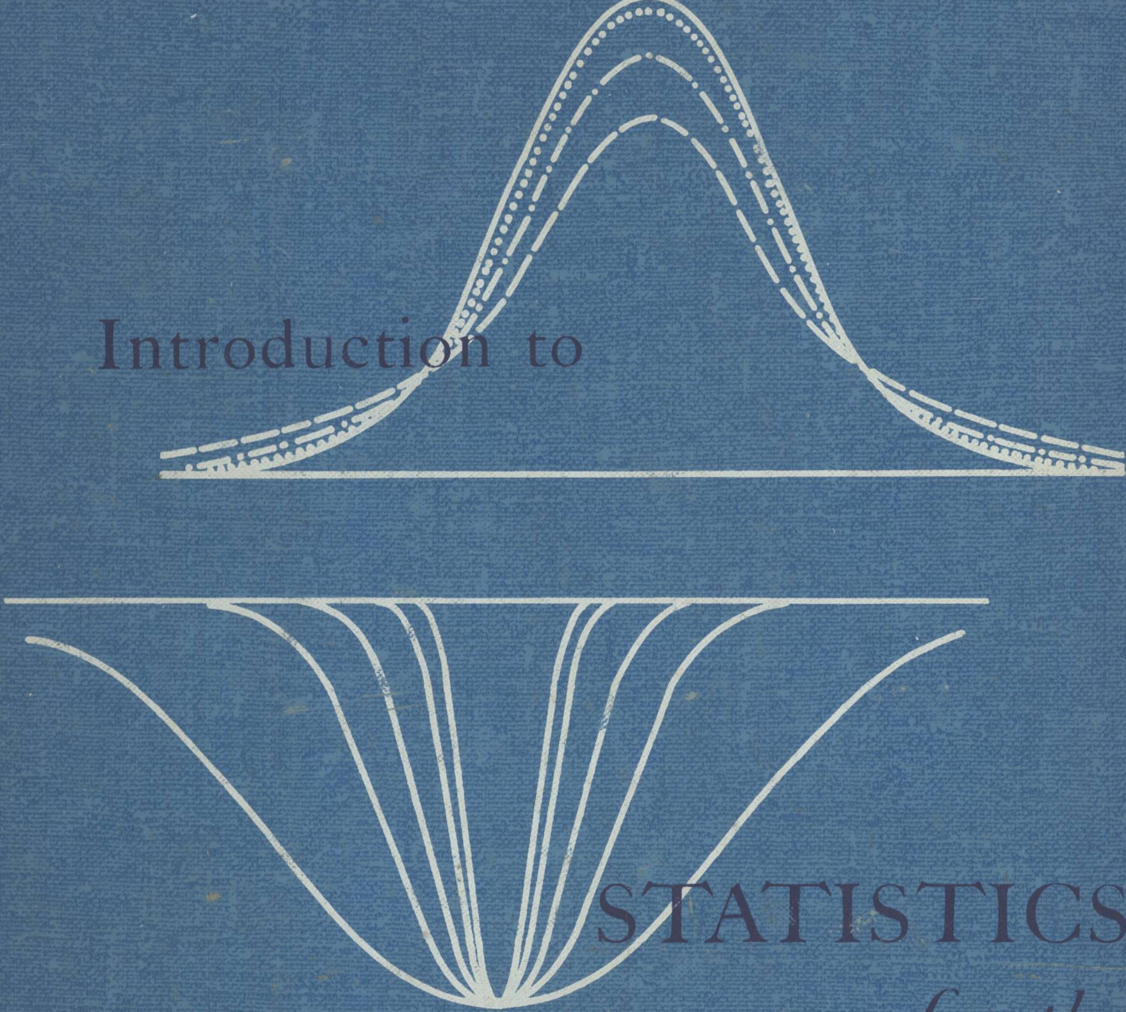

Introduction to

STATISTICS *for the* BEHAVIORAL SCIENCES

HARDYCK *and* PETRINOVICH

INTRODUCTION TO

STATISTICS FOR THE BEHAVIORAL SCIENCES

SECOND EDITION

CURTIS D. HARDYCK
University of

LEWIS F. PETRINOVICH
University of California, Riverside

W. B. Saunders Company: West Washington Square
Philadelphia, PA 19105

12 Dyott Street
London, WC1A 1DB

833 Oxford Street
Toronto, Ontario M8Z 5T9, Canada

Library of Congress Cataloging in Publication Data

Hardyck, Curtis D

Introduction to statistics for the behavioral sciences.

(Saunders books in psychology)

Includes index.

1. Psychometrics. I. Petrinovich, Lewis F., joint author. II. Title.

BF39.H265 1975 150'.1'82 74-21013

ISBN 0-7216-4521-6

Introduction to Statistics for the Behavioral Sciences ISBN 0-7216-4521-6

Last digit is the print number: 9 8 7 6 5 4 3 2 1

Saunders Books in Psychology

LEWIS F. PETRINOVICH and ROBERT D. SINGER, *Consulting Editors*

Gorsuch: *Factor Analysis*

Johnson: *Aggression in Man and Animals*

Kaufman: *Introduction to the Study of Human Behavior*

L'Abate and Curtis: *Teaching the Exceptional Child*

Sattler: *Assessment of Children's Intelligence*

Singer and Singer: *Psychological Development in Children*

Skinner: *Neuroscience—A Laboratory Manual*

Stotland and Canon: *Social Psychology—A Cognitive Approach*

Wallace: *Psychology—A Social Science*

Hardyck and Petrinovich: *Understanding Research in the Social Sciences*

1976

W. B. SAUNDERS COMPANY

Philadelphia, London, Toronto

PREFACE TO THE SECOND EDITION

The mortality rate for introductory statistics texts seems to be rather high. Of the astonishingly large number of texts introduced, relatively few seem to survive to a second edition. We are grateful to the many individuals who used the first edition and found it suitable (despite its many faults) and we hope that this edition will be as well received.

We have attempted to correct some of the problems of the first edition, first among which was a large number of typographical errors. There are few things more infuriating to either instructors or students than books containing trivial but misleading errors. We have made every effort to insure that this edition is as free of errors as possible, and would like to express our gratitude to the many who wrote to us and pointed out some of the errors in the earlier edition.

Several sections have been changed in this edition. The treatment of analysis of variance has been revised extensively (and, in our view, improved), and we have included a discussion of complex analysis of variance. The chapters on regression and correlation have been redone completely and we believe they also have been improved. In addition, the relationship between regression and analysis of variance is discussed. A chapter on multivariate statistical methods has been added to the book. We have not included computational methods, but have provided a preview of the techniques and methods of multivariate statistics. We are firmly of the opinion that multivariate methods will come to dominate behavioral science analyses in the future, and that early exposure to the concepts involved in these methods will do no harm but may prove to be quite adequate preparation for advanced courses.

We would like to thank Professor Roy Goldman for reviewing

and commenting on our chapter on multivariate methods, and Edith Lavin, who prepared the final manuscript copy and was of considerable help in our campaign to remove errors.

We are indebted to the Literary Executor of the late Sir Ronald Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Oliver and Boyd, Ltd., Edinburgh, for permission to reprint Table A-9 from their book, *Statistical Tables for Biological, Agricultural and Medical Research.*

CURTIS HARDYCK
LEWIS PETRINOVICH

CONTENTS

CHAPTER ONE

INTRODUCTION

It is hard to underestimate the extent to which statistics influence our lives. National political figures worry about public opinion statistics to such an extent that full time pollsters are retained to monitor constantly changes in public opinion. A drop (or rise) in the popularity of the President of the United States is a front page item for most newspapers. Stock market speculators anxiously scan the "Dow-Jones averages" for clues on how to invest. A large manufacturer of processed foods reads a statistical report on the declining birth rate in the United States and decides not to expand a factory preparing canned baby food. A gambler tries out a new system for winning at roulette and discovers (after spending a great deal of money) that his system does not work. Statisticians employed by gambling casinos anticipated his system long ago and adjusted the casino rules to insure that he had, at best, a million to one chance of beating the casino.

Almost everyone living in modern society is surrounded by statistics and has his life affected by decisions that are made on the basis of statistical findings. The advantages gained from at least an elementary understanding of how these decisions are made are justification enough for attempting to understand statistical concepts.

An ability to think in statistical terms and a knowledge of statistical concepts are perhaps the best safeguards a consumer can have against being tricked into making foolish decisions on the basis of advertising slogans calculated to impress (or mislead) a prospective buyer. An enjoyable presentation of the uses and abuses of statistics is given by Darrell Huff in his book *How to Lie with Statistics*.* Huff's book is

---

*Huff, Darrell: *How to Lie with Statistics*. New York, Norton, 1954.

recommended to anyone who is curious about the extent to which misuses of statistics exist in our everyday lives.*

Uses of Statistics in the Behavioral Sciences

Anyone preparing for a career in the behavioral sciences must understand statistical methods in order to comprehend the basic research literature in his field. Almost all conclusions concerning experimental data or survey results are based on statistical tests. If one is to understand how these conclusions are reached and is to be able to critically evaluate them, then it is essential that the logical steps involved in arriving at these conclusions be understood. Individuals in applied fields must have a knowledge of statistical principles to enable them to understand and to evaluate the research reports relevant to their interests. The research worker must have a thorough knowledge of statistics if he is to design experiments and to evaluate the results of such experiments.

In the behavioral sciences data are gathered to resolve questions regarding substantive issues. The experimenter who gathers such data must begin with a careful consideration of the types of statistical methods he will apply to the data, in order that the measurements he makes be in the proper form to permit the intended analysis. On the basis of these statistical analyses the experimenter usually decides whether or not his observed outcomes could reasonably be expected on the basis of chance variability. If he decides that the differences, or the relationships, may have been due to chance he will conclude, essentially, that his experimental manipulations did not result in an effect that could be detected with his procedures and measuring instruments.

Methods of Teaching Statistics

There have been two primary approaches to teaching statistics. One approach is to present the material in a cook book form. The formulas and the statistics are interpreted for the student much as a recipe in a cook book is interpreted. The student is shown how to

*Almost any evening spent watching television will produce examples of the misuse of statistics. Note the next toothpaste, or aspirin, or whatever, advertisement that uses a phrase such as "Out of all dentists surveyed, four out of five recommended brushing with Bilge toothpaste." The number of dentists actually surveyed is not mentioned. The use of the percentage of people in favor of something is a similar gimmick, and equally worthless unless you know the numbers of people entering into the percentages.

take various ingredients (defined by a set of symbols) and how to mix them into a recipe (the formula). If the recipe is followed closely enough, a successful dish (calculation) will be the outcome. This approach has the advantage that almost any student can master at least the basic computational rules. However, even the good student prepared by this teaching method can only plug a set of numbers into formulas and mechanically solve those problems with which he has had some prior experience and instruction. The ability to apply the logic of statistical techniques to problems not previously encountered tends to be lacking in students taught by this approach.

The other conventional approach produces a much better trained individual and certainly improves the student's level of comprehension of statistical methods. At the same time, however, it demands much more of him. Here, the underlying logic of statistical methods is presented by means of mathematical proofs, algebraic demonstrations, and extended abstract accounts. The student is usually assumed to have the mathematical background sufficient to understand such material (a situation that greatly limits the audience to which such a text can be addressed). Alternatively, an attempt can be made to present both the underlying logic and the mathematical bases of statistics in such a way that only the simplest knowledge of algebra is required. In our opinion this is a difficult, if not impossible, task.

We believe that still another approach can be fruitful. Modern computing methods and the availability of digital computers have made it possible to demonstrate by concrete examples what previously could be shown only mathematically. In this text we will demonstrate some basic principles using computer generated examples to demonstrate some principles of modern statistical theory. This approach results in a minimum of mathematical explanation and algebraic derivation, but allows us to present the logic of statistical inference in an easily comprehensible manner.

We are emphasizing the basic logic because an intelligent use of statistical procedures demands more than an inflexible application of a fixed set of procedures. It is necessary to understand statistical logic in order to be able to plan and carry out research; failure to do so may result in data being collected that are worthless because of an inability to understand statistical considerations. Also, the type of statistical methods available to us influences the research problem selected and the manner in which it is approached. If a problem is so complex that we can find no way to untangle it for study, we have to move to a simpler type of problem or try to deal with only limited aspects of the more complex problem. Any researcher who is not aware of these considerations about the basic procedures of statistics is severely limited in what he can accomplish. He will be unable to ask certain kinds of questions and will not be able to understand the answers provided by other research workers.

Suggestions to the Student

At this point, we would like to make two suggestions to the student who will be using this text. First, we suggest that you keep in mind that statistical theory is a well integrated set of principles and postulates. At times you will be asked to accept certain statements without proof until explanation can proceed to a point where the logic of the statements can be demonstrated. We have adopted this approach rather than attempt to present the entire basis of statistical theory in the initial stages. We will warn you of situations where you are asked to accept something temporarily without proof. Similarly, you will be warned when the relationships among certain concepts might not be clear. You should understand the basic principles of statistical theory and the logic involved in making inferences by the time you finish this book. If you are unwilling to take some things on faith and wish to know the nature of the relationships that are presented to you, the necessary references to obtain a more complete understanding are provided in the text, but this will demand a great deal of effort on your part.

The second consideration involves the differentiation of statistics and pure mathematics. If your experience is similar to that of the majority of students taking a statistics course for the first time, your expectations are probably determined by the courses you have taken in mathematics—whether high school algebra, college algebra, or introductory mathematics. In short, you may expect a simple, exact, and straightforward solution to a problem once you understand the mechanics and rationale by which the solution can be obtained. Unfortunately, statistics is not so simple. Statistical methods require the application of mathematical techniques to problems in the real world (such as the interpretation of data collected in a scientific experiment), and require both the elements of mathematical procedure and that of human judgment. Consequently, you must accept the idea that, after finishing a set of statistical calculations according to a set of prescribed rules, you must exercise a degree of subjective judgment and make reference to arbitrary conventions agreed upon by statisticians in order to interpret your results. This sequence of events will occur over and over again in the material you will encounter in this course. We cannot emphasize too strongly that statistics is a branch of mathematics applied to problems of the real world. If you limit your statistical work to events in which you can be guided by rules or equations, you will find that you are limiting yourself to a clerical or, at most, to a primitive descriptive function.

A primary use of statistical methods in the behavioral sciences is to evaluate the results of experiments and to provide guidelines an experimenter can use to interpret the results. Accordingly, a heavy emphasis is placed on experimentation as well as on the process of interpreting statistics and making inferences on the basis of these statistics. It should be emphasized, however, that statistical methods are of even greater importance when we consider data that have been gathered in natural situations that do not involve the use of experimental methods. Much of what we know about human behavior has been learned from the study of people's attitudes and from what they do in natural situations as opposed to the experimental laboratory. The problems involved in interpreting data obtained in such complex natural situations demand an even greater statistical sophistication than those faced by the laboratory scientist.

CHAPTER TWO

FUNDAMENTAL CONCEPTS

In this chapter some of the basic concepts and assumptions of statistical theory will be presented and briefly developed. The topics to be presented follow a logical sequence, although the student must accept some basic premises on faith rather than learning the underlying justification at this time. The material to be presented in this chapter is basic to the understanding of statistical methods and will be referred to time and time again. Consequently, effort spent in understanding the material in this chapter will be well repaid in terms of time and effort saved in later chapters. Three fundamental concepts will be introduced: *random sampling*, the *normal curve*, and *probability*. In addition, some aspects of their interrelationships will be presented.

Random Sampling
Normal Curve
Probability

Measurements and Observations

Assume that we have a set of *observations* represented by numbers. These *observations* (the term we will use throughout the text to refer to any numerical measurement or score) could represent a variety of things: the heights or weights of people of a certain age and sex; the incomes earned by people in a given profession; or the scores on an examination such as the National Merit Scholarship Examination. Suppose that we have 1000 observations. If we were to take these 1000 numbers and place them on a graph with the values of the numbers arranged from the lowest to the highest along the horizontal axis and the frequency with which they occur along the vertical axis, we would construct a graph such as is shown in Figure 2–1. As a matter of fact, if these numbers represented any of a variety of observations in the real world, we would expect them to be distributed in a shape somewhat similar to the observations in Figure 2–1.

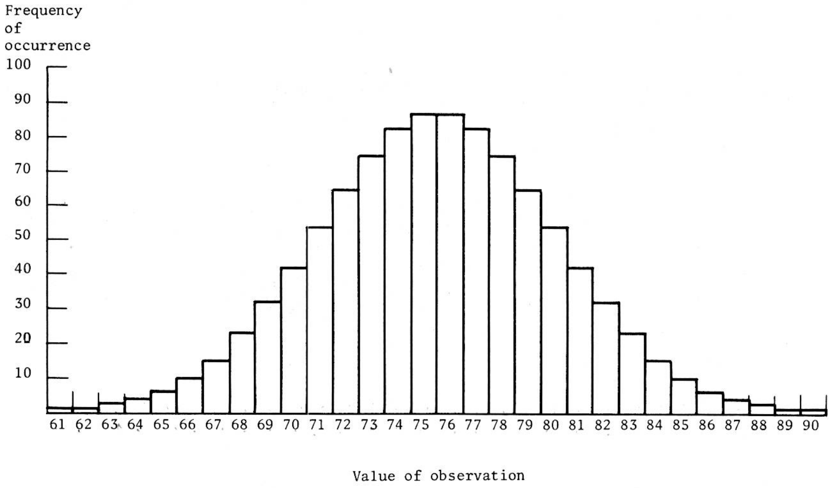

Figure 2–1. Normal distribution of 1000 observations.

There may be some skepticism as to whether real observations ever look anything like those in Figure 2–1. Figure 2–2 is a graph of scores on one of the standard tests of intelligence, the Stanford-Binet. Figure 2–3 is a graph of the weight gain of 100 pigs over a period of 20 days. These rather different examples are presented to emphasize

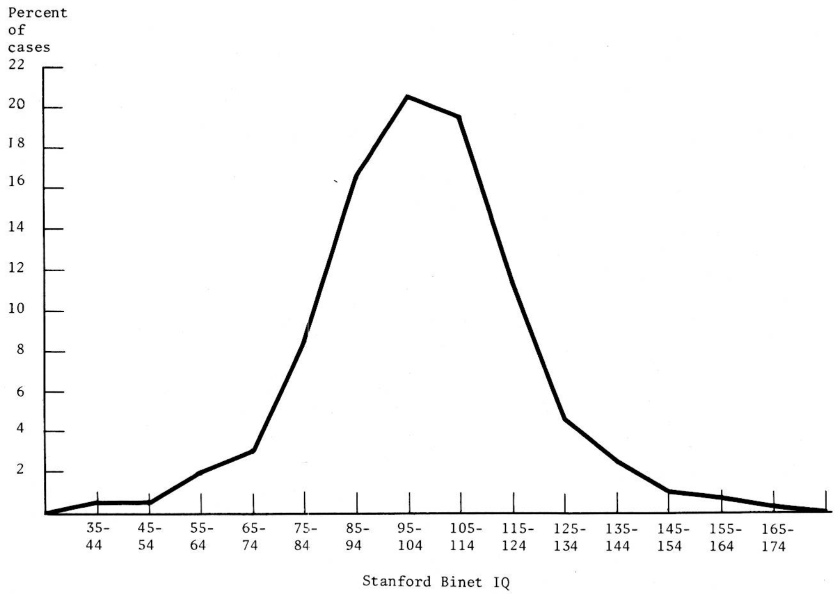

Figure 2–2. Stanford-Binet IQ's of 2904 unselected children between the ages of two and eighteen. (From *Measuring Intelligence* by L. M. Terman and M. A. Merrill. Copyright 1937, by Houghton Mifflin Company and reprinted with their permission.)

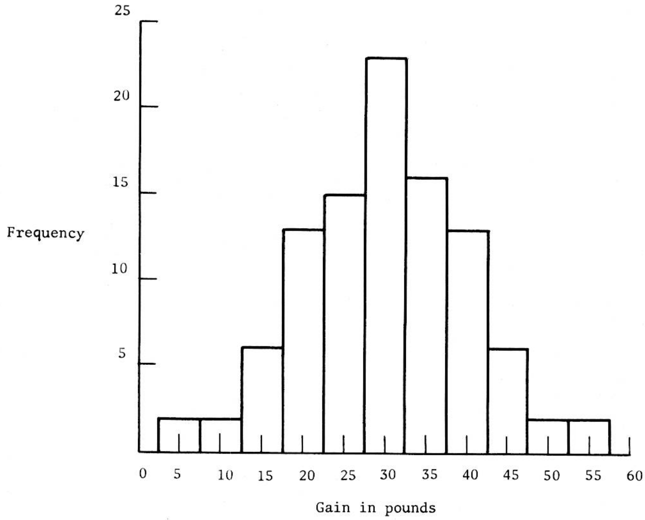

Figure 2–3. Graphical representation of gain in weight of 100 pigs over a 20-day period. (Reproduced by permission from *Statistical Methods*, 5th edition, by George W. Snedecor, © 1956 by The Iowa State University Press.)

the point that a wide variety of observations in the real world may be very similar to the demonstration set of 1000 numbers.

GRAPHICAL METHODS

The conventional method of graphing observations is to represent the values of the observations along the horizontal axis (also called the *x* axis, or *abscissa*), and the frequency with which observations occur along the vertical axis (also called the *y* axis, or *ordinate*). Figure 2–1, 2–2, and 2–3 represent three possible variations in graphical representation.

Figures 2–1 and 2–3 are *histograms* (graphical representations in which the height of a bar is used to represent the frequency with which a given value occurs). Figure 2–2 is a *frequency polygon*. In this type of graphical representation points are placed to indicate the frequency of occurrence of given values and these points are then connected by lines. Such a method provides a continuous line enclosing the observations. The basic principle underlying all methods is the same—a change in area or size is used to represent the frequency with which a given value occurs.

Note that Figures 2–1, 2–2, and 2–3 also differ in the arrangement of the values along the horizontal axis. These units are known as *class intervals*. In Figure 2–1, the size of the class interval is 1, for Figure 2–2, the class interval is 10, and for Figure 2–3, the class interval is 5.

What determines the number of class intervals? This is largely arbitrary, since the nature of the scores to be grouped must be taken into consideration. Too fine a grouping destroys the purpose of grouping scores and too coarse a grouping will tend to lose information. In general, 10 to 15 intervals are usually satisfactory. As far as the

In general 10 to 15 intervals are usually satisfactory