STATISTICAL METHODS

Snedecor and Cochran

SIXTH EDITION

Sixth Edition

METHODS

WILLIAM G. COCHRAN

Professor of Statistics
Harvard University

AMES, IOWA, U.S.A.

GEORGE W. SNEDECOR is professor emeritus of statistics, Iowa State University, where he taught from 1913 to 1958 and where he was for fourteen years director of the statistical laboratory. His writings include a body of scientific journal articles, research bulletins, and books, including *Correlation and Machine Calculation* (with H. A. Wallace), *Calculation and Interpretation of Analysis of Variance and Covariance*, and *Statistical Methods*. He holds a master of science degree from the University of Michigan, and honorary doctor of science degrees from North Carolina State University and Iowa State University. He is a member of the International Statistical Institute, past president of the American Statistical Association, and an honorary Fellow of the British Royal Statistical Society. He has served also as consultant, Human Factors Division, U.S. Navy Electronics Laboratory, San Diego, California, where he now lives.

WILLIAM G. COCHRAN is professor of statistics, Harvard University. He has served formerly on the faculties of Johns Hopkins University, North Carolina State University, and Iowa State University. He holds master of arts degrees from Glasgow University and Cambridge University and an honorary master of arts degree from Harvard University. He is past president of the American Statistical Association, the Institute of Mathematical Statistics, and the Biometric Society. His writings include many research papers in the professional journals of his field; *Sampling Techniques*, 2nd ed., 1963; and *Experimental Designs* (with Gertrude M. Cox), 2nd ed., 1957.

Printed in the U.S.A.

First edition, 1937
Second edition, 1938
Third edition, 1940

Fourth edition, 1946
Second printing, 1946
Third printing, 1948
Fourth printing, 1950
Fifth printing, 1953
Sixth Printing, 1955
Seventh printing, 1955

Fifth edition, 1956
Second printing, 1957
Third printing, 1959
Fourth printing, 1961
Fifth printing, 1962
Sixth printing, 1964
Seventh printing, 1965
Eighth printing, 1966

Sixth edition, 1967
Second printing, 1968
Third printing, 1969
Fourth printing, 1971

International Standard Book Number: 0–8138–1560–6
Library of Congress Catalog Card Number: 67–21577

Preface

In preparing the sixth edition we have kept in mind the two purposes this book has served during the past thirty years. Prior editions have been used extensively both as texts for introductory courses in statistics and as reference sources of statistical techniques helpful to research workers in the interpretation of their data.

As a text, the book contains ample material for a course extending throughout the academic year. For a one-term course, a suggested list of topics is given on the page preceding the Table of Contents. As in past editions, the mathematical level required involves little more than elementary algebra. Dependence on mathematical symbols has been kept to a minimum. We realize, however, that it is hard for the reader to use a formula with full confidence until he has been given proof of the formula or its derivation. Consequently, we have tried to help the reader's understanding of important formulas either by giving an algebraic proof where this is feasible or by explaining on common-sense grounds the roles played by different parts of the formula.

This edition retains also one of the characteristic features of the book—the extensive use of experimental sampling to familiarize the reader with the basic sampling distributions that underlie modern statistical practice. Indeed, with the advent of electronic computers, experimental sampling in its own right has become much more widely recognized as a research weapon for solving problems beyond the current skills of the mathematician.

Some changes have been made in the structure of the chapters, mainly at the suggestion of teachers who have used the book as a text. The former chapter 8 (Large Sample Methods) has disappeared, the retained material being placed in earlier chapters. The new chapter 8 opens with an introduction to probability, followed by the binomial and Poisson distributions (formerly in chapter 16). The discussion of multiple regression (chapter 13) now precedes that of covariance and multiple covariance (chapter 14).

Chapter 16 contains two related topics, the analysis of two-way classifications with unequal numbers of observations in the sub-classes and the analysis of proportions in two-way classifications. The first of these topics was formerly at the end of a long chapter on factorial arrangements; the second topic is new in this edition. This change seemed advisable for two reasons. During the past twenty years there has been a marked increase in observational studies in the social sciences, in medicine and public health, and in operations research. In their analyses, these studies often involve the handling of multiple classifications which present complexities appropriate to the later sections of the book.

Finally, in response to almost unanimous requests, the statistical tables in the book have been placed in an Appendix.

A number of topics appear for the first time in this edition. As in past editions, the selection of topics was based on our judgment as to those likely to be most useful. In addition to the new material on the analysis of proportions in chapter 16, other new topics are as follows:

- The analysis of data recorded in scales having only a small number of distinct values (section 5.8);
- In linear regression, the prediction of the independent variable X from the dependent variable Y, sometimes called linear calibration (section 6.14);
- Linear regression when X is subject to error (section 6.17);
- The comparison of two correlated estimates of variance (section 7.12);
- An introduction to probability (section 8.2);
- The analysis of proportions in ordered classifications (section 9.10);
- Testing a linear trend in proportions (section 9.11);
- The analysis of a set of 2×2 contingency tables (section 9.14);
- More extensive discussion of the effects of failures in the assumptions of the analysis of variance and of remedial measures (sections 11.10–11.13);
- Recent work on the selection of variates for prediction in multiple regression (section 13.13);
- The discriminant function (sections 13.14, 13.15);
- The general method of fitting non-linear regression equations and its application to asymptotic regression (sections 15.7–15.8).

Where considerations of space permitted only a brief introduction to the topic, references were given to more complete accounts.

Most of the numerical illustrations continue to be from biological investigations. In adding new material, both in the text and in the examples to be worked by the student, we have made efforts to broaden the range of fields represented by data. One of the most exhilarating features of statistical techniques is the extent to which they are found to apply in widely different fields of investigation.

High-speed electronic computers are rapidly becoming available as a routine resource in centers in which a substantial amount of data are analyzed. Flexible standard programs remove the drudgery of computation. They give the investigator vastly increased power to fit a variety of mathematical models to his data; to look at the data from different points of view; and to obtain many subsidiary results that aid the interpretation. In several universities their use in the teaching of introductory courses in statistics is being tried, and this use is sure to increase.

We believe, however, that in the future it will be just as necessary that the investigator learn the standard techniques of analysis and understand their meaning as it was in the desk machine age. In one respect, computers may change the relation of the investigator to his data in an unfortunate way. When calculations are handed to a programmer who translates them into the language understood by the computer, the investigator, on seeing the printed results, may lack the self-assurance to query or detect errors that arose because the programmer did not fully understand what was wanted or because the program had not been correctly debugged. When data are being programmed it is often wise to include a similar example from this or another standard book as a check that the desired calculations are being done correctly.

For their generous permission to reprint tables we are indebted to the late Sir Ronald Fisher and his publishers, Oliver and Boyd; to Maxine Merrington, Catherine M. Thompson, Joyce N. May, E. Lord, and E. S. Pearson, whose work was published in Biometrika; to C. I. Bliss, E. L. Crow, C. White, and the late F. Wilcoxon; and to Bernard Ostle and his publishers, The Iowa State University Press. Thanks are due also to the many investigators who made data available to us as illustrative examples, and to teachers who gave helpful advice arising from their experience in using prior editions as a text. The work of preparing this edition was greatly assisted by a contract between the Office of Naval Research, Navy Department, and the Department of Statistics, Harvard University. Finally, we wish to thank Marianne Blackwell, Nancy Larson, James DeGracie and Richard Mensing for typing or proofreading, and especially Holly Lasewicz for her help at many stages of the work, including the preparation of the Indexes.

George W. Snedecor
William G. Cochran

A SHORT COURSE IN THE ELEMENTS OF STATISTICAL METHOD

Table of contents

Chapter 1. Sampling of Attributes

Chapter 2. Sampling From a Normally Distributed Population

Chapter 3. Experimental Sampling From a Normal Population

Chapter 4. The Comparison of Two Samples

Chapter 5. Shortcut and Non-parametric Methods

Chapter 6. Regression

Chapter 10. One-Way Classifications. Analysis of Variance

Chapter 11. Two-Way Classifications

Chapter 12. Factorial Experiments

Chapter 13. Multiple Regression

Chapter 14. Analysis of Covariance

Chapter 15. Curvilinear Regression

STATISTICAL METHODS

Sampling of attributes

1.1—Introduction. The subject matter of the field of statistics has been described in various ways. According to one definition, statistics deals with techniques for collecting, analyzing, and drawing conclusions from data. This description helps to explain why an introduction to statistical methods is useful to students who are preparing themselves for a career in one of the sciences and to persons working in any branch of knowledge in which much quantitative research is carried out. Such research is largely concerned with gathering and summarizing observations or measurements made by planned experiments, by questionnaire surveys, by the records of a sample of cases of a particular kind, or by combing past published work on some problem. From these summaries, the investigator draws conclusions that he hopes will have broad validity.

The same intellectual activity is involved in much other work of importance. Samples are extensively used in keeping a continuous watch on the output of production lines in industry, in obtaining national and regional estimates of crop yields and of business and employment conditions, in the auditing of financial statements, in checking for the possible adulteration of foods, in gauging public opinion and voter preferences, in learning how well the public is informed on current issues, and so on.

Acquaintance with the main ideas in statistical methodology is also an appropriate part of a general education. In newspapers, books, television, radio, and speeches we are all continuously exposed to statements that draw general conclusions: for instance, that the cost of living rose by 0.3% in the last month, that the smoking of cigarettes is injurious to health, that users of "Blank's" toothpaste have 23% fewer cavities, that a television program had 18.6 million viewers. When an inference of this kind is of interest to us, it is helpful to be able to form our own judgment about the truth of the statement. Statistics has no magic formula for doing this in all situations, for much remains to be learned about the problem of making sound inferences. But the basic ideas in statistics assist us in thinking clearly about the problem, provide some guidance about the conditions that must be satisfied if sound inferences are to be made, and enable us to detect many inferences that have no good logical foundation.

1.2—Purpose of this chapter. Since statistics deals with the collection, analysis, and interpretation of data, a book on the subject might be expected to open with a discussion of methods for collecting data. Instead, we shall begin with a simple and common type of data already collected, the replies to a question given by a sample of the farmers in a county, and discuss the problem of making a statement from this sample that will apply to all farmers in the county. We begin with this problem of making inferences beyond the data because the type of inference that we are trying to make governs the way in which the data must be collected. In earlier days, and to some extent today also, many workers did not appreciate this fact. It was a common experience for statisticians to be approached with: Here are my results. What do they show? Too often the data were incapable of showing anything that would have been of interest to an investigator, because the method of collecting the data failed to meet the conditions needed for making reliable inferences beyond the data.

In this chapter, some of the principal tools used in statistics for making inferences will be presented by means of simple illustrations. The mathematical basis of these tools, which lies in the theory of probability, will not be discussed until later. Consequently, do not expect to obtain a full understanding of the techniques at this stage, and do not worry if the ideas seem at first unfamiliar. Later chapters will give you further study of the properties of these techniques and enhance your skill in applying them to a broad range of problems.

1.3—The twin problems of sampling. A *sample* consists of a small collection from some larger aggregate about which we wish information. The sample is examined and the facts about it learned. Based on these facts, the problem is to make correct inferences about the *aggregate* or *population*. It is the sample that we observe, but it is the population which we seek to know.

This would be no problem were it not for ever-present variation. If all individuals were alike, a sample consisting of a single one would give complete information about the population. Fortunately, there is endless variety among individuals as well as their environments. A consequence is that successive samples are usually different. Clearly, the facts observed in a sample cannot be taken as facts about the population. Our job then is to reach appropriate conclusions about the population despite sampling variation.

But not every sample contains information about the population sampled. Suppose the objective of an experimental sampling is to determine the growth rate in a population of young mice fed a new diet. Ten of the animals are put in a cage for the experiment. But the cage gets located in a cold draught or in a dark corner. Or an unnoticed infection spreads among the mice in the cage. If such things happen, the growth rate in the sample may give no worthwhile information about that in the population of normal mice. Again, suppose an interviewer in an opinion