# BASIC STATISTICS for the BEHAVIORAL SCIENCES

2nd edition

Gary W. Heiman

# BASIC STATISTICS FOR THE BEHAVIORAL SCIENCES

Sponsoring Editor: Rebecca J. Dudley
Senior Associate Editor: Jane Knetzger
Senior Project Editor: Charline Lake
Production/Design Coordinator: Jennifer Waddell
Senior Manufacturing Coordinator: Priscilla J. Bailey

Cover design: Janet Theurer, Theurer Briggs Design
Cover image: "Crowd IV," Diana Ong

MYSTAT and SYSTAT are registered trademarks of SYSTAT, Inc. SPCC/PC+ Studentware Plus™ and the SPSS/PC+™ system are trademarks of SPCC, Inc. All other product names or trademarks are the property of their respective owners.

Copyright © 1996 by Houghton Mifflin Company. All rights reserved.

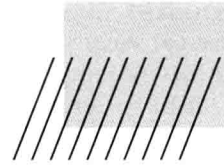
No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system without the prior written permission of Houghton Mifflin Company unless such copying is expressly permitted by federal copyright law. Address inquiries to College Permissions, Houghton Mifflin Company, 222 Berkeley Street, Boston, MA 02116-3764.

Printed in the U.S.A.

Library of Congress Catalog Card Number: 95-76952

Student Text ISBN: 0-395-74528-4
Examination Copy ISBN: 0-395-76581-1

3456789-DW-99 98 97 96



# Preface

Many of the undergraduates who enroll in my statistics course have a weak background in mathematics and some degree of "math phobia." By the end of the course, these same students must understand and perform the descriptive and inferential statistical procedures commonly used in behavioral research. The problem is that textbooks often dwell on the remarkable things statisticians can do with statistics and say too little about the things *researchers* commonly do. Such books present a catalog of procedures but do not explain the conceptual purposes of these procedures. Although students can then compute an answer on demand, they do not understand why they should perform the procedure nor what their answer reveals about the data. Therefore, I wanted my students to have a textbook that takes their needs more fully into account: a book that *explains*—clearly, patiently, and with an occasional touch of humor—the way a good teacher does.

## My Objectives

In writing this book, I pursued five objectives.

**1. Take a conceptual-intuitive approach** The text emphasizes the context in which statistics are used to make sense out of data. Each procedure is introduced in the context of a simple study with readily understandable goals. I focus on the purpose of research as examining the relationships between variables, then delineate the procedures for describing and inferring such relationships, and finally return to the conceptual purpose and interpretation of the study. Throughout, I provide students with simplified ways to think about statistical concepts and to see how these concepts translate into practical procedures for answering practical questions.

**2. Present statistics within an understandable research context** Many of the text's early examples involve simple variables and research questions taken from everyday life, so that students have an intuitive feel for the meaning of the scores and relationships discussed. In later chapters, along with students' developing statistical thinking, examples become more technical and "psychological." Virtually all examples and study questions involve specific variables and research questions, instead of generic data.

**3. Deal directly and positively with student weaknesses in mathematics** The text presents no formulas or statistical statements without explanation. Formulas are introduced in terms of what they accomplish, and an example of each is worked out completely, step by step. To further reduce the apparent complexity of statistics, I have stressed the similarities among different procedures, showing how, despite slight variations in computations, they have similar components and answer similar questions.

**4. Introduce new terms and concepts in an integrated way** I have strived to tie each new concept and procedure to previous material, briefly reviewing that material in every possible instance, including in the new chapter openings. Throughout, difficult concepts are presented in small chunks, which are then built into a foundation and later elaborated on.

**5. Create a text that students will enjoy as well as learn from** To make the text readable and engaging, I repeatedly point out the everyday usefulness of statistics. I have also tried to convey my own excitement about statistics and to dispel the notion that statistics (and statisticians) are boring. One can take a discipline seriously yet still recognize its quirks and foibles and have fun with it.

## Organization of the Text

In Part 1, *Getting Started*, Chapter 1 serves as a brief preface for the student and reviews basic math and graphing techniques. Chapter 2 then introduces the terminology, logic, and goals of statistics within the context of behavioral research.

Chapters 3 through 6 make up Part 2, *Descriptive Statistics* (along with a discussion of linear interpolation in Appendix A). I placed the chapter dealing with $z$-scores (Chapter 6) after the chapters on central tendency and variability (Chapters 4 and 5) so that these building blocks would be fresh in students' minds when discussing $z$-scores. I included sampling distributions and computing a $z$-score for a sample mean in this chapter so that students can later see how the logic of inferential statistics is essentially that of computing $z$-scores.

Part 3, *Describing Relationships*, consists of Chapters 7 and 8, in which correlation and regression are introduced as descriptive procedures, with emphasis on interpreting the correlation coefficient and the variance accounted for. (The point-biserial correlation is included to provide a bridge to measures of effect size in later chapters.) I placed the chapters dealing with correlation and linear regression before the inferential statistical procedures because otherwise it is confusing to introduce these procedures while discussing their inferential tests. Substantial cautions are given, however, about the need for performing inferential procedures on correlation coefficients, and when discussed later, they are presented as a logical variation of significance testing of means.

Part 4, *Inferential Statistics*, begins with Chapter 9, although extensive groundwork is laid in the previous chapters, with strong emphasis on understanding the proportion of the area under the normal curve. Chapter 9 introduces probability and previews hypothesis testing, focusing on using the normal curve to compute probability, with the goal of making decisions about the representativeness of

此为试读,需要完整PDF请访问: www.ertongbook.com

sample means. In Chapter 10, hypothesis testing is formalized using the *z*-test. Chapter 11 presents the single-sample *t*-test, the confidence interval for a single mean, and significance testing of correlation coefficients. Chapter 12 covers two-sample *t*-tests and effect size. Chapter 13 introduces the one-way, between-subjects ANOVA, including *post hoc* tests for equal and unequal *n*'s, eta squared, and omega squared. (The one-way within-subjects ANOVA is described in Appendix C.) Chapter 14 deals with the two-way between-subjects ANOVA, *post hoc* tests for main effects and for unconfounded comparisons in an interaction, as well as graphing and interpreting interactions. Chapter 15 covers the one-way and two-way chi square, as well as the nonparametric versions of all previous parametric tests (with appropriate *post hoc* tests and measures of effect size).

The text is also designed to serve as a reference book for students. To that end, I've included such procedures as the formulas for transforming a raw score into a percentile and vice versa, for the semi-interquartile range, for T-scores, for the $F_{max}$ test, for several types of confidence intervals, and for a rather extensive collection of nonparametric procedures. The instructor can skip the more uncommon procedures, however, without disrupting the discussion of the major procedures.

The text strives to teach students how to interpret their data—not just to report that a result is significant. Thus, I have emphasized such topics as plotting and interpreting graphs and understanding the relationships demonstrated by research. I've also included practical discussions of power and measures of effect size. These discussions occur at the end of a section or chapter so that instructors wishing to skip these topics can easily do so.

## Pedagogical Format and Features

A number of features have been built into the book to enhance its usefulness as both a tool for study and a reference.

- New chapter-opening pedagogy encourages review of previous material and provides learning goals.
- "More Statistical Notation" sections introduce new statistical notations at the beginning of the chapter in which they are used, not before, and to reduce student confusion, they are introduced separately from the conceptual issues presented in the chapter.
- Each important procedural point is emphasized by a "STAT ALERT," a summary reminder set off from the text about the calculation or interpretation of a statistic.
- Computational formulas are highlighted throughout the text in color.
- Key terms are highlighted in bold, reviewed in the chapter summary, and listed in the end-of-text glossary. Many mnemonics and analogies are used to promote retention and understanding.
- Graphs and diagrams are thoroughly explained in captions and fully integrated into the discussion.
- "Finally" sections at the end of each chapter provide advice, cautions, and ways to integrate material from different chapters.

- Each "Chapter Summary" provides a substantive review of the material, not merely a list of the topics covered.
- Conceptual and procedural questions, as well as computational problems, are provided at the end of each chapter. Odd-numbered problems (with final and intermediate answers in Appendix E) provide students with a solid review of the material, and even-numbered problems (with answers in the Instructor's Resource Manual) can be used as assigned homework.
- A Summary of Formulas is provided at the end of each chapter, for quick reference. In addition, a list of formulas is shrinkwrapped with each student text for use in closed-book exams.
- A glossary of symbols appears on the inside back cover. New reference tables on the inside front cover provide guidelines for selecting the various descriptive and inferential procedures discussed in the text based on the type of data or research design employed.

## New Features in the Second Edition

The first edition was quite well received, and the various reviewers and users suggested little in the way of substantial change. Therefore, my focus was to tighten the conceptual presentations and to incorporate several new explanatory techniques that I developed since the first edition. Throughout, greater emphasis is placed on explaining how to use statistics to mentally "envision" a set of data and to "think" in statistical terms. Further, discussion of the mathematical steps in computing the various procedures has been streamlined, although I have kept the text very friendly to students with math phobia. In addition, I made the following specific changes:

- *New chapter-opening pedagogy* Each chapter now begins by identifying the previously discussed concepts that students should review before continuing, followed by learning goals for the chapter.
- *Expanded practice problems* The number of practice problems at the end of each chapter has increased by fifty percent, with an emphasis on both computational problems and challenging conceptual problems. In addition, practice problems are now provided in the appendices dealing with interpolation, probability, and the one-way within-subjects ANOVA.
- *New software appendix* Appendix F, "Introduction to Statistics Software Packages," provides the student with simple step-by-step instructions for performing all the major statistical procedures discussed in the text using one of three data-analysis software programs: HMSTAT (packaged free with each student text), MYSTAT, and SPSS/PC+ Studentware Plus™. Appendix F provides directions for installing the programs, entering data, running the various procedures, and interpreting the output.

Key changes also have been made to each chapter: Chapter 1 now provides instructions on how students should read a statistics text and study for exams. Chapter 2 has been revamped to emphasize that psychological research is based on the two goals of demonstrating a relationship in sample data and using samples

to draw inferences about the population. Chapter 3 includes an improved explanation of using the area under the normal curve to describe scores. Chapter 4 more clearly takes students through the important steps of using a mean score to describe obtained scores, to predict future scores, to convert raw scores to deviation scores, and to estimate the population mean. Chapter 5 now covers using inflection points on a normal distribution to identify the location of the scores that are one standard deviation above or below the mean. Chapter 6 features more discussion of how the normal curve, *z*-scores, and the central limit theorem allow us to envision the important characteristics of any set of data. Chapters 7 and 8 explain more fully the distinction between a correlational and experimental research design, as well as the similarities and differences in the statistical procedures used in each.

Chapters 9 and 10 clearly set the stage for inferential procedures, with an improved introduction to using a sampling distribution to describe the means we expect if we are representing a particular population and a description of how inferential procedures are ultimately designed to limit Type I and Type II errors. Chapter 11 contains expanded discussion of how researchers maximize statistical power in one-sample experiments and correlational studies. Chapter 12 presents a clearer description of using a squared correlation coefficient to describe effect size and mentions the latest American Psychological Association requirements for reporting effect size with all significant results. Chapter 13 includes an expanded discussion of how the between-groups mean square is an estimate of both error and treatment variance. Chapter 14 features improved graphics showing how the diagram of a two-way design is built from the diagrams of two one-way designs. Chapter 15 has been streamlined, especially in presenting the interpretation of a two-way chi square.

## Supplementary Materials

Supporting the text are several ancillaries for students and instructors:

- *Free Data-Analysis Software* A data-analysis program called HMSTAT, custom-tailored to the text by David Abbott of the University of Central Florida, is packaged free with each student text. This menu-driven program can accept and store data, perform all the procedures discussed in the text, and be operated by students with minimal computer background. Directions for using HMSTAT appear in the new Appendix F, prepared by Michael Gayle of the State University of New York at New Paltz, and the Instructor's Resource Manual offers suggestions for integrating computers into the course. The appendix also provides instructions for using MYSTAT and SPSS/PC+ Studentware Plus™ statistical programs.
- *Student Workbook and Study Guide* Additional practice problems are available in the Student Workbook and Study Guide, which I personally revised. Each chapter contains a review of objectives, terms, and formulas, a programmed review, conceptual and computational problems (with answers), and a set of multiple-choice questions similar to those in the Instructor's Resource Manual. A final chapter, called "Getting Ready for the Final

Exam," facilitates student integration of the entire course. Answers to all questions are now provided in each workbook chapter.

- *Instructor's Resource Manual with Test Questions* This supplement, revised by Kay Smith of Brigham Young University, contains approximately 750 test items and problems, as well as suggestions for classroom activities, discussion, and use of statistical software. It also includes answers to the even-numbered end-of-chapter problems from this book. The test items are also available on computer disk for IBM and Macintosh computers.

## Acknowledgments

I gratefully acknowledge the help and support of many professionals associated with Houghton Mifflin Company. In particular, I want to thank Michael DeRocco, Rebecca Dudley, Jane Knetzger, and Karen Donovan. A special thanks to Larissa Semanchuk. Finally, I am grateful to the following reviewers who, in evaluating all or parts of this text at one stage or another, provided invaluable feedback.

Bernard C. Beins, Ithaca College
James I. Chumbley, University of Massachusetts, Amherst
Richard P. Deshon, Michigan State University
George M. Diekhoff, Midwestern State University
Susan L. Donaldson, University of Southern Indiana
Warren Fass, University of Pittsburgh at Bradford
Nancy Forger, University of Massachusetts, Amherst
Michael C. Gayle, SUNY New Paltz
Brett W. Pelham, University of California, Los Angeles
Tom Pyle, Eastern Washington University
George A. Raymond, Providence College
Robert A. Reeves, Augusta College
Raymond R. Reno, University of Notre Dame
Kay H. Smith, Brigham Young University
Steven M. Specht, Lebanon Valley College
Lori L. Temple, University of Nevada, Las Vegas
Philip Tolin, Central Washington University
George Whitehead, Salisbury State University
Leonard Williams, Rowan College
Stephanie Gray Wilson, Seton Hall University

Gary W. Heiman

# Brief Contents


Preface xxi

*PART 1*

**GETTING STARTED**

1 Introduction 2
2 Statistics and the Research Process 16

*PART 2*

**DESCRIPTIVE STATISTICS: DESCRIBING SAMPLES AND POPULATIONS**

3 Frequency Distributions and Percentiles 46
4 Measures of Central Tendency: The Mean, Median, and Mode 81
5 Measures of Variability: Range, Variance, and Standard Deviation 113
6 *z*-Score Transformations and the Normal Curve Model 144

*PART 3*

**DESCRIBING RELATIONSHIPS**

7 Describing Relationships Using Correlations 180
8 Linear Regression and Predicting Variability 212

*PART 4*

**INFERENTIAL STATISTICS**

9 Probability: Making Decisions About Chance Events 244
10 Overview of Statistical Hypothesis Testing: The *z*-Test 271
11 Significance Testing of a Single Sample Mean or a Correlation Coefficient: The *t*-Test 307
12 Significance Testing of Two Sample Means: The *t*-Test 342
13 One-Way Analysis of Variance: Testing the Significance of Two or More Sample Means 379
14 Two-Way Analysis of Variance: Testing the Means from Two Independent Variables 414
15 Nonparametric Procedures for Frequency Data and Ranked Data 452

*APPENDICES*

A Interpolation 489
B Additional Formulas for Computing Probability 493
C One-Way Within-Subjects Analysis of Variance 499
D Statistical Tables 509
E Answers to Odd-Numbered Practice Problems 531
F Introduction to Statistics Software Packages 543

Glossary 596

Index 607

# Contents

Preface xxi

PART 1

GETTING STARTED

1 Introduction 2

Some Answers to Your Questions and Concerns About Statistics 2
What Are Statistics? 2
What Do Psychologists Do with Statistics? 2
But I'm Not Interested in Research; I Just Want to Help People! 3
But I Don't Know Anything About Research! 3
What If I'm Not Very Good at Statistics? 3
But Statistics Aren't Written in English! 4
What If I'm Not Very Good at Math? 4
So All I Have to Do Is Learn How to Compute the Answers? 4
What Should Be My Goal? 4
All Right, So How Do I Learn Statistics? 5
What Should I Know About This Book? 5



此为试读,需要完整PDF请访问: www.ertongbook.com

Review of Mathematics Used in Statistics 6
Basic Statistical Notation 7
*Identifying mathematical operations 7* *Determining the order of mathematical operations 7* *Solving equations to find the answer 8*
Rounding 8
Transformations 9
*Proportions 9* *Percents 10*
Creating Graphs 10
Finally 12
Chapter Summary 12
Practice Problems 13

2
**Statistics and the Research Process 16**

The Logic of Scientific Research 17
Variables 17
Relationships Between Variables 18
*Strength of a relationship 20* *No relationship 21* *Graphing relationships 21*
Using Relationships to Discover Laws 23
Samples and Populations 24
Drawing Inferences About a Population 25
Representativeness of a Sample 26
*Random sampling 27* *Unrepresentative samples 27*
Using Statistical Procedures to Analyze Data 28
Descriptive Statistics 29
Inferential Statistics 30
Statistics and Parameters 31
The Characteristics of a Study 31
Research Designs 32
Experiments 32
*The independent variable 32* *The dependent variable 33* *Drawing conclusions from experiments 34* *The problem of causality 35*
Correlational Studies 36
*Drawing conclusions from correlational studies 36* *Again, the problem of causality 36*
Types of Variables 37
*The four types of measurement scales 37* *Discrete and continuous scales 38*
Finally 39
Chapter Summary 40
Practice Problems 41

PART 2

## DESCRIPTIVE STATISTICS: DESCRIBING SAMPLES AND POPULATIONS


3

**Frequency Distributions and Percentiles 46**

More Statistical Notation 47

Creating Simple Frequency Distributions 47

Presenting Simple Frequency in a Table 48

Graphing a Simple Frequency Distribution 49

*Bar graphs 49* *Histograms 49* *Frequency polygons 51*

Types of Simple Frequency Distributions 52

The Normal Distribution 52

*Overlapping normal distributions 54* *The normal curve model 55* *Approximations to the normal distribution 55*

Other Common Frequency Polygons 56

*Skewed distributions 56* *Bimodal and rectangular distributions 57* *Distributions of real data versus ideal distributions 58*

Creating Relative Frequency Distributions 59

Presenting Relative Frequency in a Table 59

Graphing a Relative Frequency Distribution 60

Finding Relative Frequency Using the Normal Curve 61

Creating Cumulative Frequency Distributions 63

Presenting Cumulative Frequencies in a Table 64

Graphing a Cumulative Frequency Distribution 64

Computing Percentile 65

Finding Percentile Using the Area Under the Normal Curve 65

Creating Grouped Frequency Distributions 67

Real Versus Apparent Limits 69

Graphing Grouped Distributions 70

Determining the Score at a Given Percentile 71

Finding a Percentile for a Given Score 74

Finally 76

Chapter Summary 76

Practice Problems 78

Summary of Formulas 80

## 4
## Measures of Central Tendency: The Mean, Median, and Mode 81

More Statistical Notation 82
What Is Central Tendency? 82
The Mode 84
Uses of the Mode 84
Problems with the Mode 86
The Median 86
Uses of the Median 88
Problems with the Median 88
The Mean 88
Uses of the Mean 89
Problems with the Mean 91
Transformations and the Mean 93
Deviations Around the Mean 94
Using the Mean to Interpret Research 95
Using the Mean to Describe Individual Scores in a Sample 95
Using the Mean to Predict Individual Scores 96
Using the Mean to Describe a Score's Location in a Distribution 97
Using the Sample Mean to Describe the Population Mean 99
Summarizing the Results of Experiments 100
Summarizing a Relationship Using Measures of Central Tendency 101
Graphing the Results of an Experiment 103
*Line graphs 103* *Bar graphs 104*
Inferring the Relationship in the Population 106
Finally 107
Chapter Summary 108
Practice Problems 109
Summary of Formulas 112

## 5
## Measures of Variability: Range, Variance, and Standard Deviation 113

More Statistical Notation 114
The Concept of Variability 114
The Range 116
The Semi-interquartile Range 117
Understanding Variance and Standard Deviation 118

Describing the Sample Variance 119
Computational Formula for the Sample Variance 121
Interpreting Variance 122
Describing the Sample Standard Deviation 123
Computational Formula for the Sample Standard Deviation 123
Interpreting the Standard Deviation 124
Applying the Standard Deviation to the Normal Curve 125
Avoiding Some Common Errors in Computing $S_X$ 128
Mathematical Constants and the Standard Deviation 128
The Population Variance and the Population Standard Deviation 129
Estimating the Population Variance and Population Standard Deviation 130
*Computational formula for the estimated population variance 132*
*Computational formula for the estimated population standard deviation 133*
Interpreting the Estimated Population Variance and Standard Deviation 133
Variance Is the Error in Predictions 134
Estimating the Error in Predictions in the Population 135
Accounting for Variance 136
Finally 138
Chapter Summary 139
Practice Problems 140
Summary of Formulas 142

6
**z-Score Transformations and the Normal Curve Model 144**

More Statistical Notation 145
Understanding z-Scores 145
Describing a Score's Relative Location as a z-Score 147
Computing z-Scores 148
Computing a Raw Score When z Is Known 149
How Variability Influences z-Scores 150
Interpreting z-Scores: The z-Distribution 151
Creating a z-Distribution 151
Characteristics of the z-Distribution 152
Transforming z-Scores 153
Using the z-Distribution to Compare Different Distributions 154
Plotting Different z-Distributions on the Same Graph 155

Using the *z*-Distribution to Determine the Relative Frequency of Raw Scores 156
The Standard Normal Curve 157
Applying the Standard Normal Curve Model 159
*Finding percentile rank for a raw score 160* *Finding a raw score at a given percentile rank 161* *Using the z-table 162*
Using the Standard Normal Curve Model to Define Psychological Attributes 164
Using the Standard Normal Curve Model to Describe Samples 165
The Sampling Distribution of Means 165
The Standard Error of the Mean 168
Calculating a *z*-score for a Sample Mean 170
Using the Sampling Distribution to Determine Relative Frequency of Sample Means 171
Finally 172
Chapter Summary 173
Practice Problems 174
Summary of Formulas 177

PART 3

DESCRIBING RELATIONSHIPS

7
Describing Relationships Using Correlations 180

More Statistical Notation 181
Understanding Correlational Research 182
Correlation and Causality 182
Distinguishing Characteristics of Correlation 183
Plotting Correlational Data: The Scatterplot 184
The Correlation Coefficient 186
Types of Relationships 186
Linear Relationships 186
Nonlinear Relationships 188
How the Correlation Coefficient Describes the Type of Relationship 189
Strength of the Relationship 189
Perfect Association 190
Intermediate Association 191
Zero Association 193

Computing the Correlation Coefficient 194
The Pearson Correlation Coefficient 195
*Computing the Pearson correlation coefficient 195*
The Spearman Rank-Order Correlation Coefficient 197
Computing the Spearman correlation coefficient 198
*Tied ranks 199*
The Point-Biserial Correlation Coefficient 200
*Computing the point-biserial correlation coefficient 200*
The Restriction of Range Problem 202
Correlations in the Population 203
Finally 204
Chapter Summary 205
Practice Problems 206
Summary of Formulas 211

8
**Linear Regression and Predicting Variability 212**

More Statistical Notation 213
Understanding Linear Regression 213
Summarizing the Scatterplot Using the Regression Line 213
Predicting Scores Using the Regression Line 215
The Linear Regression Equation 216
Computing the Slope 218
Computing the *Y*-Intercept 219
Describing the Linear Regression Equation 219
Plotting the Regression Line 220
Using the Regression Equation to Predict *Y* Scores 220
Error When the Linear Regression Equation Is Used to Predict Scores 221
Computing the Variance of the *Y* Scores Around *Y'* 222
Computing the Standard Error of the Estimate 224
Assumptions of Linear Regression 226
Strength of the Relationship and Amount of Prediction Error 227
Predicting Variability: The Proportion of Variance Accounted For 230
Understanding the Variance Accounted For 231
Computing the Proportion of Variance Accounted For 232
Using *r* to Compute the Proportion of Variance Accounted For 233
Using the Variance Accounted For 234
A Word About Multiple Correlation and Regression 235
Finally 236
Chapter Summary 236
Practice Problems 237
Summary of Formulas 240

此为试读,需要完整PDF请访问: www.ertongbook.com