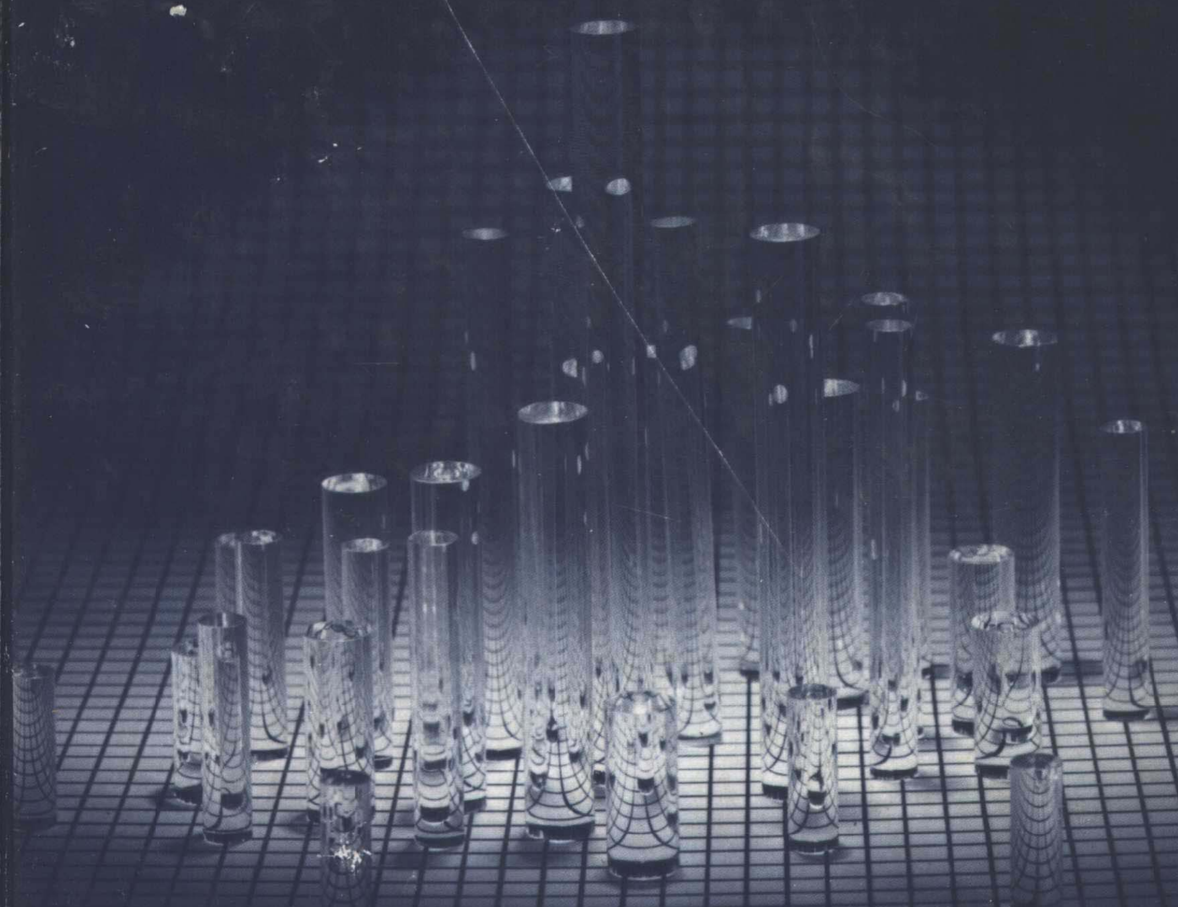
STATISTICAL ANALYSIS FOR ADMINISTRATIVE DECISIONS



FOURTH EDITION

STATISTICAL ANALYSIS FOR ADMINISTRATIVE DECISIONS

CHARLES T. CLARK
Professor of Business Statistics
The University of Texas at Austin

LAWRENCE L. SCHKADE
Professor and Chairman
Department of Systems Analysis
The University of Texas at Arlington

M28

Published by

SOUTH-WESTERN PUBLISHING CO.

CINCINNATI   WEST CHICAGO, ILL.   DALLAS   PELHAM MANOR, N.Y.   PALO ALTO, CALIF.

ISBN: 0-538-13280-9

Library of Congress Catalog Card Number: 82-60363

2 3 4 5 6 7 8 D 0 9 8 7 6 5 4 3

Printed in the United States of America

PREFACE

The objective of this edition of STATISTICAL ANALYSIS FOR ADMINISTRATIVE DECISIONS is basically the same as that of the third edition in that it has been written for use as a text for courses on modern statistical concepts and their application to problems of administration in business, industry, government, and other types of organizations and for use as a reference for administrators who wish to use statistical methods in making decisions. The concepts and methods presented in this edition range from elementary to advanced in degree of sophistication and comprise a significant portion of the methodology that is called quantitative methods. The level of presentation presumes no mathematical preparation beyond college algebra and no previous background in digital computers. Topics that are more advanced than these basic levels are developed fully in the book.

The writing of this edition has provided the opportunity to improve upon the third edition by presenting more recent data in illustrations and problems, altering the emphasis for selected topics, adding the presentation of new concepts, simplifying some of the notation, and increasing the readability of the manuscript. Thanks to the interest and kindness of a large number of students, instructors, and administrators who used the third edition, many valuable comments and suggestions were contributed, and these were utilized in the preparation of this edition.

The increase in the use of statistical and other quantitative methods and the application of computers in the decision-making process in recent years has been dramatic. This edition reflects the trend by presenting statistical concepts that are used in making administrative decisions and by illustrating how digital computers are used in the application of statistical methods. The availability of computers for problem solving has led to the development of new statistical methods and to the modification of some of the traditional manual computational methods. A selection of these new statistical methods is included in this edition. It is not necessary, however, that a computer be used in conjunction with this book, for the topical presentation is designed to be used either with or without a computer.

The style of presentation reflects an approach to the exposition of statistical concepts that highlights clarity and understandability, and it has been developed on the basis of extensive combined teaching and consulting experience. For example, to clarify further the differences in and applications of descriptive statistics and parametric and nonparametric methods, greater emphasis is placed on the quality of data and levels of measurement in this edition. The discussion of selected topics has been condensed and the organization of topics has been changed slightly to enhance readability.

The organization of this book is similar to that of the preceding edition. The symbols and notation in this book are consistent with those recommended by the American Statistical Association. Study questions and many new problems are included at the end of each chapter. An extensive group of statistical tables are included in the appendix. The examples and problems in this book have been selected to present understandable and realistic situations. The study questions are designed to focus attention on important concepts and relationships developed in the discussions. Solutions to most of the odd-numbered problems are provided in the back of the book as reinforcements to the reader when solving the problems at the end of each chapter. The selected readings contain brief comments concerning other books that offer similar or more extensive presentations of the topics included in each of the chapters.

The authors are indebted to many persons for their interest and encouragement in the preparation of this manuscript. In particular, appreciation is expressed to Dr. John R. Stockton, who suggested the writing of this book; to Pearl Clark for her valuable assistance in editing the manuscript; and to Dr. Bawa Singh and David Schkade for their valuable comments.

The authors are also indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Cambridge, to Dr. Frank Yates, F.R.S., Rothamsted, and to Messrs. Oliver and Boyd, Ltd., Edinburgh, for their permission to reprint Tables III and IV from their book *Statistical Tables for Biological, Agricultural and Medical Research,* and Tables III and V-A from *Statistical Methods for Research Workers;* to Dr. Stephen P. Shao, Old Dominion University, and the publisher for permission to reprint Table B from his book *Mathematics of Finance;* to the Editors of the *Annals of Mathematical Statistics* for their permission to publish data included in Table 12.2; and to Dr. Robert Schlaifer and the President and Fellows of Harvard College for permission to publish the tables of the Unit Normal Loss Function that appear in *Probability and Statistics for Business Decisions* by Robert Schlaifer, published by the McGraw-Hill Book Company in 1959.

CONTENTS

APPENDIX TABLES

PART ONE

STATISTICAL DATA AND DESCRIPTIVE MEASURES

CHAPTER

1

THE NATURE OF STATISTICS

Statistics is a discipline which is made up of a body of methods and techniques for quantitatively expressing and interpreting knowledge. It is especially designed to treat the complexities and uncertainties of the business and administrative world. Today, more and more administrative and scientific decisions are made on the basis of information presented in quantitative form.

As organizations grow in size and complexity, the administrator is forced to make many decisions, some of which involve millions of dollars. Actions must be taken on the basis of the facts available about the organization and the economic, political, and physical aspects of the external environment that affect the organization.

Modern statistical methods provide one of the most useful and necessary sets of tools for decision making. It is the purpose of this book to introduce the basic statistical techniques, to present some of the more advanced methods for statistical decision making, and to give students an understanding of the power of statistical analysis in helping to obtain and summarize useful information from data.

Important applications of probability and statistics are found in almost all fields of endeavor. Success stories involving the use of statistical analysis can be found in such diverse areas as health and medicine, government, space programs, business, and weather research, to name only a few. The following examples are mentioned briefly to whet the student's appetite for what is to come.

In 1954 the largest and most expensive statistical experiment in the history of medicine was carried out to test the effectiveness of the Salk vaccine as protection against paralysis or death from poliomyelitis. This experiment involved over a million young children, cost more than $5 million, and led to a notable triumph in the battle against disease. Statistical techniques are now frequently used to test the safety and effectiveness of new drugs before they are marketed.

The 1936 national election provided for the first time an excellent example of how election polling can be used as a kind of "acid test" on the effectiveness of statistical techniques used in opinion measurement now applied in marketing. An election represents one of the few situations in which the forecast produced by the statistician can be compared to the actual (voting) results. Using scientifically selected samples, the Gallup Poll has made estimates of the percent of the total vote to be received by each presidential candidate prior to each of the last 12 national elections held since 1948. The final votes varied from those estimates by an average of only 1.6 percent.

Probability and statistics have played an important role in the development of *reliability theory*. The Apollo space vehicle would not have been possible without the ability to test the final complex system statistically rather than to test it by performing a mission. This same reliability theory developed for the space program is now being applied to the manufacture of household appliances, automobiles, telephones, power supplies, and so on. For example, the five-year guarantee still given by some automobile manufacturers resulted from their determination of the reliability of components included in the guarantee.

The Japanese automobile industry is yet another excellent example of the application of reliability theory. One reason they have had such tremendous success in selling automobiles all over the world has been their reputation for reliability. The Japanese have achieved this high reliability through the vigorous application of statistical quality control to every step in the manufacturing process.

The great variability in rainfall in the United States from one year to the next has always been a major problem for the nation's farmers. The need for accurate quantitative predictions of rainfall amounts has led to the use of a technique called *multiple discriminant analysis* to predict the probability of precipitation. Statisticians were also called on to design and conduct experimental programs to establish whether or not various cloud-seeding techniques can reliably increase precipitation to meet the needs of a growing and thirsty population.

THE DECISION-MAKING PROCESS IN ADMINISTRATION

The decision-making process in business follows essentially the same basic steps used for problem solving in physics, engineering, or chemistry. In business and in the social sciences the variables are often more numerous and more difficult to measure and control than in the physical sciences, but the steps are the same. They are:

1. Simplification
2. Building a decision model
3. Testing the model
4. Using the model to find a solution

Simplification

The ancient Greeks in the fourth, fifth, and sixth centuries before Christ made a great discovery. They learned how mathematical reasoning could be used to simplify and explain the things they observed in nature. When they saw that the heavenly bodies are spheres, the surfaces of lakes are flat, light travels in a straight line, and the sides of a house form a rectangle, they began to study lines, planes, circles, triangles, and rectangles as abstractions.

In their study of geometry the Greeks observed that certain basic facts are obvious. A straight line is the shortest distance between two points in a plane. Or, all the points on the circumference of a circle are equidistant from the center. They reasoned that if some new facts could be derived, these facts would apply to all those physical objects with the same basic properties. If the area of a triangle could be shown by reasoning to be one half the base times the altitude, then a carpenter could use this abstract idea to determine how much lumber is needed to enclose the end of a gable roof.

By dealing first with abstractions that strip away all the nonessential details of a situation, it is possible to reason with the remaining basic relationships and to use this reasoning to cover a multitude of complex cases. Better still, the basic reasoning may produce other meaningful information that is entirely unforeseen. For example, the power company executive who uses a multiple regression equation to predict the need for electric power each of the 24 hours of the day, may get highly accurate predictions. The equation considers relationships between variables which would not be evident without this type of statistical analysis.

Building a Decision Model

Simplification is but the first step in the decision-making process. The decision maker next turns to the task of taking the essential factors in the problem and arranging them in a *model*, which is a representation of reality designed to explain it and used to predict or to control it.

While the mathematician leads in the use of abstract concepts and models, mathematical ways of thinking about essential relationships have long been used by the physical scientist. This kind of thinking is rapidly being adopted by the social scientist and the business executive.

Example: Mathematicians who see an object dropped from a tower can estimate with great accuracy the time it will take the object to reach the ground. They use a simple model,

$$d = 16t^2$$

where d is the distance in feet from the top of the tower to the ground and t is the time in seconds it takes the object to fall. Mathematicians may disregard the resistance of air and the weight of the object since they know that when air resistance is not considered, all objects fall at the same rate under the pull of gravity. The entire physical problem is now a simple algebra problem; they need only to insert a value for d in the equation and solve for t.

From this example, several characteristics of a model can be pointed out:

1. It is a simplified representation of an actual situation.
2. It need not be complete or exact in all respects.
3. It concentrates on the most essential relationships and ignores the less essential ones.
4. It is more easily understood than the empirical situation and, hence, permits the problem to be more readily solved with a minimum of time and effort.
5. It can be used again and again for similar problems or the model can be modified if necessary to solve new problems with added complications. The model will often work reasonably well even under conditions which are not altogether ideal.

The statistical formulas used in this book can be thought of as mathematical models capable of providing the decision maker with useful tools for the important and arduous task of making decisions. Some of these models are quite simple, while others are complex.

A great deal of exciting work is being done on the frontiers of modern business decision making. Those individuals charged with the responsibility of guiding the destiny of large corporations have been faced with a growing problem in recent years. Too often they find that facts gathered in the traditional manner are so long in preparation that the decisions they are designed to guide must be made before all the facts are available. To solve this problem, research and planning personnel in many large organizations are working to develop complex models of their firm or industry, programmed for a large computer and capable of giving approximate answers to far-reaching questions long before these same answers could come through conventional organizational channels. This is model building in its most sophisticated form—*simulation.*

Testing the Model

The real test of any model is whether it predicts outcomes with usable accuracy. If the formula $d = 16t^2$ does not predict accurately the time it will take an object dropped from a tower to hit the ground, the formula must be replaced by one that can predict with the needed accuracy. An oversimplification of the empirical situation may have led to the development of a model that ignores elements essential to its functioning as a predictor. Even if all the essential elements have been identified, the model must be tested to see if the correct relationships between these essentials have been established. Certainly, as models become more complex, a great deal of testing may be necessary to establish that they will work.

In the field of statistics many of the models used are ones that have withstood the test of time and are known to be reliable within given limits. The task of the student is to understand what the limitations of a given model are and how it can be used to produce the required results. A simple model such as a least-squares trend equation has been used for many years and its characteristics are well known. Some of the more sophisticated models are still being tested in order to learn what they will do.

Using the Model to Find a Solution

One of the difficulties encountered in a brief discussion of models is that there are many kinds of models and they have many uses. Furthermore, they involve many different levels of abstraction. In statistics a model may be used to describe the trend component in a time series or to determine if a manufacturing process is producing an acceptable product.

The statistician may use a model to determine which variables are pertinent to changes occurring in some dependent variable and to what extent each shares in the process. At a higher level of abstraction, there are simulation models that operate on a problem-solving sequence in which the input for one stage is the output from a previous stage. Such problems are solved through the application of the so-called *Monte Carlo* methods.

The very nature of human thought is such that model building, model testing, and model application to solve problems are an integral part of any organized thought process.

WHAT IS STATISTICS?

Much of the confusion that arises in the public mind about statistics comes from the fact that the word *statistics* has two meanings.

1. When the administrator asks for the most recent statistics on population and housing for a state, the request is for facts. These facts are in quantitative form and are, strictly speaking, *statistical data*.
2. When a business manager asks a quality control engineer to explain how statistics can be used to determine whether a shipment of parts should be accepted for use or returned to the supplier as unsatisfactory after inspecting only a few parts, the manager is asking about a *statistical method*.

To the scientist, the engineer, the business executive, or any other person engaged in problem-solving activities, the second sense of the word statistics is of equal importance. It refers to the vast and ever-growing body of methods for collecting, summarizing, analyzing, and interpreting quantitative facts. These techniques are a part of the scientific method and can be applied in many fields of endeavor. In this book the prime emphasis will be on the application of statistical methods to the solution of administrative and business problems.

The word *statistic* should also be mentioned here for completeness. Throughout this text, when the word statistic is used in the singular, it will have a meaning completely different from the two previously given. Statistic will be used to refer to an arithmetic mean, a median, a standard deviation, or some other descriptive measure computed from a sample.

Descriptive and Inductive Statistics

The word statistics has been defined as meaning both "statistical data" and "statistical methods." A further distinction is often made between descriptive statistics and inductive statistics. The term *descriptive statistics* is confined to the treatment of data for the purpose of describing their characteristics. This term is distinct from the term *inductive statistics*, which involves making forecasts, estimations, or judgments about some larger group of data than that actually observed or about some future happening based on a study of historical data. The arithmetic mean of a sample of observations is a descriptive measure, but if it is used to estimate the arithmetic mean of the larger group from which the sample was drawn, inductive techniques are involved.

The foundation of inductive statistics is probability. Some examples of inductive techniques are the use of probability theory to estimate the likelihood that a sample has been drawn from a given universe to forecast sales or to predict the action of one variable based on its relationship with other variables. The whole idea of using probability theory to solve scientific or business problems on a formal, wide-scale basis is relatively recent but is of rapidly growing importance.

Example: To point up the distinction between descriptive statistics and inductive statistics, imagine a manufacturer who is studying a report made by the quality control department of a shipment of parts received by the firm for an assembly operation. The shipment is from one of the manufacturer's regular suppliers and contains 10,000 parts. The inspectors have checked 200 of these parts drawn at random from the shipment.

The inspection shows 12 out of the 200 parts examined are defective. This is 6 percent of the sample total. As long as the executive is interested only in this sample percent of defective parts, he or she is dealing with descriptive statistics.

Imagine further that there is a contract between the supplier and the manufacturer which states that the manufacturer will not accept any shipment of parts with more than 5 percent defective. Since the sample percent of defectives is 6 percent, the manufacturer may decide that the percent of defectives in the entire shipment is too high and may reject the entire lot. Just as soon as the manufacturer generalizes about the proportion of defective parts of the entire shipment of 10,000 from the results obtained from an inspection of only 200 of the parts, the focus has moved into the domain of inductive statistics. The manufacturer has generalized about the quality of the group when having information about only a part of it.

The generalization that the manufacturer must make in the example above is not an easy one. The manufacturer must recognize that the inspectors may have gotten a "bad" sample which has 6 percent defectives. If another sample of 200 parts were selected, "chance" might dictate that it would contain only 4 percent defectives. If the difference between the sample percent and the critical 5 percent for the whole shipment had been large, the decision might be easier to make. Also, if the sample had been much larger, it would be reasonable to assume that it might be more representative of the whole group. Finally, the way in which the sample was drawn is important. If only parts from the top of each box had been selected, these might be the ones most subject to damage. When a sample experiment has been properly designed using appropriate statistical techniques, and then the data are scientifically selected to avoid bias, the findings of the study may give both informative and reliable answers to difficult questions.

In discussing models, the formula

$$d = 16t^2$$

was used to describe an important law of falling bodies, which determines when an object will hit the ground. By using Newton's laws of motion, one can also compute exactly where the falling object will hit the ground beneath the tower. On the other hand, if a six-sided die is rolled on the table, there are no laws to predict whether the top face will stop as a one, two, three, four, five, or six. However, it is possible to predict that if the die is rolled a great many times, the one will appear on top about one-sixth of the time. The distinction that is made between these two situations is the distinction between a *causal law* and a *statistical law*.

Even though the causal law permits an estimate of the exact spot where the object will fall, the problem is not that simple. If an experiment is carried out repeatedly, it will be observed that the actual hits will form a certain pattern about the predicted spot. Only if all physical conditions could be perfectly controlled in the empirical experiment would the dispersion of the actual hits be reduced to zero. The law is called "causal" because it is theoretically possible to control simultaneously all the physical conditions and to compute mathematically an exact answer.

In the case of the rolling die, the Newtonian laws of motion are still in effect, but the ability to predict the outcome of a single roll is not present. For one roll, or even a few rolls, the pattern of the outcomes is not predictable. Thus, it is necessary to fall back on the frequency with which different values occur in a great many trials to compute the probability that a one will occur in any one trial. The prediction in this case is based on a "statistical law."

In a sense there are statistical laws that allow for controlled conditions in such a manner that they permit the expression of a causal law. On the other hand, there are statistical laws, such as the laws for rolling a die, in which the conditions cannot be controlled. These laws must continue to be expressed as statistical laws. It is with the latter group that this book is concerned.

Populations and Samples

Another distinction that must be made early in a preliminary discussion of the field of statistics is the difference between a population and a sample. A *population* or a *universe* (the terms are used synonymously) can be defined as each and every member of some group. The group that constitutes a population can be determined in many ways. For example, the population of City A may be defined as all those persons living within the city limits. In this case a political boundary is used to designate the group. The population of sophomore students at College B may be defined as all the students having 30 or more but fewer than 60 hours of college course credit and registered in College B during the current semester. The definition of the population should be clear and complete.

Any descriptive measure of the characteristics of a population is called a *parameter*. For example, the total number of persons living in City A or the average age of the sophomores registered in College B is a parameter.

Once the population has been defined, a *sample* can be described as some of the members of the population. Some of the residents of City A or some of the sophomores at College B would constitute a sample. There are, of course, many possible samples of a given size that might be drawn from any population of any size.

As has already been mentioned, a *statistic* is a descriptive measure of some characteristic of a sample. For example, the average age of a sample of sophomores from College B is a statistic.

Much of the work in inductive statistics is concerned with the problems involved in using a statistic to estimate a parameter.

VARIABLES

A basic building block with which the statistician deals is the variable. Statistical data are the result of successive observations of some characteristic of a group. The characteristic being observed is the *variable*. The observations, which are recorded as the corresponding magnitudes or numbers, are the *values of the variable*.

A variable in which the several possible magnitudes differ by clearly defined steps is called a *discrete variable*. A discrete variable can assume only certain values, often integers, and no intermediate values. For example, the number of rooms in a house or the number of automobiles sold during the month of August is a discrete variable.

When the values of a variable are obtained by measuring from a continuous scale, the variable is said to be a *continuous variable*. Such measures as weight, time, distance, or volume involve measurement on a continuous scale. For any two measurements, no matter how close together, a third measurement can always be found that lies between the first two if a more precise measurement is taken. A continuous variable always has an infinite number of values that need not be whole numbers but which may be carried to as many decimal places as the accuracy of the measurement will justify.

A variable that is derived as a ratio of two other variables, such as income per capita, miles per hour, or units per day of production, is called a *derived variable*.

The individual values of a variable in a population are designated by a capital X, and the number of observations is designated by a capital N. The N observations may be designated as

$$X_1\ X_2,\ .\ .\ .\ ,\ X_N.$$

The values of a variable in a sample are designated by a lowercase x, and the number of observations is designated by a lowercase n. The n observations may be designated as