STATISTICS

A TOOL FOR SOCIAL RESEARCH

SECOND EDITION ▪ JOSEPH F. HEALEY

Statistics: A Tool for Social Research

Second Edition

Joseph F. Healey
Christopher Newport College

Wadsworth Publishing Company
Belmont, California
A Division of Wadsworth, Inc.

Sociology Editor: Serina Beauparlant
Editorial Assistant: Marla Nowick
Production Editor: Gary Mcdonald
Managing Designer: Andrew Ogus
Print Buyer: Karen Hunt
Designer: Lisa Mirski
Copy Editor: Greg Gullickson
Technical Illustrators: Evanell Towne, Jeanne Schreiber
Compositor: G&S Typesetters, Austin
Cover: Cynthia Bassett-Bogue

Printed in the United States of America 85

3 4 5 6 7 8 9 10—94 93 92 91

Library of Congress Cataloging-in-Publication Data

Healey, Joseph F., 1945–
Statistics, a tool for social research / Joseph F. Healey.—2nd ed.
p. cm.
ISBN 0-534-12450-X
1. Social sciences—Statistical methods. 2. Statistics. I. Title.
HA29.H39 1990
519.5'024301—dc20 89-48052
CIP

Preface

Teaching statistics to sociology and other social science majors is, to say the least, a challenge. Students typically enter the classroom with a wide range of mathematical backgrounds and an equally diverse set of career goals. They are often genuinely puzzled about the relevance of statistics for their professional development and rather less than ecstatic about having the opportunity to gain some mastery of this subject matter. These elements of varied levels of preparation and need, combined with a certain reluctance to plunge headlong into social statistics, make this course a pedagogical task of significant proportions.

This textbook was written to meet the challenge of introducing today's social science majors to statistical analysis while directly addressing these instructional problems. The text makes minimal assumptions about the mathematical background of the students (the ability to read a simple formula is sufficient preparation for virtually all of the material in the text), and a variety of special features have been integrated into the text to assist students to successfully analyze data. The text has been written especially for sociology and social work programs but is sufficiently flexible to be used in statistics courses in political science or in any program with a social science base and an applied focus (for example, public administration, criminal justice, urban studies, and gerontology).

The text is written at a level intermediate between the more rigorous and sophisticated texts on one hand and the mere "cookbook" on the other. That is, while I have not sacrificed comprehensive coverage or statistical correctness, the theoretical and mathematical explanations of why statistics "do what they do" are kept at an elementary level, as is appropriate in a first exposure to social statistics. For example, I do not treat formal probability theory per se. Rather, the background necessary for an understanding of inferential statistics is introduced, informally and intuitively, in Chapters 5 and 6 while considering the concepts of the normal curve and the sampling distribution. The text makes no claim that statistics are "fun" or that the material can be mastered without considerable effort. At the same time, students are not overwhelmed with abstract proofs and mathematical theory, which at this level needlessly frustrate the learning experience.

My major goal is basic statistical literacy. The text is designed to provide a solid foundation in statistical analysis and to prepare students to be intelligent consumers of social research. More specifically, I believe that basic statistical literacy can be defined in terms of three interrelated qualities and, as a way of further describing the nature of this text, I would like to list each of these qualities and briefly summarize how the text is designed to develop them.

Computational Competence. At a minimum, students should emerge from their first course in statistics with the ability to perform elementary forms of data analysis—to execute a series of calculations and arrive at the correct answer. Since students in social science statistics courses frequently

do not have strong quantitative backgrounds, I have included a number of features to help students cope with computations:

- *Step-by-step computational algorithms* are provided for each statistic.
- *Extensive problem sets* are provided at the end of each chapter. For the most part, these problems use fictitious data and are designed for relative ease of computation.
- *Solutions* to odd-numbered problems are provided so that students may check their answers.
- *An introduction to SPSS^x and SPSS/PC+* is included to give students access to the computational power of the computer. This feature is explained in more detail below.

An Appreciation of Statistics. A statistically literate person can do much more, of course, than merely calculate correct answers. Such a person understands the relevance of statistics for social research, can select an appropriate statistic for a given purpose and a given set of data, and can analyze and interpret the meaning of that statistic. This textbook begins to develop these qualities, within the constraints imposed by the introductory nature of the course, in the following ways:

- *The relevance of statistics.* Chapter 1 includes a discussion of the role of statistics in social research and stresses the usefulness of these techniques as ways of analyzing and manipulating data and answering research questions. Each example problem is framed in the context of a research problem. A question is posed and then, with the aid of a statistic, answered. The relevance of statistics for answering questions is thus stressed throughout the text. This central theme of usefulness is further reinforced by a series of boxes labeled "Applications," each of which illustrates some specific way statistics can be used to answer questions.

  Finally, the great majority of the end-of-chapter problems are labeled by the social science discipline or subdiscipline from which they are drawn. The following abbreviations are used as labels: *SOC* for *sociology*, *SW* for *social work*, *PS* for *political science*, *CJ* for *criminal justice*, *PA* for *public administration*, and *GER* for *gerontology*. By identifying problems with specific disciplines, students can more easily see the relevance of statistics to their own academic interests. (Not incidentally, they will also see that the disciplines have a large subject matter in common.)
- *Selecting appropriate statistics.* A series of flowcharts are included to help students select appropriate statistics. These flowcharts have two components. Decision points are represented by diamonds and information by rectangles. The selection process is represented in general terms on the in-

side front cover and at the beginning of each part. Chapters begin with detailed flowcharts that, based on a consideration of the purpose of the analysis, the format of the data, and the level of measurement criterion, lead students to specific formulas or sections of the chapter. I have found these flowcharts very helpful in eliminating much of the confusion about "when to do what" that often characterizes beginning students.

- *Interpreting statistics.* After selecting and computing a statistic, students still face difficulties in understanding what the statistic means. The ability to interpret statistics can be developed only by exposure and experience. To provide exposure, I have been careful, in the example problems, always to express the meaning of the statistic in terms of the original research question. To provide experience, the end-of-chapter problems almost always call for an interpretation of the statistic calculated. To provide examples, many of the Answers to Odd-Numbered Problems in the back of the text are expressed in words as well as numbers.

The Ability to Read the Professional Social Science Literature. The statistically literate person can comprehend and critically appreciate research reports written by others. The development of this quality is a particular problem at the introductory level because of the marked disparity between the concise language of the professional researcher and the rather wordy vocabulary of the classroom. To help bridge this gap, I have included a series of boxes labeled "Reading Statistics." These begin in Chapter 1 and appear every two or three chapters. In each box, I briefly describe the reporting style typically used for the statistic in question and try to alert students about what to expect when they approach the professional literature.

Additional Features. A number of other features make the text more meaningful for students and more useful for instructors:

- *Readability.* The writing style is informal and accessible to students without ignoring the traditional vocabulary of statistics. Problems and examples have been written to maximize student interest and to focus on issues of concern and significance. For the more difficult material (such as hypothesis testing), students are first walked through an example problem before being confronted by formal terminology and concepts. Each chapter ends with a summary of major points and a glossary of important concepts. A glossary of symbols inside the back cover can be used for quick reference.
- *Organization and coverage.* The text is divided into four parts, with most of the coverage devoted to univariate descriptive statistics, inferential statistics, and bivariate measures of association. The distinction between description and inference is introduced in the first chapter and maintained throughout the text.

In selecting statistics for inclusion, I have tried to strike a balance between the essential concepts with which students must be familiar and the amount of material students can reasonably be expected to learn in their first (and perhaps only) statistics course, while bearing in mind that different instructors will naturally wish to stress different aspects of the subject. Thus, the text covers a full gamut of the usual statistics, with each chapter broken into subsections so that instructors may choose the particular statistics they wish to include.

- *Review of mathematical skills.* Appendix H provides a comprehensive review of all of the mathematical skills that will be used in this text. Students who are inexperienced or out of practice with mathematics may want to study this section early in the course and/or refer to it as needed. A self-test is included in Appendix H so students may check their level of preparation for the course.

- *Computer applications and realistic data.* In order to help students learn to take advantage of the power of the computer, this text contains an introduction to SPSS^{X} and SPSS/PC+. These are state-of-the-art statistical packages and may not be available to all students. However, instructors may request a data diskette from Wadsworth that contains a variety of data on a sample of 1000 respondents from the National Opinion Research Council's General Social Survey for 1976 and 1986. The data from the 1986 sample are used at the end of almost every chapter to illustrate SPSS/PC+ applications, and the data set is described in detail in Appendix G. Thus, even if SPSS/PC+ is not available to students, they can still analyze this data with whatever statistical package is locally available and compare their results with those presented in the text. Also, the GSS data bases give students an opportunity to utilize their data-analysis skill on "real-life" data.

- *Instructor's Manual.* The Instructor's Manual includes learning objectives, chapter summaries, a test item file of multiple-choice questions, answers to even-numbered computational problems, and step-by-step solutions to selected problems.

The second edition includes numerous additions to the first. Some of these changes are mentioned in the discussion above; others include more end-of-chapter problems, a new chapter on analysis of variance, the goodness-of-fit test for chi square, and a section on the t test for matched samples. As with the first edition, my goal is to provide a comprehensive, flexible, and student-oriented text that will provide a challenging first exposure to social statistics.

Joseph F. Healey

Acknowledgments

This text has been in development, in one form or another, for more than eight years. An enormous number of people have made contributions, both great and small, to this project and, at the risk of inadvertently omitting someone, I am bound at least to attempt to acknowledge my many debts. I have had the pleasure of developing this second edition with the expert assistance of two different editors at Wadsworth, and I thank both Sheryl Fullerton and Serina Beauparlant for their professional advice and cheerful support. I would also like to acknowledge the contributions of Bob Podstepny of Wadsworth during the early stages of the project, the excellent editing of Greg Gullickson, and the assistance of Andrew Ogus and Gary Mcdonald.

Much of whatever integrity and quality this book has is a direct result of the contributions of colleagues who reviewed the manuscript during the various stages of writing. I have been consistently impressed by their sensitivity to the needs of the students, and I would like to thank Karen Lynch Frederick, Saint Anselm College; Marsha S. Harman, Long Beach State University; Ted Jitodai, San Francisco State University; Janet Kelly-Moen, University of North Dakota; David J. Maume, Jr., University of Cincinnati; Duane R. Monette, Northern Michigan University; Ira Wasserman, Eastern Michigan University; and Paul R. Wozniak, Western Kentucky University. Whatever failings are contained in the text are, of course, my responsibility and are probably the results of my occasional decisions not to follow the advice of the reviewers.

I would like to thank the instructors who made statistics understandable to me (Professors Satoshi Ito, Noelie Herzog, and Ed Erikson), all of my colleagues at Christopher Newport College for their support and encouragement (especially Professors F. Samuel Bauer, Richard Butchko, Robert Durel, Ruth Kernodle, Cheryl Mathews, Lea Pellet, Grace Stuckey, and William Winter), and, in particular, all of my students for their patience and support. For their help in the preparation of the first edition, I thank Wendy Turrissinni, Phyllis Snead Yeaw, Frank Marcucci, Evelyn Green, Christy Smith, Patsy Trail, Debbie Lane, Patricia Stuart, Marsha Butchko, and Elena Grose. I suppose it is a sign of the times that I have no one to thank but my word-processing program for assistance in typing this second edition. I would be very remiss, however, if I did not acknowledge the superb secretarial and organizational talents of Louann Ivins. Also, 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 Appendixes B, C and D from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974).

Finally, I want to acknowledge the support of my family and rededicate this work to them. I am fortunate to be a member of an extended family that is remarkable in many ways. Although I cannot list everyone (my family is also remarkably large), I would like to thank my mother, Alice Healey, and my aunt, Hazel Adams, for all they have done for me. I owe most special thanks to my sons, Kevin and Christopher, and my wife, Judith Bessor Healey, for their patience and support throughout this project.

J. F. H.

Brief Contents

Detailed Contents

1 Introduction

1.1 THE ROLE OF STATISTICS IN SCIENTIFIC INQUIRY

Students often approach their first course in statistics with questions about the value of the subject matter. What, after all, do numbers and statistics have to do with understanding people and society? In a sense, this entire book will attempt to answer this question, and the value of statistics will become clear as we move from chapter to chapter. For now, the importance of statistics can be demonstrated, in a preliminary way, by briefly reviewing the research process as it operates in the social sciences. These disciplines are scientific in the sense that social scientists attempt to verify their ideas and theories through research. Broadly conceived, research is any process by which information is systematically and carefully gathered for the purpose of answering questions, examining ideas, or testing theories. Research is a disciplined inquiry that can take numerous forms. Statistical analysis is relevant only for those research projects where the information collected is represented by numbers. Numerical information of this sort is called **data**, and the sole purpose of statistics is to manipulate and analyze data. **Statistics**, then, are a set of mathematical techniques used by social scientists to organize and manipulate data for the purpose of answering questions and testing theories.

What is so important about learning how to manipulate data? On one hand, let me admit that some of the most important and enlightening works in the social sciences do not utilize any statistical techniques. There is nothing magical about data and statistics. The mere presence of numbers guarantees nothing about the quality of a scientific inquiry. On the other hand, data can be the most trustworthy kind of information available to the researcher and, consequently, deserve special attention. Data that have been carefully collected and thoughtfully analyzed are the strongest, most objective foundations for building theory and enhancing understanding. Without a firm base in data, the social sciences would lose the right to the name *science* and would be of far less value to humanity.

Thus, the social sciences rely heavily on data-gathering for the advancement of knowledge. Let me be very clear about one point: it is never enough merely to gather data (or, for that matter, any kind of information). Even the most objective and carefully collected numerical information does not and cannot speak for itself. The researcher must be able to use statistics effectively. To be useful, the data must be organized, evaluated, and analyzed. Without a good understanding of the principles of statistical analysis, the re-

searcher will be unable to make sense of the data. Without the appropriate application of statistical techniques, the data will remain mute and useless.

I want to stress two points with respect to the role of statistics in the research process. First, statistical analysis is crucial to the social sciences as science. Simply put, without statistics, quantitative research is impossible. Only by the application of statistical techniques can mere data help us to understand the social world better. Second, somewhat paradoxically, the role of statistics is limited. To explain, research can be conceptualized as a process involving several mutually interdependent stages. In the formulation stage, the investigator develops a statement of a problem or a question that the research project will be designed to answer. The research question may come to the researcher from any number of different sources including theory, previous research, or hunches. Once the research question has been formulated, the process enters a second phase, when decisions about research design and methodology are made. The investigator now decides such things as how many cases will be tested, what kinds of cases will be included, and how the cases are to be tested.

Only at the end of this second phase of the process—after all the cases have been tested and all the data gathered—do statistics become directly relevant. Before any statistical analysis can be legitimately applied, however, the first two phases of the process must have been successfully completed. If the researcher has asked poorly conceived questions or has made serious errors of design or method, then even the most sophisticated statistical analysis is valueless. As useful as they can be, statistics cannot substitute for rigorous conceptualization, detailed and careful planning, or creative use of theory. Statistics cannot salvage a poorly conceived or designed research project. They cannot make sense out of garbage.

By the same token, of course, inappropriate use of statistics can render an otherwise useful project useless. Only by successfully completing all phases of the process can a quantitative research project hope to contribute to social understanding. A good knowledge of the uses and limitations of statistics is as essential to the education of the social scientist as is training in theory and methodology.

1.2 THE GOALS OF THIS TEXT

In the preceding section, I argued that statistics are a crucial part of the process by which scientific investigations are carried out and that, therefore, some training in statistical analysis is a crucial component in the education of every social scientist. In this section, we will address the questions of how much training is necessary and what the purposes of that training are. First of all, I take the point of view that statistics are tools. They can be a very useful means of increasing our knowledge of the social world, but they are not ends in themselves. Thus, we will not take a "mathematical" approach to the sub-