

**FOURTH
EDITION**

Fundamentals of Statistical Reasoning in Education

Theodore Coladarci | Casey Cobb

WILEY

Fundamentals of Statistical Reasoning in Education

Fourth Edition

Theodore Coladarci

University of Maine

Casey D. Cobb

University of Connecticut

WILEY

PUBLISHER	JAY O'CALLAGHAN
ACQUISITIONS EDITOR	ROBERT JOHNSTON
ASSISTANT EDITOR	BRITTANY CHEETHAM
MARKETING MANAGER	MARGARET BARRETT
COVER DESIGNER	KENJI NGIENG
ASSOCIATE PRODUCTION MANAGER	JOYCE POH
PRODUCTION EDITOR	JOLENE LING
PRODUCTION MANAGEMENT SERVICES	MPS LIMITED

This book was set in 10/12 TimesTenLTStd by MPS Limited. Cover and text printed and bound by Lightning Source.

This book is printed on acid free paper.

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

Copyright © 2014, 2011, 2008, 2004 John Wiley & Sons, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA 01923, website www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201)748-6011, fax (201)748-6008, website <http://www.wiley.com/go/permissions>.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return mailing label are available at www.wiley.com/go/returnlabel. If you have chosen to adopt this textbook for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local sales representative.

Library of Congress Cataloging-in-Publication Data

Fundamentals of statistical reasoning in education / Theodore Coladarci, Casey D. Cobb. – Fourth edition.

pages cm

Includes index.

ISBN 978-1-118-42521-3 (pbk.)

1. Educational statistics. I. Coladarci, Theodore.

LB2846.F84 2013

370.2'1–dc23

2013024353

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To our students

PREFACE

Fundamentals of Statistical Reasoning in Education 4e, like the first three editions, is written largely with students of education in mind. Accordingly, we draw primarily on examples and issues found in school settings, such as those having to do with instruction, learning, motivation, and assessment. Our emphasis on educational applications notwithstanding, we are confident that readers will find *Fundamentals 4e* of general relevance to other disciplines in the behavioral sciences as well.

Our overall objective is to provide clear and comfortable exposition, engaging examples, and a balanced presentation of technical considerations, all with a focus on conceptual development. Required mathematics call only for basic arithmetic and an elementary understanding of simple equations. For those who feel in need of a brushup, we provide a math review in Appendix A. Statistical procedures are illustrated in step-by-step fashion, and end-of-chapter problems give students ample opportunity for practice and self-assessment. (Answers to roughly half of these problems are found in Appendix B.) Almost all chapters include an illustrative case study and a “Reading the Research” section showing how a particular concept or procedure appears in the research literature. The result is a text that should engage all students, whether they approach their first course in statistics with confidence or apprehension.

You will find several changes in *Fundamentals 4e*:

- Guided by instructor feedback, we now use \hat{Y} instead of Y' as the symbol for the predicted value of Y in our treatment of regression analysis. Further, we now explicitly address the meaning, and importance, of residuals in regression analysis.
- A difficult decision for us was whether—and if so, how—to integrate computer applications into *Fundamentals*. In the end, we agreed that it is neither our intention nor our place to teach students in a first course the use of statistical software. If the computer is to be introduced in such a course, it is for the instructor to decide how to go about doing so. That said, we chose to illustrate (on the supporting website) the use of *SPSS*, a statistical package that enjoys considerable popularity in the education research community. Sample output provided with commentary, all within the context of the statistical procedures and tests covered in *Fundamentals 4e*. For the student who has access to *SPSS* and wishes to replicate our results (or simply explore this software further), we also provide a link to the data on which our applications are based.
- All chapters have benefited from the careful editing, along with the occasional clarification or elaboration, that one should expect of a new edition.

Fundamentals 4e is still designed as a “one semester” book. We intentionally sidestep topics that few introductory courses cover (e.g., factorial analysis of variance, repeated measures analysis of variance, multiple regression). At the same time, we incorporate effect size and confidence intervals throughout, which today are regarded as essential to good statistical practice.

Instructor's Guide

A guide for instructors can be found on the Wiley website at www.wiley.com/college/coladarci. This guide contains:

- Suggestions for adapting *Fundamentals 4e* to one's course.
- Helpful Internet resources on statistics education.
- The remaining answers to end-of-chapter problems.
- SPSS output, with commentary (and supporting data sets), presented within the context of the various statistical procedures and tests in *Fundamentals 4e*.
- An extensive bank of multiple-choice items.
- Supplemental material (“FYI”) providing elaboration or further illustration of procedures and principles in the text (e.g., the derivation of a formula, the equivalence of the t test, and one-way ANOVA when $k = 2$).

Acknowledgments

For their thoughtful comments and suggestions toward the preparation of *Fundamentals 4e*, we are grateful to Russel Almond, Florida State University; Betsy Becker, Florida State University; Carolyn L. Pearson, University of Arkansas at Little Rock; Tufan Tiglioglu, Alvernia University; and Shihfen Tu, University of Maine. The following reviewers gave invaluable feedback on earlier editions of *Fundamentals*, the positive result of which is reflected in the present edition as well: Terry Ackerman, University of Illinois, Urbana; Deb Allen, University of Maine; Tasha Beretvas, University of Texas at Austin; Shelly Blozis, University of Texas at Austin; Elliot Bonem, Eastern Michigan State University; David L. Brunsmma, University of Alabama in Huntsville; Daniel J. Calcagnettie, Fairleigh Dickinson University; David Chattin, St. Joseph's College; Grant Cioffi, University of New Hampshire; Stephen Cooper, Glendale Community College; Brian Doore, University of Maine; David X. Fitt, Temple University; Shawn Fitzgerald, Kent State University; Gary B. Forbach, Washburn University; Roger B. Frey, University of Maine; Jane Halpert, DePaul University; Larry V. Hedges, Northwestern University; Mark Hoyert, Indiana University Northwest; Jane Loeb, University of Illinois, Larry H. Ludlow, Boston College; David S. Malcolm, Fordham University; Terry Malcolm, Bloomfield College; Robert Markley, Fort Hayes State University; William Michael, University of Southern California; Wayne Mitchell, Southwest Missouri State University; David Mostofsky, Boston University; Ken Nishita, California State University at Monterey Bay; Robbie Pittman, Western Carolina University; Phillip A. Pratt, University of Maine; Katherine Prenovost, University of Kansas; Bruce G. Rogers, University of Northern Iowa; N. Clayton Silver, University of Nevada; Leighton E. Stamps, University of New Orleans; Irene Trenholme, Elmhurst College; Shihfen Tu, University of Maine; Gail Weems, University of Memphis; Kelly Kandra, University of North Carolina at Chapel Hill; James R. Larson, Jr., University of Illinois at Chicago; Julia Klausili, University of Texas at Dallas; Hiroko Arikawa, Forest Institute of Professional Psychology; James

Petty, University of Tennessee at Martin; Martin R. Deschenes, College of William and Mary; Kathryn Oleson, Reed College; Ward Rodriguez, California State University, Easy Bay; Gail D. Hughes, University of Arkansas at Little Rock; and Lea Witta, University of Central Florida.

We wish to thank John Moody, Derry Cooperative School District (NH); Michael Middleton, University of New Hampshire; and Charlie DePascale, National Center for the Improvement of Educational Assessment, each of whom provided data sets for some of the case studies.

We are particularly grateful for the continued support and encouragement provided by Robert Johnston of John Wiley & Sons, and to Brittany Cheetham, Kristen Mucci, Preeti Longia Sinha, Jolene Ling, and all others associated with this project.

**Theodore Coladarci
Casey D. Cobb**

CONTENTS

Chapter 1	Introduction	1	Chapter 3	Graphic Representation	34			
1.1	Why Statistics?	1	3.1	Why Graph Data?	34			
1.2	Descriptive Statistics	2	3.2	Graphing Qualitative Data: The Bar Chart	34			
1.3	Inferential Statistics	3	3.3	Graphing Quantitative Data: The Histogram	35			
1.4	The Role of Statistics in Educational Research	4	3.4	Relative Frequency and Proportional Area	39			
1.5	Variables and Their Measurement	5	3.5	Characteristics of Frequency Distributions	41			
1.6	Some Tips on Studying Statistics	8	3.6	The Box Plot	44			
			3.7	Summary	45			
PART 1			Chapter 4			Central Tendency	52	
DESCRIPTIVE STATISTICS			13	Chapter 5			Variability	66
Chapter 2	Frequency Distributions	14	4.1	The Concept of Central Tendency	52	5.1	Central Tendency Is Not Enough: The Importance of Variability	66
2.1	Why Organize Data?	14	4.2	The Mode	52	5.2	The Range	67
2.2	Frequency Distributions for Quantitative Variables	14	4.3	The Median	53	5.3	Variability and Deviations From the Mean	68
2.3	Grouped Scores	15	4.4	The Arithmetic Mean	54	5.4	The Variance	69
2.4	Some Guidelines for Forming Class Intervals	17	4.5	Central Tendency and Distribution Symmetry	57	5.5	The Standard Deviation	70
2.5	Constructing a Grouped-Data Frequency Distribution	18	4.6	Which Measure of Central Tendency to Use?	59			
2.6	The Relative Frequency Distribution	19	4.7	Summary	59			
2.7	Exact Limits	21						
2.8	The Cumulative Percentage Frequency Distribution	22						
2.9	Percentile Ranks	23						
2.10	Frequency Distributions for Qualitative Variables	25						
2.11	Summary	26						

viii Contents

5.6	The Predominance of the Variance and Standard Deviation	71
5.7	The Standard Deviation and the Normal Distribution	72
5.8	Comparing Means of Two Distributions: The Relevance of Variability	73
5.9	In the Denominator: n Versus $n - 1$	75
5.10	Summary	76

Chapter 6 Normal Distributions and Standard Scores 81

6.1	A Little History: Sir Francis Galton and the Normal Curve	81
6.2	Properties of the Normal Curve	82
6.3	More on the Standard Deviation and the Normal Distribution	82
6.4	z Scores	84
6.5	The Normal Curve Table	87
6.6	Finding Area When the Score Is Known	88
6.7	Reversing the Process: Finding Scores When the Area Is Known	91
6.8	Comparing Scores From Different Distributions	93
6.9	Interpreting Effect Size	94
6.10	Percentile Ranks and the Normal Distribution	96
6.11	Other Standard Scores	97
6.12	Standard Scores Do Not “Normalize” a Distribution	98
6.13	The Normal Curve and Probability	98
6.14	Summary	99

Chapter 7 Correlation 106

7.1	The Concept of Association	106
7.2	Bivariate Distributions and Scatterplots	106
7.3	The Covariance	111
7.4	The Pearson r	117

7.5	Computation of r : The Calculating Formula	118
7.6	Correlation and Causation	120
7.7	Factors Influencing Pearson r	122
7.8	Judging the Strength of Association: r^2	125
7.9	Other Correlation Coefficients	127
7.10	Summary	127

Chapter 8 Regression and Prediction 134

8.1	Correlation Versus Prediction	134
8.2	Determining the Line of Best Fit	135
8.3	The Regression Equation in Terms of Raw Scores	138
8.4	Interpreting the Raw-Score Slope	141
8.5	The Regression Equation in Terms of z Scores	141
8.6	Some Insights Regarding Correlation and Prediction	142
8.7	Regression and Sums of Squares	145
8.8	Residuals and Unexplained Variation	147
8.9	Measuring the Margin of Prediction Error: The Standard Error of Estimate	148
8.10	Correlation and Causality (Revisited)	152
8.11	Summary	153

PART 2

INFERENCE STATISTICS 163

Chapter 9 Probability and Probability Distributions 164

9.1	Statistical Inference: Accounting for Chance in Sample Results	164
9.2	Probability: The Study of Chance	165
9.3	Definition of Probability	166
9.4	Probability Distributions	168

9.5	The OR/addition Rule	169
9.6	The AND/Multiplication Rule	171
9.7	The Normal Curve as a Probability Distribution	172
9.8	“So What?”—Probability Distributions as the Basis for Statistical Inference	174
9.9	Summary	175

Chapter 10 Sampling Distributions 179

10.1	From Coins to Means	179
10.2	Samples and Populations	180
10.3	Statistics and Parameters	181
10.4	Random Sampling Model	181
10.5	Random Sampling in Practice	183
10.6	Sampling Distributions of Means	184
10.7	Characteristics of a Sampling Distribution of Means	185
10.8	Using a Sampling Distribution of Means to Determine Probabilities	188
10.9	The Importance of Sample Size (n)	191
10.10	Generality of the Concept of a Sampling Distribution	193
10.11	Summary	193

Chapter 11 Testing Statistical Hypotheses About μ When σ Is Known: The One-Sample z Test 199

11.1	Testing a Hypothesis About μ : Does “Homeschooling” Make a Difference?	199
11.2	Dr. Meyer’s Problem in a Nutshell	200
11.3	The Statistical Hypotheses: H_0 and H_1	201
11.4	The Test Statistic z	202
11.5	The Probability of the Test Statistic: The p Value	203

11.6	The Decision Criterion: Level of Significance (α)	204
11.7	The Level of Significance and Decision Error	207
11.8	The Nature and Role of H_0 and H_1	209
11.9	Rejection Versus Retention of H_0	209
11.10	Statistical Significance Versus Importance	210
11.11	Directional and Nondirectional Alternative Hypotheses	212
11.12	The Substantive Versus the Statistical	214
11.13	Summary	215

Chapter 12 Estimation 222

12.1	Hypothesis Testing Versus Estimation	222
12.2	Point Estimation Versus Interval Estimation	223
12.3	Constructing an Interval Estimate of μ	224
12.4	Interval Width and Level of Confidence	226
12.5	Interval Width and Sample Size	227
12.6	Interval Estimation and Hypothesis Testing	228
12.7	Advantages of Interval Estimation	230
12.8	Summary	230

Chapter 13 Testing Statistical Hypotheses About μ When σ Is Not Known: The One-Sample t Test 235

13.1	Reality: σ Often Is Unknown	235
13.2	Estimating the Standard Error of the Mean	236
13.3	The Test Statistic t	237
13.4	Degrees of Freedom	238
13.5	The Sampling Distribution of Student’s t	239
13.6	An Application of Student’s t	242

X Contents

13.7	Assumption of Population Normality	244
13.8	Levels of Significance Versus p Values	244
13.9	Constructing a Confidence Interval for μ When σ Is Not Known	246
13.10	Summary	247

Chapter 14 Comparing the Means of Two Populations: Independent Samples 253

14.1	From One μ (μ) to Two	253
14.2	Statistical Hypotheses	254
14.3	The Sampling Distribution of Differences Between Means	255
14.4	Estimating $\sigma_{\bar{x}_1 - \bar{x}_2}$	257
14.5	The t Test for Two Independent Samples	259
14.6	Testing Hypotheses About Two Independent Means: An Example	260
14.7	Interval Estimation of $\mu_1 - \mu_2$	262
14.8	Appraising the Magnitude of a Difference: Measures of Effect Size for $\bar{X}_1 - \bar{X}_2$	264
14.9	How Were Groups Formed? The Role of Randomization	268
14.10	Statistical Inferences and Nonstatistical Generalizations	269
14.11	Summary	270

Chapter 15 Comparing the Means of Dependent Samples 278

15.1	The Meaning of “Dependent”	278
15.2	Standard Error of the Difference Between Dependent Means	279
15.3	Degrees of Freedom	281
15.4	The t Test for Two Dependent Samples	281
15.5	Testing Hypotheses About Two Dependent Means: An Example	283
15.6	Interval Estimation of μ_D	286
15.7	Summary	287

Chapter 16 Comparing the Means of Three or More Independent Samples: One-Way Analysis of Variance 294

16.1	Comparing More Than Two Groups: Why Not Multiple t Tests?	294
16.2	The Statistical Hypotheses in One-Way ANOVA	295
16.3	The Logic of One-Way ANOVA: An Overview	296
16.4	Alison’s Reply to Gregory	299
16.5	Partitioning the Sums of Squares	300
16.6	Within-Groups and Between-Groups Variance Estimates	303
16.7	The F Test	304
16.8	Tukey’s “HSD” Test	306
16.9	Interval Estimation of $\mu_i - \mu_j$	308
16.10	One-Way ANOVA: Summarizing the Steps	309
16.11	Estimating the Strength of the Treatment Effect: Effect Size (ω^2)	311
16.12	ANOVA Assumptions (and Other Considerations)	312
16.13	Summary	313

Chapter 17 Inferences About the Pearson Correlation Coefficient 322

17.1	From μ to ρ	322
17.2	The Sampling Distribution of r When $\rho = 0$	322
17.3	Testing the Statistical Hypothesis That $\rho = 0$	324
17.4	An Example	324
17.5	In Brief: Student’s t Distribution and Regression Slope (b)	326
17.6	Table E	326
17.7	The Role of n in the Statistical Significance of r	328
17.8	Statistical Significance Versus Importance (Again)	329

17.9	Testing Hypotheses Other Than $\rho = 0$	329
17.10	Interval Estimation of ρ	330
17.11	Summary	332

Chapter 18 Making Inferences From Frequency Data 338

18.1	Frequency Data Versus Score Data	338
18.2	A Problem Involving Frequencies: The One-Variable Case	339
18.3	χ^2 : A Measure of Discrepancy Between Expected and Observed Frequencies	340
18.4	The Sampling Distribution of χ^2	341
18.5	Completion of the Voter Survey Problem: The χ^2 Goodness-of-Fit Test	343
18.6	The χ^2 Test of a Single Proportion	344
18.7	Interval Estimate of a Single Proportion	345
18.8	When There Are Two Variables: The χ^2 Test of Independence	347
18.9	Finding Expected Frequencies in the Two-Variable Case	348
18.10	Calculating the Two-Variable χ^2	350
18.11	The χ^2 Test of Independence: Summarizing the Steps	351
18.12	The 2×2 Contingency Table	352
18.13	Testing a Difference Between Two Proportions	353
18.14	The Independence of Observations	353
18.15	χ^2 and Quantitative Variables	354
18.16	Other Considerations	355
18.17	Summary	355

Chapter 19 Statistical “Power” (and How to Increase It) 363

19.1	The Power of a Statistical Test	363
19.2	Power and Type II Error	364
19.3	Effect Size (Revisited)	365
19.4	Factors Affecting Power: The Effect Size	366

19.5	Factors Affecting Power: Sample Size	367
19.6	Additional Factors Affecting Power	368
19.7	Significance Versus Importance	369
19.8	Selecting an Appropriate Sample Size	370
19.9	Summary	373

Epilogue A Note on (Almost) Assumption-Free Tests 379

References 380

Appendix A Review of Basic Mathematics 382

A.1	Introduction	382
A.2	Symbols and Their Meaning	382
A.3	Arithmetic Operations Involving Positive and Negative Numbers	383
A.4	Squares and Square Roots	383
A.5	Fractions	384
A.6	Operations Involving Parentheses	385
A.7	Approximate Numbers, Computational Accuracy, and Rounding	386

Appendix B Answers to Selected End-of-Chapter Problems 387

Appendix C Statistical Tables 408

Glossary 421

Index 427

Useful Formulas 433

CHAPTER 1

Introduction

1.1 Why Statistics?

An anonymous sage once defined a statistician as “one who collects data and draws confusions.” Another declared that members of this tribe occupy themselves by “drawing mathematically precise lines from unwarranted assumptions to foregone conclusions.” And then there is the legendary proclamation attributed (by Mark Twain) to the 19th-century British statesman Benjamin Disraeli: “There are three kinds of lies: lies, damned lies, and statistics.”

Are such characterizations justified? Clearly we think not! Just as every barrel has its rotten apples, there are statisticians among us for whom these sentiments are quite accurate. But they are the exception, not the rule. While there are endless reasons explaining why statistics is sometimes viewed with skepticism (math anxiety? mistrust of the unfamiliar?), there is no doubt that when properly applied, statistical reasoning serves to illuminate, not obscure. In short, our objective in writing this book is to acquaint you with the proper applications of statistical reasoning. As a result, you will be a more informed and critical patron of the research you read; furthermore, you will be able to conduct basic statistical analyses to explore empirical questions of your own.

Statistics merely formalizes what humans do every day. Indeed, most of the fundamental concepts and procedures we discuss in this book have parallels in everyday life, if somewhat beneath the surface. You may notice that there are people of different ages (“variability”) at Eric Clapton concerts. Because Maine summers are generally warm (“average”), you don’t bring a down parka when you vacation there. Parents from a certain generation, you observe, tend to drive Volvo station wagons (“association”). You believe that it is highly unlikely (“probability”) that your professor will take attendance two days in a row, so you skip class the day after attendance was taken. Having talked for only a few minutes (“sample”) with a person you just met, you conclude that you like him (“generalization,” “inference”). After getting a disappointing meal at a popular restaurant, you wonder whether it was just an off night for the chef or the place actually has gone downhill (“sampling variability,” “statistical significance”).

We could go on, but you probably get the point: Whether you are formally crunching numbers or simply going about life, you employ—consciously or not—the fundamental concepts and principles underlying statistical reasoning.

So what does formal statistical reasoning entail? As can be seen from the two-part structure of this book, statistical reasoning has two general branches: descriptive statistics and inferential statistics.

1.2 Descriptive Statistics

Among first-year students who declare a major in education, what proportion are male? Female? Do those proportions differ between elementary education and secondary education students? Upon graduation, how many obtain teaching positions? How many go on to graduate school in education? And what proportion end up doing something unrelated to education? These are examples of questions for which **descriptive statistics** can help to provide a meaningful and convenient way of characterizing and portraying important features of the data.¹ In the examples above, *frequencies* and *proportions* will help to do the job of statistical description.

The purpose of **descriptive statistics** is to organize and summarize data so that the data are more readily comprehended.

What is the average age of undergraduate students attending American universities for each of the past 10 years? Has it been changing? How much? What about the Graduate Record Examination (GRE) scores of graduate students over the past decade—has that average been changing? One way to show the change is to construct a graph portraying the average age or GRE score for each of the 10 years. These questions illustrate the use of *averages* and *graphs*, additional tools that are helpful for describing data.

We will explore descriptive procedures in later chapters, but for the present let's consider the following situation. Professor Tu, your statistics instructor, has given a test of elementary mathematics on the first day of class. She arranges the test scores in order of magnitude, and she sees that the distance between the highest and lowest scores is not great and that the class average is higher than normal. She is pleased because the general level of preparation seems to be good and the group is not exceedingly diverse in its skills, which should make her teaching job easier. And you are pleased, too, for you learn that your performance is better than that of 90% of the students in your class. This scenario illustrates the use of more tools of descriptive statistics: the *frequency distribution*, which shows the scores in ordered arrangement; the *percentile*, a way to describe the location of a person's score relative to scores of others in a group; and the *range*, which measures the variability of scores.

Because they each pertain to a single variable—age, GRE scores, and so on—the preceding examples involve **univariate** procedures for describing data. But often researchers are interested in describing data involving two characteristics of a person (or object) simultaneously, which call for **bivariate** procedures. For example, if you had information on 25 people concerning how many friends each person has (popularity) and how outgoing each person is (extroversion), you could see whether popularity and extroversion are related. Is popularity greater among people with higher levels of extroversion and, conversely, lower among people lower in extroversion? The *correlation coefficient* is a bivariate statistic that

¹We are purists with respect to the pronunciation of this important noun (“day-tuh”) and its plural status. Regarding the latter, promise us that you will recoil whenever you hear an otherwise informed person utter, “The data is. . . .” Simply put, data *are*.

describes the nature and magnitude of such relationships, and a *scatterplot* is a helpful tool for graphically portraying these relationships.

Regardless of how you approach the task of describing data, never lose sight of the principle underlying the use of descriptive statistics: The purpose is to organize and summarize data so the data are more readily comprehended and communicated. When the question “Should I use statistics?” comes up, ask yourself, “Would the story my data have to tell be clearer if I did?”

1.3 Inferential Statistics

What is the attitude of taxpayers toward, say, the use of federal dollars to support private schools? As you can imagine, pollsters find it impossible to put such questions to *every* taxpayer in this country! Instead, they survey the attitudes of a random **sample** of taxpayers, and from that knowledge they estimate the attitudes of taxpayers as a whole—the **population**. Like any estimate, this outcome is subject to random “error” or **sampling variation**. That is, random samples of the same population don’t yield identical outcomes. Fortunately, if the sample has been chosen properly, it is possible to determine the magnitude of error that is involved.

The second branch of statistical practice, known as **inferential statistics**, provides the basis for answering questions of this kind. These procedures allow one to account for chance error in drawing inferences about a larger group, the population, on the basis of examining only a sample of that group. A central distinction here is that between **statistic** and **parameter**. A statistic is a characteristic of a sample (e.g., the proportion of *polled* taxpayers who favor federal support of private schools), whereas a parameter is a characteristic of a population (the proportion of *all* taxpayers who favor such support). Thus, statistics are used to estimate, or make inferences about, parameters.

Inferential statistics permit conclusions about a population, based on the characteristics of a sample of the population.

Another application of inferential statistics is particularly helpful for evaluating the outcome of an experiment. Does a new drug, Melo, reduce hyperactivity among children? Suppose that you select at random two groups of hyperactive children and prescribe the drug to one group. All children are subsequently observed the following week in their classrooms. From the outcome of this study, you find that, on average, there is less hyperactivity among children receiving the drug.

Now some of this difference between the two groups would be expected even if they were treated alike in all respects, because of chance factors involved in the random selection of groups. As a researcher, the question you face is whether the obtained difference is within the limits of chance sampling variation. If certain assumptions have been met, statistical theory can provide the basis for an answer. If you find that the obtained difference is larger than that can be accounted for by chance alone, you will infer that other factors (the drug being a strong candidate) must be at work to influence hyperactivity.

This application of inferential statistics also is helpful for evaluating the outcome of a correlational study. Returning to the preceding example concerning the

relationship between popularity and extroversion, you would appraise the obtained correlation much as you would appraise the obtained difference in the hyperactivity experiment: Is this correlation larger than what would be expected from chance sampling variation alone? If so, then the traits of popularity and extroversion may very well be related in the population.

1.4 *The Role of Statistics in Educational Research*

Statistics is neither a beginning nor an end. A problem begins with a question rooted in the substance of the matter under study. *Does Melo reduce hyperactivity? Is popularity related to extroversion?* Such questions are called **substantive questions**.² You formulate the question, which is informed by a careful consideration of relevant theory, associated research, and due regard for the field of educational practice. You then decide on the appropriate methodology for exploring the question empirically—that is, using data.

Once you have diligently crafted the substantive question—and only then—it is now time for statistics to play a part. Let's say your study calls for averages (as in the case of the hyperactivity experiment). You calculate the average for each group and raise a **statistical question**: Are the two averages so different that sampling variation alone cannot account for the difference? Statistical questions differ from substantive questions in that the former are questions about a statistical index—in this case, the average. If, after applying the appropriate statistical procedures, you find that the two averages are so different that it is not reasonable to believe chance alone could account for it, you have made a **statistical conclusion**—a conclusion about the statistical question you raised.

Now back to the substantive question. If certain assumptions have been met and the conditions of the study have been carefully arranged, you may be able to conclude that the drug *does* make a difference, at least within the limits tested in your investigation. This is your final conclusion, and it is a **substantive conclusion**. Although the substantive conclusion derives partly from the statistical conclusion, other factors must be considered. As a researcher, therefore, you must weigh both the statistical conclusion *and* the adequacy of your methodology in arriving at the substantive conclusion.

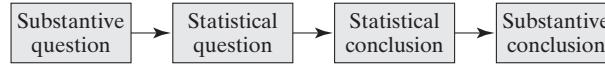
It is important to see that although there is a close relationship between the substantive question and the statistical question, the two are not identical. You will recall that a statistical question always concerns a statistical property of the data (e.g., an average or a correlation). Often, alternative statistical questions can be applied to explore the particular substantive question. For instance, one might ask whether the *proportion* of students with very high levels of hyperactivity differs beyond the limits of chance variation between the two conditions. In this case, the statistical question is about a different statistical index: the proportion rather than the average.

Thus, part of the task of mastering statistics is to learn how to choose among, and sometimes combine, different statistical approaches to a particular substantive question. When designing a study, the consideration of possible statistical analyses

²The substantive question also is called the research question.

to be performed should be situated in the course of refining the substantive question and developing a plan for collecting relevant data.

To sum up, the use of statistical procedures is always a middle step; they are a technical means to a substantive end. The argument we have presented can be illustrated as follows:



1.5 Variables and Their Measurement

Descriptive and inferential statistics are applied to **variables**.

A **variable** is a characteristic (of a person, place, or thing) that takes on different values.

Variables in educational research often (but not always) reflect characteristics of *people*—academic achievement, age, leadership style, intelligence, educational attainment, beliefs and attitudes, and self-efficacy, to name a few. Two nonpeople examples of variables are school size and brand of computer software. Although simple, the defining characteristic of a variable—*something that varies*—is important to remember. A “variable” that doesn’t vary sufficiently, as you will see later, can sabotage your statistical analysis.³

Statistical analysis is not possible without numbers, and there cannot be numbers without **measurement**.

Measurement is the process of assigning numbers to the characteristics you want to study.

For example, “20 years” may be the measurement for the characteristic, *age*, for a particular person; “115” may be that person’s measurement for *intelligence*; on a scale of 1 to 5, “3” may be the *sociability* measurement for this person; and because this hypothetical soul is female, perhaps she arbitrarily is assigned a value of “2” for *sex* (males being assigned “1”).

But numbers can be deceptive. Even though these four characteristics—age, intelligence, sociability, and sex—all have been expressed in numerical form, the numbers differ considerably in their underlying properties. Consequently, these numbers also differ in how they should be interpreted and treated. We now turn to a more detailed consideration of a variable’s properties and the corresponding implications for interpretation and treatment.

³If this statement perplexes you, think through the difficulty of determining the relationship between, say, “school size” and “academic achievement” if *all* of the schools in your sample were of an identical size. How could you possibly know whether there is a relationship (correlation) between academic achievement and school size? True, this may be an absurd scenario. After all, who would ask such a question if all schools were the same size?! But the problem of insufficient variability is far more subtle in practice, as you will see in Chapter 7.

Qualitative Versus Quantitative Variables

Values of **qualitative variables** (also known as **categorical variables**) differ in *kind* rather than in *amount*. Sex is a good example. Although males and females clearly are different in reproductive function (a qualitative distinction), it makes no sense to claim one group is either “less than” or “greater than” the other in this regard (a quantitative distinction).⁴ And this is true even if the arbitrary measurements suggest otherwise! Other examples of qualitative variables are college major, marital status, political affiliation, county residence, and ethnicity.

In contrast, the numbers assigned to **quantitative variables** represent differing *quantities* of the characteristic. Age, intelligence, and sociability, which you saw above, are examples of quantitative variables: A 40-year-old is “older than” a 10-year-old; an IQ of 120 suggests “more intelligence” than an IQ of 90; and a child with a sociability rating of 5 presumably is more sociable than the child assigned a 4. Thus, the values of a quantitative variable differ in amount. As you will see shortly, however, the properties of quantitative variables can differ greatly.

Scales of Measurement

In 1946, Harvard psychologist S. S. Stevens wrote a seminal article on **scales of measurement**, in which he introduced a more elaborate scheme for classifying variables. Although there is considerable debate regarding the implications of his typology for statistical analysis (e.g., see Gaito, 1980; Stine, 1989), Stevens nonetheless provided a helpful framework for considering the nature of one’s data. A variable, Stevens argued, rests on one of four scales: **nominal**, **ordinal**, **interval**, or **ratio**.

Nominal scales Values on a nominal scale merely “name” the category to which the object under study belongs. As such, interpretations must be limited to statements of kind rather than amount. (A qualitative variable thus represents a nominal scale.) Take ethnicity, for example, which a researcher may have coded 1 = Italian, 2 = Irish, 3 = Asian, 4 = Hispanic, 5 = African American, and 6 = Other.⁵ It would be perfectly appropriate to conclude that, say, a person assigned “1” (Italian, we trust) is different from the person assigned “4” (Hispanic), but you cannot demand more of these data. For example, you could not claim that because $3 < 5$, Asian is “less than” African American; or that an Italian, when added to an Asian, begets an Hispanic (because $1 + 3 = 4$). The numbers don’t mind, of course, but it still makes no sense. The moral throughout this discussion is the same: One should remain forever mindful of the variable’s underlying scale of measurement and the kinds of interpretations and operations that are sensible for that scale.

Ordinal scales Unlike nominal scales, values on an ordinal scale can be “ordered” to reflect differing degrees or amounts of the characteristic under study. For example, rank ordering students based on when they completed an in-class exam would reflect an ordinal scale, as would ranking runners according to when

⁴Although males and females, on average, do differ *in amount* on any number of variables (e.g., height, strength, annual income), the scale in question is no longer sex. Rather, it is the scale of the other variable on which males and females are observed to differ.

⁵Each individual must fall into only one category (i.e., the categories are *mutually exclusive*), and the five categories must represent all ethnicities included among the study’s participants (i.e., the categories are *exhaustive*).

they crossed the finish line. You know that the person with the rank of 1 finished the exam sooner, or the race faster, than individuals receiving higher ranks.⁶ But there is a limitation to this additional information: The only relation implied by ordinal values is “greater than” or “less than.” One cannot say *how much* sooner the first student completed the exam compared to the third student, or that the difference in completion time between these two students is the same as that between the third and fourth students, or that the second-ranked student completed the exam in half the time of the fourth-ranked student. Ordinal information simply does not permit such interpretations.

Although rank order is the classic example of an ordinal scale, other examples frequently surface in educational research. Percentile ranks, which we take up in Chapter 2, fall on an ordinal scale: They express a person’s performance relative to the performance of others (and little more). Likert-type items, which many educational researchers use for measuring attitudes, beliefs, and opinions (e.g., 1 = strongly disagree, 2 = disagree, and so on), are another example. Socio-economic status, reflecting such factors as income, education, and occupation, often is expressed as a set of ordered categories (e.g., 1 = lower class, 2 = middle class, 3 = upper class) and, thus, qualifies as an ordinal scale as well.

Interval scales Values on an interval scale overcome the basic limitation of the ordinal scale by having “equal intervals.” The 2-point difference between, say, 3 and 5 on an interval scale is the same—in terms of the underlying characteristic being measured—as the difference between 7 and 9 or 24 and 26. Consider an ordinary Celsius thermometer: A drop in temperature from 30°C to 10°C is equivalent to a drop from 50°C to 30°C.

The limitation of an interval scale, however, can be found in its **arbitrary zero**. In the case of the Celsius thermometer, for example, 0°C is arbitrarily set at the point at which water freezes (at sea level, no less). In contrast, the absence of heat (the temperature at which molecular activity ceases) is roughly -273°C . As a result, you could not claim that a 30°C day is three times as warm as a 10°C day. This would be the same as saying that column A in Figure 1.1 is three times

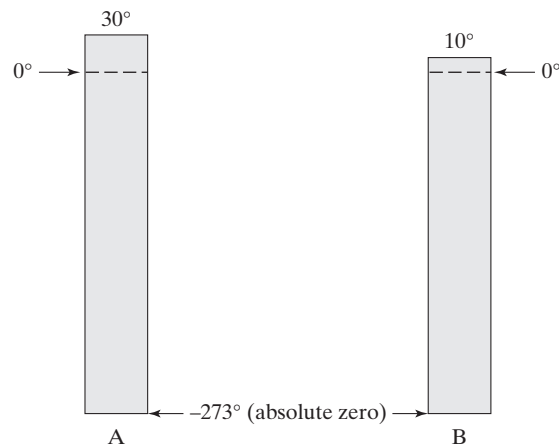


Figure 1.1 Comparison of 30°C and 10°C with the absolute zero on the Celsius scale.

⁶Although perhaps counterintuitive, the convention is to reserve *low* ranks (1, 2, etc.) for good performance (e.g., high scores, few errors, fast times).

as tall as column B. Statements involving ratios, like the preceding one, cannot be made from interval data.

What are examples of interval scales in educational research? Researchers typically regard composite measures of achievement, aptitude, personality, and attitude as interval scales. Although there is some debate as to whether such measures yield truly interval data, many researchers (ourselves included) are comfortable with the assumption that they do.

Ratio scales The final scale of measurement is the ratio scale. As you may suspect, it has the features of an interval scale *and* it permits ratio statements. This is because a ratio scale has an **absolute zero**. “Zero” weight, for example, represents an unequivocal absence of the characteristic being measured: *no* weight. Zip, nada, nothing. Consequently, you can say that a 230-pound linebacker weighs twice as much as a 115-pound jockey, a 30-year-old is three times the age of a 10-year-old, and the 38-foot sailboat *Adagio* is half the length of 76-foot *White Wings*—for weight, age, and length are all ratio scales.

In addition to physical measures (e.g., weight, height, distance, elapsed time), variables derived from *counting* also fall on a ratio scale. Examples include the number of errors a student makes on a reading comprehension task; the number of friends one reports having; the number of verbal reprimands a high school teacher issues during a lesson; or the number of students in a class, school, or district.

As with any scale, one must be careful when interpreting ratio scale data. Consider two vocabulary test scores of 10 and 20 (words correct). Does 20 reflect twice the performance of 10? It does if one’s interpretation is limited to performance on this particular test (“You knew twice as many words on this list as I did”). However, it would be unjustifiable to conclude that the student scoring 20 has twice the *vocabulary* as the student scoring 10. Why? Because “0” on this test does not represent an absence of vocabulary; rather, it represents an absence of knowledge of the specific words on this test. Again, proper interpretation is critical with any measurement scale.

1.6 Some Tips on Studying Statistics

Is statistics a hard subject? It is and it isn’t. Learning the “how” of statistics requires attention, care, and arithmetic accuracy, but it is not particularly difficult. Learning the “why” of statistics varies over a somewhat wider range of difficulty.

What is the expected reading rate for a book about statistics? Rate of reading and comprehension differ from person to person, of course, and a four-page assignment in mathematics may require more time than a four-page assignment in, say, history. Certainly, you should not expect to read a statistics text like a novel, or even like the usual history text. Some parts, like this chapter, will go faster; but others will require more concentration and several readings. In short, do not feel cognitively challenged or grow impatient if you can’t race through a chapter and, instead, find that you need time for absorption and reflection. The formal logic of statistical inference, for example, is a new way of thinking for most people and

requires some getting used to. Its newness can create difficulties for those who are not willing to slow down. As one of us was constantly reminded by his father, “*Festina lente!*”⁷

Many students expect difficulty in the area of mathematics. Ordinary arithmetic and some familiarity with the nature of equations are needed. Being able to see “what goes on” in an equation—to peek under the mathematical hood, so to speak—is necessary to understand what affects the statistic being calculated, and in what way. Such understanding also is helpful for spotting implausible results, which allows you to catch calculation errors when they first occur (rather than in an exam). Appendix A is especially addressed to those who feel that their mathematics lies in the too-distant past to assure a sense of security. It contains a review of elementary mathematics of special relevance for study of this book. Not all these understandings are required at once, so there will be time to brush up in advance of need.

Questions and problems are included at the end of each chapter. You should work enough of these to feel comfortable with the material. They have been designed to give practice in how-to-do-it, in the exercise of critical evaluation, in development of the link between real problems and methodological approach, and in comprehension of statistical relationships. There is merit in giving some consideration to all questions and problems, even though your instructor may formally assign fewer of them.

A word also should be said about the cumulative nature of a course in elementary statistics: What is learned in earlier stages becomes the foundation for what follows. Consequently, it is most important to keep up. If you have difficulty at some point, seek assistance from your instructor. Don’t delay. Those who think matters may clear up if they wait may be right, but the risk is greater here—considerably so—than in courses covering material that is less interdependent. It can be like attempting to climb a ladder with some rungs missing, or to understand an analogy when you don’t know the meaning of all the words. Cramming, never very successful, is least so in statistics. Success in studying statistics depends on regular work, and, if this is done, relatively little is needed in the way of review before examination time.

Finally, always try to “see the big picture.” First, this pays off in computation. Look at the result of your calculation. *Does it make sense?* Be suspicious if you find the average to be 53 but most of the numbers are in the 60s and 70s. And dismiss outright any result that defies plausibility, such as a statistic that falls beyond its possible limits. Remember, the eyeball is the statistician’s most powerful tool. Second, because of the ladderlike nature of statistics, also try to relate what you are currently studying to concepts, principles, and techniques you learned earlier. Search for connections—they are there. When this kind of effort is made, you will find that statistics is less a collection of disparate techniques and more a concerted course of study. Happily, you also will find that it is easier to master!

⁷“Make haste slowly!”