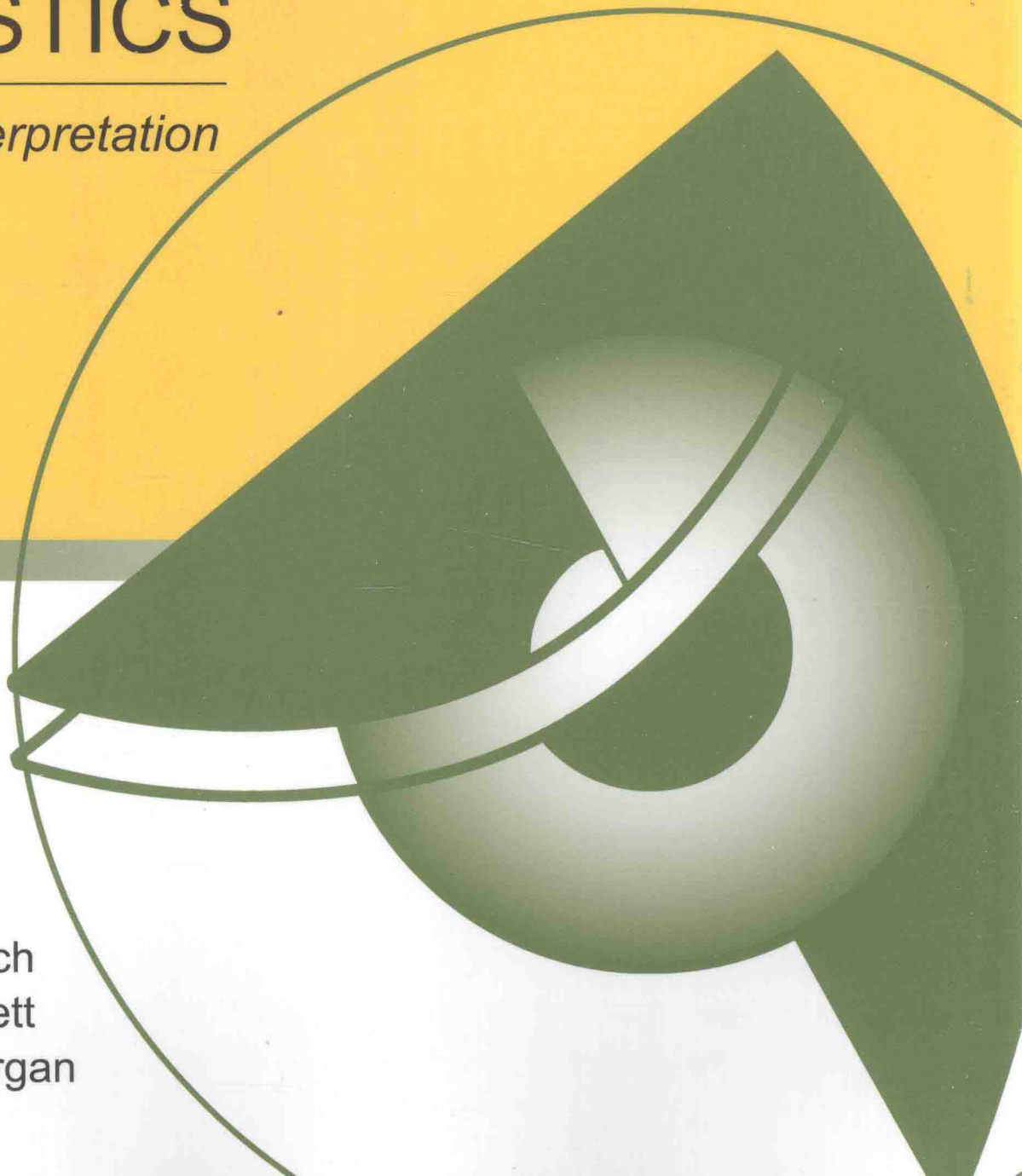
IBM SPSS FOR INTERMEDIATE STATISTICS

Use and Interpretation

Fifth Edition



Nancy L. Leech
Karen C. Barrett
George A. Morgan

ROUTLEDGE

IBM SPSS FOR INTERMEDIATE STATISTICS

Use and Interpretation

Fifth Edition

Nancy L. Leech

Karen C. Barrett

George A. Morgan

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2015
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
27 Church Road, Hove, East Sussex BN3 2FA

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data
Leech, Nancy L.
[SPSS for intermediate statistics]
IBM SPSS for intermediate statistics : use and interpretation / authored by Nancy L. Leech, Karen C. Barrett, and George A. Morgan. -- Fifth edition.
pages cm
Includes index.
ISBN 978-1-138-80117-2 (hardback) -- ISBN 978-1-84872-999-5 (pbk.) -- ISBN 978-0-203-12277-8 (ebook) 1. SPSS for Windows. 2. SPSS (Computer file) 3. Social sciences--Statistical methods--Computer programs. I. Barrett, Karen Caplovitz. II. Morgan, George A. (George Arthur), 1936- III. Title.
HA32.L444 2015
005.5'5--dc23
2014012307

ISBN: [978-1-138-80117-2] (hbk)
ISBN: [978-1-84872-999-5] (pbk)
ISBN: [978-0-203-12277-8] (ebk)

Typeset in Times New Roman

Publisher's Note
This book has been prepared from camera-ready copy provided by the authors.

MIX
Paper from responsible sources
FSC
www.fsc.org
FSC® C014174

Printed and bound in the United States of America by Sheridan Books, Inc. (a Sheridan Group Company).

PREFACE

This book is designed to help students learn how to analyze and interpret research data with intermediate statistics. It is intended to be a supplemental text in an intermediate statistics course in the behavioral sciences, social sciences, or education and it can be used in conjunction with any mainstream text. We have found that the book makes IBM SPSS easy to use so that it is not necessary to have a formal, instructional computer lab; you should be able to learn how to use SPSS on your own with this book. Access to the SPSS program and some familiarity with Windows® is all that is required. Although SPSS is quite easy to use, there is such a wide variety of options and statistics that knowing which ones to use and how to interpret the printouts can be difficult. This book is intended to help with these challenges.

SPSS 22 and Earlier Versions

We use SPSS 22 from IBM SPSS in this book; except for enhanced tables and graphics, there are only minor differences from SPSS version 10 to 22. We expect future Windows and Mac versions to be similar. Our students have used earlier editions of this book with all of the recent versions of SPSS; most of the procedures and outputs are the same or very similar. We point out some of the changes at various points in the text.

In addition to various SPSS modules that may be available at your university, there are versions available to students that you can rent for 6 or 12 months online. You can get information about available products for graduate students by visiting http://www-01.ibm.com/software/analytics/spss/products/statistics/gradpack/. *IBM SPSS Statistics Standard GradPack* enables you to do all the statistics in this book (except for data imputation in chapter 13), those in our *IBM SPSS for Introductory Statistics,* 5th edition (Morgan, Leech, Gloeckner, & Barrett, 2013), and many others.

Goals of This Book

This book demonstrates how to produce a variety of statistics that are usually included in intermediate statistics courses, plus some (e.g., reliability measures, canonical correlation, and multilevel models) that are unusual in intermediate statistics books. We also show imputation methods to deal with missing data, and we show how one can see if a variable is a mediator or moderator. These methods may be useful in your research. Our goal is to describe the use and interpretation of these statistics as much as possible in nontechnical, jargon-free language.

Helping you learn how to choose the appropriate statistics, interpret the outputs, and develop skills in writing about the meaning of the results are the main goals of this book. Thus, we have included material on:

1. How the appropriate **choice of a statistic** is based in good part on the design of the research and the measurement level of the variables.
2. How to use SPSS to **help answer research questions**.
3. How to **interpret SPSS outputs**.
4. How to **write about the outputs** in the Results section of a paper.

This information will help you develop skills that cover all steps in the research process: design, data collection, data entry, data analysis, interpretation of outputs, and writing results. The modified high school and beyond data set (HSB) used for many of the chapters in this book is similar to one you might have for a thesis, dissertation, or research project. Therefore, we think it can serve as a model for your analysis. The Web site, www.routledge.com/9781848729995 contains the HSB data file and several other data sets that are used in the book and for the extra SPSS problems at the end of chapters. However, **you will need to have access to or purchase the SPSS program.**

To make the text more readable, we have chosen not to cite many references in the text; however, we have provided a short bibliography, "For Further Reading," of some of the books and articles that we have found useful. We assume that most students will use this book in conjunction with a class that has a statistics textbook; it will help you to read more about each statistic before doing the assignments.

Our companion book, Morgan et al. (2013), *IBM SPSS for Introductory Statistics: Use and Interpretation* (5th ed.), also published by Routledge/Taylor & Francis, is on the *For Further Reading* list at the end of this book. To learn more about that book feel free to visit www.routledge.com/9781848729827. Our introductory book provides an extended discussion of how to interpret and write about introductory statistics, including ones such as *t* tests, chi-square, and correlation. A brief review of such basic statistics is provided in Appendix B of this book.

Instructional Features

Several user-friendly features of this book include:

1. The **key SPSS windows** that you see when performing the statistical analyses. This has been helpful to "visual learners."
2. The **outputs** for almost all analyses that we have done so you can see what you will get, after some editing in SPSS to make the outputs fit better on the pages.
3. **Callout boxes** on the outputs that point out parts of the output to focus on and indicate what they mean.
4. For each output, boxed **interpretation sections** will help you understand the output.
5. Specially developed flow charts and tables to help you **select an appropriate inferential statistic** and help you **interpret statistical significance and effect sizes** (in Chapter 5). That chapter also provides an extended example of how to identify and write a research problem, several research questions, and a results paragraph for a *t* test and bivariate regression.
6. For the statistics in Chapters 3–4 and 6–13, an example of **how to write about the output** and make a table for a thesis, dissertation, or research paper in APA format.
7. **Interpretation questions** that stimulate you to think about the information in the chapter and outputs.
8. Several **extra SPSS problems** at the end of each chapter, except Chapters 1 and 5, for you to run with SPSS and discuss.
9. Information (in Appendix A) on how to **get started with SPSS** and some other useful commands.
10. A brief review (Appendix B) of **basic statistics**.
11. Appendix C provides examples of how to write **research problems** and **research questions or hypotheses.**
12. **Answers** to the odd-numbered interpretation questions (Appendix D). Answers to the even-numbered questions are available to instructors.
13. **Several data sets are available on the book Website** www.routledge.com/9781848729995. These realistic data sets provide you with data to be used to solve the chapter problems and the extra SPSS problems at the end of the chapters. Also on the website are three other files: (a) a **Quick Reference Guide (QRG)** to commonly used SPSS procedures, (b) a document, **Making APA Tables and Figures**, describing how to make tables in APA format, and (c) a file to use with the syntax for Canonical Correlation in Chapter 7.
14. An **Instructor Resource Web site** is available to course instructors who request access from the publisher. To request access, please visit the website below. It contains aids for teaching the course, including PowerPoint® slides, the answers to the even–numbered interpretation questions, and extra SPSS problems. Students will benefit from the chapter outlines and study guides. The study guide portion includes a list of key concepts to remember and define after reading each chapter. Researchers who purchase copies for their personal use can access the data files by visiting www.routledge.com/9781848729995.

Overview of the Chapters

Our approach in this book is to present how to use and interpret IBM SPSS in the context of proceeding as if the HSB data were the actual data from your research project. However, before starting the SPSS assignments, we have two introductory chapters. The first chapter is an introduction and review of research design and how it would apply to analyzing the HSB data. In addition, this chapter includes a review of measurement levels and descriptive statistics. Chapter 2 discusses rules for coding data, exploratory data analysis (EDA), and assumptions. Much of what is done in these chapters involves preliminary analyses to get ready to answer the research questions that you might investigate in a report.

Chapters 3 and 4 present methods for assessing the reliability and validity of your data. Chapter 3 covers how to compute Cronbach's alpha, test–retest, and interobserver reliability, including intraclass correlation coefficients. Chapter 4 presents one method of assessing validity; including exploratory factor analysis and principal components analysis. Again, these statistical methods are often used to prepare your data so it will be ready to use to help answer your research questions.

Chapter 5 provides a brief overview of research designs (between groups and within subjects). This chapter provides flow charts and tables useful for selecting an appropriate statistic. Also included is an overview of how to interpret and write about the results of two basic inferential statistics. This chapter includes not only testing for statistical significance but also discussions of power and effect size measures, including guidelines for interpretation.

Chapters 6–12 are designed to help you answer a number of research questions. Solving the problems in these chapters should give you a good idea of some of the intermediate statistics that can be computed with IBM SPSS Statistics. Hopefully, seeing how the research questions, design and measurement lead naturally to the choice of statistics will become apparent after using this book. In addition, it is our hope that interpreting what you get back from the computer will become clearer after doing these assignments, studying the outputs, answering the interpretation questions, and doing the extra SPSS problems.

Finally, Chapter 13 describes the problem of having missing or incomplete data when computing statistics, and we describe data imputation methods to help with this problem. We apply this to a problem using multilevel regression, and an extra problem shows how to do a more basic analysis after multiple imputation (a paired t-test).

Our Approach to Research Questions, Measurement, and Selection of Statistics

In Chapters 1 and 5, our approach is somewhat nontraditional because we have found that students have a great deal of difficulty with some aspects of research and statistics but not others. Most can learn formulas and "crunch" the numbers quite easily and accurately with a calculator or with a computer. However, many have trouble knowing what statistics to use and how to interpret the results. They do not seem to have a "big picture" or see how research questions, design and measurement influence data analysis. Part of the problem is inconsistent terminology. For these reasons, we have tried to present a semantically consistent and coherent picture of how research design leads to three basic kinds of research questions (difference, associational, and descriptive) which, in turn, lead to three kinds or groups of statistics with the same names. We realize that these and other attempts to develop and utilize a consistent framework are both nontraditional and somewhat of an oversimplification. However, we think the framework and consistency pay off in terms of student understanding and the ability to actually use statistics to answer the research questions. Instructors who are not persuaded that this framework is useful can skip Chapter 1 and the first part of Chapter 5 and still have a book that helps their students use and interpret SPSS.

Major Changes and Additions to This Edition

In addition to updating the text, SPSS windows, and outputs to IBM SPSS 22, we have also attempted to correct any typos in the 4th edition and clarify some passages. In Chapter 2, we added research questions and directions for conducting each statistic to assist the reader in conducting exploratory data analysis. In the chapter on reliability (Chapter 3) we included an "example of how to write about" each problem, and we added a problem on using interclass correlation (ICC). Chapter 5 includes an expanded discussion on effect sizes to include information on confidence intervals of effect sizes. Chapter 6 includes new information on part and partial correlations and how they are interpreted, and we have added forward and backward elimination, which are other useful multiple regression methods. Chapter 7 is new; it includes a discussion of how to know if a variable is a mediator or moderator, and it also includes canonical correlation which in earlier editions was part of the chapter on MANOVA. Chapter 12 was revised extensively. Chapter 13 on how to deal with missing data is new; it provides an extensive description of how to do and interpret multiple imputation. We also expanded the appendix about Getting Started with IBM SPSS (Appendix A) to include several useful procedures that were not discussed in the body of the text. Finally, because many of the SPSS outputs are long and complex, we have split many of our Interpretation sections so that the interpretations come closer to the output table(s) that they discuss.

We have checked the format of the examples of how to write about the outputs to be consistent with APA format in the 6th edition (2010) of the *Publication Manual of the American Psychological Association*. Although this edition of our SPSS book was written using version 22, this SPSS program is sufficiently similar to prior versions of this software that we feel you should be able to use this book with earlier and later versions as well.

Bullets, Arrows, Bold, and Italics

To help you do the problems with SPSS, we have developed some conventions. We use bullets to indicate actions in SPSS windows that you will take. For example:

- Highlight *gender* and *math achievement*.
- Click on the arrow to move the variables into the right-hand box.
- Click on **Options** to get Fig. 2.16.
- Check **Mean, Std Deviation**, **Minimum**, and **Maximum**.
- Click on **Continue**.

Note that the words in italics are variable names and words in bold are words that you will see in the SPSS windows and utilize to produce the desired output. In the text such bolded words are spelled and capitalized as you see them in SPSS. Bold also is used to identify key terms when they are introduced, defined, or important to understanding.

The words you will see in the pull-down menus are given in bold with arrows between them. For example:

- Select **Analyze → Descriptive Statistics → Frequencies**.

(This means pull down the Analyze menu, then slide your cursor down to Descriptive Statistics, over to Frequencies, and click.)

Occasionally, we have used underlines to <u>emphasize critical points or commands</u> that have sometimes been confusing for students.

We have tried hard to make this book accurate and clear so that it could be used by students and professionals to learn to compute and interpret statistics without the benefit of a class. However, we find

that there are always some errors and places that are not totally clear. Thus, we would like for you to help us identify any grammatical or statistical errors and to point out places that need to be clarified. Please send suggestions to *nancy.leech@ucdenver.edu*.

Acknowledgments

This SPSS book is consistent with and could be used as a supplement for Gliner, Morgan, and Leech (2009), *Research Methods in Applied Settings: An Integrated Approach to Design and Analysis,* or Morgan, Gliner, and Harmon (2006), *Understanding and Evaluating Research in Applied and Clinical Settings*. Information about both books can be found at www.routledge.com/9781848729995. In fact, some sections of Chapters 1 and 5 have been only slightly modified from these texts. For this we thank Jeff Gliner, coauthor of those books. Orlando Griego was an author on our first SPSS book; this revision still shows the imprint of his student-friendly writing style.

We would like to acknowledge the assistance of the many students in our education and human development classes who have used earlier versions of this book and provided helpful suggestions for improvement. We could not have completed the task or made it look so good without our technology consultant, Don Quick, and our word processor, Madison Myers. Several other previous student workers were key to earlier versions. Jikyeong Kang, Bill Sears, LaVon Blaesi, Mei-Huei Tsay, and Sheridan Green assisted with classes and the development of materials for the DOS and earlier Windows versions of the assignments. Laura Jensen, Lisa Vogel, Andrea Weinberg, James Lyall, Joan Anderson, Pam Cress, Joan Clay, Yasmine Andrews, Madison Myers, and Sophie Nelson helped with writing or editing parts of this or earlier editions. Jeff Gliner, Jerry Vaske, Jim zumBrunnen, Laura Goodwin, David MacPhee, Gene Gloeckner, James O. Benedict, Barry Cohen, John Ruscio, Tim Urdan, and Steve Knotek provided reviews of earlier editions and/or suggestions for improving the text. Carolyn Springer, Jay Parkes, Joshua Watson, and John Rugutt provided helpful reviews for the 3rd edition. Marianne Fallon, Craig A. Johnson, and several anonymous reviewers provided critiques of the fourth edition that we have used to try to improve this edition. Don Quick and John Cumming updated helpful appendixes for this edition. Bob Fetsch and Ray Yang provided helpful feedback on the readability and user friendliness of the text. Finally, the patience of our spouses (Grant, Terry, and Hildy) and families enabled us to complete the task, without too much family strain.

Nancy L. Leech
University of Colorado Denver

Karen Caplovitz Barrett
Colorado State University

George A. Morgan
Colorado State University

About the Authors

Nancy L. Leech is Professor of Education and Human Development at the University of Colorado Denver. She received her Ph.D. in Education with an emphasis on research and statistics from Colorado State University in 2002. She currently teaches master's and PhD level courses in research, statistics, and measurement. Her area of research is promoting new developments and better understanding in applied qualitative, quantitative, and mixed methodologies.

Karen C. Barrett is Professor of Human Development & Family Studies at Colorado State University. She received her Ph.D. in Developmental Psychology from University of Denver. Her research is focused on social and emotional development, especially the early development of emotion regulation and self-regulation; early development of "social emotions" such as shame, guilt, and pride; cultural and contextual influences on these; and early intervention to promote emotional and regulatory competence. She has taught research methods and statistics at the graduate and undergraduate level.

George A. Morgan is an Emeritus Professor of Education and Human Development at Colorado State University. He received his Ph.D. in child development and psychology from Cornell University. In addition to writing the textbooks on SPSS and research methods mentioned in the preface, he has advised many Ph.D. students in education and related fields. Over the past 35 years, he has conducted a program of research on children's motivation to master challenging tasks and solve problems. He is currently working with colleagues on developing a computer-based motivational assessment for school readiness and a revision of the widely used Dimensions of Mastery Questionnaire (DMQ) so that it will be even more useful with children in at least English, Chinese, Spanish, and Hungarian speaking countries.

Table of Contents

Preface viii

About the Authors xiii

1. Introduction 1
   - Research Problems and Variables
   - Research Hypotheses and Research Questions
   - A Sample Research Problem: The Modified High School and Beyond (HSB) Study
   - Research Questions for the Modified HSB Study
   - Frequency Distributions
   - Levels of Measurement
   - Descriptive Statistics
   - Conclusions About Measurement and the Use of Statistics
   - The Normal Curve
   - Interpretation Questions

2. Data Coding and Exploratory Analysis (EDA) 25
   - Rules for Data Coding
   - Exploratory Data Analysis (EDA)
   - Statistical Assumptions
   - Checking for Errors and Assumptions With Ordinal and Scale Variables
   - Problem 2.1: Descriptive Statistics for Ordinal and Scale Variables
   - Problem 2.2: Boxplots for Ordinal and Scale Variables
   - Problem 2.3: Boxplots Split by a Dichotomous Variable
   - Problem 2.4: Using Tables for EDA with Dichotomous Variables
   - Problem 2.5: Using Frequency Tables
   - Problem 2.6: Bar Charts
   - Problem 2.7: Histograms and Frequency Polygons
   - Problem 2.8: Matrix Scatterplots
   - Problem 2.9: Transforming Variables
   - Interpretation Questions
   - Extra SPSS Problems

3. Several Measures of Reliability 53
   - Problem 3.1: Cronbach's Alpha for the Motivation Scale
   - Problems 3.2 and 3.3: Cronbach's Alpha for the Competence and Pleasure Scales
   - Problem 3.4: Test–Retest Reliability Using Correlation
   - Problem 3.5: Intraclass Correlation Coefficients (ICC)
   - Problem 3.6: Cohen's Kappa With Nominal Data
   - Interpretation Questions
   - Extra SPSS Problems

4. Exploratory Factor Analysis and Principal Components Analysis 68
   - Problem 4.1: Factor Analysis on Math Attitude Variables
   - Problem 4.2: Principal Components Analysis on Achievement Variables
   - Interpretation Questions
   - Extra SPSS Problems

5. Selecting and Interpreting Inferential Statistics 84
Selection of Inferential Statistics
The General Linear Model
Interpreting the Results of a Statistical Test
A Review of How to Select and Interpret Basic Inferential Statistics
Review of Writing About Your Outputs
Interpretation Questions

6. Multiple Regression 109
Problem 6.1: Using the Simultaneous Method to Compute Multiple Regression
Problem 6.2: Simultaneous Regression Correcting Multicollinearity
Problem 6.3: Hierarchical Multiple Linear Regression
Problem 6.4: Forward Multiple Linear Regression
Problem 6.5: Backward Elimination Multiple Linear Regression
Interpretation Questions
Extra SPSS Problems

7. Mediation, Moderation, and Canonical Correlation 144
Mediation
Moderation
Canonical Correlation
Problem 7.1: Computing Statistical Mediation
Problem 7.2: Moderation
Problem 7.3: Moderation: Graphing the Slopes
Problem 7.4: Canonical Correlation
Interpretation Questions
Extra SPSS Problems

8. Logistic Regression and Discriminant Analysis 167
Problem 8.1: Logistic Regression
Problem 8.2: Hierarchical Logistic Regression
Problem 8.3: Discriminant Analysis
Interpretation Questions
Extra SPSS Problems

9. Factorial ANOVA and ANCOVA 188
Problem 9.1: Factorial (Two-Way) ANOVA
Problem 9.2: Post Hoc Analyses of a Significant Interaction
Problem 9.3: Analyses of Covariance (ANCOVA)
Interpretation Questions
Extra SPSS Problems

10. Repeated-Measures and Mixed ANOVAs 213
The Product Data Set
Problem 10.1: Repeated-Measures ANOVA
Problem 10.2: The Friedman Nonparametric Test for Several Related Samples
Problem 10.3: Mixed ANOVA
Interpretation Questions
Extra SPSS Problems

11. Multivariate Analysis of Variance (MANOVA) 233
Problem 11.1: GLM Single-Factor Multivariate Analysis of Variance
Problem 11.2: GLM Two-Factor Multivariate Analysis of Variance
Problem 11.3: Mixed MANOVA
Interpretation Questions
Extra SPSS Problems

12. Multilevel Linear Modeling/Hierarchical Linear Modeling 256
Problem 12.1: Unconditional Level 1 Repeated-Measures Model
Problem 12.2: Repeated Measures with Level 2 Predictor
Problem 12.3: Unconditional Individuals-Nested-in-Schools Model
Problem 12.4: Conditional Individuals-Nested-in-Schools Model with Level 1 Covariate
Interpretation Questions
Extra SPSS Problems

13. Missing Data and Multiple Imputation 292
Randomness of Missing Data
Problem 13.1: Patterns of the Missing Data
Problem 13.2: Restructuring and Imputing the Data
Problem 13.3: Mixed Models Analysis after Imputed Data
Interpretation Questions
Extra SPSS Problems

Appendices

A. Getting Started With SPSS and Other Useful Procedures 312
by Don Quick and Madison Myers
B. Review of Basic Statistics 328
by John M. Cumming and Andrea Weinberg
C. Writing Research Problems and Questions 342
D. Answers to Odd Interpretation Questions 346

For Further Reading 355

Index 357

CHAPTER 1

Introduction

This chapter will review important information about measurement and descriptive statistics and provide an overview of the expanded high school and beyond (HSB) data set, which will be used in this chapter and throughout the book to demonstrate the use and interpretation of the several statistics that are presented. First, we provide a brief review of some key terms, as we will use them in this book.

Research Problems and Variables

Research Problems

The research process begins with an issue or problem of interest to the researcher. This **research problem** is a statement that asks about the relationships between two or more variables. Almost all research studies have *more* than two variables.

Variables

Variables are key elements in research. A **variable** is defined as a characteristic of the participants or situation for a given study that has several values in that study. A variable must be able to vary or have different values or levels in the study.[1] For example, *gender* is a variable because it has two levels, female or male. *Age* is a variable that has a large number of values. *Type of treatment/intervention* (or *type of curriculum*) is a variable if there is more than one treatment or a treatment and a control group. *Number of days to learn something or to recover from an ailment* are common measures of the effect of a treatment and, thus, are also variables. Similarly, *amount of mathematics knowledge* is a variable because it can vary from none to a lot. If a concept has only one value in a particular study, it is not a variable; it is a constant. Thus, ethnic group is not a variable if all participants are European-American. Gender is not a variable if all participants in a study are female.

In quantitative research, variables are defined operationally and are commonly divided into **independent variables** (active or attribute), **dependent variables**, and **extraneous variables**. Each of these topics will be dealt with briefly in the following sections.

Operational definitions of variables. An operational definition describes or defines a variable in terms of the operations or techniques used to make it happen or measure it. When quantitative researchers describe the variables in their study, they specify what they mean by demonstrating how they measured the variable. Demographic variables like age, gender, or ethnic group are usually measured simply by asking the participant to choose the appropriate category from a list, or using school or government records. Types of treatment (or curriculum) are usually operationally defined much more extensively by describing what was done during the treatment or new curriculum. Likewise, abstract concepts like mathematics knowledge, self-concept, or mathematics anxiety need to be defined operationally by spelling out in some detail how they were measured in a particular study. To do this, the investigator may provide sample questions, append the actual instrument, or provide a reference where more information can be found.

---

[1] To help you, we have identified the SPSS variable names and labels using italics (e.g., *gender* and *ethnic*). Sometimes italics are also used to emphasize a word. We have put in bold the terms used in the SPSS windows and outputs (e.g., **SPSS Data Editor**) and other key terms when they are introduced, defined, or are important to understanding. Underlines are used to emphasize critical points. Bullets precede instructions about SPSS actions (e.g., click, highlight).

Independent Variables

We refer to two types of independent variables: **active** and **attribute**. It is important to distinguish between these types when we discuss the results of a study.

Active or manipulated independent variables. An active independent variable is a variable, such as a workshop, new curriculum, or intervention, at least one level of which is given to a group of participants, ideally within a specified period of time during the study. For example, a researcher might investigate a new kind of therapy compared to the traditional treatment. A second example might be to study the effect of a new teaching method, such as cooperative learning, on student performance. In these two examples, the variable of interest was something that was *given to* the participants. Thus, active independent variables are *given* to the participants during the study but are not necessarily given or manipulated by the experimenter. The independent variable may be given by a clinic, school, or someone other than the investigator, but from the participants' point of view the situation was manipulated. Ideally, the treatment is given *after* the study was planned so that there can be a pretest and so participants can be randomly assigned to receive it or not. If there is no pretest, or no random assignment, internal validity of the study will be in jeopardy. Other researchers have similar but, perhaps, slightly different definitions of active independent variables. **Randomized experimental** and **quasi-experimental** studies have an active independent variable. An active independent variable is a necessary but not sufficient condition to make cause-and-effect conclusions; the clearest causal conclusions can be drawn from well-controlled randomized experiments when participants are *assigned randomly* to *one of two or more* groups, at least one of which will receive the intervention.

Attribute or measured independent variables. A variable that cannot be or was not manipulated, yet is a major focus of the study presumed by the researchers to be a predictor of or influence on one or more dependent variables, can be called an attribute independent variable. For attribute independent variables, the values of the independent variables are preexisting attributes of the persons or their ongoing environment that are not systematically changed during the study. For example, gender, age, ethnic group, IQ, and self-esteem can be used as attribute independent variables. Studies with only attribute independent variables are called **nonexperimental** studies.

In keeping with SPSS, but unlike authors of some research methods books, we do not restrict the term independent variable to those variables that are manipulated or active. We define an independent variable more broadly to include any predictors, antecedents, or *presumed* causes or influences under investigation in the study. Attributes of the participants, as well as active independent variables, fit within this definition. For the social sciences and education, attribute independent variables are especially important. Type of disability or level of disability may be the major focus of a study. Disability certainly qualifies as a variable since it can take on different values even though they are not *given during* the study. For example, cerebral palsy is different from Down syndrome, which is different from spina bifida, yet all are disabilities. Also, there are different levels of the same disability. Before we begin our study, participants already have defining characteristics or attributes that place them into one of two or more categories. Examples of these could be different disabilities or different levels of self-esteem. Thus, often we are interested in studying variables that are not manipulated during the study, even by other persons, schools, or clinics.

Other labels for the independent variable. SPSS uses a variety of terms in addition to independent variable, for example, **factor** (Chapters 9, 10, 11, and 12,), and **covariate** (Chapters 9 and 12). In other cases (Chapters 3 and 4), neither SPSS nor statisticians make a distinction between the independent and dependent variable; they just label them **variables**. For example, there is no independent variable for a correlation or chi-square. However, even for chi-square, we think it is sometimes educationally useful to

think of one variable as the independent variable and the other as the outcome (dependent variable), as is the case in analysis of variance (ANOVA) or multivariate analysis of variance (MANOVA).

Values of the independent variable. SPSS uses the term **values** to describe the several options or values of a variable. These values are *not* necessarily ordered, and several other terms — **categories**, **levels**, **groups**, or **samples** — are sometimes used interchangeably with the term values, especially in statistics books. Suppose that an investigator is performing a study to investigate the effect of a treatment. One group of participants is assigned to the treatment group. A second group does not receive the treatment. The study could be conceptualized as having one independent variable (*treatment type)*, with two values or levels (*treatment* and *no treatment*). The independent variable in this example would be classified as an active independent variable. Now, suppose instead that the investigator was primarily interested in comparing two different treatments but decided to include a third no-treatment group as a control group in the study. The study still would be conceptualized as having one active independent variable (*treatment type*) but with three values or levels (the two treatment conditions and the control condition). As an additional example, consider *gender*, which is an attribute independent variable with two values, *male* and *female*.

Note that in SPSS each variable is given a **variable label** that assists the researcher in understanding the variable. Moreover, each value or level is assigned a number used by SPSS to compute statistics, and these numbers can be assigned **value labels** that indicate what the numbers stand for (e.g., 1 = *male* and 2 = *female*). It is especially important to use value labels when the variable is **nominal** (i.e., when the values of the variable are just names and, thus, are not ordered), so that you can know the category to which each of the numbers refers.

Dependent Variables

The **dependent variable** is assumed to measure or assess the effect of the independent variable. It is thought of as the presumed outcome or criterion. Dependent variables are often test scores, ratings on questionnaires, readings from instruments (e.g., electrocardiogram, galvanic skin response), or measures of physical performance. Dependent variables, like independent variables, must have at least two values; most dependent variables have many values, varying from low to high.

SPSS also uses a number of other terms in addition to dependent variable. **Dependent list** is used in cases where you can do the same statistic several times, for a list of dependent variables (e.g., in one-way ANOVA). **Grouping variable** is used in Chapter 8 for discriminant analysis.

Extraneous Variables

These are variables (also called nuisance variables or, in some designs, covariates) that are not of primary interest in a particular study but could influence the dependent variable. Environmental factors (e.g., temperature or distractions), time of day, and characteristics of the experimenter, teacher, or therapist are some possible extraneous variables that need to be controlled. SPSS does not use the term extraneous variable. However, sometimes such variables are controlled using statistics that are available in SPSS.

Research Hypotheses and Research Questions

Research hypotheses are predictive statements about the relationship between variables. **Research questions** are similar to hypotheses, except that they do not entail specific predictions and are phrased in question format. For example, one might have the following research question: "Is there a difference in students' scores on a standardized test if they took two tests in one day versus taking only one test on each of two days?" A research hypothesis regarding the same issue might be: "Students who take only one test per day will score *better* on standardized tests than will students who take two tests in one day."

We divide research questions into three broad types: **difference**, **associational**, and **descriptive** as shown in the middle of Fig 1.1. The figure also shows the general and specific purposes and the general types of statistics for each of these three types of research question.

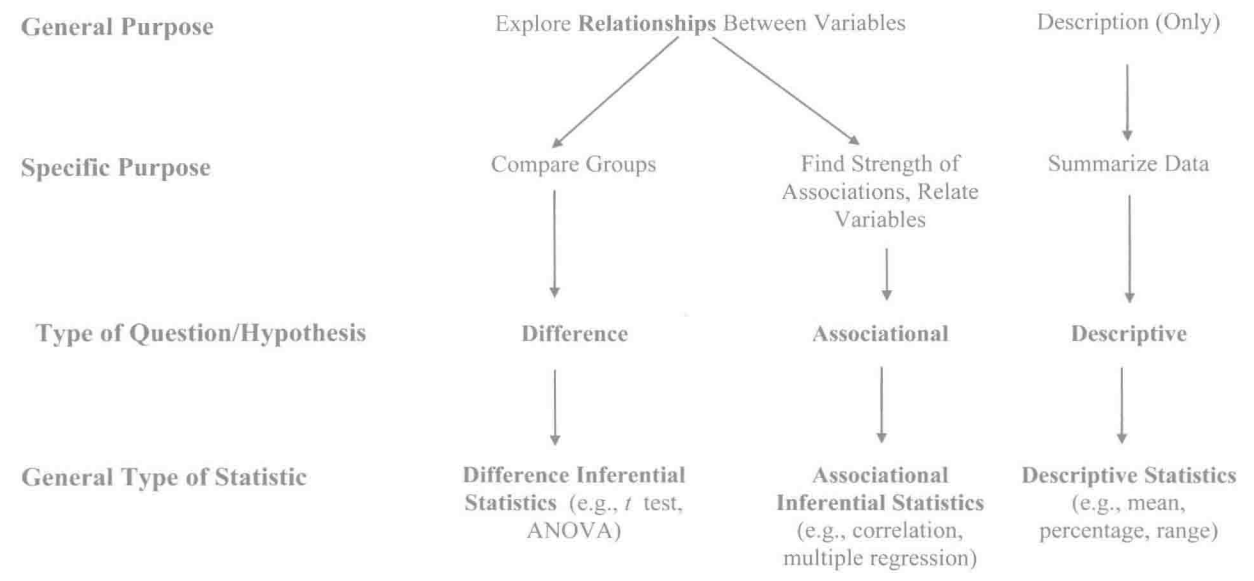


Fig. 1.1. Schematic diagram showing how the purpose and type of research question correspond to the general type of statistic used in a study.

Difference research questions. For these questions, we compare scores (on the dependent variable) of two or a few groups (e.g., males and females). Each group is composed of individuals who have the same value or level of the independent variable and who potentially (and hopefully!) have a different level of the dependent variable than do individuals in the other group(s). This type of question attempts to discover whether or not groups are the same on the dependent variable in the larger population.

Associational research questions are those in which two or more variables are associated or related. This approach usually involves an attempt to see how two or more variables covary or how one or more variables enable one to predict another variable. When two variables covary positively, individuals who have high scores on one variable also tend to have high scores on the other variable; other individuals tend to have relatively low scores on both variables or middling scores on both variables.

Descriptive research questions are not answered with inferential statistics. They merely describe or summarize data without trying to generalize to a larger population of individuals.

Figure 1.1 shows that both difference and associational questions or hypotheses are similar in that they explore the relationships between variables.[2] Note that difference and associational questions differ in specific purpose and the kinds of statistics they use to answer the question.

[2]This similarity is in agreement with the statement by statisticians that all common parametric inferential statistics are relational. We use the term associational for the second type of research question rather than relational or correlational to distinguish it from the *general purpose* of both difference and associational questions/hypotheses,

Remember that research questions are similar to hypotheses, but they are stated in question format. We think it is advisable to use the question format when one does not have a clear directional prediction and also for the descriptive approach. As implied by Fig. 1.1, it is acceptable to phrase any research question that involves two or more variables as whether there is a relationship between the variables (e.g., "Is there a relationship between *gender* and *math achievement*?" or "Is there a relationship between *anxiety* and *GPA*?"). However, we think that phrasing the question as a difference or association is desirable because it helps one choose an appropriate statistic and interpret the result. For some examples of how to write research questions, see the section in this chapter, "Research Questions for the Modified HSB study."

Complex Research Questions

Most research questions posed in this book involve more than two variables at a time. We call such questions and the appropriate statistics **complex**. Some of these statistics are called **multivariate** in other texts, but there is not a consistent definition of multivariate in the literature. We provide examples of how to write complex research questions in the chapter pertaining to each complex statistic.

In a factorial ANOVA, there are two (or a few) independent variables and one dependent variable. We will see, in Chapter 9, that when you do one factorial ANOVA there are actually three (or more) research questions. This *set* of three questions can be considered a complex difference question because the study has two independent variables. Likewise, complex associational questions are used in studies with more than one independent variable considered together.

Table 1.1 expands our overview of research questions to include both basic and complex questions of each of the three types: **descriptive**, **difference**, and **associational**. The table also includes references to other chapters in this book and examples of the types of statistics that we include under each of the six types of questions.

A Sample Research Problem: The Modified High School and Beyond (HSB) Study

The SPSS file name of the data set used with this book is *hsbdataNew.sav*; it stands for high school and beyond data. It is based on an actual national sample of data from more than 28,000 high school students collected some years ago. Our data set includes a sample of 75 students drawn randomly from the larger population. The data that we use for this sample include school outcomes such as grades and the number of mathematics courses of different types that the students took in high school. Also, there are several kinds of standardized test data and demographic data such as gender and mother's and father's education. To provide an example of questionnaire data, we have included 14 questions about mathematics attitudes. These data were developed for this book and, thus, are not really the math[3] attitudes of the 75 students in this sample; however, they are based on real data gathered by one of the authors to study motivation. Also, we made up data for *religion*, *ethnic group*, *SAT-math and retests for the visualization and mosaic pattern test scores*, which are somewhat realistic overall. These inclusions enable us to do some additional statistical analyses.

---

which is to study relationships. Also we wanted to distinguish between correlation, as a specific statistical technique, and the broader type of associational question and that group of statistics.

[3] We have decided to use the short version of mathematics (i.e., math) throughout the book to save space and because it is used in common language.