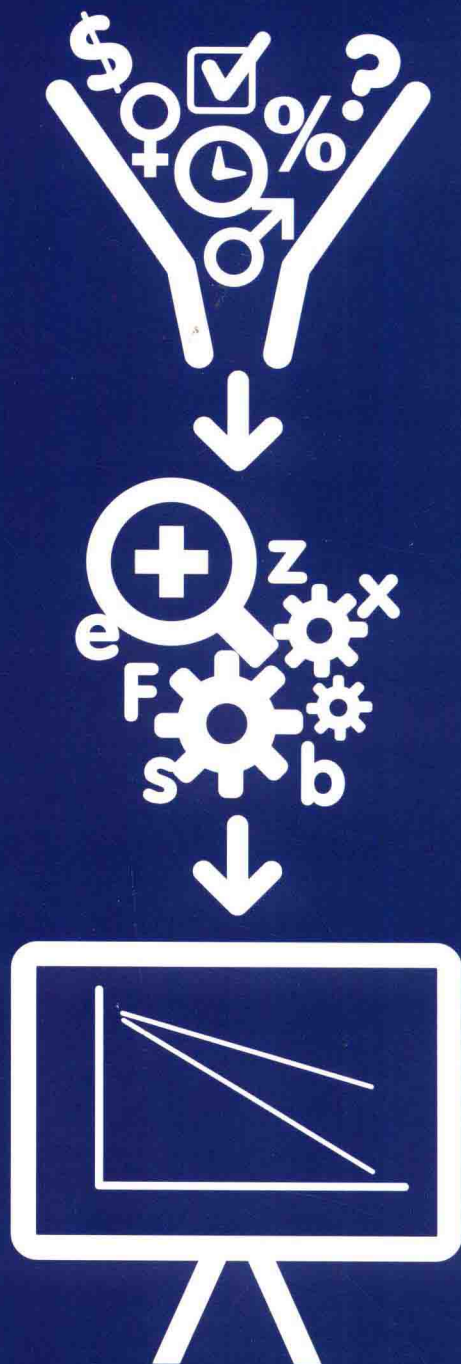


SECOND EDITION

SOCIAL STATISTICS

Managing Data,
Conducting Analyses,
Presenting Results



Thomas J. Linneman

Visit the companion website at:
www.routledge.com/cw/Linneman



SOCIAL STATISTICS
MANAGING DATA,
CONDUCTING ANALYSES,
PRESENTING RESULTS



Second Edition

Thomas J. Linneman

 **Routledge**
Taylor & Francis Group
NEW YORK AND LONDON

First published 2014
by Routledge
711 Third Avenue, New York, NY 10017

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

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

© 2014 Taylor & Francis

The right of Thomas J. Linneman to be identified as author of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

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

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Linneman, Thomas John.

Social statistics : managing data, conducting analyses, presenting results /
by Thomas J. Linneman. — Second Edition.

pages cm. — (Contemporary sociological perspectives)

Includes bibliographical references and index.

1. Social sciences—Statistical methods. 2. Statistics. I. Title.

HA29.L83118 2014

519.5—dc23 2013029016

ISBN: 978-0-415-66146-1 (hbk)

ISBN: 978-0-415-66147-8 (pbk)

ISBN: 978-0-203-07342-1 (ebk)

Typeset in Times New Roman

by Apex CoVantage, LLC



SUSTAINABLE
FORESTRY
INITIATIVE

Certified Chain of Custody

At Least 20% Certified Forest Content

www.sfiprogram.org

SFI-00712

SOCIAL STATISTICS

Many fundamentally important decisions about social life are a function of how well we understand and analyze *data*. This sounds so obvious but it is so misunderstood. Social statisticians struggle with this problem in their teaching constantly. This book and its approach are the ally and a support for all instructors who want to accomplish this hugely important teaching goal.

This innovative text for undergraduate social statistics courses is (as one satisfied instructor put it), a “breath of fresh air.” It departs from convention by not covering some techniques and topics that have been in social statistics textbooks for 30 years but that are no longer used by social scientists today. It also *includes* techniques that conventional wisdom has previously thought to be the province of graduate level courses.

Linneman’s text is for those instructors looking for a thoroughly “modern” way to teach quantitative thinking, problem-solving, and statistical analysis to their students . . . an undergraduate social statistics course that recognizes the increasing ubiquity of analytical tools in our data-driven age and therefore the practical benefit of learning how to “do statistics,” to “present results” effectively (to employers as well as instructors), and to “interpret” intelligently the quantitative arguments made by others.

Thomas J. Linneman is Associate Professor of Sociology at the College of William and Mary in Williamsburg, Virginia. He teaches courses on statistics, social change, sexualities, and the media. At William and Mary, he has been the recipient of the Thomas Jefferson Teaching Award, the highest teaching honor given annually to a younger member of the faculty. The citation for his award noted that Linneman has developed a reputation among his students as a demanding professor but one who genuinely cares about them. His teaching evaluations for his statistics course are regularly a standard deviation above the mean.

Contemporary Sociological Perspectives

Edited by Doug Hartmann, University of Minnesota, Valerie Jenness, University of California, Irvine and Jodi O'Brien, Seattle University

This innovative series is for all readers interested in books that provide frameworks for making sense of the complexities of contemporary social life. Each of the books in this series uses a sociological lens to provide current critical and analytical perspectives on significant social issues, patterns and trends. The series consists of books that integrate the best ideas in sociological thought with an aim toward public education and engagement. These books are designed for use in the classroom as well as for scholars and socially curious general readers.

Published:

Political Justice and Religious Values by Charles F. Andrain

GIS and Spatial Analysis for the Social Sciences by Robert Nash Parker and Emily K. Asencio

Hoop Dreams on Wheels: Disability and the Competitive Wheelchair Athlete by Ronald J. Berger

The Internet and Social Inequalities by James C. Witte and Susan E. Mannon

Media and Middle Class Mom: Images and Realities of Work and Family by Lara Descartes and Conrad Kottak

Watching T.V. Is Not Required: Thinking about Media and Thinking about Thinking by Bernard McGrane and John Gunderson

Violence Against Women: Vulnerable Populations by Douglas Brownridge

State of Sex: Tourism, Sex and Sin in the New American Heartland by Barbara G. Brents, Crystal A. Jackson & Kate Hausbeck

Sociologists Backstage: Answers to 10 Questions About What They Do by Sarah Fenstermaker and Nikki Jones

Gender Circuits by Eve Shapiro

Surviving the Holocaust: A Life Course Perspective by Ronald Berger

Transforming Scholarship: Why Women's and Gender Studies Students Are Changing Themselves and the World by Michelle Berger and Cheryl Radeloff

Stargazing: Celebrity, Fame, and Social Interaction by Kerry Ferris and Scott Harris

The Senses in Self, Society, and Culture by Phillip Vannini, Dennis Waskul and Simon Gottschalk

Who Lives, Who Dies, Who Decides? by Sheldon Ekland-Olson

Surviving Dictatorship by Jacqueline Adams

The Womanist Idea by Layli Maparyan

Social Theory Re-Wired: New Connections to Classical and Contemporary Perspectives by Wesley Longhofer and Daniel Winchester

Religion in Today's World: Global Issues, Sociological Perspectives, by Melissa Wilcox

Life and Death Decisions: The Quest for Morality and Justice in Human Societies, Sheldon Ekland-Olson

Understanding Deviance: Connecting Classical and Contemporary Perspectives, Tammy L. Anderson

Titles of Related Interest:

The Connected City by Zachary Neal

Regression Analysis for the Social Sciences by Rachel A. Gordon

Applied Statistics for the Social and Health Sciences by Rachel A. Gordon

GIS and Spatial Analysis: A Tool for the Social Sciences by Robert Parker and Emily Asencio

TABLE OF CONTENTS IN DETAIL

<i>Preface</i>	xxix
<i>Acknowledgments</i>	xxxvii

Chapter 1:	Life in a Data-Laden Age: Finding and Managing Datasets	1
	Introduction: details the ways in which this book is unique	1
	What Data Look Like: introduces the forms that data take in the rawest of ways, and what variables and categories are	3
	Making the Data Work for You: covers the very important ways in which we manage data, such as recoding and indexing	6
	Our Datasets: describes the datasets that are used throughout the book in examples and exercises: the General Social Survey, the American National Election Studies, the World Values Survey, and surveys from the Pew Internet & American Life Project	9
	Other Great Datasets: briefly describes other commonly used datasets, such as the Panel Study of Income Dynamics, and the National Longitudinal Study of Adolescent Health	11
	GSS Example: An Index of Work Hostility: illustrates how to recode and combine variables to develop an index of the hostility workers experience at their jobs	13
	New Forms of Data: makes the case that there are many new forms of data available for analysis due to the explosion of media availability, old and new	17
	Levels of Measurement: carefully delineates the differences among different types of variables and why knowing these differences are important	18

	Major Types of Statistical Procedures: covers the three types of statistics that the book covers: descriptive, inferential, and explanatory	22
	Literature Example: Wikipedia as a Data Source: describes a piece of research that argues that wiki-led organizations play by different rules than most organizations, and uses contested Wikipedia pages to make the argument	23
	Literature Example: IMDb as a Data Source: provides an example of how researchers used the vast database of movies, actors, and directors to study social networks	24
	Conclusion	25
	SPSS Demonstrations: introduces students to SPSS, and covers very important ways to work with data, such as recoding and computing new variables	25
	From Output to Presentation: recommends what to say about a dataset when you're using it in a presentation or paper	40
	Exercises: examine such topics as American attitudes toward wiretapping, concern over China's economic expansion, and social desirability bias in a survey of teenagers	41
Chapter 2:	The Art of Visual Storytelling: Creating Accurate Tables and Graphs	46
	Introduction: makes the case for the importance of tables and graphs, and knowing how to make good ones that don't lie	46
	Tables with One Variable: Frequency Distributions: takes students step by step through the process of building a clear summary of a single variable	47
	GSS Example: Number of Children: walks students through a typical frequency distribution from SPSS, such as the difference between percents and valid percents	48
	Grouped Frequency Distributions: describes when it might be preferable to present a variable's data in grouped form	49
	Tables with Two Variables: Crosstabulations: offers a step-by-step demonstration of how to construct a table with an independent variable and a dependent variable	50
	GSS Example: Spanking and Child-Rearing Goals: uses crosstabs to show how Americans' propensity to spank their children is related to specific goals	53
	GSS Example: Education and Internet Access: models for students every possible mistake you could make, resulting in a horrible crosstab; then reconstructs it correctly	55

Tables with Three Variables: Elaboration in Crosstabs: covers the very important process of elaborating a relationship, showing where a relationship is stronger or weaker, exists or doesn't exist	57
GSS Example: Chivalry, Age, and Gender: uses elaboration to examine who is most likely to give up a seat for another person	58
GSS Example: Racial Differences over Time: offers an example of how a relationship changes over time, using elaboration to show this	60
GSS Example: Gender, Unemployment, and Happiness: investigates whether men or women are more affected by unemployment	61
Graphs with One Variable: examines how to choose among bar, line, and pie graphs	63
Graphs with Two Variables: covers clustered bar graphs and stacked bar graphs	66
Graphs with Three Variables: covers 3-D bar graphs and plotted pie graphs	68
Tufte's Lie Factor: offers a classic way to judge the extent to which a graphic accurately represents the change in data	70
GSS Example: Support for Marijuana Legalization: compares two bar graphs that represent the same data in different ways, leading to different emphases	73
Literature Example: Changing Racial Classification: describes an article in a top research journal that presented a misleading graph	75
Conclusion	77
SPSS Demonstrations: goes through table and graph construction in numerous ways	77
From Output to Presentation: shows how crosstabs don't usually appear in pure crosstab form in professional settings	86
Exercises: explore such issues as perceptions of racial discrimination, the effect of caring for a sick child on one's quality of life, and how many parents say their children don't use social networking sites, when actually they do	87
Chapter 3: Summarizing Center and Diversity: Basic Descriptive Statistics	92
Introduction: discusses in general the field of descriptive statistics	92
Three Ways of Thinking about the Center: using a hypothetical example of housing prices, covers the mean, median, and mode	93
GSS Example: TV Watching: gives an example of how, just based on a few descriptive statistics, we can develop a pretty good guess regarding what the distribution for a variable looks like	96

Procedures for Finding the Median: covers the basics of median finding with even and odd numbers of cases	97
Finding the Centers with a Frequency Distribution: goes through the procedures for locating the mean, median, and mode when the data are in frequency distribution format	98
Measures of Center and Levels of Measurement: explains the connections among descriptive statistics and the levels of measurement covered in Chapter One	100
Close Relatives of the Median: Quartiles and Percentiles: describes why we might want to look at the 25th, 50th, and 75th percentiles when examining our data	101
GSS Example: TV Watching Rerun: shows how to find the quartiles for the frequency distribution we examined earlier in the chapter	101
Envisioning Variation: introduces the very important concept of variation, showing how it exists throughout our daily lives	102
Assessing Variation among Groups: covers how to find the range, the interquartile range, and, most importantly, the variance	103
GSS Example: Educational Attainment: offers an example of how the variance for education differs among racial groups	108
Visual Variation: The Shapes of Distributions: uses a hypothetical series of examples to show how we can observe different levels of variation graphically	109
Assessing the Variation of Individual Cases: introduces the standard deviation and z -scores, and shows how to calculate them and talk about them	111
GSS Example: Internet Usage among Racial Groups: illustrates the uniqueness of various types of people with regard to how many hours per week they use the Internet	113
Finding s and s^2 with a Frequency Distribution: explains how the formulas change once we're looking at the data in grouped form	115
Variation When There Is No Mean: introduces the Index of Qualitative Variation	116
GSS Example: Attitudes toward Government Spending: using the IQV, shows how variation in attitudes has changed over time for numerous social issues	123
Other Measures of Diversity: briefly discusses two common measures: the index of dissimilarity and the Gini coefficient	126
Literature Example: Cost of Medical Care: describes a student-led research project that exposed frightening levels of variation in medical costs	127
Conclusion	128

	SPSS Demonstration: shows students numerous ways to get descriptive statistics in SPSS, and how to represent them in graph form	128
	From Output to Presentation: models the typical way descriptive statistics are presented in professional settings	135
	Exercises: examine such topics as gender differences in attitudes toward feminists, variation in health for people in different income brackets, and teen texting habits	136
Chapter 4:	Using Sample Crosstabs to Talk about Populations: The Chi-Square Test	141
	Introduction: discusses the importance of sampling, the difference between populations and samples, and what an inference is	141
	A Series of Hypothetical Crosstabulations: illustrates different levels of relationship in a crosstab, from perfect to none, and how this relates to the inference we're trying to make based on these results	143
	Calculating Expected Frequencies: shows, based on what we've observed in our actual data, what these really mean (i.e., what are we actually expecting?)	146
	Introducing Chi-Square: goes through the calculation of the chi-square value step by step, using one of the hypothetical crosstabs we've already covered	151
	Statistical Significance: covers one of the most important concepts in social statistics, as it determines what we can say about our population based on our sample crosstab; explains, in reference to the hypothetical crosstabs, the idea of type one and type two error; shows how to use the chi-square table	153
	The Effect of Sample Size: drives home the point that the chi-square procedure and statistical significance are reliant on sample size	156
	Chi-Square with Larger Crosstabs: covers the concept of degrees of freedom and what, in the context of a crosstab, this actually means	157
	GSS Example: Gun Ownership and Sex: goes through the entire chi-square procedure again, this time using real data	160
	GSS Example: Age and Cynicism: emphasizes the importance of looking out for a few cells that are causing a high chi-square value and statistical significance	162
	The Language of Hypothesis Testing: introduces the oft-used language of null and alternative hypotheses	164
	The Chi-Square Distribution: shows, now that we know what chi-square does, what is really going on in terms of a probability	

	distribution; builds a chi-square distribution step by step to make clear the connection between this distribution and statistical significance	165
	GSS Example: Catholic Confidence in Organized Religion: combines chi-square with elaboration to show how time, gender, and confidence are related	171
	GSS Example: Guns, Age, and Sex: returns to an earlier example and shows how elaboration and chi-square can be used to summarize an interesting story	174
	Literature Example: Real Research about Poop, Really: goes over a fascinating example in which researchers used crosstabs and chi-square to illuminate differences based on sex and sexual orientation with regard to college students' paranoia over public restrooms	175
	Literature Example: Obesity in the United States and France: summarizes a cross-cultural study of media coverage of obesity, using crosstabs and chi-square to show differences in how this social problem is framed in the two countries	178
	Conclusion	180
	SPSS Demonstration: goes through how to conduct a chi-square test in SPSS and interpret the output	181
	From Output to Presentation: models how to present chi-square results in a professional format	184
	Exercises: examine such topics as the relationship between smoking and sexual orientation, the phenomenon of customer showrooming in retail stores, and parental propensity for friending their children on social networks	185
Chapter 5:	Using a Sample Mean or Proportion to Talk about a Population: Confidence Intervals	189
	Introduction: makes connections between what we did in the chi-square chapter and where we are headed in this chapter	189
	Sampling Distributions of Sample Means: illustrates the step-by-step building of a sampling distribution, showing the probabilities of pulling various sample means	190
	The Standard Error: makes very clear, based on the previous example, what the standard error really is, why it might be useful, and how we can estimate it	200
	Claims about the Population Mean: introduces the first use of these new ideas: using sample data to refute a claim someone makes about a population mean	203
	GSS Example: TV Watching among Young Women: puts to use the population claim procedure using real data	207

	Confidence Intervals: introduces the second use of these new ideas: based on sample data, confidently predicting where the population mean can fall	207
	GSS Example: Police Violence: builds confidence intervals for various racial groups with regard to their support for police use of violence	212
	GSS Example: Job Stress and Satisfaction: shows how variation in a sample can affect confidence interval width	215
	Confidence Intervals with Proportions: modifies the confidence interval procedure to build margins of error around sample proportions	215
	GSS Example: Support for Marijuana Legalization: puts the student in the situation of working for a politician speaking on this issue before two different age groups	218
	Literature Example: Black Characters in Film: recounts a content analysis of recent popular films that feature certain types of black characters, and uses confidence intervals to do it	218
	Conclusion	220
	SPSS Demonstrations: cover the important and often tricky procedure of selecting cases, as well as covering confidence intervals	220
	From Output to Presentation: shows how professionals often present confidence interval results	227
	Exercises: examine such topics as Internet news consumption, Hispanics' acquisition of health insurance, and teenagers engaging in "sexting"	228
Chapter 6:	Using Multiple Sample Means to Talk about Populations: <i>t</i> -Tests and ANOVA	231
	Introduction: makes connections between the previous two chapters and where this chapter will take the student	231
	A Different Kind of Sampling Distribution: builds, step by step, a sampling distribution of sample mean differences to illustrate the key concept of the chapter	232
	Testing Differences between Two Means: The <i>t</i> -test: introduces the essence of the <i>t</i> -test and how the language of hypothesis testing is used with the test	235
	GSS Example: Back to TV Watching: goes through the entire <i>t</i> -test procedure using real data	235
	Looking More Closely at the Formula: manipulates the results from the previous example to illustrate how the <i>t</i> -test formula really works	237

	GSS Example: Suicide, Age, and Political Party: compares strong Democrats and strong Republicans for three age groups on an index of suicide support	238
	Testing the Difference among More than Two Means: ANOVA: walks students through the ANOVA procedure, pointing out its similarities to the <i>t</i> -test and chi-square test	240
	A Comparative Graphical Approach to Understanding ANOVA: uses a series of three ANOVAs to show how variation affects the outcome	245
	GSS Example: Attitude versus Behavior about Housework: uses two ANOVAs to see if men's and women's attitudes affect their actual housework behavior	247
	Interchapter Connection: ANOVA and Chi-Square: describes how these two related tests both set up hypothetical situations to which to compare the actual data	250
	GSS Example: Internet Use, Race, and Gender: shows that, similar to the chi-square test, a single group can be responsible for a statistically significant ANOVA	250
	Literature Example: Overdoing Gender: recounts an article that uses <i>t</i> -tests to show that when men's gender is threatened, they react in very interesting ways	251
	Literature Example: Activism through the Life Course: covers an article that compares three groups of people who had different levels of activism in college in order to see how active they are a generation later	253
	Conclusion . . . with Interchapter Connections: reviews how to choose which technique to use based on what types of variables you have	255
	SPSS Demonstrations: covers <i>t</i> -tests and ANOVA	255
	From Output to Presentation: models how to present a series of <i>t</i> -test results	259
	Exercises: addresses such questions as: do those who vote trust government more than those who don't vote? Do women use technology to shop more than men? Does residential location affect health?	260
Chapter 7:	Give Me One Good Reason Why: Bivariate Correlation and Regression	264
	Introduction: starts the hypothetical example that runs through the chapter: the relationship between men's heights and incomes	264
	Linear Equations: reviews the basics of linear equations, such as constants and slopes	265

Calculating the Regression Equation: takes things step by step to calculate the slope and constant	267
Calculating the Correlation Coefficient: describes how to find the correlation coefficient and what this number means	273
The Effects of an Outlier: modifies one of the original data points from the previous example to show that, particularly in a small dataset, an outlier can have a huge effect on the regression results	276
Explained Variation: covers one of the most important concepts in regression, and does so in a way that explains exactly what explained variation is	277
Example: Forecasting with Regression: shows how regression can be used to forecast into the future, using data regarding movie grosses	281
GSS Example: Education and Income: uses real data to analyze a classic relationship using simple regression	285
GSS Example: Income and Hours to Relax: switches income to independent variable status to see if it affects how much time people have to relax, emphasizing that sometimes a lack of effect is the most interesting finding of all	288
GSS Example: Explaining Cynicism: uses several demographic variables, one at a time, of course, to explain why some people are more cynical than others	288
GSS Example: Intergenerational Family Size: runs the same regression for several groups, foreshadowing a more advanced technique later in the book	289
Literature Example: Support for the War on Terror: covers an article that uses correlations to examine relationships between authoritarianism and support for various civil-rights-restricting policies	290
Literature Example: Physical Attractiveness: recounts an article that uses correlations to examine, for both men and women, which factors are related to people's ratings of others' physical attractiveness	291
Conclusion	293
SPSS Demonstration: Creating a Scatterplot: shows how to create a visual representation of the relationship between two ratio-level variables	293
Exercises: analyze such topics as traditionalism, environmentalism, and AIDS	295

Chapter 8:	Using Sample Slopes to Talk about Populations: Inference and Regression	303
	Introduction: links regression-related inference with the types of inference already covered in the book	303
	One More Sampling Distribution: builds one last sampling distribution, showing how the repeated random sampling of slopes gives us another useful probability distribution	304
	From Standard Error to a <i>t</i>-Value to a <i>p</i>-Conclusion: introduces the standard error of the slope, and how to use it to conduct a <i>t</i> -test to determine the statistical significance of our sample slope, making connections to previous similar techniques, and emphasizing the importance of sample size	306
	GSS Example: Education and Sexual Activity: investigates the relationship between education and amount of sex, looking at different age groups with differing sample sizes	311
	GSS Example: Income and Sexual Activity: drives home the point that statistical significance can differ from substantive significance	313
	GSS Example: Work Time and Sexual Activity: provides another example how a lack of a statistically significant finding can still be thought of as interesting	314
	Literature Example: Grade Point Averages: recounts an article that tried to find evidence for a commonly assumed relationship: how much a student studies and her G.P.A.	314
	Literature Example: Family Size and Grades: tells the story of an article that investigates whether kids from large families have lower grades, and provides another distinction between statistical and substantive significance	316
	Conclusion	317
	SPSS Demonstration: Regression: walks through the SPSS Regression procedure and how to make sense of the output	317
	Creating a Correlation Matrix: shows how to run and read a set of correlations in the form of a matrix	320
	From Output to Presentation: presents the first example of what will soon become a theme: how to present regression results in a format similar to that in professional research	321
	Exercises: investigate such topics as the relationship between age and concern about privacy, the relationships among health conditions, and kids' perceptions of their parents' cyberblocking behavior	322

Chapter 9:	It's All Relative: Dichotomies as Independent Variables in Regression	326
	Introduction: recalls the importance of levels of measurement and why we may want to use variables other than those at the ratio level in regression	326
	Dichotomies: shows how, once you recode a dichotomy into 0 and 1, interpreting the slope of a dichotomy is quite simple	327
	Interchapter Connection: <i>t</i> -Tests versus Regression: reruns the previous regression example as a <i>t</i> -test, showing the similarities and differences between <i>t</i> -tests and regression	328
	Categorical Variables: introduces the technique of reference grouping, where we create dichotomies for all but one of the categories of a (typically) nominal-level variable, leaving one group as the reference category	329
	GSS Example: Variation in STEM Achievement: investigates variation in science and math achievement using sex, race, and political views as independent variables	331
	GSS Example: TV Watching and Work Status: shows how to use various work status variables as a set of reference-group variables	333
	GSS Example: Happiness and Partnership Status: investigates variation in happiness based on a set of reference-group variables regarding partnership status	334
	GSS Example: Party Identification and Political Knowledge: uses a set of reference-group variables to see if strength of political affiliation affects self-reported political knowledge	335
	Literature Example: Gender and Housework: examines an article that uses ratio-level variables, dichotomies, and reference groups to explain variation in time spent on housework	337
	Literature Example: Tracking Changes over Time: reports from an article that uses reference groups to show the changes over time in attitudes toward homosexuality	339
	Conclusion	341
	SPSS Demonstration: Reference Grouping: walks through the process of creating a set of reference-group variables and running a regression with them	341
	From Output to Presentation: continues the reporting of regression results in tabular form, showing how it is really useful when using reference groups	343
	Exercises: examine such issues as political efficacy, conspiracy theories, and economic optimism	344