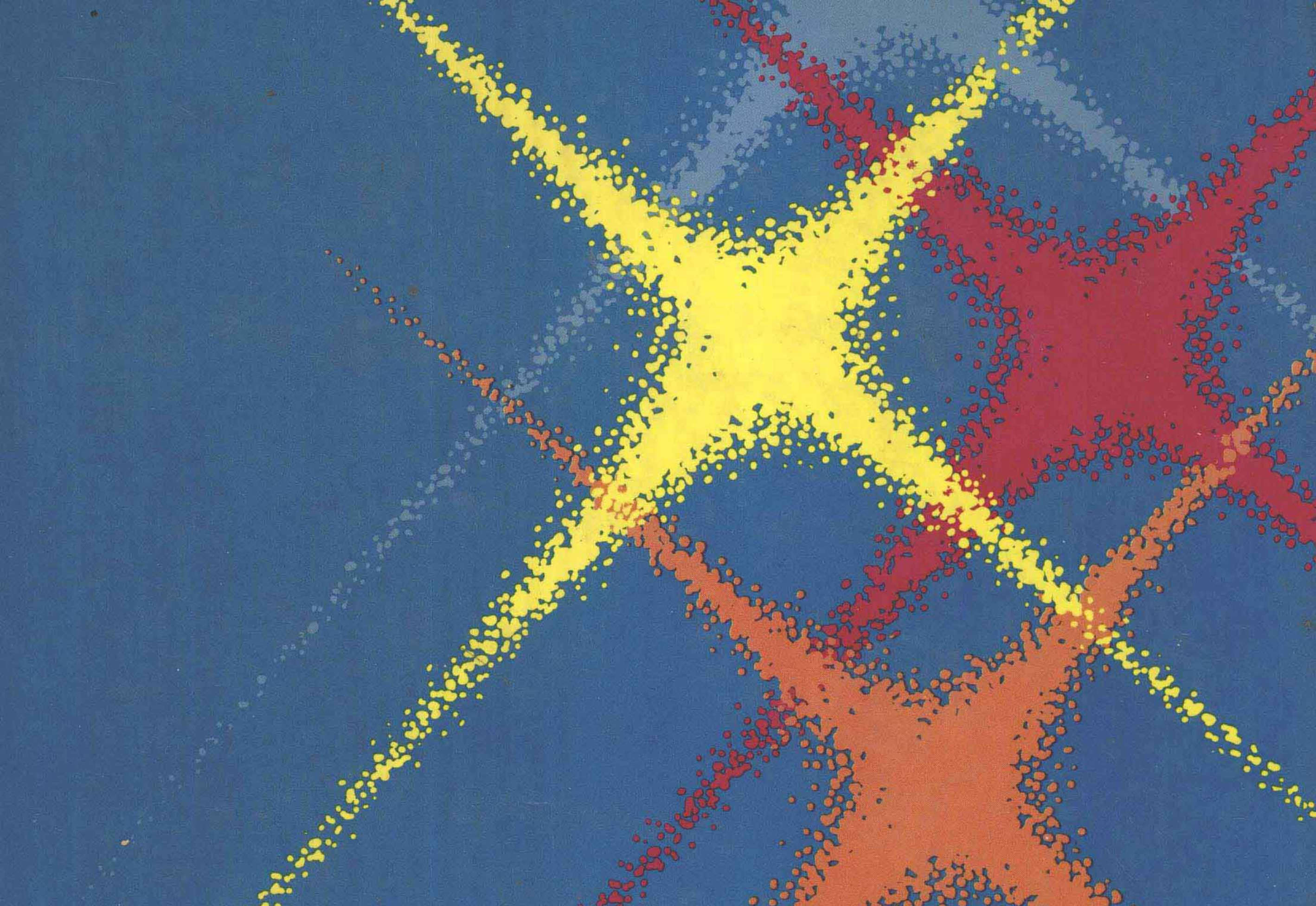

LEVIN / RUBIN

# APPLIED ELEMENTARY STATISTICS

# APPLIED ELEMENTARY STATISTICS

*Library of Congress Cataloging in Publication Data*

LEVIN, RICHARD I
Applied elementary statistics.

Bibliography: p.
Includes index.
1. Statistics. I. Rubin, David S., joint author.
II. Title
QA276.12.L48 1980 519.5 79-20520
ISBN 0-13-040113-7



10 9 8 7 6 5 4 3 2 1

This is a Special Projects book, designed and edited by
Maurine Lewis, *director*
Ray Keating, *manufacturing buyer*
*Page layouts:* Judith A. Matz
*Cover:* Olympia Shahbaz

Prentice-Hall International, Inc., *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall of Canada, Ltd., *Toronto*
Prentice Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall of Southeast Asia Pte. Ltd., *Singapore*
Whitehall Books Limited, *Wellington, New Zealand*

# To the student

This is a short Preface, because most of what is usually said in the Preface is in the first chapter: the goals of the book, its methods, and a sneak preview of the chapters. We've both taught beginning courses in statistics for many years, and we've listened to complaints from students about "imponderable" textbooks. We've answered thousands of questions beginning, "What do the authors mean here?" and we've sympathized with those for whom most textbooks on this subject bring on an anxiety attack.

So we wrote our own statistics book. We started off with the idea that we would try to explain concepts well enough so that the student could lean on the book instead of the teacher. Many of our explanations may be long-winded, and sometimes we use three examples to make a point when perhaps one would do, but neither of us believes it possible to overexplain statistics. So if you'll bear with our surplus examples and copious explanations, we think we can learn statistics together and enjoy doing it!

Our approach uses intuitive explanations, extensions of something you already know, instead of complicated statistical proofs. When something cannot be explained in this commonsense manner, we leave it out.

You will find nearly a thousand text problems and exercises in our book. They cover the widest possible range of topic areas, all the way from abortion to X-rays—nearly one hundred different facets of our lives. If you want to get a quick idea of these different areas, turn to the front endpapers and look at the Index of Applications. Many entries, such as biology, economics, education, government, medicine, and psychology, are fields we encounter or read about every day. But some are newer concerns, such as communication, computers, consumer protection, ecology, energy, environment, housing, nuclear power, population, and solar heat, where a lot of the action is today.

Lots of people made inputs into our book, and a few deserve a special vote of thanks: Carolyn Ezzell, whose creativity concerning problems is everywhere; Harry Gaines, our editor, whose keen insights pick up what students and instructors really want in a book; Lisa Levin, whose indexing abilities are truly amazing, and Maurine Lewis, whose editorial and artistic ideas are always exciting.

We are grateful, too, to our reviewers, each of whom leaves a special imprint on the book: Professor George Casella, Rutgers—The State University; Professor Susan S. Lenker, Central Michigan University; Professor Albert Liberi, Westchester Community College; Professor Norman Locksley, Prince George's Community College; Professor Richard C. Orr, State University of New York at Oswego; Professor J. Roger Teller, Georgetown University, and Professor Stephen B. Vardeman, Purdue University. We are also 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 and IV from their book, *Statistical Tables for Biological, Agricultural, and Medical Research* (6th edition, 1974).

We hope you like what we've put together, and we hope it helps make your study of statistics interesting, nearly painless, and even a little fun!

*Chapel Hill, N.C.*

DICK LEVIN
DAVE RUBIN

# Contents

# 1 Introduction

This book was written for students taking statistics for the first time. A glance at this chapter should convince any concerned citizen and future decision maker that a working knowledge of basic statistics will be quite useful in coping with the complex problems of our society. Your first look will also convince you that this book is dedicated to helping you acquire that knowledge with virtually no previous formal mathematical training and with no pain at all.

Different meanings of statistics depending on use

## 1 DEFINITIONS

The word *statistics* means different things to different people. To a football fan, statistics are the information about rushing yardage, passing yardage, and first downs, given at halftime. To an administrator in the Environmental Protection Agency, statistics may be information about the quantity of pollutants being released into the atmosphere. To a school principal, statistics are information on absenteeism, test scores, and teacher salaries. To a medical researcher investigating the effects of a new drug, statistics are evidence of the success of his research efforts. And to a college student, statistics are the grades made on all the quizzes in a course this semester.

Each of these people is using the word *statistics* correctly, yet each uses it in a slightly different way and for a somewhat different purpose. *Statistics* is a word that can refer to quantitative data (such as wheat yield per acre) or to a field of study (you can, for example, major in statistics).

Today, statistics and statistical analysis are used in nearly every profession. For decision makers, in particular, statistics have become a most valuable tool.

Origin of the word

## 2 HISTORY

The word *statistik* comes from the Italian word *statista* (meaning "statesman"). It was first used by Gottfried Achenwall (1719–1772), a professor at Marlborough and Göttingen. Dr. E. A. W. Zimmerman introduced the word *statistics* into England. Its use was popularized by Sir John Sinclair in his work, *Statistical Account of Scotland 1791–1799.* Long before the eighteenth century, however, people had been recording and using data.

Early government records

Official government statistics are as old as recorded history. The Old Testament contains several accounts of census taking. The governments of ancient Babylonia, Egypt, and Rome gathered detailed records of populations and resources. In the Middle Ages, governments began to register the ownership of land. In A.D. 762, Charlemagne asked for detailed descriptions of church-owned properties. Early in the ninth century, he completed a statistical enumeration of the serfs attached to the land. About 1086, William the Conqueror ordered the writing of the *Domesday Book,* a record of the ownership, extent, and value of the lands of England. This work was England's first statistical abstract.

Because of Henry VII's fear of the plague, England began to register its dead in 1532. About this same time, French law required the clergy to register baptisms, deaths, and marriages. During an outbreak of the plague in the late 1500s, the English government started publishing weekly death statistics. This practice continued, and by 1632 these *Bills of Mortality* listed births and deaths by sex. In 1662, Captain John Graunt used thirty years of these Bills to make predictions about the number of persons who would die from various diseases and the proportion of male and female births that could be expected. Summarized in his work, *Natural and Political Observations . . . Made Upon the Bills of Mortality,* Graunt's study was a pioneer effort in statistical analysis. For his achievement using past records to predict future events, Graunt was made a member of the original Royal Society.

An early prediction from statistics

The history of the development of statistical theory and practice is a lengthy one. We have only begun to list the people who have made significant contributions to this field. Later we will encounter others whose names are now attached to specific laws and methods. Many people have brought to the study of statistics refinements or innovations that, taken together, form the theoretical basis of what we will study in this book.

## 3 SUBDIVISIONS WITHIN STATISTICS

Decision makers apply some statistical technique to virtually every branch of public and private enterprise. These techniques are so diverse that statisticians commonly separate them into two broad categories: *descriptive statistics* and *inferential statistics.* Let's use examples to understand the difference between the two.

Descriptive statistics

Suppose a professor computes an average grade for one history class. Since statistics describe the performance of that one class but do not make a generalization about several classes, we can say that the professor is using *descriptive* statistics. Graphs, tables, and charts that display data so that they are easier to understand are all examples of descriptive statistics.

Inferential statistics

Suppose now that the history professor decides to use the average grade achieved by one history class to estimate the average grade achieved in all ten sections of the same history course. The process of estimating this average grade would be a problem in *inferential* statistics. Statisticians also refer to this category as *statistical inference.* Obviously, any conclusion the professor makes about the ten sections of the course will be based on a generalization that goes far beyond the data for the original history class; and the generalization may not be completely valid, so the professor must state how likely it is to be true. Similarly, statistical inference involves generalizations and statements about the *probability* of their validity.

Decision theory

The methods and techniques of statistical inference can also be used in a branch of statistics called *decision theory.* Knowledge of decision theory is very helpful for decision makers because it is used to make decisions under conditions of uncertainty—when, for example, a state government social service agency cannot specify precisely the demand for its services or when the chairman of the English department at your school must schedule faculty teaching assignments without knowing precisely the student enrollment for next fall.

Chapter 2

## 4 CHAPTER PREVIEW

Chapters 2, 3, and 4 will introduce the concepts and techniques of descriptive statistics. Chapter 2 examines two methods for describing a collection of items: tables and graphs. If you have ever heard someone give a long-winded economic report (such as how many dues are owed by all eighty club members) and wished for a quick graphic display to ease the pain, you already have an appreciation of what's to come in Chapter 2.

Chapter 3

Chapter 3 focuses on special ways to describe a collection of items, particularly the way observations tend to cluster or bunch up. Here, we shall encounter some familiar terms, such as the concept of an average. If the basketball coach at your university says the average height of the members of his team is 6′11″, he is really saying that there is a tendency for the heights of the players to bunch up around 6′11″. For a basketball team with this much height, you intuitively know that the chances of a winning season are quite good—even before you formally study statistics. In Chapter 3, we'll also study the mean, the median, and the mode—all ways of measuring and locating data.

Chapter 4

Chapter 4 finishes our study of descriptive statistics by looking at methods that enable us to measure the tendency for a group of data to spread out, or disperse. Suppose an airline requires that its pilots average 6′ in height, and you recruit a 4′ person and an 8′ person to apply for jobs. You would not get much praise for your efforts even though these two unusual persons do average 6′. Instead, the airline is likely to reject both of your candidates because their heights are too far from the desired average. In this situation the 6′ average is an inadequate summary description of your two candidates. Chapter 4 provides better descriptions of variability.

Chapter 5

Chapter 5 introduces the basic concepts of *probability* (or chance) and, with Chapter 6, gives us a foundation for our study of statistical inference, which will follow in later chapters. Here, we shall examine methods of calculating and using probabilities under various conditions. If you are one of 200 students in a class and the professor seems to call on you each time the class meets, you might accuse that professor of not calling on students at random. If, on the other hand, you are one student in a class of 8 and you never prepare for class, assuming that the professor will not get around to you, then you may be the one who needs to examine probability ideas a bit more.

In Chapter 6, we are concerned with probability *distributions,* that is, with the various ways in which data array themselves when we graph them. Here again we are laying the foundation for later work in statistical inference. You may already have a notion about probability distributions if you have dealt with the concept of the bell-shaped curve in psychology or mathematics courses. If this term means nothing to you, but you are a male who wears a size 16 EE shoe or a female who wears a size 3 AAAA shoe, you may have an intuitive idea about probability distributions, too. Each time you try to get fitted and can't, you probably wish the shoe store manager would order a larger distribution of shoes. If the manager thinks in terms of correct probability distributions, however, he will not order such unusual sizes and won't be able to accommodate people with very large or very small feet. Chapter 6

Most consequential decisions in the public and private sectors are made under conditions of uncertainty because decision makers seldom have complete information about what the future will bring. Also introduced in Chapter 6 is *statistical decision theory,* those methods that are useful when we must decide among alternatives despite uncertain conditions.

Statistical sampling, the subject of Chapter 7, is a systematic approach to selecting a few elements (a *sample*) from an entire collection of data (a *population*) in order to make some inferences about the total collection. Here we shall learn methods that help to ensure that the samples we collect are actually representative of the entire collection. If you have ever examined a peach on the top of a basket, bought the whole basket on the basis of that peach's condition, and then found the bottom of the basket filled with overripe fruit, you already have a good (if somewhat expensive) understanding of statistical sampling and the need for better sampling methods. Chapter 7

Chapters 8 and 9 deal with statistical inference. In Chapter 8, we shall learn to *estimate* the characteristics of a population by observing the characteristics of a sample. Two characteristics of special interest will be how a population tends to "bunch up" and how it spreads out. Chapter 8

The subject of Chapter 9 is *hypothesis testing.* Here we are trying to determine when it is reasonable to conclude, from analysis of a sample, that the entire population possesses a certain property and when it is not reasonable to reach such a conclusion. Suppose a student purchases a $500 second-hand car from a dealer who advertises "our cars are the finest, most dependable automobiles in town." If the car's repair bills during the first month are $600, that one-car sample may cause the student to conclude that the dealer's population of used cars is probably not as advertised. Chapter 9 will allow us to test and evaluate larger samples than those available to the buyer of the used car. Chapter 9

Chapter 10 discusses two statistical techniques: chi-square tests and analysis of variance. *Chi-square tests* are useful in analyzing more than two Chapter 10 populations. They can be helpful in political situations to determine whether, in fact, there is a difference between the proportions of the population supporting two opposing candidates in data taken from different states or geographic regions. Chi-square tests also enable us to determine whether a group of data that we think could be described by the bell-shaped curve actually does conform to that graphic pattern. *Analysis of variance,* the second subject of Chapter 10, is used to test the difference between several sample means. It is a method the Environmental Protection Agency might use when evaluating five series of tests on the same model car. This method can help the agency answer the question, "Are the miles per gallon results really the same, or do they only appear to be?"

Chapter 11

If your university used your high school grade point average and your college board scores to predict your college grade point average, it may have used the technique of *regression analysis,* one of the subjects of Chapter 11. And if you have heard the statement that there is a high correlation between smoking and lung cancer, then the word *correlation* (another topic in Chapter 11) is no stranger to you. *Correlation analysis* is used to measure the degree of association between two variables.

Chapter 12

In Chapters 7 to 11, then, we shall learn how statisticians take samples from populations and attempt to reach conclusions from those samples. But how can we handle cases in which we do not know what kind of population we are sampling (that is, when we do not know the shape of the population distribution)? In these cases, we can often apply the techniques of *nonparametric statistics* presented in Chapter 12.

For students, not statisticians

**5 STRATEGY, ASSUMPTIONS, AND APPROACH** This book is designed to help you get the feel of statistics—what it is, how and when to apply statistical techniques to decision making situations, and how to interpret the results you get. Since we are not writing for professional statisticians, this book is tailored to the backgrounds and needs of college students who, as future citizens, probably accept the fact that statistics can be of considerable help to them in their future occupations but are, very likely, apprehensive about studying the subject.

Symbols are simple and explained

In this book, we discard mathematical proofs in favor of intuitive ones. You will be guided through the learning process by reminders of what you already know, by examples with which you can identify, and by a step-by-step process instead of statements like "it can be shown" or "it therefore follows."

As you thumb through this book and compare it with other basic statistics textbooks, you will notice a minimum of mathematical notation. In the past, the complexity of the notation has intimidated many students who got lost in the symbols, even though they were motivated and intellectually capable of understanding the ideas. Each symbol and formula that is used is explained in detail, not only at the point at which it is introduced but also in a section at the end of the chapter.

No math beyond simple algebra required

If you felt reasonably comfortable when you finished your college algebra course, you have enough background to understand *everything* in this book. Nothing beyond basic algebra is either assumed or used. This book's goals are for you to be comfortable as you learn and for you to get a good intuitive grasp of statistical concepts and techniques. As a member of either a private or a public organization, you will need to know when statistics can help your decision process and which tools to use. If you do need statistical help, you can find a statistical expert to handle the details.

Text problems cover a wide variety of situations

The problems used to introduce material in the chapters and the exercises at the end of each section within the chapter are drawn from a wide variety of situations you are already familiar with or are likely to confront quite soon. You will see problems involving public education, social services, health systems, environmental protection, consumer advocacy, history, psychology, biology, anthropology, sociology, political science, economics, and the private sector.

In each problem situation, a decision maker is attempting to use statistics productively. Helping you become comfortable doing exactly that is the goal of this book.

# 2 Arranging data to convey meaning: tables and graphs