

The background of the cover is a complex, abstract mosaic. It consists of numerous small, irregular squares and rectangles in a variety of colors, including shades of red, orange, blue, grey, and white. The overall effect is a textured, pixelated surface that resembles a digital or artistic collage.

Social Statistics Using MicroCase[®]

WILLIAM FOX

Social Statistics Using MicroCase®

William Fox
Skidmore College

MicroCase Corporation
Bellevue, Washington

Cover Painting: *Puzzle* by Susan Miller-Thompson

MicroCase® and *ShowCase*® are registered trademarks of MicroCase Corporation.

Library of Congress Cataloging-Publication Data

Fox, William, 1942-

Social statistics using MicroCase/William Fox.

p. cm.

Includes bibliographical references (p.) and index.

ISBN 0-922914-10-9

1. MicroCase. 2. Social sciences—Statistical methods—Computer programs. I. Title.

HA32.F68 1992

92-15365

300'.2854—dc20

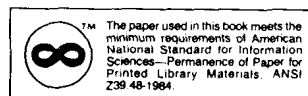
CIP

Copyright © 1992 by MicroCase Corporation

All rights reserved. No part of this book may be reproduced in any form without permission in writing from the publisher, except by a reviewer who wishes to quote brief passages in connection with a review written for broadcast or for inclusion in a magazine or newspaper. For information contact MicroCase Corporation, 1301 120th Avenue N.E., Bellevue, Washington 98005. Phone (206) 635-0293.

Manufactured in the United States of America

10 9 8 7 6 5 4 3



Preface

Like most statistics teachers, I have often thought about writing a statistics text. That's understandable, of course. We all learn from experience what works well in our courses and how best to handle statistical topics. But until recently I never got serious about writing a stat text. The reason was that the world just didn't need another one. There are scads of excellent statistics texts already available. My bookshelves are filled with them.

Then MicroCase® came along, offering a reason—in fact, an incentive—to write an introductory statistics textbook. MicroCase is a terrific computer program for doing statistics. It is wonderfully friendly yet plenty powerful enough to do any statistical analysis that 99 out of 100 undergraduates are likely ever to carry out. For that matter, it can handle the statistical needs of most of the rest of us too. Here is a statistics package easily accessible to computer novices but appreciated as well by more experienced users.

I use MicroCase in my classes and find it works better than any statistical package I ever adopted (and I have used a lot of them). Students learn statistics—and substantive social science too, for that matter—with minimal distraction from the computer itself. No memorization of commands. No involvement with Byzantine operating systems. No inundation with esoteric options. Just learning statistics (or whatever). Just doing data analysis.

Here then was a reason to write a stat text, one using MicroCase as an aid to learning statistics. I have tried to write a text as good as the software it integrates. And I have tried too to write a book consistent with MicroCase's philosophy. MicroCase is a no nonsense, pragmatic piece of software. It is written to do data analysis and get the job done efficiently, competently, intelligently. I hope this text does the same.

What to Expect in This Text

This book covers basic statistical methods. Although you will not be a full-fledged statistician after finishing *Social Statistics Using MicroCase*, you will know enough statistics to evaluate critically others' uses of basic statistical techniques and to analyze data yourself. In other words, you'll be both an informed consumer and a responsible producer of statistics. You will also be well-prepared to study more advanced statistics should you choose to do so. In short, you won't learn everything about statistics from this text, but you will learn a lot.

I have written this text for students in the social sciences and related disciplines such as social work, criminal justice, public administration, and education. I assume you need to carry out real analyses using real data, so I do not avoid the problems you and I find in the "real world"—missing information, skewed data, outliers, variables that need to be collapsed, and so on. In this text we'll face the messiness that makes the social world so challenging to study and so much fun to live in. This text relies heavily on data from sources like the U.S. Census, *Uniform Crime Reports*, and an actual national survey, so we can't avoid facing the real issues and problems that come with real data.

I also assume you need to present your statistical analyses to others (your instructor now, employers or a larger public later in your life), so I include guidelines for writing up statistical analyses, formatting tables, and constructing graphs. You don't need to be a professional statistician to communicate statistical findings effectively to others. I think you'll be able to do so after finishing this text.

So this text does not skimp on nitty-gritty details about handling real problems in the real world. But I am also concerned that you understand statistical principles and reasoning. You need this understanding to know what statistical procedures are right for given situations and to interpret statistical analyses sensibly. Therefore, I'll go beyond formulas and procedures to explain why we are doing what we do.

You will find no computational formulas in this book to facilitate calculations. There is, of course, a place for computational formulas. They belong in a Museum of Statistical Antiquities in a glass case next to slide rules to remind us how unpleasant and tedious statistics used to be.

But this text does not shy away from formulas and calculations. To the contrary, we will learn to do calculations “by hand” before using the friendly computer to do this mostly routine work for us. Calculating statistics with our human brains is essential for understanding why statistics work the way they do and what the computer’s faster brain does for us. However, we will use definitional rather than computational formulas. Although less convenient and more tedious, definitional formulas describe what statistics are really about and thus are the best way to understand how statistics works. The *Doing Statistics Using MicroCase* workbook accompanying this text invites you to do hand calculations before carrying out computer analyses.

But in the last analysis (literally), the computer exercises in the workbook are even more important than the “by hand” exercises. I have written this text and workbook for you to use actively with a computer. You should, in fact, read much of this text while using a computer.

You’ll find this textbook conversational in style. That’s deliberate. It’s the way I teach. My students and I interact in the classroom and, within the limits imposed by print, there is no reason *you and I should not interact via this text*. So I hope you will not take my use of the second-person “you” as an unwarranted familiarity, nor my frequent use of “we” as either an affectation or a ploy. You and I—*we*—are learning statistics together even though you are learning stat for the first time and I am learning it for the umpteenth time.

When you are finished with this book, I hope you will begin a lifetime using statistics effectively. I trust you will be an intelligent critic of others’ applications of statistics and will use statistics intelligently yourself. But beyond these “practical” goals, I hope you will appreciate the beauty, elegance, and grace of statistics. Like other magnificent achievements of human reason, statistics at its core is an aesthetic endeavor.

What You Need to Bring to This Text

You are probably interested in what you need to know *before* you begin to learn statistics. I assume you know basic arithmetic—adding, subtracting, multiplying, and dividing. You should also be reasonably comfortable with fractions and decimals, squares and square roots, and simple equations. (Andrew R. Baggaley’s *Mathematics for Introductory Statistics* can help you brush up.)

Whatever else you need to know, you’ll learn along the way. All

the better if you have had a course in research methods, but that is not essential. Likewise, I hope you have had several substantive courses in the social sciences, but those too are not really necessary. Frankly, I assume more motivation and good will on your part than I do prior knowledge.

I trust that you have a good statistics teacher. I think good texts matter (that's why I've worked hard on this one), but reading even the best text is less important than learning from a knowledgeable and enthusiastic teacher.

Equipment Needs

You'll need an IBM personal computer or a fully compatible clone that operates under MS or PC DOS 3.2 or higher. The IBM computer may be a PC, XT, AT, or PS/2 model. The computer must have at least 512K of memory and either one 3.5" or two 5.25" disk drives. Your computer needs a graphics card; a color graphics card is not necessary but will allow you to enjoy the color displays produced by MicroCase. A dot matrix printer is fine for printing everything you need to print.

I've written this text and the *Doing Statistics Using MicroCase* workbook for you to use with Student MicroCase, but you can use the full MicroCase instead if that is available at your school. These two versions of MicroCase work almost exactly the same. The only difference is that Student MicroCase does not have the data entry capability or some more advanced data manipulation and analysis features found in the full version. Student MicroCase is able to do everything presented in this text except for the advanced statistical techniques briefly described in Chapter 12.

Oh yes, you also need a calculator. An inexpensive one is fine as long as it finds square roots. Some graph paper will also prove useful.

Some Acknowledgments

Most of this book was written while on sabbatical leave from Skidmore College. I'm fortunate indeed to teach at a college so committed to the liberal arts, high-quality teaching, and support of its faculty . . . and so cognizant that these three commitments are one.

This text began with my former colleague Richard Rosenfeld, now at the University of Missouri at St. Louis. Rick extolled the

virtues of MicroCase, gently nagged me to try it, suggested the need for an accompanying stat text, and then actively supported my project at critical points. Rick also reviewed drafts, suggesting numerous changes that I have incorporated into this final text. Without Rick I would not have written this book.

Comments from Michael C. Kearl of Trinity University, David Royse of the University of Kentucky, and an anonymous reviewer were most useful. Jennifer Fox, John Fox, Bree Pasternack, Geoff Rishell, Thomas Schmeling, and Deborah Sutherland helped in diverse ways, not all statistical. My sister Phyllis Hriso also helped in very special ways that she best understands and that I really appreciate.

The MicroCase folks—Steve Bird, Dave Smetters, and especially Lynne Roberts and Rod Stark—eagerly welcomed this book and did everything they could to aid its progress. They developed Student MicroCase, offered advice, handled critiques of drafts, and provided MicroCase displays. Most important of all, however, they offered enthusiasm.

I received all sorts of assistance from Skidmore College's Computer Services. Its staff must be the user friendliest bunch of computer personnel anywhere. Leo Geoffrion and John Danison were especially helpful to me.

Thanks too to my students over the years. True, they couldn't stop me from trying different teaching strategies on them. But they need not have been so good-natured about it nor have taught me so much, and for that I am grateful.

I want to thank Elton Jackson, Professor of Sociology at Indiana University, for teaching me statistics. On the first day of class Elton promised a new way of seeing the world. He delivered on that promise. I learned statistics and much more from Elton Jackson.

I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Dr. Frank Yates, F.R.S., and the Longman Group Ltd, London, for permission to reprint parts of Tables IV and V from their book *Statistical Tables for Biological, Agricultural and Medical Research*.

But most of all, I want to thank my wife Collette. She's done much more than make this book possible.

Any Comments or Suggestions?

I want very much to hear from you if you have any helpful feedback concerning this text. Let me know what you like and don't like about

it, what works well and what can be improved, what might be left out and what needs to be added. If you have access to a computer network like Bitnet or Internet, write me at my Internet address: statprof@scott.skidmore.edu. Or you can write me by "snail mail" at the following address:

William Fox
Department of Sociology, Anthropology,
and Social Work
Skidmore College
Saratoga Springs, NY 12866

One final word before we start learning statistics: I had a great time writing this book. I hope you enjoy it just as much.

Bill Fox
Skidmore College

Table of Contents

Preface	xiii
----------------------	------

Part I: Introduction and Univariate Analyses

Chapter 1: Statistics and Variables	3
1.1 An Overview of Statistics.....	3
1.2 Samples and Populations.....	6
1.3 Variables	7
1.4 Levels of Measurement.....	8
1.5 Continuous and Discrete Variables	12
1.6 Aggregate Data	13
1.7 The General Social Survey	15
1.8 The USA Data Set	16
1.9 MicroCase.....	17
1.10 Ideas and Thinking	18
1.11 Playing with Data.....	19
1.12 How We Will Learn Statistics.....	20
1.13 Summing Up Chapter 1.....	21
Chapter 2: Frequency and Percentage Distributions	25
2.1 Frequency Distributions.....	25
2.2 Percentage Distributions	27
2.3 Cumulative Distributions.....	29
2.4 Creating Sensible and Well-Formatted Tables.....	31
2.5 Pie Charts and Bar Graphs.....	33
2.6 Outliers	36
2.7 Opening a Data File in MicroCase	39

2.8	Univariate Analysis Using MicroCase	43
2.9	Mapping Ecological Variables.....	50
2.10	Getting Help and Escaping in MicroCase.....	55
2.11	Leaving MicroCase.....	56
2.12	Summing Up Chapter 2.....	56
Chapter 3: Codebooks and Data Manipulation.....		59
3.1	Codebooks.....	59
3.2	Selecting Subsets of Cases.....	64
3.3	Collapsing Variables.....	66
3.4	Guidelines for Collapsing Variables.....	71
3.5	Excluding Missing Data.....	72
3.6	Fully Ordering Values of Ordinal Variables.....	75
3.7	Summing Up Chapter 3.....	76
Chapter 4: Averages and Measures of Variation.....		77
4.1	Mode.....	77
4.2	Median.....	79
4.3	Mean.....	84
4.4	Which to Use—Mode, Median, or Mean?.....	87
4.5	Measures of Variation: Variances and Standard Deviations.....	90
4.6	Shapes of Distributions.....	95
4.7	The Normal Distribution.....	96
4.8	Standard Scores (Z-Scores).....	98
4.9	Finding Averages and Measures of Variation Using MicroCase.....	99
4.10	Using Univariate Statistics with Aggregate Data.....	102
4.11	Some Cautions Using Univariate Statistics.....	103
4.12	Summing Up Chapter 4.....	106
 Part II: Bivariate Analyses		
Chapter 5: Bivariate Tabular Analysis.....		111
5.1	Bivariate Frequency Tables.....	112
5.2	Bivariate Percentage Tables.....	115
5.3	How to Read Percentage Tables.....	117
5.4	Tabular Analysis Using MicroCase.....	119
5.5	Positive, Negative, and Curvilinear Relationships.....	124
5.6	Format Conventions for Bivariate Tables.....	127
5.7	A Caution about Bivariate Tables Based on Small Ns.....	130
5.8	Association Does <i>Not</i> Imply Causation.....	131
5.9	Summing Up Chapter 5.....	132

Chapter 6: The Chi-Square Test of Statistical Significance	135
6.1 The Logic of Tests of Statistical Significance	135
6.2 The Chi-Square Test	137
6.3 The Chi-Square Test Using MicroCase	143
6.4 Problems with Expected Frequencies Less Than 5	144
6.5 Statistical Significance Does Not Mean Substantive Significance.....	145
6.6 Significance Tests on Population Data	146
6.7 Summing Up Chapter 6.....	147
Chapter 7: Measures of Association for Tabular Data	149
7.1 Measures of Association.....	149
7.2 Chi-Square Based Measures for Nominal Variables: C, V, and ϕ	151
7.3 Lambda	154
7.4 Choosing a Nominal Measure of Association	157
7.5 Measures of Association for Ordinal Variables: Gamma	158
7.6 Somers' Dyx and Dxy	161
7.7 Tau-b and Tau-c.....	164
7.8 MicroCase's Display of Ordinal Measures	166
7.9 Summing Up Chapter 7.....	167
Chapter 8: Analysis of Variance	171
8.1 The Logic of Analysis of Variance	171
8.2 The Correlation Ratio (E^2)	181
8.3 The ANOVA Table.....	182
8.4 Analysis of Variance Using MicroCase.....	183
8.5 Two-Way Analysis of Variance (and Beyond)	188
8.6 The t-Test for the Difference Between Two Means	189
8.7 Association Does <i>Not</i> Imply Causation.....	189
8.8 Summing Up Chapter 8.....	190
Chapter 9: Regression and Correlation	193
9.1 Scatterplots	195
9.2 Regression	201
9.3 The Correlation Coefficient (r)	205
9.4 Interpretation of r^2 as the Proportion of Variation Explained	208
9.5 Association <i>Still</i> Does Not Imply Causation	210
9.6 Linear and Nonlinear Relationships.....	210
9.7 Test of Significance for a Correlation Coefficient	211
9.8 Scatterplots Using MicroCase.....	213
9.9 Regression Coefficients and Correlations Using MicroCase	218
9.10 Summing Up Chapter 9.....	221

Part III: Multivariate and Advanced Analyses

Chapter 10: Multivariate Tabular Analysis	227
10.1 The Logic of Causal Relationships	227
10.2 Spurious Relationships	229
10.3 Some Terminology	232
10.4 Examples of Spurious Relationships	233
10.5 Replication	234
10.6 Somewhere Between Explanation and Replication	235
10.7 Specification	237
10.8 Suppressor Variables	237
10.9 Controlling for an Intervening Variable	240
10.10 Partial Gamma	241
10.11 Summing Up Elaboration	242
10.12 Multivariate Tabular Analysis Using MicroCase	243
10.13 The Relationship of Multivariate Analysis to Experimental Design	250
10.14 Summing Up Chapter 10	251
Chapter 11: Multiple Regression and Correlation	253
11.1 Extending the Regression Model	253
11.2 The Partial Correlation Coefficient	257
11.3 Extending Partial Correlations to Additional Control Variables	261
11.4 Partial Correlations Using MicroCase	262
11.5 The Multiple Correlation Coefficient	265
11.6 Multiple Correlation Using MicroCase	267
11.7 Significance Tests for Partial and Multiple Correlations	270
11.8 Dummy Variables	272
11.9 Summing Up Chapter 11	275
Chapter 12: A Glance at More Advanced Statistical Techniques	277
12.1 Analysis of Covariance	277
12.2 Curve Fitting	278
12.3 Regression Models	279
12.4 Factor Analysis	281
12.5 Using Existing Data Files	282
12.6 Creating New Data Files	283
12.7 File Management Features	284
12.8 Editing Existing Files	285
12.9 Listing or Printing Variable Values	285
12.10 Miscellaneous MicroCase Features	286
12.11 Summing Up Chapter 12	286

Appendices

Appendix A: The General Social Survey	291
Appendix B: What You Need to Know about Computers.....	297
Appendix C: MicroCase Data Sets	305
Appendix D: Tables	309
Table 1: The Chi-Square Distribution.....	309
Table 2: The F Distribution	310
Glossary	313
Bibliography	323
Index.....	327

PART I Introduction and Univariate Analyses

**CHAPTER 1
Statistics and Variables**

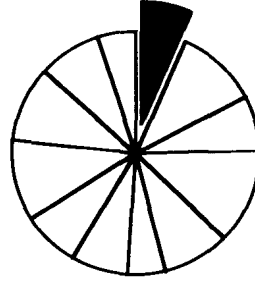
**CHAPTER 2
Frequency and Percentage
Distributions**

**CHAPTER 3
Codebooks and Data
Manipulation**

**CHAPTER 4
Averages and Measures of
Variation**

CHAPTER 1

Statistics and Variables



In this introductory chapter I will describe what statistics is, what variables are, and what data are like. I'll then offer an overview of how we will go about learning basic statistics using a friendly computer program called MicroCase to analyze interesting sets of data for the 50 states and a national sample of Americans.

1.1 An Overview of Statistics

Sometimes we use the word *statistics* to refer to numbers that summarize information quantitatively. We say that Joe DiMaggio had "good stats" with his lifetime batting average of .325, or that statistics show that violent crime increased 5.5 per cent last year, or that the average attendance at Yankee Stadium was 32,505 per game last summer, or that 56 per cent of Americans surveyed think the President is doing a good job. These sorts of quantitative, numerical summaries are indeed statistics and we encounter them often. Modern societies float in a sea of statistics. And although we often complain about statistical information ("They're trying to reduce everything to a number!"), we know we could not function very well without statistics. We need quantitative summaries—statistics—to understand the world clearly and to make decisions intelligently.

But *statistics* also means the *methods* used to calculate these summary numbers and to generalize from them. In this sense statistics is the set of procedures used to calculate batting averages or describe crime trends or determine average attendance or generalize from a sample of Americans to the entire electorate. In this text I'll mainly use statistics in this second sense to mean methods for