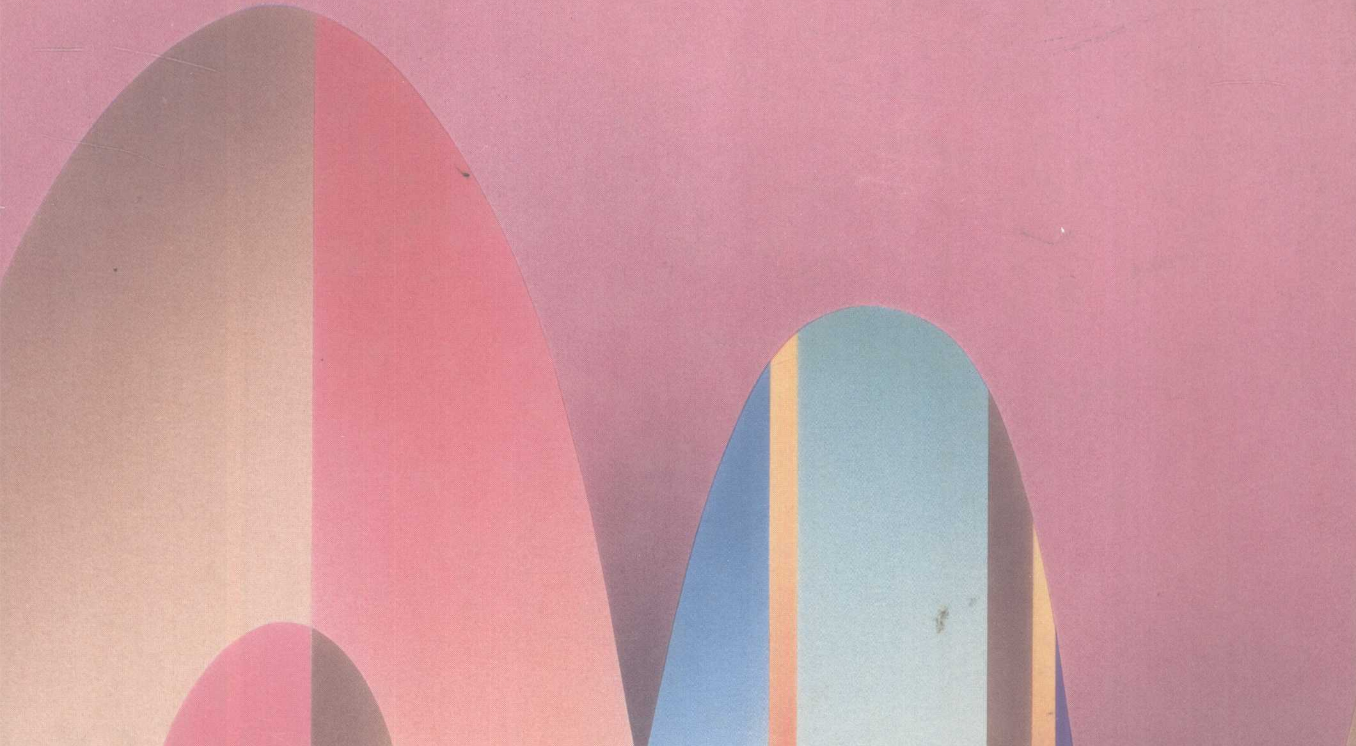

FUNDAMENTAL STATISTICS FOR BEHAVIORAL SCIENCES

ROBERT B. McCALL

FIFTH EDITION

FUNDAMENTAL STATISTICS FOR BEHAVIORAL SCIENCES

FIFTH EDITION

Robert B. McCall
University of Pittsburgh

Under the General Editorship of
Jerome Kagan
Harvard University

Harcourt Brace Jovanovich College Publishers
Fort Worth Philadelphia San Diego
New York Orlando Austin San Antonio
Toronto Montreal London Sydney Tokyo

COPYRIGHTS AND ACKNOWLEDGMENTS

LONGMAN GROUP Tables B, C, G, and K are taken from Tables III, VII, IV, and XXXIII of Fisher and Yates: *Statistical Tables for Biological, Agricultural and Medical Research* published by Longman Group Ltd., London (previously published by Oliver and Boyd Ltd., Edinburgh) and by permission of the authors and publishers.

ISBN: 0-15-529476-8

Library of Congress Catalog Card Number: 89-85302

Printed in the United States of America

PREFACE

A veteran scholar of statistics once remarked that there are two ways to teach statistics—accurately or understandably. Disproving that "either-or" statement is the challenge that has guided my writing of this book.

As a textbook for the first course in applied statistics, *Fundamental Statistics for Behavioral Sciences* is used primarily by students majoring in psychology, education, and other behavioral sciences. In writing for this audience, my earliest and most basic decision was to emphasize the purpose, rationale, and application of important statistical concepts over rote memorization and the mechanical application of formulas. I believe that students at the introductory level, whether or not they plan to take advanced courses in statistics, are served better by a book that fosters an understanding of statistical logic than by one that stresses mechanics.

When understanding is emphasized, elementary statistics is neither dull nor mathematically difficult. *Fundamental Statistics for Behaviorial Sciences* does not require much background in mathematics. The student need be familiar only with the thinking patterns learned in high school algebra and geometry; all relevant terms and operations are reviewed in Appendix 1. To be sure, the book contains many computations and problems to solve, but most statistical formulas rely heavily on simple arithmetic—addition, subtraction, multiplication, division, and the taking of square roots—and can be worked out quickly with the aid of a hand calculator. In addition, I have kept the data for computational problems simple so that the emphasis remains on the rationale and outcome of techniques instead of on calculation for its own sake.

Understanding a statistical concept or strategy can be a formidable mental exercise, but once it is done, details that torment mechanical learners become obvious deductions. The goal of understanding concepts is not hard to reach if students also understand as much of the mathematical reasoning as is within their grasp. Whenever possible, I have explained in mathematically simple terms the logic that undergirds the basic concepts and techniques, although a few items require advanced mathematics and must therefore be taken on faith at this level. Beginning in Chapter 3, optional tables that show full algebraic derivations and proofs supplement the text explanations. These tables can be omitted without loss of continuity.

For the sake of students, and contrary to the traditional practice of mathematical writers, I have included and explained every step in proofs, however "obvious." I have also avoided excessive use of symbols, since symbols require an extra mental translation and thus often confuse students. Deviation scores $(X_i - \overline{X})$, for instance, are *not* abbreviated by x. Further, each new symbol is carefully introduced and is frequently accompanied by its verbal equivalent.

Even so, many students are intimidated by a course in statistics, and numerous factors are frequently blamed—math phobia, the symbolism, and so forth. Actually, I believe the main problem is that students must not expect to study this course in the same way they study so-called "reading" courses. Simply reading statistics, even if you understand it, does not guarantee that you can work statistical problems. Most students must practice problem solving. Also, when the text asks the student to look at a figure or table or to follow an algebraic manipulation, most students really need to do it—and not move on until they understand the point being made. A little patience and discipline can go a long way toward mastering this course.

Anyone who has analyzed his or her own data knows the anticipation that accompanies the final calculation of the r, t, or F hidden in a mass of numbers that took months to gather. Students, too—even though it is not their own data they are analyzing—can experience the excitement of seeing meaning emerge as they manipulate an apparently patternless collection of numbers. Yet they sometimes fail to see the fascination of statistical analysis because it is presented more as a numbers game played in a vacuum than as a crucial part of the scientific investigation of real phenomena.

For this reason, many end-of-chapter exercises and examples in the text are drawn from actual studies (modified for numerical simplicity). For example, the distinction between a correlation and a difference between means is demonstrated through findings that related the IQs of adopted children to those of their biological and adoptive mothers, and Freedman's work on the feeding behavior of dogs reared under indulged and disciplined conditions is used to illustrate interaction effects in the two-factor analysis of variance. Although most of these studies were performed by psychologists, many of them concern developmental and educational issues of interest to future teachers and school administrators.

As another means of showing that numbers can have real applications, I have tried to give students a feel for the behavior of a statistic by providing several data sets that display obvious contrasts. Before calculating an analysis of variance in Chapter 10, for example, I present a set of random numbers, introduce a main effect, and then add an interaction treatment effect. Many exercises ask students to alter a given set of scores in some way and to observe how the change affects the value of a statistic.

This Fifth Edition includes a number of changes, most of them suggested by users of previous editions. The most obvious alteration is the resequencing of some chapters. Parts 1 and 2 contain the basic topics, while Part 3 now includes more advanced material. In addition, certain sections throughout the text have been rewritten to be both clearer and briefer, more exercises have been provided in the earlier chapters, and a few new figures have been added to aid some explanations. Finally, some guidance is offered to students who have access to hand calculators and computers with statistical programs, but this is limited by the great variety of these computational aids.

In the Study Guide, the main section of each chapter is a semiprogrammed unit that reviews the basic terms, concepts, and computational routines described in the corresponding textbook chapter. The presentation is stripped of excess detail, and its tone is very concrete and applied. One feature that students have found particularly valuable has been retained: guided computational examples—step-by-step outlines that show how to organize the operations required in complicated calculations and that clarify the logical and computational details.

Individual instructors emphasize different aspects of elementary statistics, and many must select a subset of chapters that can be covered in the available time. The organization of the text is straightforward. Part 1 (Chapters 1–6) presents descriptive statistics, including central tendency, variability, relative position, regression, and correlation. Part 2 (Chapters 7–10) deals with elementary inferential statistics, including sampling distributions, the logic of hypothesis testing, elementary parametric tests, and simple analysis of variance. For most courses, these are the core chapters. Part 3 contains more advanced topics that instructors can select to supplement the core material. Most commonly, the chapter on nonparametrics is assigned, but those instructors teaching a more intense course may also add the chapters on research design, probability, and two-factor analysis of variance.

In addition to the colleagues and friends who contributed their advice to earlier editions, I want to thank Mark Appelbaum of the Department of Psychology at Vanderbilt University. He was my chief technical consultant, helping to blend my personal style with the demands of formal theory. I am indebted to the many whose commitment to the course and to the book led them to communicate directly with me over the years; their comments, criticisms, and suggestions stimulated many of the improvements in this new edition.

The revision of the Fifth Edition in particular has been guided by the suggestions of Willard D. Larkin, University of Maryland, and the following reviewers: Janice F. Adams, Middle Tennessee State University; Jack Haynes, University of North Texas; Bruce G. Rogers, University of Northern Iowa; and James Wardop, University of Illinois at Urbana-Champaign. Further, the job of producing a revision was made considerably easier by the assistance of Michael Carriger, who checked the exercises in the text and the Study Guide, and Cari Lakatosh, Kimberly Mason, Peggy Maloney, and Carol Kenderes, who prepared the manuscript and helped with the proofs.

I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group Ltd., London, for permission to reprint Tables III, IV, VIII, and part of XXXIII from their book *Statistical Tables for Biological, Agricultural and Medical Research* (Sixth Edition, 1974).

Special thanks go to my wife, Rozanne, for her encouragement and for the absence of complaints while she was temporarily widowed for this cause.

Robert B. McCall

CONTENTS

PART 1

DESCRIPTIVE STATISTICS

One use of statistical methods is to organize, summarize, and describe quantitative information. Such techniques are called **descriptive statistics**. Baseball batting averages, the rate of inflation, and the degree of relationship between heredity and IQ test performance are three examples. This part of the text presents some elementary descriptive statistics.

CHAPTER 1

THE STUDY OF STATISTICS

THE NEWS IS FULL OF STATISTICS: "Approximately one-fourth of the infants born in the country had unmarried mothers; 10% were born to teenagers. . . . Last month inflation rose a seasonally adjusted .3% to an annual rate of 4.5%. . . . Pitcher Dwight Gooden's ERA is now 2.95. . . . The probability of rain today is 60%."

Such quantities are "statistics," and an elementary knowledge of them is useful to everyone. But the academic study of statistics is much more than these common percentages, rates, averages, or probabilities, and a detailed knowledge of this field is essential to anyone who wants to read or, especially, to conduct research in psychology, education, sociology, economics, or any of many other subjects.

As a field of study, statistics consists of a set of procedures for organizing, describing and interpreting measurements and for drawing conclusions and making inferences about what is generally true for an entire group when only a few members of the group are actually measured. This chapter explains why the study of statistics is important, discusses the purpose of statistics, and presents the kinds of measurements typically made in the social and behavioral sciences.

WHY STUDY STATISTICS?

A knowledge of statistics is useful for two main purposes: to help understand the common use of statistics in the news, in our jobs, and elsewhere in our daily lives, and to be able to read and understand scientific articles and conduct research in the social and behavioral sciences.

Everyday Use

All of us have a basic understanding of simple percentages, rates, and averages. For example, if we read that the average household income in the United States in 1985 was $29,066, we assume that someone added up the incomes for all the households in the country and divided that total by the number of such families to give the average family income of $29,066.

But many more statistical quantities exist than averages, for example. Suppose the same newspaper article reports that the *median* household income was only $23,618. What is the *median*? Why is it different than the average? Which value, the average or the mean, is the best index of typical family income? (See Chapter 3.)

Sometimes two simple statistics seem to lead to opposite conclusions. For example, you might read in a magazine that the IQs of adopted children are more closely correlated with those of their biological parents than with those of the parents who reared them. This seems to suggest that heredity plays a more important role in the development of IQ than environment. But the magazine may also report that the same study using the same parents and children found that the IQs of the adoptive children were closer in value to those of their rearing parents than to those of their biological parents, a fact that seems to suggest that environment is more important than heredity. How can both these facts be true, and what do they say about the contributions of heredity and environment to IQ? (See Chapter 6.)

Or you may wake up one morning and hear on the radio that the probability of rain is 40%, but you look outside and it is already raining. Or you might hear that the probability of rain that afternoon is 70%, so you decide not to use the free tickets you were given to the football game. But, in fact, it doesn't rain at all. In both cases you decide the weather forecaster is incompetent, but actually the forecaster has been correct in predicting the weather and in using the concept of probability. (See Chapters 7 and 12.)

So, even the statistics commonly used in everyday life are often more complicated than they first appear. An introduction to statistics will help you understand and interpret these statistics.

Scientific Use

While statistics are useful to everyone, they are crucial if you want to read a scientific paper, study a science, or conduct scientific research, and this is especially true if the science is behavioral. Undergraduate majors in psychology, for example, are required to take a course in statistics—which may be the only reason you are reading this—and many graduate programs in education, sociology, economics, and other sciences require statistical training. Why?

Behavioral scientists rely heavily on statistics because almost all numerical information, or data, in the social sciences contains variability. **Variability** refers to the fact that the scores or measurement values obtained in a study differ from one another, even when all the subjects in the study are assessed under the same circumstances. If your teacher were to give a test to all the students in your statistics class, you would not all obtain the same score. Similarly, some children diagnosed as hyperactive improve their classroom behavior while others do not after being placed on drug medication or after being put through an intensive behavioral regimen at home and at school. This dissimilarity in scores and outcomes is variability.

Variability exists in behavioral data for at least three reasons. First, the units (usually people) that behavioral scientists study are rarely identical to one another. A chemist or a physicist assumes that each of the units under study (molecules, atoms, electrons, and so forth) is identical in its composition and behavior to every other unit of its kind. But the behavioral scientist cannot assume that all people will respond identically in a given situation. In fact, the behavioral scientist can count on the fact that they will *not* respond in the same way. The first major source of variability, then, is **individual differences** in the behavior of different subjects.

Second, behavioral scientists cannot always measure the attribute or behavior they wish to study as accurately as they would like. Scores on a classroom examination are supposed to be a measure of how much students have learned, but even a good test has flaws that make it an imperfect measure of learning. Perhaps some questions are ambiguous, or maybe one part of the course material is emphasized more on the test than other parts, thereby favoring those students who happened to study more thoroughly the relevant parts. These and other factors that influence scores or performance and that are associated with how the measurements are made constitute **measurement error**. Almost all behavioral measurements contain some measurement error, which contributes to variability.

Third, even a single unit (person, animal) usually will not respond exactly the same way on two different occasions. If you give a person several opportunities to rate the attractiveness of advertising displays or the degree of aggressiveness in filmed episodes of each of 15 nursery-school children, he or she will most likely not assign the same scores on each occasion. This source of variability is called **unreliability**.

Individual differences, measurement error, and unreliability are the major reasons that it is often difficult to draw accurate conclusions from behavioral science data. It is the task of statistics to quantify the variability in a set of measurements; to describe the data for a group of subjects, despite the variability inherent in the measurements; and to derive precisely stated and consistent decisions about the results by quantifying the uncertainty produced by variability.

DESCRIPTIVE AND INFERENTIAL STATISTICS

Statistics is the study of methods for describing and interpreting quantitative information, including techniques for organizing and summarizing data and techniques for making generalizations and inferences from data. The first of these two broad classes of methods is called descriptive statistics, and the second is called inferential statistics.

Descriptive statistics refers to procedures for organizing, summarizing, and describing quantitative information, which is called data. Most people are familiar with some statistics of this sort. The baseball fan is accustomed to checking over a favorite player's batting average; the sales manager relies on charts showing the sales distribution and cost efficiency of an enterprise; the head of a household may consult tables showing the average domestic expenditures of families of comparable size and income; and the actuary possesses charts outlining the life expectancies of people in various professions. These are relatively simple statistical tools for characterizing data, but additional techniques are available to describe such things as the extent to which measured values deviate from one another and the relationship between individual differences in performance of two different kinds. For example, one can describe statistically the degree of relationship between the scores of a group of students on a college entrance examination and their later grades in college.

The second major class of statistics, **inferential statistics**, includes methods for making inferences about a larger group of individuals on the basis of data actually collected on a much smaller group. For example, suppose a researcher wanted to know if administering the medication, Ritalin, helps children who have been diagnosed as having attention-deficit disorder with hyperactivity (ADD/H) to accomplish more of their school work in class. Without medication, such children typically have difficulty paying attention to tasks over even moderate periods of time, and therefore they often do not finish assignments. They also have special difficulty with problems that require them to perform a sequence of steps to obtain the answer. Ritalin presumably slows them down so they can sustain their attention over longer periods, complete more work, and work more accurately.

Suppose, with their parents' permission, 40 third-grade children with the ADD/H diagnosis are randomly divided into two groups. One group is given Ritalin treatment under the care of a physician while the other group is administered a *placebo*, a look-alike drug that actually has no effect on the children. Then, after a period of one month to adjust to the treatment, the number of correctly worked problems assigned in class is counted over a four-week period. The Ritalin group worked an average of 63 problems correctly while the placebo group worked an average of 56. The research question is: Does Ritalin improve school performance, at least under these circumstances?

Notice that the question is, "*Does* Ritalin improve performance?" not "*Did* Ritalin improve performance?" The word *does* is much more general than the word *did*. The results already show that the Ritalin group *did* better than the placebo group, 63 *vs*. 56. The scientific question is whether this difference is likely to happen again if two new groups of children are studied, and then again with two more groups, and again and again until all children with ADD/H are studied. In short, *does* Ritalin have this effect for *all* children? "All children with ADD/H" constitutes the **population** of children with ADD/H, while the specific group of 40 children studied constitutes a **sample** from this population. So the scientific question translates into whether Ritalin has an effect in the population, a conclusion that the scientist must *infer* from the results of the sample.

Of course, inferring what would be true for the population of all ADD/H children on the basis of results from a small sample of 40 represents something of a guess, and we feel some uncertainty about drawing such a conclusion. For example, 63 problems is definitely more than 56, but we know that if we repeated the experiment the results for the two groups would not be precisely the same. Perhaps the averages for the treated and nontreated groups would be 60 *vs*. 58, 67 *vs*. 54, 60 *vs*. 60, or even 59 *vs*. 64—the opposite result—respectively. Obviously, if the observed difference between the two sample groups was very large, say 80 *vs*. 20, we would feel more confident that the Ritalin group would do better sample after sample, that is, in the population. Conversely, if the actual observed difference between the two sample groups was quite small, say 64 *vs*. 62, then we would not feel confident that this result would occur consistently in the population. But most of the time the differences between the groups are not this large or small.

How do we make a decision in the face of this uncertainty? The strategy in science is similar to what you might do to decide whether a coin is fair. First, you tentatively suppose that the coin is indeed fair. If it is, then over countless flips one should expect that half, or 50%, would turn up heads. Second, you cannot flip the coin countless numbers of times, so you take a sample, say 100 flips, to see how close the results are to 50–50 (that is, no difference). If the result is actually 52 *vs*. 48, you might conclude, "That result is quite likely to occur if the coin is fair, so I have no compelling reason to believe the coin is biased." But, if the result is 65 *vs*. 35, then you might conclude, "That result is very unlikely to occur if the coin is fair, so I believe the coin is biased."

The scientist essentially follows the same strategy. First, it is presumed that Ritalin *has no effect*, just as we presumed above that the coin was fair. If this is true, then over countless samples we should expect no difference between the Ritalin and the placebo groups. Second, we take a sample, conduct the experiment, and observe the difference between the Ritalin and the placebo groups, just as we flipped the coin 100 times. If the difference is small and likely to occur if no differences actually exist in the population, then there is no compelling evidence to think Ritalin has an effect. But if the observed difference is large and not likely to occur if no differences actually exist in the population, then the conclusion would be that Ritalin does improve performance in the population.

But how "unlikely" does the difference have to be before the scientist concludes that Ritalin has an effect in the population? This is where statistical procedures, some of which are taught in Chapters 8–14, are used. Such procedures quantify that likelihood—they assign it a number, called **probability**, that ranges from 1.00 (it is certain) to .00 (it is impossible) that the observed result could occur even if no differences actually exist in the population. If the probability is high that such a result could occur, then one has no evidence that Ritalin improves performance. But if the probability is small, then perhaps Ritalin does improve performance.

The probability determined by statistical procedures is almost never 1.00 or .00; it is always a value in between. So behavioral scientists have agreed on