N. M. DOWNIE
R. W. HEATH

BASIC STATISTICAL METHODS

FOURTH EDITION

BASIC STATISTICAL METHODS

Fourth Edition

N. M. Downie

Purdue University

R. W. Heath

Formerly Stanford University

HARPER & ROW, PUBLISHERS

New York Evanston San Francisco London

Sponsoring Editor: George A. Middendorf
Project Editor: Cynthia Hausdorff
Designer: T. F. Funderburk
Production Supervisor: Robert A. Pirrung

BASIC STATISTICAL METHODS, Fourth Edition

Library of Congress Cataloging in Publication Data

Downie, Norville Morgan, 1910-
Basic statistical methods.

Bibliography: p.
1. Statistics. I. Heath, R. W., joint author.
II. Title.
HA29.D68 1974 519.5 73-13299
ISBN 0-06-042731-0

BASIC STATISTICAL METHODS

Preface

This edition, like its three predecessors, is written to meet the need of the beginning student in the social sciences for a short, clear, elementary statistics book. We believe that such a book should treat the computation, interpretation, and application of commonly used statistics. No extensive attempt has been made to derive formulas or to involve statistical theory, since the mathematical background of the typical user of this book precludes effective presentation of these topics.

Like the earlier editions, this book is essentially divided into three parts. The first nine chapters present descriptive statistics, and the next seven are an introduction to statistical inference. The third part consists of two unrelated chapters, first an introduction to test theory and construction, and second, a look at the more frequently used distribution-free statistical tests.

Many portions of the fourth edition have been rewritten, and new material has been added. Altogether, we have attempted to present an up-to-date elementary text for a rapidly developing field.

Problems designed to offer practice in the techniques discussed in each chapter appear throughout the book. The answers appear in Appendix P. At the request of many users of this textbook, the senior author has written a separate study guide for use with the text.

Appendixes A to N contain material that will become more and more useful to the student as he progresses through the book. Appendix A, after the student becomes familiar with it, will save hours of computational time. The other appendixes, B through N, are associated with the statistical concepts and tests introduced in the text.

Many sincere thanks are due to the various authors and publishers who gave us permission to use the material that appears in Appendixes A to N. Special acknowledgments appear at the bottom of each appendix.

We are especially indebted to the late Ronald A. Fisher of Cambridge, to Dr. Frank Yates of Rothamsted, and to Messrs. Oliver and Boyd, Ltd., of Edinburgh, for permission to reprint Appendixes C, D, and F from their book *Statistical Tables for Biological, Agricultural, and Medical Research*.

N. M. Downie
R. W. Heath

Contents

Introduction
Chapter 1

Students often wonder what statistics is and why they should bother to study the subject. For hundreds of years people have been collecting statistics. An early tribal chief, for instance, had so many armed warriors, so many horses, took a certain number of the enemy. Today we have vast quantities of data associated with sports, the stock market, traffic, law enforcement, and hundreds of other human activities. From one point of view then statistics may be considered collections of data associated with human enterprises. In a more limited sense, each individual is a statistic. From the viewpoint of a life insurance company, each of us is a statistic.

Statistics may also be considered to be a method that can be used to analyze data—that is to organize and make sense out of a large amount of material. It is with this manipulation of data that we shall be concerned in this book. Statistical methodology may be looked upon as being of three types—descriptive, correlational, and inferential.

To illustrate the three types suppose that we use the entering freshman class of a large university. Each student's folder will contain the scores of the admissions tests used by the university, such as the college boards, a high school transcript, results of a physical examination, and, perhaps, scores based upon interest and personality inventories. Taken as a whole these folders present a mass of information about the freshmen. To learn about the freshmen all these data must be studied. Let us look at a few things that we can do with them. Suppose that we use only the scores on the verbal and mathematical parts of the Scholastic Aptitude Test. We might summarize all of these scores into two distributions. We might draw graphs that would show the differences in the scores of males and females, the differences among individuals in the several schools of the university, or the differences among students from varying backgrounds. Then we could find an average score on each of the two parts of the test for the whole group and for the different subgroups. If we are concerned with the relationship of individual scores to these averages, we can change these raw scores into another type of a more meaningful nature. Centiles or standard *scores* show how an individual in the group stands in reference to others. All of the foregoing operations are included in what we call *descriptive statistics* because they give us information, or describe, the sample we are studying.

We could also examine the scores on the verbal part of this test and see if they are related to scores on the mathematical part. This is called computing a correlation coefficient. We could also correlate these test scores with the first semester grades made by these freshmen, or with their high school ranks, or with the various scores on the interest and personality inventories. The results of correlational work are useful in making predictions of future behavior. If we know that a relationship exists between two variables, then scores on one may be used to predict scores on the other. In statistics this study of prediction is referred to as regression analysis. The results of correlational analysis are used to study the reliability and validity of educational and psychological tests. *Correlational analysis,* then, is a major part of statistical methodology.

Third we have *inferential statistics.* Usually when samples are studied, the investigator is interested in going beyond the sample and making an inference about the population from which the sample was drawn. Populations are frequently so large that the only way their characteristics will ever be known is through the study of samples drawn systematically from the population. It follows then that from measures of averages and variability based upon samples we make inferences about the size of the same traits in the population. The use of inferential statistics is basic to experimental research in all branches of science.

There are reasons for studying statistics other than knowing how to use the subject in a research task. A knowledge of statistics is basic to the intelligent reading of a research article or a modern text in science. Without the background one would get from a first course in statistics, these accounts of modern science are unintelligible. Statistics is also of use to the informed instructor in building and analyzing tests and in preparing grades. Statistics may also contribute to the general education of a consumer. Modern advertising makes all sorts of claims, often bolstering them with impressive statistics. The intelligent consumer looks critically at these claims and the statistics used to support them.

A BRIEF HISTORY OF STATISTICS

Statistics has a long and venerable history. Perhaps the earliest use of statistics was when an ancient chief counted the number of effective warriors that he had or the number he would need to defeat his enemy, or when he figured how much might judiciously be collected in taxes. In later times, statistics were used to report death rates in the great London plague and in the study of natural resources. These uses of statistics, which encompass a broad field of activity referred to as "state arithmetic," are purely descriptive in nature.

In the seventeenth and eighteenth centuries mathematicians were asked by gamblers to develop principles that would improve the chances of winning at cards and dice. The two most noted mathematicians who became involved in this, the first major study of probability, were Bernoulli and DeMoivre. In the 1730s DeMoivre developed the equation for the normal curve. Important work on probability was conducted in the first two decades of the nineteenth century by two other mathematicians, LaPlace and Gauss. Their work was an application of probability principles to astronomy.

Through the eighteenth century statistics was mathematical, political, and governmental. In the early nineteenth century, a famous Belgian statistician, Quetelet, applied statistics to investigations of social and educational problems. Walker (1929)[1] credits Quetelet with developing statistical theory as a general method of research applicable to any observational science. Beyond any doubt, the individual who had the greatest effect upon the introduction and use of statistics in the social sciences was Francis Galton. In the course of his long life he made notable contributions in the fields of heredity and eugenics, psychology, anthropometry, and statistics. Our present understanding of correlation, the measure of agreement between two variables, is credited to him. The mathematician Pearson collaborated with Galton in later years and was instrumental in developing many of the correlation and regression formulas that are in use today. Among Galton's contributions was the development of centiles or percentiles.

The famous American psychologist James McKeen Cattell studied in Europe in the 1880s and contacted Galton and other European statisticians. On his return to the United States he and his students, including E. L. Thorndike, began to apply statistical methods to psychological and educational problems. The influence of these men was great; in a few years theoretical and applied statistics courses were commonly taught in American universities.

In the twentieth century new techniques and methods were applied to the study of small samples. The major contributions in small-sample theory were made by the late R. A. Fisher, an English statistician. Although most of his methods were developed in an agricultural or biological setting, it was not long before social scientists recognized the utility of Fisher's methods and made use of his ideas. Today statistics is the major methodological tool of the research worker in the social sciences. The student interested in the history of statistics is referred to a brief but thorough article written by Dudycha and Dudycha (1972).

[1] Complete references are given at back of book.

A NOTE TO THE STUDENT

It is probable that some students using the text may have found mathematics unpleasant or difficult in the past. Since this book is concerned with numbers and formulas, they may fear that the subject is complicated and obscure. These students should set their minds at ease. One does not have to be a genius, or fond of mathematics, to learn the methods presented here. We believe that a grasp of the elements of seventh- and eighth-grade arithmetic will suffice. The only further prerequisites are the ability to take square roots and to subtract negative numbers. It is possible to brush up on these with a little practice.

A more important problem for most students is that of precision or accuracy. It has been our experience that most errors come not from using the wrong principle but from carelessness in the simple operations of addition and subtraction.

Statistics is one of those subjects that cumulates. One topic leads to another and the second is built upon the first. The work has to be kept up to date. If knowledge of statistics is built on an incomplete foundation, the whole structure will surely topple. The problems found at the end of each chapter will help to test the stability of the student's progress.

A Review of Fundamentals
Chapter 2

In this chapter we shall be concerned with two topics: a review of simple arithmetic and elementary algebra and a discussion of the fundamental nature of measurement. Many of the errors that occur in statistical computations are not caused by a lack of knowledge of statistics but by mistakes in very simple arithmetic. Hence we shall start with a rapid review of the rules that must be followed if computations are to be correct.

REVIEW OF ARITHMETIC

Decimals

ADDITION AND SUBTRACTION. When adding or subtracting decimals, align the numbers so that the decimal point of each number is directely below that of the number above.

To add 3.094, 235.67, and 45.7, we align the numbers like this:

$$\begin{array}{r} 3.094 \\ 235.67 \\ 45.7 \\ \hline 284.464 \end{array}$$

MULTIPLICATION. In multiplying decimals, the product has as many decimal places as there are in both multiplicand and the multiplier taken together, as shown below:

$$\begin{array}{r} 1.072 \\ \times\ .02 \\ \hline .02144 \end{array} \qquad \begin{array}{r} .00007 \\ \times\ .2 \\ \hline .000014 \end{array} \qquad \begin{array}{r} 1.2 \\ \times\ 1.2 \\ \hline 1.44 \end{array}$$

DIVISION. When two decimals are divided, the number of decimal places in the quotient is equal to the number of decimal places in the dividend minus the number of places in the divisor when there is no remainder.

$$\frac{.012}{.3} = .04 \qquad \frac{2.0648}{.2} = 10.324 \qquad \frac{.008}{8} = .001$$

Fractions

ADDITION AND SUBTRACTION. When two or more fractions are to be added or subtracted, they must first be reduced to fractions having the same or a common denominator, as below:

$$\frac{1}{2} + \frac{1}{3} = \frac{3}{6} + \frac{2}{6} = \frac{5}{6} \qquad 2\frac{3}{4} + \frac{1}{2} = \frac{11}{4} + \frac{2}{4} = \frac{13}{4} = 3\frac{1}{4}$$

$$\frac{3}{8} - \frac{3}{16} = \frac{6}{16} - \frac{3}{16} = \frac{3}{16} \qquad \frac{x}{y} + \frac{a}{b} = \frac{xb}{yb} + \frac{ya}{yb} = \frac{xb + ya}{yb}$$

MULTIPLICATION. To multiply fractions, multiply all the numerators and place this quantity over the product of all of the denominators, as shown:

$$\frac{1}{2} \times \frac{2}{3} = \frac{2}{6} = \frac{1}{3}$$

$$\left(\frac{3}{4}\right)\left(\frac{2}{3}\right)\left(\frac{4}{6}\right)\left(\frac{4}{5}\right) = \frac{96}{360} = \frac{4}{15}$$

When multiplying fractions, considerable time is saved if terms common to both the numerator and the denominator are canceled. This is equivalent to dividing the numerator and denominator of a fraction by the same number. The size of the product is not changed, Let us rework the last example:

$$\frac{\overset{1}{\cancel{3}}}{\underset{1}{\cancel{4}}} \times \frac{\overset{1}{\cancel{2}}}{\underset{1}{\cancel{3}}} \times \frac{\overset{1}{\cancel{4}}}{\underset{3}{\cancel{6}}} \times \frac{4}{5} = \frac{4}{15}$$

First we cancel the first 3 in the numerator with the 3 in the denominator. Next the first 4 in the numerator is canceled by the 4 in the denominator. The 2 in the numerator is divided into the 6 in the denominator, leaving a 3 in that position. Then we have $1 \times 1 \times 1 \times 4$ in the numerator and $1 \times 1 \times 3 \times 5$ in the denominator. The product of all terms in the numerator is equal to 4; the product of those in the denominator is 15. Our answer is again $4/15$.

DIVISION. To divide one fraction by another, invert the fraction which is the divisor and proceed as in multiplication.

$$\frac{3}{4} \div \frac{2}{3} = \frac{3}{4} \times \frac{3}{2} = \frac{9}{8} = 1\frac{1}{8}$$

$$2\frac{3}{4} \div \frac{11}{7} = \frac{11}{4} \times \frac{7}{11} = \frac{7}{4} = 1\frac{3}{4}$$

$$\frac{x}{y} \div \frac{a}{b} = \frac{x}{y} \times \frac{b}{a} = \frac{xb}{ya}$$

Negative Numbers

ADDITION. To add numbers, all of which are negative, add the numbers in the usual fashion and place a minus sign in front of the sum.

$$\begin{array}{r} -6 \\ -8 \\ -12 \\ \hline -26 \end{array}$$

When the signs are mixed and there are only two numbers, that is, one negative and one positive, subtract the smaller from the larger and give the remainder the sign of the larger.

$$\begin{array}{r} -6 \\ 8 \\ \hline 2 \end{array} \qquad \begin{array}{r} -22 \\ 28 \\ \hline 6 \end{array} \qquad \begin{array}{r} 56 \\ -72 \\ \hline -16 \end{array} \qquad \begin{array}{r} 19 \\ -30 \\ \hline -11 \end{array}$$

When adding a series of numbers with different signs, add all the positive numbers and then the negative ones and combine the results as above.

$$\begin{array}{r} -4 \\ -7 \\ 8 \\ 13 \\ -12 \\ -5 \\ \hline 21 \\ -28 \\ \hline -7 \end{array}$$

SUBTRACTION. To subtract a negative number from another number, change its sign and proceed as in addition.

$$\begin{array}{r} 12 \\ -(-8) \\ \hline 20 \end{array} \qquad \begin{array}{r} -22 \\ -(-8) \\ \hline -14 \end{array} \qquad \begin{array}{r} -4.48 \\ -(-8.24) \\ \hline 3.76 \end{array}$$

MULTIPLICATION. When two numbers have the same sign, either positive or negative, the product of the two numbers is positive. When the two numbers have different signs, one positive and one negative, the product of the two numbers is negative.

$$\begin{array}{r}6\\ \times 2\\ \hline 12\end{array}\qquad \begin{array}{r}-6\\ \times(-2)\\ \hline 12\end{array}\qquad \begin{array}{r}6\\ \times(-2)\\ \hline -12\end{array}\qquad \begin{array}{r}-6\\ \times(2)\\ \hline -12\end{array}$$

DIVISION. As in multiplication, when a positive or negative number is divided by a number of the same sign, the quotient is always positive. When the dividend and divisor are of unlike signs, the quotient is always negative.

$$\frac{6}{2}=3 \qquad \frac{-6}{-2}=3 \qquad \frac{6}{-2}=-3 \qquad \frac{-6}{2}=-3$$

Use of Zero

The chief rule to remember when using zero is: When any number is multiplied by zero, the product is zero.

$$2 \times 0 = 0$$

$$(.5)(3.55)(0)(4976) = 0$$

Exponents

We shall use exponents only in a limited way in elementary statistics, but the student should know what an exponent is and what it means. For example, in 2^3 the 3 is the exponent, and it means to multiply $2 \times 2 \times 2$ or to raise 2 to the third power.

$$3^2 = 3 \times 3 = 9$$

$$4^3 = 4 \times 4 \times 4 = 64$$

$$x^4 = (x)(x)(x)(x)$$

Removing Parentheses and Simplifying

Sometimes it is necessary to simplify a complex term. The general rule is to start by performing operations so that the parentheses located on the inside can be removed.

$$\begin{aligned} & [(12 + 4)4] - [(3 + 10) + (6 \times -12)] \\ &= [(16)4] - [(13) + (-72)] \\ &= 64 - (-59) \\ &= 64 + 59 \\ &= 123 \end{aligned}$$

Proportions and Percentages

A proportion, the symbol for which is p, is defined as a part of a whole. If a pie is cut into six equal parts, each slice is a proportion and we can write that p = $\frac{1}{6}$ or .167.

To use another example, suppose that in a given class of 400 students, 40 receive A's as their final grade; 100, B's; 150, C's; 70, D's; and 40, F's. What proportion received each letter grade?

	N		p	*P*
A	40	$\frac{40}{400}$	= .10	10
B	100	$\frac{100}{400}$	= .25	25
C	150	$\frac{150}{400}$	= .375	37.5
D	70	$\frac{70}{400}$	= .175	17.5
F	40	$\frac{40}{400}$	= .10	10
N =	400		1.000	100.0

It should be noted that the sum of the proportions for a given example is always 1 and the maximum value of any single proportion is 1.

A percentage is obtained by multiplying a proportion by 100. The symbol for a percentage is *P*. For our example, the corresponding percentages are shown in the column at the right. It will be noted that the percentages for our data add up to 100.

A word of warning should be given about percentages and proportions. When the number of cases is small, percentages are unstable. That is, a change in one case can cause a relatively large change in the percentage. For example, when there are ten cases, a change in one case causes a change of 10 in terms of percent. It might be desirable to follow a rule that when the number of cases is less than 100 the use of percentages should be avoided. In fairness to the reader of the results of a study, the number of cases on which a percentage is computed should always be reported with the percentage.

A recent article stated that there was an increase of 132 percent in the number of new teachers of Russian between one year and the next, whereas there was only a 16.5 percent increase in the number of new high school teachers of English. It should have been noted in the article that there were 11,966 new teachers of English, but only 65 new teachers of Russian.

Rounding Numbers

In rounding numbers to the nearest whole number or to the nearest decimal place, we proceed as follows:

To the nearest whole number	7.2 = 7
	7.8 = 8
To the nearest tenth	7.17 = 7.2
	7.11 = 7.1
	.09 = .1
To the nearest hundredth	7.177 = 7.18
	.674 = .67
	1.098 = 1.10

The general rule is that if the last digit is less than 5, it is dropped; if the last digit is more than 5, the preceding digit is raised to the next higher digit. The only complication arises when numbers end in 5. There is a general rule for this case. When the digit preceding the 5 is an odd number, this digit is raised to the next higher one; when it is an even number, the 5 is dropped. The following examples illustrate this rule:

8.875 = 8.88
8.05 = 8.0
5.25 = 5.2
66.975 = 66.98

Situations like the following arise:

$$\frac{37}{52}(3)$$

There are two ways of simplifying this. The first is to divide 37 by 52 and then multiply the quotient by 3. Or the numerator, 37, could be multiplied by 3 and the product then divided by 52 or multiplied by the reciprocal of 52. The second method is preferred because only one rounding operation is necessary.

SIGNIFICANT DIGITS. One question that frequently arises in recording numbers is How many digits should we have in our answers? As a general rule the answer should have only one digit more than exists in the raw data. For example, if we have a series of test scores, each of which contains two digits, then ordinarily we would have no more than three digits in the average or mean which we compute from these data. There is nothing to be gained in computing these averages to five or six decimal places. No meaningful accuracy is obtained from these large decimals. As a matter of fact, such large decimals mean nothing when computed on the basis of two-place numbers. A good rule is to have one more significant digit in the answer than was present in the original numbers. Here are some examples of the number of significant digits in a series of numbers.