



INTRODUCTORY

Statistics

for the BEHAVIORAL
SCIENCES

SIXTH EDITION

JOAN WELKOWITZ | BARRY H. COHEN | ROBERT B. EWEN



C32
W446
E.6

Introductory Statistics for the Behavioral Sciences

Sixth Edition

Joan Welkowitz

New York University

Barry H. Cohen

New York University

Robert B. Ewen

Gulliver Preparatory School



E2007000869



WILEY

John Wiley & Sons, Inc.

This book is printed on acid-free paper. ☺

Copyright © 2006 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

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 Section 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, (978) 750-8400, fax (978) 646-8600, or on the web at 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, (201) 748-6011, fax (201) 748-6008, or online at <http://www.Wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If legal, accounting, medical, psychological or any other expert assistance is required, the services of a competent professional person should be sought.

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc. is aware of a claim, the product names appear in initial capital or all capital letters. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

For general information on our other products and services please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books. For more information about Wiley products, visit our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Welkowitz, Joan.

Introductory statistics for the behavioral sciences / Joan Welkowitz, Barry H. Cohen, Robert B. Ewen. — 6th ed.
p. cm.

Various multi-media instructional resources are available to supplement the text.

Includes bibliographical references and index.

ISBN-13: 978-0-471-73547-2 (cloth)

ISBN-10: 0-471-73547-7 (cloth)

1. Social sciences—Statistical methods. 2. Psychometrics. 3. Sociology—Statistical methods.
4. Education—Statistical methods.

I. Cohen, Barry H., 1949– . II. Ewen, Robert B., 1940– . III. Title.

HA29.W445 2006

519.5024'3—dc22

2006013970

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Introductory Statistics for the Behavioral Sciences

Sixth Edition

This book is dedicated to
our students—past, present, and future—
to Walter, Julie, Larry, David, Sara, and Ray,
to Lori, Michael, and Melissa,
to Judy and Meredith,

*and especially to the memory of Joan Welkowitz,
our mentor and brilliant, much-loved friend.
She will be greatly missed by many family members,
friends, colleagues, and former students.*

Preface

This sixth edition represents a major revision of a text that was first published in 1971, and has been in print continuously ever since. However, our purpose in this edition is the same as it was in the very first edition: to introduce and explain statistical concepts and principles clearly and in a highly readable fashion, assuming minimal mathematical sophistication, but avoiding a “cook-book” approach to methodology.

In this edition we have moved to a new publisher, John Wiley & Sons, and replaced our dear, departed colleague, Jack Cohen, with another co-author of the same last name. Although not directly related to Jack in a genetic sense, Barry Cohen began teaching statistics at New York University while Jack Cohen was still doing the same, and has been influenced accordingly. Before then, he served as a teaching assistant to Joan Welkowitz at NYU more than twenty years ago. One advantage of joining John Wiley & Sons is that this text can now share supplementary materials with Barry Cohen’s graduate-level statistics text, *Explaining Psychological Statistics*, also published by Wiley. (See Barry Cohen’s statistics web page: www.psych.nyu.edu/cohen/stattext.html.)

There are five major changes in this edition.

1. The Inclusion of Exercises in the Text

Most of the numerical exercises from the Study Guide prepared by Bob Ewen for the fifth edition of this text have been moved to the end of the appropriate chapters in this edition. Most of the conceptual workbook problems have been moved to a separate section following the exercises in each chapter, labeled Thought Questions. Answers to selected exercises appear in an appendix to the text; answers to all of the exercises (including the thought questions) are in the instructor’s manual, available from the Wiley web page for this text (instructors will need to set up a password to access this manual). For instructors who have used the Study Guide from the previous edition, the instructor’s manual includes a table that will tell them where to find each of those problems in this edition. A student study guide is also available from the Wiley web page. (Links to the Wiley web page can always be found on the Cohen statistics web page.)

2. The Addition of Computer Exercises and a “Bridge to SPSS” Section at the End of Every Chapter

Although we believe that the performance of some simple hand calculations is vital for a complete understanding of the statistical procedures in this text, we

also recognize that most instructors these days require their students to learn how to calculate basic statistics with Microsoft Excel, or one of the comprehensive statistical software packages (e.g., SPSS, SAS, Minitab). Therefore, we have included one large data set in the appendix, which can also be downloaded in several useful formats from the web, and end-of-chapter computer exercises, all of which refer to this data set. The use of a single data set for all of the computer exercises facilitates comparisons among seemingly disparate, yet nonetheless closely related statistical procedures.

Because SPSS for Windows (SPSS, Chicago, Illinois) seems to be the most often used statistical package among our intended audience, we have included instructions that will help students to solve the computer exercises with SPSS, and to understand the output that SPSS creates. This built-in SPSS guide cannot take the place of a more general introductory guide to the SPSS package, and students will still need instructions concerning access to SPSS at their particular college or university. However, our Bridge to SPSS sections will ensure that students can connect the statistical terms that SPSS uses, which are not always universally recognized, with the language being used in this text.

It should be possible to solve most, if not all, of our computer exercises using any major statistical software package, but instructors using packages other than SPSS will need to provide their students with some guidance in translating their output into terms that are compatible with this text. Our Bridge to SPSS sections are based on version 14.0 of SPSS for Windows, the current version at the time of publication of this edition, but users of SPSS versions as early as 8.0 should find these sections just as applicable. If newer versions of SPSS for Windows, released before the next edition of this text, require any modifications of our Bridge to SPSS sections, these will be posted on the Cohen statistics web page.

3. The Addition of an Entire Chapter on Multiple Comparisons

The material on Fisher's protected *t* tests and LSD has been moved from the one-way ANOVA chapter in the fifth edition to a chapter of its own in this edition, and has been joined by Tukey's HSD, the modified LSD test, and the Bonferroni adjustment. The relative advantages and disadvantages of these alternatives are explained.

4. The Addition of a Chapter on Repeated-Measures ANOVA

This new chapter extends the scope of this text into an area that is now covered commonly in introductory statistics texts. This chapter appears *after* the two-way ANOVA, and can therefore take advantage of both the formulas and concepts of the two-way ANOVA to explain the workings of the repeated-measures and randomized-blocks ANOVA procedures. The new chapter concludes with a brief description of the computation and use of the two-way mixed-design ANOVA.

5. The Reorganization of Material from the Previous Edition

An important theme of this new edition has been the drawing of connections between statistical procedures from different chapters. Towards that end, we have reanalyzed, whenever feasible, the data sets of earlier chapters by the methods of subsequent chapters. Moreover, our text examples frequently refer to, as a unifying theme, the same hypothetical experiment that we used for our computer exercises.

Another important, and somewhat related, theme has been continuity. For example, the material on discrete probability (in chapter 8 of the fifth edition) was interrupting the flow between the description of standardized (z) scores, and their use with respect to the areas of the normal distribution. Consequently, we moved this material to a new chapter, just before “Chi-Square Tests,” to create an introduction to nonparametric statistics. Also, in accord with the suggestion of an anonymous reviewer, we reversed the order of chapters 6 and 7 from the fifth edition, so that the chapter on standardized scores directly precedes the description of the normal curve, and the chapter on exploratory data summaries follows directly after the chapters on central tendency and variability.

What has not changed in this edition is our continuing commitment to teaching our readers to understand the rationales that underlie the various statistical procedures explained herein, and to interpret and apply the results of those procedures without falling prey to common conceptual errors. Our emphasis therefore remains on the fundamental logic and proper use of descriptive and inferential statistics. This text was ahead of its time in highlighting the advantages of using confidence intervals and effect size measures, and making simple power calculations accessible to students at an introductory level. It is our hope that this revision brings our text not just up-to-date, but moves it once again to the forefront of statistical education for students of the behavioral sciences.

Acknowledgments

Thanks are due to our many encouraging friends and relatives, to colleagues and reviewers who made many useful comments on previous editions, and to two reviewers who gave us helpful suggestions for this latest edition: Dr. Marie S. Hammond, Tennessee State University; and Dr. Jasia Pietrzak, Columbia University. Most of all, we wish to thank our many students who, throughout the years, have provided invaluable feedback on our teaching of statistics, as well as on earlier editions of this text, and its accompanying workbook.

For the very existence of this latest edition, we owe our thanks to Patricia Rossi (senior editor) and Isabel Pratt (associate editor); without the vision and understanding of the editors and publishers at John Wiley & Sons, this revision would not have been published. We are also grateful to those responsible for the look and design of this text—especially, Linda Witzling (senior production editor, Wiley) and Susan Dodson (Graphic Composition, Inc., Athens, GA). Finally, we owe a debt of gratitude to Leona Gizzi for tirelessly typing the new chapters and sections from handwritten copy, and to Ihno Lee for her careful preparation of the answer key both in the appendix of this edition and in the accompanying instructor's manual.

Joan Welkowitz
Barry H. Cohen
Robert B. Ewen

Postscript

It is with much sadness that we note the passing of our senior author, Joan Welkowitz, early in 2006. She was very excited about the whole framework of this new edition, and remained involved in every stage of this revision, until the final copyediting. This edition is a tribute to her lifelong dedication to the teaching of statistics, and more generally, the mentoring of students of psychology. We will miss her greatly.

Barry H. Cohen
Robert B. Ewen

Glossary of Symbols

Numbers in parentheses indicate the chapter in which the symbol first appears.

a_{YX}	Y-intercept of linear regression line for predicting Y from X (12)
α	criterion (or level) of significance; probability of Type I error (9)
α_{EW}	the experiment-wise alpha (16)
α_{pc}	the alpha per comparison (16)
b_{YX}	slope of linear regression line for predicting Y from X (12)
β	probability of Type II error (9)
$1 - \beta$	power (14)
cf	cumulative frequency (2)
χ^2	statistic following the chi square distribution (20)
D	difference between two scores or ranks (11)
\bar{D}	mean of the Ds (11)
d	effect size involving two populations (14)
df	degrees of freedom (9)
df_B	degrees of freedom between groups (15)
df_W	degrees of freedom within groups (15)
df_1	degrees of freedom for factor 1 (17)
df_2	degrees of freedom for factor 2 (17)
$df_{1 \times 2}$	degrees of freedom for interaction (17)
δ	delta (14)
η^2	eta squared (15)
f	effect size involving multiple populations (15)
f	frequency (2)
f_e	expected frequency (20)
f_o	observed frequency (20)
F	statistic following the F distribution (15)
g	effect size involving two samples (14)
h	interval size (3)

H	statistic following the Kruskal—Wallis test (21)
$H\%$	percent of subjects in all intervals higher than the critical one (3)
H_0	null hypothesis (10)
H_1	alternative hypothesis (10)
HSD	Tukey's Honestly Significant Difference (16)
i	case number (1)
$I\%$	percent of subjects in the critical interval (3)
k	a constant (1)
k	number of groups (or the last group) (15)
$L\%$	percent of subjects in all intervals below the critical one (3)
LRL	lower real limit (3)
LSD	Fisher's Least Significant Difference (16)
Mdn	median (4)
MS	mean square (15)
MS_B	mean square between groups (15)
MS_W	mean square within groups (15)
MS_1	mean square for factor 1 (17)
MS_2	mean square for factor 2 (17)
$MS_{1 \times 2}$	mean square for interaction (17)
μ	population mean (4)
N_T	total number of subjects or observations (1)
N_i	number of observations or subjects in group i (15)
π	hypothetical population proportion (10)
p	observed sample proportion (10)
$P(A)$	probability of event A (19)
PR	percentile rank (3)
ϕ	phi coefficient (20)
ϕ_C	Cramér's ϕ (20)
q	studentized range statistic (16)
r_C	matched pairs rank biserial correlation coefficient (21)
r_G	Glass rank biserial correlation coefficient (21)
r_{pb}	point-biserial correlation coefficient (13)
r_s	Spearman rank-order correlation coefficient (21)
r_{XY}	sample Pearson correlation coefficient between X and Y (12)
\bar{R}	mean of a set of ranks (21)
ρ_{XY}	population correlation coefficient between X and Y (12)

s	sample standard deviation (5)
s^2	population variance estimate (5)
s_D^2	variance of the D s (11)
s_{pooled}^2	pooled variance (11)
$s_{\bar{X}}$	standard error of the mean (10)
$s_{\bar{X}_1 - \bar{X}_2}$	standard error of the difference (11)
$s_{Y'}$	estimate of $\sigma_{Y'}$ obtained from a sample (12)
$Score_p$	score corresponding to the p th percentile (3)
STB	sum of frequencies below the critical interval (3)
SS	sum of squares (15)
SS_T	total sum of squares (15)
SS_B	sum of squares between groups (15)
SS_W	sum of squares within groups (15)
SS_1	sum of squares for factor 1 (17)
SS_2	sum of squares for factor 2 (17)
$SS_{1 \times 2}$	sum of squares for interaction (17)
Σ	summation sign (1)
σ	population standard deviation (5)
σ^2	population variance (5)
σ_p	standard error of a sample proportion (10)
σ_T	standard error of the ranks of independent samples (21)
σ_{T_M}	standard error of the ranks of matched samples (21)
$\sigma_{\bar{X}}$	standard error of the mean when σ is known (10)
$\sigma_{Y'}$	standard error of estimate for predicting Y (12)
t	statistic following the t distribution (10)
T	T score (7)
T_E	expected sum of the ranks (21)
T_i	sum of ranks in group i (21)
x	deviation score (4)
X'	predicted X score (12)
\bar{X}	sample mean (4)
\bar{X}_i	mean of group i (15)
\bar{X}_G	grand mean (15)
Y'	predicted Y score (12)
z	standard score (7)

Contents

Preface	xv
Acknowledgments	xix
Glossary of Symbols	xxi

Part I **Descriptive Statistics** **1**

Chapter 1 **Introduction** **3**

Why Study Statistics?	4
Descriptive and Inferential Statistics	5
Populations, Samples, Parameters, and Statistics	6
Measurement Scales	6
Independent and Dependent Variables	8
Sara's Study	9
Summation Notation	10
Summary	16
Exercises	17
Thought Questions	20
Computer Exercises	21
Bridge to SPSS	21

Chapter 2 **Frequency Distributions and Graphs** **23**

The Purpose of Descriptive Statistics	24
Regular Frequency Distributions	25
Cumulative Frequency Distributions	26
Grouped Frequency Distributions	27
Graphic Representations	30
Shapes of Frequency Distributions	35
Summary	37
Exercises	38
Thought Questions	39
Computer Exercises	40
Bridge to SPSS	40

Chapter 3	Transformed Scores I: Percentiles	42
	Interpreting a Raw Score	43
	Definition of Percentile and Percentile Rank	43
	Computational Procedures	44
	Deciles, Quartiles, and the Median	52
	Summary	52
	Exercises	53
	Thought Questions	54
	Computer Exercises	54
	Bridge to SPSS	54
Chapter 4	Measures of Central Tendency	56
	Introduction	57
	The Mean	58
	The Median	64
	The Mode	66
	Summary	66
	Exercises	67
	Thought Questions	67
	Computer Exercises	68
	Bridge to SPSS	68
Chapter 5	Measures of Variability	69
	The Concept of Variability	70
	The Range	72
	The Semi-Interquartile Range	73
	The Standard Deviation and Variance	74
	Summary	80
	Exercises	82
	Thought Questions	83
	Computer Exercises	83
	Bridge to SPSS	84
Chapter 6	Additional Techniques for Describing Batches of Data	85
	Numerical Summaries	86
	Graphic Summaries	88
	Summary	91

Exercises	91
Thought Questions	92
Computer Exercises	92
Bridge to SPSS	92

Chapter 7 Transformed Scores II: z and T Scores 94

Interpreting a Raw Score	95
Rules for Changing \bar{X} and σ	96
Standard Scores (z Scores)	98
T Scores and SAT Scores	100
IQ Scores	102
Summary	103
Exercises	104
Thought Questions	106
Computer Exercises	106
Bridge to SPSS	106

Chapter 8 The Normal Distribution 108

Introduction	109
Score Distributions	110
Parameters of the Normal Distribution	111
Tables of the Standard Normal Distribution	111
Characteristics of the Normal Curve	112
Illustrative Examples	113
Summary	119
Exercises	120
Thought Questions	121
Computer Exercises	121
Bridge to SPSS	121

Part II Basic Inferential Statistics 123

Chapter 9 Introduction to Statistical Inference 125

Introduction	126
The Goals of Inferential Statistics	127

Sampling Distributions	128
The Standard Error of the Mean	132
The z Score for Sample Means	135
Null Hypothesis Testing	137
Assumptions Required by the Statistical Test for the Mean of a Single Population	144
Summary	144
Exercises	146
Thought Questions	148
Computer Exercises	149
Bridge to SPSS	149

Chapter 10 The One-Sample t Test and Interval Estimation 150

The Statistical Test for the Mean of a Single Population When σ Is Not Known: The t Distributions	151
Interval Estimation	155
The Standard Error of a Proportion	159
Summary	162
Exercises	164
Thought Questions	165
Computer Exercises	166
Bridge to SPSS	166

Chapter 11 Testing Hypotheses about the Difference between the Means of Two Populations 167

The Standard Error of the Difference	169
Estimating the Standard Error of the Difference	173
The t Test for Two Sample Means	174
Confidence Intervals for the Difference of Two Population Means	177
Using the t Test for Two Sample Means: Some General Considerations	179
Measuring Size of an Effect	181
The t Test for Matched Samples	182
Summary	188
Exercises	191
Thought Questions	193
Computer Exercises	195
Bridge to SPSS	195

Chapter 12	Linear Correlation and Prediction	197
	Introduction	198
	Describing the Linear Relationship between Two Variables	201
	Interpreting the Magnitude of a Pearson r	210
	When Is It Important That Pearson's r be Large?	212
	Testing the Significance of the Correlation Coefficient	214
	Prediction and Linear Regression	217
	Measuring Prediction Error: The Standard Error of Estimate	225
	Summary	228
	Exercises	230
	Thought Questions	233
	Computer Exercises	234
	Bridge to SPSS	235
	Appendix: Equivalence of the Various Formulas for r	236
Chapter 13	The Connection between Correlation and the t Test	241
	Introduction	242
	The Point-Biserial Correlation Coefficient	243
	The Proportion of Variance Accounted For in Your Samples	246
	Estimating the Proportion of Variance Accounted For in the Population	247
	Summary	249
	Exercises	250
	Thought Questions	251
	Computer Exercises	252
	Bridge to SPSS	252
Chapter 14	Introduction to Power Analysis	255
	Introduction	256
	Concepts of Power Analysis	257
	The Test of the Mean of a Single Population	259
	The Significance Test of the Proportion of a Single Population	264
	The Significance Test of a Pearson r	266
	Testing the Difference between Independent Means	267
	Testing the Difference between the Means of Two Matched Populations	272
	Choosing a Value for d for a Power Analysis Involving Independent Means	273