,$$

since the coordinates are measured from the means in terms of the standard deviations as units.

Hence substituting from (xiii) we have:

$$r=\underset{s,s'}{S}\left(\frac{n_{ss'}}{\bar{n}_{ss'}}\{\mathfrak{H}_sT_0\mathfrak{H}_{s'}T_0'+r\mathfrak{H}_sT_1\mathfrak{H}_{s'}T_1'+\ldots+r^p\mathfrak{H}_sT_p\mathfrak{H}_{s'}T_p'+\ldots\}\right)\quad\text{(xvi)}.$$

Here $\bar{n}_{ss'}$ must be substituted from (x) and we have finally

$$r=\underset{s,s}{S}\left(\frac{n_{ss'}}{N}\left\{\frac{\mathfrak{H}_sT_0\mathfrak{H}_{s'}T_0'+r\mathfrak{H}_sT_1\mathfrak{H}_{s'}T_1'+\ldots+r^p\mathfrak{H}_sT_p\mathfrak{H}_{s'}T_p'+\ldots}{\mathfrak{H}_s\tau_0\mathfrak{H}_{s'}\tau_0'+r\mathfrak{H}_s\tau_1\mathfrak{H}_{s'}\tau_1'+\ldots+r^p\mathfrak{H}_s\tau_p\mathfrak{H}_{s'}\tau_p'+\ldots}\right\}\right)\quad\text{(xvi) bis.}$$

This equation based upon the product-moment method of finding r is clearly likely to be very complicated, and although it can be proved that the product-moment method is the "best" method of finding r when we are dealing with a series of quantitatively measured individuals, we have no certainty that it is the best method in the present case of broad categories. It may indeed be questioned whether another method now to be considered cannot be shown to be better or at least equally efficacious.

Let us consider for a moment what we have in view. We observe $n_{ss'}$ as the frequency of the sth-s'th cell; we find that with a given correlation r the frequency of this cell would be $\bar{n}_{ss'}$ on the assumption that the frequency surface is the normal frequency surface corresponding to the observed marginal totals. Accordingly the most probable value to give to r would be that which made

$$\chi^2=\underset{s,s'}{S}\frac{(\bar{n}_{ss'}-n_{ss'})^2}{\bar{n}_{ss'}}=\text{minimum},$$

or, what is the same thing,

$$\underset{s,s'}{S}\left(\frac{n^2_{ss'}}{\bar{n}_{ss'}}\right)=\text{minimum}.$$

This leads us, differentiating with regard to r, to

$$\underset{s,s'}{S}\left\{\left(\frac{n_{ss'}}{\bar{n}_{ss'}}\right)^2\frac{d\bar{n}_{ss'}}{dr}\right\}=0,$$

or, writing at length, our equation for r is:

$$\underset{s,s'}{S}\left\{\left(\frac{n_{ss'}}{N}\right)^2\frac{\mathfrak{H}_s\tau_1\mathfrak{H}_{s'}\tau_1'+2r\mathfrak{H}_s\tau_2\mathfrak{H}_{s'}\tau_2'+\ldots+pr^{p-1}\mathfrak{H}_s\tau_p\mathfrak{H}_{s'}\tau_p'+\ldots}{(\mathfrak{H}_s\tau_0\mathfrak{H}_{s'}\tau_0'+r\mathfrak{H}_s\tau_1\mathfrak{H}_{s'}\tau_1'+r^2\mathfrak{H}_s\tau_2\mathfrak{H}_{s'}\tau_2'+\ldots+r^p\mathfrak{H}_s\tau_p\mathfrak{H}_{s'}\tau_p'+\ldots)^2}\right\}=0\quad\text{(xvii)}.$$

Neither (xvi) nor (xvii) are very readily solved. Probably the easiest way will be to obtain an approximate value of r by existing methods either from a good fourfold table, or from contingency, and then evaluate (xvi) or (xvii) for values of r, one well above and one well below this result, so that the real value of r lies

between the two. A linear interpolation will probably suffice in most cases to determine r with sufficient accuracy.

It will be observed that what we are trying to do is to fit a normal correlation surface to a series of cell frequencies. We may do this by equating product-moments, or actual cell frequencies properly weighted. The factors $\frac{n_{ss'}}{\bar{n}_{ss'}}$ and $\left(\frac{n_{ss'}}{\bar{n}_{ss'}}\right)^2$ come into our equations as a form of weights. When $n_{ss'}$ is small as compared with $\bar{n}_{ss'}$ that cell will contribute less to the general equations for r, and when $n_{ss'}$ is large as compared with $\bar{n}_{ss'}$, the contribution will be considerable. If the observed results were closely normal then $n_{ss'}$ would be nearly $\bar{n}_{ss'}$. If we might assume the differences of $n_{ss'}$ and $\bar{n}_{ss'}$ so small as to be negligible we should have:

$$r = \underset{s,s'}{S}(\mathfrak{H}_s T_0 \mathfrak{H}_{s'} T_0' + r\mathfrak{H}_s T_1 \mathfrak{H}_{s'} T_1' + \ldots + r^p \mathfrak{H}_s T_p \mathfrak{H}_{s'} T_p' + \ldots \quad \text{(xvi) ter,}$$

and

$$0 = \underset{s,s'}{S}(\mathfrak{H}_s \tau_1 \mathfrak{H}_{s'} \tau_1' + 2r\mathfrak{H}_s \tau_2 \mathfrak{H}_{s'} \tau_2' + \ldots + pr^{p-1}\mathfrak{H}_s \tau_p \mathfrak{H}_{s'} \tau_p' + \ldots \text{(xvii) bis,}$$

instead of (xvi) bis and (xvii). These equations it will be found are identically satisfied. Hence our values for r from (xvi) and (xvii) depend on $\bar{n}_{ss'}$ differing from $n_{ss'}$.

(3) We now proceed to illustrate the application of these results.

Stature of Father and Son.

The following table gives a correlation table for the inheritance of stature in Father and Son made up in broad categories corresponding to eye-colour groups*. Upon this material we shall be able to test our correlations and our graph against those found by definite numerical groupings.

Stature of Father (Broad Categories).

Stature of Son (Broad Categories).	1	2	3	4	5	6	7	Totals
1′	4	22	7	—	1	—	—	34
2′	23	154	84	26	8	6	—	301
3′	8	87	75	66	22	24	2	284
4′	1	29	36	37	14	14	6	137
5′	—	18	27	26	11	18	5	105
6′	—	9	26	19	7	29	8	98
7′	—	3	9	6	6	10	7	41
Totals	36	322	264	180	69	101	28	1000

The positive direction of x is from left to right and of y vertically downwards. It will suffice to take the τ's to five decimal figures but it will be needful to go further with the τ's if the T's are to be taken correctly to five figures from (viii). The general reduction formula for the T's is:

$$T_t(x) = \frac{T_{t-1}(x)(t-2)\sqrt{t-1}\,x^3 - T_{t-2}(x)(t-3)((t-1)x^2+1)}{\sqrt{t(t-1)}(x^2(t-2)+1)} \quad \ldots\text{(xviii),}$$

or,

$$= \frac{q_t}{x^2(t-2)+1}\left\{\frac{x^3}{p_{t-1}}T_{t-1}(x) - (x^2(t-1)+1)\frac{t-3}{t-2}T_{t-2}(x)\right\} \quad \text{(xviii) bis.}$$

* See *Biometrika*, Vol. IX. p. 220.

Hence if T_0 and T_1 be found accurately the remaining T's can be determined as accurately as we please without reference to the τ's.

But,
$$T_0 = \tau_1 = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{.....(xix),}$$

$$T_1 = \sqrt{2}\,\tau_2 + \tau_0 = \frac{x}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \quad \text{.....(xx).}$$

Hence the tables of ordinates and areas of the normal curve readily provide T_0 and T_1 to seven decimal places, and (xviii) provides the higher T's. These were cut down to five figures and an approximate check on their values obtained by (viii).

As a matter of fact if r is of the order ·50 we cannot hope to obtain more than three figure accuracy in r without going to higher τ- and T-functions than the sixth, especially when using (xvii). But three figures in the correlation are usually adequate and the labour of computing is much increased if higher functions are used. Such must, however, be used if the correlation be sensibly higher than ·50.

The following table gives the $\frac{1}{2}(1+\alpha)$'s, h's, H's, $\bar{x}$'s, τ's, $\vartheta\tau$'s, T's, and ϑT's for the x-variate.

TABLE I.

$\frac{1}{2}(1+\alpha)$	0	·036	·358	·622	·802	·871	·972	1·000
h	$-\infty$	−1·79912	−·36381	+·31074	+·84879	+1·13113	+1·91104	$+\infty$
$H=\tau_1$	0	·07908	·37340	·38014	·27827	·21042	·06425	0
$\bar{x}_s=\frac{H_{s-1}-H_s}{\frac{1}{2}(a_s-a_{s-1})}$	−2·19667	−·91404	−·02553	+·56594	+·98333	+1·44723	+2·29464	
τ_0	0	−·036	−·358	−·622	−·802	−·871	−·972	−1
$\tau_1=T_0$	0	+·07908	+·37340	+·38014	+·27827	+·21042	+·06425	0
τ_2	0	−·10060	−·09606	+·08353	+·16701	+·16830	+·08682	0
τ_3	0	+·07221	−·13226	−·14021	−·03176	+·02401	+·06952	0
τ_4	0	−·00688	+·07952	−·07001	−·10990	−·08359	+·01634	0
τ_5	0	−·04291	+·07579	+·08432	−·02041	−·05839	−·03270	0
τ_6	0	+·03654	−·06933	+·06182	+·07319	+·03408	−·03744	0
$\vartheta\tau_0$	−·036	−·322	−·264	−·180	−·069	−·101	−·028	
$\vartheta\tau_1$	+·07908	+·29432	+·00674	−·10187	−·06785	−·14617	−·06425	
$\vartheta\tau_2$	−·10060	+·00454	+·17959	+·08348	+·00129	−·08148	−·08682	
$\vartheta\tau_3$	+·07221	−·20447	−·00795	+·10845	+·05577	+·04555	−·06956	
$\vartheta\tau_4$	−·00688	+·08640	−·14953	−·03989	+·02631	+·09993	−·01634	
$\vartheta\tau_5$	−·04291	+·11870	+·00853	−·10473	−·03798	+·02569	+·03270	
$\vartheta\tau_6$	+·03654	−·10587	+·13115	+·01137	−·03911	−·07152	+·03744	
T_1	0	−·17827	−·49385	−·50388	−·56581	−·63299	−·84922	−1
T_2	0	+·23690	+·29898	+·29475	+·33853	+·33915	+·21135	0
T_3	0	−·18799	−·00734	+·00466	+·06947	+·12432	+·18307	0
T_4	0	+·04848	−·09506	−·09186	−·10916	−·08255	+·06601	0
T_5	0	+·07412	+·07799	−·00511	−·06648	−·10343	−·05518	0
T_6	0	−·08325	+·05616	+·05364	+·05379	+·01471	−·08491	0
ϑT_0	+·07908	+·29432	+·00674	−·10187	−·06786	−·14617	−·06425	
ϑT_1	−·17827	−·31558	−·01003	−·06193	−·06718	−·21622	−·15078	
ϑT_2	+·23690	+·06208	−·00422	+·04377	+·00063	−·12780	−·21135	
ϑT_3	−·18799	+·18065	+·01200	+·06481	+·05485	+·05874	−·18307	
ϑT_4	+·04848	−·14354	+·00320	−·01731	+·02662	+·14856	−·06601	
ϑT_5	+·07412	−·06613	−·01310	−·06137	−·03695	+·04825	+·05518	
ϑT_6	−·08325	+·13941	−·00253	+·00015	−·03907	−·09962	+·08491	

The following table gives the corresponding quantities $\frac{1}{2}(1+\alpha')$'s, k's, K's, $\bar{y}''$s, τ''s, $\vartheta\tau_1''$s T''s and $\vartheta T''$s for the y-variate.

TABLE II.

$\frac{1}{2}(1+\alpha')$	0	·034	·335	·619	·756	·861	·959	1·000
k	$-\infty$	−1·82501	−·42615	+·30286	+·69349	+1·08482	+1·73920	$+\infty$
$K=\tau_1'$	0	+·07545	+·36431	+·38106	+·31367	+·22149	+·08792	0
$\bar{y}_s=\frac{K_{s-1}-K_s}{\frac{1}{2}(\alpha_s'-\alpha_{s-1}')}$	−2·21916	·95968	−·05896	+·49188	+·87789	+1·36301	+2·14436	
τ_0'	0	−·034	−·335	−·619	−·756	−·861	−·959	−1
$\tau_1'=T_0'$	0	+·07545	+·36431	+·38106	+·31367	+·22149	+·08792	0
τ_2'	0	−·09737	−·10978	+·08160	+·15382	+·16990	+·10812	0
τ_3'	0	+·07179	−·12172	−·14130	−·06647	+·01599	+·07268	0
τ_4'	0	−·00929	+·08932	−·06851	−·11185	−·08942	+·00077	0
τ_5'	0	−·04057	+·06463	+·08551	+·00990	−·05411	−·04815	0
τ_6'	0	+·03702	−·07647	+·06061	+·08449	+·04134	−·03475	0
$\vartheta\tau_0'$	−·034	−·301	−·284	−·137	−·105	−·098	−·041	
$\vartheta\tau_1'$	+·07545	+·28886	+·01675	−·06739	−·09218	−·13357	−·08792	
$\vartheta\tau_2'$	−·09737	−·01241	+·19138	+·07222	+·01608	−·06178	−·10812	
$\vartheta\tau_3'$	+·07179	−·19351	−·01958	+·07483	+·08246	+·05669	−·07268	
$\vartheta\tau_4'$	−·00929	+·09861	−·15783	−·04334	+·02243	+·09019	−·00077	
$\vartheta\tau_5'$	−·04057	+·10520	+·02088	−·07561	−·06401	+·00596	+·04815	
$\vartheta\tau_6'$	+·03702	−·11349	+·13708	+·02388	−·04315	−·07609	+·03475	
T_1'	0	−·17170	−·49025	−·50360	−·53847	−·62073	−·80610	−1
T_2'	0	+·23105	+·30439	+·29416	+·32847	+·34094	+·25021	0
T_3'	0	−·18723	−·01151	+·00432	+·04271	+·11544	+·18882	0
T_4'	0	+·05286	−·09892	−·09140	−·11081	−·08901	+·03768	0
T_5'	0	+·06989	+·01240	−·00474	−·04316	−·09869	−·08340	0
T_6'	0	−·08412	+·05897	+·05326	+·06263	+·02276	−·08010	0
$\vartheta T_0'$	+·07545	+·28886	+·01675	−·06739	−·09218	−·13358	−·08792	
$\vartheta T_1'$	−·17170	−·31855	−·01334	−·03488	−·08225	−·18537	−·19391	
$\vartheta T_2'$	+·23105	+·07334	−·01023	+·03431	+·01247	−·09072	−·25021	
$\vartheta T_3'$	−·18723	+·17572	+·01583	+·03839	+·07273	+·07338	−·18882	
$\vartheta T_4'$	+·05286	−·15178	+·00752	−·01941	+·02179	+·12670	−·03768	
$\vartheta T_5'$	+·06989	−·05749	−·01714	−·03842	−·05553	+·10529	+·08340	
$\vartheta T_6'$	−·08412	+·14309	−·00571	+·00937	−·03988	−·10286	+·08010	

From Tables I and II we can find from (x) the value of $\bar{n}_{ss}/N$ for any given value of r, and by equating $\bar{n}_{ss}/N$ to n_{ss}/N we should have an equation to determine the correlation r from that cell alone. The weighted mean of these 49 r's would be Ritchie-Scott's polychoric correlation coefficient. But the labour would be immense*.

We are now in a position to give the product of $\vartheta_s\tau_p\vartheta_{s'}\tau_p'$: see Table III, p. 138.

There are certain checks on the accuracy of this table, namely

$$S_{ss'}\vartheta_s\,\tau_p\,\vartheta_{s'}\,\tau_p' = 0 \text{ except for } p=0, \text{ when it} = 1.$$

* We are not underrating the large amount of arithmetic of the present process. It is not likely to be often repeated, and the sole purpose of publishing all these tables for an individual case is to impress the reader with that fact; while at the same time illustrating the actual numerical processes. The amount of arithmetic, great as it is, is relatively small compared with that of solving and weighting the resulting r's in the case of a 49-cell table.

TABLE III.

Values of $\vartheta_s \tau_p \vartheta_{s'} \tau_p'$.

s'	p	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$	p	s
	0	+·001,224	+·010,948	+·008,976	+·006,120	+·002,346	+·003,434	+·000,952	0	
	1	+·005,967	+·022,206	+·000,509	−·007,686	−·005,119	−·011,029	−·004,848	1	
	2	+·009,795	−·000,442	−·017,487	−·008,128	−·000,126	+·007,934	+·008,454	2	
1	3	+·005,184	−·014,679	−·000,571	+·007,786	+·004,004	+·003,270	−·004,994	3	1
	4	+·000,064	−·000,803	+·001,389	+·000,371	−·000,244	−·000,928	+·000,152	4	
	5	+·001,741	−·004,816	−·000,346	+·004,249	+·001,541	−·001,042	−·001,327	5	
	6	+·001,353	−·003,919	+·004,855	+·000,421	−·001,448	−·002,648	+·001,386	6	
	0	+·010,836	+·096,922	+·079,464	+·054,180	+·020,769	+·030,401	+·008,428	0	
	1	+·022,843	+·085,017	+·001,947	−·029,426	−·019,599	−·042,223	−·018,559	1	
	2	+·001,248	−·000,056	−·002,229	−·001,036	−·000,016	+·001,011	+·001,077	2	
2	3	−·013,973	+·039,567	+·001,538	−·020,986	−·010,792	−·008,814	+·013,461	3	2
	4	−·000,678	+·008,520	−·014,745	−·003,934	+·002,594	+·009,854	−·001,611	4	
	5	−·004,514	+·012,487	+·000,897	−·011,018	−·003,995	+·002,703	+·003,440	5	
	6	−·004,147	+·012,015	−·014,884	−·001,290	+·004,439	+·008,117	−·004,249	6	
	0	+·010,224	+·091,448	+·074,976	+·051,120	+·019,596	+·028,684	+·007,952	0	
	1	+·001,325	+·004,930	+·000,113	−·001,706	−·001,136	−·002,448	−·001,076	1	
	2	−·019,253	+·000,869	+·034,370	+·015,976	+·000,247	−·015,594	−·016,616	2	
3	3	−·001,414	+·004,004	+·000,156	−·002,123	−·001,092	−·000,892	+·001,362	3	3
	4	+·001,086	−·013,637	+·023,600	+·006,296	−·004,153	−·015,772	+·002,579	4	
	5	−·000,896	+·002,478	+·000,178	−·002,187	−·000,793	+·000,536	+·000,683	5	
	6	+·005,009	−·014,513	+·017,978	+·001,559	−·005,361	−·009,804	+·005,132	6	
	0	+·004,932	+·044,114	+·036,168	+·024,660	+·009,453	+·013,837	+·003,836	0	
	1	−·005,329	−·019,834	−·000,454	+·006,865	+·004,572	+·009,850	+·004,330	1	
	2	−·007,265	+·000,328	+·012,970	+·006,029	+·000,093	−·005,884	−·006,270	2	
4	3	+·005,403	−·015,300	−·000,595	+·008,115	+·004,173	+·003,409	−·005,205	3	4
	4	+·000,298	−·003,745	+·006,481	+·001,729	−·001,140	−·004,331	+·000,708	4	
	5	+·003,244	−·008,975	−·000,645	+·007,919	+·002,872	−·001,942	−·002,472	5	
	6	+·000,873	−·002,528	+·003,132	+·000,272	−·000,934	−·001,708	+·000,894	6	
	0	+·003,780	+·033,810	+·027,720	+·018,900	+·007,245	+·010,605	+·002,940	0	
	1	−·007,290	−·027,130	−·000,621	+·009,390	+·006,254	+·013,474	+·005,923	1	
	2	−·001,678	+·000,073	+·002,888	+·001,342	+·000,021	−·001,310	−·001,396	2	
5	3	+·005,954	−·016,861	−·000,656	+·008,943	+·004,599	+·003,756	−·005,736	3	5
	4	−·000,154	+·001,938	−·003,354	−·000,895	+·000,590	+·002,241	−·000,367	4	
	5	+·002,747	−·007,598	−·000,546	+·006,704	+·002,431	−·001,644	−·002,093	5	
	6	−·001,577	+·004,568	−·005,659	−·000,491	+·001,688	+·003,086	−·001,616	6	
	0	+·003,528	+·031,556	+·025,872	+·017,640	+·006,762	+·009,898	+·002,744	0	
	1	−·010,563	−·039,312	−·000,900	+·013,607	+·009,063	+·019,524	+·008,582	1	
	2	+·006,215	−·000,280	−·011,095	−·005,157	−·000,080	+·005,034	+·005,364	2	
6	3	+·004,094	−·011,591	−·000,451	+·006,148	+·003,162	+·002,582	−·003,943	3	6
	4	−·000,621	+·007,792	−·013,486	−·003,598	+·002,373	+·009,013	−·001,474	4	
	5	−·000,256	+·007,107	+·000,051	−·000,624	−·000,226	+·000,153	+·000,195	5	
	6	−·002,780	+·008,056	−·009,979	−·000,865	+·002,976	+·005,442	−·002,849	6	
	0	−·001,476	+·013,202	+·010,824	+·007,380	+·002,829	+·004,141	+·001,148	0	
	1	−·006,953	−·025,877	−·000,593	+·008,956	+·005,965	+·012,851	+·005,649	1	
	2	+·010,877	−·000,491	−·019,417	−·009,026	−·000,139	+·008,810	+·009,387	2	
7	3	−·005,248	+·014,861	+·000,578	−·007,882	−·004,053	−·003,311	+·005,056	3	7
	4	+·000,005	−·000,067	+·000,115	+·000,031	−·000,020	−·000,077	+·000,013	4	
	5	−·002,066	+·005,715	+·000,411	−·005,043	−·001,829	+·001,237	+·001,575	5	
	6	+·001,270	−·003,679	+·004,557	+·000,395	−·001,359	−·002,485	+·001,301	6	

Applying these tests we find:

$$S_{ss'}(\Im_s \tau_0 \Im_{s'} \tau_0') = \quad 1{\cdot}000{,}000, \quad S_{ss'}(\Im_s \tau_1 \Im_{s'} \tau_1') = +\ {\cdot}000{,}001,$$
$$S_{ss'}(\Im_s \tau_2 \Im_{s'} \tau_2') = +\ {\cdot}000{,}001, \quad S_{ss'}(\Im_s \tau_3 \Im_{s'} \tau_3') = +\ {\cdot}000{,}003,$$
$$S_{ss'}(\Im_s \tau_4 \Im_{s'} \tau_4') = -\ {\cdot}000{,}002, \quad S_{ss'}(\Im_s \tau_5 \Im_{s'} \tau_5') = +\ {\cdot}000{,}001,$$

and $$S_{ss'}(\Im_s \tau_6 \Im_{s'} \tau_6') = +\ {\cdot}000{,}002,$$

results as close as we should expect, when we take into account the fact that our $\Im\tau$'s were only to five figure accuracy, and our products to six.

The meaning of Table III should be quite intelligible; namely, for example:

$$\cdot 084 = \frac{n_{3,2}}{N} = {\cdot}079{,}464 + {\cdot}001{,}947\, r - {\cdot}002{,}229\, r^2$$
$$+ {\cdot}001{,}538 r^3 - {\cdot}014{,}745 r^4 + {\cdot}000{,}897 r^5 - {\cdot}014{,}884 r^6 + \ldots \quad \ldots(\text{xxi})$$

is the equation which will give the correlation coefficient r as deduced from the (3, 2) cell. If r be given any other value the right hand of the above expression is equal to the contents of the (3, 2) cell for a normal correlation surface of correlation coefficient r having the observed marginal totals.

Thus far the arithmetic is absolutely comparable with that needed for Ritchie-Scott's "polychoric r." We should have to solve the 49 equations, and then calculate—the stiffest part of the work—the probable errors of the 49 correlation coefficients which are the roots of these equations. Using these probable errors as our weighting data, we should find a mean coefficient. Our purpose is to replace the weighting and the solution of the 49 equations by the solution of a single equation. It will be noticed that both Ritchie-Scott's and our methods have an undesirable limitation, for we both assume the marginal totals to be those of the normal correlation surface. Actually in our case we ought to treat the marginal totals as unknown, or select $h_1, h_2, h_3, \ldots h_q, k_1, k_2, k_3, \ldots k_{q'}$ as well as r to give as closely as possible the observed frequencies. Now the τ's and consequently the T's and $\Im\tau$'s and $\Im T$'s all depend upon the h's and k's and the equations obtained by making

$$S_{ss'}\left(\frac{n^2_{ss'}}{\bar{n}_{ss'}}\right) = \text{minimum}$$

do not appear to lend themselves to any reasonably brief system of solutions. We were compelled therefore to introduce the admittedly limited form of solution, i.e. the determination of the best normal correlation surface subject to the restriction of its having the same marginal totals as the observed frequency surface. We consider this a practically necessary but none the less grave restriction.

We next proceeded to determine the value of $\bar{n}_{ss'}/n_{ss'}$ and $(\bar{n}_{ss'}/n_{ss'})^2$ for certain selected values of r in order to build up equation (xvii) and solve it by interpolation. The values chosen were: 0·45, 0·50 and 0·55. These cover the range within which we anticipate the solution of (xvii) for r will lie. We need also the value of the numerator in (xvii), i.e.

$$\nu_{ss'} = \Im_s \tau_1 \Im_{s'} \tau_1' + 2r \Im_s \tau_2 \Im_{s'} \tau_2' + 2r^2 \Im_s \tau_3 \Im_{s'} \tau_3' + \ldots,$$

for the same three values of r. These results are given in Table IV.

TABLE IV. *Values of* $(\bar{n}_{ss'}/n_{ss'})$, $(\bar{n}_{ss'}/n_{ss'})^2$ *and* $\nu_{ss'}$.

s'	Function		$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	$\bar{n}_{ss'}/n_{ss'}$	(a)	1·602,750	·879,954	·814,714	∞	·388,000	∞	∞
		(b)	1·846,000	·902,000	·705,571	∞	·266,000	∞	∞
		(c)	2·115,500	·916,409	·587,857	∞	·174,000	∞	∞
	$(\bar{n}_{ss'}/n_{ss'})^2$	(a)	2·568,806	·774,321	·663,759	∞	·150,544	∞	∞
		(b)	3·407,716	·813,604	·497,830	∞	·070,756	∞	∞
		(c)	4·475,340	·839,805	·345,576	∞	·030,276	∞	∞
	$\nu_{ss'}$	(a)	+ ·018,462	+ ·011,177	− ·014,603	− ·009,218	− ·002,733	− ·002,747	− ·000,336
		(b)	+ ·020,480	+ ·008,113	− ·015,910	− ·008,382	− ·002,154	− ·001,929	− ·000,218
		(c)	+ ·022,694	+ ·004,477	− ·017,013	− ·007,243	− ·001,519	− ·001,228	− ·000,168
2	$\bar{n}_{ss'}/n_{ss'}$	(a)	·867,348	·905,539	·944,250	1·478,500	1·379,000	1·887,333	∞
		(b)	·894,565	·944,630	·939,833	1·383,615	1·215,375	1·544,667	∞
		(c)	·915,130	·986,935	·933,333	1·278,500	1·043,500	1·213,333	∞
	$(\bar{n}_{ss'}/n_{ss'})^2$	(a)	·752,293	·820,001	·891,608	2·185,962	1·901,641	3·562,026	∞
		(b)	·800,247	·892,326	·883,286	1·914,390	1·477,136	2·385,996	∞
		(c)	·837,463	·974,041	·871,110	1·634,562	1·088,892	1·472,177	∞
	$\nu_{ss'}$	(a)	+ ·013,846	+ ·116,000	− ·005,963	− ·046,943	− ·025,552	− ·041,623	− ·009,764
		(b)	+ ·011,084	+ ·125,051	− ·009,011	− ·051,854	− ·026,828	− ·040,529	− ·007,913
		(c)	+ ·007,767	+ ·135,874	− ·013,006	− ·057,659	− ·028,171	− ·038,864	− ·005,940
3	$\bar{n}_{ss'}/n_{ss'}$	(a)	·857,750	1·075,552	1·108,280	·812,500	·854,818	·984,375	2·194,000
		(b)	·751,875	1·076,195	1·138,747	·823,410	·844,773	·930,333	1·846,500
		(c)	·635,750	1·075,448	1·175,027	·835,909	·831,636	·866,000	1·486,500
	$(\bar{n}_{ss'}/n_{ss'})^2$	(a)	·735,735	1·156,812	1·228,285	·660,156	·730,714	·968,994	4·813,636
		(b)	·565,316	1·158,196	1·296,745	·678,004	·713,641	·865,519	3·409,562
		(c)	·404,178	1·156,588	1·380,688	·698,744	·691,618	·749,956	2·209,682
	$\nu_{ss'}$	(a)	− ·016,095	+ ·002,075	+ ·041,770	+ ·013,402	− ·003,847	− ·023,749	− ·013,555
		(b)	− ·017,786	+ ·000,037	+ ·049,827	+ ·015,435	− ·005,038	− ·028,268	− ·014,205
		(c)	− ·019,311	− ·002,805	+ ·059,278	+ ·017,601	− ·006,601	− ·033,522	− ·014,539
4	$\bar{n}_{ss'}/n_{ss'}$	(a)	1·634,000	1·155,997	1·078,222	·808,892	·850,571	1·225,786	·671,833
		(b)	1·260,000	1·096,966	1·098,417	·837,135	·877,714	1·239,929	·627,333
		(c)	·917,000	1·030,862	1·121,944	·869,568	·907,429	1·250,000	·570,000
	$(\bar{n}_{ss'}/n_{ss'})^2$	(a)	2·669,956	1·336,098	1·162,563	·654,306	·723,471	1·502,551	·451,360
		(b)	1·587,600	1·203,334	1·206,520	·700,795	·770,382	1·537,424	·393,547
		(c)	·840,889	1·062,676	1·258,758	·756,149	·823,427	1·562,500	·324,900
	$\nu_{ss'}$	(a)	− ·007,715	− ·032,319	+ ·013,434	+ ·019,505	+ ·007,261	+ ·004,459	− ·004,625
		(b)	− ·007,215	− ·036,132	+ ·015,696	+ ·022,370	+ ·007,947	+ ·003,430	− ·006,095
		(c)	− ·006,471	− ·040,720	+ ·018,237	+ ·025,717	+ ·008,735	+ ·002,185	− ·007,680
5	$\bar{n}_{ss'}/n_{ss'}$	(a)	∞	1·114,278	1·028,556	·934,423	·960,545	·935,111	·946,600
		(b)	∞	1·006,167	1·027,185	·969,000	1·008,273	·978,944	·944,400
		(c)	∞	·890,389	1·024,148	1·007,692	1·061,727	1·025,111	·927,400
	$(\bar{n}_{ss'}/n_{ss'})^2$	(a)	∞	1·241,615	1·057,927	·873,146	·922,647	·874,433	·896,052
		(b)	∞	1·012,372	1·055,109	·938,961	1·016,614	·958,331	·891,891
		(c)	∞	·792,793	1·048,879	1·015,443	1·127,264	1·050,853	·860,071
	$\nu_{ss'}$	(a)	− ·004,797	− ·037,653	− ·000,381	+ ·017,025	+ ·009,967	+ ·015,398	+ ·000,440
		(b)	− ·003,957	− ·040,252	− ·001,134	+ ·018,995	+ ·011,095	+ ·016,166	− ·000,916
		(c)	− ·002,988	− ·043,158	− ·002,230	+ ·021,305	+ ·012,465	+ ·017,113	− ·002,508
6	$\bar{n}_{ss'}/n_{ss'}$	(a)	∞	1·461,333	·867,077	1·216,474	1·604,286	·701,931	·906,500
		(b)	∞	1·224,000	·830,576	1·245,526	1·693,714	·754,966	·969,125
		(c)	∞	·988,111	·786,077	1·273,789	1·791,000	·812,828	1·028,375
	$(\bar{n}_{ss'}/n_{ss'})^2$	(a)	∞	2·135,494	·751,823	1·479,809	2·573,734	·492,707	·821,742
		(b)	∞	1·498,176	·689,856	1·551,335	2·868,667	·569,974	·939,203
		(c)	∞	·976,363	·617,917	1·622,538	3·207,681	·660,689	1·057,555
	$\nu_{ss'}$	(a)	− ·003,069	− ·042,728	− ·017,170	+ ·011,165	+ ·012,060	+ ·029,542	+ ·010,202
		(b)	− ·002,189	− ·042,658	− ·020,931	+ ·010,905	+ ·013,028	+ ·032,069	+ ·009,778
		(c)	− ·001,381	− ·042,197	− ·025,479	+ ·010,572	+ ·014,219	+ ·035,116	+ ·009,152
7	$\bar{n}_{ss'}/n_{ss'}$	(a)	∞	·961,333	·747,556	1·462,667	·845,000	1·140,500	·870,286
		(b)	∞	·705,000	·648,556	1·411,167	·865,000	1·235,000	1·003,143
		(c)	∞	·491,000	·542,000	1·337,333	·877,000	1·331,000	1·150,286
	$(\bar{n}_{ss'}/n_{ss'})^2$	(a)	∞	·924,161	·558,840	2·139,395	·714,025	1·300,740	·757,398
		(b)	∞	·497,025	·420,625	1·991,392	·748,225	1·525,225	1·006,296
		(c)	∞	·241,081	·293,764	1·788,460	·769,129	1·771,561	1·323,158
	$\nu_{ss'}$	(a)	− ·000,633	− ·016,551	− ·017,086	− ·004,935	+ ·002,845	+ ·018,719	+ ·017,641
		(b)	− ·000,417	− ·014,160	− ·018,536	− ·007,468	+ ·001,950	+ ·019,060	+ ·019,571
		(c)	− ·000,309	− ·011,471	− ·019,787	− ·010,293	+ ·000,873	+ ·019,302	+ ·021,685

The values (*a*), (*b*), (*c*) refer respectively to $r=0{\cdot}45$, $0{\cdot}50$, $0{\cdot}55$.

Having obtained $(\bar{n}_{ss'}/n_{ss'})^2$ and $\nu_{ss'}$ for the trial values of r, it is only a matter of adding $\nu_{ss}/(\bar{n}_{ss'}/n_{ss'})^2$ for all values of s and s' on the machine in order to obtain:

$$u = S_{ss'}\{\nu_{ss'}/(\bar{n}_{ss'}/n_{ss'})^2\}.$$

The values obtained were:

$r=$	0·45	0·50	0·55
$u=$	+ ·157,074	+ ·012,276	− ·209,976

Whence by inverse interpolation* we find:

$$u = 0 \text{ for } r_p = \cdot 5034,$$

which is "polychoric r" as based upon Equation (xvii). We shall compare later the value for r as found by other processes. But the above value is clearly well in accord with the usual result for paternal correlation in man.

Table V gives the working values of $\nu_{ss'}/(\bar{n}_{ss'}/n_{ss'})^2$.

TABLE V. *Values of* $\nu_{ss'}/(\bar{n}_{ss'}/n_{ss'})^2$†.

s'	r	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	(*a*)	+ ·007,186	+ ·014,435	− ·022,001	0	− ·018,159	0	0
	(*b*)	+ 006,010	+ ·009,972	− ·031,958	0	− ·030,424	0	0
	(*c*)	+ ·005,071	+ ·005,331	− ·049,232	0	− ·050,132	0	0
2	(*a*)	+ ·018,405	+ ·141,463	− ·006,688	− ·021,475	− ·013,436	− ·011,685	0
	(*b*)	+ ·013,851	+ ·140,141	− ·010,202	− ·027,086	− ·018,162	− ·016,986	0
	(*c*)	+ ·009,274	+ ·139,495	− ·014,930	− ·035,275	− ·025,871	− ·026,399	0
3	(*a*)	− ·021,876	+ ·001,794	+ ·034,007	+ ·020,301	− ·005,265	− ·024,509	− ·002,816
	(*b*)	− ·031,462	+ ·000,032	+ ·038,425	+ ·022,755	− ·007,060	− ·032,660	− ·004,166
	(*c*)	− ·047,779	− ·002,425	+ ·042,934	+ ·025,189	− ·009,544	− ·044,832	− ·006,579
4	(*a*)	− ·002,890	− ·024,189	+ ·011,556	+ ·029,810	+ ·010,036	+ ·002,968	− ·010,247
	(*b*)	− ·004,543	− ·030,027	+ ·013,009	+ ·031,921	+ ·010,316	+ ·002,231	− ·015,487
	(*c*)	− ·007,695	− ·038,318	+ ·014,488	+ ·034,010	+ ·010,608	+ ·001,398	− ·023,039
5	(*a*)	0	− ·030,326	− ·000,360	+ ·019,498	− ·010,803	+ ·017,610	+ ·000,491
	(*b*)	0	− ·039,760	− ·001,075	+ ·020,230	+ ·010,914	+ ·016,869	− ·001,027
	(*c*)	0	− ·054,439	− ·002,126	+ ·020,981	+ ·011,058	+ ·016,285	− ·002,916
6	(*a*)	0	− ·020,008	− ·022,838	+ ·007,545	+ ·004,686	+ ·059,958	+ ·012,415
	(*b*)	0	− ·028,474	− ·030,341	+ ·007,029	+ ·004,542	+ ·056,264	+ ·010,411
	(*c*)	0	− ·043,219	− ·041,234	+ ·006,516	+ ·004,433	+ ·053,151	+ ·008,654
7	(*a*)	0	− ·017,910	− ·030,574	− ·002,307	+ ·003,984	+ ·014,391	+ ·023,291
	(*b*)	0	− ·028,491	− ·044,067	− ·003,750	+ ·002,606	+ ·012,497	+ ·019,449
	(*c*)	0	− ·047,576	− ·067,356	− ·005,755	+ ·001,135	+ ·010,895	+ ·016,389

$S(a) = + \cdot 157,074, \quad S(b) = + \cdot 012,276, \quad S(c) = - \cdot 209,976.$

* The formula used was *Casus* I or $z_\theta = z_0 + \frac{1}{2}\theta(\Delta z_{-1} + \Delta z_0) + \frac{1}{2}\theta^2\delta^2 z_0$, the solution of the quadratic giving θ.

† The table suggests, *a posteriori*, that we should have got quite reasonable results from linear interpolation; we have: from (*a*) and (*b*) $r = \cdot 5042$; from (*a*) and (*c*) $r = \cdot 4928$, and from (*b*) and (*c*) $r = \cdot 5025$, as against our ·5034. It should be noticed that the values in Table V are not always in agreement in the last figure with those obtained by dividing $\nu_{ss'}$ in Table IV by the $(\bar{n}_{ss'}/n_{ss'})^2$ of that table, because the somewhat more accurate process was adopted of multiplying $\nu_{ss'}$ by $n^2_{ss'}$ and then dividing by $\bar{n}^2_{ss'}$. Still the physical meanings of $\bar{n}_{ss'}/n_{ss'}$ and $(\bar{n}_{ss'}/n_{ss'})^2$ are so prominent in the work that it seemed desirable to register their values.

Before we consider the graph due to this solution, let us investigate the value of r to be found from (xvi). The values of $\bar{n}_{ss'}/n_{ss'}$ are already provided in Table IV, but we need a table corresponding to Table III giving the product $\mathfrak{Y}_s T_p \mathfrak{Y}_{s'} T_p'$ instead of the product $\mathfrak{Y}_s \tau_p \mathfrak{Y}_{s'} \tau_p'$. This is provided in Table VI. Further if

$$\kappa_{ss'} = \mathfrak{Y}_s T_0 \mathfrak{Y}_{s'} T_0' + r \mathfrak{Y}_s T_1 \mathfrak{Y}_{s'} T_1' + r^2 \mathfrak{Y}_s T_2 \mathfrak{Y}_{s'} T_2' + \ldots,$$

Table VII (p. 143) provides $\kappa_{ss'}$ for the same three values of r, i.e. 0·45, 0·50 and 0·55. Finally Table VIII (p. 143) gives $\kappa_{ss'}/(\bar{n}_{ss'}/n_{ss'})$, whence by summing we obtain

$$v = r - S_{ss'} \{\kappa_{ss'}/(\bar{n}_{ss'}/n_{ss'})\},$$

for the three cases.

Using the same interpolation formula as before in order to discover the value of r for which $v = 0$ we find:

$$r = \cdot 5204.$$

There is thus a difference of ·0170 between the two methods. The probable error found for the product-moment r is ·0160 and the result by the usual product-moment process may be given:

$$r = \cdot 5189 \pm \cdot 0160.$$

Thus either of the values reached by the methods of this paper differ by less than the probable error from the true product-moment value.

(4) If we work out the results by mean square contingency we find:

$$C_2 = \cdot 480{,}690,$$

and the class index correlations are*:

$$\text{For fathers: } r_{Cf} = \cdot 962{,}329.$$
$$\text{For sons: } r_{Cs} = \cdot 964{,}523.$$

Hence correlation from mean square contingency

$$r = C_2/(r_{Cf} r_{Cs}) = \cdot 5179,$$

which is in excellent agreement with the product-moment value.

It would therefore be quite reasonable for such a table as the present to use mean square contingency and class index corrections, and save the heavy labour of Equation (xvi bis) or (xvii). At the same time we cannot assert that this process would always be equally satisfactory for tables with but few broad categories and with much higher correlation.

Our two processes seem to give values slightly in defect and in excess of the true value of r, and we might use their mean, i.e. ·5118, to obtain our graph. We shall, however, first proceed to compare the actual results of solving (xiv) and substituting in (xv) with the result of such approximative processes.

Table IX (p. 145) gives the products of $\mathfrak{Y}_s T_p \mathfrak{Y}_{s'} \tau_p'$ and will therefore enable us by aid of Table IV (p. 140) which gives the values of $\bar{n}_{ss'}/n_{ss'}$ to obtain $\bar{h}_s$. for any value of r. Let

$$\lambda_{ss'} = \mathfrak{Y}_s T_0 \mathfrak{Y}_{s'} \tau_0' + r \mathfrak{Y}_s T_1 \mathfrak{Y}_{s'} \tau_1' + r^2 \mathfrak{Y}_s T_2 \mathfrak{Y}_{s'} \tau_2' + \ldots \ldots \ldots \text{(xxii)}.$$

* Using the values of $\bar{x}_s$ and $\bar{y}_s$ in Tables I and II respectively.

TABLE VI.

Values of $\mathfrak{H}_s T_p \mathfrak{H}_{s'} T_p'$.

s'	p	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$	p	s'
	0	+·005,966	+·022,207	+·000,509	−·007,686	−·005,120	−·011,028	−·004,848	0	
	1	+·030,608	+·054,185	+·001,722	+·010,633	+·011,536	+·037,126	+·025,890	1	
	2	+·054,736	+·014,342	−·000,976	+·010,114	+·000,145	−·029,528	−·048,833	2	
1	3	+·035,199	−·033,825	−·002,246	−·012,136	−·010,270	−·010,999	+·034,276	3	1
	4	+·002,562	−·007,587	+·000,169	−·000,915	+·001,407	+·007,853	−·003,489	4	
	5	+·005,180	−·004,622	−·000,915	−·004,289	−·002,583	+·003,372	+·003,856	5	
	6	+·007,003	−·011,727	+·000,213	−·000,013	+·003,287	+·008,380	−·007,142	6	
	0	+·022,842	+·085,018	+·001,948	−·029,426	−·019,601	−·042,222	−·018,559	0	
	1	+·056,787	+·100,528	+·003,195	+·019,728	+·021,402	+·068,878	+·048,033	1	
	2	+·017,375	+·004,553	−·000,310	+·003,210	+·000,046	−·009,373	−·015,501	2	
2	3	−·033,035	+·031,745	+·002,108	+·011,389	+·009,639	+·010,323	−·032,169	3	2
	4	−·007,358	+·021,786	−·000,486	+·002,627	−·004,040	−·022,548	+·010,019	4	
	5	−·004,261	+·003,802	+·000,753	+·003,528	+·002,124	−·002,774	−·003,172	5	
	6	−·011,913	+·019,949	−·000,362	+·000,022	−·005,591	−·014,255	+·012,150	6	
	0	+·001,324	+·004,929	+·000,113	−·001,706	−·001,136	−·002,448	−·001,076	0	
	1	+·002,378	+·004,211	+·000,134	+·000,826	+·000,896	+·002,885	+·002,012	1	
	2	−·002,423	−·000,635	+·000,043	−·000,448	−·000,006	+·001,307	+·002,162	2	
3	3	−·002,976	+·002,860	+·000,190	+·001,026	+·000,868	+·000,930	−·002,898	3	3
	4	+·000,365	−·001,080	+·000,024	−·000,130	+·000,200	+·001,118	−·000,497	4	
	5	−·001,271	+·001,134	+·000,225	+·001,052	+·000,634	−·000,827	−·000,946	5	
	6	+·000,475	−·000,796	+·000,014	−·000,001	+·000,223	+·000,569	−·000,485	6	
	0	−·005,329	−·019,833	−·000,455	+·006,865	+·004,573	+·009,850	+·004,330	0	
	1	+·006,217	+·011,006	+·000,350	+·002,160	+·002,343	+·007,541	+·005,259	1	
	2	+·008,127	+·002,130	−·000,145	+·001,502	+·000,022	−·004,385	−·007,251	2	
4	3	−·007,217	+·006,935	+·000,461	+·002,488	+·002,106	+·002,255	−·007,028	3	4
	4	−·000,941	+·002,786	−·000,062	+·000,336	−·000,517	−·002,883	+·001,281	4	
	5	−·002,847	+·002,540	+·000,503	+·002,358	+·001,419	−·001,854	−·002,120	5	
	6	−·000,780	+·001,306	−·000,024	+·000,001	−·000,366	−·000,934	+·000,796	6	
	0	−·007,289	−·027,130	−·000,622	+·009,390	+·006,255	+·013,473	+·005,923	0	
	1	+·014,662	+·025,956	+·000,825	+·005,094	+·005,526	+·017,784	+·012,402	1	
	2	+·002,953	+·000,774	−·000,053	+·000,546	+·000,008	−·001,593	−·002,635	2	
5	3	−·013,673	+·013,139	+·000,873	+·004,714	+·003,990	+·004,273	−·013,315	3	5
	4	+·001,057	−·003,128	+·000,070	−·000,377	+·000,580	+·003,238	−·001,439	4	
	5	−·004,116	+·003,672	+·000,727	+·003,408	+·002,052	−·002,679	−·003,064	5	
	6	+·003,320	−·005,559	+·000,101	−·000,006	+·001,558	+·003,973	−·003,386	6	
	0	−·010,563	−·039,314	−·000,901	+·013,607	+·009,064	+·019,524	+·008,582	0	
	1	+·033,046	+·058,500	+·001,860	+·011,480	+·012,454	+·040,082	+·027,952	1	
	2	−·021,493	−·005,632	+·000,383	−·003,971	−·000,057	+·011,595	+·019,175	2	
6	3	−·013,795	+·013,256	+·000,880	+·004,756	+·004,025	+·004,311	−·013,433	3	6
	4	+·006,142	−·018,186	+·000,406	−·002,193	+·003,372	+·018,822	−·008,364	4	
	5	+·001,134	−·001,011	−·000,200	−·000,939	−·000,565	+·000,738	+·000,844	5	
	6	+·008,563	−·014,340	+·000,260	−·000,016	+·004,019	+·010,247	−·008,733	6	
	0	−·006,952	−·025,876	−·000,593	+·008,956	+·005,966	+·012,851	+·005,649	0	
	1	+·034,867	+·061,193	+·001,945	+·012,009	+·013,028	+·041,927	+·029,238	1	
	2	−·059,276	−·015,532	+·001,057	−·010,953	−·000,157	+·031,978	+·052,883	2	
7	3	+·035,498	−·034,111	−·002,265	−·012,238	−·010,357	−·011,092	+·034,567	3	7
	4	−·001,827	+·005,409	−·000,121	+·000,652	−·001,003	−·005,599	+·002,488	4	
	5	+·006,181	−·005,515	−·001,092	−·005,118	−·003,082	+·004,024	+·004,602	5	
	6	−·006,668	+·011,167	−·000,202	+·000,202	−·003,130	−·007,980	+·006,801	6	

TABLE VII. *Values of* $\kappa_{ss'}$.

s'	r	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	(a)	+·034,290	+·045,918	+·000,873	−·002,076	−·000,798	−·000,849	−·000,094
	(b)	+·039,786	+·047,855	+·000,831	−·001,549	−·000,541	−·000,496	−·000,036
	(c)	+·045,903	+·049,468	+·000,762	−·001,097	−·000,350	−·000,251	−·000,001
2	(a)	+·048,425	+·135,200	+·003,506	−·018,688	−·009,255	−·013,278	−·002,561
	(b)	+·050,671	+·142,180	+·003,719	−·017,061	−·007,957	−·010,554	−·001,722
	(c)	+·052,617	+·149,704	+·003,946	−·015,291	−·006,630	−·008,054	−·001,089
3	(a)	+·001,628	+·006,926	+·000,205	−·001,317	−·000,633	−·000,765	−·000,039
	(b)	+·001,526	+·007,188	+·000,223	−·001,252	−·000,545	−·000,509	+·000,040
	(c)	+·001,386	+·007,465	+·000,245	−·001,175	−·000,444	−·000,235	+·000,096
4	(a)	−·001,641	−·013,645	−·000,278	+·008,425	+·005,826	+·012,401	+·004,608
	(b)	−·001,250	−·012,657	−·000,247	+·008,726	+·006,019	+·012,553	+·004,294
	(c)	−·000,903	−·011,563	−·000,211	+·009,071	+·006,233	+·012,663	+·003,892
5	(a)	−·001,344	−·014,202	−.000,165	+·012,270	+·009,181	+·021,659	+·009,613
	(b)	−·000,940	−·012,484	−·000,085	+·012,745	+· 09,643	+·022,682	+·009,562
	(c)	−·000,625	−·010,689	+·000,007	+·013,278	+·010,160	+·023,755	+·009,352
6	(a)	−·000,958	−·013,805	+·000,109	+·018,295	+·015,185	+·041,172	+·023,419
	(b)	−·000,584	−·011,207	+·000,258	+·018,782	+·016,036	+·044,362	+·025,040
	(c)	−·000,328	−·008,749	+·000,419	+·019,263	+·016,957	+·047,837	+·026,557
7	(a)	−·000,047	−·004,380	+·000,263	+·010,961	+·010,729	+·036,961	+·032,908
	(b)	−·000,076	−·003,086	+·000,316	+·010,574	+·010,938	+·040,073	+·038,215
	(c)	+·000,159	−·002,067	+·000,348	+·010,019	+·011,027	+·043,208	+·044,126

TABLE VIII. *Values of* $\kappa_{ss'}/(\bar{n}_{ss'}/n_{ss'})$.

s'	r	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	(a)	+·021,394	+·052,182	+·001,072	0	−·002,057	0	0
	(b)	+·021,552	+·053,054	+·001,178	0	−·002,033	0	0
	(c)	+·021,698	+·053,980	+·001,296	0	−·022,011	0	0
2	(a)	+·055,831	+·149,303	+·003,713	−·012,640	−·006,711	−·007,035	0
	(b)	+·056,643	+·150,514	+·003,957	−·012,331	−·006,547	−·006,833	0
	(c)	+·057,497	+·151,686	+·004,228	−·011,960	−·006,354	−·006,638	0
3	(a)	+·001,898	+·006,439	+·000,185	−·001,621	−·000,741	−·000,777	−·000,018
	(b)	+·002,029	+·006,679	+·000,196	−·001,521	−·000,645	−·000,547	+·000,022
	(c)	+·002,180	+·006,941	+·000,209	−·001,406	−·000,534	−·000,271	+·000,065
4	(a)	−·001,004	−·011,805	−·000,258	+·010,415	+·006,850	+·010,117	+·006,859
	(b)	−·000,992	−·011,538	−·000,225	+·010,424	+·006,858	+·010,124	+·006,845
	(c)	−·000,985	−·011,217	−·000,188	+·010,432	+·006,869	+·010,130	+·006,828
5	(a)	0	−·012,745	−·000,160	+·013,131	+·009,558	+·023,162	+·010,155
	(b)	0	−·012,407	−·000,083	+·013,153	+·009,564	+·023,170	+·010,125
	(c)	0	−·012,005	−·000,007	+·013,177	+·009,569	+·023,173	+·010,084
6	(a)	0	−·009,447	+·000,126	+·015,039	+·009,465	+·058,655	+·025,835
	(b)	0	−·009,156	+·000,311	+·015,079	+·009,468	+·058,760	+·025,838
	(c)	0	−·008,854	+·000,533	+·015,123	+·009,468	+·058,853	+·025,824
7	(a)	0	−·004,556	+·000,352	+·007,494	+·012,697	+·032,408	+·037,813
	(b)	0	−·004,377	+·000,487	+·007,493	+·012,645	+·032,448	+·038,095
	(c)	0	−·004,210	+·000,642	+·007,492	+·012,574	+·032,463	+·038,361

$S(a) = +·510,573$, $S(b) = +·517,476$, $S(c) = +·524,735$,

$v_a = -·060,573$, $v_b = -·017,476$, $v_c = +·025,265$.

TABLE IX.

Values of $\mathfrak{Y}_s T_p \mathfrak{Y}_{s'} \tau_p'$.

s'	p	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$	p	s'
	0	−·002,689	−·010,007	−·000,229	+·003,464	+·002,307	+·004,970	+·002,185	0	
	1	−·013,450	−·023,810	−·000,757	−·004,673	−·005,069	−·016,314	−·011,377	1	
	2	−·023,067	−·006,044	+·000,411	−·004,262	−·000,061	+·012,444	+·020,579	2	
1	3	−·013,496	+·012,969	+·000,861	+·004,653	+·003,938	+·004,217	−·013,142	3	1
	4	−·000,450	+·001,333	−·000,030	+·000,161	−·000,247	−·001,380	+·000,613	4	
	5	−·003,007	+·002,683	+·000,531	+·002,490	+·001,499	−·001,958	−·002,239	5	
	6	−·003,082	+·005,161	−·000,094	+·000,006	−·001,446	−·003,688	+·003,143	6	
	0	−·023,802	−·088,590	−·002,030	+·030,662	+·020,425	+·043,996	+·019,339	0	
	1	−·051,494	−·091,158	−·002,898	−·017,889	−·019,407	−·062,458	−·043,556	1	
	2	−·002,940	−·000,770	+·000,052	−·000,543	−·000,008	+·001,586	+·002,623	2	
2	3	+·036,379	−·034,958	−·002,322	−·012,542	−·010,614	−·011,368	+·035,425	3	2
	4	+·004,781	−·014,154	+·000,316	−·001,707	+·002,625	+·014,650	−·006,510	4	
	5	+·007,797	−·006,957	−·001,378	−·006,456	−·003,887	+·005,076	+·005,805	5	
	6	+·009,448	−·015,822	+·000,287	−·000,017	+·004,434	+·011,306	−·009,636	6	
	0	−·022,458	−·083,587	−·001,915	+·028,931	+·019,271	+·041,511	+·018,247	0	
	1	−·002,986	−·005,286	−·000,168	−·001,037	−·001,125	−·003,622	−·002,526	1	
	2	+·045,338	+·011,880	−·000,808	+·008,377	+·000,120	−·024,459	−·040,449	2	
3	3	+·003,681	−·003,537	−·000,235	−·001,269	−·001,074	−·001,150	+·003,584	3	3
	4	−·007,651	+·022,655	−·000,506	+·002,732	−·004,201	−·023,447	+·010,419	4	
	5	+·001,548	−·001,381	−·000,273	−·001,281	−·000,772	+·001,007	+·001,152	5	
	6	−·011,412	+·019,111	−·000,346	+·000,021	−·005,356	−·013,656	+·011,639	6	
	0	−·010,833	−·040,322	−·000,924	+·013,956	+·009,296	+·020,025	+·008,802	0	
	1	+·012,013	+·021,267	+·000,676	+·004,173	+·004,528	+·014,571	+·010,161	1	
	2	+·017,109	+·004,483	−·000,305	+·003,161	+·000,045	−·009,230	−·015,264	2	
4	3	−·014,068	+·013,518	+·000,898	+·004,850	+·004,105	+·004,396	−·013,699	3	4
	4	−·002,101	+·006,221	−·000,139	+·000,750	−·001,154	−·006,439	+·002,861	4	
	5	−·005,604	+·005,000	+·000,990	+·004,640	+·002,794	−·003,648	−·004,172	5	
	6	−·001,988	+·003,329	−·000,060	+·000,004	−·000,933	−·002,379	+·002,028	6	
	0	−·008,303	−·030,904	−·000,708	+·010,696	+·007,125	+·015,347	+·006,746	0	
	1	+·016,433	+·029,090	+·000,925	+·005,709	+·006,193	+·019,931	+·013,899	1	
	2	+·003,809	+·000,998	−·000,068	+·000,704	+·000,010	−·002,055	−·003,399	2	
5	3	−·015,502	+·014,897	+·000,989	+·005,344	+·004,523	+·004,844	−·015,096	3	5
	4	+·001,087	−·003,220	+·000,072	−·000,388	+·000,597	+·003,332	−·001,481	4	
	5	−·004,744	+·004,233	+·000,838	+·003,928	+·002,365	−·003,088	−·003,532	5	
	6	+·003,592	−·006,016	+·000,109	−·000,007	+·001,686	+·004,299	−·003,664	6	
	0	−·007,749	−·028,843	−·000,661	+·009,983	+·006,650	+·014,324	+·006,297	0	
	1	+·023,811	+·042,152	+·001,340	+·008,272	+·008,974	+·028,881	+·020,140	1	
	2	−·014,636	−·003,835	+·000,261	−·002,704	−·000,039	+·007,896	+·013,057	2	
6	3	−·010,657	+·010,241	+·000,680	+·003,674	+·003,110	+·003,330	−·010,378	3	6
	4	+·004,372	−·012,946	+·000,289	−·001,561	+·002,401	+·013,399	−·005,954	4	
	5	+·000,442	−·000,394	−·000,078	−·000,366	−·000,220	+·000,288	+·000,329	5	
	6	+·006,335	−·010,608	+·000,192	−·000,012	+·002,973	+·007,580	−·006,461	6	
	0	−·003,242	−·012,067	−·000,276	+·004,177	+·002,782	+·005,993	+·002,634	0	
	1	+·015,673	+·027,746	+·000,882	+·005,445	+·005,907	+·019,010	+·013,257	1	
	2	−·025,614	−·006,712	+·000,457	−·004,733	−·000,068	+·013,818	+·022,851	2	
7	3	+·013,663	−·013,130	−·000,872	−·004,711	−·003,987	−·004,270	+·013,305	3	7
	4	−·000,037	+·000,111	−·000,002	+·000,013	−·000,020	−·000,114	+·000,051	4	
	5	+·003,569	−·003,184	−·000,631	−·002,955	−·001,779	+·002,323	+·002,657	5	
	6	−·002,893	+·004,845	−·000,088	+·000,005	−·001,358	−·003,462	+·002,951	6	

TABLE X.

Values of $\vartheta_s \tau_p \vartheta_{s'} T_p'$.

s'	p	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$	p	s'
	0	−·002,716	−·024,296	−·019,919	−·013,581	−·005,206	−·007,621	−·002,113	0	
	1	−·013,578	−·050,535	−·001,157	+·017,491	+·011,650	+·025,097	+·011,032	1	
	2	−·023,244	+·001,049	+·041,494	+·019,288	+·000,298	−·018,826	−·020,060	2	
1	3	−·013,520	+·038,284	+·001,489	−·020,306	−·010,442	−·008,529	+·013,024	3	1
	4	−·000,364	+·004,567	−·007,904	−·002,108	+·001,391	+·005,282	−·000,864	4	
	5	−·002,999	+·008,296	+·000,596	−·007,320	−·002,654	+·001,795	+·002,285	5	
	6	−·003,074	+·008,906	−·011,032	−·000,956	+·003,290	+·006,016	−·003,149	6	
	0	−·010,399	−·093,014	−·076·260	−·051,995	−·019,931	−·029,175	−·008,088	0	
	1	−·025,191	−·093,756	−·002,147	+·032,451	+·021,614	+·046,563	+·020,467	1	
	2	−·007,378	+·000,333	+·013,171	+·006,123	+·000,095	−·005,976	−·006,367	2	
2	3	+·012,689	−·035,930	−·001,397	+·019,057	+·009,800	+·008,004	−·012,223	3	2
	4	+·001,044	−·013,114	+·022,696	+·006,054	−·003,993	−·015,167	+·002,480	4	
	5	+·002,467	−·006,824	−·000,490	+·006,021	+·002,183	−·001,477	−·001,880	5	
	6	+·005,229	−·015,149	+·018,767	+·001,627	−·005,596	−·010,234	+·005,357	6	
	0	−·000,603	−·005,392	−·004,421	−·003,014	−·001,155	−·001,691	−·000,469	0	
	1	−·001,055	−·003,927	−·000,090	+·001,359	+·000,905	+·001,950	+·000,857	1	
	2	+·001,029	−·000,046	−·001,837	−·000,854	−·000,013	+·000,833	+·000,888	2	
3	3	+·001,143	−·003,237	−·000,126	+·001,717	+·000,883	+·000,721	−·001,101	3	3
	4	−·000,052	+·000,650	−·001,125	−·000,300	+·000,198	+·000,752	−·000,123	4	
	5	+·000,736	−·002,035	−·000,146	+·001,795	+·000,651	−·000,440	−·000,561	5	
	6	−·000,209	+·000,605	−·000,749	−·000,065	+·000,223	+·000,408	−·000,214	6	
	0	+·002,426	+·021,699	+·017,790	+·012,130	+·004,650	+·006,806	+·001,887	0	
	1	−·002,758	−·010,265	−·000,235	+·003,553	+·002,366	+·005,098	+·002,241	1	
	2	−·003,451	+·000,156	+·006,161	+·002,864	+·000,044	−·002,795	−·002,979	2	
4	3	+·002,772	−·007,849	−·000,305	+·004,163	+·002,141	+·001,749	−·002,670	3	4
	4	+·000,134	−·001,677	+·002,902	+·000,774	−·000,511	−·001,939	+·000,317	4	
	5	+·001,648	−·004,560	−·000,328	+·004,023	+·001,459	−·000,987	−·001,256	5	
	6	+·000,342	−·000,992	+·001,229	+·000,107	−·000,366	−·000,670	+·000,351	6	
	0	+·003,318	+·029,682	+·024,335	+·016,592	+·006,360	+·009,310	+·002,581	0	
	1	−·006,504	−·024,207	−·000,554	+·008,379	+·005,581	+·012,022	+·005,284	1	
	2	−·001,254	+·000,057	+·002,239	+·001,041	+·000,016	−·001,016	−·001,082	2	
5	3	+·005,252	−·014,872	−·000,578	+·007,888	+·004,056	+·003,313	−·005,059	3	5
	4	−·000,150	+·001,883	−·003,259	−·000,869	+·000,573	+·002,178	−·000,356	4	
	5	+·002,383	−·006,592	−·000,474	+·005,816	+·002,109	−·001,427	−·001,816	5	
	6	−·001,457	+·004,222	−·005,230	−·000,453	+·001,560	+·002,852	−·001,493	6	
	0	+·004,809	+·043,011	+·035,264	+·024,044	+·009,217	+·013,491	+·003,740	0	
	1	−·014,659	−·054,559	−·001,249	+·018,884	+·012,578	+·027,096	+·011,910	1	
	2	+·009,127	−·000,412	−·016,293	−·007,574	−·000,117	+·007,392	+·007,876	2	
6	3	+·005,299	−·015,004	−·000,583	+·007,958	+·004,092	+·003,342	−·005,104	3	6
	4	−·000,872	+·010,946	−·018,945	−·005,054	+·003,333	+·012,661	−·002,070	4	
	5	−·000,656	+·001,815	+·000,130	−·001,602	−·000,581	+·000,393	+·000,500	5	
	6	−·003,758	+·010,889	−·013,490	−·001,169	+·004,023	+·007,356	−·003,851	6	
	0	+·003,165	+·028,310	+·023,211	+·015,825	+·006,066	+·008,880	+·002,462	0	
	1	−·015,334	−·057,071	−·001,307	+·019,753	+·013,157	+·028,344	+·012,459	1	
	2	+·025,172	−·001,136	−·044,936	−·020,888	−·000,323	+·020,387	+·021,724	2	
7	3	−·013,635	+·038,608	+·001,501	−·020,478	−·010,531	−·008,601	+·013,134	3	7
	4	+·000,259	−·003,256	+·005,635	+·001,503	−·000,991	−·003,766	+·000,616	4	
	5	−·003,579	+·009,899	+·000,711	−·008,734	−·003,167	+·002,142	+·002,727	5	
	6	+·002,927	−·008,480	+·010,505	+·000,911	−·003,133	−·005,729	+·002,999	6	

We shall proceed to calculate $\lambda_{ss'}$ for three values of r which lie near the probable value of r as found from each column. We will take these as ·45, ·50 and ·55; from these values we shall obtain $\bar{h}_s$. for each column from (xiv) and interpolating the real $\bar{h}_s$. between them find the corresponding columnar r, which will be then substituted in (xv) by aid of Table X to obtain the columnar mean $\bar{k}_s$.. Table XI gives the values of $\lambda_{ss'}$ for $r = ·45$, ·50 and ·55, and Table XII the resulting values of $\bar{h}_s$.

TABLE XI. *Values of* $\lambda_{ss'}$ *for* $r = ·45$, ·50 *and* ·55.

s'	r	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	(*a*)	−·014,742,01	−·020,616,58	−·000,400,18	+·000,974,70	+·000,377,97	+·000,409,54	+·000,044,95
	(*b*)	−·017,038,00	−·021,554,08	−·000,383,88	+·000,731,59	+·000,258,31	+·000,246,06	+·000,015,95
	(*c*)	−·019,587,49	−·022,373,22	−·000,356,40	+·000,518,95	+·000,168,60	+·000,136,31	−·000,003,29
2	(*a*)	−·043,836,23	−·133,792,74	−·003,545,25	+·021,169,83	+·010,795,76	+·015,963,45	+·003,258,21
	(*b*)	−·045,046,53	−·140,080,50	−·003,775,08	+·019,705,30	+·009,504,62	+·012,993,41	+·002,268,84
	(*c*)	−·045,869,07	−·146,859,24	−·004,026,98	+·018,090,53	+·008,150,14	+·010,141,50	+·001,500,21
3	(*a*)	−·014,665,26	−·082,820,10	−·002,204,29	+·030,133,27	+·018,460,19	+·033,767,07	+·009,791,12
	(*b*)	−·012,764,50	−·082,030,73	−·002,275,94	+·030,479,17	+·018,233,88	+·031,794,16	+·008,188,80
	(*c*)	−·010,711,23	−·080,956,49	−·002,360,54	+·030,869,67	+·017,938,38	+·029,455,85	+·006,551,72
4	(*a*)	−·003,450,60	−·028,237,21	−·000,587,66	+·017,032,32	+·011,713,27	+·024,762,35	+·009,092,34
	(*b*)	−·002,645,25	−·026,280,92	−·000,528,69	+·017,630,94	+·012,084,98	+·024,998,89	+·008,434,25
	(*c*)	−·001,920,26	−·024,106,94	−·000,459,56	+·018,316,53	+·012,492,17	+·025,139,70	+·007,601,99
5	(*a*)	−·001,562,59	−·016,357,80	−·000,196,08	+·013,951,10	+·010,408,16	+·024,456,57	+·010,780,30
	(*b*)	−·001,096,19	−·014,410,34	−·000,106,48	+·014,492,89	+·010,926,94	+·025,583,17	+·010,698,56
	(*c*)	−·000,731,63	−·012,372,26	−·000,003,49	+·015,100,01	+·011,507,01	+·026,761,82	+·010,435,95
6	(*a*)	−·000,728,92	−·010,344,20	+·000,068,82	+·013,421,77	+·011,082,88	+·029,840,53	+·016,766,62
	(*b*)	−·000,448,58	−·008,432,81	+·000,177,88	+·013,793,06	+·011,705,64	+·032,119,62	+·017,871,20
	(*c*)	−·000,255,73	−·006,613,75	+·000,295,92	+·014,164,31	+·012,382,26	+·034,601,52	+·018,889,99
7	(*a*)	−·000,090,63	−·002,150,92	+·000,121,52	+·005,185,57	+·005,018,14	+·016,965,98	+·014,515,02
	(*b*)	−·000,037,11	−·001,530,11	+·000,149,03	+·005,035,92	+·005,142,06	+·018,430,12	+·016,770,70
	(*c*)	−·000,000,75	−·001,037,56	+·000,167,89	+·004,808,83	+·005,217,99	+·019,928,67	+·019,271,47

TABLE XII. *Values of* $\bar{h}_s$. *for Columns.*

r	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
·45	−2·19299	−·92114	−·02549	+·56650	+·98336	+1·45054	+2·30569
·50	−2·18505	−·91848	−·02538	+·56616	+·98315	+1·44900	+2·29880
·55	−2·17567	−·91485	−·02521	+·56580	+·98286	+1·44685	+2·28999
Actual $\bar{h}_s$.	−2·19667	−·91404	−·02553	+·56594	+·98333	+1·44723	+2·29464
Extra- or Interpolated r	·4229	·5599	·4167	·5309	·4585	·5426	·5249

We have thus the values of r found from each column*.

We now turn to Table X and calculate in exactly the same way the values of

$$\lambda'_{ss'} = \vartheta_s \tau_0 \vartheta_{s'} T_0' + r \vartheta_s \tau_1 \vartheta_{s'} T_1' + \ldots + r^p \vartheta_s \tau_p \vartheta_{s'} T_p' + \ldots,$$

for the r peculiar to each column for that column. We thus obtain Table XIII.

TABLE XIII.

Values of $\lambda'_{ss'}$ for r of each Vertical Column.

s'	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	−·013,707,54	−·044,362,34	−·013,376,98	−·002,394,74	−·000,770,04	−·000,212,73	−·000,006,10
2	−·021,315,40	−·153,841,07	−·074,192,36	−·029,418,02	−·009,240,63	−·006,036,02	−·000,641,41
3	−·000,771,58	−·008,202,77	−·004,826,27	−·002,225,87	−·000,633,67	−·000,217,60	+·000,030,11
4	+·000,880,64	+·014,176,58	+·018,829,61	+·015,680,02	+·005,954,01	+·008,797,09	+·001,837,83
5	+·000,759,52	+·013,488,41	+·024,318,82	+·022,680,28	+·009,395,75	+·016,257,72	+·004,194,22
6	+·000,584,54	+·011,211,78	+·031,232,08	+·032,630,32	+·015,526,73	+·032,207,15	+·011,205,66
7	+·000,127,47	+·002,739,83	+·015,206,16	+·017,131,65	+·010,878,46	+·028,515,80	+·017,104,71
$\bar{k}_{s.}$	−·963,72	−·507,66	−·010,29	+·303,04	+·425,95	+·789,11	+1·238,75

The values in Table XIII divided by $\bar{n}_{ss'}/n_{ss'}$ from Table XIV and summed for each column give, on multiplication by $N/n_{s.}$, the $\bar{k}_{s.}$ of the last row of the table.

To obtain Table XIV we must return to Equation (x), use the appropriate r for the column and the values in Table III of $\vartheta_s \tau_p \vartheta_{s'} \tau_p'$. Taking σ_x and σ_y as units of the horizontal and vertical variates we can plot $\bar{k}_{s.}$ in Table XIII to $\bar{h}_{s.}$ from Table XII and so obtain the regression line as formed by the means of each column, and set against it the regression lines as found from polychoric $r_p = ·5034$, or $·5204$.

TABLE XIV.

Values of $\dfrac{\bar{n}_{ss'}}{n_{ss}}$ for columnar Values of r.

s'	$s=1$	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$
1	1·481,160	·918,247	·881,901	∞	·366,258	∞	∞
2	·850,270	·995,746	·946,290	1·319,975	1·351,776	1·261,565	∞
3	·910,673	1·075,087	1·090,803	·830,945	·853,291	·876,250	1·668,232
4	1·846,116	1·016,776	1·066,432	·856,640	·855,012	1·248,823	·600,417
5	∞	·866,469	1·028,801	·992,395	·968,288	1·018,112	·937,952
6	∞	·980,885	·887,677	1·263,099	1·619,040	·803,917	·999,089
7	∞	·454,171	·808,851	1·368,315	·848,918	1·316,691	1·074,517

* The mean value of r weighted with the column totals is ·5022 which is in reasonable accord with (i.e. within the probable error of) the results on p. 142.

This is done in Diagram I. But what we actually desire is to compare the observations and the regression lines as given by the present polychoric method with those obtained by product-moment methods.

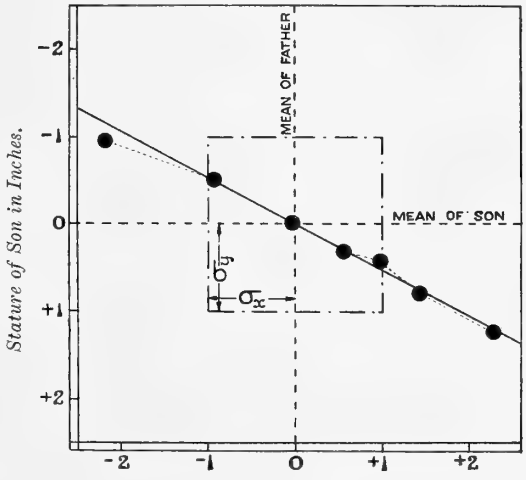


Diagram I.

Our actual data from which the table on p. 135 was obtained are given in Table XV. The following are the values of the constants in inches:

Mean Stature of Father: $\bar{x} = 67''{\cdot}878$.

Mean Stature of Son: $\bar{y} = 68''{\cdot}845$.

Standard Deviation of Father: $\sigma_x = 2''{\cdot}6576$.

Standard Deviation of Son: $\sigma_y = 2''{\cdot}6885$.

Correlation of Father and Son: $r = {\cdot}5189 \pm {\cdot}0160$.

In Diagram II the regression line (slope, ·5245) with means of the arrays as dark circles is given. Against this we have put as hollow circles the values of $\bar{h}_s$. and $\bar{k}_s$. multiplied by their respective S.D.'s to indicate the result as worked out in the present paper. The closeness of the polychoric coefficient ·5204 and the product-moment coefficient does not permit of two regression lines being drawn. It will be seen that the fit to the observations by use of broad categories and the polychoric method is really quite as satisfactory as the fit by the product-moment method. But the amount of arithmetical work is incomparably greater by the former, even if it be less than Ritchie-Scott's process with 49 cells would be.

Accordingly we now proceeded to investigate the extent to which approximations shortening the arithmetic would introduce serious error. The first question to be answered is: To what extent in finding the means $\bar{k}_s$. of the arrays is it needful to use the actual value of the correlation coefficient as found for each column? In order to test this we proceeded to find the $\bar{k}_s$. for each columnar

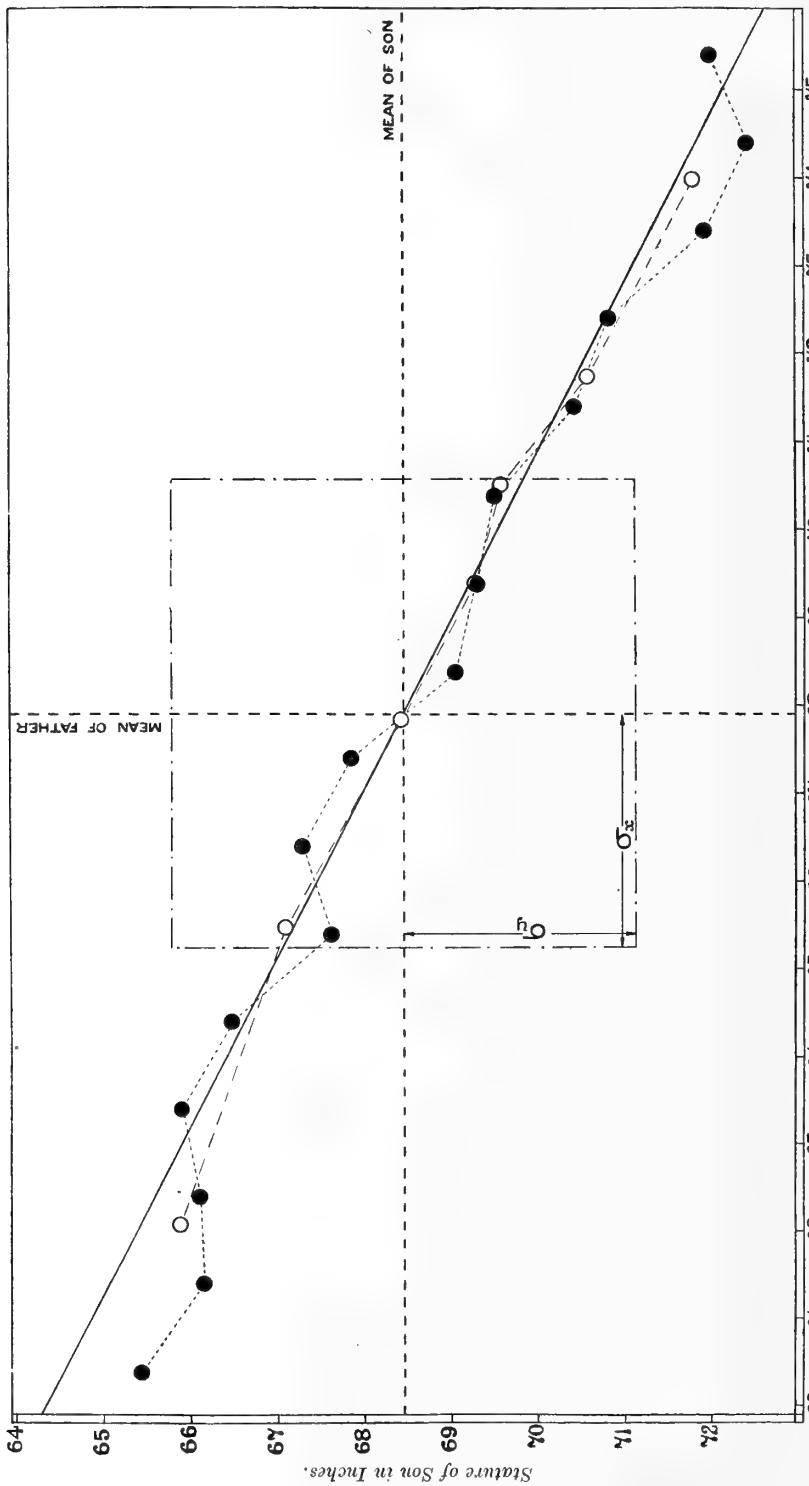

MEAN OF SON
MEAN OF FATHER
σ_x
σ_y
64 65 66 67 68 69 70 71 72
60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
Stature of Son in Inches.
Stature of Father in Inches.

Diagram II.

TABLE XV.

Correlation of Stature in 1000 *pairs, Father and Son.*

Stature of Father.

Stature of Son.	59″·875—	60″·875—	61″·875—	62″·875—	63″·875—	64″·875—	65″·875—	66″·875—	67″·875—	68″·875—	69″·875—	70″·875—	71″·875—	72″·875—	73″·875—	74″·875—	75″·875—	Totals
59″·875—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	2
60″·875—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1
61″·875—	1	—	—	—	3	—	1	1	—	—	—	—	—	—	—	—	—	6
62″·875—	—	—	2	4	4	3	1	3	2	—	1	—	—	—	—	—	—	20
63″·875—	1	—	1	5	6	5	9	2	2	1	—	—	—	—	—	—	—	32
64″·875—	2	3	3	5	11	11	10	17	4	1	2	—	1	—	—	—	—	70
65″·875—	1	1	2	6	9	10	20	17	15	7	6	—	—	—	—	—	—	94
66″·875—	1	—	6	4	11	24	21	28	10	12	7	4	1	—	—	—	—	129
67″·875—	—	2	2	7	9	20	16	33	27	26	20	13	6	—	—	—	—	181
68″·875—	1	—	1	4	1	12	13	10	22	26	24	6	2	2	1	—	—	125
69″·875—	—	—	—	—	5	11	15	18	18	23	18	13	4	4	1	1	—	131
70″·875—	—	—	—	—	2	5	4	13	12	12	13	8	7	3	1	—	—	80
71″·875—	—	—	—	—	—	4	1	7	7	9	9	9	7	3	1	—	—	57
72″·875—	—	—	—	—	1	2	—	—	13	4	2	9	1	1	1	2	—	36
73″·875—	—	—	—	—	1	1	—	—	4	1	5	4	3	2	—	—	—	21
74″·875—	—	—	—	—	—	—	—	—	2	2	—	1	—	2	1	—	—	8
75″·875—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
76″·875—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	2
77″·875—	—	—	—	—	—	—	—	—	—	1	1	—	1	—	—	—	—	3
78″·875—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1
Totals	7	6	17	36	63	109	111	149	139	125	109	67	34	18	7	3	0	1000

array for the *same* correlation coefficient, and we took for the value of that coefficient ·5000, somewhat under the value found by either polychoric coefficient.

Table XVI gives our results. It involved finding a new series of values for $\lambda'_{ss'}$, but those for $\bar{n}_{ss'}/n_{ss'}$ have already been computed under (*b*) in Table IV. The results are given in terms of inches.

TABLE XVI.

Columnar Means by Different Processes.

s	$\bar{h}_{s.} \times \sigma_x$	$\bar{k}_{s.} \times \sigma_y$	$\bar{k}_{s.} \times \sigma_y$	$\bar{k}_{s.} \times \sigma_y$	Common base
		Each column its own r	Each column for $r = ·50$	Each column assumed Normal	
1	−5·8379	−2·5881	−2·6498	−2·4809	2′
2	−2·4292	−1·3633	−1·3531	−1·4357	3′
3	− ·0678	− ·0276	− ·0176	− ·0701	3′
4	+1·5040	+ ·8138	+ ·8087	+ ·7632	3′
5	+2·6133	+1·1439	+1·1511	+1·0866	3′+4′
6	+3·8462	+2·1192	+2·1122	+2·1744	4′+5′
7	+6·0982	+3·3267	+3·3194	+3·1109	5′

An examination of the fourth column of Table XVI shows us that we have not for practical purposes seriously modified the columnar means by using $r = \cdot 50$ instead of the individual value for each column. This is illustrated in Diagram III, where except in the case of the first array there is hardly daylight between the two series of points.

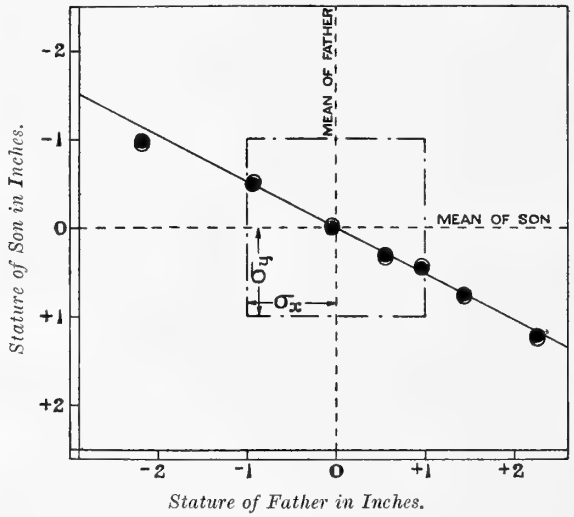


Diagram III.

In Diagram III the hollow circles give the means with r obtained for each column, the nearly superposed dark circles the means with $r = \cdot 5000$.

The solution of the problem therefore falls back on Equations (x), (xvi†) and (xv). We should still have to calculate $\mathfrak{Y}_s \tau_p$, $\mathfrak{Y}_s \tau_p'$, $\mathfrak{Y}_s T_p$ and $\mathfrak{Y}_s T_p'$, but we should only need the three series of products $\mathfrak{Y}_s \tau_p \mathfrak{Y}_{s'} \tau_p'$, $\mathfrak{Y}_s T_p \mathfrak{Y}_{s'} T_p'$ and $\mathfrak{Y}_s \tau_p \mathfrak{Y}_{s'} T_p'$, and to obtain $\bar{k}_s$. it would be adequate to use a value of r for which $\bar{n}_{ss'}/n_{ss'}$ had been found for the final interpolation. Still this involves very lengthy arithmetic, and we naturally crave for a still easier process. The present full working out of a numerical example enables us for the first time really to test the adequacy of an easier method of dealing with such polychoric tables which has been long in use as an approximate method in the Biometric Laboratory.

(6) It is clear that if we could find the means of the columnar arrays, we could readily obtain the correlation and the regression line by aid of the correlation ratio corrected for class index. The whole problem accordingly turns on a ready means of reaching—at any rate—an approximate value of the mean of a columnar array. This array is the slice between two parallel planes of a normal correlation surface.

In the case of a surface of zero correlation

$$Z = Z_0 e^{-\frac{1}{2}(X^2 + Y^2)/a^2},$$

the slice between X_1 and X_2 has for its volume on dY

$$\int_{X_1}^{X_2} e^{-\frac{1}{2}X^2/a^2}\,dX\, e^{-\frac{1}{2}Y^2/a^2}\,dY;$$

the slice is therefore given by the normal curve:

$$\text{Ordinate} = \text{const.} \times e^{-\frac{1}{2}Y^2/a^2}.$$

It seems therefore not unreasonable after the surface of revolution is stretched and slid into a correlation surface to assume the slice to be still approximately a normal curve. Unfortunately the determination of the best mean and standard deviation for normal material given in broad categories does not admit of very easy solution. What we need is the difference between the means of a columnar array and of a marginal frequency as a multiple of the standard deviation of the latter. We shall obtain results differing more or less from each other according to the individual broad category we take as the basis of comparison between σ_s the standard deviation of the sth slice and σ_y the standard deviation of the marginal frequency. In fact the range of any broad category or of any combination of broad categories, except the tail categories, can be made a means of linking up σ_s and σ_y. A little experience, however, shows (a) that it is undesirable to find the σ_s of any array from a category of small frequency, and (b) that for arrays of small total frequency symmetrical tripartite divisions as far as feasible are the best*. The last column in Table XVI shows the system selected for each of our columnar arrays.

Take, for example, $s = 5$, the columnar array may be taken on the base of 3′ and 4′ categories as

1′ + 2′	9		335
3′ + 4′	36	and compared with	421
5′ + 6′ + 7′	24		244
Totals	69		1000

as the corresponding marginal distribution. The corresponding proportional frequencies up to the dichotomic planes are: $\begin{matrix}\cdot1304\\ \cdot6521\end{matrix}$ and $\begin{matrix}\cdot3350\\ \cdot7560\end{matrix}$. The distances of the mean† from the two dichotomic planes in the first case are

$$-1\cdot1245\,\sigma_5 \text{ and } +\cdot3910\,\sigma_5,$$

and in the second case

$$-\cdot4261\,\sigma_y \text{ and } +\cdot6935\,\sigma_y,$$

where σ_5 is the standard deviation of the normal curve assumed to represent the columnar array 5. Accordingly the range of 3′ + 4′ categories

$$= 1\cdot5155\,\sigma_5 = 1\cdot1196\,\sigma_y,$$

which gives σ_5 in terms of σ_y.

* The probable error of a standard deviation found in this way is discussed in *Biometrika*, Vol. XIII. p. 129.

† Found from the Probability Integral Table.

Hence the distance between the means is

$$\begin{aligned} &\cdot6935\sigma_y - \cdot3910\sigma_5 \\ &= \{\cdot6935 - \cdot3910 \times 1\cdot1196/1\cdot5155\}\,\sigma_y \\ &= \cdot4046\sigma_y \\ &= 1\cdot0866, \text{ if we introduce the value of } \sigma_y. \end{aligned}$$

This and the corresponding values are recorded in the fifth column of Table XVI. It will be seen that these values approximate to those in the third column, the greatest differences being in the small first and last arrays.

Of course in actually working with material solely given in broad categories we use the value ·4046, treating σ_y as our unit of measurement. The means of the columnar arrays can be found with great ease and with considerable approximation by this method.

If we now proceed to take the mean of our means duly weighted with their frequencies, we find it to be $-\cdot0510$,—not a very serious divergence from zero. However, we subtract it* from the means in the fifth column of Table XVI, multiply the squares of the remainders by the corresponding frequencies, sum and divide by the square of σ_y. Thus we obtain

$$\eta^2 = \frac{1\cdot818,8034}{7\cdot211,9103} = \cdot25210144,$$

or:

$$\eta = \cdot502148.$$

If we divide by the class index correlation of the x-variate, i.e. ·962,329†, we obtain

$$\eta = \cdot5218,$$

which correlation ratio we may take to be the correlation coefficient and compare with our polychoric coefficient ·5204 (p. 142). Clearly although our means as found by the hypothesis of normal distribution of the columnar arrays agree only approximately with the polychoric means of the third column of Table XVI, they lie practically on the same regression line, as Diagram IV indicates. We conclude, therefore, that in this case as probably in many like cases, it is quite adequate to obtain the means of the columnar arrays by treating them as normal distributions, then determining their correlation ratio and correcting it for the class index. The corresponding regression line with the means of the columnar arrays indicated will be for many purposes an adequate graph showing the general nature of the correlation.

The general purpose of this paper has now been fulfilled; it has been shown how a general polychoric coefficient covering all the data provided in a given contingency table may be found, and how a graph may be drawn representing such a table effectively. At the same time such a process is very laborious and probably will not be lightly undertaken or only in cases of grave uncertainty. The method

* Correlation ratio without subtraction = ·5222.

† See p. 142.

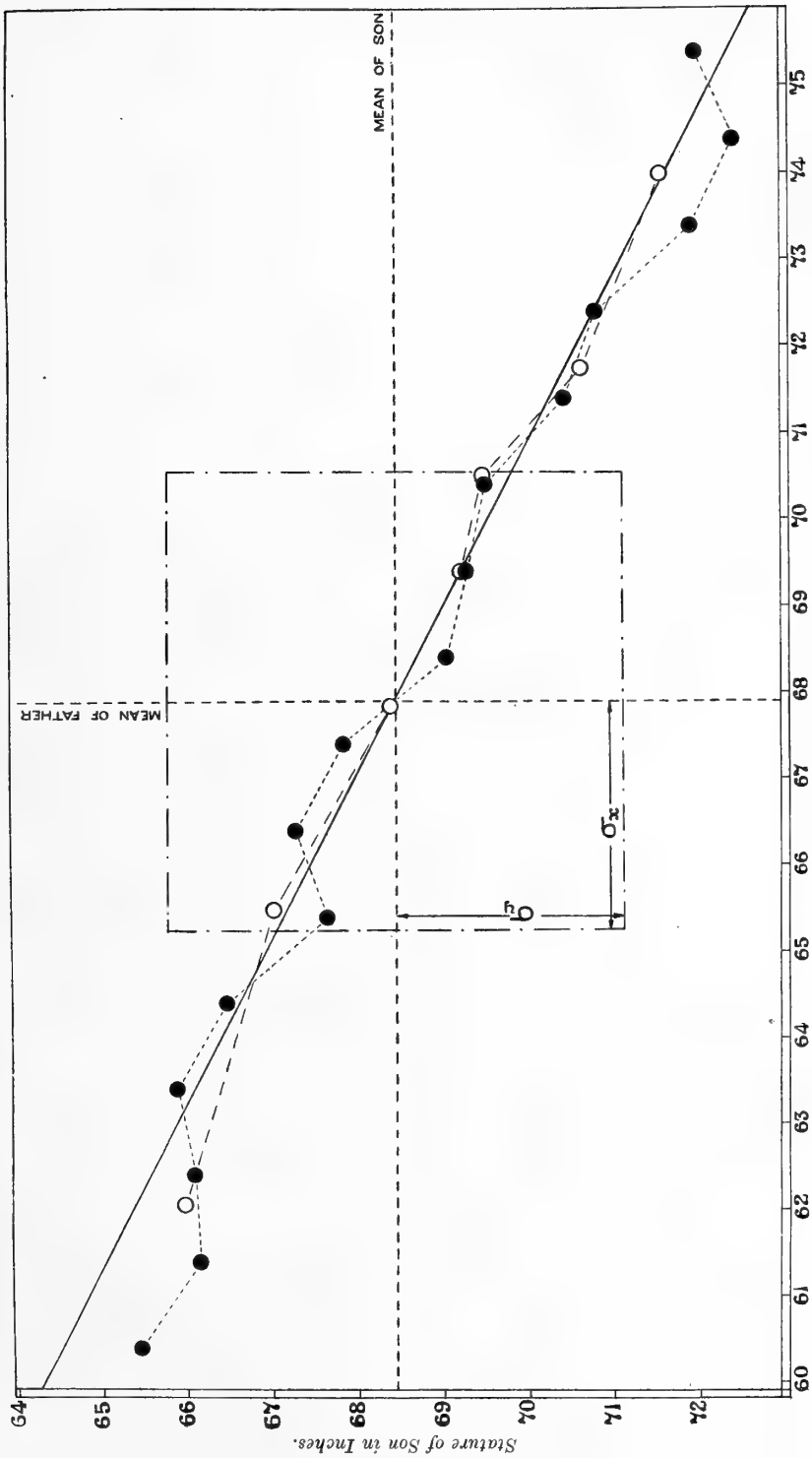

MEAN OF SON
MEAN OF FATHER
σ_x
σ_y
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
64
65
66
67
68
69
70
71
72
Stature of Father in Inches.
Stature of Son in Inches.

Diagram IV.

is one of fitting the "best" normal surface to the data subject to the limitation that the marginal totals are exactly reproduced, and this limits the generality.

An example has been given of the process, but it is seen from this example that the heavy arithmetic does not lead us to any more accurate value for the correlation than far simpler methods. Thus:

Correlation from product-moment	$= \cdot 5189 \pm \cdot 0160.$
Polychoric Correlation Coefficient "Best Fit"	$= \cdot 5034.$
Polychoric Correlation Coefficient "Product Moment"	$= \cdot 5204.$
Mean Square Contingency, Corrected for Class Indices	$= \cdot 5179.$
Correlation Ratio from means of arrays	$= \cdot 5218.$

The latter method, which has been long in use in the Biometric Laboratory, is thus, when used with due precaution, seen to be justified by the theoretically preferable polychoric method. If a method could be discovered of finding uniquely the mean of a columnar array, *using all its cells at the same time*, this method would still more effectively replace the polychoric correlation coefficient.

ON EXPANSIONS IN TETRACHORIC FUNCTIONS.

By JAMES HENDERSON, M.A., B.Sc.

(1) We define the *tetrachoric function* of order s to be $\tau_s(x)$, where

$$\tau_s(x) = \frac{1}{\sqrt{s!}} \left(-\frac{d}{dx}\right)^{s-1} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \quad \text{.....(i).}$$

Other writers have adopted various other values for the external numerical factor but this is immaterial. The factor $\frac{1}{\sqrt{s!}}$ was chosen because it gives an extremely simple expression for the volume of a quadrant of the normal bivariate frequency surface, and because for tabulating the numerical values of the functions it is necessary to have some reduction factor of this kind to keep them of manageable size. We can usually drop the argument x and speak of τ_s. The values of τ_s for $s=1$ up to $s=6$ are tabled to five decimal places in the book, *Tables for Statisticians and Biometricians**, for values of $\frac{1}{2}(1-\alpha)$ (which is really τ_0, when the argument is negative) from ·000 to ·500 at intervals of ·001. With a different multiplier they have been tabled by Charlier† to four decimal places only for $s=1$, 4 and 5 ($x=$ ·00 to 3).

The general form of the tetrachoric function of order s is

$$\tau_s(x) = \frac{1}{\sqrt{s!}} \left\{ x^{s-1} - \frac{(s-1)(s-2)}{2.1!} x^{s-3} + \frac{(s-1)(s-2)(s-3)(s-4)}{2^2.2!} x^{s-5} - \text{etc.} \right\} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{.....(ii),}$$

that is, the ordinate of the normal curve of errors multiplied by a polynomial of degree $(s-1)$. τ_1 is simply the ordinate of the normal curve, while τ_0 is the area of the tail of the normal curve up to a given abscissa x, with the addition of an arbitrary constant. This constant may be so selected that $\tau_0 = \int_{-\infty}^{x} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx$, and will be found from the tables of the probability integral. It will be equal to $\frac{1}{2}(1+\alpha)$, if x is positive and $\frac{1}{2}(1-\alpha)$, if x be negative in the usual notation. Accordingly the expansion of a function of x, $f(x)$ in a series of tetrachoric functions, is really the expansion of the difference of the function and a multiple of the probability integral in terms of

$$\left(c_0 + c_1 \frac{x}{\sigma} + c_2 \frac{x^2}{\sigma^2} + \ldots\right) e^{-\frac{1}{2}x^2/\sigma^2},$$

where σ and c_0, c_1, c_2 ... are at our choice.

* Cambridge University Press, p. 1, and pp. 42—51.

† *Vorlesungen über die Grundzüge der mathematischen Statistik*, 1920.

The real reason for adopting

$$c_0'\tau_0 + c_1'\tau_1 + c_2'\tau_2 + c_3'\tau_3 + \ldots,$$

instead of the above expression, is that the calculation of the constants $c_0', c_1', c_2' \ldots$ is more direct than that of $c_0, c_1, c_2 \ldots$ because the tetrachoric functions are semi-orthogonal functions*. It will be seen that the problem of expansion in tetrachoric functions is closely related to a theorem of Laplace. If U be a unimodal function of x within the range under discussion and the integral $I = \int U dx$ be required, Laplace transfers to the mode m as origin so that $x = m + \xi$ and writes U in the following form:

$$U = U_m e^{-\frac{1}{2}a_0\xi^2}(1 + a_3\xi^3 + a_4\xi^4 + \ldots).$$

He extends the limits to ∞ in both directions by supposing $U = 0$ outside the given range and in the integration applies the well-known values of $\int_{-\infty}^{\infty} \xi^s e^{-\frac{1}{2}a_0\xi^2} d\xi$, i.e. zero if s be odd, and again if s be even $(= 2r)$,

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\xi^2/\sigma^2}\, \xi^{2r}\, d\xi = (2r-1)(2r-3)\ldots 3 \,.\, 1 \,.\, \sqrt{2\pi}\, \sigma^{2r+1}.$$

It will be seen that Laplace is really proceeding by expansion in tetrachoric functions as the process is precisely the same whatever be the limits of the integral of U. Following Laplace we develop our function in "incomplete normal moment functions," i.e. $\int_{-\infty}^{x} \frac{x^s e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx$†; it is better to use tetrachoric functions. The series in tetrachoric functions seems to converge slightly better than that in incomplete normal moment functions.

If we have

$$F(x) = a_0\tau_1 + a_1\tau_2 + a_2\tau_3 + \ldots + a_{s-1}\tau_s + \ldots,$$

then, assuming we may integrate the right-hand side of this equation term by term (i.e. assuming uniform convergence) between x and ∞,

$$\int F(x)\, dx = a_0\tau_0 + \frac{a_1\tau_1}{\sqrt{2}} + \frac{a_2\tau_2}{\sqrt{3}} + \ldots,$$

since

$$\int_x^{\infty} \tau_s dx = \frac{\tau_{s-1}}{\sqrt{s}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(iii)}.$$

* A series of functions $f_1(x), f_2(x) \ldots f_s(x) \ldots f_{s'}(x)$ is orthogonal if $\int f_s(x) f_{s'}(x)\, dx = 0$ when s and s' are not equal, the integration being throughout the range. They are semi-orthogonal if

$$\int f_s(x) f_{s'}(x)\, \phi(x)\, dx = 0,$$

$\phi(x)$ being a function of x peculiar to the series. In other words a system is orthogonal if the sums of the products of different order functions vanish *without weighting* for x. A system is semi-orthogonal if we require to weight the values of x to obtain the vanishing of the product sum. This weighting is the great disadvantage of semi-orthogonal functions. In our case of the tetrachoric functions the weighting factor is $e^{\frac{1}{2}x^2}$ or the tails of series are excessively weighted.

† Discussed *Biometrika*, Vol. VI. p. 59. Tables of these functions up to $s = 10$ are given in *Tables for Statisticians*, pp. 22—3.

Let $\tau_s = \frac{1}{\sqrt{s!}} p_{s-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, where p_{s-1} is the polynomial in x of degree $(s-1)$ in (ii).

Let $\tau_{s'}$ be another tetrachoric function and suppose s' is greater than s. Then

$$\int_{-\infty}^{\infty} \tau_s \tau_{s'} e^{\frac{1}{2}x^2} dx = \frac{1}{\sqrt{s!}} \frac{1}{\sqrt{s'!}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} p_{s-1} \left(-\frac{d}{dx}\right)^{s'-1} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx.$$

Now since $\tau_s(\infty)$ and $\tau_s(-\infty)$ will always be zero owing to the exponential factor ($s > 0$) we can integrate by parts transferring the $\frac{d}{dx}$ from the exponential to the polynomial, therefore

$$\int_{-\infty}^{\infty} \tau_s \tau_{s'} e^{\frac{1}{2}x^2} dx = \frac{1}{\sqrt{s!}\sqrt{s'!}} \frac{1}{\sqrt{2\pi}} \left[\left\{ -p_{s-1} \left(-\frac{d}{dx}\right)^{s'-2} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \right\}_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \left(-\frac{d}{dx}\right)^{s'-2} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \frac{d}{dx} p_{s-1} \, dx \right].$$

The integrated part at every step vanishes at the limits and ultimately

$$\int_{-\infty}^{\infty} \tau_s \tau_{s'} e^{\frac{1}{2}x^2} dx = \frac{1}{\sqrt{s!}\sqrt{s'!}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{d^{s'-1}}{dx^{s'-1}} p_{s-1} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx.$$

Since p_{s-1} is a polynomial of degree $(s-1)$ and s' is $> s$ the differential of the polynomial vanishes, i.e.

$$\int_{-\infty}^{\infty} \tau_s \tau_{s'} e^{\frac{1}{2}x^2} dx = 0, \quad s \neq s' \quad \text{.....(iv).}$$

If $s' = s$ then the differential of p_{s-1} reduces to $(s-1)!$ so that

$$\int_{-\infty}^{\infty} \tau_s^2 e^{\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \frac{(s-1)!}{s!} \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{s}, \quad s > 0 \quad \text{.....(v).}$$

These equations (iv) and (v), which give the fundamental properties of the tetrachoric functions, enable us to expand any function $F(x)$ in terms of tetrachoric functions if we can find the value of the integral

$$\int_{-\infty}^{\infty} F(x) \tau_s e^{\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{s!}} p_{s-1} F(x) \, dx \quad \text{.....(vi).}$$

Since p_{s-1} is an integral function of x, this amounts to saying that we can expand any function of which we are able to determine the successive moment-coefficients.

The *practical value* of the functional expansion when obtained is, however, a very different matter. That depends on the convergency of the series and our experience has shown us that in the most common cases the convergency is so slight or non-existent as to render the expansion idle.

The matter is a very important one for Thiele*, Edgeworth† and Charlier‡ have proposed to treat skew frequency distributions by a process, which amounts to the same thing as the expansion by tetrachoric functions.

An attempt made many years ago§ to expand Incomplete Γ- and B-functions by Laplace's method in Incomplete Moment Functions convinced Professor Pearson that little was to be gained by a series expansion in the form of a polynomial multiplied by the ordinate of a normal curve. A variant of this method, that of expressing Incomplete Γ- and B-functions in a series of tetrachoric functions, was tried a year ago and it was found that except for a small distance round the mode this method of expressing a frequency distribution was quite ineffectual. The matter is of considerable importance because quite recently a Scandinavian actuary in America‖ has been analysing mortality curves by tetrachoric functions and asserts not only that they give a good fit but apparently believes that each function of the series has some natural physiological meaning! It is quite possible to represent the survivors of 100,000 persons born in the same year of life by a Fourier's series from 0 to 100 years but one would hardly claim any special physiological significance for the individual periodic terms¶. Such a series however is far easier to deal with in later treatment, such as differencing, than a series in tetrachoric functions.

For the numerical calculation of the tetrachoric functions the difference equation of these functions is invaluable, i.e.

$$\tau_s = x\beta_s\tau_{s-1} - \gamma_s\tau_{s-2},$$

where x is the argument of the functions and

$$\beta_s = \frac{1}{\sqrt{s}}, \quad \gamma_s = \frac{s-2}{\sqrt{s(s-1)}}.$$

Tables of β_s and γ_s are given in *Tables for Statisticians* (p. 1 of introduction) to five decimal places for $s = 7$ to $s = 24$ (the first six tetrachoric functions being given on pp. 42—51) and in *Biometrika*, Vol. XIV. p. 130 to 7 decimal places.

For our work β_s and γ_s were required to 7 places (sometimes to 8) to obtain the requisite accuracy. The procedure consists in calculating τ_1, which is equal to $\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$, directly to the required degree of accuracy and then by means of the tables referred to above the higher tetrachoric functions are obtained in rapid succession on the machine for a given value of the argument. In the testing of our tetrachoric series seven-place accuracy was aimed at so that it was necessary to calculate τ_1 to eight places, which was done with the help of Vega's ten-figure logarithms.

* *Forlaesninger over Almindelig Iagttagelseslaere*, Kjöbenhaven, 1889.

† *Royal Soc. Proc.* Vol. LVI. p. 271, and in many papers, *Journal of R. Statistical Society.*

‡ *Vorlesungen über die Grundzüge der mathematischen Statistik* (Hamburg, 1920), p. 67.

§ *Biometrika*, Vol. VI. p. 68, 1908.

‖ Arne Fisher, *Casualty, Actuarial and Statistical Society of America. Proceedings*, Vol. IV. Part I. No. 9.

¶ A normal curve, for example, is quite adequately represented by two or three periodic terms; see *Phil. Trans.* Vol. CLXXXVI. A, p. 355, 1895.

(2) It is well known that a wide range of frequency distributions can be adequately represented by one or other of the curves

$$\left.\begin{aligned} y &= y_0\, e^{-\gamma x/a}\left(1+\frac{x}{a}\right)^{p-1} \quad \text{.....}(a)\\ \text{and}\qquad y &= y_0\left(1+\frac{x}{a_1}\right)^{m_1-1}\left(1-\frac{x}{a_2}\right)^{m_2-1} \quad \text{.....}(b)\end{aligned}\right\} \quad \text{(vii).}$$

By a change of origin and the appropriate stretch or squeeze these may be reduced to

$$\left.\begin{aligned} y &= y_0\, x^{p-1}\, e^{-x} \quad \text{.....}(a)\\ \text{and}\qquad y &= y_0\, x^{m_1-1}(1-x)^{m_2-1} \quad \text{.....}(b)\end{aligned}\right\} \quad \text{(vii) bis.}$$

Now, generally, it is not the ordinates of these curves which are required but the areas of certain portions, or in other words the probability integrals of these skew curves. The total range for (vii) bis (a) is 0 to ∞ and for (b) is 0 to 1; since

$$\int_0^\infty x^{p-1}\, e^{-x}\, dx = \Gamma(p)$$

and

$$\int_0^1 x^{m_1-1}(1-x)^{m_2-1}\, dx = \mathrm{B}(m_1, m_2),$$

we may take these probability integrals to be

$$I(p, v) \qquad = \frac{1}{\Gamma(p)}\int_0^v v^{p-1}\, e^{-v}\, dv$$

and

$$\mathrm{B}(v, m_1, m_2) = \frac{1}{\mathrm{B}(m_1, m_2)}\int_0^v v^{m_1-1}(1-v)^{m_2-1}\, dv,$$

which are the ratios of the incomplete to the complete Γ- and B-functions.

The equations on p. 158 show us that if either of the frequency functions (vii) is expressible in a series of tetrachoric functions their probability integrals (assuming convergence) will also be. Now there is no doubt that a large mass of material does not differ practically from the forms in (vii) and accordingly if the above probability integrals cannot be adequately expressed in a series of tetrachoric functions, we may be certain that tetrachoric functions do not furnish a suitable method of representing skew frequency. Accordingly our problem reduces itself to the following one: Can $I(p, v)$ and $\mathrm{B}(v, m_1, m_2)$, or the Incomplete Γ- and B-functions, be represented with adequate convergency by a series of tetrachoric functions? After examination of the numerical and graphical results obtained, we are obliged to conclude that the answer to this question is in the negative.

(3) Let us first consider the expansion in tetrachoric functions of the function

$$y = x^{p-1}\, e^{-x}/\Gamma(p) \quad \text{.....(viii).}$$

In expanding this expression there are at least two methods, which we ought to consider, and one may have advantages over the other as far as convergency is

concerned. It may be expanded with regard: (i) to the mean and the standard deviation, or (ii) to the mode in the manner of Laplace*.

(i) The mean of the function (viii) is easily found to be at $x = p$, the mode is at $x = p - 1$ and the standard deviation is $\sqrt{p}$.

Referring to the mean as origin the function becomes

$$y = \frac{(\xi + p)^{p-1} e^{-(\xi+p)}}{\Gamma(p)} \quad \text{.....(ix).}$$

Let $$y = \phi(-D)\frac{e^{-\xi^2/2p}}{\sqrt{2\pi}}, \text{ where } D = \frac{d}{dz} \text{ and } z = \frac{\xi}{\sqrt{p}} \quad \text{.....(x).}$$

Except for a numerical factor the right-hand side is a series of tetrachoric functions.

Let $$\phi(-D) = c_0 - c_1 D + c_2 D^2 \ldots (-1)^s c_s D^s + \ldots.$$

The function $\phi(-D)$ has to be determined, i.e. we require to find the successive c's:

$$\phi(-D)\left\{\frac{e^{-\xi^2/2p}}{\sqrt{2\pi}}\right\} = \phi(-D)\left\{\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\right\}$$
$$= c_0\tau_1 + c_1\sqrt{2!}\,\tau_2 + c_2\sqrt{3!}\,\tau_3 + \ldots + c_{s-1}\sqrt{s!}\,\tau_s + \ldots \quad \text{(xi).}$$

To determine the c's. With the origin at the mean the function y must be taken as zero from $-\infty$ to $-p$, while from $-p$ to $+\infty$ it is given by (ix). The c's will be obtained most easily by multiplying both sides of (x) by $e^{\theta\xi}$ and equating the coefficients of powers of θ on both sides of the equation, i.e. we make all the moments of the two expressions for the curve the same, for the coefficient of θ^s on either side is the sth moment†. Thus

$$\int_{-\infty}^{\infty} y e^{\theta\xi}\, d\xi = \int_{-\infty}^{\infty} e^{\theta\xi}\, \phi(-D)\frac{e^{-\xi^2/2p}}{\sqrt{2\pi}}\, d\xi;$$

but $y = 0$ from $x = -\infty$ to $-p$.

Accordingly $$\int_{-p}^{\infty} \frac{e^{\theta\xi}(\xi + p)^{p-1} e^{-(\xi+p)}}{\Gamma(p)}\, d\xi = \int_{-\infty}^{\infty} e^{\theta\xi}\, \phi(-D)\frac{e^{-\xi^2/2p}}{\sqrt{2\pi}}\, d\xi.$$

Now $x = p + \xi$ and $z = \xi/\sqrt{p}$.

Thus $$\int_0^{\infty} \frac{e^{\theta(x-p)} x^{p-1} e^{-x}}{\Gamma(p)}\, dx = \sqrt{p}\int_{-\infty}^{\infty} e^{(\theta\sqrt{p}\, z)}\, \phi(-D)\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\, dz.$$

The left-hand side is equal to

$$e^{-p\theta}\int_0^{\infty} \frac{x^{p-1} e^{-x(1-\theta)}}{\Gamma(p)}\, dx$$
$$= e^{-p\theta}\int_0^{\infty} \frac{u^{p-1}}{(1-\theta)^{p-1}}\frac{e^{-u}}{\Gamma(p)}\frac{du}{(1-\theta)} \qquad \text{Let } x(1-\theta) = u.$$
$$= e^{-p\theta}(1-\theta)^{-p}.$$

* Laplace's method is really an expansion in incomplete normal moment functions but as we have seen (p. 158) these may be replaced by tetrachoric functions.

† We owe this elegant method of determining the c's to Mr H. E. Soper. Originally the c's were determined by use of the fundamental property of the tetrachoric functions but that method, while leading to the same result, is more laborious.

To find the value of the integral on the right-hand side, consider the term $c_s(-D)^s$ in the function $\phi(-D)$. Its contribution to the integral is

$$c_s \int_{-\infty}^{\infty} e^{\theta' z} \left(-\frac{d}{dz}\right)^s \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\, dz,$$

where $\theta' = \theta\sqrt{p}$.

On integrating by parts the term between limits vanishes owing to the factor $e^{-\frac{1}{2}z^2}$. Hence the integral

$$= c_s \theta' \int_{-\infty}^{\infty} e^{\theta' z} \left(-\frac{d}{dz}\right)^{s-1} \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\, dz,$$

and ultimately

$$= c_s \theta'^s \int_{-\infty}^{\infty} \frac{e^{\theta' z - \frac{1}{2}z^2}}{\sqrt{2\pi}}\, dz$$

$$= c_s \theta'^s \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}(z-\theta')^2 + \frac{1}{2}\theta'^2}}{\sqrt{2\pi}}\, dz$$

$$= c_s \theta'^s e^{\frac{1}{2}\theta'^2}.$$

Therefore the whole integral on the right is

$$\phi(\theta')\, e^{\frac{1}{2}\theta'^2},$$

i.e. $$e^{-p\theta}(1-\theta)^{-p} = \sqrt{p}\,\phi(\sqrt{p}\theta)\, e^{\frac{1}{2}p\theta^2},$$

or $$\sqrt{p}\,\phi(\sqrt{p}\theta) = e^{-p\theta - \frac{1}{2}p\theta^2}(1-\theta)^{-p},$$

and $$\phi(\sqrt{p}\theta) = c_0 + c_1(\sqrt{p}\theta) + c_2(\sqrt{p}\theta)^2 + \ldots + c_s(\sqrt{p}\theta)^s + \ldots$$

$$= c_0 + c_1'\theta + c_2'\theta^2 + \ldots\ldots\ldots + c_s'\theta^s + \ldots,$$

where $$c_s' = c_s(\sqrt{p})^s \quad \text{or} \quad c_s = c_s'(\sqrt{p})^{-s}.$$

Now $$e^{-p\theta - \frac{1}{2}p\theta^2}(1-\theta)^{-p}$$

$$= e^{-p\theta - \frac{1}{2}p\theta^2 - p\log(1-\theta)}$$

$$= e^{-p\theta - \frac{1}{2}p\theta^2 + p\theta + \frac{1}{2}p\theta^2 + \frac{1}{3}p\theta^3 + \frac{1}{4}p\theta^4 + \ldots}$$

$$= e^{\frac{1}{3}p\theta^3 + \frac{1}{4}p\theta^4 + \frac{1}{5}p\theta^5 + \ldots}$$

$$= b_0 + b_1\theta + b_2\theta^2 + b_3\theta^3 + b_4\theta^4 + \ldots,$$

where

$$b_0 = 1,\ b_1 = b_2 = 0,\ b_3 = \tfrac{1}{3}p,\ b_4 = \tfrac{1}{4}p,\ b_5 = \tfrac{1}{5}p,\ b_6 = \tfrac{1}{6}p + \tfrac{1}{2}(\tfrac{1}{3}p)^2 = \tfrac{1}{18}p(p+3), \text{ etc.}$$

But $\sqrt{p}c_s' = b_s$, therefore

$$c_0 = \frac{1}{\sqrt{p}},\ c_1' = c_2' = 0,\ c_3' = \tfrac{1}{3}\sqrt{p},\ c_4' = \tfrac{1}{4}\sqrt{p},\ c_5' = \tfrac{1}{5}\sqrt{p}, \text{ etc.}$$

so that $$c_0 = \frac{1}{\sqrt{p}},\ c_1 = c_2 = 0,\ c_3 = \tfrac{1}{3}\frac{1}{p},\ c_4 = \tfrac{1}{4}\frac{1}{p\sqrt{p}},\ c_5 = \tfrac{1}{5}\frac{1}{p^2}, \text{ etc.}$$

For numerical purposes these coefficients are much more usefully obtained in the following way:

Let $$e^{-p\theta - \frac{1}{2}p\theta^2}(1-\theta)^{-p} = b_0 + b_1\theta + b_2\theta^2 + \ldots \text{ etc.}$$

Take the differential of the logarithms of both sides; then

$$(-p - p\theta + p/1-\theta)(b_0 + b_1\theta + b_2\theta^2 + \ldots + b_s\theta^s + \ldots) = b_1 + 2b_2\theta + \ldots + sb_s\theta^{s-1} + \ldots,$$

i.e. $$p\theta^2(b_0 + b_1\theta + \ldots + b_s\theta^s + \ldots) = (1-\theta)(b_1 + 2b_2\theta + \ldots + sb_s\theta^{s-1} + \ldots).$$

Equating coefficients of θ^s we have

$$pb_{s-2} = (s+1)\,b_{s+1} - sb_s,$$

i.e.
$$b_{s+1} = \frac{1}{s+1}\{sb_s + pb_{s-2}\} \quad \text{.....(xii).}$$

By this difference formula successive b's can be found very quickly if b_0, b_1, b_2 are known and we have already found these.

Now
$$c_s = (\sqrt{p})^{-s}\, c_s'$$
$$= (\sqrt{p})^{-s}\, b_s (\sqrt{p})^{-1} = b_s/(\sqrt{p})^{s+1},$$

or
$$b_s = (\sqrt{p})^{s+1}\, c_s.$$

Substituting in (xii)

$$(\sqrt{p})^{s+2}\, c_{s+1} = \frac{1}{(s+1)}\{s(\sqrt{p})^{s+1}\, c_s + p(\sqrt{p})^{s-1}\, c_{s-2}\}$$
$$= \frac{(\sqrt{p})^{s+1}}{(s+1)}\{sc_s + c_{s-2}\},$$

or
$$c_{s+1} = \frac{1}{\sqrt{p}}\,\frac{1}{(s+1)}\{sc_s + c_{s-2}\} \quad \text{.....(xiii).}$$

This formula gives us very readily the coefficients of $\phi(-D)$ and thus the expansion is obtained.

We had, Equation (xi),

$$\frac{x^{p-1}e^{-x}}{\Gamma(p)} = \phi(-D)\,\frac{e^{-\xi^2/2p}}{\sqrt{2\pi}} = c_0\tau_1 + c_1\sqrt{2!}\,\tau_2 + c_2\sqrt{3!}\,\tau_3 + \ldots + c_s\sqrt{(s+1)!}\,\tau_{s+1} + \ldots,$$

and all the c's are known since $c_0 = \frac{1}{\sqrt{p}}$, $c_1 = c_2 = 0$.

To find the area under the curve (xi) up to abscissa x, remembering that the left-hand side is zero from $\xi = -\infty$ to $-p$,

$$\int_{-p}^{\xi} \frac{(\xi+p)^{p-1}e^{-(\xi+p)}}{\Gamma(p)}\,d\xi = \int_{-\infty}^{\xi} \phi(-D)\,\frac{e^{-\xi^2/2p}}{\sqrt{2\pi}}\,d\xi,$$

i.e.
$$\int_0^x \frac{x^{p-1}e^{-x}}{\Gamma(p)}\,dx = \sqrt{p}\int_{-\infty}^{z} \phi(-D)\,\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\,dz$$
$$= \sqrt{p}\int_{-\infty}^{z} (c_0\tau_1 + c_1\sqrt{2!}\,\tau_2 + \ldots + c_{s-1}\sqrt{s!}\,\tau_s + \ldots)\,dz.$$

Now
$$\int_{-\infty}^{z} \tau_s\,dz = -\frac{\tau_{s-1}}{\sqrt{s}},$$

therefore

$$\int_0^x \frac{x^{p-1}e^{-x}}{\Gamma(p)}\,dx = \sqrt{p}\left[c_0\int_{-\infty}^{z}\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\,dz - c_1\frac{\sqrt{2!}\,\tau_1}{\sqrt{2}} - c_2\frac{\sqrt{3!}\,\tau_2}{\sqrt{3}} - \ldots - c_{s-1}\frac{\sqrt{s!}\,\tau_{s-1}}{\sqrt{s}} - \ldots\right]$$

$$= \tfrac{1}{2}(1+\alpha_z) - \sqrt{p}\,\{c_1\tau_1 + c_2\sqrt{2!}\,\tau_2 + \ldots + c_{s-1}\sqrt{(s-1)!}\,\tau_{s-1} + \ldots\} \text{ since } c_0 = \frac{1}{\sqrt{p}}.$$

Therefore finally

$$\int_0^x \frac{x^{p-1}\,e^{-x}}{\Gamma(p)}\,dx = \tfrac{1}{2}(1+\alpha_z) - a_3\tau_3 - \ldots - a_s\tau_s - \ldots \quad \ldots\ldots\ldots\ldots\text{(xiv)},$$

(since $c_1 = c_2 = 0$) where $a_s = \sqrt{p}\,c_s\sqrt{s!}$.

Now $c_{s+1} = \dfrac{1}{\sqrt{p}}\,\dfrac{1}{s+1}\{sc_s + c_{s-2}\}$ from equation (xiii),

i.e.
$$\frac{a_{s+1}}{\sqrt{(s+1)!}} = \frac{1}{\sqrt{p}}\,\frac{1}{(s+1)}\left\{s\,\frac{a_s}{\sqrt{s!}} + \frac{a_{s-2}}{\sqrt{(s-2)!}}\right\},$$

therefore
$$a_{s+1} = \frac{1}{\sqrt{p}}\,\frac{1}{(s+1)}\{s\sqrt{(s+1)}\,a_s + \sqrt{(s+1)(s)(s-1)}\,a_{s-2}\}$$

$$= \sqrt{\frac{s}{p(s+1)}}\{\sqrt{s}\,a_s + \sqrt{(s-1)}\,a_{s-2}\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xv)},$$

where
$$a_0 = 1, \quad a_1 = a_2 = 0.$$

The argument of $\frac{1}{2}(1+\alpha)$ and of the tetrachoric functions is $\xi/\sqrt{p}$, which equals $\dfrac{x-p}{\sqrt{p}} = z$, say.

Since the terms τ_1 and τ_2 do not appear one might hope that only a few terms of the expansion (xiv) would be required to obtain a sufficiently accurate result.

$\frac{1}{2}(1+\alpha_z)$ is the ordinary probability integral at z.

Note that if x is less than p, i.e. z is negative, $\frac{1}{2}(1-\alpha_z)$ must be used instead of $\frac{1}{2}(1+\alpha_z)$ and the tetrachoric functions of even order must be taken of opposite sign to those for positive z such as are given in the tables. The odd order functions are the same for positive and negative z:

$$\tau_{2s}(z) = -\tau_{2s}(-z), \quad \tau_{2s+1}(z) = \tau_{2s+1}(-z).$$

Obviously we could get the area of any portion of the curve between $x = x_1$ and $x = x_2$ by subtracting two expressions like (xiv) for z_1 and z_2.

The general expression for $\dfrac{x^{p-1}\,e^{-x}}{\Gamma(p)}$ is

$$\begin{aligned}
\frac{x^{p-1}\,e^{-x}}{\Gamma(p)} &= \frac{1}{\sqrt{p}}\tau_1 + \frac{1}{p}\,\frac{\sqrt{4!}}{3}\tau_4 + \frac{1}{p\sqrt{p}}\,\frac{\sqrt{5!}}{4}\tau_5 \\
&+ \frac{1}{p^2}\,\frac{\sqrt{6!}}{5}\tau_6 + \frac{1}{p^2\sqrt{p}}\,\frac{\sqrt{7!}}{6}\left(\frac{p+3}{3}\right)\tau_7 \\
&+ \frac{1}{p^3}\,\frac{\sqrt{8!}}{7}\left\{\frac{7p+12}{12}\right\}\tau_8 + \frac{1}{p^3\sqrt{p}}\,\frac{\sqrt{9!}}{8}\left\{\frac{47p+60}{60}\right\}\tau_9 \\
&+ \frac{1}{p^4}\,\frac{\sqrt{10!}}{9}\left\{\frac{p^2}{18} + \frac{19}{20}p + 1\right\}\tau_{10} + \frac{1}{p^4\sqrt{p}}\,\frac{\sqrt{11!}}{10}\left\{\frac{5}{36}p^2 + \frac{153}{140}p + 1\right\}\tau_{11} \\
&+ \frac{1}{p^5}\,\frac{\sqrt{12!}}{11}\left\{\frac{341}{1440}p^2 + \frac{341}{280}p + 1\right\}\tau_{12} \\
&\qquad + \frac{1}{p^5\sqrt{p}}\,\frac{\sqrt{13!}}{12}\left\{\frac{p^3}{162} + \frac{493}{1440}p^2 + \frac{3349}{2520}p + 1\right\}\tau_{13} + \ldots,
\end{aligned}$$

and

$$\int_0^x \frac{x^{p-1}e^{-x}}{\Gamma(p)}dx = \tfrac{1}{2}(1+\alpha_z) - \frac{1}{\sqrt{p}}\frac{\sqrt{4!}}{3\sqrt{4}}\tau_3 - \frac{1}{p}\frac{\sqrt{5!}}{4\sqrt{5}}\tau_4$$
$$- \frac{1}{p\sqrt{p}}\frac{\sqrt{6!}}{5\sqrt{6}}\tau_5 - \frac{1}{p^2}\frac{\sqrt{7!}}{6\sqrt{7}}\left(\frac{p+3}{3}\right)\tau_6 - \frac{1}{p^2\sqrt{p}}\frac{\sqrt{8!}}{7\sqrt{8}}\left(\frac{7p+12}{12}\right)\tau_7$$
$$- \frac{1}{p^3}\frac{\sqrt{9!}}{8\sqrt{9}}\left\{\frac{47p+60}{60}\right\}\tau_8 - \frac{1}{p^3\sqrt{p}}\frac{\sqrt{10!}}{9\sqrt{10}}\left\{\frac{p^2}{18}+\frac{19}{20}p+1\right\}\tau_9$$
$$- \frac{1}{p^4}\frac{\sqrt{11!}}{10\sqrt{11}}\left\{\frac{5}{36}p^2+\frac{153}{140}p+1\right\}\tau_{10} - \frac{1}{p^4\sqrt{p}}\frac{\sqrt{12!}}{11\sqrt{12}}\left\{\frac{341}{1440}p^2+\frac{341}{280}p+1\right\}\tau_{11}$$
$$- \frac{1}{p^5}\frac{\sqrt{13!}}{12\sqrt{13}}\left\{\frac{p^3}{162}+\frac{493}{1440}p^2+\frac{3349}{2520}p+1\right\}\tau_{12} - \ldots$$

$$= \tfrac{1}{2}(1+\alpha_z) - \frac{\cdot 8164,9658}{\sqrt{p}}\tau_3 - \frac{1\cdot 2247,4487}{p}\tau_4$$
$$- \frac{2\cdot 1908,9023}{p\sqrt{p}}\tau_5 - \frac{1\cdot 4907,1198}{p^2}(p+3)\tau_6$$
$$- \frac{\cdot 8451,5425}{p^2\sqrt{p}}(7p+12)\tau_7 - \frac{\cdot 4183,3001}{p^3}(47p+60)\tau_8$$
$$- \frac{\cdot 3718,4890}{p^3\sqrt{p}}(10p^2+171p+180)\tau_9 - \frac{\cdot 1511,8579}{p^4}(175p^2+1377p+1260)\tau_{10}$$
$$- \frac{\cdot 0569,8743}{p^4\sqrt{p}}(2387p^2+12276p+10080)\tau_{11}$$
$$- \frac{\cdot 0201,0408}{p^5}(560p^3+31059p^2+120564p+90720)\tau_{12} - \ldots.$$

(ii) *Laplacian Form of Expansion.*

This is an expansion with regard to the mode or maximum ordinate as origin. The mode of $y = \dfrac{x^{p-1}e^{-x}}{\Gamma(p)}$ is at $x=(p-1)$, so that it will be easier to deal with y in the form

$$y = \frac{x^{p'}e^{-x}}{\Gamma(p'+1)},$$

where $p'=(p-1)$.

Let $x=p'+\xi$, i.e. take the mode as origin. Then as before we require to find $\phi(-D)$ so that

$$\frac{(p'+\xi)^{p'}e^{-(p'+\xi)}}{\Gamma(p'+1)} = \phi(-D)\frac{e^{-\frac{1}{2}\xi^2/p'}}{\sqrt{2\pi p'}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xvi)},$$

where $D=\dfrac{d}{dz}$ and $z=\dfrac{\xi}{\sqrt{p'}}$.

The introduction of $\sqrt{p'}$ in the denominator simplifies the integration a little.

Proceeding as before:

$$\int_{-p'}^{\infty} \frac{e^{\theta\xi}(p'+\xi)^{p'}e^{-(p'+\xi)}}{\Gamma(p'+1)}\,d\xi = \int_{-\infty}^{\infty} e^{\theta\xi}\,\phi(-D)\,\frac{e^{-\frac{1}{2}\xi^2/p'}}{\sqrt{2\pi p'}}\,d\xi,$$

i.e.
$$\int_{0}^{\infty} \frac{e^{\theta(x-p')}x^{p'}e^{-x}}{\Gamma(p'+1)}\,dx = \int_{-\infty}^{\infty} e^{\theta\sqrt{p'}z}\,\phi(-D)\,\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\,dz,$$

or
$$\frac{e^{-p'\theta}}{(1-\theta)^{p'+1}} = \int_{-\infty}^{\infty} e^{(\theta\sqrt{p'})z}\,\phi(-D)\,\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\,dz,$$

and
$$e^{-p'\theta}(1-\theta)^{-(p'+1)} = \phi(\theta\sqrt{p'})\int_{-\infty}^{\infty}\frac{e^{\theta\sqrt{p'}z-\frac{1}{2}z^2}}{\sqrt{2\pi}}\,dz$$

$$= \phi(\theta\sqrt{p'})\,e^{\frac{1}{2}p'\theta^2}\int_{-\infty}^{\infty}\frac{e^{-\frac{1}{2}(z-\theta\sqrt{p'})^2}}{\sqrt{2\pi}}\,dz$$

$$= \phi(\theta\sqrt{p'})\,e^{\frac{1}{2}p'\theta^2};$$

therefore
$$\phi(\theta\sqrt{p'}) = e^{-p'\theta-\frac{1}{2}p'\theta^2}(1-\theta)^{-(p'+1)} \quad \text{.....(xvii).}$$

Now if
$$\phi(-D) = c_0 - c_1D + c_2D^2 + \ldots + (-1)^s c_sD^s + \ldots,$$

$$\phi(\theta\sqrt{p'}) = c_0 + c_1(\theta\sqrt{p'}) + c_2(\theta\sqrt{p}\,)^2 + \ldots + c_s(\theta\sqrt{p'})^s + \ldots$$

$$= c_0 + c_1'\theta + c_2'\theta^2 + \ldots + c_s'\theta^s + \ldots,$$

where:
$$c_s' = c_s(\sqrt{p'})^s \quad \text{or} \quad c_s = c_s'(\sqrt{p'})^{-s},$$

$$e^{-p'\theta-\frac{1}{2}p'\theta^2}(1-\theta)^{-(p'+1)} = e^{-p'\theta-\frac{1}{2}p'\theta^2-(p'+1)\log(1-\theta)}$$

$$= e^{-p'\theta-\frac{1}{2}p'\theta^2+(p'+1)\left(\theta+\frac{\theta^2}{2}+\frac{\theta^3}{3}+\ldots\right)}$$

$$= e^{\theta+\frac{1}{2}\theta^2+\frac{(p'+1)}{3}\theta^3+\frac{(p'+1)}{4}\theta^4+\ldots}$$

$$= c_0 + c_1'\theta + c_2'\theta^2 + \ldots,$$

where:
$$c_0 = 1,\ c_1' = 1,\ c_2' = 1,\ c_3' = \tfrac{1}{6} + \frac{p'+1}{3} + \tfrac{1}{2} = \frac{p'+3}{3},$$

and generally by differentiating

$$c_s' = c'_{s-1} + \frac{p'}{s}\,c'_{s-3},$$

or
$$c_s(\sqrt{p'})^s = c_{s-1}(\sqrt{p'})^{s-1} + \frac{p'}{s}\,c_{s-3}(\sqrt{p'})^{s-3};$$

thus
$$c_s = \frac{1}{\sqrt{p'}}\left\{c_{s-1} + \frac{1}{s}\,c_{s-3}\right\} \quad \text{.....(xviii),}$$

where
$$c_0 = 1,\ c_1 = \frac{1}{\sqrt{p'}},\ c_2 = \frac{1}{p'}.$$

Therefore

$$\frac{(p'+\xi)^{p'}e^{-(p'+\xi)}}{\Gamma(p'+1)} = \{c_0 - c_1D + c_2D^2 - \ldots(-1)^s c_sD^s - \ldots\}\left(\frac{e^{-\xi^2/2p'}}{\sqrt{2\pi p'}}\right)$$

$$= \frac{1}{\sqrt{p'}}\{c_0\tau_1 + \sqrt{2!}\,c_1\tau_2 + \sqrt{3!}\,c_2\tau_3 + \ldots + \sqrt{(s+1)!}\,c_s\tau_{s+1} + \ldots\}$$

To find the area up to abscissa x we have

$$\int_{-p'}^{\xi} \frac{(p'+\xi)^{p'} e^{-(p'+\xi)}}{\Gamma(p'+1)} d\xi = \frac{1}{\sqrt{p'}} \int_{-\infty}^{\xi} \{c_0\tau_1 + \sqrt{2\,!}\, c_1\tau_2 + \dots + \sqrt{(s+1)\,!}\, c_s\tau_{s+1} + \dots\} d\xi$$

$$= \int_{-\infty}^{z} \{c_0\tau_1 + \sqrt{2\,!}\, c_1\tau_2 + \dots + \sqrt{(s+1)\,!}\, c_s\tau_{s+1} + \dots\} dz$$

$$= \tfrac{1}{2}(1+\alpha_z) - c_1\tau_1 - \sqrt{2\,!}\, c_2\tau_2 - \sqrt{3\,!}\, c_3\tau_3 - \dots - \sqrt{s\,!}\, c_s\tau_s - \dots \text{ as } c_0 = 1,$$

i.e. $$\int_0^x \frac{x^{p'} e^{-x}}{\Gamma(p'+1)} dx = \tfrac{1}{2}(1+\alpha_z) - a_1'\tau_1 - a_2'\tau_2 - a_3'\tau_3 - \dots - a_s'\tau_s - \dots,$$

where $$a_s' = c_s\sqrt{s\,!}.$$

Substituting in (xviii) to obtain the difference equation for the a''s we have

$$\frac{a'_s}{\sqrt{s\,!}} = \frac{1}{\sqrt{p'}}\left\{\frac{a'_{s-1}}{\sqrt{(s-1)\,!}} + \frac{1}{s}\frac{a'_{s-3}}{\sqrt{(s-3)\,!}}\right\};$$

therefore $$a_s' = \frac{1}{\sqrt{p'}}\frac{1}{\sqrt{s}}\{s a'_{s-1} + \sqrt{(s-1)(s-2)}\, a'_{s-3}\} \quad \text{..............(xix)},$$

and $$a_0' = 1, \quad a_1' = \frac{1}{\sqrt{p'}}, \quad a_2' = \frac{\sqrt{2}}{p'}.$$

By this formula the a's are readily obtained numerically. It is to be noted that in this case the terms in τ_1 and τ_2 do not vanish, as they did in the expansion from the mean. The argument of $\frac{1}{2}(1+\alpha_z)$ and of the tetrachoric functions is $\frac{x-p'}{\sqrt{p'}}$, and the remarks with regard to sign made above must be again observed.

Coefficients in the expansion from the mode:

$$a_0' = 1, \quad a_1' = \frac{1}{\sqrt{p'}}, \quad a_2' = \frac{\sqrt{2\,!}}{p'},$$

$$a_3' = \frac{\sqrt{3\,!}}{p'\sqrt{p'}}\left\{\frac{p'+3}{3}\right\}, \quad a_4' = \frac{\sqrt{4\,!}}{p'^2}\left\{\frac{7p'+12}{12}\right\}, \quad a_5' = \frac{\sqrt{5\,!}}{p'^2\sqrt{p'}}\left\{\frac{47p'+60}{60}\right\},$$

$$a_6' = \frac{\sqrt{6\,!}}{p'^3}\left\{\frac{p'^2}{18} + \frac{19}{20}p' + 1\right\}, \quad a_7' = \frac{\sqrt{7\,!}}{p'^3\sqrt{p'}}\left\{\frac{5}{36}p'^2 + \frac{153}{140}p' + 1\right\},$$

$$a_8' = \frac{\sqrt{8\,!}}{p'^4}\left\{\frac{341}{1440}p'^2 + \frac{341}{280}p' + 1\right\}, \quad a_9' = \frac{\sqrt{9\,!}}{p'^4\sqrt{p'}}\left\{\frac{p'^3}{162} + \frac{493}{1440}p'^2 + \frac{3349}{2520}p' + 1\right\},$$

....

$$a_s' = \frac{1}{\sqrt{p'}}\left\{\sqrt{s}\, a'_{s-1} + \sqrt{\frac{(s-1)(s-2)}{s}}\, a'_{s-3}\right\}.$$

We note that the coefficients of powers of θ in the functions

$$\phi(\theta) = e^{p\left\{\frac{\theta^3}{3} + \frac{\theta^4}{4} + \frac{\theta^5}{5} + \dots\right\}}$$

and $$\phi'(\theta) = e^{\theta + \frac{1}{2}\theta^2 + \frac{p+1}{3}\theta^3 + \frac{p+1}{4}\theta^4 + \dots}$$

(in the expansion from the mode we had p' for p in $\phi'(\theta)$) are closely related.

Then if c_n is the coefficient of θ^n in $\phi(\theta)$ and c_n' is the coefficient of θ^n in $\phi'(\theta)$,

$$c_n = \frac{p}{n} c'_{n-3}.$$

(4) In the last expansion it might seem possible to get rid of the terms in τ_1 and τ_2 by breaking away from Laplace and expanding with regard to $e^{-\frac{1}{2}\xi^2/q}$ instead of $e^{-\frac{1}{2}\xi^2/p'}$; then choose q to give us the desired result. In Laplace's form of the modal expansion the exponential term is $e^{\frac{1}{2}\left(\frac{d^2u}{dx^2}\right)_{m_0}\xi^2}$, where $u = \log y$ and $\left(\frac{d^2u}{dx^2}\right)_{m_0}$ means the value of $\frac{d^2u}{dx^2}$ at the mode.

Now
$$y = \frac{x^{p'} e^{-x}}{\Gamma(p'+1)},$$
$$u = \log_e y = p' \log_e x - x - \log_e \Gamma(p'+1),$$
$$\frac{du}{dx} = \frac{p'}{x} - 1,$$
$$\frac{d^2u}{dx^2} = -\frac{p'}{x^2};$$

therefore
$$\left(\frac{d^2u}{dx^2}\right)_{\text{Mode}} = -\frac{p'}{p'^2} = -\frac{1}{p'}.$$

If $\frac{x^{p'} e^{-x}}{\Gamma(p'+1)} = \phi(-D)\frac{e^{-\xi^2/2q}}{\sqrt{2\pi q}}$ where $D = \frac{d}{dz}$ and $z = \frac{\xi}{\sqrt{q}}$, we have to find q, so that either the τ_1 or τ_2 term or both will vanish.

By proceeding as before equation (xvii) becomes

$$\phi(\theta\sqrt{q}) = e^{-p'\theta - \frac{1}{2}q\theta^2}(1-\theta)^{-(p'+1)}$$
$$= e^{-p'\theta - \frac{1}{2}q\theta^2 - (p'+1)\log(1-\theta)}.$$

The term in τ_2 will vanish if $q = p' + 1$ which is the square of the standard-deviation from the mean, but τ_1 will still be left. However, it does not seem likely that any advantage will be gained by departing from Laplace's form of the exponential term.

Having found the two expansions from the mean and the mode respectively we shall now proceed to examine the behaviour of the series by numerical calculation, but before doing so we shall endeavour to find a similar series for the Incomplete B-function.

(5) To expand $\int_0^x \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx$ in terms of tetrachoric functions about the mean.

The mean is at
$$x = p/(p+q).$$

The standard deviation is
$$\sigma = \frac{\sqrt{pq}}{(p+q)\sqrt{p+q+1}}.$$

Take origin at the mean; then $x = p/(p+q) + \xi$. Let

$$\frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} = \phi(-D)\frac{e^{-\frac{1}{2}\xi^2/\sigma^2}}{\sqrt{2\pi\sigma}} \quad \text{.....(xx),}$$

where
$$D = \frac{d}{dy},\ y = \frac{\xi}{\sigma}.$$

As in the case of the Incomplete Γ-function multiply each side by $e^{\theta\xi}$ and integrate. The limits of the integral on the left-hand side will be $x=0$ and $x=1$, as we take the value of the integral outside these limits to be zero.

The ξ limits will therefore be $-p/(p+q)$ for $x=0$ and $q/(p+q)$ for $x=1$. Then

$$\int_{-p/p+q}^{q/p+q} \frac{e^{\theta\xi}(\xi+p/(p+q))^{p-1}}{\mathrm{B}(p,q)}\{1-(\xi+p/(p+q))\}^{q-1}\,d\xi=\int_{-\infty}^{\infty} e^{\theta\xi}\,\phi(-D)\frac{e^{-\frac{1}{2}\xi/\sigma^2}}{\sqrt{2\pi}\sigma}\,d\xi,$$

i.e.
$$\int_0^1 \frac{e^{\theta(x-p/(p+q))}x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx=\int_{-\infty}^{\infty} e^{(\theta\sigma)y}\,\phi(-D)\frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\,dy,$$

and
$$e^{-\theta p/p+q}\int_0^1 \frac{e^{\theta x}x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx=\phi(\theta\sigma)\int_{-\infty}^{\infty}\frac{e^{(\theta\sigma)y-\frac{1}{2}y^2}}{\sqrt{2\pi}}\,dy$$

$$=\phi(\theta\sigma)\int_{-\infty}^{\infty}\frac{e^{-\frac{1}{2}(y-\theta\sigma)^2+\frac{1}{2}\theta^2\sigma^2}}{\sqrt{2\pi}}\,dy$$

$$=\phi(\theta\sigma)\,e^{\frac{1}{2}\theta^2\sigma^2} \quad \text{.....(xxi).}$$

Now

$$\int_0^1 \frac{x^{p-1}(1-x)^{q-1}e^{\theta x}}{\mathrm{B}(p,q)}\,dx$$

$$=\int_0^1 \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\left\{1+\theta x+\frac{\theta^2x^2}{2!}+\frac{\theta^3x^3}{3!}+\ldots+\frac{\theta^sx^s}{s!}+\ldots\right\}dx$$

$$=\frac{\mathrm{B}(p,q)}{\mathrm{B}(p,q)}+\theta\frac{\mathrm{B}(p+1,q)}{\mathrm{B}(p,q)}+\frac{\theta^2}{2!}\frac{\mathrm{B}(p+2,q)}{\mathrm{B}(p,q)}+\ldots+\frac{\theta^s}{s!}\frac{\mathrm{B}(p+s,q)}{\mathrm{B}(p,q)}+\ldots.$$

But
$$\frac{\mathrm{B}(p+s,q)}{\mathrm{B}(p,q)}=\frac{p(p+1)\ldots(p+s-1)}{(p+q)(p+q+1)\ldots(p+q+s-1)},$$

therefore
$$\int_0^1 \frac{x^{p-1}(1-x)^{q-1}e^{\theta x}}{\mathrm{B}(p,q)}\,dx=1+\theta\frac{p}{p+q}+\frac{\theta^2}{2!}\frac{p(p+1)}{(p+q)(p+q+1)}$$

$$+\ldots+\frac{\theta^s}{s!}\frac{p(p+1)\ldots(p+s-1)}{(p+q)(p+q+1)\ldots(p+q+s-1)}+\ldots.$$

From equation (xxi)

$$\phi(\theta\sigma)=e^{-\frac{p\theta}{p+q}-\frac{1}{2}\theta^2\sigma^2}\left\{1+\theta\frac{p}{p+q}+\frac{\theta^2}{2!}\frac{p(p+1)}{(p+q)(p+q+1)}\right.$$

$$\left.+\ldots+\frac{\theta^s}{s!}\frac{p(p+1)\ldots(p+s-1)}{(p+q)(p+q+1)\ldots(p+q+s-1)}+\ldots\right\}\text{.....(xxii).}$$

Let
$$\phi(-D)=a_0-a_1D+a_2D^2-\ldots(-1)^s a_sD^s+\ldots,$$

$$\phi(\theta\sigma)=a_0+a_1(\theta\sigma)+a_2(\theta\sigma)^2+a_3(\theta\sigma)^3+\ldots+a_s(\theta\sigma)^s+\ldots$$

$$=c_0+c_1\theta+c_2\theta^2+\ldots+c_s\theta^s+\ldots,$$

where
$$c_s=a_s\sigma^s.$$

By equating coefficients of powers of θ in equation (xxii) the coefficients in $\phi(\theta\sigma)$ can be obtained in terms of p and q, for

$$\sigma^2=\frac{pq}{(p+q)^2(p+q+1)}.$$

Obviously

$$c_0 = 1,$$

$$c_1 = -p/(p+q) + p/(p+q) = 0,$$

$$c_2 = \frac{1}{2!}\frac{p^2}{(p+q)^2} - \frac{1}{2}\frac{pq}{(p+q)^2(p+q+1)} + \frac{1}{2!}\frac{p(p+1)}{(p+q)(p+q+1)} - \frac{p^2}{(p+q)^2}$$

$$= \frac{1}{2}\left\{\frac{p}{p+q}\right\}\left\{\frac{p+1}{(p+q+1)} - \frac{p}{p+q} - \frac{q}{(p+q)(p+q+1)}\right\}$$

$$= \frac{1}{2}\left\{\frac{p}{p+q}\right\}\left\{\frac{q}{(p+q)(p+q+1)} - \frac{q}{(p+q)(p+q+1)}\right\}$$

$$= 0.$$

Similarly the other c's can be determined but the work becomes more and more laborious as we go on.

Unfortunately, as far as the numerical work is concerned, we have failed after many attempts to find a relation connecting successive c's, similar to that found in the case of the Incomplete Γ-function. At first it was thought that the following treatment would facilitate the numerical calculation of these coefficients.

Let $$e^{-\frac{p\theta}{p+q} - \frac{1}{2}\sigma^2\theta^2} = b_0 + b_1\theta + b_2\theta^2 + \ldots + b_s\theta^s + \ldots,$$

then $$-p/(p+q)\,\theta - \tfrac{1}{2}\sigma^2\theta^2 = \log_e\{b_0 + b_1\theta + b_2\theta^2 + \ldots + b_s\theta^s + \ldots\}.$$

Differentiate this and then equate coefficients of powers of θ:

$$(b_0 + b_1\theta + b_2\theta^2 + \ldots + b_s\theta^s + \ldots)(-p/(p+q) - \sigma^2\theta)$$
$$= b_1 + 2b_2\theta + 3b_3\theta^2 + \ldots + sb_s\theta^{s-1} + \ldots.$$

Equate coefficients of θ^{s-1}:

$$sb_s = -p/(p+q)\,b_{s-1} - \sigma^2 b_{s-2};$$

therefore
$$b_s = -\frac{1}{s}\left[\frac{p}{p+q}b_{s-1} + \sigma^2 b_{s-2}\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(xxiii)}.$$

This formula enables us to calculate the b's very rapidly on the machine when $p/(p+q)$ and σ^2 have been determined.

From equation (xxii)

$$c_0 + c_1\theta + c_2\theta^2 + \ldots + c_s\theta^s + \ldots = (b_0 + b_1\theta + b_2\theta^2 + \ldots)$$
$$\left\{1 + \theta\frac{p}{p+q} + \frac{\theta^2}{2!}\frac{p(p+1)}{(p+q)(p+q+1)} + \ldots + \ldots\right\}.$$

Equate coefficients of θ^s:

$$c_s = b_0\frac{1}{s!}\frac{p(p+1)\ldots(p+s-1)}{(p+q)(p+q+1)\ldots(p+q+s-1)}$$
$$+ b_1\frac{1}{(s-1)!}\frac{p(p+1)\ldots(p+s-2)}{(p+q)(p+q+1)(p+q+s-2)} + \ldots + b_{s-1}\frac{1}{1!}\frac{p}{p+q} + b_s,$$

i.e.
$$c_s = \sum_{r=0}^{s} b_r\frac{1}{(s-r)!}\frac{p(p+1)\ldots(p+s-r-1)}{(p+q)(p+q+1)\ldots(p+q+s-r-1)}.$$

The b's, having been calculated previously by (xxiii), this last formula gives a fairly rapid way of calculating the c's, at least the earlier c's. Then

$$a_s = \frac{1}{\sigma^s} \sum_{r=0}^{s} b_r \frac{1}{(s-r)!} \frac{p(p+1)\dots(p+s-r-1)}{(p+q)(p+q+1)\dots(p+q+s-r-1)} \quad (a_0 = 1) \dots \text{(xxiv)}.$$

What we require generally is the area represented by $\int_0^x x^{p-1}(1-x)^{q-1}\,dx$:

$$\int_0^x \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx = \int_{-\infty}^{\xi} \phi(-D)\frac{e^{-\frac{1}{2}\xi^2/\sigma^2}}{\sqrt{2\pi}\,\sigma}\,d\xi$$

$$= \int_{-\infty}^{y} \phi(-D)\frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\,dy,$$

i.e. $$\int_0^x \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx$$

$$= \int_{-\infty}^{y} \left\{ a_0 \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}} - a_1 D \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}} + a_2 D^2 \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}} - \dots (-1)^s D^s \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}} - \dots \right\} dy$$

$$= \int_{-\infty}^{y} a_0 \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\,dy - a_1 \left[\frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\right]_{-\infty}^{y} + a_2 \left[D\frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\right]_{-\infty}^{y} - \dots + a_s \left[(-1)^s D^{s-1}\frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}\right]_{-\infty}^{y} - \dots$$

$$= \tfrac{1}{2}(1+\alpha) - a_1\sqrt{1!}\,\tau_1 - a_2\sqrt{2!}\,\tau_2 - a_3\sqrt{3!}\,\tau_3 - \dots - a_s\sqrt{s!}\,\tau_s - \dots \quad (a_0 = 1)$$

$$= \tfrac{1}{2}(1+\alpha) - a_1'\tau_1 - a_2'\tau_2 - a_3'\tau_3 - \dots - a_s'\tau_s - \dots,$$

where $$a_s' = a_s\sqrt{s!}.$$

Then $$a_s' = \frac{\sqrt{s!}}{\sigma^s} \sum_{r=0}^{s} b_r \frac{1}{(s-r)!} \frac{p(p+1)\dots(p+s-r-1)}{(p+q)(p+q+1)\dots(p+q+s-r-1)} \dots \text{(xxv)}.$$

Now c_1 and c_2 are equal to zero, so that a_1', a_2' are zero. Thus there are no terms in τ_1 and τ_2. The argument of the tetrachoric functions and of $\frac{1}{2}(1+\alpha)$ is y, which is equal to $\dfrac{\xi}{\sigma} = \dfrac{x - p/(p+q)}{\sigma}$. On applying the above formula for a_s', we were greatly disappointed to find, that with the b's to 8 decimal places the expression under the summation sign in the examples used commenced with 4 or 5 zeros after the decimal point. As $\sqrt{s!}$ and $\left(\dfrac{1}{\sigma}\right)^s$ both increase with s $\left(\dfrac{1}{\sigma}\right.$ being in our case $\left.> 1\right)$ accuracy to the seventh place in our a''s could not be obtained. Accordingly the formula actually used was of a different type.

Let $$\frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} = \sum_{1}^{\infty} c_s \tau_s,$$

where the argument of the tetrachoric function is again $\dfrac{\xi}{\sigma} = \dfrac{x - p/(p+q)}{\sigma}$.

Multiply both sides by τ_s, weighting by the factor $e^{\frac{1}{2}\xi^2/\sigma^2}$, and integrate from $-\infty$ to $+\infty$, the left-hand side being taken as zero outside $x = 0$ and $x = 1$.

Then $$\int_{-p/p+q}^{q/p+q} \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} \tau_s e^{\frac{1}{2}\xi^2/\sigma^2}\,d\xi = \int_{-\infty}^{\infty} \tau_s \sum_{1}^{\infty} c_s \tau_s e^{\frac{1}{2}\xi^2/\sigma^2}\,d\xi.$$

Since $\int_{-\infty}^{\infty} \tau_s \tau_{s'} e^{\frac{1}{2}\xi^2/\sigma^2} d\xi = 0$ only the term in τ_s^2 will be left on the right-hand side, i.e.

$$\int_{-\infty}^{\infty} \tau_s \overset{\infty}{S}_1 c_s \tau_s e^{\frac{1}{2}\xi^2/\sigma^2} d\xi = c_s \int_{-\infty}^{\infty} \tau_s^2 e^{\frac{1}{2}\xi^2/\sigma^2} d\xi.$$

Putting $\xi/\sigma = y,\ d\xi = \sigma dy,$

we have $$\int_{-p/p+q}^{q/p+q} \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} \tau_s e^{\frac{1}{2}\xi^2/\sigma^2} d\xi = c_s \int_{-\infty}^{\infty} \tau_s^2 e^{\frac{1}{2}y^2} \sigma dy$$

$$= \frac{c_s \sigma}{s\sqrt{2\pi}},$$

i.e. $$c_s = \frac{s\sqrt{2\pi}}{\sigma} \int_0^1 \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} \tau_s e^{\frac{1}{2}\left(\frac{x-p/(p+q)}{\sigma}\right)^2} dx$$

$$= \frac{s\sqrt{2\pi}}{\sigma} \frac{1}{\sqrt{s}!\sqrt{2\pi}} \int_0^1 \left\{ \left(\frac{x-p/(p+q)}{\sigma}\right)^{s-1} - \frac{(s-1)(s-2)}{2.1!} \left(\frac{x-p/(p+q)}{\sigma}\right)^{s-3} \right.$$

$$\left. + \frac{(s-1)(s-2)(s-3)(s-4)}{2^2.2!} \left(\frac{x-p/(p+q)}{\sigma}\right)^{s-5} - \ldots \right\} \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} dx$$

$$= \frac{s}{\sigma\sqrt{s}!} \int_0^1 \left\{ \left(\frac{x-p/(p+q)}{\sigma}\right)^{s-1} - \frac{(s-1)(s-2)}{2.1!} \left(\frac{x-p/(p+q)}{\sigma}\right)^{s-3} \right.$$

$$\left. + \frac{(s-1)(s-2)(s-3)(s-4)}{2^2.2!} \left(\frac{x-p/(p+q)}{\sigma}\right)^{s-5} - \ldots \right\} \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} dx$$

.....(xxvi).

The integral for any particular value of s reduces to a series of B-functions and so c_s is found.

The area up to abscissa x is generally required:

i.e. $$\int_0^x \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} dx.$$

But $$\int_0^x \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} dx = \int_{-\infty}^{\xi} \overset{\infty}{S}_1 (c_s \tau_s) d\xi$$

$$= \sigma \int_{-\infty}^{y} \overset{\infty}{S}_1 (c_s \tau_s) dy.$$

Now $$\int_{-\infty}^{y} \tau_s dy = -\frac{1}{\sqrt{s}} \tau_{s-1},$$

$$\therefore \int_0^x \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} dx = \sigma \int_{-\infty}^{y} c_1 \tau_1 dy - \sigma \left\{ c_2 \frac{1}{\sqrt{2}} \tau_1 \right.$$

$$\left. + c_3 \frac{1}{\sqrt{3}} \tau_2 + c_4 \frac{1}{\sqrt{4}} \tau_3 + \ldots + c_s \frac{1}{\sqrt{s}} \tau_{s-1} + \ldots \right\}$$

$$= \sigma c_1 \int_{-\infty}^{y} \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}} dy - a_1\tau_1 - a_2\tau_2 - \ldots - a_s\tau_s - \ldots,$$

where $a_s = c_{s+1} \frac{\sigma}{\sqrt{s+1}}$, $= \sigma c_1 \frac{1}{2}(1+\alpha) - a_1\tau_1 - a_2\tau_2 - a_3\tau_3 - \ldots.$

If we put $s=1$, $s=2$, $s=3$ in the above formula (xxvi) for c_s,

$$c_1 = \frac{1}{\sigma},$$

$$c_2 = \frac{2}{\sigma\sqrt{2!}}\int_0^1 \left(\frac{x-p/(p+q)}{\sigma}\right)\frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx$$

$$= \frac{2}{\sigma^2\sqrt{2!}}\,\frac{1}{\mathrm{B}(p,q)}\{\mathrm{B}(p+1,q)-p/(p+q)\,\mathrm{B}(p,q)\}$$

$$= 0,$$

$$c_3 = \frac{3}{\sigma\sqrt{6}}\int_0^1 \left[\frac{1}{\sigma^2}\{x^2-2xp/(p+q)+p^2/(p+q)^2\}-\frac{2.1}{2}\right]\frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx$$

$$= \frac{3}{\sigma^3\sqrt{6}}\left[\frac{1}{\mathrm{B}(p,q)}\right]\left[\mathrm{B}(p+2,q)-2\frac{p}{p+q}\,\mathrm{B}(p+1,q)\right.$$
$$\left.+\frac{p^2}{(p+q)^2}\,\mathrm{B}(p,q)-\sigma^2\,\mathrm{B}(p,q)\right]$$

$$= \frac{3}{\sigma^3\sqrt{6}}\left\{\frac{p(p+1)}{(p+q)(p+q+1)}-2\frac{p^2}{(p+q)^2}+\frac{p^2}{(p+q)^2}-\frac{pq}{(p+q)^2(p+q+1)}\right\}$$

$$= 0,$$

as obtained before by the other method.

The terms in τ_1 and τ_2 do not exist, so that the expansion becomes:

$$\int_0^x \frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx = \tfrac{1}{2}(1+\alpha)-a_3\tau_3-a_4\tau_4-\ldots-a_s\tau_s-\ldots,$$

where
$$a_s = \frac{\sigma}{\sqrt{s+1}}\cdot c_{s+1}$$

$$= \frac{\sigma}{\sqrt{s+1}}\cdot\frac{(s+1)}{\sqrt{(s+1)!}}\,\sigma\int_0^1\left\{\left(\frac{x-p/(p+q)}{\sigma}\right)^s-\frac{s(s-1)}{2.1!}\left(\frac{x-p/(p+q)}{\sigma}\right)^{s-2}\right.$$
$$\left.+\frac{s(s-1)(s-2)(s-3)}{2^2.2!}\left(\frac{x-p/(p+q)}{\sigma}\right)^{s-4}-\ldots\right\}\frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx$$

$$= \frac{1}{\sqrt{s!}}\int_0^1\left\{\left(\frac{x-p/(p+q)}{\sigma}\right)^s-\frac{s(s-1)}{2.1!}\left(\frac{x-p/(p+q)}{\sigma}\right)^{s-2}\right.$$
$$\left.+\frac{s(s-1)(s-2)(s-3)}{2^2.2!}\left(\frac{x-p/(p+q)}{\sigma}\right)^{s-4}-\ldots\right\}\frac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)}\,dx$$

.....(xxvii).

The argument for the tetrachoric functions and for $\frac{1}{2}(1+\alpha)$ is $\dfrac{x-p/(p+q)}{\sigma}$. If this is negative then we must take $\frac{1}{2}(1-\alpha)$.

From the above expression (xxvii) the coefficients of the expansion can be determined both algebraically and numerically, but for the higher coefficients the algebraic work becomes exceedingly heavy. It is to be remembered that

$$\sigma^2 = \frac{pq}{(p+q)^2(p+q+1)}.$$

Suppose $(p+q)=m$; then $\sigma^2 = \dfrac{pq}{m^2(m+1)}$.

The coefficients a_1, a_2, ... etc. are given below:

$$a_1 = 0, \quad a_2 = 0, \quad a_3 = \frac{1}{\sqrt{3!}} 2! \sqrt{\frac{m+1}{pq}} \frac{(m-2p)}{(m+2)},$$

$$a_4 = \frac{1}{\sqrt{4!}} 3! \frac{1}{(m+2)(m+3)} \left\{ \frac{m^2(m+1)}{pq} - (5m+6) \right\},$$

$$a_5 = \frac{1}{\sqrt{5!}} 4! \sqrt{\frac{m+1}{pq}} \frac{(m-2p)}{(m+2)(m+3)(m+4)} \left\{ \frac{m^2(m+1)}{pq} - (7m+12) \right\},$$

$$a_6 = \frac{1}{\sqrt{6!}} 5! \frac{1}{(m+2)(m+3)(m+4)(m+5)} \left\{ \frac{m^4(m+1)^2}{p^2q^2} + \frac{m^2(m+1)}{3pq}(m^2-32m-60) - \tfrac{2}{3}(2m^3-41m^2-154m-120) \right\},$$

$$a_7 = \frac{1}{\sqrt{7!}} 6! \sqrt{\frac{m+1}{pq}} \frac{(m-2p)}{(m+2)(m+3)(m+4)(m+5)(m+6)} \times \left\{ \frac{m^4(m+1)^2}{p^2q^2} + \frac{m^2(m+1)}{12pq}(7m+15)(m-20) - \tfrac{5}{12}\{7m^3-59m^2-342m-360\} \right\},$$

$$a_8 = \frac{1}{\sqrt{8!}} 7! \frac{1}{(m+2)(m+3)(m+4)(m+5)(m+6)(m+7)} \times \left\{ \frac{m^6(m+1)^3}{p^3q^3} + \frac{m^4(m+1)^2}{60p^2q^2}\{47m^2-853m-2100\} + \frac{m^2(m+1)}{30pq}\{-251m^3+1503m^2+9974m+10920\} + \tfrac{1}{60}\{1271m^4-1697m^3-44512m^2-104364m-65520\} \right\}.$$

An additional coefficient a_9 was calculated for one of our examples, but it was not considered worth while working it out algebraically.

The coefficients in the tetrachoric expansion obtained by this latter method, that is, by using the property of tetrachoric functions as semi-orthogonal functions, are identical with those obtained from the first method, which consisted in equating moments of the functions on both sides of the equation. Thus we are led to the same expansion in both cases.

(6) The numerical results are certainly interesting but from the utility point of view they are not very satisfactory. Tables I—VIII contain these results in a convenient form; the values of the coefficients a_s and a_s', the tetrachoric functions, the successive terms ($-a_s\tau_s$ and $-a_s'\tau_s$) and the values of the series up to the term containing τ_s are given. It is to be noted that the coefficients do not appear in Tables II and IV but as these are the same as in Tables I and III respectively it

was not necessary to repeat them. In all these tables, in the row $s=0$ we have placed $\frac{1}{2}(1-\alpha)$ in the column containing the tetrachoric functions and it is only necessary to draw attention to the fact that in the next column the negative sign in $-a_s\tau_s$ does not apply to the first term $\frac{1}{2}(1-\alpha)$. The tables will then be easily

TABLE I.

$$\int_0^{29\cdot4} \frac{x^{48}e^{-x}}{\Gamma(49)}\,dx, \quad z=-2\cdot8\ *.$$

s	a_s	Tetrachoric Functions τ_s	Terms in Series $-a_s\tau_s$	Value of Series up to term τ_s
0	1·00000000	+ ·0025551	+ ·0025551	·0025551
1	0·00000000	+ ·00791545	—	—
2	0·00000000	− ·01567180	—	—
3	0·11664237	+ ·02210325	− ·0025782	− ·0000231
4	0·02499479	− ·02189644	+ ·0005473	·0005242
5	0·00638743	+ ·01259137	− ·0000804	·0004438
6	0·03228531	+ ·00159776	− ·0000516	·0003922
7	0·01785148	− ·01140536	+ ·0002036	·0005958
8	0·00840223	+ ·01000967	− ·0000841	·0005117
9	0·01470566	+ ·00006659	− ·0000010	·0005107
10	0·01282194	− ·00849985	+ ·0001090	·0006197
11	0·00895618	+ ·00711870	− ·0000638	·0005560
12	0·01042333	+ ·00164419	− ·0000171	·0005388
13	0·01079260	− ·00754632	+ ·0000814	·0006203
14	0·00962776	+ ·00418464	− ·0000403	·0005800
15	0·01015854	+ ·00374438	− ·0000380	·0005419
16	0·01102777	− ·00640271	+ ·0000706	·0006125
17	0·01128126	+ ·00094254	− ·0000106	·0006019
18	0·01209893	+ ·00523425	− ·0000633	·0005386
19	0·01345974	− ·00422873	+ ·0000569	·0005955
20	0·01483350	− ·00218561	+ ·0000324	·0006279
21	0·01660082	+ ·00525599	− ·0000873	·0005407
22	0·01901932	− ·00110396	+ ·0000210	·0005617
23	0·02196131	− ·00426227	+ ·0000936	·0006553
24	0·02561864	+ ·00346981	− ·0000889	·0005664
25	0·03033429	+ ·00205905	− ·0000625	·0005039
26	0·03631783	− ·00439701	+ ·0001597	·0006636
27	0·04391748	+ ·00042653	− ·0000187	·0006449
28	0·05371111	+ ·00393217	− ·0002112	·0004337
29	0·06638572	− ·00244866	+ ·0001626	·0005963
30	0·08285325	− ·00248099	+ ·0002056	·0008018

True value ·0005850.

understood, but, in order that a better appreciation of the results may be obtained, the value of the series up to a certain term has been plotted against the number of that term. A line, drawn across the paper and corresponding to the true value of the integral, shows how much the value of the series is in excess or defect of the true value of the integral. The various points have been joined by continuous wavy lines but, of course, these lines have no real physical meaning. However, by joining the points, the graph will, we think, convey a better idea of the variation

* $z=\dfrac{x-p}{\sqrt{p}}=\dfrac{29\cdot4-49}{7}=-2\cdot8.$

of the values of the series than a set of isolated points would. Figures 1—7 correspond to the data given in Tables I—VII.

Now in the case of the Incomplete Γ-function we obtained two expansions, with respect to the mean and the mode respectively, and the graphs tell us which of these two gives us the better approximation. Figs. 1 and 3 (Tables I and III) show the variations in the values of the series for $\int_0^{29\cdot4} \frac{x^{48}e^{-x}}{\Gamma(49)}dx$ from the mean and the mode respectively, while Figs. 2 and 4 give us similar information for $\int_0^{42} \frac{x^{48}e^{-x}}{\Gamma(49)}dx$.

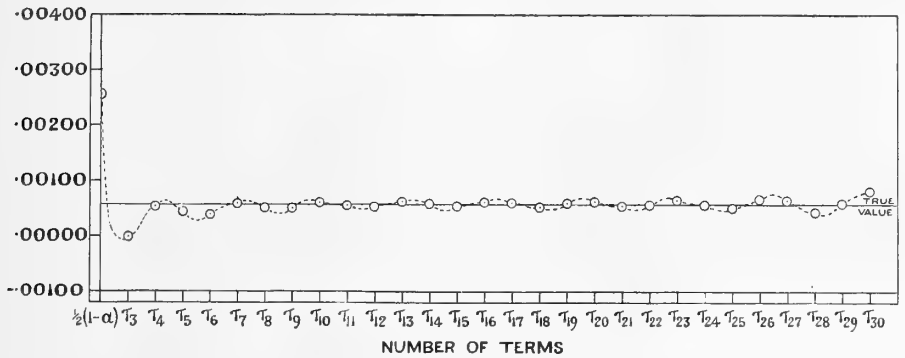


Fig. 1.

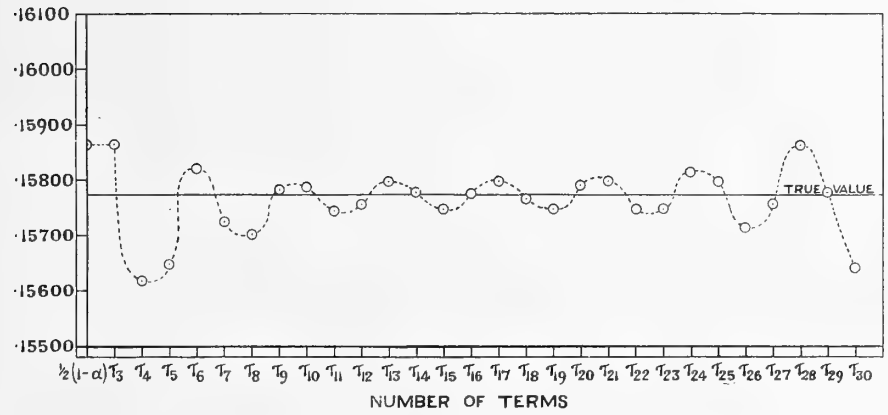


Fig. 2.

It will be seen that in Fig. 1 the points are much closer to the 'true value' line than in Fig. 3 (and similarly in Fig. 2 they are closer than in Fig. 4) so that the expansion from the mean seems to give a better approximation than that from the mode and it has the additional advantage that the terms in τ_1 and τ_2 are missing. Besides, it seems more natural to expand these normal curve functions in terms of the mean and standard deviation. For comparison purposes the graphs are all on the same scale. The graphs for the mode and the mean behave in a very

TABLE II. $\int_0^{42} \frac{x^{48}e^{-x}}{\Gamma(49)}\,dx, \quad z=-1$*.

s	Tetrachoric Functions τ_s	Terms in the Series $-a_s\tau_s$	Value of Series up to term τ_s
0	+ ·1586553	+ ·1586553	·1586553
1	+ ·24197074	·0000000	—
2	− ·17109916	·0000000	—
3	·00000000	·0000000	—
4	+ ·09878417	− ·0024691	·1561862
5	− ·04417762	+ ·0002822	·1564684
6	− ·05410632	+ ·0017468	·1582152
7	+ ·05453404	− ·0009735	·1572417
8	+ ·02410087	− ·0002025	·1570392
9	− ·05302190	+ ·0007797	·1578189
10	− ·00355664	+ ·0000456	·1578645
11	+ ·04657133	− ·0004172	·1574474
12	− ·01034833	+ ·0001079	·1575553
13	− ·03814548	+ ·0004117	·1579669
14	+ ·01939964	− ·1001868	·1577802
15	+ ·02921077	− ·0002967	·1574834
16	− ·02483411	+ ·0002739	·1577573
17	− ·02054429	+ ·0002318	·1579891
18	+ ·02755708	− ·0003334	·1576556
19	+ ·01256341	− ·0001691	·1574865
20	− ·02825493	+ ·0004191	·1579057
21	− ·00548187	+ ·0000910	·1579967
22	+ ·02745951	− ·0005223	·1574744
23	− ·00060803	+ ·0000134	·1574878
24	− ·02558848	+ ·0006555	·1581433
25	+ ·00568862	− ·0001726	·1579707
26	+ ·02297227	− ·0008343	·1571364
27	− ·00978859	+ ·0004299	·1575663
28	− ·01987296	+ ·0010674	·1586337
29	+ ·01296514	− ·0008607	·1577730
30	+ ·01649808	− ·0013669	·1564061

True value ·1577387.

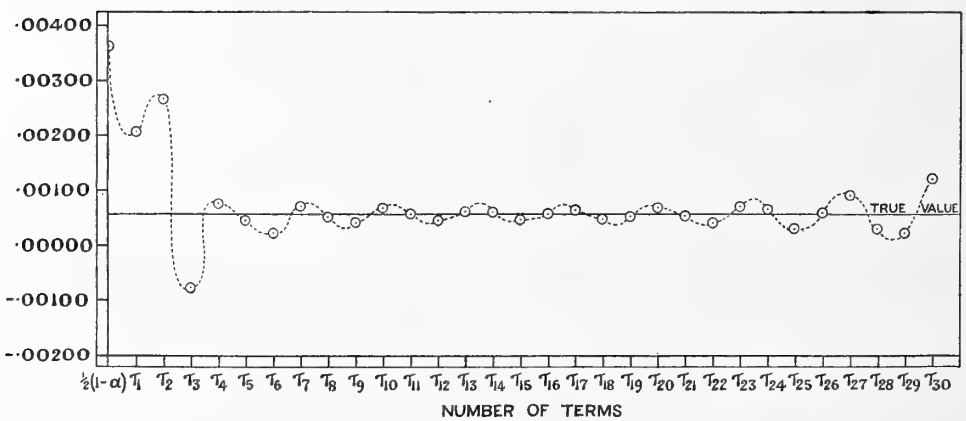


Fig. 3.

* $z=\frac{x-p}{\sqrt{p}}=\frac{42-49}{7}=-1.$

similar manner; for, if we regard the graphs as a wave, it will be noticed that at first the amplitude of the wave is big, decreases gradually up to a term in the neighbourhood of τ_{20} and thereafter increases more and more rapidly. This can be explained fairly easily; as s increases the tetrachoric functions τ_s do not increase or decrease steadily but vary in sign and remain of the same order of magnitude. The coefficients a_s vary in much the same way (except that they are all positive) up to a certain point and then begin to increase very fast. In equation (xv) we had

$$a_{s+1} = \sqrt{\frac{s}{p\,(s+1)}}\,\{\sqrt{s}\,a_s + \sqrt{(s-1)}\,a_{s-2}\},$$

i.e. a_{s+1} is of order $\sqrt{\frac{s}{p}}\,\{a_s + a_{s-2}\}$, so that as s increases there comes a time when $\sqrt{s}$ overcomes the reducing effect of $\frac{1}{\sqrt{p}}$ and then the coefficients will continually

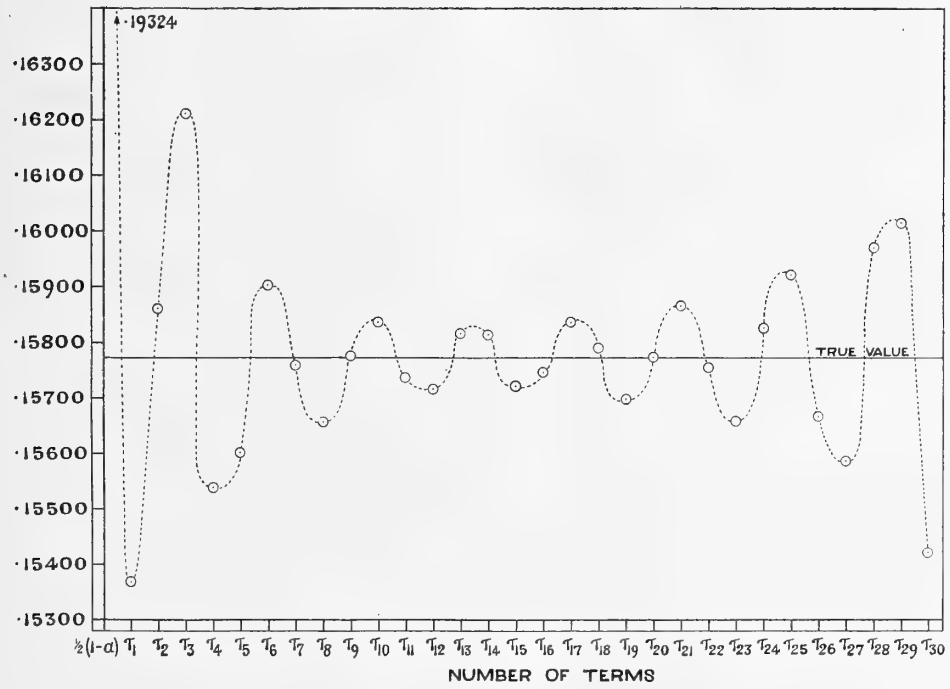


Fig. 4.

increase. For higher values of p this turning point will not be arrived at so soon and the points will hang closer to the 'true value' line for a greater number of terms, but it does not seem likely that the values of the series will tend to a definite limit. The equation for the modal expansion coefficients is a similar one and these coefficients behave in the same way.

Turning our attention to the expansions from the mean, Fig. 1 (and Fig. 3 to a less extent) would seem to suggest that the tetrachoric series gives quite a good approximation to the value of the integral. Although some of the points are very

TABLE III.

$$\int_0^{29\cdot4} \frac{x^{48}e^{-x}}{\Gamma(49)}\,dx, \quad z = -2\cdot6846788^{*}.$$

s	a_s'	Tetrachoric Functions τ_s	Terms in Series $-a_s'\tau_s$	Value of Series up to term τ_s
0	1·00000000	+ ·0036296	+ ·0036296	·0036296
1	0·14433757	+ ·01085979	− ·0015675	·0020621
2	0·02946278	− ·02061573	+ ·0006074	·0026695
3	0·12521683	+ ·02752089	− ·0034461	− ·0007766
4	0·06166251	− ·02503988	+ ·0015440	·0007675
5	0·02648957	+ ·01160193	− ·0003073	·0004601
6	0·04236301	+ ·00557065	− ·0002360	·0002241
7	0·03460284	− ·01460370	+ ·0005053	·0007295
8	0·02288715	+ ·00939504	− ·0002150	·0005144
9	0·02516285	+ ·00363988	− ·0000916	·0004229
10	0·02488684	− ·01101274	+ ·0002741	·0006969
11	0·02136289	+ ·00579095	− ·0001237	·0005732
12	0·02167770	+ ·00509737	− ·0001105	·0004627
13	0·02272772	− ·00713419	+ ·0001621	·0006249
14	0·02256727	+ ·00058475	− ·0000132	·0006117
15	0·02351439	+ ·00599464	− ·0001410	·0004707
16	0·02546065	− ·00455186	+ ·0001158	·0005865
17	0·02739094	− ·00248832	+ ·0000682	·0006547
18	0·02996700	+ ·00573797	− ·0001720	·0004827
19	0·03360181	− ·00124666	+ ·0000419	·0005246
20	0·03803862	− ·00454994	+ ·0001731	·0006977
21	0·04355963	+ ·00382135	− ·0001665	·0005312
22	0·05068120	+ ·00204641	− ·0001037	·0004275
23	0·05968962	− ·00471305	+ ·0002813	·0007089
24	0·07107603	+ ·00066657	− ·0000474	·0006615
25	0·08566837	+ ·00406751	− ·0003485	·0003130
26	0·10443752	− ·00276906	+ ·0002892	·0006022
27	0·12866399	− ·00240728	+ ·0003097	·0009119
28	0·16018283	+ ·00383980	− ·0006151	·0002969
29	0·20147298	+ ·00036667	− ·0000739	·0002230
30	0·25589543	− ·00382481	+ ·0009788	·0012017

True value ·0005850.

near to the 'true value' line, the approximation is not really a good one. The important question for us is: To how many decimal places does the series give the result correct? On going through the tables it will be found that there is no value of the series up to the sth term giving the result correct to more than three or four places. We now come to the real trouble. Suppose a frequency function is expanded in tetrachoric series, how are we to know at what term to stop so as to obtain the most accurate result? If the value of an integral is required, the true value is wanted. In our work we chose integrals of which the value was already known. From Figs. 1—4 it is easily seen that we have as good an approximation at the

* $z = \dfrac{x - p'}{\sqrt{p'}} = \dfrac{29\cdot4 - 48}{\sqrt{48}} = -2\cdot6846788.$

TABLE IV. $\int_0^{42} \frac{x^{48}e^{-x}}{\Gamma(49)}dx, \quad z=-\cdot8660254$*.

s	Tetrachoric Functions τ_s	Terms in Series $-a_s'\tau_s$	Value of Series up to term τ_s
0	·1932381	·1932381	·1932381
1	+ ·27418875	− ·0395757	·1536624
2	− ·16790564	+ ·0049470	·1586093
3	− ·02798427	+ ·0035041	·1621134
4	+ ·10905792	− ·0067248	·1553886
5	− ·02346554	+ ·0006216	·1560102
6	− ·07134833	+ ·0030225	·1590328
7	+ ·04145828	− ·0014346	·1575982
8	+ ·04451198	− ·0010188	·1565794
9	− ·04705083	+ ·0011839	·1577634
10	− ·02465039	+ ·0006135	·1583768
11	+ ·04681171	− ·0010000	·1573768
12	+ ·00975248	− ·0002114	·1571654
13	− ·04356976	+ ·0009902	·1581556
14	+ ·00140962	− ·0000318	·1581238
15	+ ·03877058	− ·0009117	·1572122
16	− ·00966795	+ ·0002462	·1574583
17	− ·03323150	+ ·0009102	·1583686
18	+ ·01562623	− ·0004683	·1579003
19	+ ·02744360	− ·0009222	·1569781
20	− ·01974338	+ ·0007510	·1577291
21	− ·02171195	+ ·0009458	·1586749
22	+ ·02237974	− ·0011342	·1575407
23	+ ·01622818	− ·0009687	·1565720
24	− ·02382475	+ ·0016934	·1582654
25	− ·01111122	+ ·0009519	·1592173
26	+ ·02431475	− ·0025394	·1566779
27	+ ·00643169	− ·0008275	·1558504
28	− ·02404492	+ ·0038516	·1597020
29	− ·00222729	+ ·0004487	·1601507
30	+ ·02317774	− ·0059311	·1542196

True value ·1577387.

TABLE V. $\int_0^{\cdot5} \frac{x^{14}(1-x)^4}{\mathrm{B}(15,\,5)}dx, \quad y=-2\cdot6457513, \quad p=15, \quad q=5, \quad m=20$†.

s	a_s	Tetrachoric Functions τ_s	Terms in Series $-a_s\tau_s$	Value of Series up to term τ_s
0	1·00000000	·0040751	·0040751	·0040751
3	− ·19638608	+ ·02950904	+ ·0057952	·0098703
4	+ ·01452267	− ·02602453	+ ·0003780	·0102482
5	+ ·03818545	+ ·01099737	− ·0004199	·0098283
6	+ ·05515045	+ ·00712711	− ·0003931	·0094352
7	− ·01389639	− ·01561177	− ·0002170	·0092183
8	− ·03609105	+ ·02031787	+ ·0007333	·0099516

True value ·0096054.

* $z=\frac{x-p'}{\sqrt{p'}}=\frac{42-48}{\sqrt{48}}=-\cdot8660254.$

† $y=\frac{x-\frac{p}{p+q}}{\sigma}=\frac{\cdot5-\cdot75}{\cdot09449112}=-2\cdot6457513.$

TABLE VI.

$$\int_0^{\cdot 5} \frac{x^3 (1-x)^{\frac{1}{2}}}{\mathrm{B}(4, \frac{3}{2})}\, dx, \quad y = -1{\cdot}3010412, \quad p = 4, \quad q = \tfrac{3}{2}, \quad m = 5\tfrac{1}{2}*.$$

s	a_s	Tetrachoric Functions τ_s	Terms in Series $-a_s\tau_s$	Value of Series up to terms τ_s
0	1·00000000	·0966212	·0966212	·0966212
3	− ·28327885	+ ·04839695	+ ·0137098	·1103310
4	− ·01400852	+ ·05941568	+ ·0008323	·1111633
5	+ ·16688842	− ·06703628	+ ·0111876	·1223509
6	+ ·05349154	− ·00778490	+ ·0004164	·1227673
7	− ·05325140	+ ·05554783	+ ·0029580	·1257253
8	− ·09445982	− ·01930950	− ·0018240	·1239013
9	− ·00063525	− ·03745046	− ·0000238	·1238775

True value ·1188790.

TABLE VII.

$$\int_0^{\cdot 1} \frac{x^3 (1-x)^{\frac{1}{2}}}{\mathrm{B}(4, \frac{3}{2})}\, dx, \quad y = -3{\cdot}59087385\dagger.$$

s	a_s	Tetrachoric Functions τ_s	Terms in Series $-a_s\tau_s$	Value of Series up to term τ_s
0	1·00000000	·0001648	·0001648	·0001648
3	− ·28327885	+ ·00307042	+ ·0008698	·0010346
4	− ·01400852	− ·00458580	− ·0000642	·0009704
5	+ ·16688842	+ ·00530458	− ·0008853	·0000851
6	+ ·05349154	− ·00442734	+ ·0002368	·0003219
7	− ·05325140	+ ·00191632	+ ·0001020	·0004239
8	− ·09445982	+ ·00111687	+ ·0001055	·0005294
9	− ·00063525	− ·00291774	− ·0000019	·0005275

True value ·00023603.

5th or 6th term as at the 15th, say, and better than at the 30th. Of course, one might calculate the various terms till the sums became more or less steady, take the mean of these sums after the steady stage is reached and use that as the value required. This process, however, will not give a greater accuracy than three or four decimal places correct and very likely the result will not be so good as that. Besides which it is difficult to give such an arbitrary weighting of terms a theoretical justification. Thus it seems that the tetrachoric series is not at all suitable for the representation of the Incomplete Γ-function.

* $y = \dfrac{x - \frac{p}{p+q}}{\sigma} = \dfrac{\cdot 5 - \frac{8}{11}}{\cdot 17468526} = -1{\cdot}3010412.$

† $y = \dfrac{x - \frac{p}{p+q}}{\sigma} = \dfrac{\cdot 1 - \frac{8}{11}}{\cdot 17468526} = -3{\cdot}59087385.$

When we consider the tables and graphs for the Incomplete B-function, the results are certainly no better than in the case of the Incomplete Γ-function. Unfortunately, owing to the lack of a difference formula connecting the successive coefficients, we only calculated a few terms, but the behaviour of the graphs is similar to that of the graphs of the Incomplete Γ-function. Fig. 5 is very like Figs. 1—4 but Figs. 6 and 7 are rather different. In Fig. 5 the integral is $\int_0^{\cdot 5} \frac{x^{14}(1-x)^4}{\mathrm{B}(15, 5)}\,dx$, where p is of high value and q is of moderate size. In Figs. 6 and 7 the integral is $\int_0 \frac{x^3(1-x)^{\frac{1}{2}}}{\mathrm{B}(4, \frac{3}{2})}\,dx$, where the upper limits are ·5 and ·1 respectively. Here p is 4 and q is $\frac{3}{2}$. It seems in the incomplete Γ- and B-functions that the points come nearer the 'true value' line for the tail of the integral than if the upper limit is near the mode.

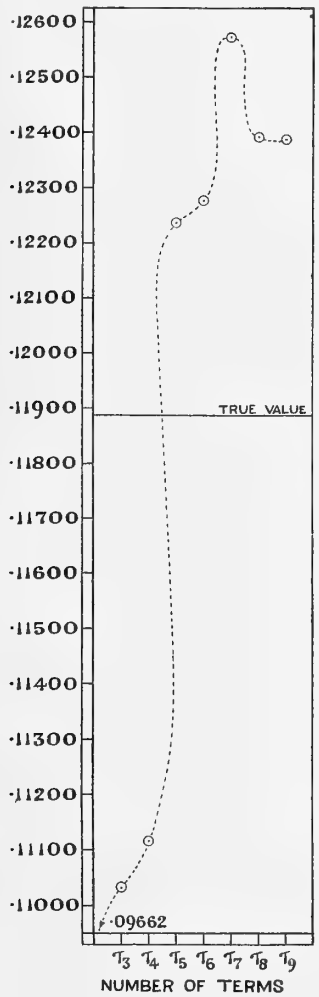


Fig. 6.

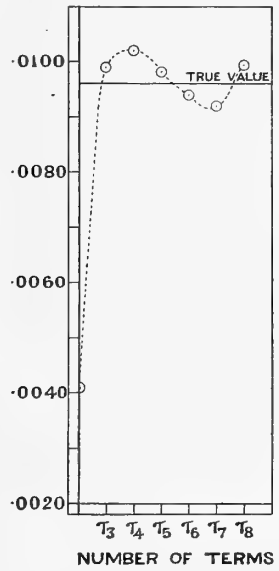


Fig. 5.

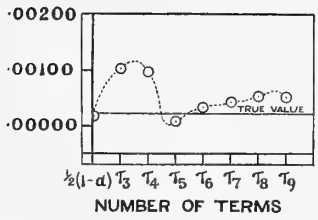


Fig. 7.

Table VIII gives the results for $\int_0^{49} \frac{x^{48}e^{-x}}{\Gamma(49)}dx$ and, since $z = \frac{x-p}{\sqrt{p}} = \frac{49-49}{7} = 0$ for the expansion from the mean, all the tetrachoric functions of even order vanish. It will be observed that the values of the series vary in a similar fashion to the others and not one of these gives the result correct to more than four decimal places.

TABLE VIII.

$$\int_0^{49} \frac{x^{48}e^{-x}}{\Gamma(49)}dx, \quad z = 0^*.$$

(Expansion with regard to the Mean.)

s	a_s	Tetrachoric Functions τ_s	Terms in Series $-a_s\tau_s$	Value of Series up to term τ_s
0	1·00000000	·5000000	·5000000	·5000000
3	·11664237	− ·1628675	+ ·0189973	·5189973
5	·00638743	+ ·1092549	− ·0006979	·5182994
7	·01785148	− ·0842920	+ ·0015047	·5198041
9	·01470566	+ ·0695373	− ·0010226	·5187815
11	·00895618	− ·0596711	+ ·0005345	·5193160
13	·01079260	+ ·0525526	− ·0005672	·5187488
15	·01015854	− ·0471442	+ ·0004789	·5192277

True value ·5189993.

After a careful study of the tables and graphs we are forced to the conclusion that a tetrachoric series is of no practical utility as a representation of skew frequency curves such as $y = y_0x^{p-1}e^{-x}$ and $y = y_0x^{m_1-1}(1-x)^{m_2-1}$, and although it may be rash to generalise from our results on these two types it would seem that such a series cannot be generally suitable to represent skew frequency distributions. Moreover, the types, which have been discussed, are of common occurrence and for these the expansion is certainly futile.

The true values of the incomplete Γ-function were taken from *Tables of the Incomplete Γ-function* which will be shortly issued by H.M. Stationery Office. The values of the incomplete B-function were determined by direct calculation; the power of $(1-x)$ was expanded and the result readily obtained with the help of the relation

$$\mathrm{B}(p, q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}.$$

In his *Vorlesungen über die Grundzüge der mathematischen Statistik* (Hamburg, 1920) Charlier, when dealing with skew frequency curves, gives as the general equation for the skew frequency curves of his Type A

$$\tfrac{1}{5}Y = \phi_0 + \beta_3\phi_0''' + \beta_4\phi_0^{\mathrm{iv}} + \beta_5\phi_0^{\mathrm{v}} + \ldots$$

* $z = \frac{x-p}{\sqrt{p}} = \frac{49-49}{7} = 0.$

where $\phi_0 = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}X^2}$ and ϕ_0''', ϕ_0^{iv}, ϕ_0^{v}, ... are the third, fourth, fifth, etc. differential of coefficients ϕ_0, i.e. Y is really expressed in a series of tetrachoric functions, or

$$Y = 5\{\tau_1 - \beta_3\sqrt{4!}\,\tau_4 + \beta_4\sqrt{5!}\,\tau_5 - \beta_5\sqrt{6!}\,\tau_6 - \ldots\}.$$

β_3, β_4, β_5, etc. along with M (the mean) and σ Charlier calls the 'characteristics' of the distribution curve. Now he seems to think that generally the coefficients β_3 and β_4* will only be required and so he has tabled $\phi_0(x)$, $\frac{d^3\phi_0}{dx^3}$, $\frac{d^4\phi_0}{dx^4}$ for $x = \cdot 00$ to 3·00 at intervals of ·01 and also for $x = 4$ (Tables III, IV and V on pp. 123—125) to four decimal places. With the series up to β_4 the theoretical Y-coordinate will be found, according to Charlier, but from our experience of tetrachoric functions we are exceedingly sceptical about the accuracy of such a result. In fact, we feel certain that the approximation will not be a good one. If the frequency curve be little different from the normal then possibly the approximation would not be very bad.

The above investigation was undertaken by me at the suggestion of Professor Pearson and I am indebted to him for several hints. My grateful thanks are due to Miss I. M^{c}Learn for her assistance in the preparation of the diagrams.

* Charlier defines the 'skewness' S to be $S = 3\beta_3$ and the 'excess' E to be $E = 3\beta_4$.

MISCELLANEA.

I. On the χ^2 test of Goodness of Fit.

By KARL PEARSON, F.R.S.

In a paper published in the *Philosophical Magazine* for July 1900, pp. 157—175, I dealt with the following problem: A very large population is sampled, say, the population $n_1, n_2, \ldots n_s, \ldots n_p$ with total N, and any individual sample is $m_1, m_2, \ldots m_s, \ldots m_p$, total M. The "probable constitution" is given by:

$$m_1' = \frac{M}{N} n_1, \quad m_2' = \frac{M}{N} n_2, \quad \ldots \; m_s' = \frac{M}{N} n_s, \quad \ldots \; m_p' = \frac{M}{N} n_p.$$

If a large number of samples of size M are taken, what is the distribution of variations from the "probable constitution" in these samples?

I showed that if the distribution of categories were such that no category contained a few isolated units, then the distribution depended on the calculation of $\chi^2 = S_1^p \frac{(m_s - m_s')^2}{m_s'}$, and provided a value for the probability P that samples would not diverge more than any given sample from the "probable constitution." This process is now familiar to statisticians as the χ^2, P test.

The sole limiting conditions were that the samples should be random, and each should be of the same size M.

In some cases the "probable constitution" (m' series) can be found at once because the distribution of the sampled population is known *a priori*. In other cases the values of the m' series have to be approximated to, and such approximations are the general rule in all discussions of probable error.

We say for example that the standard deviation of the mean of a sample taken from an indefinitely large population of size N and standard deviation σ is $\sigma/\sqrt{n}$, where n is the size of the sample.

We say that the standard deviation of second moment-coefficients of samples of size n is

$$\frac{\sqrt{\mu_4 - \mu_2^2}}{\sqrt{n}},$$

where $\mu_2 (= \sigma^2)$ and μ_4 are the second and fourth moment-coefficients of the population sampled. In fact every constant of the sample has a probable error determinable in terms of the constants of the sampled population. All these distributions of deviations from "probable constitution" are true for perfectly general but random samples of size n drawn from our indefinitely large population.

But unfortunately in a considerable number of cases that sampled population is unknown to us; we have no direct means of finding μ_2, μ_4, etc. What accordingly do we do? Why we replace the constants of the sampled population by those calculated from the sample itself, as the best information we have. And the justification of this proceeding is not far to seek. μ_s as found for the sample will only differ from the μ_s of the sampled population by terms of the order $1/\sqrt{n}$; for example if we are not dealing with *small* samples, and σ' be the standard deviation of the sample, σ' differs from σ by terms of the order $\sigma/\sqrt{2n}$ and accordingly the standard deviation of the mean is written $\sigma'/\sqrt{n}$ when it is really $\sigma/\sqrt{n}$. This method of treating probable errors is universal in the case of fair sized samples to-day and scarcely needs justification. In writing the

sample values of the constants for those of the sampled population, we do not in any way alter our original supposition that we are considering the distribution of random samples of size n. We have still $p-1$ degrees of freedom, if we have p categories of frequency.

The process of substituting sample constants for sampled population constants does *not* mean that we select out of possible samples of size n, those which have precisely the same values of the constants as the individual sample under discussion. Clearly the given sample has definite moment-coefficients, and if there be p frequency categories the first $p-1$ moment-coefficients together with the size n of the sample would suffice to fix all the frequencies of the p categories*. Hence no deviations from the "probable constitution" would be possible if we confined our attention to samples of n tied to the constants of the given sample! In using the constants of the given sample to replace the constants of the sampled population, we in no wise restrict the original hypothesis of free random samples tied down only by their definite size. We certainly do not by using sample constants reduce in any way the random sampling degrees of freedom.

What we actually do is to replace the accurate value of χ^2, which is unknown to us, and cannot be found, by an approximate value, and we do this with precisely the same justification as the astronomer claims, when he calculates his probable error on his observations, and not on the mean square error of an infinite population of errors which is unknown to him. The whole of this matter was very fully discussed (pp. 164—7) in my original paper dealing with the χ^2, P test.

The above re-description of what seem to me very elementary considerations would be unnecessary had not a recent writer in the *Journal of the Royal Statistical Society*† appeared to have wholly ignored them. He considers that I have made serious blunders in not limiting my degrees of freedom by the number of moments I have taken; for example he asserts (p. 93) that if a frequency curve be fitted by the use of four moments then the n' of the tables of goodness of fit should be reduced by 4. I hold that such a view is entirely erroneous, and that the writer has done no service to the science of statistics by giving it broad-cast circulation in the pages of the *Journal of the Royal Statistical Society.*

What he would obtain if he placed this restriction on his samples is not the χ^2 for the distribution of samples of size n, but of samples which give definite moments. The absurdity of this manner of approach is at once obvious, if as I have suggested, we consider the p first-moments, as there is no reason why we should not do,—for these are just as much "fixed" as the first four—and the conclusion must be that we can learn nothing at all about variation from our sample; for we have p frequency groups and p-tying conditions.

When we wish to find the probable error of a mean or a standard deviation, we do *not* start by fixing down these characters to their values in the individual sample; we suppose them to take all the possible values they could take by sampling, and after we have reached our measure of variation we then put into our formula the sampled values, to give an approximate value to the functions reached, because we are in ignorance of the real values in the sampled population.

The writer in the *Journal of the Royal Statistical Society* speaks as if I applied χ^2 to a contingency table *starting* by fixing the marginal totals. As far as I am aware I am not guilty of this. My conception of contingency is very different from my conception of χ^2. I started my conception of contingency with the idea not of a random sample, but with the idea that some function of frequencies alone without regard to their relation to the measured characters would lead to the value of the correlation. Naturally I started from the deviation of the individual cell contents from the same cell contents on the basis of independent probability, as determined by the marginal totals. There was no question of sampling in the matter. In now fairly usual notation I termed

$$m_{ss'} - \frac{m_{s\bullet} m_{\bullet s'}}{M}$$

* This is Thiele's method of representing frequency distributions.

† Vol. LXXXV. p. 87, 1922.

the cell contingency and after playing about with such cell contingencies for a time succeeded in finding a function ϕ^2 of them which for indefinitely fine grouping for a bi-variate normal frequency distribution gave the correlation r as :

$$r=\sqrt{\frac{\phi^2}{1+\phi^2}},$$

where

$$\phi^2=\frac{1}{M}S\frac{\left(m_{ss'}-\frac{m_{s\bullet}m_{\bullet s'}}{M}\right)^2}{\frac{m_{s\bullet}m_{\bullet s'}}{M}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\alpha).$$

I see no reason for confusing this ϕ^2 as a measure of correlation with the χ^2 which is a measure of variability in the samples of constant size drawn from an indefinitely large population. It was different in its origin, as far as I am concerned, and different in its use. It is only when we come to consider the probable error of ϕ^2 that we have to distinguish between (a) the actual marginal totals of the sample and (b) the probable constitution of the marginal totals as deduced from an indefinitely large sampled population.

There are, as those who have read *Biometrika** will recognise, considerable difficulties about determining the probable error of ϕ^2, where

$$1+\phi^2=S\left(\frac{m^2_{ss'}}{m_{s\bullet}m_{\bullet s'}}\right),$$

and the determination of the mean ϕ^2 and of the standard deviation of ϕ^2 involves very troublesome analysis.

So laborious is the arithmetic involved that for ordinary statistical use it became doubtful whether it would not be better to define ϕ^2 as the mean squared contingency measured not from the marginal totals of the sample, but from the "probable constitution" of the marginal totals of the sample as deduced from the sampled population. In this case if

$$m'_{ss'}=\frac{M}{N}n_{ss'}, \qquad m'_{s\bullet}=\frac{M}{N}n_{s\bullet}, \qquad m'_{\bullet s'}=\frac{M}{N}n_{\bullet s'},$$

$$\phi^2=S\frac{\left(m_{ss'}-\frac{m'_{s\bullet}m'_{\bullet s'}}{M}\right)^2}{M.\frac{m'_{s\bullet}m'_{s'}}{M}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\beta)$$

or,

$$1+\phi^2=S\left(\frac{m^2_{ss'}}{m'_{s\bullet}m'_{\bullet s'}}\right);$$

with this change of definition the probable error and mean of ϕ^2 are more easily obtainable, and in this case *for the first time*, $M\phi^2$ can be looked upon as equivalent to a χ^2.

The form (a) from my standpoint cannot be treated as a χ^2, because it is not the deviation-measure of a given sample from the sampled population. Nor again is (β) the deviation-measure of the sample from the sampled population, unless we assume that population to have zero contingency, i.e. $m'_{ss'}=m'_{s\bullet}m'_{\bullet s'}/M$.

But χ^2 may in the form (β) be treated as a deviation-measure of the actual sample from an artificial sampled population, which differs from the actual population in having no correlation or contingency, but having the same marginal distributions of the two characters.

The moment, however, we assume form (β) for our contingency we are giving, what we clearly must give, absolute freedom to the marginal totals of our samples. The sole limit on our sample is its total size M. But when we come to actually calculating ϕ^2 for the individual sample, or the mean value or the standard deviation (i.e. probable error) of ϕ^2 for a series of samples, we have only one course open to us, if we do not know the constants of the sampled population, we must insert the marginal totals of the individual sample of which we have cognizance in place of the

* Vol. v. p. 191, Vol. x. p. 570, Vol. xi. p. 570, and Vol. xii. p. 259.

unknown values of the sampled population. Thus (α) and (β) provide ultimately the same ϕ^2, but the probable error of ϕ^2 and the mean value of ϕ^2 will be different in the two cases. In the first case we vary our marginal totals with the sample as they obviously would vary in practice. In the second case we define our ϕ^2 to be a deviation from the independent probability of an artificial population, we do not keep the marginal totals of the sample fixed any more than in (α). But if we think in terms of χ^2 (and not ϕ^2) we appear to do so because ultimately we have to take our marginal probabilities as those of the sample in default of a knowledge of any better values.

This point seems to me well illustrated in what my critic in the *Journal of the Royal Statistical Society* has to say on p. 90 of his paper about Messrs Greenwood and Yule's use of χ^2 for a fourfold table. He asserts that they ought to have entered the table of goodness of fit with $n'=2$. The problem before them was whether their fourfold tables could possibly be samples of bi-variate independent probability distributions. Each sample from such a distribution would have perfectly free cell frequencies m_{11}, m_{12}, m_{21}, m_{22}, subject to the sole binding condition that

$$m_{11}+m_{12}+m_{21}+m_{22}=M.$$

The proper χ^2 is given by

$$\chi^2=\frac{\left(m_{11}-\frac{m'_{1\cdot}m'_{\cdot 1}}{M}\right)^2}{\frac{m'_{1\cdot}m'_{\cdot 1}}{M}}+\frac{\left(m_{12}-\frac{m'_{1\cdot}m'_{\cdot 2}}{M}\right)^2}{\frac{m'_{1\cdot}m'_{\cdot 2}}{M}}+\frac{\left(m_{21}-\frac{m'_{2\cdot}m'_{\cdot 1}}{M}\right)^2}{\frac{m'_{2\cdot}m'_{\cdot 1}}{M}}+\frac{\left(m_{22}-\frac{m'_{2\cdot}m'_{\cdot 2}}{M}\right)^2}{\frac{m'_{2\cdot}m'_{\cdot 2}}{M}} \quad ...(\gamma),$$

and this has three degrees of freedom and is what Messrs Yule and Greenwood desired to find, and they properly used the value of P for $n'=4$.

Then like the astronomer, who finding the probable error of his mean to be $\cdot 67449\sigma/\sqrt{M}$ and not knowing the σ of his sampled population, puts it equal to the σ of his observations, so Messrs Yule and Greenwood very properly replaced the marginal totals of their unknown population by those of their sample, but very properly did not replace $n'=4$ by $n'=2$!.

But says my critic*, if they had, they would have got the same measure of improbability as if they had compared the difference of percentages ! Quite so, and obviously so ; for in taking percentages they have actually fixed their marginal totals taking 100 of each class and thus for the first time confined their attention to a limited class of samples, not the random sample of size M, which has not its marginal totals fixed. We have, indeed, reduced our degrees of freedom by two in taking ratios.

When we consider generally the χ^2 for a fourfold table to measure the improbability of a sample we are really comparing the special sample

a	b	$a+b$	with	a'	b'	$a'+b'$
c	d	$c+d$		c'	d'	$c'+d'$
$a+c$	$b+d$	M		$a'+c'$	$b'+d'$	M

the general population, where in the latter case $a'd'=c'b'$.

Now the mean square contingency of the first of these tables is

$$\phi^2=\frac{1}{M}\left\{\frac{\left(a-\frac{(a+b)(a+c)}{M}\right)^2}{\frac{(a+b)(a+c)}{M}}+\frac{\left(b-\frac{(a+b)(b+d)}{M}\right)^2}{\frac{(a+b)(b+d)}{M}}+\frac{\left(c-\frac{(a+c)(c+d)}{M}\right)^2}{\frac{(a+c)(c+d)}{M}}+\frac{\left(d-\frac{(c+d)(b+d)}{M}\right)^2}{\frac{(c+d)(b+d)}{M}}\right\}$$

$$=\left\{\frac{a^2}{(a+b)(a+c)}+\frac{b^2}{(a+b)(b+d)}+\frac{c^2}{(a+c)(c+d)}+\frac{d^2}{(c+d)(b+d)}-1\right\}$$

$$=\frac{(ab-cd)^2}{(a+b)(a+c)(b+d)(c+d)}.$$

* *Loc. cit.* p. 90.

But the χ^2 is

$$= \frac{\left(a - \frac{(a'+b')(a'+c')}{M}\right)^2}{\frac{(a'+b')(a'+c')}{M}} + \frac{\left(b - \frac{(a'+b')(b'+d')}{M}\right)^2}{\frac{(a'+b')(b'+d')}{M}} + \frac{\left(c - \frac{(a'+c')(c'+d')}{M}\right)^2}{\frac{(a'+c')(c'+d')}{M}} + \frac{\left(d - \frac{(c'+d')(b'+d')}{M}\right)^2}{\frac{(c'+d')(b'+d')}{M}}$$

$$= M\left\{\frac{a^2}{(a'+b')(a'+c')} + \frac{b^2}{(a'+b')(b'+d')} + \frac{c^2}{(a'+c')(c'+d')} + \frac{d^2}{(c'+d')(b'+d')} - 1\right\},$$

there being *three* degrees of freedom or we must take $n' = 4$ in calculating the probability P, this may be written

$$\chi^2 = \frac{1}{M}\left\{\frac{a^2}{p'_{\cdot 1}p'_{1\cdot}} + \frac{b^2}{p'_{\cdot 1}p'_{2\cdot}} + \frac{c^2}{p'_{\cdot 2}p'_{1\cdot}} + \frac{d^2}{p'_{\cdot 2}p'_{2\cdot}} - 1\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots(\delta),$$

where $p'_{\cdot 1}$, $p'_{\cdot 2}$, $p'_{1\cdot}$ and $p'_{2\cdot}$ are the four percentage numbers of the marginal categories in the sampled population. Now we do not know these percentages in that population and we do what every physicist, every astronomer, and—till I saw the paper by my critic in the *Journal of the Statistical Society* I should have said—every statistician does, supply the unknown constants from the sample, which leads us to

$$\chi^2 = \frac{M(ab-cd)^2}{(a+b)(a+c)(b+d)(c+d)} = M\phi^2$$

as used in my memoir of 1912*.

The problem I had and still have in view is the variability in samples of definite size—with no other restriction than sample size. The solution of that problem is absolutely comparable with that of any discussion of the probability of an observed result in the theory of probable errors. We have in the bulk of such cases constants involved which concern the distribution in an unknown population, and we supply those constants from the sample itself.

As I have already noted the probable error of a mean is

$$\frac{\cdot 67449\sqrt{\mu_2' - \mu_1'^2}}{\sqrt{M}}.$$

By this we understand that the means of samples restricted solely by their size M from an indefinitely large population of moment-coefficients μ_1', μ_2' about a fixed origin will have a variability determined by the above formula. But when we proceed to give both μ_1' and μ_2' the values determined from the sample we know, we do *not* add in the manner of my Royal Statistical Society critic, "but in doing so the type of samples is reduced to those having the mean and standard deviation of the sample." If we did, this selection of samples would clearly have no variation of mean or standard deviation at all! In fact probable errors would be meaningless, unless we drew our samples from a population already fully known to us, in which case we should not in 99 °/₀ of cases want to sample it at all.

In the same way when we use the marginal totals of the sample in formulae like (δ) we do not thereby reduce our samples to those having constant marginal totals, we merely take the best approximation available to the proper value of χ^2, and the fact that χ^2, as found from the sample, is only an approximation to the true χ^2 was fully recognised and discussed in my original memoir in the *Philosophical Magazine*.

It only remains to say that the following sentence of my critic's paper seems to me based upon a fallacious principle and apparently flows from a disregard of the nature of probable errors in general.

"It should be pointed out that certain of Pearson's *Tables for Statisticians and Biometricians*, namely Tables XVII, XIX and XX, together with XXII (*Abac* to determine r_P') are all calculated

* On a novel method of regarding the association of two variates classed solely in alternative categories. *Drapers' Company Research Memoirs*, Cambridge University Press.

on the assumption that $n'=4$ in fourfold tables, and consequently should not be used when, as is almost always the case, the marginal totals are obtained from the data" (*loc. cit.* p. 91).

I hold those tables are quite correctly calculated for $n'=4$, and those who attempt to modify them by assuming $n'=2$ will be dealing with an entirely different problem. Namely, they will be considering not the improbability of the given sample as one of all possible samples of the given size, which it really is, but one of the indefinitely smaller number of samples that have fixed marginal totals. We do not find the probable error of r for a tetrachoric table* on the assumption that the marginal totals are fixed. We find it on the assumption that the marginal totals also vary from sample to sample, and when we have found it, then we substitute in the result the values of not only the marginal totals, but the cell-contents, a, b, c, d of the sample itself for those of the unknown population. With χ^2 we go through an exactly similar process of reasoning. If by this procedure we in some mysterious manner tied our degrees of freedom down to the values of the cell-contents used in our formula and adopted from our sample there could be no probable error for r, for the values of a, b, c, and d are all required and used. I trust my critic will pardon me for comparing him with Don Quixote tilting at the windmill; he must either destroy himself, or the whole theory of probable errors, for they are invariably based on using sample values for those of the sampled population unknown to us. For example here is an argument for Don Quixote of the simplest nature: In the sth category of a population N the frequency is n_s, a sample shows m_s in a total M. The standard deviation of this frequency is

$$\sqrt{M \frac{n_s}{N}\left(1-\frac{n_s}{N}\right)}.$$

But we don't know the population sampled and accordingly obtain an approximate value of the above standard deviation by writing for $\frac{n_s}{N}$, $\frac{m_s}{M}$ and taking for the standard deviation of m_s $\sqrt{m_s\left(1-\frac{m_s}{M}\right)}$. In doing this it is not a question even of using a marginal total, we have used a cell frequency found from our sample. We have therefore according to our critic reduced our possibilities of freedom by selecting out of all possible samples those with m_s in the sth cell—this is exactly parallel to our reducing our freedom by "fixing" marginal proportions or moment-coefficients. But if m_s be fixed, it is ridiculous to talk of a variation of the m_s frequency. Therefore either $m_s=0$ or $m_s=M$, or the usual theory and practice of probable errors are wholly at fault. I think this will illustrate what I mean by Don Quixote and the windmill.

II.

Is Tuberculosis to be regarded from the Aetiological Standpoint as an acute disease of Childhood? By Dr Kr. F. Andvord (Christiania). *Tubercle*, Vol. III. No. 3, December, 1921.

This paper is, we must confess, unconvincing. The author holds that in a community that has long been subject to tuberculosis the time of infection should be fixed in the infantile years for the great majority of cases and consequently we should protect children for the first three or four years from infection.

As evidence of his views he takes a graph of what he calls a "population frame" which is really the well-known "number living in a stationary population" (l_x) and represents within this graph the numbers dying from tuberculosis and the numbers who have suffered from it at each age. We are doubtful if his graphs for deaths are correctly drawn. They are made to rise suddenly for about a year and then fall till age 7 but we suspect that they should fall from birth till age 7. We cannot justify his chart (No. VIII) which gives the whole population and the

* *Phil. Trans.* Vol. 195 A, p. 14.

tubercular population. The non-tubercular found by this chart actually increase after age 17 for many years so that the non-tubercular not only have no mortality but are increased by some process of resurrection! Admittedly the chart is hypothetical but as it stands it calls for amendment.

Dr Andvord's remark that "one would hardly gather from these per-thousand curves," i.e. from rates of mortality for various ages, "that, as is really the case, more persons die from tuberculosis in the first and second years of life than in any subsequent age period" seems to betray an inexperience in matters related to a life table: this weakness is shown elsewhere, e.g. p. 102, where deaths are stated without populations and without reference to age distributions.

Dr Andvord may have other evidence in support of his views but the article under review does not justify them statistically; we think every point he brings out could be explained as well on other hypotheses. He cannot, moreover, completely prove his case till he has studied communities which become subject to infection after having been kept free from it. For if his theory be correct, the measures he proposes would necessarily produce such a community.

W. Palin Elderton.

The Cambridge University Press, Fetter Lane, London, E.C. 4, and their Agents, are now the sole agents for the sale of the following publications of the Galton and Biometric Laboratories, University of London:

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 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. A Co-operative Study. *[In preparation.*

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.*

X. A Monograph on the Long Bones of the English Skeleton. By KARL PEARSON, F.R.S., and JULIA BELL, M.A. Part I. *The Femur [of Man].* Text I and Atlas of Plates I. *Price 30s. net.*

XI. A Monograph on the Long Bones of the English Skeleton. By KARL PEARSON, F.R.S., and JULIA BELL, M.A. Part I, Section II. *The Femur [with Special Reference to other Primate Femora.]* Text II and Atlas of Plates II. *Price 40s. 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 6s. net.* Sold only with complete sets.

II. A First Study of the Statistics of Pulmonary Tuberculosis (Inheritance). By KARL PEARSON, F.R.S. *Price 6s. net.* Sold only with complete sets.

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. *[In preparation.*

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.*

Technical Series.

I. On a Theory of the Stresses in Crane and Coupling Hooks with Experimental Comparison with Existing Theory. By E. S. ANDREWS, B.Sc. Eng., assisted by KARL PEARSON, F.R.S. *Issued. Price 3s. net.*

II. On some Disregarded Points in the Stability of Masonry Dams. By L. W. ATCHERLEY, assisted by KARL PEARSON, F.R.S. *Issued. Price 7s. net.* Sold only with complete sets.

III. On the Graphics of Metal Arches with special reference to the Relative Strength of Two-pivoted, Three-pivoted and Built-in Metal Arches. By L. W. ATCHERLEY and KARL PEARSON, F.R.S. *Issued. Price 5s. net.*

IV. On Torsional Vibrations in Axles and Shafting. By KARL PEARSON, F.R.S. *Issued. Price 6s. net.*

V. An Experimental Study of the Stresses in Masonry Dams. By KARL PEARSON, F.R.S., and A. F. CAMPBELL POLLARD, assisted by C. W. WHEEN, B.Sc. Eng., and L. F. RICHARDSON, B.A. *Issued. Price 7s. net.*

VI. On a Practical Theory of Elliptic and Pseudo-elliptic Arches, with special reference to the ideal Masonry Arch. By KARL PEARSON, F.R.S., W. D. REYNOLDS, B.Sc. Eng., and W. F. STANTON, B.Sc. Eng. *Issued. Price 4s. net.*

VII. On the Torsion resulting from Flexure in Prisms with Cross-sections of Uni-axial Symmetry only. By A. W. YOUNG, ETHEL M. ELDERTON and KARL PEARSON, F.R.S. *Issued. Price 7s 6d. net*

Drapers' Company Research Memoirs—(*cont.*).

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. [*Out of print.*

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, Ph.D. *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.*

X. **The Science of Man. Its Needs and its** Prospects. By KARL PEARSON, F.R.S. Being the Presidential Address to Section H of the British Association, 1920. *Price* 1*s.* 6*d. net.*

Eugenics Laboratory Publications

MEMOIR SERIES.

I. **The Inheritance of Ability.** By EDGAR SCHUSTER, D.Sc., Formerly Galton Research Fellow, and ETHEL M. ELDERTON, Galton Scholar. *Price* 4*s. net.* Sold only with complete sets.

II. **A First Study of the Statistics of** Insanity and the Inheritance of the Insane Diathesis. By DAVID HERON, D.Sc., Formerly Galton Research Fellow. *Price* 3*s. net.*

III. **The Promise of Youth and the** Performance of Manhood. By EDGAR SCHUSTER, D.Sc., Formerly Galton Research Fellow. *Price* 2*s.* 6*d. net.*

IV. **On the Measure of the Resemblance** of First Cousins. By ETHEL M. ELDERTON, Galton Research Fellow, assisted by KARL PEARSON, F.R.S. *Price* 3*s.* 6*d. net.*

V. **A First Study of the Inheritance of** Vision and of the Relative Influence of Heredity and Environment on Sight. By AMY BARRINGTON and KARL PEARSON, F.R.S. *Price* 4*s. net.*

VI. **Treasury of Human Inheritance** (Pedigrees of physical, psychical, and pathological Characters in Man). Parts I and II (double part). (Diabetes insipidus, Split-Foot, Polydactylism, Brachydactylism, Tuberculosis, Deaf-Mutism, and Legal Ability.) *Price* 14*s. net.*

VII. **On the Relationship of Condition of** the Teeth in Children to Factors of Health and Home Environment. By E. C. RHODES, B.A. *Price* 9*s. net.*

VIII. **The Influence of Unfavourable Home** Environment and Defective Physique on the Intelligence of School Children. By DAVID HERON, M.A., D.Sc., Formerly Galton Research Fellow. [*Out of print.*

IX. **The Treasury of Human Inheritance** (Pedigrees of physical, psychical, and pathological Characters in Man). Part III. (Angioneurotic Oedema, Hermaphroditism, Deaf-Mutism, Insanity, Commercial Ability.) *Price* 6*s. net.*

X. **The Influence of Parental Alcoholism** on the Physique and Intelligence of the Offspring. By ETHEL M. ELDERTON, assisted by KARL PEARSON. *Second Edition. Price* 4*s. net.*

XI. **The Treasury of Human Inheritance** (Pedigrees of physical, psychical, and pathological Characters in Man). Part IV. (Cleft Palate, Hare-Lip, Deaf-Mutism, and Congenital Cataract.) *Price* 10*s. net.*

XII. **The Treasury of Human Inheritance** (Pedigrees of physical, psychical, and pathological Characters in Man). Parts V and VI. (Haemophilia.) *Price* 15*s. net.*

XIII. **A Second Study of the Influence of** Parental Alcoholism on the Physique and Intelligence of the Offspring. By KARL PEARSON, F.R.S., and ETHEL M. ELDERTON. *Price* 4*s. net.*

XIV. **A Preliminary Study of Extreme** Alcoholism in Adults. By AMY BARRINGTON and KARL PEARSON, F.R.S., assisted by DAVID HERON, D.Sc. *Price* 4*s. net.*

XV. **The Treasury of Human Inheritance.** Dwarfism, with 49 Plates of Illustrations and 8 Plates of Pedigrees. *Price* 15*s. net.*

XVI. **The Treasury of Human Inheritance.** Prefatory matter and indices to Vol. I. With Frontispiece Portraits of Sir Francis Galton and Ancestry. *Price* 3*s. net.*

XVII. **A Second Study of Extreme Alco-** holism in Adults. With special reference to the Home-Office Inebriate Reformatory data. By DAVID HERON, D.Sc. *Price* 5*s. net.*

XVIII. **On the Correlation of Fertility with** Social Value. A Cooperative Study. *Price* 6*s. net.*

XIX—XX. **Report on the English Birthrate.** Part I. England, North of the Humber. By ETHEL M. ELDERTON, Galton Research Fellow. *Price* 9*s. net.*

XXI. **The Treasury of Human Inheritance.** Vol. II (Nettleship Memorial Volume). Anomalies and Diseases of the Eye. Part I. [*Just ready.*

Eugenics Laboratory Publications—(*cont.*).

Vol. I of *The Treasury of Human Inheritance* (VI+IX+XI+XII+XV+XVI of the above memoirs) may now be obtained bound in buckram, price 57*s.* 6*d. net.* Buckram cases for binding can be purchased at 3*s.* 9*d.* with impress of the bust of Sir Francis Galton. An engraved portrait of Sir Francis Galton can be obtained by sending a postal order for 3*s.* 6*d.* to the Secretary to the Laboratory, University College, London, W.C.

LECTURE SERIES. *Price 1s. net each* (*Nos. III, X, XI, XII excepted*).

I. **The Scope and Importance to the** State of the Science of National Eugenics. By KARL PEARSON, F.R.S. Third Edition.

II. **The Groundwork of Eugenics.** By KARL PEARSON, F.R.S. Second Edition.

III. **The Relative Strength of Nurture and** Nature. Much enlarged Second Edition. Part I. The Relative Strength of Nurture and Nature. (Second Edition revised.) By ETHEL M. ELDERTON. Part II. Some Recent Misinterpretations of the Problem of Nurture and Nature. (First Issue.) By KARL PEARSON, F.R.S. *Price* 2*s. net.*

IV. **On the Marriage of First Cousins.** By ETHEL M. ELDERTON.

V. **The Problem of Practical Eugenics.** By KARL PEARSON, F.R.S. Second Edition.

VI. **Nature and Nurture, the Problem of** the Future. By KARL PEARSON, F.R.S. Second Edition.

VII. **The Academic Aspect of the Science** of National Eugenics. By KARL PEARSON, F.R.S.

VIII. **Tuberculosis, Heredity and Environment.** By KARL PEARSON, F.R.S.

IX. **Darwinism, Medical Progress and Eugenics.** The Cavendish Lecture, 1912. By KARL PEARSON, F.R.S.

X. **The Handicapping of the First-born.** By KARL PEARSON, F.R.S. *Price* 2*s. net.*

XI. **National Life from the Standpoint of** Science. (Third Issue.) By KARL PEARSON, F.R.S. *Price* 1*s.* 6*d. net.*

XII. **The Function of Science in the Modern** State. (New Issue.) By KARL PEARSON, F.R.S. *Price* 2*s. net.*

XIII. **Sidelights on the Evolution of Man.** By KARL PEARSON, F.R.S. *Price* 3*s. net.*

The Chances of Death and other Studies in Evolution

BY KARL PEARSON, F.R.S.

GALTON PROFESSOR, UNIVERSITY OF LONDON

VOL. I

1. The Chances of Death.
2. The Scientific Aspect of Monte Carlo Roulette.
3. Reproductive Selection.
4. Socialism and Natural Selection.
5. Politics and Science.
6. Reaction.
7. Woman and Labour.
8. Variation in Man and Woman.

VOL. II

9. Woman as Witch. Evidences of Mother-Right in the Customs of Mediaeval Witchcraft.
10. Ashiepattle, or Hans seeks his Luck.
11. Kindred Group Marriage.
    - Part I. Mother Age Civilisation.
    - Part II. General Words for Sex and Kinship.
    - Part III. Special Words for Sex and Relationship.
12. The German Passion Play: A Study in the Evolution of Western Christianity.

Reissue. *Price* **30/-** *net.*

The following works prepared in the Biometric Laboratory can be obtained from H.M. Stationery Office.

The English Convict, A Statistical Study. By CHARLES GORING, M.D. *Text.* Price 9*s.* *Tables of Measurements* (printed by Convict-Labour). Price 5*s.*

The English Convict. An Abridgment, with an Introduction by KARL PEARSON, F.R.S. *Price* 3*s.*

Tables of the Incomplete Γ-Function. Edited with an Introduction by KARL PEARSON, F.R.S. *Price* £2. 2. 0*d. or by Post* £2. 2*s.* 9*d.*

UNIVERSITY OF LONDON, UNIVERSITY COLLEGE

DEPARTMENT OF APPLIED STATISTICS

The Biometric Laboratory

(assisted by a grant from the Worshipful Company of Drapers)

Until the phenomena of any branch of knowledge have been submitted to measurement and number it cannot assume the status and dignity of a science. FRANCIS GALTON.

Under the direction of Professor KARL PEARSON, F.R.S. Assistants: Dr JULIA BELL, E. S. PEARSON, B.A.; Crewdson Benington Student in Anthropometry: G. MORANT, B.Sc., Hon. Research Assistant, J. HENDERSON, M.A.

This laboratory provides a complete training in modern statistical methods and is especially arranged so as to assist research workers engaged on biometric problems.

The Francis Galton Eugenics Laboratory

National Eugenics is the study of agencies under social control, that may improve or impair the racial qualities of future generations, either physically or mentally.

The Laboratory was founded by Sir FRANCIS GALTON and is under the supervison of Professor KARL PEARSON, F.R.S. Galton Research Fellow: ETHEL M. ELDERTON; Reader in Medical Statistics: M. GREENWOOD, M.R.C.P., M.R.C.S. Medical Officer: PERCY STOCKS, M.A., M.D. Assistants: E. C. RHODES, M.A., M. NOEL KARN, M. MOUL and J. O. IRWIN, B.A. Secretary: M. B. CHILD.

It was the intention of the Founder, that the Laboratory should serve (i) as a storehouse of statistical material bearing on the mental and physical conditions in man, and the relation of these conditions to inheritance and environment; (ii) as a centre for the publication or other form of distribution of information concerning National Eugenics; (iii) as a school for training and assisting research-workers in special problems in Eugenics.

Now Ready, Cambridge University Press

New Series, TRACTS FOR COMPUTERS

This series will endeavour to supply a gap in statistical literature, namely "first aid" to the professional computer.

No. I. Tables of the Digamma and Trigamma Functions. By ELEANOR PAIRMAN, M.A. *Price 3s. net.*

Tables for summing $S = \sum_{i=1}^{\infty} \frac{1}{(p_1 i + q_1)(p_2 i + q_2) \ldots (p_n i + q_n)}$ where the p's and q's are numerical factors.

No. II. On the Construction of Tables and on Interpolation. Part I. Univariate Tables. By KARL PEARSON, F.R.S. *Price 3s. 9d. net.*

No. III. On the Construction of Tables and on Interpolation. Part II. Bivariate Tables. By KARL PEARSON, F.R.S. *Price 3s. 9d. net.*

No. IV. Tables of the Logarithms of the Complete Γ-function to Twelve Figures. Reprint with Portrait of A. M. LEGENDRE. *Price 3s. 9d. net.*

No. V. Table of Coefficients of Everett's Central-Difference Interpolation Formula. By A. J. THOMSON, B.Sc. *Price 3s. 9d. net.*

No. VI. Smoothing. By E. C. RHODES, B.A. *Price 3s. 9d. net.*

No. VII. The numerical Evaluations of the Incomplete B-Function or of the Integral $\int_0^x x^{p-1}(1-x)^{q-1}\,dx$ for Ranges of x between 0 and 1. By H. E. SOPER, M.A. *Price 3s. 9d. net.*

No. VIII. Table of the Logarithms of the Complete Γ-Function (to ten decimal places) for Argument 2 to 1200 beyond Legendre's Range (Argument 1 to 2). By EGON S. PEARSON, B.A. *Price 3s. 9d. net.*

BIOMETRIKA. Vol. XIV, Parts I and II

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 400 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 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 French, Italian or German. In the last case the manuscript should be in Roman not German characters. Russian contributors may use Russian but their papers will be translated into English before publication.

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, 40*s. net* per volume (packing and postage 4*s.*): single numbers 15*s. net* (postage 1*s.*). Owing to the scarcity of early volumes, the following rates must now be charged: Volumes I.—II in wrappers, sold only with complete sets, 60*s. net* each, Vols. III—XIII in wrappers, 40*s. net* each. Index to Vols I to V, 2*s. net*. Standard buckram cases for binding, price 3*s.* per volume. Subscriptions should be sent direct to the Secretary (Miss M. B. Child), Biometric Laboratory, University College, London, W.C. 1, to whom orders for series and single copies should be addressed. All cheques should be crossed "*Biometrika Account.*"

PRINTED IN ENGLAND BY J. B. PEACE, M.A., AT THE CAMBRIDGE UNIVERSITY PRESS.