

772940

UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
STACKS

Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

<http://www.archive.org/details/dataanalysis540brig>

Faculty Working Papers

DATA ANALYSIS

Gerald D. Brighton, Professor, Department of
Accountancy
Robert H. Michaelson, Graduate Student in
Accountancy

#540

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

January 16, 1970

DATA ANALYSIS

Gerald D. Brighton, Professor, Department of
Accountancy

Robert H. Michaelson, Graduate Student in
Accountancy

#540

Summary:

In any comprehensive research project, there are essentially five steps. First, one starts with a literature review with regard to a particular research question. Second, one seeks to develop a theory. Third, the research question is finalized, frequently in the form of a hypothesis to be tested. Fourth, data are collected. Fifth, the subject matter of this paper, the data are analyzed in order to come to a resolution of the research question. There are two general approaches to analyzing research data. If the data were gathered concerning a "research question," a description of the data may be sufficient. However, if the data were gathered to accept or reject a formal hypothesis, statistical analysis is usually in order. This paper briefly surveys the principal data analysis methodologies that are available.

1911
1912
1913
1914
1915

| Year | 1911 | 1912 | 1913 | 1914 | 1915 |
|------|------|------|------|------|------|
| Jan | 100 | 100 | 100 | 100 | 100 |
| Feb | 100 | 100 | 100 | 100 | 100 |
| Mar | 100 | 100 | 100 | 100 | 100 |
| Apr | 100 | 100 | 100 | 100 | 100 |
| May | 100 | 100 | 100 | 100 | 100 |
| Jun | 100 | 100 | 100 | 100 | 100 |
| Jul | 100 | 100 | 100 | 100 | 100 |
| Aug | 100 | 100 | 100 | 100 | 100 |
| Sep | 100 | 100 | 100 | 100 | 100 |
| Oct | 100 | 100 | 100 | 100 | 100 |
| Nov | 100 | 100 | 100 | 100 | 100 |
| Dec | 100 | 100 | 100 | 100 | 100 |

I. Introduction

In any comprehensive research project, there are essentially five steps. First, one starts with a literature review with regard to a particular research question. Second, one seeks to develop a theory. Third, the research question is finalized, frequently in the form of a hypothesis to be tested. Fourth, data are collected. Fifth, the subject matter of this paper, the data are analyzed in order to come to a resolution of the research question.

There are two general approaches to analyzing research data. If the data were gathered concerning a "research question," a description of the data may be sufficient. Also, data may be so obvious that a statistical test is not really necessary. However, if the data were gathered to accept or reject a formal hypothesis, statistical analysis is usually in order.

Since this is a brief paper intended only to remind the aspiring researcher of the principal data analysis methodologies that are available, it of necessity includes some "technical jargon" that is not fully explained. To help in the solution of that potential problem, we have cited a general reference just below and are citing specific references in each section of the paper.

General reference: Kerlinger, Fred N., Foundations of Behavioral Research, New York: Holt, Rinehart, & Winston, 1973.

II. Nonstatistical Methods

Once data have been collected concerning a study, it will always be necessary to arrange these data in a logical manner, whether or not statistical analysis is employed, so that the relationships of the research problem can be studied. If statistical analysis is not employed,

the data are still analyzed in a descriptive manner, often with the use of frequency distributions, crossbreaks, and graphs. Once the data are properly analyzed, interpretation can proceed so that conclusions can be drawn concerning the research relationships studied.

In addition to informal description, formalized methods of nonstatistical data analysis have been developed. This formalization concerns the manner in which the data are measured, categorized, ordered, and summarized. For example, scales are employed to measure variables. A scale is a set of symbols or numerals so constructed that they can be assigned by rule to individuals to whom the scale is applied, the assignment being indicated by the individual's possession of whatever attribute the scale is supposed to measure. As an example of a particular application, a scaling technique that has been used to analyze the decisions of Tax Court judges is a cumulative attitude scale called the Guttman scale.

Guttman scaling postulates that if the responses to a set of items referring to an attitude area can be arranged in certain specified ways, that area is scalable and a respondent's rank makes it possible to predict exactly which items were answered favorably. If a set of data is scalable, scalogram analysis will establish an order among the data and pinpoint deviations from the group. This method is simpler than sophisticated statistical methods, but does not provide their detailed specification of pivotal factors.

Reference: Edwards, A., Techniques of Attitude Scale Construction,
New York: Appleton, 1957.

Again, even if statistical methods are being applied, it is always necessary first to sort and arrange the data in some logical order. This may be as simple as tables of data. It is a mistake, therefore, to think that nonstatistical analysis methods are not important. The real point is that increasingly they are not sufficient.

The emphasis in research on statistical methods is because statistical methods are a more scientific approach and consequently increase the acceptability of the results. The remainder of this discussion concerns statistical methods of data analysis.

III. Modern Statistical Methods

A. Situations With Only One Independent and Dependent Variable (or Only Two Groups)

The most common situations with one independent and dependent variable include comparison of a mean, variance, correlation coefficient, or proportion of a sample with the same property of another sample or of the population. Because of the simplicity in experimental design that is required for such statistics, they are often inapplicable.

Reference: Glass, G., and Stanley, J., Statistical Methods in Education and Psychology, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.

B. Situations With More Than One Independent or Dependent Variable (or More Than Two Groups)

In more complex situations, three basic approaches are available: analysis of variance, regression analysis, and factor analysis. Each of these approaches is superior to the others in a particular set of circumstances. In the remainder of this discussion, the question of when each of these approaches is preferred will be briefly sketched.

1. Analysis of variance (anova)

Analysis of variance is a method of identifying, breaking down, and testing for statistical significance the variances in the dependent variable that come from different sources of variation. According to Kerlinger (general reference, p. 238), analysis of variance is the preferable method of data analysis for the following reasons. It:

- (1) Permits us to test several hypotheses at one time.
- (2) Permits us to test hypotheses that cannot be tested in any other way, at least with precision.
- (3) Gives insight into research approaches and methods by focusing sharply and constantly on variance thinking, by making clear the close relationship between research problems and statistical methods and inference, and by clarifying the structure of research design.

Although analysis of variance would virtually always be preferable if it were applicable, applicability is a serious problem. In general, anova is best suited to experimental research in which the subjects can be randomly assigned to cells (groups), the group n's thus kept equal and the assumptions behind the method more or less satisfied (Kerlinger, general reference, pp. 268-269). Anova is not nearly so well suited to ex post (data retrieval or field study) research, or experimental research that uses a number of nonexperimental (attribute) variables, for the following reasons:

- (1) In such cases, interaction can be caused by some extraneous, unwanted, uncontrolled effect.
- (2) If the n's in the cells of a factorial design are not equal or proportionate, the independence of the independent variables is

impaired. While adjustments can be made, they are awkward and not too satisfactory. In nonexperimental anova, the n's get beyond the control of the researcher. Also, in experiments with more than one categorical variable (like race and sex), n's almost necessarily become unequal.

A brief description of the various types of anova follows: .

a. Factorial Anova

Factorial anova is the statistical method that analyses the independent and interactive effects of two or more independent variables on a dependent variable.

Reference: Kirk, R., Experimental Design: Procedures for the Behavioral Sciences, Belmont, Calif.: Brooks/Cole, 1968.

b. Nonparametric Anova

Nonparametric statistics use properties of data other than the strictly quantitative. These statistics are based on properties of data that can be tested against chance expectation: rank, range, periodicity, distribution, etc.

Reference: Siegel, Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill, 1956.

c. Analysis of Covariance (Ancova)

Ancova is used when it is necessary to study groups as they are; when subjects cannot be matched or assigned at random. Ancova tests the significance of the differences between means of final experimental data by taking into account the correlation between the dependent variable and one or more covariates, and by adjusting initial mean differences in the experimental groups.

Reference: Kirk, R., Experimental Design: Procedures for the Behavioral Sciences, Belmont, Calif.: Brooks/Cole, 1968.

d. Multivariate Anova

Multivariate anova is the generalization of anova to any number of independent variables and any number of dependent variables.

Reference: Tatsuoka, M., Multivariate Analysis: Techniques for Educational and Psychological Research, New York: Wiley, 1971.

2. Regression Analysis

Regression analysis is a method for studying the effects and the magnitude of the effects of one or more independent variables on one or more dependent variables using principles of correlation and regression. According to Kerlinger (general reference, p. 268), regression analysis is the best method for analyzing data when the unequal n problem (discussed under anova) appears. This problem is most common with ex post (data retrieval or field study) research, or experimental research that uses a number of nonexperimental (attribute) variables.

A brief description of the various types of regression analysis follows:

a. Multiple Regression Analysis

Multiple regression analysis is a method for studying the effects and the magnitudes of the effects of more than one independent variable on one dependent variable using principles of correlation and regression. This analysis can be applied to cross-sectional data, time-series data, or a combination of both.

Reference: Kerlinger, F., and Pedhazur, E., Multiple Regression in Behavioral Research, New York: Holt, Rinehart and Winston, 1973.

b. Discriminant Analysis

A discriminant function is a regression equation with a dependent variable that represents group membership. The function maximally

discriminates the members of the group; it tells us to which group each member probably belongs. In short, if we have two or more independent variables and the members of say, three groups, the discriminant function gives the "best" prediction, in the least squares sense, of the correct group membership of each member of the sample. The discriminant function, then, can be used to assign individuals to groups on the basis of their scores on two or more measures. The discriminant approach generally provides the most accurate predictions in the sense of quantity right divided by predictions made.

Reference: Tatsuoka, M., Discriminant Analysis: The Study of Group Differences, Champaign, Ill.: Institute for Personality and Ability Testing, 1970.

c. Canonical Correlation

This is simply multiple regression analysis with more than one dependent variable. There are limitations in the interpretation of the results it yields.

Reference: Cooley, W., and Lohnes, P., Multivariate Data Analysis, New York: Wiley, 1971.

3. Factor Analysis

Factor analysis is a method for determining the number and nature of the underlying variables among larger numbers of measures. It serves the cause of scientific parsimony. It reduces the multiplicity of tests and measures toward greater simplicity. It tells us, in effect, what tests or measures belong together--which ones virtually measure the same thing, and how much they do so. It thus is a technique to reduce the number of variables with which the scientist must cope.

Reference: Harman, H., Modern Factor Analysis, 2nd ed., Chicago: University of Chicago Press, 1967.

IV. Conclusion

Obviously, no one method of data analysis is appropriate in all circumstances. The best method must be determined on a project by project basis using the following considerations:

1. Nature of the data available. See discussion under anova.
2. Intent or essence of the hypothesis or research question.
3. Feasibility.
4. Cost.
5. Skill of the researcher.

Finally, since data analysis is the last of five steps in the research process, it will need to be appropriate to the establishing of the validity of the data resulting from the preceding steps.





UNIVERSITY OF ILLINOIS-URBANA



3 0112 060296396