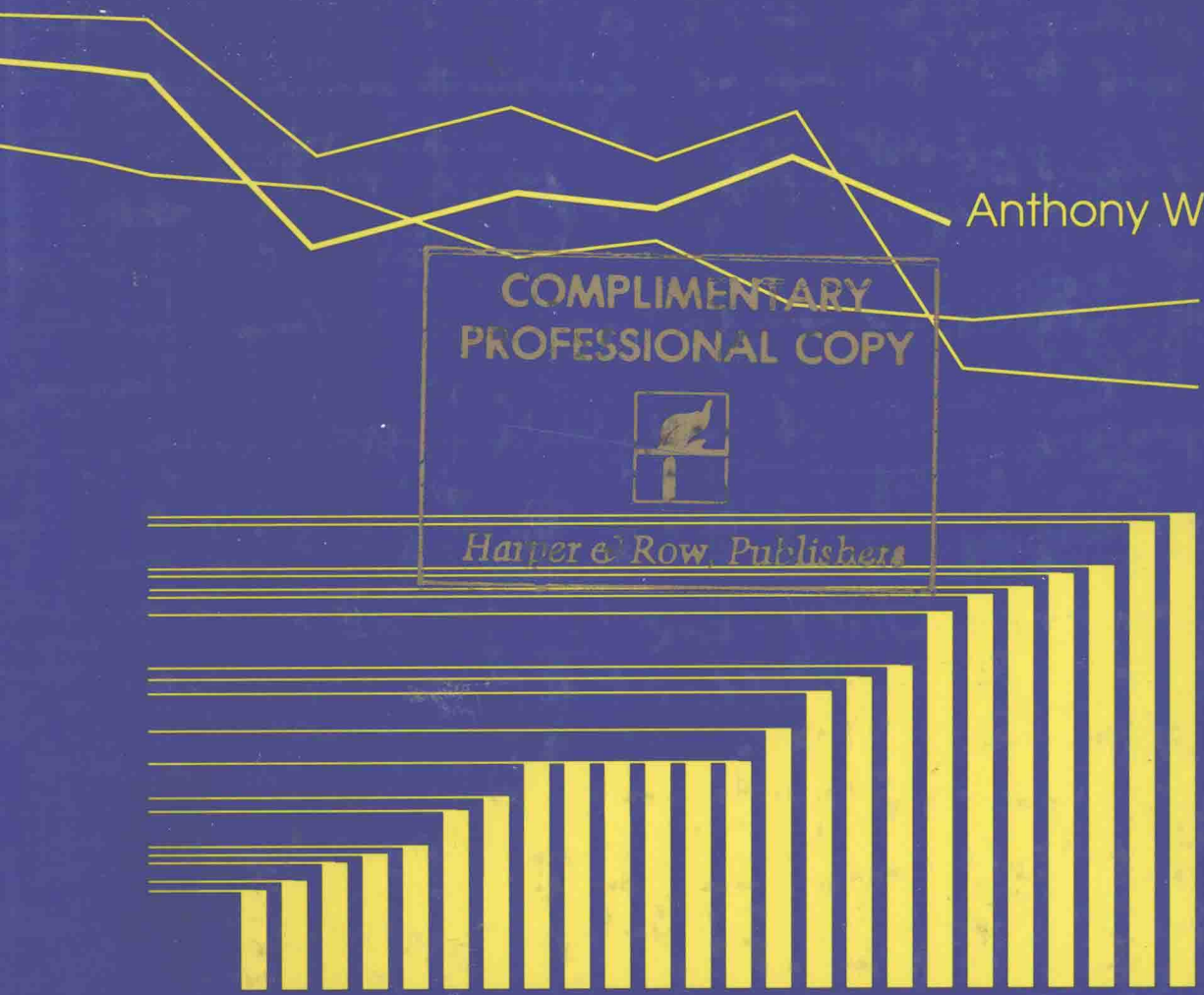
Statistics for the Social Sciences

With Computer Applications

Anthony Walsh



STATISTICS FOR THE SOCIAL SCIENCES

With Computer Applications

Anthony Walsh
Boise State University

HARPER & ROW, PUBLISHERS, New York
Grand Rapids, Philadelphia, St. Louis, San Francisco,
London, Singapore, Sydney, Tokyo

Sponsoring Editor: Alan McClare
Project Editor: Donna Conte
Art Direction and Cover Coordination: Heather A. Ziegler
Cover Design: Jung Sok Yo
Production: Beth Maglione

STATISTICS FOR THE SOCIAL SCIENCES: With Computer Applications

Library of Congress Cataloging-in-Publication Data

Walsh, A. (Anthony)
Statistics for the social sciences : with computer-based applications / Anthony Walsh.
p. cm.
Includes bibliographical references.
ISBN 0-06-046894-7
1. Social sciences—Statistical methods. 2. SPSS X (Computer program) 3. SAS (Computer program) I. Title.
HA29.W33574 1990 89-26814
300'.01'5195—dc20 CIP

90 91 92 93 9 8 7 6 5 4 3 2 1

Preface

Statistics is a fascinating field of study, especially when it is used in exploring "real-world" problems. I treat statistics as a laboratory science. Students learn the logic and computation of the statistics in lectures and then apply what they have been taught in the computer lab.

The teaching and learning of statistics have changed tremendously over the last few decades. As high-speed computers, hooked to individual terminals, replaced card sorters and key punchers, we became able to do increasingly sophisticated things with our data. Unfortunately, statistics texts, on the whole, have not kept pace with the increasing capabilities.

The major advantage of this book is the opportunity for students to gain "hands on" experience with real data. Instructors who use this book may obtain four computer-ready data sets of floppy disks to facilitate this practical experience. These data sets contain a total of 153 variables, ranging from the self-esteem of multiple-sclerosis patients to the sex acts of convicted criminals, and from IQ scores of juvenile delinquents to measures of support of the Supreme Court by the elderly. The range of variables provided are of interest to a variety of social science disciplines.

I introduce the theory and logic underlying each statistical technique through narrative and simple computational examples. It is my firm belief that it is still absolutely necessary for students to solve small problems with their calculators in order to gain a grasp of the techniques. In most cases, both definitional and computational formulas are given. I then use either an SAS* or SPSS-X Data Analysis System** computer printout

---

*SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA. Output from SAS procedures is printed with the permission of SAS Institute Inc., Cary, NC, USA. Copyright © 1983.

**SPSS-X is a registered trademark of SPSS Inc. of Chicago, Illinois, for its proprietary computer software. Output from SPSSx procedures is printed with the permission of SPSS Inc. of Chicago, IL.

to further explore the technique. The printouts are based on one of the data sets provided with the text. The proper interpretation of these computer printouts can be confusing, even to advanced students, but we go through each printout in a step-by-step fashion. This is the first introductory text that integrates theoretical and computational material with "guided tours" through the intricacies of a computer printout.

This method of teaching statistics gives students the opportunity to gain closure with the statistical concepts developed in the classroom by using the computer to run their own jobs. In my own classes, I have students choose a topic from one of the data sets and follow it all the way through from the simple descriptive level to multiple regression. This continuity of topics gives students a strong intuitive grasp of the utility of statistical analysis at the various levels of sophistication (for instance, they may find that a conclusion arrived at via bivariate analysis does not hold up under multivariate analysis). This methodology goes a long way in helping students to grasp the interrelatedness of the statistics that he or she will be using. This text's emphasis on continuity and interrelatedness of the various statistical techniques leads many students to appreciate the symmetry and beauty of statistics.

Another strong point of this text is the presentation of both SAS and SPSSx computer package programs throughout. Although I do not claim to have condensed into one book all the information contained in a dozen expensive SPSSx and SAS manuals, much essential introductory information is provided. Students are given the precise instructions necessary to run their statistical jobs in both languages and the information necessary to interpret the output. A special chapter on the use of the computer in social science research has been included for those who are using the computer in conjunction with the text (see Chapter 16).

I have included only the most popular and frequently used statistics in this text, excluding, for example, the largely meaningless average deviation. An in-depth understanding of the major statistical tools is superior to a nodding acquaintance with a smorgasbord of techniques, few of which are ever seen in professional journals. Class time is limited; the students' understanding of the major tools of their trade should not be. On the other hand, some statistical techniques not usually found in introductory texts are presented where they further the interpretation of traditional techniques. For instance, the odds ratio is presented in conjunction with chi-square, and eta is presented in conjunction with the t test and ANOVA.

A strong emphasis is placed on correlation and regression techniques. A perusal of any quantitative social science journal will convince the reader that these techniques, by a wide margin, are the most frequently used of all statistics. I have also included a chapter on advanced statistics (path analysis, logit regression, and factor analysis), since these techniques are frequently encountered in the literature today (see Chapter 15). Emphasis in Chapter 15 is placed on the logic of the techniques, providing students with the instructions to obtain these statistics from the computer and providing them with an interpretation of the computer printout.

A Note on Using This Text

Statistics for the Social Sciences (With Computer Applications) is a comprehensive and flexible text that is suitable for a wide variety of teaching styles and orientations toward the subject matter. Although I use the computer in my introductory classes, many

instructors may not desire to integrate computer applications into their course. Having students run computer jobs and requiring them to interpret what they have found is time consuming and necessarily detracts from other things that some instructors may feel are more important for beginning students. This book can stand on its own as a traditional instructional text without the computer material, which is presented in addition to, not instead of, the usual statistical material. Even if the computer is not actually used by students, instructors may find the discussions that are related to the printouts to be a useful teaching tool.

In recognition of the wide variety of statistics classes that an instructor may teach, I have developed four suggested teaching outlines for those using this book based on: (1) the teaching preference (with and without the computer) for a one-semester course, (2) a two-semester sequence, and (3) a one-semester graduate course. These suggested outlines follow. The topics suggested for omission under certain course outline suggestions are usually the more advanced topics associated with a particular technique.

Suggested Outlines for Various Types of Courses

The following is a suggested outline for a one-semester introductory course in which a computer is not used and for which a broad survey of basic techniques is desired.

Chapter 1
Chapter 2
Chapter 3
Chapter 4
Chapter 5
Chapter 6
Chapter 7 (probably omitting multiple comparison tests)
Chapter 9
Chapter 10 (probably omitting certain techniques)
Chapter 11
Chapter 12 (probably omitting certain topics)
Chapter 13

The following is a suggested outline for a one-semester introductory course that emphasizes practical research, in which a computer is used in most chapters. (With this emphasis, Chapters 1 through 5 may have to receive less consideration than is optimally desirable.)

Chapter 1
Chapter 2
Chapter 3 (computer exercise: central tendency and dispersion)
Chapter 4
Chapter 5
Chapter 6 (computer exercise: t test)

Chapter 7	(optional chapter)
Chapter 9	(computer exercise: chi-square)
Chapter 10	(computer exercise: crosstabulation)
Chapter 11	
Chapter 12	(probably omitting certain topics) (computer exercise: bivariate correlation and regression)

The following is a suggested outline for a two-semester introductory sequence.

Chapter 1	
Chapter 2	(computer exercise: histogram and bar chart)
Chapter 3	(computer exercise: central tendency and dispersion)
Chapter 4	
Chapter 5	
Chapter 6	(computer exercise: t test)
Chapter 7	(computer exercise: ANOVA)
Chapter 8	(computer exercise: two-way ANOVA)
Chapter 9	(computer exercise: chi-square)
Chapter 10	(computer exercise: crosstabulation)
Chapter 11	(computer exercise: elaboration techniques and partial gamma)
Chapter 12	(computer exercise: bivariate correlation and regression)
Chapter 13	(computer exercise: multiple regression)
Chapter 14	(optional)
Chapter 15	(optional)

The following is a suggested outline for a one-semester graduate course. (Chapters 1 through 5 may be addressed briefly for "brush-up" purposes.)

Chapter 1	
Chapter 2	
Chapter 3	(computer exercise: central tendency and dispersion)
Chapter 4	
Chapter 5	
Chapter 6	(computer exercise: t test)
Chapter 7	(computer exercise: ANOVA)
Chapter 8	(computer exercise: two-way ANOVA)
Chapter 9	(computer exercise: chi-square)
Chapter 10	(computer exercise: crosstabulation)
Chapter 11	(computer exercise: elaboration techniques and partial gamma)
Chapter 12	(computer exercise: bivariate correlation and regression)
Chapter 13	(computer exercise: multiple regression)

Chapter 14 (some sections may be omitted)

Chapter 15 (computer exercise: logit regression)

I would like to acknowledge a debt of gratitude to professor Paul Hatab of the Department of Mathematics at Boise State University for his thoughtful reading and criticism of the manuscript. We spent many hours together discussing a number of statistical topics. I would also like to acknowledge my former statistical mentors: Dr. Ira Wasserman of Eastern Michigan University, Dr. Neil Palmer of the University of Toledo, and Drs. Kenneth Rothrock and Richard Zeller of Bowling Green State University.

I would also like to thank the following professors who reviewed part or all of the manuscript in various stages of its development. Their suggestions were important and I appreciate their efforts.

Kenneth J. Mietus, Western Illinois University

Michael Hughes, Virginia Polytechnic and State University

Kenneth Wilson, East Carolina University

Gregory L. Weiss, Roanoke College

Ernest Kurnow, New York University

Debra Friedman, University of Arizona

Satish Sharma, University of Nevada, Las Vegas

Ronald S. Edari, University of Wisconsin

Thanks are also extended to my sponsoring editor, Alan McClare, for having faith in this project and to my project editor, Donna Conte, for her scrupulous attention to detail. Thanks also to the students in my statistics classes who suffered through various drafts of this book (especially Grace Balazs, who did a lot of proofreading).

Special thanks to all members of my immediate family—my wife, Patricia Ann, my sons, Robert and Michael, and their wives, Dee and Sharron—for their support throughout this project.

Anthony Walsh

Contents

Chapter 1

Introduction to Statistical Analysis

1.1 WHY STUDY STATISTICS?

As civilizations become more complex and technical, it becomes increasingly necessary for the thinking of the average person to become more complex and technical. The Industrial Revolution was the beginning of the end of a long period in history in which people could feel comfortably at home in their culture with only strong backs and the authority of received opinion to sustain them. It soon became imperative for citizens to understand, interpret, and analyze the written words of their culture. People who did not have the opportunity or the ability to adjust to this new requirement for full cultural participation were, by and large, condemned to lives of poverty and ignorance. They were considered to be "illiterate" and assigned society's least meaningful roles. H. G. Wells, writing in the nineteenth century, stated that "statistical thinking will one day be as necessary for efficient citizenship as the ability to read or write." I believe that day has arrived.

We hear much talk today about computer literacy, and much dire talk about the consequences of being left behind in the race to acquire it. The computer is a tool, however, a repository of the techniques that enable us more quickly and thoroughly to understand, interpret, and analyze data about phenomena. Its function is analogous to that of the library, a tool and a repository of the written word. To use the library, one must first learn the symbols of the written word that convey knowledge to us. If one cannot read, the library is not a tool; it is just another building. If one does not understand the symbols generated by the computer, it too is not a tool—it is just another fancy electronic gadget. It is our belief that an understanding of statistics will be as important to the educated person of the future as reading was to our great grandfathers, and for much the same reason.