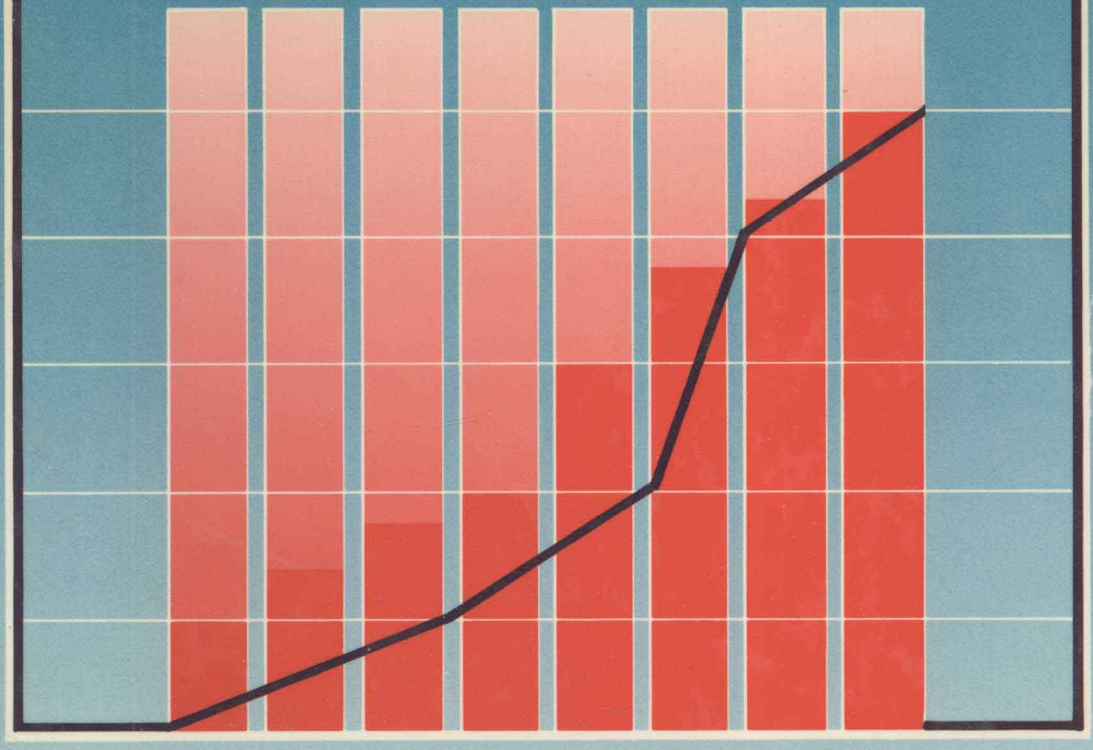
AN INTRODUCTION TO STATISTICAL TECHNIQUES FOR SOCIAL RESEARCH



ANTHONY A. HICKEY

An Introduction to Statistical Techniques for Social Research

Anthony A. Hickey
George Mason University

RANDOM HOUSE  NEW YORK

First Edition
987654321

Library of Congress Cataloging-in-Publication Data
Hickey, Anthony A. (Anthony Andrew)
An introduction to statistical techniques for social research.

Bibliography: p.
Includes index.
1. Social sciences—Statistical methods. 2. Social sciences—Methodology. I. Title.
HA29.H62 1986 519.5 85–19422
ISBN 0–394–32843–4

Manufactured in the United States of America

Designer: Glen M. Edelstein
Cover Design: Celine Brandis

PREFACE

My first encounter with statistics was an unmitigated disaster. I was in a class with two hundred likewise confused souls; since the professor was little more than a small figure at a podium on a stage, I turned to the textbook to help me through the course. The book was written by a theoretical statistician and filled with mathematical proofs. To make a long and painful story short, I eked out a gentleman's C and swore off numbers.

Later, in graduate school, I was introduced to statistics as an integral element of sociological reasoning. This introduction convinced me that statistics can be taught to people with weak mathematical backgrounds in a clear and logical manner. I learned to really enjoy the analytical process and to appreciate good teaching. I owe much of what I know of statistics and the methodology of communicating it to Joe D. Francis of Cornell University. He should see some of himself in this book.

The field of statistics has changed quite a bit in the twelve years that I have been responsible for teaching introductory statistics to sociology, social work, and anthropology students. First and foremost, the computer has removed much of the tedium in calculating statistics. The use of social science statistical packages has allowed students to concentrate on the logic and interpretation of statistical analysis while letting the computer "crunch the numbers." Students still need to calculate basic statistics, but hand-held calculators have made these tasks much easier, and complex techniques are relegated to the computer.

This textbook has three primary objectives: the development of statistical literacy, the building of research skills, and the preparation of students for further study in social science research techniques. Statistical literacy is becoming an important prerequisite for the understanding of the world around us. We must be able to "read" our culture and since so much information is presented in statistical terms, the ability to interpret and critique what we are presented is absolutely necessary. I'm not just talking about the professional literature. Many of the examples students will encounter in this book are drawn from the popular press. To a consumer of information, statistical literacy is almost as important as the ability to read.

Social science majors need to develop basic research skills for two reasons. First, I have observed that professors who teach specialized upper level courses very often expect their students to become engaged in research as part of the course requirements. Students who have mastered the basics of statistics have a great advantage in preparing class reports and seeing the application of course material to the real world. The second reason that statistics is an

important part of the social science curriculum is that it teaches students skills that are increasingly in demand among employers. I am always pleased when students return to the campus and tell me that their research skills helped them to find positions that they found rewarding.

The development of research skills involves learning more than just the "how-to" of statistical techniques. The why and the so what of statistics are equally important. To that end, the text approaches statistics as a research tool with a methodological context and so in several areas methodological and research design issues are presented along with the statistical issues. In this way, the utility of statistics as well as the mechanics are emphasized.

The third objective of the book is to provide students with a good preparation for courses in more advanced statistical techniques, research methods, and computer science. Whether these courses are taken as part of an undergraduate career or a continuation of studies in graduate school, introductory statistics should provide the proper background and this book is designed to help in that regard.

The features of the book reflect the above objectives. The emphasis is on the conceptual elements of statistics rather than the mathematical underpinnings and refinements. Students will need to use basic algebra in following the presentations and working the problems, but I have avoided mathematical proofs and developments. Without compromising soundness or comprehensiveness, the presentation is designed to be nonthreatening, clear, and "user-friendly." I have avoided using too much technical jargon and tried to engage the student in the research enterprise.

One of the difficulties in learning (and teaching) statistics is seeing the context and utility of statistical thinking. In order to keep statistics in context, I have used real examples and problems from the research literature that demonstrate the application of statistical techniques to theoretical issues that students encounter in their other courses. The Applications sections that are found at the end of each chapter show how statistical techniques are used in both very familiar situations or as extensions of the basic techniques to somewhat more complex situations.

Statistics cannot be learned simply by reading a textbook and therefore the problems that follow each chapter are both conceptual and operational. These problems are designed to encourage students not just to practice the necessary calculations but also to think about problems in a statistical framework. The student workbook that is available to accompany the text follows the same format with additional problems that involve computer-based analyses using SPSS and a data set that is provided as an appendix.

One last feature of this text is the somewhat heavy emphasis on methodology and research design. This material is included for several reasons. I expect my students to be able to interpret the results of their calculations and in order to do that, they must be able to see why a technique should be utilized (the research question) and what it means. I have found that students find this easier when the interplay between statistics and research design is made

explicit. It is impossible to completely integrate research methods and statistics in one textbook but I hope that covering some methodological issues along with the statistical material will help foster a more intuitive and less mechanistic understanding of the research process.

The final chapters of the text present techniques that are sometimes left out of introductory statistics classes but are found more and more in the professional literature. While the presentation is almost entirely conceptual, it is designed to help students read the journals and see the more interesting applications of statistical reasoning to complex theoretical problems. I hope that the book will be a helpful reference after the course is over; to that end, I have included a section on further reading for each chapter.

I am grateful to many individuals who helped me in the writing of the text. Ten years of Sociology 221 students have convinced me of the need for the book and helped me see where different approaches should be used to get various ideas across. The reviewers who took the time to read and react to various drafts of the chapters deserve special recognition. Their reactions, both positive and negative, helped me in terms of both the clarity of the writing and pointing out errors in presentation. Any remaining errors are mine, of course, but I would like to acknowledge the following people who offered helpful suggestions in reviewing the manuscript: Kenneth J. Berry, Colorado State University; William R. Kelly, University of Texas at Austin; Sally K. Ward, University of New Hampshire; Delbert J. Ervin, Illinois State University; Melissa A. Hardy, Florida State University; Art St. George, University of New Mexico; Gillian Stevens, University of Illinois at Urbana-Champaign; Alvin C. Dorse, North Texas State University; Lawrence Hamilton, University of New Hampshire; Patricia Gwartney-Gibbs, University of Oregon; Jim Mitchell, East Carolina University; Donald A. Tennant, Oklahoma State University; Margaret Platt Jendrek, Miami University; Wade Clark Roof, University of Massachusetts at Amherst; John F. Crowther, California State University at Northridge; Richard A. Zeller, Bowling Green State University; Ramon E. Henkel, University of Maryland; Paul Wilken, University of Virginia; Clifford Clogg, Pennsylvania State University; Theodore Greenstein, Southern Illinois University; Kathleen Napalino, Cleveland State University; Ronald Little, Utah State University; and Jeffrey Sobol, Gettysburg College.

The people at Random House have been a continuing source of encouragement throughout the project. I would like to thank Tom Kitt, Roth Wilkofsky, Dorchen Leidholdt, and, in particular, Bert Lummus.

My colleagues at George Mason deserve special mention. Joe Scimecca encouraged me to take on the project and provided an invaluable sounding board. Averett Tombes has been my boss for the past three years and has given me the time and encouragement which were absolutely necessary. Tom Dietz and Max Schlueter both used earlier drafts in their courses, and I thank them (and their students). Their comments and suggestions were above and beyond the call of duty.

My final acknowledgment is the most important. Jo Ann Hickey is both

colleague and wife. Her support and assistance has been a constant throughout the project. She critiqued every chapter, argued me out of some questionable ideas, and provided many good ones. Much of the book and all of the workbook is a reflection of her intelligence and common sense. It is to her and our children, Michelle and Tom, that this book is dedicated.

Anthony Andrew Hickey
Fairfax, Virginia

CONTENTS

1

Statistics and Measurement

1

INTRODUCTION: A PRELIMINARY WORD TO THE STUDENT

It is a fact of life that most social science students who find themselves in a statistics course are there because it is required. It is also commonplace that most students in this situation look forward to a semester of "fear and loathing." Several reasons are given for this attitude:

1. "Statistics is math and I can't do math." When pushed, students may even announce that one reason they are majoring in sociology or social work is that they are "people people" and not "numbers people."
2. "Statistics is like Latin—a dead language." Any course *can* be boring and out of touch with reality, but no subject has to be.
3. "Statistics is an impossible course—my GPA can't absorb the bad grade that I'll get."

My experience in teaching statistics to students of sociology, anthropology, and social work at the undergraduate level has convinced me that none of the above statements need be true. While many textbooks offer a mathematical approach to statistics, this book requires no more than basic algebra. If you can add, subtract, multiply, and divide you will able to follow the text. If you have a calculator you will be able to work along with the presentations and applications of the statistical techniques. (In more advanced courses the computer can provide the same mechanical assistance.)

The goal of this text is to develop an understanding of statistics that emphasizes not only the "how-to" but also the "why" and the "when" and the "so-what." We will approach statistics by considering examples chosen from research journals in the disciplines of sociology, anthropology, and social work. Once you have worked through problems that concern familiar topics, the mystery of statistics will be removed and the importance and practicality of statistics will be demonstrated. To learn statistics conceptually and operationally involves studying the concepts *and* working problems, and you should be prepared to work hard and to work often. Statistics cannot be "crammed"; because the material proceeds logically and later concepts build on earlier ones, the only way through this course is to study and work regularly and not fall behind. Alternative approaches generally lead to disaster to grades and psyches.

The reward for this persistence and perseverance will be a *working* knowledge of statistics. A working knowledge is important in order to handle the dual research needs of performing statistical analyses and identifying garbage passed off as social research. To reach our goal, we will have to make certain assumptions throughout the course. We will concentrate on statistical techniques while assuming "methodological goodness," that is, that sampling is properly designed and carried out, that measurement of concepts is reasonably error-free, and that the theory being tested is reasonable. You will be introduced to these problems in other courses. While they are interesting, they would detract from our topic. We will assume that the data are sound and concentrate on the statistical techniques used in sociological, anthropological, and social work research. It is worth mentioning, however, that the application of good statistical techniques will not salvage a research project that is deficient in conceptualization, measurement, sampling, or other areas.

STATISTICS AS A RESEARCH TOOL

The words "statistics" and "data" are not dirty words. Statistics and data are useful in making research reports clear, and they provide a basis for the conclusions of a research study. In this sense statistics is more than just a collection of facts and figures; it is the application of specific techniques of analysis and presentation to particular research problems. Statistical analysis involves a process of reasoning that is neither mechanical nor encyclopedic. We need to understand the assumptions of various techniques and to be able to calculate the basic formulas. But we also need to know when to use these techniques and how to interpret our results. My statistics professor used to call this "keen salty judgment," and we will stress this process of reasoning and the questions of why and when as we learn how to apply statistical techniques to research problems.

The research process has been described as a series of steps. While different authors vary the sequence somewhat, a fairly basic outline is:

1. *Research Question:* What are we studying and what do we want to know? For instance, we might be interested in the use of different types of medical services by different types of households in a suburban county in 1985. All concepts used in the study need to be carefully defined.
2. *Hypotheses:* Hypotheses are really "educated guesses" about what we will find and can take various forms, from the simple to the fairly complex. For our example, we might guess that families with small children use more medical services than families without small children. Most research studies have several hypotheses to be tested.
3. *Data Gathering:* How will the data be collected and from whom? For our example we might take a random sample of households from our county and then interview them with a questionnaire.
4. *Analysis:* What techniques will be used to test the hypotheses under study? We might calculate the frequency of visits to doctors for our two types of families and compare the average number of visits.
5. *Interpretation:* What do the results indicate? By analyzing several aspects of medical service use and testing several hypotheses we may derive quite a lot of information, but we have to decide what it means.
6. *Theory:* Can we fit our research results into a theoretical framework? Does our study support previous studies of medical service use? Why did we find what we found? What other hypotheses could we test that would increase our understanding of why different groups have different patterns of use of services? If our research question was derived from a theory, this stage would involve deciding if the theory is supported by our data or if the theory needs to be revised.

It is useful to think about the process of research as a circular model (see Figure 1.1). Each stage of the research leads to another stage, and the process is continuous. The fitting of our results to theory may suggest other contexts in which to test the theory. Our results may contradict the theory, so alternative hypotheses are suggested and new theoretical interpretations are constructed. Often the data themselves are a starting point in the research process. A researcher may observe some recurring phenomena and attempt to explain them in a theoretical context. The point is, research never has a real starting point or ending point. Knowledge is accumulated through both deduction and induction. We try to explain the world around us by testing theories with