

# PREFACE TO THE SECOND EDITION

Instructors of introductory social statistics face an unenviable quandary. They want to give their students the skills they need to succeed in the real world of social research, but they realize that if they push their students too far, they risk losing them altogether. Some instructors understandably surrender to this latter concern, opting to teach their students the more basic statistical procedures. Unfortunately, such procedures are seldom used in the real world. If instructors do decide to introduce their students to more contemporary techniques, they encounter course materials that were not developed with the introductory student in mind. This was the position in which I found myself a number of years ago, and I ultimately reached a decision to do something to remedy this dilemma. *Social Statistics: Managing Data, Conducting Analyses, Presenting Results* is my solution. It is the first statistics text that ventures to cover both classic and contemporary techniques in an approachable way that will engage the typical introductory student and make her eager rather than anxious to study this wide array of techniques.

If you compare the table of contents with those of other introductory statistics texts, you will see some similarities and some major differences. The first half of the book contains, on the surface, many of the similarities. The early chapters include many of the topics that one might find in other books: tables and graphs, measures of central tendency and variation, probability distributions, chi-square tests, confidence intervals,  $t$ -tests, ANOVA, and bivariate regression. I cover these topics innovatively and efficiently in order to prepare the students for the rest of the book. In the second half of the book, students gain significant exposure to a variety of multiple regression techniques that they will find in the real worlds of social research: reference groups, nested modeling, standardized effects, interaction effects, logistic regression, path analysis,

and nonlinearity. In stark contrast to many books with such coverage, I handle these topics at a level that introductory statistics students will find approachable and engaging. For most beginning statistics students in the social sciences, this is the one and only statistics course they will take. If they use *Social Statistics: Managing Data, Conducting Analyses, Presenting Results*, they will leave the course with a strong and varied set of skills that will serve them well as they try to navigate the social science literature or acquire a job.

Although some of these regression techniques may appear in other introductory books, they often do so only as afterthoughts, covered in the most cursory of ways in the final chapter of the book. Unfortunately, this is exactly the point at which students need more explanation, not less. I cover these techniques with a significant—though not overwhelming—level of depth. I explain each technique using unique graphics, visual analogies, and real-world examples. The clear emphasis is on interpretation: given a regression model, or having created one of his or her own, what steps should a student take to make sense of what the model is telling him or her? Combined with their instructor’s assistance, this book gets students to the point where they can translate a wide variety of statistical results, whether they are reading social science literature or making a presentation at their job. It guides students through the entire statistical research process: from working with data to get it ready for analysis, conducting the analyses (both by hand and with SPSS), and moving from raw SPSS output to professional presentation. Each chapter ends with graphical, step-by-step SPSS demonstrations, followed by short “from output to presentation” sections that teach students how to present results in clear and compelling ways.

Some instructors may be rightfully dubious about the possibility of introducing their students to some of these techniques. Yet I maintain that, with the help of this book, this is completely possible. I use several strategies to accomplish this. Each chapter includes several simple examples that convey the key aspects of the technique at hand. Most of these examples use data from the General Social Survey (GSS), primarily from 2012, but occasionally from other years. Many chapters contain “interchapter connections” that show how techniques are related to one another, and illustrate that some of the more advanced techniques can be considered extensions of more basic techniques. These connections also help the student through the challenging task of choosing the appropriate technique given a research situation. Each chapter ends with an example or two from the social science literature, showing how social researchers used the chapter’s technique in an interesting way. I guide the students through these examples, showing them how to decipher tables that, at first, seem daunting. I make further use of the literature in a unique appendix that features descriptions of 86 social science journal articles from a variety of academic fields. I have vetted these articles, including only those that have statistical results that won’t overwhelm introductory students. For each article, I offer a brief description, talk about the techniques the

authors use to make their points (and what pitfalls to watch out for when reading their results), and end with a few questions for the student about the article's use of statistics.

The book emphasizes visual learning in order to make contemporary techniques more approachable. A series of innovative Excel-based live demonstrations and PowerPoint-based animations make many of the techniques come to life. For example, the Excel-based regression demonstration can, in a brief moment, show students the effect of an outlier on a regression line. A PowerPoint animation walks students through one of the book's path models in order to show the power of indirect effects. Instructors are welcome to integrate these demonstrations and animations into their lectures. There are also innovative videos that demonstrate SPSS procedures. These and other helpful instructor support materials (such as detailed answers to all of the end-of-chapter exercises, and a variety of exam questions) can be found on a companion website at URL: [www.routledge.com/cw/Linneman](http://www.routledge.com/cw/Linneman)

For the end-of-chapter exercises, I use more real-world data from five fascinating datasets: the 2012 American National Election Studies, the 2005 World Values Survey, and three datasets from the Pew Internet & American Life Project (on consumption, health, and cyberbullying). Thus, the end-of-chapter exercises are designed for students of varied interests: sociology, political science, marketing, public health, education, criminal justice, and global studies. Here are some examples that illustrate the range of exercise topics:

- With the exercises from the 2012 American National Election Studies, students explore such questions as “Do voters trust the government more than nonvoters?” and “What propels people to be involved in their communities?”
- With exercises from the 2013 PewShop dataset, students explore such questions as “Do smartphones allow people of all ages to engage in technology-enabled shopping experiences?” and “Do income disparities account for technology consumption differences among racial groups?”
- With exercises from the 2012 PewHealth dataset, students explore such questions as “Do men and women use the Internet to seek health information at the same rate?” and “Do people of all ages use Internet information when discussing their healthcare options with their doctors?”
- With exercises from the 2011 PewKids dataset, students explore such questions as “Do children who have been cyberbullied engage in more empathetic behavior toward the cyberbullied than those who have not?” and “What role does parental age play in the level at which parents monitor their children's technology use?”
- With exercises from the 2005 World Values Survey dataset, students explore such questions as “Is the relationship between health and happiness, on a country-by-country level, linear or nonlinear?” and “What role does societal trust play in citizens' desire for authoritarian leadership?”

At every turn, the book gives students opportunities to understand how researchers use social statistics in the real world, and to conduct and present their own analyses, just as they will be expected to do in their own research in academics or employment.

## **CHAPTERS OF THE BOOK AND WHAT IS NEW TO THIS EDITION**

---

- Chapter 1 is all about forms of data: what do they look like and how do you work with them? Since many students may have never even seen a dataset, I describe how you construct a basic dataset and how you can get it into the shape you want through recoding, computing, and indexing. For example, there is a step-by-step GSS example about constructing an index of workplace hostility. I talk about several of the most innovative and extensive data collection efforts in the social sciences. I discuss how we live in an age of endless data, which presents us with myriad research opportunities, and offer literature examples of researchers using Internet-based data (Wikipedia and the Internet Movie Database) to conduct interesting research projects.
- Chapter 2 covers table construction with one, two, or three variables. I also cover basic graphing, with an emphasis on how to create a graph that accurately represents the data. Examples in this chapter include the effect of childrearing goals on parents' propensity to engage in spanking, and the effects of gender and age on chivalrous behavior. The chapter ends with a fascinating article about racial classification that appeared in a recent issue of a top social science journal, yet featured a greatly exaggerated graph.
- Chapter 3 covers, using a wide variety of unique graphic-based explanations, the basic descriptive statistics: mean, median, mode, variance, and standard deviation. Given that qualitative diversity is of paramount importance, I also provide extensive coverage of the index of qualitative variation, as well as some coverage of the index of dissimilarity and the Gini Coefficient. Examples in this chapter include variation in Internet use by race, and changing attitudes over time toward government spending on health care, the military, and the environment. A literature example highlights the extensive variation in medical costs for a single surgical procedure.
- Chapter 4 is the first of four chapters in the book that cover inferential techniques. In each of these four chapters, I discuss in depth how each technique is based on a probability distribution, showing how such distributions are actually created and what they really mean. In this chapter, I cover inference with crosstabs—the chi-square test—using a creative discussion of statistical significance. I emphasize, through a unique graphic, the effect that sample size can have on chi-square

results. The chapter's examples include the relationship between age and cynicism and the relationships among age, gender, and gun ownership. Both chi-square literature examples involve the body: one covers gender differences in flatulence habits, whereas the other compares how the French and American media treat the obesity epidemic.

- Chapter 5 is the second inference chapter. By hand, I build a sampling distribution and show graphically what the standard error of the mean really is. With regard to applications, this chapter covers testing a population claim and building confidence intervals. Examples in this chapter include attitudes toward police use of violence and the relationship between job stress and job satisfaction. The literature example regards how a researcher used confidence intervals to study how blacks are portrayed in a random sample of contemporary films.
- Chapter 6 is the third inference chapter, and in it, I cover  $t$ -tests and ANOVA. I construct by hand a sampling distribution of sample mean differences, and I go into significant depth regarding how the tests' formulas actually work. I introduce "interchapter connections," which show students how various techniques are similar or different, and help them understand how to choose among techniques. The examples for the chapter involve the relationships among political party, age, and attitudes toward suicide, and the connection between attitudes toward gender equality in the household and actual behavior within the household. The  $t$ -test literature example is on gender overcompensation. The ANOVA literature example studies activism through the life course.
- Chapter 7 covers simple bivariate correlation and regression. The graphical examples fully explain the important concept of explained variation. By examining movie grosses over time, I show how regression can be used in forecasting. Other examples include the effects of income on relaxation time, and intergenerational effects on family size. The literature examples cover attitudes toward relinquishing civil liberties in the age of terror, and the correlations among gender, body size, and physical attractiveness.
- Chapter 8 is the final chapter on inference. By building one last sampling distribution, I graphically illustrate what the standard error of the slope represents and how we use it to gauge a regression slope's statistical significance. I emphasize the relationship between sample size and statistical significance, and teach students to think critically about the distinction between statistical and substantive significance. Examples in this chapter examine how level of sexual activity is affected by educational achievement, income, and hours worked. Both literature examples involve grades: looking first at the effect of studying at the college level, second at the effect of family size at the grade-school level.
- Chapter 9 involves the use of various types of variables as independent variables in a regression equation. After covering how to interpret slopes for dichotomous variables, I show in a step-by-step fashion how to use multiple dichotomies to

create a set of reference-group variables. I also include an interchapter connection linking  $t$ -tests with dichotomous slopes. The examples investigate demographic effects on STEM (science, technology, engineering, and mathematics) achievement, partnership-status effects on happiness, and the relationship between political party and political knowledge. The literature examples show how researchers used dichotomies and reference groups to study gender differences in housework, and temporal changes in attitudes toward gay rights.

- Chapter 10 covers, with the great care that the topic warrants, the very important concept of controlling. I start with some analogies, illustrating how the concept of controlling is actually imbued in our everyday lives. I walk students through the typical tabular construction of a series of nested regression models. I offer an interchapter connection, using the same data to create both an elaborated crosstab and a nested regression model. I show how to judge improvement from model to model, and why it is important to keep sample size constant from model to model. Examples in this chapter involve explaining racial differences in attitudes toward state assistance and gender and in religion's effects on attitudes toward same-sex parenting. The literature examples examine the grade gap between whites and blacks and the media effects on attitudes toward crime.
- Chapter 11 covers the meaning behind standardized coefficients, or betas. Rather than just handing the students the simple formula for calculating betas, I take them through an in-depth explanation so that they can develop a full understanding of what the betas really are and why they are important. I include an interchapter connection that links betas to  $z$ -scores. Examples involve religiosity and attitudes toward abortion and the male/female differences in what determines life satisfaction. The literature examples cover the topics of school discipline and of country music's effect on suicide rates.
- Chapter 12 covers one of the most prominent techniques in current social science literature: interaction effects. I first make an interchapter connection that illustrates how interaction has similarities to elaborated crosstabs. Then, I show students how to work through examples to develop a full understanding of the interaction. Examples in this chapter examine the interaction effect between sex and number of children on relaxation time, the interaction effect between race and education on attitudes toward Muslim civil rights, and the interaction effect between race and sex on religiosity. Literature examples involve the interaction of gender and religious participation on black political activity, and the interaction of gender and work hours on level of family guilt.
- Chapter 13 explains the difference between regular regression and logistic regression. Without becoming bogged down in the math going on behind the scenes, I show students how to run numerous examples with a logistic regression model in order to understand the probabilities they are calculating. Because so many logistic results are presented as odds ratios, I explain how to interpret such results.

Dichotomous dependent variables in the examples include home ownership, support for gun control, interracial friendships, giving to charity, and condom usage. Literature examples are on the topics of presidential disapproval and global warming.

- Chapter 14 deals with path analysis. Although more esoteric techniques have emerged, I find that path analysis remains a very useful way for students to visualize indirect effects. I describe how to construct and interpret a path model, and in an interchapter connection, I link path analysis and nested models. To this end, I bring back the same-sex parenting example from an earlier chapter and revise it into a path model. There are also examples concerning drinking behavior, political party identification, and intergenerational socioeconomic status effects. The literature examples involve student activism, and emotion work in the service industry.
- Chapter 15 covers simple non-linear relationships and basic log transformations. I include a detailed and graphical explanation of how these nonlinear slopes work. For the examples, I use age's non-linear effect on income, education's nonlinear effect on income, income's non-linear effect on political party, and income's non-linear effect on financial satisfaction. The literature examples involve gendered occupations, and congressional effectiveness.
- Chapter 16 ends the book with a brief look forward, telling students what they might want to look out for as they enter the world of social research. I offer examples of two common regression-related problems: outliers and multicollinearity. Then I very briefly introduce several common advanced techniques that they might encounter in the social research literature, techniques used for specific types of variables (ordered logistic, multinomial), types of samples (multilevel modeling), and types of situations (structural equation modeling, hazard modeling).

I began this preface with a longstanding problem: many of our introductory statistics students do not gain exposure to the techniques they need to know. At academic conferences, from individual discussions to packed workshops on how to transform the introductory statistics course, I have witnessed concern about this situation. Many instructors want to make this type of change, but they simply haven't known how to accomplish it. *Social Statistics: Managing Datasets, Conducting Analyses, Presenting Results* provides instructors with a proven way to achieve this change in their courses. The book markedly improved my own course: I was able to help my students achieve a greater level of understanding of these techniques than ever before. From their reduced stress levels over the material to the improved quality of their class presentations, I witnessed positive change in a number of important ways. I also have heard from other instructors who used the book that students have responded very positively to it and that it has improved their courses. If given the right tools, instructors can teach

their students these contemporary techniques. I believe such changes in the introductory social statistics course are not only possible but also necessary in our data-filled world. We must give students the foundation they need to succeed in their courses, their research, their jobs, and their lives. It is my sincere belief that this book will help us accomplish these goals.

**Taylor & Francis**  
Not for distribution