

Cambridge University Press

978-1-107-46122-2 - Statistics Using IBM SPSS: An Integrative Approach: Third Edition

Sharon Lawner Weinberg and Sarah Knapp Abramowitz

Frontmatter

[More information](#)

STATISTICS USING IBM SPSS, THIRD EDITION

Written in a clear and lively tone, *Statistics Using IBM SPSS* provides a data-centric approach to statistics with integrated SPSS (version 22) commands, ensuring that students gain both a deep conceptual understanding of statistics and practical facility with the leading statistical software package. With one hundred worked examples, the textbook guides students through statistical practice using real data and avoids complicated mathematics. Numerous end-of-chapter exercises allow students to apply and test their understanding of chapter topics, with detailed answers available online. The third edition has been updated throughout and includes a new chapter on research design, new topics (including weighted mean, resampling with the bootstrap, the role of the syntax file in workflow management, and regression to the mean), and new examples and exercises. Student learning is supported by a rich suite of online resources, including answers to end-of-chapter exercises, real data sets, PowerPoint slides, and a test bank.

Sharon Lawner Weinberg is Professor of Applied Statistics and Psychology and former Vice Provost for Faculty Affairs at New York University. She has authored numerous articles, books, and reports on statistical methods, statistical education, and evaluation, as well as in applied disciplines, such as psychology, education, and health. She is the recipient of several major grants, including a recent grant from the Sloan Foundation to support her current work with NYU colleagues to evaluate the New York City Gifted and Talented programs.

Sarah Knapp Abramowitz is Professor of Mathematics and Computer Science at Drew University. She received her PhD in Mathematics Education from New York University and is an Associate Editor of the *Journal of Statistics Education*.

Cambridge University Press

978-1-107-46122-2 - Statistics Using IBM SPSS: An Integrative Approach: Third Edition

Sharon Lawner Weinberg and Sarah Knapp Abramowitz

Frontmatter

[More information](#)

Statistics Using IBM SPSS

AN INTEGRATIVE APPROACH Third Edition

SHARON LAWNER WEINBERG

New York University

SARAH KNAPP ABRAMOWITZ

Drew University

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107461222

© Sharon Lawner Weinberg and Sarah Knapp Abramowitz 2015

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2015

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Weinberg, Sharon L.

[Statistics using SPSS]

Statistics using IBM SPSS : an integrative approach / Sharon Lawner Weinberg, New York University, Sarah Knapp Abramowitz, Drew University. – Third edition.
pages cm

Previous edition: Statistics using SPSS : an integrative approach / Sharon L. Weinberg, Sarah Knapp Abramowitz (Cambridge University Press, 2008).

Includes bibliographical references and index.

ISBN 978-1-107-46122-2 (pbk.)

1. Mathematical statistics – Data processing. 2. SPSS (Computer file) 3. Mathematical statistics – Textbooks. I. Abramowitz, Sarah Knapp, 1967– II. Title.

QA276.W4423 2015

519.50285–dc23 2015017084

ISBN 978-1-107-46122-2 Paperback

Additional resources for this publication at www.cambridge.org/weinberg3.

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

Cambridge University Press

978-1-107-46122-2 - Statistics Using IBM SPSS: An Integrative Approach: Third Edition

Sharon Lawner Weinberg and Sarah Knapp Abramowitz

Frontmatter

[More information](#)

To our families

Cambridge University Press

978-1-107-46122-2 - Statistics Using IBM SPSS: An Integrative Approach: Third Edition

Sharon Lawner Weinberg and Sarah Knapp Abramowitz

Frontmatter

[More information](#)

Contents

<i>Preface</i>	<i>page xv</i>
<i>Acknowledgments</i>	<i>xix</i>
1 INTRODUCTION	1
The Role of the Computer in Data Analysis	1
Statistics: Descriptive and Inferential	2
Variables and Constants	3
The Measurement of Variables	3
Discrete and Continuous Variables	8
Setting a Context with Real Data	11
Exercises	12
2 EXAMINING UNIVARIATE DISTRIBUTIONS	20
Counting the Occurrence of Data Values	20
When Variables Are Measured at the Nominal Level	20
Frequency and Percent Distribution Tables, 20 • Bar Graphs, 21 • Pie Graphs, 23	
When Variables Are Measured at the Ordinal, Interval, or Ratio Level	25
Frequency and Percent Distribution Tables, 25 • Stem-and-Leaf Displays, 27 • Histograms, 30 • Line Graphs, 33	
Describing the Shape of a Distribution	36
Accumulating Data	38
Cumulative Percent Distributions	38
Ogive Curves	38
Percentile Ranks	39
Percentiles	40
Five-Number Summaries and Boxplots	43
Summary of Graphical Selection	49
Exercises	49
3 MEASURES OF LOCATION, SPREAD, AND SKEWNESS	65
Characterizing the Location of a Distribution	65
The Mode	65
The Median	69
The Arithmetic Mean	70
Interpreting the Mean of a Dichotomous Variable, 72 • The Weighted Mean, 73	
Comparing the Mode, Median, and Mean	74

	Characterizing the Spread of a Distribution	76
	The Range and Interquartile Range	79
	The Variance	80
	The Standard Deviation	83
	Characterizing the Skewness of a Distribution	84
	Selecting Measures of Location and Spread	86
	Applying What We Have Learned	86
	Exercises	90
4	RE-EXPRESSING VARIABLES	99
	Linear and Nonlinear Transformations	99
	Linear Transformations: Addition, Subtraction, Multiplication, and Division	100
	The Effect on the Shape of a Distribution	101
	The Effect on Summary Statistics of a Distribution	102
	Common Linear Transformations	106
	Standard Scores	107
	z-Scores	109
	Using z-Scores to Detect Outliers, 111 • Using z-Scores to Compare Scores in Different Distributions, 112 • Relating z-Scores to Percentile Ranks, 115	
	Nonlinear Transformations: Square Roots and Logarithms	115
	Nonlinear Transformations: Ranking Variables	123
	Other Transformations: Recoding and Combining Variables	124
	Recoding Variables	124
	Combining Variables	126
	Data Management Fundamentals – The Syntax File	126
	Exercises	130
5	EXPLORING RELATIONSHIPS BETWEEN TWO VARIABLES	138
	When Both Variables Are at Least Interval-Leveled	138
	Scatterplots	139
	The Pearson Product Moment Correlation Coefficient	147
	Interpreting the Pearson Correlation Coefficient, 152 • The Correlation Scale Itself Is Ordinal, 153 • Correlation Does Not Imply Causation, 153 • The Effect of Linear Transformations, 154 • Restriction of Range, 154 • The Shape of the Underlying Distributions, 155 • The Reliability of the Data, 155	
	When at Least One Variable Is Ordinal and the Other Is at Least Ordinal:	
	The Spearman Rank Correlation Coefficient	155
	When at Least One Variable Is Dichotomous: Other Special Cases of the Pearson Correlation Coefficient	157
	The Point Biserial Correlation Coefficient: The Case of One at Least Interval and One Dichotomous Variable	157
	The Phi Coefficient: The Case of Two Dichotomous Variables	162
	Other Visual Displays of Bivariate Relationships	167
	Selection of Appropriate Statistic/Graph to Summarize a Relationship	170
	Exercises	171
6	SIMPLE LINEAR REGRESSION	183
	The “Best-Fitting” Linear Equation	183
	The Accuracy of Prediction Using the Linear Regression Model	190

CONTENTS

ix

The Standardized Regression Equation	191
<i>R</i> as a Measure of the Overall Fit of the Linear Regression Model	191
Simple Linear Regression When the Independent Variable Is Dichotomous	196
Using <i>r</i> and <i>R</i> as Measures of Effect Size	199
Emphasizing the Importance of the Scatterplot	199
Exercises	201
7 PROBABILITY FUNDAMENTALS	210
The Discrete Case	210
The Complement Rule of Probability	212
The Additive Rules of Probability	213
First Additive Rule of Probability	213
Second Additive Rule of Probability	214
The Multiplicative Rule of Probability	215
The Relationship between Independence and Mutual Exclusivity	218
Conditional Probability	218
The Law of Large Numbers	220
Exercises	220
8 THEORETICAL PROBABILITY MODELS	223
The Binomial Probability Model and Distribution	223
The Applicability of the Binomial Probability Model	228
The Normal Probability Model and Distribution	232
Using the Normal Distribution to Approximate the Binomial Distribution	238
Exercises	239
9 THE ROLE OF SAMPLING IN INFERENCE STATISTICS	245
Samples and Populations	245
Random Samples	246
Obtaining a Simple Random Sample	247
Sampling with and without Replacement	249
Sampling Distributions	250
Describing the Sampling Distribution of Means Empirically	251
Describing the Sampling Distribution of Means Theoretically:	
The Central Limit Theorem	255
Central Limit Theorem (CLT)	255
Estimators and Bias	259
Exercises	260
10 INFERENCES INVOLVING THE MEAN OF A SINGLE POPULATION	
WHEN σ IS KNOWN	264
Estimating the Population Mean, μ , When the Population Standard Deviation, σ , Is Known	264
Interval Estimation	266
Relating the Length of a Confidence Interval, the Level of Confidence, and the Sample Size	269
Hypothesis Testing	270
The Relationship between Hypothesis Testing and Interval Estimation	278
Effect Size	279

Type II Error and the Concept of Power	280
Increasing the Level of Significance, α	284
Increasing the Effect Size, δ	284
Decreasing the Standard Error of the Mean, $\sigma_{\bar{x}}$	284
Closing Remarks	285
Exercises	286
11 INFERENCES INVOLVING THE MEAN WHEN σ IS NOT KNOWN: ONE- AND TWO-SAMPLE DESIGNS	290
Single Sample Designs When the Parameter of Interest Is the Mean and σ Is Not Known	290
The t Distribution	291
Degrees of Freedom for the One-Sample t -Test	292
Violating the Assumption of a Normally Distributed Parent Population in the One-Sample t -Test	293
Confidence Intervals for the One-Sample t -Test	294
Hypothesis Tests: The One-Sample t -Test	298
Effect Size for the One-Sample t -Test	300
Two Sample Designs When the Parameter of Interest Is μ, and σ Is Not Known	303
Independent (or Unrelated) and Dependent (or Related) Samples	304
Independent Samples t -Test and Confidence Interval	305
The Assumptions of the Independent Samples t -Test	307
Effect Size for the Independent Samples t -Test	315
Paired Samples t-Test and Confidence Interval	317
The Assumptions of the Paired Samples t -Test	318
Effect Size for the Paired Samples t -Test	323
The Bootstrap	323
Summary	327
Exercises	328
12 RESEARCH DESIGN: INTRODUCTION AND OVERVIEW	346
Questions and Their Link to Descriptive, Relational, and Causal Research Studies	346
The Need for a Good Measure of Our Construct, Weight	346
• The Descriptive Study, 347	
From Descriptive to Relational Studies	348
From Relational to Causal Studies	348
The Gold Standard of Causal Studies: The True Experiment and Random Assignment	350
Comparing Two Kidney Stone Treatments Using a Non-randomized Controlled Study	351
Including Blocking in a Research Design	352
Underscoring the Importance of Having a True Control Group Using Randomization	353
Analytic Methods for Bolstering Claims of Causality from Observational Data (Optional Reading)	357
Quasi-Experimental Designs	359
Threats to the Internal Validity of a Quasi-Experimental Design	360
Threats to the External Validity of a Quasi-Experimental Design	361
Threats to the Validity of a Study: Some Clarifications and Caveats	361
Threats to the Validity of a Study: Some Examples	362
Exercises	363

CONTENTS

xi

13 ONE-WAY ANALYSIS OF VARIANCE	367
The Disadvantage of Multiple <i>t</i> -Tests	367
The One-Way Analysis of Variance	369
A Graphical Illustration of the Role of Variance in Tests on Means	369
ANOVA as an Extension of the Independent Samples <i>t</i> -Test	370
Developing an Index of Separation for the Analysis of Variance	371
Carrying Out the ANOVA Computation	372
• The Between Group Variance (MS_B), 372 • The Within Group Variance (MS_W), 373	
The Assumptions of the One-Way ANOVA	374
Testing the Equality of Population Means: The <i>F</i> -Ratio	375
How to Read the Tables and to Use the SPSS Compute Statement for the <i>F</i> Distribution	375
ANOVA Summary Table	380
Measuring the Effect Size	380
Post-hoc Multiple Comparison Tests	384
The Bonferroni Adjustment: Testing Planned Comparisons	394
The Bonferroni Tests on Multiple Measures	396
Exercises	399
14 TWO-WAY ANALYSIS OF VARIANCE	404
The Two-Factor Design	404
The Concept of Interaction	407
The Hypotheses That Are Tested by a Two-Way Analysis of Variance	412
Assumptions of the Two-Way Analysis of Variance	413
Balanced versus Unbalanced Factorial Designs	414
Partitioning the Total Sum of Squares	414
Using the <i>F</i> -Ratio to Test the Effects in Two-Way ANOVA	415
Carrying Out the Two-Way ANOVA Computation by Hand	416
Decomposing Score Deviations about the Grand Mean	420
Modeling Each Score as a Sum of Component Parts	421
Explaining the Interaction as a Joint (or Multiplicative) Effect	421
Measuring Effect Size	422
Fixed versus Random Factors	426
Post-hoc Multiple Comparison Tests	426
Summary of Steps to Be Taken in a Two-Way ANOVA Procedure	432
Exercises	437
15 CORRELATION AND SIMPLE REGRESSION AS INFERENTIAL TECHNIQUES	445
The Bivariate Normal Distribution	445
Testing Whether the Population Pearson Product Moment Correlation Equals Zero	448
Using a Confidence Interval to Estimate the Size of the Population Correlation Coefficient, ρ	451
Revisiting Simple Linear Regression for Prediction	454
Estimating the Population Standard Error of Prediction, $\sigma_{Y X}$	455
Testing the <i>b</i> -Weight for Statistical Significance	456
Explaining Simple Regression Using an Analysis of Variance Framework	459
Measuring the Fit of the Overall Regression Equation: Using <i>R</i> and R^2	462
Relating R^2 To $\sigma^2_{Y X}$	463

Testing R^2 for Statistical Significance	463
Estimating the True Population R^2 : The Adjusted R^2	464
Exploring the Goodness of Fit of the Regression Equation: Using Regression Diagnostics	465
Residual Plots: Evaluating the Assumptions Underlying Regression	467
Detecting Influential Observations: Discrepancy and Leverage	470
Using SPSS to Obtain Leverage	472
Using SPSS to Obtain Discrepancy	472
Using SPSS to Obtain Influence	473
Using Diagnostics to Evaluate the Ice Cream Sales Example	474
Using the Prediction Model to Predict Ice Cream Sales	478
Simple Regression When the Predictor Is Dichotomous	478
Exercises	479
16 AN INTRODUCTION TO MULTIPLE REGRESSION	491
The Basic Equation with Two Predictors	492
Equations for b , β , and $R_{Y,12}$ When the Predictors Are Not Correlated	493
Equations for b , β , and $R_{Y,12}$ When the Predictors Are Correlated	494
Summarizing and Expanding on Some Important Principles of Multiple Regression	496
Testing the b -Weights for Statistical Significance	501
Assessing the Relative Importance of the Independent Variables in the Equation	503
Measuring the Drop in R^2 Directly: An Alternative to the Squared Part Correlation	504
Evaluating the Statistical Significance of the Change in R^2	504
The b -Weight as a Partial Slope in Multiple Regression	505
Multiple Regression When One of the Two Independent Variables Is Dichotomous	508
The Concept of Interaction between Two Variables That Are at Least Interval-Levelled	514
Testing the Statistical Significance of an Interaction Using SPSS	516
Centering First-Order Effects to Achieve Meaningful Interpretations of b -Weights	521
Understanding the Nature of a Statistically Significant Two-Way Interaction	521
Interaction When One of the Independent Variables Is Dichotomous and the Other Is Continuous	524
Exercises	528
17 NONPARAMETRIC METHODS	539
Parametric versus Nonparametric Methods	539
Nonparametric Methods When the Dependent Variable Is at the Nominal Level	540
The Chi-Square Distribution (χ^2)	540
The Chi-Square Goodness-of-Fit Test	542
The Chi-Square Test of Independence	547
• Assumptions of the Chi-Square Test of Independence	550
Fisher's Exact Test	552
• Calculating the Fisher Exact Test by Hand Using the Hypergeometric Distribution	554
Nonparametric Methods When the Dependent Variable Is Ordinal-Levelled	558
Wilcoxon Sign Test	558
The Mann-Whitney U Test	561
The Kruskal-Wallis Analysis of Variance	565
Exercises	567

CONTENTS

xiii

<i>Appendix A</i>	<i>Data Set Descriptions</i>	573
<i>Appendix B</i>	<i>Generating Distributions for Chapters 8 and 9 Using SPSS Syntax</i>	586
<i>Appendix C</i>	<i>Statistical Tables</i>	587
<i>Appendix D</i>	<i>References</i>	588
<i>Appendix E</i>	<i>Solutions to Exercises</i>	592
<i>Appendix F</i>	<i>The Standard Error of the Mean Difference for Independent Samples: A More Complete Account (Optional)</i>	593
	<i>Index</i>	595

Cambridge University Press

978-1-107-46122-2 - Statistics Using IBM SPSS: An Integrative Approach: Third Edition

Sharon Lawner Weinberg and Sarah Knapp Abramowitz

Frontmatter

[More information](#)

Preface

This text, in its third edition, capitalizes on the widespread availability of menu-driven software packages to create a course of study that links good statistical practice to the analysis of real data, and the many years of our combined experience teaching statistics to undergraduate students at a liberal arts university and to graduate students at a large research university from a variety of disciplines including education, psychology, sociology, health, and policy analysis. The third edition continues to embrace and be motivated by several important guiding principles.

First, and perhaps most important, we believe that a good data analytic plan must serve to uncover the story behind the numbers, what the data tell us about the phenomenon under study. To begin, a good data analyst must know his/her data well and have confidence that it satisfies the underlying assumptions of the statistical methods used. Accordingly, we emphasize the usefulness of diagnostics in both graphical and statistical form to expose anomalous cases, which might unduly influence results, and to help in the selection of appropriate assumption-satisfying transformations so that ultimately we may have confidence in our findings. We also emphasize the importance of using more than one method of analysis to answer fully the question posed and understanding potential bias in the estimation of population parameters. In keeping with these principles, the third edition contains an even more comprehensive coverage of essential topics in introductory statistics not covered by other such textbooks, including robust methods of estimation based on resampling using the bootstrap, regression to the mean, the weighted mean, Simpson's Paradox, counterfactuals and other topics in research design, and data workflow management using the SPSS syntax file. A central feature of the book that continues to be embraced in the third edition is the integration of SPSS in a way that reflects practice and allows students to learn SPSS along with each new statistical method.

Second, because we believe that data are central to the study of good statistical practice, the textbook's website contains several data sets used throughout the text. Two are large sets of real data that we make repeated use of in both worked-out examples and end-of-chapter exercises. One data set contains forty-eight variables and five hundred cases from the education discipline; the other contains forty-nine variables and nearly forty-five hundred cases from the health discipline. By posing interesting questions about variables in these large, real data sets (e.g., is there a gender difference in eighth graders' expected income at age thirty?), we are able to employ a more meaningful and contextual approach to the introduction of statistical methods and to engage students more actively in the learning process. The repeated use of these data sets also contributes to creating a more cohesive presentation of statistics; one that links different methods of analysis to each other and

avoids the perception that statistics is an often-confusing array of so many separate and distinct methods of analysis, with no bearing or relationship to one another.

Third, we believe that the result of a null hypothesis test (to determine whether an effect is real or merely apparent), is only a means to an end (to determine whether the effect being studied is important or useful), rather than an end in itself. Accordingly, in our presentation of null hypothesis testing, we stress the importance of evaluating the magnitude of the effect if it is deemed to be real, and of drawing clear distinctions between statistically significant and substantively significant results. Toward this end, we introduce the computation of standardized measures of effect size as common practice following a statistically significant result. While we provide guidelines for evaluating, in general, the magnitude of an effect, we encourage readers to think more subjectively about the magnitude of an effect, bringing into the evaluation their own knowledge and expertise in a particular area.

Fourth, a course in applied statistics should not only provide students with a sound statistical knowledge base but also with a set of data analytic skills. Accordingly, we have incorporated the latest version of SPSS, a popularly-used statistical software package, into the presentation of statistical material using a highly integrative approach. SPSS is used to provide students with a platform for actively engaging in the learning process associated with what it means to be a good data analyst by allowing them to apply their newly-learned knowledge to the real world of applications. This approach serves also to enhance the conceptual understanding of material and the ability to interpret output and communicate findings.

Finally, we believe that a key ingredient of an introductory statistics text is a lively, clear, conceptual, yet rigorous approach. We emphasize conceptual understanding through an exploration of both the mathematical principles underlying statistical methods and real world applications. We use an easy-going, informal style of writing that we have found gives readers the impression that they are involved in a personal conversation with the authors. And, we sequence concepts with concern for student readiness, reintroducing topics in a spiralling manner to provide reinforcement and promote the transfer of learning.

New to the third edition are the addition of other essential topics in introductory statistics, including robust methods of estimation based on resampling using the bootstrap, regression to the mean, the weighted mean and Simpson's Paradox, counterfactuals, potential sources of bias in the estimation of population parameters based on the analysis of data from quasi-experimental designs, other issues related to research design contained in a new chapter on research design, the importance of the SPSS Syntax file in workflow management, an expanded bibliography of references to relevant books and journal articles, and many more end-of-chapter exercises, with detailed answers on the textbook's website. Along with topics from the second edition, such as data transformations, diagnostic tools for the analysis of model fit, the logic of null hypothesis testing, assessing the magnitude of effects, interaction and its interpretation in two-way analysis of variance and multiple regression, and nonparametric statistics, the third edition provides comprehensive coverage of essential topics in introductory statistics. In so doing, the third edition gives instructors flexibility in curriculum planning and provides students with more advanced material for future work in statistics. Also new is a companion website that includes copies of the data sets and other ancillary materials. These materials are available at www.cambridge.org/weinberg3appendix under the Resources tab.

PREFACE

xvii

The book, consisting of seventeen chapters, is intended for use in a one- or two-semester introductory applied statistics course for the behavioral, social, or health sciences at either the graduate or undergraduate level, or as a reference text as well. It is not intended for readers who wish to acquire a more theoretical understanding of mathematical statistics. To offer another perspective, the book may be described as one that begins with modern approaches to Exploratory Data Analysis (EDA) and descriptive statistics, and then covers material similar to what is found in an introductory mathematical statistics text, such as for undergraduates in math and the physical sciences, but stripped of calculus and linear algebra and instead grounded in data examples. Thus, theoretical probability distributions, The Law of Large Numbers, sampling distributions, and The Central Limit Theorem are all covered, but in the context of solving practical and interesting problems.

Cambridge University Press

978-1-107-46122-2 - Statistics Using IBM SPSS: An Integrative Approach: Third Edition

Sharon Lawner Weinberg and Sarah Knapp Abramowitz

Frontmatter

[More information](#)

Acknowledgments

This book has benefited from the many helpful comments of our New York University and Drew University students, too numerous to mention by name, and from the insights and suggestions of several colleagues. For their help, we would like to thank (in alphabetical order) Chris Apelian, Gabriella Belli, Patricia Busk, Ellie Buteau, John Daws, Michael Karchmer, Steve Kass, Jon Kettenring, Linda Lesniak, Kathleen Madden, Joel Middleton, Robert Norman, Eileen Rodriguez, and Marc Scott. Of course, any errors or shortcomings in the third edition remain the responsibility of the authors.