

FIFTH EDITION

Intro Stats

Richard D. De Veaux

Williams College

Paul F. Velleman

Cornell University

David E. Bock

Cornell University

Preface ix

Index of Applications xxii

PART I Exploring and Understanding Data**1 Stats Starts Here 1**

1.1 What Is Statistics? • 1.2 Data • 1.3 Variables • 1.4 Models

2 Displaying and Describing Data 17

2.1 Summarizing and Displaying a Categorical Variable • 2.2 Displaying a Quantitative Variable • 2.3 Shape • 2.4 Center • 2.5 Spread

3 Relationships Between Categorical Variables—Contingency Tables 64

3.1 Contingency Tables • 3.2 Conditional Distributions • 3.3 Displaying Contingency Tables • 3.4 Three Categorical Variables

4 Understanding and Comparing Distributions 95

4.1 Displays for Comparing Groups • 4.2 Outliers • 4.3 Re-Expressing Data: A First Look

5 The Standard Deviation as a Ruler and the Normal Model 122

5.1 Using the Standard Deviation to Standardize Values • 5.2 Shifting and Scaling • 5.3 Normal Models • 5.4 Working with Normal Percentiles • 5.5 Normal Probability Plots

Review of Part I: Exploring and Understanding Data 154

PART II Exploring Relationships Between Variables**6 Scatterplots, Association, and Correlation 183**6.1 Scatterplots • 6.2 Correlation • 6.3 Warning: Correlation $\neq$ Causation • *6.4 Straightening Scatterplots**7 Linear Regression 196**7.1 Least Squares: The Line of “Best Fit” • 7.2 The Linear Model • 7.3 Finding the Least Squares Line • 7.4 Regression to the Mean • 7.5 Examining the Residuals • 7.6 R^2 —The Variation Accounted for by the Model • 7.7 Regression Assumptions and Conditions

*Indicates optional sections.

8 Regression Wisdom 233

8.1 Examining Residuals • **8.2** Extrapolation: Reaching Beyond the Data • **8.3** Outliers, Leverage, and Influence • **8.4** Lurking Variables and Causation • **8.5** Working with Summary Values • ***8.6** Straightening Scatterplots—The Three Goals • ***8.7** Finding a Good Re-Expression

9 Multiple Regression 276

9.1 What Is Multiple Regression? • **9.2** Interpreting Multiple Regression Coefficients • **9.3** The Multiple Regression Model—Assumptions and Conditions • **9.4** Partial Regression Plots • ***9.5** Indicator Variables

Review of Part II: Exploring Relationships Between Variables 307**PART III Gathering Data****10 Sample Surveys 319**

10.1 The Three Big Ideas of Sampling • **10.2** Populations and Parameters • **10.3** Simple Random Samples • **10.4** Other Sampling Designs • **10.5** From the Population to the Sample: You Can't Always Get What You Want • **10.6** The Valid Survey • **10.7** Common Sampling Mistakes, or How to Sample Badly

11 Experiments and Observational Studies 343

11.1 Observational Studies • **11.2** Randomized, Comparative Experiments • **11.3** The Four Principles of Experimental Design • **11.4** Control Groups • **11.5** Blocking • **11.6** Confounding

Review of Part III: Gathering Data 367**PART IV From the Data at Hand to the World at Large****12 From Randomness to Probability 373**

12.1 Random Phenomena • **12.2** Modeling Probability • **12.3** Formal Probability • **12.4** Conditional Probability and the General Multiplication Rule • **12.5** Independence • **12.6** Picturing Probability: Tables, Venn Diagrams, and Trees • **12.7** Reversing the Conditioning and Bayes' Rule

13 Sampling Distribution Models and Confidence Intervals for Proportions 410

13.1 The Sampling Distribution Model for a Proportion • **13.2** When Does the Normal Model Work? Assumptions and Conditions • **13.3** A Confidence Interval for a Proportion • **13.4** Interpreting Confidence Intervals: What Does 95% Confidence Really Mean? • **13.5** Margin of Error: Certainty vs. Precision • ***13.6** Choosing the Sample Size

14 Confidence Intervals for Means 441

14.1 The Central Limit Theorem • **14.2** A Confidence Interval for the Mean • **14.3** Interpreting Confidence Intervals • ***14.4** Picking Our Interval up by Our Bootstraps • **14.5** Thoughts About Confidence Intervals

15 Testing Hypotheses 472

15.1 Hypotheses • 15.2 P-Values • 15.3 The Reasoning of Hypothesis Testing
 • 15.4 A Hypothesis Test for the Mean • 15.5 Intervals and Tests • 15.6 P-Values and
 Decisions: What to Tell About a Hypothesis Test

16 More About Tests and Intervals 507

16.1 Interpreting P-Values • 16.2 Alpha Levels and Critical Values • 16.3 Practical vs.
 Statistical Significance • 16.4 Errors

Review of Part IV: From the Data at Hand to the World at Large 532

PART V Inference for Relationships**17** Comparing Groups 541

17.1 A Confidence Interval for the Difference Between Two Proportions • 17.2 Assumptions
 and Conditions for Comparing Proportions • 17.3 The Two-Sample z -Test: Testing for the
 Difference Between Proportions • 17.4 A Confidence Interval for the Difference Between
 Two Means • 17.5 The Two-Sample t -Test: Testing for the Difference Between Two Means
 • *17.6 Randomization Tests and Confidence Intervals for Two Means • *17.7 Pooling
 • *17.8 The Standard Deviation of a Difference

18 Paired Samples and Blocks 565

18.1 Paired Data • 18.2 The Paired t -Test • 18.3 Confidence Intervals for Matched Pairs
 • 18.4 Blocking

19 Comparing Counts 610

19.1 Goodness-of-Fit Tests • 19.2 Chi-Square Test of Homogeneity • 19.3 Examining the
 Residuals • 19.4 Chi-Square Test of Independence

20 Inferences for Regression 642

20.1 The Regression Model • 20.2 Assumptions and Conditions • 20.3 Regression
 Inference and Intuition • 20.4 The Regression Table • 20.5 Multiple Regression Inference
 • 20.6 Confidence and Prediction Intervals • *20.7 Logistic Regression • *20.8 More
 About Regression

Review of Part V: Inference for Relationships 685

Parts I–V Cumulative Review Exercises 698

Appendices

A Answers A-1 • B Credits A-47 • C Indexes A-51
 • D Tables and Selected Formulas A-63