

THIS IS A USED BOOK

This book was originally distributed as a sample copy by the publisher, for academic review. It was (then) purchased by a used book dealer and resold as used. This allows you a substantial savings. All the chapters and pages are included.

Fundamentals of Social Statistics

Second Edition

Kirk W. Elifson · Richard P. Runyon · Audrey Haber



Fundamentals of Social Statistics

Second Edition

Kirk W. Elifson

Georgia State University

Richard P. Runyon and Audrey Haber

McGraw-Hill Publishing Company

New York St. Louis San Francisco Auckland Bogotá Caracas Hamburg
Lisbon London Madrid Mexico Milan Montreal New Delhi Oklahoma City
Paris San Juan São Paulo Singapore Sydney Tokyo Toronto

FUNDAMENTALS OF SOCIAL STATISTICS

Copyright © 1990, 1982 by McGraw-Hill, Inc. All rights reserved.

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 data base or retrieval system, without the prior written permission of the publisher.

1234567890 DOC DOC 9543210

ISBN 0-07-557292-3

This book was set in Serif by Ruttle, Shaw & Wetherill, Inc.

The editors were Phillip A. Butcher, Bertrand W. Lummus, and Eleanor Castellano; the designer was Chuck Carson;

the production supervisor was Birgit Garlasco.

Cover photograph by Masahiro Sano, courtesy of The Stock Market.

R. R. Donnelley & Sons Company was printer and binder.

Library of Congress Cataloging-in-Publication Data

Elifson, Kirk W., (date).

Fundamentals of social statistics / Kirk W. Elifson, Richard P. Runyon, and Audrey Haber. —2nd ed.

p. cm.

Includes bibliographical references.

ISBN 0-07-557292-3. —ISBN 0-07-019460-2 (instructor's manual)

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

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

HA29.E482 1990

519.5—dc20

89-13055

Fundamentals of Social Statistics

Preface

While an undergraduate, I enrolled in the first of many statistics courses. At that time the course frequently left me and my classmates confused and floundering to the extent that we learned little about statistics; however, I was unable to express the faults I found with the assigned text. The book lacked interesting and relevant examples, omitted computational steps, introduced but did not define new concepts, and seldom presented the relevance or logic of the techniques it purported to teach. I was frequently juggling numbers that were not placed into a meaningful context rather than engaging in any real problem solving.

My first experience with statistics would have been far more pleasant had the instructor assigned a book that was less vulnerable to these criticisms. One of my primary reasons for writing this book was to create a text that is not deficient in these areas. Often, I have found that my experience as a statistics professor has been shaped by my initial statistics course. When a textbook presents the subject clearly, logically, and completely—not in abstract terms, but through concrete examples—students understand better and enjoy the material more. Students *can* learn statistics and have a satisfying experience, provided the material is presented appropriately. I hope that you find this text far more useful than the one I first used.

This book is the second edition of a major revision of the best-selling behavioral science statistics textbook, *Fundamentals of Behavioral Statistics*, by Richard P. Runyon and Audrey Haber. Their book is now in the sixth edition and has been used by over one-quarter of a million students. Rewriting one of the most widely acclaimed behavioral science textbooks for use by social scientists was a challenge I could not resist. I was asked to preserve the strength of the book that had made it such a success with students, professors, and practitioners, yet write a text that would meet the specific needs of those in sociology and the related areas of criminology, family and urban studies, political science, and social welfare.

Student and faculty response to the first edition has been very satisfying. This current edition features a broader emphasis on the interpretation of data. To illustrate, Case Examples have been prominently displayed throughout the text. These examples are excerpted from a rich and varied selection of contemporary research in the social sciences. With each of these Case Examples, raw data are presented and analyzed in terms of the topics featured in a given chapter. Later, they are subjected to more advanced analyses. The effect of

these Case Examples is to bring cohesion to statistical topics that are often perceived as separate and discrete. Statistics in Action, another new feature, presents a recently published article or other research including design, research goals, and raw data. The boxed examples demonstrate statistical analysis at a level appropriate to student understanding, then offer additional data for analysis followed by solutions. The same study is often repeated later in the text with more advanced analyses, so students can observe the continuity and connectedness of statistical analysis.

The chapter on probability has undergone considerable revision, including the addition of tree and Venn diagrams, as well as conditional and joint probabilities. A new chapter, “Statistical Inference with Ordinally Scaled Variables,” has been added. It discusses such topics as the Mann–Whitney *U*-test, nonparametric tests involving correlated samples, the sign test, and the Wilcoxon matched-pairs signed-rank test. A new section on goals of statistical research has been added in Chapter 1, Kendall’s tau-*b* has been added to Chapter 8, and approximately 20 percent of the exercises at the end of each chapter are new.

Features of the Book

A number of important features have been included in this text. Key terms appear in boldface, and terms that require emphasis have been italicized. Visual devices, such as charts, figures, and graphs are also incorporated to ensure minimum understanding by the student. New equations are discussed fully rather than mechanically applied, and ample examples are initially provided to ensure comprehension.

Each chapter begins with a content listing to provide the student with an overview of the included material and ends with a glossary of key terms and chapter summary. A student workbook has been designed to provide review practice and feedback for the student. The workbook incorporates a programmed review of terms, symbols, and concepts; selected computational exercises for application purposes; and test questions. A software statistics disk for IBM microcomputers is available for students to learn statistical concepts by calculating basic statistical measures.

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

Acknowledgments

I owe a debt of gratitude to the many colleagues and professors who have taught me statistics. From my students I have also learned much about statistics and how it should be taught.

Several persons at McGraw-Hill were particularly helpful. Bertrand W. Lummus encouraged and supported me throughout the development of the second edition. Phillip Butcher served as the book's editor as it was being completed. Eleanor Castellano ably coordinated the book's editing and production, and Elaine Honig copyedited the manuscript.

Good reviewers are invaluable to an author. The following persons read all or part of the manuscript and offered valuable suggestions: Esther Heffernan, Edgewood College; James F. Iaccino, Illinois Benedictine College; Tai Shick Kang, SUNY–Buffalo; Edward Nelson, California State University–Fresno; and J. D. Robson, University of Arkansas at Little Rock.

Finally, I wish to thank Claire for her invaluable support and Kristin and Shelley, my daughters, for their love and unfailing interest in the project.

Kirk W. Elifson

Contents

Preface	xi
1 Introduction	
1 The Definition of Statistical Analysis	2
1.1 What Is Statistics?	3
1.2 Definitions of Terms Commonly Used in Statistics	7
1.3 Descriptive Statistics	10
1.4 Inferential Statistics	12
1.5 The Goals of Research	13
1.6 A Word to the Student	15
Summary	16
Terms to Remember	17
Exercises	18
2 Basic Mathematical Concepts	22
2.1 Introduction	23
2.2 Summation Rules	24
2.3 Types of Numbers and Scales	27
2.4 Continuous and Discrete Variables	33
2.5 Rounding	35
2.6 Ratios	37
2.7 Proportions, Percentages, and Rates	38
Summary	43
Terms to Remember	43
Exercises	44
2 Descriptive Statistics	
3 Frequency Distributions and Graphing Techniques	54
3.1 Grouping of Data	54
3.2 Cumulative Frequency and Cumulative Percentage Distributions	55
3.3 Graphing Techniques	59
3.4 Misuse of Graphing Techniques	61
3.5 Nominally Scaled Variables	62
3.6 Ordinally Scaled Variables	63
3.7 Interval- and Ratio-Scaled Variables	63
3.8 Forms of Frequency Curves	69

3.9	Other Graphic Representations	72
	Summary	73
	Terms to Remember	74
	Exercises	75
4	Percentiles	82
4.1	Introduction	83
4.2	Cumulative Percentiles and Percentile Rank	84
4.3	Percentile Rank and Reference Group	90
4.4	Centiles, Deciles, and Quartiles	90
	Summary	92
	Terms to Remember	92
	Exercises	93
5	Measures of Central Tendency	96
5.1	Introduction	97
5.2	The Arithmetic Mean (\bar{X})	99
5.3	The Median (Md_n)	107
5.4	The Mode (M_o)	109
5.5	Comparison of Mean, Median, and Mode	110
5.6	The Mean, Median, Mode, and Skewness	114
	Summary	115
	Terms to Remember	116
	Exercises	116
6	Measures of Dispersion	122
6.1	Introduction	123
6.2	The Range	124
6.3	The Interquartile Range	125
6.4	The Mean Deviation	125
6.5	The Variance (s^2) and Standard Deviation (s)	128
6.6	Interpretation of the Standard Deviation	134
	Summary	134
	Terms to Remember	135
	Exercises	136
7	The Standard Deviation and the Standard Normal Distribution	142
7.1	Introduction	143
7.2	The Concept of Standard Scores	143
7.3	The Standard Normal Distribution	148
7.4	Characteristics of the Standard Normal Distribution	148
7.5	Transforming Raw Scores to z Scores	150
7.6	Illustrative Problems	152
7.7	The Standard Deviation as an Estimate of Error and Precision	158
7.8	Interpreting the Standard Deviation	159

Summary	160
Terms to Remember	160
Exercises	160
8 An Introduction to Contingency Tables	166
8.1 Introduction	167
8.2 Dependent and Independent Variables	167
8.3 The Bivariate Contingency Table	168
8.4 Percentaging Contingency Tables	170
8.5 Existence, Direction, and Strength of a Relationship	172
8.6 Introduction to Measures of Association for Contingency Tables	177
8.7 Nominal Measures of Association: Lambda	178
8.8 Ordinal Measures of Association	182
8.9 Goodman and Kruskal's Gamma	184
8.10 Somer's d	189
8.11 Kendall's tau- b	190
Summary	191
Terms to Remember	192
Exercises	193
9 Correlation	198
9.1 The Concept of Correlation	199
9.2 Calculation of Pearson's r	201
9.3 Pearson's r and z Scores	204
9.4 The Correlation Matrix	206
9.5 Interpreting Correlation Coefficients	208
9.6 Spearman's Rho (r_s)	214
Summary	216
Terms to Remember	216
Exercises	217
10 Regression and Prediction	224
10.1 Introduction to Prediction	225
10.2 Linear Regression	228
10.3 Residual Variance and Standard Error of Estimate	238
10.4 Explained and Unexplained Variation	240
10.5 Correlation and Causation	243
Summary	245
Terms to Remember	246
Exercises	246
11 Multivariate Data Analysis	253
11.1 Introduction	254
11.2 The Multivariate Contingency Table	255
11.3 Partial Correlation	258

11.4	Multiple Regression Analysis	264
	Summary	272
	Terms to Remember	273
	Exercises	274
3	Inferential Statistics: Parametric Tests of Significance	
12	Probability	282
12.1	An Introduction to Probability	283
12.2	The Concept of Randomness	285
12.3	Approaches to Probability	288
12.4	Formal Properties of Probability	290
12.5	Probability and Continuous Variables	302
12.6	Probability and the Normal-Curve Model	303
12.7	One- and Two-Tailed p Values	305
	Summary	308
	Terms to Remember	309
	Exercises	310
13	Introduction to Statistical Inference	318
13.1	Why Sample?	319
13.2	The Concept of Sampling Distributions	320
13.3	Testing Statistical Hypotheses: Level of Significance	324
13.4	Testing Statistical Hypotheses: Null Hypothesis and Alternative Hypothesis	327
13.5	Testing Statistical Hypotheses: The Two Types of Error	330
13.6	A Final Word of Caution	336
	Summary	336
	Terms to Remember	337
	Exercises	338
14	Statistical Inference and Continuous Variables	342
14.1	Introduction	343
14.2	Sampling Distribution of the Mean	346
14.3	Testing Statistical Hypotheses: Population Mean and Standard Deviation Known	350
14.4	Estimation of Parameters: Point Estimation	355
14.5	Testing Statistical Hypotheses with Unknown Parameters: Student's t	358
14.6	Estimation of Parameters: Interval Estimation	363
14.7	Confidence Intervals and Confidence Limits	365
14.8	Test of Significance for Pearson's r : One Sample Case	370
14.9	Test of Significance for Goodman's and Kruskal's Gamma	374
	Summary	375
	Terms to Remember	375
	Exercises	376

15	An Introduction to the Analysis of Variance	382
15.1	Multigroup Comparisons	383
15.2	The Concept of Sums of Squares	385
15.3	Obtaining Variance Estimates	389
15.4	Fundamental Concepts of Analysis of Variance	391
15.5	Assumptions Underlying Analysis of Variance	391
15.6	An Example Involving Three Groups	392
15.7	The Interpretation of F	394
	Summary	396
	Terms to Remember	396
	Exercises	397
4	Inferential Statistics: Nonparametric Tests of Significance	
16	Statistical Inference with Categorical Variables: Chi Square and Related Measures	404
16.1	Introduction	405
16.2	The χ^2 One-Variable Case	407
16.3	The χ^2 Test of the Independence of Variables	410
16.4	Limitations in the Use of χ^2	413
16.5	Nominal Measures of Association Based on χ^2	416
	Summary	419
	Terms to Remember	420
	Exercises	420
17	Statistical Inference: Ordinally Scaled Variables	426
17.1	Introduction	427
17.2	Mann-Whitney U -Test	428
17.3	Nonparametric Tests Involving Correlated Samples	433
17.4	The Sign Test	433
17.5	Wilcoxon Matched-Pairs Signed-Rank Test	436
	Summary	438
	Terms to Remember	438
	Exercises	439
5	Appendixes	
A	Review of Basic Mathematics	444
B	Glossary of Symbols	454
C	Tables	461
D	Glossary of Terms	506
E	References	519
6	Answers to Selected Exercises	528
	Index	567

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

chapter **1**

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.

BOX 1.1

Acquired Immunodeficiency Syndrome (AIDS) in Europe

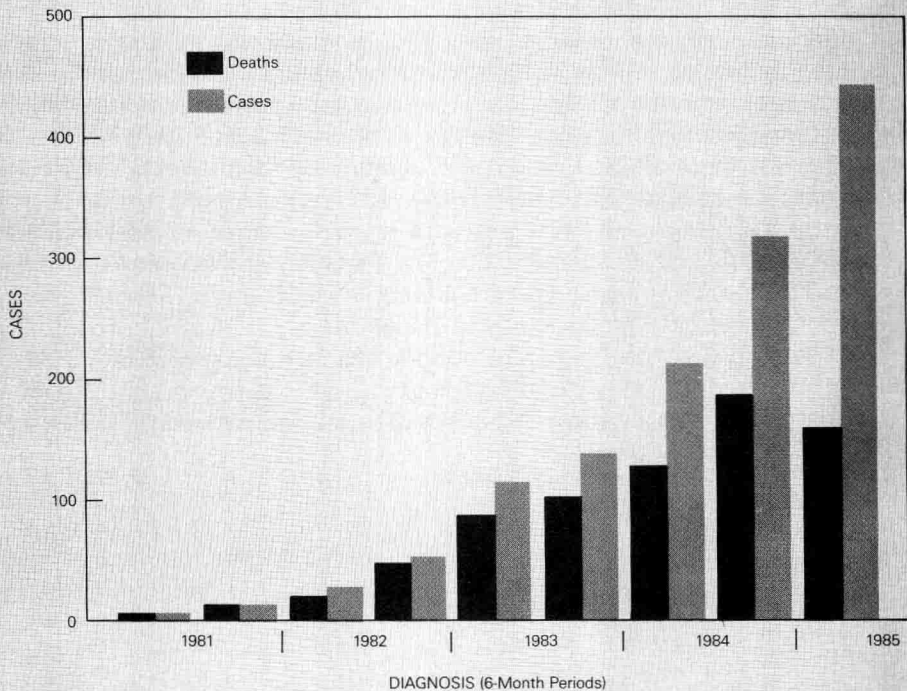
Statistics is one of the most widely used tools in the behavioral, social, medical, and physical sciences. Statistical information is collected on virtually every aspect of life and death. The resulting data are then subjected to various levels of statistical processing, inferences are drawn, and decisions are made that directly and indirectly affect our daily lives.

Often, observations are made at varying times to permit the comparison of changes over time. The records of these observations are then analyzed and the

results are summarized visually in graphic form. The data summarized in the following graph (Figure 1.1) show the number of cases and the number of deaths ascribed to acquired immunodeficiency syndrome (AIDS) in 21 European countries from 1981 through the first 6 months of 1985 [MMWR, **35**(3), 1986]. The graph leaves little to the imagination. It shows a progressive increase over time in both the number of new cases and the number of deaths from previously diagnosed cases.

Table 1.1 shows the risk groups that

Figure 1.1 Acquired immunodeficiency syndrome cases and deaths, by 6-month period of diagnosis—21 European countries, January 1, 1981 to June 30, 1985.



are identified according to the country of origin. Among native Europeans, the overwhelming majority of victims are either homosexual or bisexual. Known risk factors have been identified as any practices that involve the sharing of body fluids with individuals infected with AIDS. These include sexual intercourse, the shared use of intravenous drug paraphernalia, and blood transfusions if the donor blood contains the AIDS virus. Note that the majority of AIDS victims from the Caribbean Islands and Africa do not appear to be among risk groups that are known to be highest in Europe and the United States—homosexual and bisexual males. Rather, among African nationals AIDS appears to be more common among heterosexuals than homosexuals (124 out of 157 are males and females presumed to be heterosexual, whereas only 11 out

of 157 were identified as homosexual or bisexual). The same appears to be true of Caribbean nationals (only 4 out of 39 were identified as homosexual or bisexual).

As of this writing, there is neither a vaccine nor a cure for AIDS, nor does any breakthrough appear imminent. For the present, prevention appears to be the only available option in the efforts to stem the tide in the spread of AIDS. Large-scale educational efforts—enlisting the active participation of medical, social, and mental health professionals—are underway to convince those at greatest risk to modify many aspects of their sexual behavior [MMWR, 35(10), 1986].

Source: From Centers for Disease Control (1986) "Acquired Immunodeficiency Syndrome—Europe", *Morbidity and Mortality Weekly Report*, 35, 35–46.

Table 1.1 Acquired immunodeficiency syndrome cases, by patient risk group and geographic origin—21 European countries, through September 30, 1985

Patient risk group	Origin				Total	
	Europe	Caribbean Islands	Africa	Other	No.	(%)
1. Male homosexual or bisexual	1,031	4	11	39	1,085	(69)
2. IV drug abuser	90	—	—	—	90	(6)
3. Hemophilia patient	52	—	—	1	53	(3)
4. Transfusion recipient (without other risk factors)	30	—	5	—	35	(2)
5. 1- and 2-associated*	21	—	1	2	24	(2)
6. No known risk factor						
Male	59	24	81	3	167	(11)
Female	31	10	43	—	84	(5)
7. Unknown	16	1	16	2	35	(2)
Total	1,330 (85%)	39 (2%)	157 (10%)	47 (3%)	1,573	(100)

*Individuals were IV drug users and either homosexual or bisexual.