

*SOCIAL STATISTICS: THE BASICS AND BEYOND*

---

Preface

*PART ONE: THE BASICS*

---

- Chapter 1: Data: What They Are and What We Do With Them
- Chapter 2: Telling Visual Stories: Tables and Graphs
- Chapter 3: The Center and What Surrounds It:  
Measures of Central Tendency and Variation
- Chapter 4: Using Sample Crosstabs to Talk about Populations:  
The Chi-Square Test
- Chapter 5: Using a Sample Mean or Proportion to Talk about a Population:  
Confidence Intervals
- Chapter 6: Using Multiple Sample Means to Talk about Populations:  
T-Tests and ANOVA
- Chapter 7: Ratio-Level Relationships: Bivariate Correlation and Regression
- Chapter 8: Using Sample Slopes to Talk about Populations:  
Inference and Regression

*PART TWO: BEYOND THE BASICS*

---

- Chapter 9: Dichotomies as Independent Variables: Reference Grouping
- Chapter 10: The Logic and Power of Controlling: Nested Regression Models
- Chapter 11: Comparing and Contrasting Regression Slopes: Beta Coefficients
- Chapter 12: Catching Up or Falling Behind: Interaction Effects
- Chapter 13: Predicting Probabilities Instead of Values: Logistic Regression
- Chapter 14: Telling Visual Stories with Regression: Path Analysis
- Chapter 15: Questioning the Greatness of Straightness: Nonlinear Relationships

*APPENDICES*

---

- Appendix A: Variables and Indexes from the American National Election Studies and World Values Survey
- Appendix B: Fifty Articles That Use Statistics in Less than Scary Ways
- Appendix C: Translating Output: Writing Up and Presenting Statistical Results
- Appendix D: Statistical Tables
- Appendix E: Answers to Selected End-of-Chapter Exercises

## Preface

This book is borne out of a disjuncture I identified during my years of teaching statistics at the undergraduate level. On one side of this chasm are dozens of introductory statistics textbooks that cover a particular set of statistical techniques. On the other side is the real world, which features a considerably different set of statistical techniques. Many statistics textbooks spend too many of their pages teaching skills that students will never ever see again, and far too few of their pages covering techniques that students are likely to encounter. This futility, combined with students' considerable angst, at least partially explains why introductory statistics courses are dreaded. For years, I attempted to bridge this chasm, as many instructors do, by patchworking together a chapter from this book or that book. But often these chapters were written without beginning statistics students in mind. There had to be a better way. It is my sincere hope that *Social Statistics: The Basics and Beyond* provides this better way.

If you compare the above table of contents with those of other introductory statistics texts, you will see some similarities and some major differences. Part One of the book contains, on the surface, many of the similarities. The chapters in Part One includes many of the basics that one might find in other books: tables and graphs, measures of central tendency and variation, 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. It is in Part Two that the book significantly stands apart from most introductory texts. In Part Two, 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, interaction effects, logistic regression, path analysis, and nonlinearity. I cover these topics at a level that beginning statistics students will find approachable. 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: The Basics and Beyond*, 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.

While 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 her own, what steps should a student take to make sense of it? For many students, a statistics course is essentially a foreign-language course. This book acknowledges this, helps students become proficient in this new language, and turns them into

statistical translators in their own right. Combined with their instructor's assistance, *The Basics and Beyond* 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.

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 *The Basics and Beyond*, this is completely possible. I use several strategies to accomplish this. In each chapter, I start off with very simple examples that convey the key aspects of the technique at hand. The chapter then builds slowly in complexity, going through several more examples. Most of these examples use data from the widely reputed General Social Survey, primarily from 2006 and 2008, but occasionally from other years. Many chapters end with an example from the social science literature, showing how a social science researcher used the chapter's technique in an interesting way. I make further use of the literature in a unique appendix that features fifty interesting social science journal articles from a variety of academic fields. For each article, I offer a brief description, talk about how the authors use statistics to make their points (and what pitfalls to watch out for when reading the statistics), and end with a few questions for the student about the article's use of statistics.

For the end-of-chapter exercises, I use more real-world data. For many of the exercises, I use data from the American National Election

Studies, guiding students through analyses that either mimic GSS examples earlier in the chapter, or analyses that make use of very interesting variables unique to the ANES data. For the remainder of the exercises, I used World Values Survey data to create a small dataset where each of the forty-eight cases is a country. These exercises allow the student to see how statistics are used at a unit of analysis other than the individual. In addition, it gives the book a more global flavor.

*The Basics and Beyond* acknowledges that many of today's students are visual learners. The book includes over 250 exhibits that help students understand the material. These exhibits also model how statistical results should be communicated clearly. To this end, the book contains a unique appendix that describes how to take raw statistical output and convert it into material suitable for inclusion in research reports or presentations. It is critical that students not only understand statistics, but also how to convey them in a professional manner. A graduate who can clearly demystify statistical results is increasingly desired in a real world filled with too much information.

Another way the book emphasizes visual learning is through a series of innovative Excel-based live demonstrations and PowerPoint-based animations that make many of the techniques literally 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. Instructors are

welcome to integrate these demonstrations and animations into their lectures. Also included is a series of guided tours through a variety of actual journal articles. With these exercises, instructors can show students how researchers use statistical techniques to investigate a wide variety of fascinating topics, from personal ads to protest photographs.

With relation to statistical output, SPSS is the computer software I use throughout the book (or, using the SPSS company's recent program name change, PASW). SPSS remains the program of choice in many introductory statistics courses. Also, I viewed SPSS' point-and-click approach as preferable to other programs' reliance on programming syntax, as I want the beginning student to concentrate on the meaning of the statistical output, not the complex machinations of an advanced program. At the end of each chapter, I have included a detailed, graphically based SPSS demonstration of a procedure or set of procedures linked to that chapter.

## CHAPTERS OF THE BOOK

---

### *Part One*

As described above, Part One covers much material that other introductory statistics books cover, and does so in an efficient manner, allowing instructors to move along, relatively early in the course of a semester, to other techniques.

- Chapter 1 is all about 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 dataset and how you can get it into the shape you want through recoding, computing, and indexing. I also describe the three datasets that the book uses. An example in this chapter involves creating an index to measure Americans' attitudes toward protest behavior.
- Chapter 2 covers table construction with one, two, or three variables. I introduce crosstabulation, even with elaboration, early on in the book in order to get students excited about the interesting relationships they can find, even before they get to actual statistical procedures. 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 religion and education on attitudes toward the big bang theory, and the effect of gender and unemployment on happiness.
- Chapter 3 covers three measures of central tendency (mean, median, and mode) and three measures of variation (variance, standard deviation, and the index of qualitative variation), using a wide variety of unique graphics and explanations. Examples early in the chapter involve fertility, while examples later in the chapter regard variation in attitudes toward government spending.

- Chapter 4 is the first of four chapters in the book that cover inferential techniques, in which we speak about a large population based on results from a small sample. 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 includes examples regarding confidence in organized religion, and, interestingly juxtaposed, gender differences in flatulence habits.
- Chapter 5 is the second inference chapter. By hand, I build a sampling distribution and show graphically what the standard error really is. With regard to applications, this chapter covers testing a claim about a population and building confidence intervals. Examples in this chapter regard television viewing habits and attitudes towards the legalization of marijuana.
- Chapter 6 is the third inference chapter, and in it I cover simple t-tests and ANOVA. Whereas some books get bogged down in many different permutations of these tests, I find it both sufficient and efficient to deal with only one form of each. I do go into significant depth regarding how the tests' formulas work. The examples for the chapter test for gender differences in abortion attitudes, and racial differences in hours worked.

- Chapter 7 covers simple, bivariate regression. The graphical examples fully explain the important concept of explained variation. Using movies grosses, I show how regression can be used for forecasting. Using GSS data, I explore the relationships between wealth and relaxation time, and between education and fertility.
- Chapter 8 is the final chapter on inference. By building another sampling distribution, I carefully and graphically illustrate what the standard error of the slope represents and how we use it to gauge a slope's statistical significance. I emphasize the relationship between sample size and statistical significance, and teach students to look beyond the stars. Examples in this chapter involve the relationships between education and sexual activity, studying and grades, and family size and grades.

### *Part Two*

Now we venture beyond the basics. While the techniques in Part One are important and lay a sound foundation, Part Two covers the regression techniques that students are very likely to see in the real worlds of social research and the workplace. I cover all of these techniques at a level that is sufficient enough so that students will develop the ability to translate such results if they encounter them, and use these techniques in their own work.

- 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 then offer a graphical approach for understanding these effects. The examples investigate the various demographic factors involved in the digital divide, and housework among various types of households.
- Chapter 10 covers, with the great care that the topic warrants, the very important concept of controlling. I start with some visual analogies, and then show students how controlling allows us to debunk some common misconceptions. I walk students through the typical tabular construction of a series of nested regression models. Examples in this chapter involve explaining differences in concern over global warming, and explaining the grade gap between whites and blacks.
- Chapter 11 is a relatively brief chapter that 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. A GSS example involves determining, for various educational levels, which job factors most affect job

satisfaction. Another example concerns the effect of computer use on childhood obesity.

- Chapter 12 covers one of the most prominent techniques in current social science literature: interaction effects. While some find the understanding of such effects elusive, I have developed an extremely graphical way to explain how these effects work. I show students how to work through examples to develop a full understanding of the interaction. Of the several GSS examples in this chapter, one concerns how men and women's relaxation time is differentially affected by number of children, and another concerns the interaction of sex and education on abortion attitudes.
- Chapter 13 carefully 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 the logistic regression model in order to understand the probabilities they are calculating. Since so many logistic results are presented as odds ratios, I fully explain how to interpret such results. Dichotomous dependent variables in the examples include interracial friendships, support for evolution, and condom usage.
- 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 I make explicit connections between path analysis and nested models. Examples include models that attempt to explain drinking behavior and support for interracial marriage.

- Chapter 15 covers simple nonlinear relationships. I include a detailed and graphical explanation of how these nonlinear slopes work. I use age as a typical variable that one squares, using dependent variables such as income (which produces one type of nonlinear effect) and happiness (which produces the opposite type of nonlinear effect).

If an instructor guides her students through all or most of *The Basics and Beyond*, the students will emerge with a very useful set of statistical skills that they can immediately apply in their further studies or on the job. Yes, the book takes students further than in the past. But given its user-friendly writing style and graphical approach, the book makes this completely possible. I believe it is not only possible, but necessary in today's world, where they will face these techniques on a regular basis. Most students take a single statistics course, and this one course must take them beyond the basics.