

HERSCHEL
THERESA MARCHANT-SHAPIRO



STATISTICS

for Political Analysis

Understanding the Numbers



Statistics for Political Analysis

Understanding the Numbers

Theresa Marchant-Shapiro
Southern Connecticut State University



 SAGE


CQ PRESS

Los Angeles | London | New Delhi
Singapore | Washington DC



Los Angeles | London | New Delhi
Singapore | Washington DC

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Publisher: Charisse Kiino
Associate Editor: Nancy Matuszak
Editorial Assistant: Davia Grant
Production Editor: Olivia Weber-Stenis
Copy Editor: Liann Lech
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Dennis W. Webb
Indexer: Judy Hunt
Cover Designer: Rose Storey
Marketing Manager: Erica DeLuca

Copyright © 2015 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Printed in the United States of America

ISBN 978-1-4522-5865-2

Cataloging-in Publication data is available from the Library of Congress.

This book is printed on acid-free paper.



SFI label applies to text stock

14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

Preface

I wrote *Statistics for Political Analysis: Understanding the Numbers* at the request of my students. Every semester, when the representatives for different publishers visited my office, I asked whether they had a statistics textbook designed for political science undergraduates. Although they diligently sent me political science research methods textbooks and sociology statistics textbooks, they never had quite what I was looking for. I had two primary criteria in choosing a textbook. First, I wanted it to teach the math in a step-by-step way that my math-phobic students could follow easily. Second, I wanted it designed for political science students so that it contained all of the statistics that my students needed to learn, but not much more.

I had settled on a public administration textbook as the best fit for those two requirements. But over the years, it morphed into a graduate-level textbook that contained much I didn't need to cover. Unfortunately, for continuity purposes, cutting sections out was difficult. And I was still getting comments from my students asking, "Why can't the authors just explain it the way you do in class?" Finally, one day, a book rep, seeing my frustration, asked, "Have you ever thought about writing your own textbook?" My students pounced on the idea, and so I prepared to write a chapter a week over the next semester to post online for my students to use. The final result is *Statistics for Political Analysis*, a book that my students list on their evaluations as one of the best parts of the course.

UNDERSTANDING THE NUMBERS: STEP-BY-STEP

My top priority is to make statistics accessible. In teaching the math in this book, I include the appropriate equations in the book, but I know that my students find the equations terrifying. So immediately after giving an equation, I take the students through a step-by-step process of what they need to do to solve the equation. Usually, this means completing a work table in which each column contains the next mathematical operation. This allows students the needed practice to master the process without making the process intimidating.

In addition to teaching the math in an accessible way, I limit the topics to the ones that I see as most important for political science students to master. The book covers the basic univariate statistics and the basic measures of association, as well as both bivariate and multivariate regression. It also includes a few statistical concepts that are not core for political scientists: z-scores, means testing, and analysis of variance. I include these for pedagogical reasons. I remember as an undergraduate having my mind blown by the concept of statistical significance: It took several exposures before the concept really sunk in.

I see the same confusion in my students. As a result, the book covers means testing and analysis of variance because I see them as helpful in building the cognitive structure that students need to master the notion of statistical significance. Students are still a bit uncertain about the concept, but by the end of the semester, most of them are able to explain statistical significance on the final exam correctly.

Organization

Although my minimum requirements in looking for a textbook were accessibility and coverage, in writing my own I was able to structure it to fit my philosophy of teaching and learning. Pedagogically, each chapter of the book works to build a cognitive framework for the student to learn a particular concept. Every chapter begins with a cultural example that depicts the concept in a way that is familiar to the student. As it works through the math, it explains why statisticians take each step. The steps are then summarized briefly, followed by an example in which I model doing the math. At the end of each chapter are “Your Turn” exercises for students to actually apply what they’ve learned. The modeling in the chapter makes completing the exercises very doable.

Math courses are active by nature: You can’t learn math from lectures; you have to do problems. But I think that statistics needs to be even more active than that. Learning the math is an important step in the process of understanding what assumptions statistical tests make and what the results mean. But although it is vital to develop those intuitions, the math is only a means to the more important end of analyzing data so that we can better understand the world. As a result, I have always taught my students both how to calculate statistics mathematically and how to get statistics using SPSS. Thus, after teaching the math, each chapter proceeds to teach the equivalent SPSS commands. Once again, I build cognitive frameworks by first modeling the process of using SPSS to answer a question and then giving a similar assignment in the “Apply It Yourself” section. In each problem-based learning situation, students answer a real-world question by writing a memo based on their statistical results. Each memo serves as something of a capstone for the chapter, requiring students to both analyze data and explain their results. This practical approach allows students to see the professional relevance of the material.

In 2005, the American Statistical Association recommended using real data so that students can learn about the messiness of real data and engage in the learning process,¹ and I use real-world political data throughout the book. Students major in political science because they find politics interesting; many of them hate math. Using political data for these majors certainly does spice up the otherwise boring math. But more than that, using real data shows how understanding statistics is relevant to students’ education. I believe that, as a result, this is as much a political science textbook as it is a statistics textbook.

In Chapter 1, I discuss how important statistics are to the student professionally, academically, and civically. In one of the “Your Turn” exercises in that chapter, I ask them to look for an example of statistics being used in the news. My goal is to make students conscious of how often statistics are used in the political world. To reinforce that notion, each chapter contains a “Numbers in the News” feature. Each of these depicts a time in which a concept from the chapter was relevant to something that happened politically. My hope is

that these features will make students more sensitive to how often events depicted in the news are dependent on the statistical concepts they have learned in this course.

The goal of this book is to give students a vision of how to apply the principles of statistics to analyze real-world data. Hopefully, the accessibility of the math instructions will enable students to get beyond their fears and develop intuitions about what the numbers mean. The political discussions should keep students interested long enough to realize that this course isn't just another requirement. By the end, they should realize that they have learned valuable skills that are relevant to their studies and their professional futures.

ACKNOWLEDGMENTS

Thanks go to John Kelly for asking that fateful day, "Have you ever thought about writing your own textbook?" Even more thanks go to Michael Tweedie for his enthusiasm for the idea, without which I am not sure I would have had the courage to undertake the project. I am very appreciative for the time he spent with me during the next semester, brainstorming topics that students would find interesting as examples and then finding the data that made the examples come to life.

I'm grateful for the support of all of my students and for all of their encouragement. They were unfailingly patient with the process. In particular, Michele Kuck, Kyle Tamulevich, and Paul Vitale were very helpful in finding errors in earlier drafts.

I'm blessed with supportive colleagues at Southern Connecticut State University. When I started teaching Quantitative Analysis, many students so dreaded it that they procrastinated taking it until their last semester. I'm very grateful that my colleagues agreed to refocus their student advising so that most students now finish the course well before their senior year. This contributes immeasurably to having a positive learning environment. Thanks in particular to Art Paulson and Kevin Buterbaugh for their consistent reminders to students that what they learn in this course matters both academically and professionally.

A project of this magnitude always requires family support. Thanks to my husband Andy for always having his eye out for interesting numbers in the news. Thanks to my youngest son Abram for his patience during those evenings and Saturdays when I was focused on my computer instead of him. This book has a special connection to my third child. During the fall of 2012, I spent each Saturday morning revising a different chapter at East Haven beach while my son Isaac and his fellow future soldiers did physical training in preparation for serving their country. The process reinforced to me what an incredible generation this is. Many of my students are veterans and so have already shown their commitment to service, but my other students are equally committed to serving their communities. I feel privileged to be able to spend my days helping to prepare them to reach their goals.

I'm grateful to all those who have taught me. Stan Taylor and Dave Magleby introduced me to all these concepts in the two methods classes I took as an undergraduate. I was fortunate that my graduate years at Chicago overlapped with some excellent methodologists: Henry Brady, Art Miller, and Lutz Erbring. I owe the most to my mentor, Chris Achen, who taught me both the Gauss-Markov assumptions and that OLS is robust. I never saw him do

statistics for statistics' sake. He taught me that statistics is an incredibly valuable tool for learning about the political world.

This book benefited from an excellent editorial team at SAGE/CQ Press. Publisher Charisse Kiino and Development Editor Nancy Matuszak have been full of good suggestions and encouragement. In addition to being a great copy editor, Liann Lech has a statistical mojo that she contributed to this book. The perspectives of the reviewers were very helpful in reminding me what concepts are most important: David Damore, University of Nevada, Las Vegas; Teri Fair Platt, Suffolk University; Sean Gailmard, University of California, Berkeley; and Steve B. Lem, Kutztown University of Pennsylvania. Because of their suggestions, the book has a much clearer and tighter structure. Christopher Lawrence, Middle Georgia State College, served as the technical reviewer for the book—I am indebted to his careful work. Considering the magnitude of assistance I received from this team, I have to take sole credit for any mistakes and weaknesses that remain.

¹ Guidelines for Assessment and Instruction in Statistics Education (GAISE), "College Report," *American Statistical Association*, 2005. Accessed 13 July 2013. <http://www.amstat.org/education/gaise/>.

About the Author

Theresa Marchant-Shapiro received her PhD from the University of Chicago. Throughout her career, she has been just as interested in teaching as she is in research, and she has participated in various NSF and APSA-sponsored programs for teaching statistics and research methods. She currently teaches the methods classes at Southern Connecticut State University, where she normally receives high student evaluations—an unusual occurrence in methods classes. When not busy teaching measurement with a freshly baked loaf of bread and standard errors with M&Ms, Tess teaches classes related to mass political behavior, such as Race and Ethnicity in American Politics and Political Participation. Her research focuses on various aspects of decision making, particularly in an electoral context.

Brief Contents

Preface	xiii
About the Author	xvii
1 The Political Use of Numbers: Lies and Statistics	1
2 Measurement: Counting the Biggel-Balls	17
3 Measures of Central Tendency: That's Some Mean Baseball	51
4 Measures of Dispersion: Missing the Mark	76
5 Continuous Probability: So What's Normal Anyway?	111
6 Means Testing: Sampling a Population	145
7 Hypothesis Testing: Examining Relationships	179
8 Describing the Pattern: What Do You See?	212
9 Chi-Square and Cramer's V: What Do You Expect?	245
10 Measures of Association: Making Connections	273
11 Multivariate Relationships: Taking Control	319
12 Bivariate Regression: Putting Your Ducks in a Line	343
13 Multiple Regression: The Final Frontier	377
14 Understanding the Numbers: Knowing What Counts	417
Appendixes	430
Notes	449
Glossary	455
Index	459

Detailed Contents

Preface	xiii
About the Author	xvii
1 The Political Use of Numbers: Lies and Statistics	1
The Power of Numbers	2
The Science of Politics	3
Introductory Statistics: An Overview	7
Removing the Barriers to Understanding How Statistics Works	9
The Importance of Statistics: This Book's Approach	11
Using Data to Answer a Question	11
A Political Application: Indoctrination U.	13
YOUR TURN: USING STATISTICS	15
APPLY IT YOURSELF: ASSESS GRANTS TO POLITICAL SCIENTISTS	16
Key Terms	16
2 Measurement: Counting the Biggel-Balls	17
Finding Your Cases	18
Measure an Attribute	20
<i>Distinguish the Conceptual and Operational Differences</i>	20
<i>Articulate the Operational Measure</i>	21
Evaluate the Conceptual and Operational Definitions	25
<i>Valid Measures</i>	25
<i>Measuring Variables Reliably</i>	26
Translate Information in Numbers: Coding Your Data	29
<i>Create a Coding Sheet</i>	29
<i>Code Your Data</i>	30
Get a Frequency Distribution	31
Summarizing the Process: Measurement	33
Use SPSS to Answer a Question with Measurement	34
<i>Transfer Your Data to an Electronic Database</i>	34
<i>Clean Your Data</i>	36
<i>Get a Frequency Table</i>	38
<i>An SPSS Application: Europe and the European Union</i>	39

YOUR TURN: MEASUREMENT 48

APPLY IT YOURSELF: MEASURE THE NORM FOR CHIEF JUSTICE APPOINTMENTS 49

Key Terms 50

3 Measures of Central Tendency: That's Some Mean Baseball

51

Measures of Central Tendency 53

Mean 54

Median 58

Mode 62

Summarizing the Math: Averages 64

Calculate Averages from Raw Data 64

Calculate Averages from Tables 66

Use SPSS to Answer a Question with Averages 68

An SPSS Application: Income in Poor Places 69

YOUR TURN: MEASURES OF CENTRAL TENDENCY 72

APPLY IT YOURSELF: CALCULATE THE PERCENT OF EARNED INCOME 74

Key Terms 75

4 Measures of Dispersion: Missing the Mark

76

Ranges 78

Distance from Mean 82

Calculate the Standard Deviation from Raw Data 82

Calculate the Standard Deviation from Tabular Data 84

Summarizing the Math: Dispersion 90

Calculate the Range, Interquartile Range, and Five-Point Summary 90

Calculate the Standard Deviation from Raw Data 92

Calculate the Standard Deviation from Tabular Data 95

Calculate the Standard Deviation for a Dichotomous Variable 97

Use SPSS to Answer a Question with Measures of Dispersion 98

An SPSS Application: Law School Tuition 101

YOUR TURN: MEASURES OF DISPERSION 106

APPLY IT YOURSELF: EVALUATE GRADUATES' SALARIES 108

Key Terms 110

5 Continuous Probability: So What's Normal Anyway?

111

The Normal Curve 112

z-Scores 114

Using a z Table 115

Find the Probability of Negative z-Scores 117

Find the Probability of Different Ranges: One-Tailed 117

Find the Probability of Different Ranges: Two-Tailed 120

Finding a z-Score	122
Use Probability to Calculate z-Scores	123
Summarizing the Math: Probabilities of Normally Distributed Events	126
<i>Calculate z-Scores</i>	126
<i>Find the Probability of a z-Score</i>	128
<i>Use a Probability to Calculate a Value</i>	132
Use SPSS to Answer a Question with Continuous Probability	135
<i>An SPSS Application: Crime Data</i>	137
YOUR TURN: CONTINUOUS PROBABILITY	142
APPLY IT YOURSELF: EVALUATE THE MURDER RATE	143
Key Terms	144

6 Means Testing: Sampling a Population

145

Type I and Type II Errors	146
Means Testing	148
<i>Compare Populations with Samples</i>	148
<i>Standard Error and t-Tests</i>	152
<i>Proportions as a Special Case: Showing Employment Discrimination</i>	157
Confidence Intervals: Two-Tailed Distributions	159
<i>Confidence Intervals with Proportions</i>	162
Choose a Sample Size	163
Summarizing the Math: Sampling a Population	164
<i>Means Testing</i>	165
<i>Means Testing with Proportions</i>	167
<i>Confidence Intervals</i>	169
<i>Choosing a Sample Size</i>	170
Use SPSS to Answer a Question with Means Testing	171
<i>A Political Example: Civil War and Infant Mortality</i>	173
YOUR TURN: MEANS TESTING	176
APPLY IT YOURSELF: ASSESS MATERNAL MORTALITY RATE INCREASES	177
Key Terms	178

7 Hypothesis Testing: Examining Relationships

179

Hypothesis Testing	180
<i>Hypotheses</i>	181
<i>The Null Hypothesis</i>	183
<i>Analysis of Variance</i>	185
Summarizing the Math: Hypothesis Testing and ANOVA	194
<i>Hypothesis Testing</i>	194
<i>ANOVA</i>	195
Use SPSS to Answer a Question with ANOVA	199
<i>A Political Example: Partisanship and Support for the President</i>	202

YOUR TURN: HYPOTHESIS TESTING 208

APPLY IT YOURSELF: EXAMINE PARTISANSHIP'S EFFECT ON FEELINGS TOWARD THE DEMOCRATIC PARTY 208

Key Terms 211

8 Describing the Pattern: What Do You See? 212

Choosing the Appropriate Form of Presentation 213

Graphs: Relationships and Scales 213

Visualizing a Relationship: Contingency Tables 220

Set up a Contingency Table 221

Describe the Pattern in a Contingency Table 224

Collapse Data for a Contingency Table 225

Summarizing the Math: Graphs and Contingency Tables 228

Graphs 228

Contingency Tables 231

Use SPSS to Answer a Question Using a Contingency Table 235

An SPSS Application: Life Expectancy and Female Literacy 237

YOUR TURN: DESCRIBING THE PATTERN 242

APPLY IT YOURSELF: DETERMINE STABILITY ACROSS LEGISLATIVE SYSTEMS 243

Key Terms 244

9 Chi-Square and Cramer's V: What Do You Expect? 245

The Probability of Discrete Events 246

Probabilities of Independent Events 247

Probabilities of Contingent Events 249

Chi-Square 252

What We Expect to See 253

Comparing What We Expect with What We Actually Observe 254

Cramer's V 258

Summarizing the Math: Chi-Square and Cramer's V 260

Chi-Square 260

Cramer's V 261

Use SPSS to Answer a Question with Chi-Square and Cramer's V 264

An SPSS Application: Cross Tabulation of Religious Freedom 265

YOUR TURN: CHI-SQUARE AND CRAMER'S V 269

APPLY IT YOURSELF: ANALYZE DATA BY TYPE 272

Key Terms 272

10 Measures of Association: Making Connections 273

Basic Principles of Measures of Association 274

Pearson's r 278

Get a Visual Intuition for Correlation 278

<i>Setting up the Work Table for Pearson's r</i>	281
<i>Working through an Example: Political Rights and Civil Liberties</i>	283
Gamma	285
<i>Positive and Negative Relationships in Contingency Tables</i>	286
Gamma for Dichotomous Variables	287
Gamma for More than Two Categories	290
Kendall's Tau	296
Lambda	297
Calculating Lambda	297
<i>Working through an Example: Generations of Partisanship</i>	298
Limitations of Lambda	300
Summarizing the Math: Measures of Association	302
Pearson's r for Two Interval-Level Variables	303
Gamma for Two Ordinal-Level Variables	306
Lambda for Nominal-Level Variables	308
Use SPSS to Answer a Question with Measures of Association	310
<i>An SPSS Application: The Internet and Political Instability</i>	312
YOUR TURN: MEASURES OF ASSOCIATION	314
APPLY IT YOURSELF: MEASURE POOR STUDENT GRADUATION RATES	317
Key Terms	318

11 Multivariate Relationships: Taking Control

319

Spurious Relationships	320
<i>Antecedent Variables</i>	321
<i>Intervening Variables</i>	321
<i>Spurious Non-Relationships</i>	322
Interaction Effects	323
Three-Way Contingency Tables	324
Summarizing the Process: Setting up Three-Way Contingency Tables	329
<i>A Political Example: Abortion and Religious Fundamentalism</i>	330
Use SPSS to Answer a Question with a Three-Way Contingency Table	332
<i>An SPSS Application: Views on Legal Abortion for Rape among Catholics</i>	333
YOUR TURN: MULTIVARIATE RELATIONSHIPS	339
APPLY IT YOURSELF: ANALYZE DATA ON RACE FOR PARTISANSHIP AND INCOME	340
Key Terms	342

12 Bivariate Regression: Putting Your Ducks in a Line

343

Graph a Relationship	344
<i>Plot the Data</i>	344
<i>Find a Line</i>	346
Fit the Data with the Ordinary Least Squares Estimate of the Line	346
Find the Statistical Significance	350

Find the Strength of the Relationship	353
Use Regressions with Time Series Data	354
<i>A Political Example: Murder Rates</i>	354
Interpret Regressions with Dichotomous Independent Variables	356
<i>A Political Example: Income and Gender</i>	356
Summarizing the Math: Regression	357
<i>Describe the Pattern by Estimating the Line</i>	357
<i>Identify the Statistical Significance with the Standard Error</i>	358
<i>Evaluate the Strength of the Association with R^2</i>	359
<i>A Political Example: Birth Rates and Abortion Rates</i>	359
Use SPSS to Answer a Question with Bivariate Regression	363
<i>A Political Example: State Expenditures and Population</i>	366
YOUR TURN: BIVARIATE REGRESSION	371
APPLY IT YOURSELF: ANALYZE INFLUENCES ON CORRUPTION	374
Key Terms	376

13 Multiple Regression: The Final Frontier

377

Using Regression to Control for Other Variables	378
The Assumptions of Regression	381
<i>Gauss-Markov Assumption 1: Interval-Level Variables</i>	382
<i>Gauss-Markov Assumption 2: Linear Relationship</i>	384
<i>Gauss-Markov Assumption 3: Model Correctly Specified</i>	393
<i>Gauss-Markov Assumption 4: Non-Collinear</i>	394
<i>Gauss-Markov Assumption 5: Errors Have a Mean of Zero</i>	397
<i>Gauss-Markov Assumption 6: Errors are Homoscedastic</i>	397
Summarizing the Process: Multiple Regression	398
<i>Analyzing Multiple Regression</i>	398
<i>The Assumptions of Regression</i>	399
<i>A Political Example: Teen Pregnancy</i>	400
Use SPSS to Answer a Question with Multiple Regression	401
<i>Dummy Variables</i>	401
<i>Bivariate Scatter Plots</i>	402
<i>Residual Scatter Plots</i>	403
<i>An SPSS Application: The Relationship of Multiple Factors on State Instability</i>	404
YOUR TURN: MULTIPLE REGRESSION	412
APPLY IT YOURSELF: EVALUATE THE IMPACT OF MULTIPLE FACTORS ON THE 2012 PRESIDENTIAL ELECTION	414
Key Terms	416

14 Understanding the Numbers: Knowing What Counts

417

Measurement	418
Univariate Statistics	419

Multivariate Statistics	421
<i>Hypotheses</i>	421
<i>Describe the Pattern</i>	422
<i>Identify the Statistical Significance</i>	423
<i>Evaluate the Substantive Significance</i>	424
Keeping the Numbers Meaningful	427
Embracing the Uncertainty	428

Appendix 1: Tips for Professional Writing	430
Appendix 2: How to Use SPSS	436
Appendix 3: z Table	443
Appendix 4: t Table	445
Appendix 5: Chi-Square Table	447
Notes	449
Glossary	455
Index	459

C H A P T E R 1

The Political Use of Numbers

Lies and Statistics

In 1904, Mark Twain was busy writing his last book, an autobiography, in which he reminisced about his life. At one point, as he recalled writing *Innocents Abroad* in 1868, he became a bit mournful as he remembered that while writing it, he had been able to write 3,000 words a day. He began to feel his age as he remembered that twenty-nine years later, in 1897, he had slowed down from producing 3,000 to only 1,800 words a day, and seven years later, in 1904, his output had declined by an additional 400 words to only 1,400 words a day. But as he began to despair over the limitations of age, Twain realized that what had declined in those years was not how fast he wrote, but how much time he spent writing. He had begun by writing seven to nine hours a day, but ended with only four to five. After he realized that the number of pages had halved when the number of hours spent writing had also halved, he understood that the change in writing was not that his mind had gotten any less sharp; rather, he just didn't work as much. Twain concluded that numbers do not always speak for themselves. And then he repeated a well-known saying of the day, attributing it to Benjamin Disraeli: "There are three kinds of lies: lies, damned lies, and statistics." For Twain,¹ the aphorism was a reminder to be careful about the use of numbers.

In 1954, well-known "How to" author Darrell Huff² repeated the refrain when he wrote his best-selling statistics textbook *How to Lie with Statistics*. In it, Huff regales readers with examples of statistics used to delude the public. The interesting thing about his examples is that the lies are perpetrated by the violation of principles taught in all introductory texts. He begins the work in the same place as most statistics texts, with measurement. The ability to perform statistical analyses depends on our ability to attach numbers to concepts. How well we measure those concepts determines how accurately we can describe our world. Huff then proceeds to work his way through all the major topics addressed in any introductory statistics textbook: measures of central tendency, measures of dispersion,