Elementary Statistical Analysis

S. S. WILKS

$4

ELEMENTARY STATISTICAL ANALYSIS

By S. S. WILKS

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY

Reprinted with corrections but
without change of pagination

JANUARY 1951
SEVENTH PRINTING 1954
EIGHTH PRINTING 1956
NINTH PRINTING 1957
TENTH PRINTING 1958

PRINTED IN THE UNITED STATES OF AMERICA

Reprinted, 1961

ELEMENTARY STATISTICAL ANALYSIS

PREFACE

This book has been prepared for a one-semester basic course in elementary statistical analysis which, at Princeton, is the introductory course for all fields of statistical application, and is usually taken in the freshman year. It is especially designed for those who intend to go into the biological and social sciences. It presupposes one semester of elementary mathematical analysis covering topics such as those included in the first half of F. L. Griffin's Introduction to Mathematical Analysis. The material has been developed from two years of experience with such a course.

An effort has been made throughout the book to emphasize the role played in statistical analysis by a sample of measurements and a population from which the sample is supposed to have arisen. Only three chapters are devoted to elementary descriptive statistics of a sample of measurements. In these three chapters the idea of a population is presented on a purely intuitive basis. Probability concepts are then introduced. This makes it possible to use these basic concepts at an early stage in dealing more critically with the idea of a population and sampling from a population. Considerable attention is given to the application of sampling principles to the simpler problems of statistical inference such as determining confidence limits of population means and difference of means, making elementary significance tests, testing for randomness, etc. No attempt has been made here (in fact, there is not enough time in one semester!) to go into analysis of variance and more sophisticated problems of statistical inference. An elementary treatment of analysis of pairs of measurements including least squares methods is presented. Special effort has been made throughout the book to keep the mathematics elementary and to state specifically at which points the mathematics is too advanced to present.

The course in which this material has been used has been conducted satisfactorily (not ideally!) without the use of a computing laboratory. The problems in the exercises have been selected so that computations can be carried out effectively by the use of a small handbook of tables such as C. D. Hodgman's Mathematical Tables from Handbook of Chemistry and Physics.

The author would like to express his appreciation to: Professor R. A. Fisher and Messrs. Oliver and Boyd for permission to use the material in Table 10.2; to Professor E. S. Pearson and the Biometrika Office for permission to reprint Figures 10.1 and 13.6; to Dr. C. Eisenhart and Miss Freda S. Swed for permission to use the material in Table 12.3; and to the College Entrance Examination Board for permission to reprint Figure 13.5.

Finally, the author takes this opportunity to acknowledge the benefit of many helpful discussions he has had with his colleagues, Professors A. W. Tucker and J. W. Tukey, and Professor F. Mosteller of Harvard University during the preparation of the material. He is also indebted to Drs. K. L. Chung, R. Otter and D. F. Votaw, who assisted with the preparation of Chapters 4, 6, 7 and 8; to Mr. J. G. C. Templeton, who checked the computations and proofread the manuscript; to Mr. R. B. Murphy who drew the figures; and to Mrs. Frances Purvis who typed the manuscript.

The material in this book is still in a tentative form. Any errors or weaknesses in presentation are solely the responsibility of the author. Corrections, criticisms, and expressions of other points of view on the teaching of such a course will be gratefully received.

S. S. Wilks

Princeton, New Jersey
September
1948

CONTENTS

CHAPTER 1. INTRODUCTION

1.1 General Remarks.

To many persons the word statistics means neatly arranged tables of figures and bar charts printed in financial sections of newspapers or issued by almanac publishers and government agencies. They have the impression that statistics are figures used by persons called statisticians to prove or disprove something. There is plenty of ground for this impression. Anyone who tries to make sense out of a set of observational or experimental data is assuming the role of a statistician, no matter whether he is a business executive, a medical research man, a biologist, a public opinion poller or an economist. Some sets of data are very simple and the implications and conclusions inherent in the data are obvious. Other data, however, are complex and may trick and confuse the statistical novice, even though he may be an expert in the subject matter field from which the data came. The only way to reduce this confusion is through scientific methods of collecting, analyzing and interpreting data. Such methods have been developed and are available. The fact that expert statisticians well-versed in these methods can and do come out with sound conclusions from a given set of data which differ very little from one statistician to another is evidence that there are no real grounds for the naive claim that statistics can prove anything. Some of the most dangerously deceptive uses of statistics occur in situations where correct conclusions are drawn and which seem to depend on the statistics, when in fact, the statistics have little if anything to do with the original question. While mathematics cannot protect a person from this danger directly, familiarity with numerical analysis will make it easier to spot such hidden fallacies.

Modern statistical method is a science in itself, dealing with such questions as: How shall a program of obtaining data be planned so that reliable conclusions can be made from the data? How shall the data be analyzed? What conclusions are we entitled to draw from the data? How reliable are the conclusions? To try to present all the statistical methods that are known and used at present would be an encyclopaedic venture which would lead us deeply into statistical theory and many subject matter fields. However, there is a body of fundamental concepts and elementary methods which can be presented in a beginning course. The purpose of this course is to do just this, and to illustrate the concepts and methods on simple examples and problems from various fields.

You will ask at this juncture what kinds of situations come up which involve these fundamental concepts and elementary methods. Series or sets of raw statistical observations or measurements arise in many ways and in many different fields. The number of observations or measurements needed or feasible varies tremendously from one situation to another. In some cases, as in peacetime firing of large caliber naval guns, only a very small sample of measurements (of where the shells actually fall) can be obtained because of cost. In other cases, as for example in a Gallup poll for a presidential election, the number of observations runs into the thousands. In some situations the sample comprises the entire population of measurements or observations which could be made, particularly in census-type work where complete enumerations of populations are made. Federal and state government agencies and national associations compile data on entire populations of objects, e.g., births, deaths, automobile registrations, number of life insurance policies, etc., etc.

There are two general types of statistical observations: (1) quantitative and (2) qualitative. We shall discuss these separately.

1.2 Quantitative Statistical Observations.

By quantitative statistical observations we mean a sequence or set of numerical measurements or observations made on some or all of the objects in a specified population of objects. If the observations are made on some of the objects we call the set of observations a sample. Let us illustrate by some examples.

Suppose a men's clothing store proprietor writes down from sales slips the sizes of men's overcoats sold every other week for September and October. He would end up with a list of numbers that might run something like this: 36, 42, 44, 30, 40, 36, ... and so on for 145 numbers. The list of numbers written down constitutes a sample of sizes from the population of overcoats he has sold during September and October.

By making an analysis of a series of specimens from a certain deposit of ore, for percent of iron, a chemist might turn up with 1 measurement on each of five specimens something like this: 28.2, 27.6, 29.3, 28.2, 30.1. This is a sample of iron percentage measurements from five specimens out of an extremely large population of possible specimens from the deposit.

A record-keeping bridge player might keep track of the number of honor cards he gets in 200 bridge hands finding some such sequence as: 9, 5, 7, 2, 4, 8, 0, 3 and so on for 200 numbers. He would therefore accumulate honor card counts in a sample of 200 hands out of a population of indefinitely many hands which could be conceivably dealt under a given shuffling and cutting practice.

A quality control inspector interested in maintaining control of the inner diameter of bushings turned out by an automatic lathe would pick a bushing every 30 minutes and measure its inner diameter, obtaining some such sequence as this: 1.001", .998", .999", 1.001", 1.002", etc. He is selecting a sample of bushings out of the population of bushings being manufactured by this lathe.

A Princeton personnel researcher goes through the record cards of all 246 freshmen who took Mathematics 103 and writes down two numbers for each student; his College Board mathematics score and his final group in Mathematics 103. His sequence of pairs of numbers (arranged alphabetically with respect to students' names) might run like this (680, 3+), (740, 1-), (530, 5), (620, 3), (510, 6) and so on for 246 pairs of scores. In this case the sample would consist of all of the freshmen in the population of freshmen who took Mathematics 103.

Notice that in this last example each quantity in the sequence consists of two measurements. We could mention many examples in which the sequence would contain not only pairs, but sets of three, four or more measurements.

We could continue with dozens of such examples. It is to be noted that in every example mentioned, the series of statistical measurements may be regarded as a sample of measurements from a population of measurements. In general, there are two kinds of populations: finite populations and indefinitely large populations. For example, the undergraduates now enrolled at Princeton constitute a finite population. The licensed hunters of Pennsylvania form a finite population. The sequence of numbers of dots obtained by rolling a pair of dice indefinitely many times is an indefinitely large population. In the case of the dice, the indefinitely large population consisting of the sequence of dots is generated by successively rolling the dice indefinitely many times. The population in this case depends on various factors, such as the dice themselves (which may be slightly biased), the method of throwing them and the surface on which they are thrown. If the dice were "perfectly true", and if they were thoroughly shaken before throwing and if they were thrown on a "perfect" table top, we can imagine having an "ideal" population. We can use probability theory (to be discussed in later sections) to predict characteristics of this "ideal" population, such as the fraction of rolls of two dice in which 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 dots will occur "in the long run", the fraction of sets of three rolls of two dice in which 6, 2, 11 dots are obtained in that order, and so on. In the case of the lathe turning out bushings, we may essentially consider the population to be indefinitely large, since the population is being generated by the production of one bushing after another, with no consideration of a "last bushing" (which, as a matter of fact, sooner or later will be made). But the important thing about this population of bushings is that it actually changes because of tool wear or changes in raw material from which bushings are made or change of operators, etc. For any particular shift of operators the population may be fairly constant and a sample of inner diameters of bushings taken during that shift may be considered as a sample from an indefinitely large potential population of bushings that might be turned out under the particular conditions of that shift. Even in the case of a finite population of objects, a given sampling procedure might be such that when applied to a relatively small number of objects in the population it essentially begins to generate a population different from the finite population one thinks he is sampling. For example, if one should take every 20th residence listed in the Princeton telephone directory and call the number for information about that residence, one has, on the face of it, a sampling procedure which might be expected to yield information from which one could make accurate inferences about the population of Princeton residences with telephones. Actually, there will be a substantial number of residences for which there will be no response. If we take the sample of residences in which a response is obtained, our sampling procedure is not sampling the population of residences with telephones -- it is sampling the population of residences with telephones in which telephones are answered. These two populations of residences are actually different. For example, the second tends to have larger families and more old people and other stay-at-home types of people in them. Of course, if we make enough repeated telephone calls to the residences who did not answer the telephone originally, we would then be sampling the first population.

What is supposed to be done with samples of measurements? The main reason for keeping track of such measurements is not simply to accumulate a lot of numbers, but, in general, to try to learn something about the main features of the set of numbers -- their average, how much they vary from one another, etc., -- for the purpose of making inferences about the population from which they can be considered as having been "drawn". None of these measurement-makers want to get any more data than necessary to make these inferences. Once he has what he thinks is a pretty sound inference as to what the population is (that is, a "reasonably" accurate description of it from the sample) he can then begin to consider what ought to be done (perhaps nothing) to change it in some direction or other which will be to his advantage, or more often, to use this information elsewhere.

The clothing store proprietor can find out from a sample whether he is stocking the right distribution of sizes of overcoats; the chemist (or rather his boss) can use the results of his sample of analyses to help decide whether the iron ore is worth mining; the bridge player can satisfy his curiosity as to how frequently various numbers of honor cards are obtained (since he presumably does not want to try to figure these things out mathematically on the assumption of perfect shuffling); the quality control expert can see whether the inner diameters of his bushings are being kept within the specified tolerances and if not whether the holes are being made too large or too small and by how much; the personnel researcher can determine how high the relationship or correlation is between the College Board mathematics test and the final group in Mathematics 103 and whether it is high enough to make useful predictions as to how well each entering freshman can be expected to do on Mathematics 103 from a knowledge of his College Board mathematics score.

Evidently, condensing the sample data in some way is vital in any one of these problems. The first thing that has to be learned in statistics is how to condense the sample data and present it satisfactorily. The main thing that has to be learned is what kind of inferences or statements can be made from the sample about the population sampled and how reliable these inferences are. The simplest thing that can be done in condensing and describing samples of quantitative data is to make frequency distributions and describe them by calculating certain kinds of averages. Such quantities calculated from samples for describing samples are called statistics. Similarly, populations are described by population parameters.

Only rarely is it possible to know precisely the values of population parameters, simply because only rarely does one ever have the data for the entire population. The usual situation is that one only has a sample from the population. Hence the usual problem is to calculate statistics from the sample frequency distribution and then try to figure out from the values of these statistics what the values of the parameters of the population are likely to be. In case of extremely large samples, the statistics of properly drawn samples will have values very close to those of the corresponding population parameters. For example, the average of a very large sample of measurements "randomly drawn" from a population will be very close to the average of the entire population of measurements. But in the case of small samples the discrepancies become larger, and the problem of inferring the values of population parameters from sample statistics becomes more complicated and has to be settled by means of probability theory.

There is a source of information in sequences of observations which is particularly useful in such fields as analysis of data from scientific experiments and industrial research, development and production. This is the information contained in the way in which measurements jump about from value to value as one goes through the sequence of sample measurements in the order in which they are made. The usual frequency distribution analysis (to be studied in Chapters 2 and 3) does not take account of this information. But we shall discuss this kind of sequence analysis in Chapter 12.

1.3 Qualitative statistical observations.

By qualitative statistical observations we mean a sequence of observations in which each observation in the sample (as well as the population) belongs to one of several mutually exclusive classes which are likely to be non-numerical. Let us consider some examples.

A person tosses a coin 50 times and obtains some such sequence as H, H, T, T, H, T and so on (H=heads, T=tails). He is essentially drawing a sample of 50 tosses out of an indefinitely large population of tosses, and is making an observation on each toss as to whether it is an H or a T.

A movie producer polling agent stationed at the exit of the Princeton Playhouse asking outgoing moviegoers whether or not they liked Movie X (just seen), might get a sequence of 100 answers starting off like this: Yes, Yes, No, Yes, No, Yes, Yes and so on. (He probably wouldn't stop at this simple question, however, since he would probably at least want to know why he or she liked it or not.) The answers here are qualitative; they are either yes or no. The data accumulated are responses from a sample of moviegoers out of the population of moviegoers who saw Movie X at the Playhouse.