

Statistics for the Social Sciences

With Computer Applications

Anthony Walsh

COMPLIMENTARY
PROFESSIONAL COPY



Harper & Row, Publishers



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
Copyright © 1990 by Harper & Row, Publishers, Inc.

All rights reserved. Printed in the United States of America. No part of this book may be used or reproduced in any manner whatsoever without written permission, except in the case of brief quotations embodied in critical articles and reviews. For information address Harper & Row, Publishers, Inc., 10 East 53rd Street, New York, NY 10022.

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

300'.01'5195—dc20

89-26814
CIP

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

PREFACE xiii

CHAPTER 1 INTRODUCTION TO STATISTICAL ANALYSIS 1

- 1.1 Why Study Statistics 1
- 1.2 Thinking Statistically 2
- 1.3 Descriptive and Inferential Statistics 3
 - 1.3.1 Descriptive Statistics 3
 - 1.3.2 Inferential Statistics 4
 - 1.3.3 Statistics and Error 5
 - 1.3.4 Parametric and Nonparametric Statistics 6
 - 1.3.5 Operationalization 6
- 1.4 Measurement 7
 - 1.4.1 Nominal Level 8
 - 1.4.2 Ordinal Level 8
 - 1.4.3 Interval Level 9
 - 1.4.4 Ratio Level 9
- 1.5 Organization of This Book 10
- 1.6 Summary 11
- References 13

CHAPTER 2 DESCRIBING DATA 14

- 2.1 Presentation and Summarization of Data 14
 - 2.1.1 Types of Frequency Distributions 15

vi CONTENTS

2.1.2	Interpreting the Printout	16
2.1.3	Frequency Distribution of Grouped Data	16
2.1.4	Limits, Sizes, and Midpoints of Class Intervals	18
2.1.5	Bar Charts and Histograms	19
2.1.6	Frequency Polygon	21
2.1.7	Pie Charts	22
2.1.8	Advantage and Disadvantage of Grouping Data	22
2.1.9	Histogram with Ungrouped Data	23
2.1.10	Line Graphs	25
2.2	Numerical Summation of Data	26
2.2.1	Proportions, Percentages, and Ratios	26
2.2.2	Rates	28
2.3	Summary	29
	Appendix	31
	References	32

CHAPTER 3 CENTRAL TENDENCY AND DISPERSION 33

3.1	Measures of Central Tendency	33
3.1.1	Mode	34
3.1.2	Median	35
3.1.3	Mean	36
3.1.4	A Computer Example	38
3.1.5	Choosing a Measure of Central Tendency	40
3.2	Measures of Dispersion	42
3.2.1	Range	42
3.2.2	Standard Deviation	43
3.2.3	Index of Qualitative Variation	49
3.3	Summary	50
	Appendix	53
	Reference	53

CHAPTER 4 PROBABILITY AND THE NORMAL CURVE 54

4.1	Probability	54
4.1.1	The Multiplication Rule	55
4.1.2	The Addition Rule	56
4.1.3	Theoretical Probability Distributions	58
4.2	The Normal Curve	61
4.2.1	Different Kinds of Curves	64
4.2.2	The Standard Normal Curve	65
4.2.3	The Practical Value of the Normal Curve: The z Score	68
4.2.4	Finding the Area Under the Curve Between Two Scores	72
4.2.5	The Normal Curve and Probability	74
4.3	Summary	76
	Reference	78

CHAPTER 5 SAMPLING AND THE SAMPLING DISTRIBUTION 79

- 5.1 Sampling 79
 - 5.1.1 Simple Random Sampling 80
 - 5.1.2 Stratified Random Sampling 81
- 5.2 The Sampling Distribution 82
 - 5.2.1 Standard Error of the Sampling Distribution 86
 - 5.2.2 The Central limit Theorem 87
 - 5.2.3 Point and Interval Estimates 88
 - 5.2.4 Confidence Intervals and Alpha Levels 89
 - 5.2.5 Calculating Confidence Intervals 90
 - 5.2.6 Interval Estimates for Proportions 92
 - 5.2.7 Estimating Sample Size 93
 - 5.2.8 Estimating Sample Size for Proportions 94
- 5.3 Summary 95
- Appendix 97

CHAPTER 6 HYPOTHESIS TESTING: INTERVAL/RATIO DATA 99

- 6.1 Parametric Statistics 99
- 6.2 The Logic of Hypothesis Testing 99
- 6.3 The Evidence and Statistical Significance 101
- 6.4 Errors in Hypothesis Testing 103
- 6.5 One Sample z Test 105
- 6.6 Testing Hypotheses About Two Population Means 107
- 6.7 The t Test 109
 - 6.7.1 Degrees of Freedom 110
 - 6.7.2 The t Distribution 110
 - 6.7.3 Computing t with Small Samples 111
 - 6.7.4 Directional Hypotheses: One- and Two-tailed Tests 112
 - 6.7.5 t Test for Correlated Means 114
 - 6.7.6 The Effect of Sample Variability on Significance Testing 116
 - 6.7.7 Large-Sample t Test: A Computer Example 117
 - 6.7.8 Calculating t with Unequal Variances 119
 - 6.7.9 A Note on the Significance Test Controversy 120
- 6.8 Summary 121
- Appendix 123
- References 123

CHAPTER 7 ONE-WAY ANALYSIS OF VARIANCE 124

- 7.1 The Basic Logic of ANOVA 125
 - 7.1.1 The Idea of Variance 125
 - 7.1.2 The Advantage of ANOVA over Multiple Tests 127
 - 7.1.3 The F Distribution 127
- 7.2 An Example of ANOVA 128

7.3	Determining Statistical Significance: Mean Square and the F Ratio	132
7.4	Multiple Comparisons: The Tukey HSD Test	134
7.4.1	Computing Eta Squared	135
7.4.2	Eta Squared and the t Test	136
7.4.3	The ANOVA Table	136
7.5	A Computer Example of One-Way ANOVA	136
7.5.1	Interpreting the Printout	137
7.5.2	The Scheffe Test	139
7.6	Summary	140
	Appendix	143

CHAPTER 8 TWO-WAY ANALYSIS OF VARIANCE 144

8.1	Determining Statistical Significance	147
8.1.1	Significance Levels	148
8.1.2	Tukey Test	150
8.1.3	Eta Squared	150
8.2	The Concept of Interaction	151
8.2.1	An Example of a Significant Interaction Effect from the Computer	153
8.2.2	A Further Look at Interaction	157
8.2.3	Illustrating Interaction	159
8.3	Other Factorial Designs and Higher-Order Interactions	159
8.4	Summary	160
	Appendix	163
	Reference and Suggested Reading	164

CHAPTER 9 HYPOTHESIS TESTING WITH CATEGORICAL DATA: CHI-SQUARE 165

9.1	Table Construction	166
9.1.1	Putting Percentages in Tables	167
9.2	The Chi-Square Distribution	172
9.2.1	Yate's Correction for Continuity	173
9.2.2	Phi	173
9.2.3	Chi-Square and Goodness of Fit	174
9.2.4	A Computer Example of Chi-Square	176
9.3	The Odds Ratio	179
9.4	Chi-Square and Sample Size	181
9.4.1	How Big Should a Sample Be?	182
9.5	Kruskal-Wallis One-Way Analysis of Variance	182
9.5.1	Computing Eta Squared from H	184
9.6	Summary	184
	Appendix	187
	Reference	187

CHAPTER 10 MEASURES OF ASSOCIATION WITH NOMINAL AND ORDINAL DATA 188

- 10.1 The Idea of Association 188
 - 10.1.1 Does an Association Exist? 189
 - 10.1.2 What Is the Strength of the Association? 190
 - 10.1.3 What Is the Direction of the Association? 190
 - 10.1.4 Lambda 192
- 10.2 The Concept of Paired Cases 194
- 10.3 A Computer Example of Nominal and Ordinal Measures of Association 196
 - 10.3.1 The Computer Printout 196
 - 10.3.2 Gamma 196
 - 10.3.3 Lambda 200
 - 10.3.4 Somer's d 201
 - 10.3.5 Tau- b 202
 - 10.3.6 Tau- c 203
 - 10.3.7 Which Test of Association Should We Use? 203
- 10.4 Measures of Association Based on Chi-Square 204
 - 10.4.1 Contingency Coefficient 205
 - 10.4.2 Cramer's V 206
 - 10.4.3 Spearman's Rank Order Correlation 206
- 10.5 Summary 208
- Appendix 212
- Reference and Suggested Reading 212

CHAPTER 11 ELABORATION OF TABULAR DATA 213

- 11.1 Causal Analysis 213
- 11.2 Criteria for Causality 213
 - 11.2.1 Association 213
 - 11.2.2 Temporal Order 214
 - 11.2.3 Spuriousness 214
 - 11.2.4 Necessary Cause 215
 - 11.2.5 Sufficient Cause 215
 - 11.2.6 Necessary and Sufficient Cause 216
- 11.3 Statistical Demonstration of Cause-and-Effect Relationships 216
- 11.4 Multivariate Contingency Analysis 218
 - 11.4.1 Explanation and Interpretation 221
 - 11.4.2 Replication 222
 - 11.4.3 Specification 223
 - 11.4.4 Illustrating Elaboration Outcomes 223
 - 11.4.5 Controlling for One Variable 224
 - 11.4.6 Further Elaboration: Two Control Variables 226
 - 11.4.7 Partial Gamma 228
 - 11.4.8 Problems of Bivariate Elaboration 230
- 11.5 Summary 231
- Appendixes 234
- References and Suggested Readings 236

CHAPTER 12 BIVARIATE CORRELATION AND REGRESSION 237

- 12.1 Preliminary Investigation: The Scattergram 239
 - 12.1.1 Computing b and a 242
 - 12.1.2 The Intercept 243
 - 12.1.3 The Pearson Correlation Coefficient (r) 244
 - 12.1.4 Covariance and Correlation 246
 - 12.1.5 The Coefficient of Determination and PRE 248
 - 12.1.6 Partitioning r Squared and Sum of Squares 250
 - 12.1.7 Standard Error of the Estimate 251
 - 12.1.8 Standard Error of r 252
 - 12.1.9 Significance Testing for Pearson's r 253
 - 12.1.10 The Interrelationship of b , r , and β 254
 - 12.1.11 Summarizing Prediction Formulas 255
- 12.2 A Computer Example of Bivariate Correlation and Regression 256
 - 12.2.1 Interpreting the Output 256
- 12.3 Summary 260
- Appendix 262
- References and Suggested Readings 262

CHAPTER 13 MULTIVARIATE CORRELATION AND REGRESSION 263

- 13.1 Partial Correlation 263
 - 13.1.1 The Logic of Partial Correlation 264
 - 13.1.2 Computing Partial Correlations 265
 - 13.1.3 A Computer Example 265
 - 13.1.4 Second-Order Partials: Controlling for Two Independent Variables 269
 - 13.1.5 The Multiple Correlation Coefficient 271
- 13.2 Multiple Regression 273
 - 13.2.1 Extending the Regression Equation 273
 - 13.2.2 The Unstandardized Partial Slope 274
 - 13.2.3 The Standardized Partial Slope 276
 - 13.2.4 A Computer Example of Multiple Regression 278
 - 13.2.5 Interpreting the Printout 280
 - 13.2.6 A Visual Representation of Multiple Regression 283
 - 13.2.7 Reporting Results 284
- 13.3 Summary 284
- Appendixes 288
- References and Suggested Readings 290

CHAPTER 14 FURTHER EXPLORING CORRELATION AND REGRESSION 291

- 14.1 Dummy Variable Regression 291
 - 14.1.1 Dummy Variable Regression and Interaction 295

14.2	The Problem of Multicollinearity	297
14.3	Measurement Errors	298
14.4	Specification Errors	299
14.5	Nonlinear Correlation: Another Look at Eta Squared	299
14.5.1	Eta Squared and Linearity	301
14.5.2	Do the Data Significantly Depart from Linearity?	303
14.6	The Problem of Heteroscedasticity	304
14.7	Analysis of Residuals and the Outlier Problem	304
14.8	The Problem of Restricted Range	307
14.9	Point Biserial Correlation	309
14.10	Summary	310
	Appendixes	313
	References and Suggested Reading	315

CHAPTER 15 A BRIEF INTRODUCTION TO SOME ADVANCED STATISTICS 317

15.1	Path Analysis	317
15.2	Logit Regression	322
15.2.1	The Rationale for the Use of Logit	322
15.2.2	An Example of Logit Regression	322
15.2.3	A Computer Example of Logit Regression	324
15.2.4	The SAS Differences	327
15.3	Factor Analysis	329
15.3.1	The Correlation Matrix	331
15.3.2	The Factor Matrix	331
15.3.3	Communality	332
15.3.4	Extraction of Initial Factors	332
15.3.5	Rotation of Factors	334
15.3.6	Naming the Factors	335
15.3.7	Using Factor Analysis to Determine Scale Reliability	336
15.4	Summary	337
	Appendixes	338
	References and Suggested Readings	339

CHAPTER 16 INTRODUCTION TO THE COMPUTER AND SPSSx AND SAS LANGUAGES 340

16.1	The SPSSx Language	341
16.2	The SAS Language	345

APPENDIX A AREAS UNDER THE NORMAL CURVE: VALUES OF $A(z)$ BETWEEN ORDINATE AT MEAN (Y_0) AND ORDINATE AT z 349

APPENDIX B VALUES OF t FOR SELECTED PROBABILITIES 351

APPENDIX C VALUES OF F FOR UPPER 5% PROBABILITY (OR 5% AREA UNDER F DISTRIBUTION CURVE) 353

APPENDIX C (continued) VALUES OF F FOR UPPER 1% PROBABILITY (OR 1% AREA UNDER F DISTRIBUTION CURVE) 355

APPENDIX D TABULAR VALUES OF Q_α FOR TUKEY'S HSD TEST 357

APPENDIX E VALUES OF χ^2 FOR SELECTED PROBABILITIES 359

GLOSSARY 361

INDEX 369

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.