



Fundamentals of Social Statistics

Third Edition

KIRK ELIFSON
RICHARD P. RUNYON
AUDREY HABER

Fundamentals of Social Statistics

THIRD EDITION

Kirk W. Elifson

Georgia State University

Richard P. Runyon and Audrey Haber



Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Lisbon London Madrid
Mexico City Milan New Delhi Seoul Singapore Sydney Taipei Toronto

McGraw-Hill

A Division of The McGraw-Hill Companies

FUNDAMENTALS OF SOCIAL STATISTICS

Copyright © 1998 by The McGraw-Hill Companies, Inc. All rights reserved. Previous edition(s) © 1990 and 1982. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOC/DOC 9 0 9 8 7

ISBN 0-07-021579-0

Publisher: *Jane Vaicunas*

Sponsoring editor: *Jill Gordon*

Editorial coordinator: *Amy Smeltzley*

Marketing manager: *Sally Constable*

Project manager: *Christina Thornton-Villagomez*

Senior production supervisor: *Karen Thigpen*

Design manager: *Kiera Cunningham*

Compositor: *Interactive Composition Corporation*

Typeface: *10.5/12 Times Roman*

Printer: *R. R. Donnelley & Sons Company*

Library of Congress Cataloging-in-Publication Data

Elifson, Kirk W.

Fundamentals of social statistics / Kirk W. Elifson. —3rd ed.

p. cm.

Includes index.

ISBN 0-07-021579-0

1. Social sciences—Statistical methods. 2. Statistics.

I. Runyon, Richard P. II. Haber, Audrey. III. Title.

HA29.E482 1998

519.5—dc21

97-25763

<http://www.mhhe.com>

Fundamentals of Social Statistics

PREFACE

This book represents the collective wisdom of nearly a thousand statistics students whom I have taught. Much like myself when I was an undergraduate, students appreciate engaging, relevant, and concrete examples of statistics that are placed into a meaningful context. Their ability to learn is frequently hampered by computational steps that are omitted, the lack of clear definitions of new concepts, and the failure to include relevant presentations or the logic behind the techniques they are learning. My first encounter with statistics left me confused and floundering. I was unable to express the “faults” I found in the assigned text. This led me to write a text that students could more easily understand.

Students *can* learn statistics and have a satisfying experience at the same time, provided the material is presented appropriately. My students and I have worked hard to see that you will engage in some real problem solving, and not be asked to manipulate numbers within a meaningless context. A primary reason for writing this book was to create a text that is not a burden for students to use, one that could be appreciated even by those students who initially feared statistics. I believe that you will find this text to be very clear and useful.

Student and faculty response to the first and second editions of this textbook has been very satisfying. *Fundamentals of Social Statistics*, is widely used in social and behavioral science departments in the United States and in several English-speaking countries throughout the world. The publication of the third edition of a textbook is a great source of satisfaction to an author. In 1982, I was asked to rewrite Richard P. Runyon and Audrey Haber’s best-selling behavioral science textbook, *Fundamentals of Behavioral Statistics*; I never dreamed that 15 years later I would be completing a third edition. I have sought to develop a text that meets the specific needs of those in sociology and in the related fields of criminology, family and urban studies, political science, and social welfare.

The current edition features a broader emphasis on the interpretation of data than the previous editions. The book now incorporates computer analysis from the Statistical Package for the Social Sciences (SPSS) along with a complete explanation of the interpretation and application of data. Case Examples are prominently displayed throughout the text. These examples are excerpted from a rich and varied selection of contemporary research in the social sciences and cover topics such as HIV/AIDS, health, gender, race, welfare, and crime. Statistics are presented as part of the real world, not as figures in a vacuum. The examples and raw data are presented and analyzed using the statistical techniques featured in that particular chapter. The Case Examples bring cohesion to statistical topics that often are perceived as separate and discrete.

Statistics in Action presents recently published material or research, including design, research goals, and raw data. The boxed examples first demonstrate statistical analysis at a level appropriate to student understanding at that point in the book, then offer additional data for analysis followed by solutions. The same studies are often

repeated later in the text with more advance analyses, so students can observe the continuity and connectedness of statistical analysis.

The chapters on contingency tables, probability, and regression have undergone considerable revision; regression now includes a discussion on regression to the mean. The treatment of analysis of variance includes a more complete discussion of the *F*-test, and new sections dealing with hypothesis testing have been added to Chapters 13 and 14. Throughout the book, the examples have been updated and most exercises address current social issues, which provide students with meaningful, real-life knowledge. You also will find that the presentation of complex concepts has been clarified and simplified.

Features of the Book

A number of important features are included in this text. Key terms appear in boldface while terms that require emphasis are in italics. Visual devices—charts, figures, and graphs—help ensure maximum understanding by the student. New equations are discussed fully rather than applied mechanically, and ample examples of equations are provided initially to ensure comprehension.

Each chapter begins with an outline of the contents to provide the student with an overview of the material; each chapter ends with a glossary of key terms and summary. A completely updated student workbook has been designed by Philip Luck to provide a review and feedback for the student. The workbook incorporates study objectives for each chapter; a programmed review of terms, symbols, and concepts; selected computational exercises for application purposes; and sample test questions.

The Appendixes include a review of basic mathematics, a glossary of symbols keyed to the first page on which the symbols appear, a complete set of tables accompanied by explicit directions for their use, a master glossary of key terms and references. Other features of the book are a simplified chronological listing of equations most frequently used (inside the front and back covers), and a comprehensive index.

Acknowledgments

I owe a debt of gratitude to the many students who have participated in the revision of this edition and to the many colleagues who have contributed in ways that will always be appreciated.

Several persons at McGraw-Hill were particularly helpful in supporting me throughout this revision. Philip Butcher and Jill Gordon both encouraged and supported me as sponsoring editors at different phases of the project. Along with Amy Smeltzley, they kept me forging ahead while offering guidance when it was needed. Christina Thornton-Villagomez served as project manager who together with Charles Olson masterfully copyedited the third edition.

Charles Janet of Georgia State University provided a comprehensive review of earlier drafts. His assistance was invaluable and his keen eye and insight are much

appreciated. Several other reviewers read all or part of the manuscript and offered important suggestions: Glenn Deane, State University of New York, Albany; Robin Franck, Southwestern College; Duane R. Monette, Northern Michigan University; and J. D. Robson, University of Arkansas at Little Rock. Thanks also must go to Dawn Baunach of Georgia State University and Edward Bell of Brescia College for their welcome suggestions.

Finally, my daughters Kristin and Shelley have been influential from the first edition to the present. My greatest supporters were undoubtedly Mam, Pap, and Andy Sheldon. Most important, Claire Sterk got me started, kept me going, and was with me as a pal and scholar from beginning to end.

Kirk W. Elifson
Atlanta, Georgia

CONTENTS

PART I Introduction

1	The Definition of Statistical Analysis	3
1.1	What Is Statistics?	3
1.2	Definitions of Terms Commonly Used in Statistics	6
1.3	Descriptive Statistics	9
1.4	Inferential Statistics	10
1.5	The Goals of Research	12
1.6	A Word to the Student	14
	Summary	15
	Terms to Remember	16
	Exercises	16
2	Basic Mathematical Concepts	19
2.1	Introduction	19
2.2	Summation Rules	20
2.3	Types of Numbers and Scales	24
2.4	Continuous and Discrete Variables	30
2.5	Rounding Numbers	31
2.6	Ratios	33
2.7	Proportions, Percentages, and Rates	35
	Summary	38
	Terms to Remember	40
	Exercises	40

PART II Descriptive Statistics

3	Frequency Distributions and Graphing Techniques	49
3.1	Grouping of Data	49
3.2	Graphing Techniques	55
3.3	Misuse of Graphing Techniques	55
3.4	Nominally Scaled Variables	56
3.5	Ordinally Scaled Variables	57
3.6	Interval- and Ratio-Scaled Variables	59

3.7	Forms of Frequency Curves	62
3.8	Other Graphic Representations	63
	Summary	67
	Terms to Remember	69
	Exercises	70
4	Percentiles	76
4.1	Introduction	76
4.2	Cumulative Percentiles and Percentile Rank	77
4.3	Percentile Rank and Reference Group	82
4.4	Centiles, Deciles, and Quartiles	82
	Summary	86
	Terms to Remember	86
	Exercises	86
5	Measures of Central Tendency	90
5.1	Introduction	90
5.2	The Arithmetic Mean (\bar{X})	92
5.3	The Median (Md_n)	99
5.4	The Mode (M_o)	102
5.5	Comparison of Mean, Median, and Mode	102
5.6	The Mean, Median, Mode, and Skewness	104
	Summary	107
	Terms to Remember	108
	Exercises	108
6	Measures of Dispersion	113
6.1	Introduction	113
6.2	The Range	114
6.3	The Interquartile Range	115
6.4	The Mean Deviation	117
6.5	The Variance (s^2) and Standard Deviation (s)	118
6.6	Interpretation of the Standard Deviation	124
	Summary	125
	Terms to Remember	126
	Exercises	126
7	The Standard Deviation and the Standard Normal Distribution	130
7.1	Introduction	130
7.2	The Concept of Standard Scores	130
7.3	The Standard Normal Distribution	134
7.4	Characteristics of the Standard Normal Distribution	135
7.5	Transforming Raw Scores to z Scores	136

7.6	Illustrative Examples	139
7.7	The Standard Deviation as an Estimate of Error and Precision	144
7.8	Interpreting the Standard Deviation	145
	Summary	147
	Terms to Remember	148
	Exercises	148
8	An Introduction to Contingency Tables	152
8.1	Introduction	152
8.2	Dependent and Independent Variables	153
8.3	The Bivariate Contingency Table	153
8.4	Percentaging Contingency Tables	155
8.5	Existence, Direction, and Strength of a Relationship	160
8.6	Introduction to Measures of Association for Contingency Tables	162
8.7	Nominal Measures of Association: Lambda	165
8.8	Ordinal Measures of Association	169
8.9	Goodman and Kruskal's Gamma	171
8.10	Somer's d	175
8.11	Kendall's tau- b	176
	Summary	177
	Terms to Remember	178
	Exercises	179
9	Correlation	184
9.1	The Concept of Correlation	184
9.2	Calculation of Pearson's r	187
9.3	Pearson's r and z Scores	189
9.4	The Correlation Matrix	191
9.5	Interpreting Correlation Coefficients	194
9.6	Spearman's Rho (r_s)	198
	Summary	201
	Terms to Remember	202
	Exercises	202
10	Regression and Prediction	210
10.1	Introduction to Prediction	210
10.2	Linear Regression	211
10.3	Residual Variance and Standard Error of Estimate	224
10.4	Explained and Unexplained Variation	226
10.5	Correlation and Causation	229
	Summary	232
	Terms to Remember	232
	Exercises	233

11	Multivariate Data Analysis	237
11.1	Introduction	237
11.2	The Multivariate Contingency Table	238
11.3	Partial Correlation	241
11.4	Multiple Regression Analysis	247
	Summary	254
	Terms to Remember	254
	Exercises	255
PART III	Inferential Statistics: Parametric Tests of Significance	
12	Probability	265
12.1	Introduction to Probability	265
12.2	The Concept of Randomness	267
12.3	Approaches to Probability	269
12.4	Formal Properties of Probability	272
12.5	Probability and Continuous Variables	287
12.6	Probability and the Normal-Curve Model	287
12.7	One- and Two-Tailed p Values	289
	Summary	291
	Terms to Remember	293
	Exercises	294
13	Introduction to Statistical Inference	299
13.1	Why Sample?	299
13.2	The Concept of Sampling Distributions	301
13.3	Testing Statistical Hypotheses: Level of Significance	307
13.4	Testing Statistical Hypotheses: Null Hypothesis and Alternative Hypothesis	309
13.5	Testing Statistical Hypotheses: The Two Types of Error	313
13.6	A Final Word of Caution	318
	Summary	319
	Terms to Remember	320
	Exercises	320
14	Statistical Inference and Continuous Variables	323
14.1	Introduction	323
14.2	Sampling Distribution of the Mean	326
14.3	Testing Statistical Hypotheses: Population Mean and Standard Deviation Known	329
14.4	Estimation of Parameters: Point Estimation	334

14.5	Testing Statistical Hypotheses with Unknown Parameters: Student's t	337
14.6	Estimation of Parameters: Interval Estimation	342
14.7	Confidence Intervals and Confidence Limits	343
14.8	Test of Significance for Pearson's r : One-Sample Case	349
14.9	Test of Significance for Goodman and Kruskal's Gamma	351
	Summary	352
	Terms to Remember	353
	Exercises	354
15	An Introduction to the Analysis of Variance	358
15.1	Multigroup Comparisons	358
15.2	The Concept of Sums of Squares	359
15.3	Obtaining Variance Estimates	362
15.4	Relation between the t -Ratio and the F -Ratio	365
15.5	Fundamental Concepts of Analysis of Variance	366
15.6	Assumptions Underlying Analysis of Variance	367
15.7	An Example Involving Three Groups	367
	Summary	371
	Terms to Remember	372
	Exercises	372

PART IV Inferential Statistics: Nonparametric Tests of Significance

16	Statistical Inference with Categorical Variables: Chi-Square and Related Measures	379
16.1	Introduction	379
16.2	The χ^2 One-Variable Case	382
16.3	The χ^2 Test of the Independence of Variables	385
16.4	Limitations in the Use of χ^2	388
16.5	Nominal Measures of Association Based on χ^2	391
	Summary	395
	Terms to Remember	396
	Exercises	396
17	Statistical Inference: Ordinally Scaled Variables	402
17.1	Introduction	402
17.2	Mann-Whitney U -Test	403
17.3	Nonparametric Tests Involving Correlated Samples	408
17.4	The Sign Test	409

17.5	Wilcoxon Matched-Pairs Signed-Rank Test	411
	Summary	414
	Terms to Remember	415
	Exercises	415

PART V Appendixes

A	Review of Basic Mathematics	421
B	Glossary of Symbols	428
C	Tables	432
D	Glossary of Terms	482
E	References	490

PART VI Answers to Selected Exercises

	Answers to Selected Exercises	495
	Index	528

PART I

Introduction

- 1 The Definition of Statistical Analysis**
- 2 Basic Mathematical Concepts**

The Definition of Statistical Analysis

- 1.1 What Is Statistics?**
 - 1.2 Definitions of Terms Commonly Used in Statistics**
 - 1.3 Descriptive Statistics**
 - 1.4 Inferential Statistics**
 - 1.5 The Goals of Research**
 - 1.6 A Word to the Student**
-

1.1 WHAT IS STATISTICS?

Think for a moment of the thousands of incredibly complex things you do during the course of a day. You are absolutely unique. No one else possesses your physical features, your intellectual makeup, your personality characteristics, and your value system. Yet, like billions of others of your species, you are among the most finely tuned and enormously sophisticated statistical instruments ever devised by natural forces. Every moment of your life provides testimony to your ability to receive and process a variety of information and then to use this information instantly to determine possible courses of action.

To illustrate, imagine you are driving in heavy traffic. You are continuously observing the road conditions, noting the speed of cars in front of you compared to your own speed, the position and rate of approach of vehicles to your rear, and the presence of automobiles in the oncoming lane. If you are an alert driver, you are constantly summarizing this information—usually without words or even awareness.

Imagine next that, without warning, the driver of the car in front of you suddenly jams on the brakes. In an instant you must act upon this prior information. You must brake the car, turn left, turn right, or pray. Your brain instantly considers alternative courses of action: If you jam on the brakes, what is the possibility that you will stop in time? Is the car behind you far enough away to avoid a rear-end collision? Can you

avoid an accident by turning into the left lane or onto the right shoulder? Most of the time your decision is correct. Consequently, most of us live to a ripe old age.

In this situation, as in many others during the course of a lifetime, you have accurately assessed the possibilities and taken the right course of action. And you make such decisions thousands of times each and every day of your life. For this reason, you should regard yourself as a mechanism for making statistical decisions. In this sense you are already a statistician.

In daily living, our statistical functioning is usually informal and loosely structured. Consider the times you have contemplated the *likelihood* of someone you are attracted to, but do not know well, rejecting your invitation to have lunch. We *behave* statistically, although we may be totally unaware of the formal laws of probability, which will be presented in Chapter 12.

In this course, we will attempt to provide you with some of the procedures for collecting and analyzing data, and making decisions or inferences based on these analyses. Since we will frequently be building upon your prior experiences, you will often feel that you have made a similar analysis before: “Why, I have been calculating averages almost all my life—whenever I determine my test average in a course or the mileage my car gets,” and “I compute range whenever I figure how much my time varies on my favorite two-mile jog.” If you constantly draw upon your previous knowledge and relate course materials to what is familiar in daily life, statistics need not, and should not, be the bugaboo it is often painted to be.

What, then, is statistics all about? To many people, statistics is merely a collection of numerical facts that are expressed in terms of a summarizing statement such as: “Seven out of 10 doctors prescribe the pain reliever that is contained in Product X,” or “During his current hitting streak, Ken Griffey, Jr., hit safely in 20 of his last 45 times at bat,” or “During the Fourth of July weekend, 1986, more than 6 million New Yorkers and visitors participated in the centennial celebration of the unveiling of the Statue of Liberty.”

However, this is not the way statistics is defined by scientists. Rather, **statistics** is a method for dealing with data and involves the organization and analysis of numerical facts or observations that are collected in accordance with a systematic plan. The plan for collecting data is called the **research design**. Broadly speaking, the design of a particular study is structured to provide answers to specific questions.

The material presented in Box 1.1 allows us to answer the following questions: Are smaller percentages of American children living in traditional families today than in the past? What percent of all children now live in traditional families? Is there a difference by race of the child? Ask yourself, are any trends apparent from reviewing the data? What might be some possible explanations for the trends and racial differences that are evident in Box 1.1?

A distinction may be made between the two functions of the statistical method: descriptive statistical techniques and inferential or inductive statistical techniques.

The major purpose of **descriptive statistics** is to present information in a convenient, usable, and understandable form. The statistical information presented in Figure 1.1 and the text of Box 1.1 are examples of statistics used for descriptive purposes. **Inferential or inductive statistics**, on the other hand, is concerned with making inferences about populations that are based on samples taken from those populations.