FREEDMAN
PISANI
PURVES

Statistics

DAVID FREEDMAN
ROBERT PISANI
ROGER PURVES

UNIVERSITY OF CALIFORNIA · BERKELEY

Statistics

W · W · NORTON & COMPANY

NEW YORK · LONDON

Cartoons by Dana Fradon

W. W. Norton & Company, Inc. 500 Fifth Avenue, New York, N.Y. 10110
W. W. Norton & Company Ltd. 37 Great Russell Street, London WC1B 3NU

Library of Congress Cataloging in Publication Data
Freedman, David, 1938–
Statistics.
Includes index.
1. Mathematical statistics. I. Pisani, Robert, joint author. II. Purves, Roger, joint author. III. Title.
QA276.F683 519.5 77–29265

890

ISBN 0-393-09076-0

To Jerzy Neyman

Preface

What song the Syrens sang, or what name Achilles assumed when he hid among women, though puzzling questions, are not beyond all conjecture.

— SIR THOMAS BROWNE (ENGLAND, 1605–1682)

TO THE READER

We are going to tell you about some interesting problems which have been studied with the help of statistical methods, and show you how to use some of these methods yourself, explaining why the methods work and what to watch out for when other people use them. Mathematical notation only seems to confuse things for most people, so we are going to do it with words, charts, and tables—and hardly any *x*s or *y*s. Even when professional mathematicians read technical books, their eyes often skip over the equations despite their best efforts. What they really need is a sympathetic friend who will explain the ideas and draw the pictures behind the equations. We are trying to be that friend for those who read our book.

WHAT IS STATISTICS?

Statistics is the art of making numerical conjectures about puzzling questions.

- How should experiments be designed to measure the effects of new medical treatments?

- What causes the resemblance between parents and children, and how strong is that force?
- How is the rate of inflation measured? The rate of unemployment? How are they related?
- Why does the casino make a profit at roulette?
- How can the Gallup poll predict election results using a sample of only a few thousand people?

These are difficult issues, and statistical methods help quite a lot in analyzing them. These methods were developed over several hundred years by people who were looking for the answers to hard questions. Some of these people will be introduced later on in the book.

AN OUTLINE

Part I is about designing experiments properly, so that meaningful conclusions can be drawn from the results. It draws examples from medicine, education, and urban policy, and analyzes some studies which are badly designed and misleading. And it explains some of the questions to ask when judging the quality of a design. Most studies produce so many numbers that it is impossible to absorb them all: summaries are needed. Descriptive statistics is the art of describing and summarizing data. This branch of the subject is introduced in Part II. The discussion is continued in Part III, where the focus is on analyzing relationships—for instance, how income depends on education.

Much statistical reasoning depends on the theory of probability, discussed in Part IV. The connection is through chance models, which are developed in Part V and used in Part VII to solve some problems about measurement error and genetics. Parts VI and VIII, statistical inference, explain how to make valid generalizations from samples. Part VI deals with estimation and answers the questions: How does the Gallup poll take a sample to estimate the percentage of Democrats in the population? How far off is the estimate likely to be? Using tests of significance, Part VIII explains how to judge whether a sample confirms or denies a hypothesis about the population from which it is drawn.

Nowadays, inference is the branch of statistics most interesting to professionals in the field. However, nonstatisticians usually find descriptive statistics the more useful branch—and the one which is easier to understand.

The bare bones of the subject are presented in Chapters 1 to 6, 16 to 21, 23 and 26. After digesting these, the reader can browse anywhere; the next chapters to read might be 8, 10, and 13.

EXERCISES

The numbered sections in each chapter usually end with a set of exercises, the answers being at the end of the book. If you work these exercises as they come along and check the answers against the back of the book, you will get practice in your new skills—and find out the extent to which you have mastered them.

Every chapter except 1 and 7 ends with a set of review exercises; the book does not give answers for these exercises. When working them, you may be tempted to flip backward through the pages until you find the relevant formula. However, reading this book backward will prove very frustrating. The review exercises demand much more than formulas. They call for rough guesses and qualitative judgments, and require a good intuitive understanding of what is going on. The way to develop that understanding is to read the book forward.

Why does this book include so many exercises that cannot be solved by plugging into a formula? The reason is that few real-life statistical problems can be solved that way. Blindly plugging into statistical formulas has caused a lot of confusion. So this book teaches a different approach: thinking.

ACKNOWLEDGMENTS

The writing has been supported by the Ford Foundation (1973–1974) and by the Regents of the University of California (1974–1975). We want to thank Earl Cheit and Sanford Elberg (both of the University of California, Berkeley) for help and encouragement at critical times.

Our typist was Julia Rubalcava. She has been unreasonably patient with us and very deft indeed at converting our scribbles into typescript. We are also in debt to Peggy Darland and Beatrice Rizzolo for imposing administrative order on what would otherwise have been chaos.

Dana Fradon did the cartoons, and Dale Johnson the technical drawings. The computer graphics were handled by David Draper, Roberta Heintz, Richard Lockhart, and Rick Persons.

Very helpful reviews of the manuscript were provided by Frank Anscombe (Yale), Diccon Bancroft (Yale), Leo Breiman (UCLA), Merrill Carlsmith (Stanford), Persi Diaconis (Stanford), David Lane (Minnesota), Richard Light (Harvard), Gerry Mendelsohn (University of California, Berkeley), Tom Rothenberg (University of California, Berkeley), Bruce Rothschild (UCLA), Amos Tversky (Jerusalem), and Geoff Watson (Princeton).

Over the years, people in our courses made hundreds of useful comments on successive drafts of the book. Two of the most perceptive student critics were Carole Rohwer and Elton Sherwin.

Finally, we want to thank our editor, Donald Lamm. Somehow, he turned a permanently evolving manuscript into a book.

Contents

PART I

Design of Experiments

1

Controlled Experiments

Always do right. This will gratify some people, and astonish the rest.
—MARK TWAIN (UNITED STATES, 1835–1910)

1. THE SALK VACCINE FIELD TRIAL

A new drug is introduced. How should an experiment be designed to test its effectiveness? The basic method is *comparison*. The drug is given to subjects in a *treatment group*, but other subjects are used as *controls*—they aren't treated. Then the responses of the two groups are compared. Experience shows that subjects should be assigned to treatment or control *at random*, and the experiment should be run *double-blind:* neither the subjects nor the doctors who measure the responses should know who was in the treatment group and who was in the control group.[1] These ideas will be developed in the context of an actual field trial.[2]

The first polio epidemic hit the United States in 1916, and during the next forty years polio claimed many hundreds of thousands of victims, especially children. By the 1950s, several vaccines against this disease had been discovered. The one developed by Jonas Salk seemed the most promising. In laboratory trials, it had proved safe, and had caused the production of antibodies against polio. A large-scale field trial was needed to see whether the vaccine would protect children against polio outside the laboratory. In 1954, the Public Health Service decided to organize this kind of experiment. Nearly two million children were involved, and about half a million were vaccinated. About a million were deliberately left unvaccinated. And another half a million refused vaccination. The field trial was conducted on children in the most vulnerable age groups—grades 1, 2, and 3. It was carried out in selected school districts throughout the country, those where the risk of polio was believed to be the worst.

This is an example of the method of comparison. The treatment (vaccination) is given only to some of the subjects, who form the treatment group: the others do not get the treatment and are used as controls. The responses of the two groups can then be compared to see if the treatment makes any difference. Here, the treatment group and control group were of different sizes, but that did not matter. The investigators were going to compare the rates at which children got polio in the two groups—cases per hundred thousand. Looking at rates instead of absolute numbers adjusts for the differences in the size of the two groups.

There is a troublesome question of medical ethics here. Shouldn't all the children have been given the protection of the vaccine? One answer is that with new drugs, even after extensive laboratory testing, it is often unclear whether the benefits outweigh the risks. A field trial is needed to find out what the treatment does when it is used in the real world. Now giving the vaccine to a large number of children might seem to provide decisive evidence, even without controls. For instance, if the incidence of polio in 1954 had dropped sharply from 1953, that would seem to be proof of the effectiveness of the Salk vaccine. But it really wouldn't be. Polio is an epidemic disease, whose incidence varies a lot from year to year. In 1952, there were about 60,000 cases; in 1953, there were only half as many. Without controls, low incidence in 1954 could have meant one of two things: either the vaccine worked, or there was no epidemic that year.

The only way to find out whether the vaccine worked was to leave some children unvaccinated. Of course, children could be vaccinated only with their parents' permission. So one possible design was this: the children whose parents consented would form the treatment group and get the vaccine. The other children would form the control group. But it was known that higher-income parents would consent to treatment much more readily than lower-income parents. And this would have created a bias against the vaccine—because children of higher-income parents are more vulnerable to polio than children of lower-income parents. This seems paradoxical at first, but polio is a disease of hygiene. Children who live in less hygienic surroundings tend to contract very mild cases of polio early in childhood, while still protected by antibodies from their mother. After being infected, they generate their own antibodies, which protect them against more severe infection later. Children who live in more hygienic surroundings are less likely to contract these early mild infections, do not develop antibodies, and are less likely to be protected against severe infection later.

The statistical lesson is that to avoid bias, the treatment group and control groups should be as similar as possible—except for the treatment. That makes it possible to conclude that any difference in response between the two groups is due to the treatment, rather than some other factor. If the