Preface 11

Index of Applications 21

PART I Exploring and Understanding Data

1 Stats Starts Here 27

1.1 What Is Statistics? ◆ **1.2** Data ◆ **1.3** Variables ◆ **1.4** Models

2 Displaying and Describing Data 43

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 90

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

4 Understanding and Comparing Distributions 121

4.1 Displays for Comparing Groups ◆ **4.2** Outliers ◆ **4.3** Timeplots: Order, Please! ◆ **4.4** Re-Expressing Data: A First Look

5 The Standard Deviation as a Ruler and the Normal Model 152

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 184

PART II Exploring Relationships Between Variables

6 Scatterplots, Association, and Correlation 193

6.1 Scatterplots ◆ **6.2** Correlation ◆ **6.3** Warning: Correlation $\neq$ Causation ◆ ***6.4** Straightening Scatterplots

7 Linear Regression 226

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 265

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 308

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 339

PART III Gathering Data

10 Sample Surveys 351

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 375

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 399

PART IV Randomness and Probability

12 From Randomness to Probability 405

12.1 Random Phenomena ◆ **12.2** Modeling Probability ◆ **12.3** Formal Probability

13 Probability Rules! 423

13.1 The General Addition Rule ◆ **13.2** Conditional Probability and the General Multiplication Rule ◆ **13.3** Independence ◆ **13.4** Picturing Probability: Tables, Venn Diagrams, and Trees ◆ **13.5** Reversing the Conditioning and Bayes' Rule

14 Random Variables 445

14.1 Center: The Expected Value ◆ **14.2** Spread: The Standard Deviation ◆ **14.3** Shifting and Combining Random Variables ◆ **14.4** Continuous Random Variables

15 Probability Models 468

15.1 Bernoulli Trials ◆ **15.2** The Geometric Model ◆ **15.3** The Binomial Model ◆ **15.4** Approximating the Binomial with a Normal Model ◆ ***15.5** The Continuity Correction ◆ **15.6** The Poisson Model ◆ **15.7** Other Continuous Random Variables: The Uniform and the Exponential

Review of Part IV: Randomness and Probability 495

PART V Inference for One Parameter

16 Sampling Distribution Models and Confidence Intervals for Proportions 501

16.1 The Sampling Distribution Model for a Proportion ◆ **16.2** When Does the Normal Model Work? Assumptions and Conditions ◆ **16.3** A Confidence Interval for a Proportion ◆ **16.4** Interpreting Confidence Intervals: What Does 95% Confidence Really Mean? ◆ **16.5** Margin of Error: Certainty vs. Precision ◆ ***16.6** Choosing the Sample Size

17 Confidence Intervals for Means 532

17.1 The Central Limit Theorem ◆ **17.2** A Confidence Interval for the Mean ◆ **17.3** Interpreting Confidence Intervals ◆ ***17.4** Picking Our Interval up by Our Bootstraps ◆ **17.5** Thoughts About Confidence Intervals

18 Testing Hypotheses 563

18.1 Hypotheses ◆ **18.2** P-Values ◆ **18.3** The Reasoning of Hypothesis Testing ◆ **18.4** A Hypothesis Test for the Mean ◆ **18.5** Intervals and Tests ◆ **18.6** P-Values and Decisions: What to Tell About a Hypothesis Test

19 More About Tests and Intervals 598

19.1 Interpreting P-Values ◆ **19.2** Alpha Levels and Critical Values ◆ **19.3** Practical vs. Statistical Significance ◆ **19.4** Errors

Review of Part V: Inference for One Parameter 623

PART VI Inference for Relationships

20 Comparing Groups 630

20.1 A Confidence Interval for the Difference Between Two Proportions ◆ **20.2** Assumptions and Conditions for Comparing Proportions ◆ **20.3** The Two-Sample *z*-Test: Testing the Difference Between Proportions ◆ **20.4** A Confidence Interval for the Difference Between Two Means ◆ **20.5** The Two-Sample *t*-Test: Testing for the Difference Between Two Means ◆ ***20.6** Randomization Tests and Confidence Intervals for Two Means ◆ ***20.7** Pooling ◆ ***20.8** The Standard Deviation of a Difference

21 Paired Samples and Blocks 675

21.1 Paired Data ◆ **21.2** The Paired *t*-Test ◆ **21.3** Confidence Intervals for Matched Pairs ◆ **21.4** Blocking

22 Comparing Counts 700

22.1 Goodness-of-Fit Tests ◆ **22.2** Chi-Square Test of Homogeneity ◆ **22.3** Examining the Residuals ◆ **22.4** Chi-Square Test of Independence

23 Inferences for Regression 732

23.1 The Regression Model ◆ **23.2** Assumptions and Conditions ◆ **23.3** Regression Inference and Intuition ◆ **23.4** The Regression Table ◆ **23.5** Multiple Regression Inference ◆ **23.6** Confidence and Prediction Intervals ◆ ***23.7** Logistic Regression ◆ ***23.8** More About Regression

Review of Part VI: Inference for Relationships 775

PART VII Inference When Variables Are Related

24 Multiple Regression Wisdom 788

24.1 Cleaning and Formatting Data ◆ **24.2** Diagnosing Regression Models: Looking at the Cases ◆ **24.3** Building Multiple Regression Models

25 Analysis of Variance 823

25.1 Testing Whether the Means of Several Groups Are Equal ◆ **25.2** The ANOVA Table ◆ **25.3** Assumptions and Conditions ◆ **25.4** Comparing Means ◆ **25.5** ANOVA on Observational Data

26 Multifactor Analysis of Variance 858

26.1 A Two-Factor ANOVA Model ◆ **26.2** Assumptions and Conditions ◆ **26.3** Interactions

27 Introduction to Statistical Learning and Data Science 893

27.1 Data Science and Big Data ◆ **27.2** The Data Mining Process ◆ **27.3** Data Mining Algorithms: A Sample ◆ **27.4** Models Built from Combining Other Models ◆ **27.5** Comparing Models ◆ **27.6** Summary

Review of Part VII: Inference When Variables Are Related 925

Cumulative Review Exercises 937

Appendixes

A Answers **943** ◆ **B** Credits **997** ◆ **C** Index **999** ◆ **D** Tables and Selected Formulas **1009**