

Contents

	Preface	12
	Index of Applications	27
Part I	Exploring and Collecting Data	
Chapter 1	Data and Decisions (H&M)	33
	1.1 Data, 35 • 1.2 The Role of Data in Decision Making, 37 • 1.3 Variable Types, 40	
	• 1.4 Data Sources: Where, How, and When, 42	
	Ethics in Action	45
	From Learning to Earning	46
	Tech Support: Entering Data	47
	Brief Case: Credit Card Bank	48
Chapter 2	Visualizing and Describing Categorical Data (Dalia Research)	53
	2.1 Summarizing a Categorical Variable, 54 • 2.2 Displaying a Categorical Variable, 56	
	• 2.3 Exploring Relationships Between Two Categorical Variables: Contingency Tables, 60 • 2.4 Segmented Bar Charts and Mosaic Plots, 62 • 2.5 Three Categorical Variables, 69 • 2.6 Simpson's Paradox, 71	
	Ethics in Action	73
	From Learning to Earning	74
	Tech Support: Displaying Categorical Data	75
	Brief Case: Credit Card Bank	78
Chapter 3	Describing, Displaying, and Visualizing Quantitative Data (AIG)	88
	3.1 Visualizing Quantitative Variables, 90 • 3.2 Shape, 92 • 3.3 Center, 94	
	• 3.4 Spread of the Distribution, 96 • 3.5 Shape, Center, and Spread—A Summary, 99 • 3.6 Standardizing Variables, 99 • 3.7 Five-Number Summary and Boxplots, 101 • 3.8 Comparing Groups, 104 • 3.9 Identifying Outliers, 107	
	• 3.10 Time Series Plots, 108 • *3.11 Transforming Skewed Data, 111	
	Ethics in Action	116
	From Learning to Earning	117
	Tech Support: Displaying and Summarizing Quantitative Variables	119
	Brief Case: Detecting the Housing Bubble	122
Chapter 4	Correlation and Linear Regression (Zillow.com)	137
	4.1 Looking at Scatterplots, 138 • 4.2 Assigning Roles to Variables in Scatterplots, 141	
	• 4.3 Understanding Correlation, 142 • 4.4 Lurking Variables and Causation, 147	
	• 4.5 The Linear Model, 148 • 4.6 Correlation and the Line, 149 • 4.7 Regression to the Mean, 152 • 4.8 Checking the Model, 153 • 4.9 Variation in the Model and R^2 , 156	
	• 4.10 Reality Check: Is the Regression Reasonable? 158 • 4.11 Nonlinear Relationships, 162 • *4.12 Multiple Regression—A Glimpse Ahead, 165	
	Ethics in Action	169
	From Learning to Earning	170
	Tech Support: Correlation and Regression	171
	Brief Case: Fuel Efficiency, Cost of Living, and Mutual Funds	174
	Case Study: Paralyzed Veterans of America	187
Part II	Modeling with Probability	
Chapter 5	Randomness and Probability (Credit Reports, the Fair Isaacs Corporation, and Equifax)	189
	5.1 Random Phenomena and Probability, 190 • 5.2 The Nonexistent Law of Averages, 192	
	• 5.3 Different Types of Probability, 193 • 5.4 Probability Rules, 195 • 5.5 Joint Probability and Contingency Tables, 200 • 5.6 Conditional Probability and the General Multiplication Rule, 201 • 5.7 Constructing Contingency Tables, 204 • 5.8 Probability Trees, 205 • *5.9 Reversing the Conditioning: Bayes' Rule, 207	
	Ethics in Action	209
	From Learning to Earning	209
	Tech Support: Generating Random Numbers	211
	Brief Case: Global Markets	212

Chapter 6	Random Variables and Probability Models (Metropolitan Life Insurance Company)	222
	6.1 Expected Value of a Random Variable, 223 • 6.2 Standard Deviation of a Random Variable, 226 • 6.3 Properties of Expected Values and Variances, 229 • 6.4 Bernoulli Trials, 233 • 6.5 Discrete Probability Models, 233	
	Ethics in Action	241
	From Learning to Earning	242
	Tech Support: Random Variables and Probability Models	243
	Brief Case: Investment Options	244
Chapter 7	The Normal and Other Continuous Distributions (The NYSE)	252
	7.1 The Standard Deviation as a Ruler, 253 • 7.2 The Normal Distribution, 255 • 7.3 Normal Probability Plots, 262 • 7.4 The Distribution of Sums of Normals, 263 • 7.5 The Normal Approximation for the Binomial, 266 • 7.6 Other Continuous Random Variables, 269	
	Ethics in Action	273
	From Learning to Earning	273
	Tech Support: Probability Calculations and Plots	274
	Brief Case: Price/Earnings and Stock Value	276
Part III Gathering Data		
Chapter 8	Data Sources: Observational Studies and Surveys (Roper Polls)	284
	8.1 Observational Studies and Found Data, 285 • 8.2 Sample Surveys, 287 • 8.3 Populations and Parameters, 291 • 8.4 Common Sampling Designs, 292 • 8.5 The Valid Survey, 297 • 8.6 How to Sample Badly, 299	
	Ethics in Action	302
	From Learning to Earning	302
	Tech Support	304
	Brief Case: Market Survey Research and The GfK Roper Reports Worldwide Survey	305
Chapter 9	Data Sources: Experiments (Capital One)	312
	9.1 Randomized, Comparative Experiments, 315 • 9.2 The Four Principles of Experimental Design, 316 • 9.3 Experimental Designs, 318 • 9.4 Issues in Experimental Design, 323 • 9.5 Displaying Data from Designed Experiments, 325	
	Ethics in Action	332
	From Learning to Earning	332
	Brief Case: Design a Multifactor Experiment	334
Part IV Inference for Decision Making		
Chapter 10	Sampling Distributions and Confidence Intervals for Proportions (Marketing Credit Cards: The MBNA Story)	342
	10.1 The Distribution of Sample Proportions, 343 • 10.2 A Confidence Interval for a Proportion, 348 • 10.3 Margin of Error: Certainty vs. Precision, 353 • 10.4 Choosing the Sample Size, 357	
	Ethics in Action	362
	From Learning to Earning	362
	Tech Support: Confidence Intervals for Proportions	364
	Brief Case: Has Gold Lost its Luster? and Forecasting Demand	365
	Case Study: Real Estate Simulation	375
Chapter 11	Confidence Intervals for Means (Guinness & Co.)	376
	11.1 The Central Limit Theorem, 377 • 11.2 The Sampling Distribution of the Mean, 381 • 11.3 How Sampling Distribution Models Work, 382 • 11.4 Gosset and the <i>t</i> -Distribution, 384 • 11.5 A Confidence Interval for Means, 386 • 11.6 Assumptions and Conditions, 388 • 11.7 Visualizing Confidence Intervals for the Mean, 395	
	Ethics in Action	400
	From Learning to Earning	400
	Tech Support: Confidence Intervals for Means	402
	Brief Case: Real Estate and Donor Profiles	403

Chapter 12	Testing Hypotheses (Casting Ingots)	413
	12.1 Hypotheses, 414 • 12.2 P-Values, 416 • 12.3 The Reasoning of Hypothesis Testing, 419 • 12.4 A Hypothesis Test for the Mean, 424 • 12.5 Intervals and Tests, 430 • 12.6 P-Values and Decisions: What to Tell About a Hypothesis Test, 435	
	Ethics in Action	438
	From Learning to Earning	438
	Tech Support: Hypothesis Tests	440
	Brief Case: Real Estate and Donor Profiles	443
Chapter 13	More About Tests and Intervals (Traveler’s Insurance)	451
	13.1 How to Think About P-Values, 453 • 13.2 Alpha Levels and Significance, 458 • 13.3 Critical Values, 460 • 13.4 Confidence Intervals and Hypothesis Tests, 461 • 13.5 Two Types of Errors, 464 • 13.6 Power, 466	
	Ethics in Action	470
	From Learning to Earning	470
	Brief Case: Confidence Intervals and Hypothesis Tests	471
Chapter 14	Comparing Two Means (Visa Global Organization)	479
	14.1 Comparing Two Means, 480 • 14.2 The Two-Sample <i>t</i> -Test, 483 • 14.3 Assumptions and Conditions, 484 • 14.4 A Confidence Interval for the Difference Between Two Means, 488 • 14.5 The Pooled <i>t</i> -Test, 490 • 14.6 Paired Data, 495 • 14.7 Paired <i>t</i> -Methods, 496	
	Ethics in Action	502
	From Learning to Earning	502
	Tech Support: Comparing Two Groups	504
	Brief Case: Real Estate and Consumer Spending Patterns (Data Analysis)	508
Chapter 15	Inference for Counts: Chi-Square Tests (SAC Capital)	525
	15.1 Goodness-of-Fit Tests, 527 • 15.2 Interpreting Chi-Square Values, 531 • 15.3 Examining the Residuals, 532 • 15.4 The Chi-Square Test of Homogeneity, 534 • 15.5 Comparing Two Proportions, 538 • 15.6 Chi-Square Test of Independence, 539	
	Ethics in Action	545
	From Learning to Earning	546
	Tech Support: Chi-Square	547
	Brief Case: Health Insurance and Loyalty Program	550
	Case Study: Investment Strategy Segmentation	562
Part V	Models for Decision Making	
Chapter 16	Inference for Regression (Nambé Mills)	563
	16.1 A Hypothesis Test and Confidence Interval for the Slope, 564 • 16.2 Assumptions and Conditions, 568 • 16.3 Standard Errors for Predicted Values, 574 • 16.4 Using Confidence and Prediction Intervals, 577	
	Ethics in Action	579
	From Learning to Earning	579
	Tech Support: Regression Analysis	581
	Brief Case: Frozen Pizza and Global Warming?	583
Chapter 17	Understanding Residuals (Kellogg’s)	597
	17.1 Examining Residuals for Groups, 598 • 17.2 Extrapolation and Prediction, 601 • 17.3 Unusual and Extraordinary Observations, 603 • 17.4 Working with Summary Values, 607 • 17.5 Autocorrelation, 608 • 17.6 Transforming (Re-expressing) Data, 610 • 17.7 The Ladder of Powers, 614	
	Ethics in Action	621
	From Learning to Earning	621
	Tech Support: Examining Residuals	622
	Brief Case: Gross Domestic Product and Energy Sources	624

Chapter 18	Multiple Regression (Zillow.com)	639
	18.1 The Multiple Regression Model, 641 • 18.2 Interpreting Multiple Regression Coefficients, 643 • 18.3 Assumptions and Conditions for the Multiple Regression Model, 645 • 18.4 Testing the Multiple Regression Model, 653 • 18.5 Adjusted R^2 and the F-statistic, 655 • *18.6 The Logistic Regression Model, 657	
	Ethics in Action	664
	From Learning to Earning	665
	Tech Support: Regression Analysis	666
	Brief Case: Golf Success	668
Chapter 19	Building Multiple Regression Models (Bolliger and Mabillard)	680
	19.1 Indicator (or Dummy) Variables, 682 • 19.2 Adjusting for Different Slopes—Interaction Terms, 686 • 19.3 Multiple Regression Diagnostics, 689 • 19.4 Building Regression Models, 695 • 19.5 Collinearity, 705 • 19.6 Quadratic Terms, 708	
	Ethics in Action	713
	From Learning to Earning	714
	Tech Support: Building Multiple Regression Models	715
	Brief Case: Building Models	717
Chapter 20	Time Series Analysis (Whole Foods Market®)	729
	20.1 What Is a Time Series? 731 • 20.2 Components of a Time Series, 731	
	• 20.3 Smoothing Methods, 734 • 20.4 Summarizing Forecast Error, 739	
	• 20.5 Autoregressive Models, 741 • 20.6 Multiple Regression–Based Models, 748	
	• 20.7 Choosing a Time Series Forecasting Method, 758 • 20.8 Interpreting Time Series Models: The Whole Foods Data Revisited, 759	
	Ethics in Action	760
	From Learning to Earning	760
	Tech Support: Time Series	763
	Brief Case: U.S. Trade with the European Union	763
	Case Study: Health Care Costs	777
Part VI	Analytics	
Chapter 21	Introduction to Big Data and Data Mining (Paralyzed Veterans of America)	778
	21.1 Data Mining and the Big Data Revolution, 779 • 21.2 The Data Mining Process, 783 • 21.3 Data Mining Algorithms: A Sample, 789 • 21.4 Models Built from Combining Other Models, 797 • 21.5 Comparing Models, 800 • 21.6 Summary, 806	
	Ethics in Action	807
	From Learning to Earning	807
Part VII	Online Topics	
Chapter 22	Quality Control (Sony)	22-1
	22.1 A Short History of Quality Control, 22-3 • 22.2 Control Charts for Individual Observations (Run Charts), 22-7 • 22.3 Control Charts for Measurements: \bar{X} and R Charts, 22-10 • 22.4 Actions for Out-of-Control Processes, 22-16 • 22.5 Control Charts for Attributes: p Charts and c Charts, 22-22 • 22.6 Philosophies of Quality Control, 22-25	
	Ethics in Action	22-27
	From Learning to Earning	22-27
	Tech Support: Quality Control Charts	22-29
	Brief Case: Laptop Touchpad Quality	22-30
Chapter 23	Nonparametric Methods (i4cp)	23-1
	23.1 Ranks, 23-2 • 23.2 The Wilcoxon Rank-Sum/Mann-Whitney Statistic, 23-3	
	• 23.3 Kruskal-Wallis Test, 23-7 • 23.4 Paired Data: The Wilcoxon Signed-Rank Test, 23-10 • *23.5 Friedman Test for a Randomized Block Design, 23-13 • 23.6 Kendall's Tau: Measuring Monotonicity, 23-14 • 23.7 Spearman's Rho, 23-15 • 23.8 When Should You Use Nonparametric Methods? 23-16	
	Ethics in Action	23-17
	From Learning to Earning	23-18
	Tech Support: Nonparametric Methods	23-19
	Brief Case: Real Estate Reconsidered	23-20

Chapter 24	Decision Making and Risk (Data Description, Inc.)	24-1
	<ul style="list-style-type: none"> 24.1 Actions, States of Nature, and Outcomes, 24-2 • 24.2 Payoff Tables and Decision Trees, 24-3 • 24.3 Minimizing Loss and Maximizing Gain, 24-4 • 24.4 The Expected Value of an Action, 24-5 • 24.5 Expected Value with Perfect Information, 24-7 • 24.6 Decisions Made with Sample Information, 24-7 • 24.7 Estimating Variation, 24-9 • 24.8 Sensitivity, 24-11 • 24.9 Simulation, 24-12 • 24.10 More Complex Decisions, 24-14 	
	Ethics in Action	24-14
	From Learning to Earning	24-15
	Brief Case: Texaco-Pennzoil and Insurance Services, Revisited	24-16
Chapter 25	Analysis of Experiments and Observational Studies	25-1
	<ul style="list-style-type: none"> 25.1 Analyzing a Design in One Factor—The One-Way Analysis of Variance, 25-2 • 25.2 Assumptions and Conditions for ANOVA, 25-6 • *25.3 Multiple Comparisons, 25-9 • 25.4 ANOVA on Observational Data, 25-11 • 25.5 Analysis of Multifactor Designs, 25-12 	
	From Learning to Earning	25-21
	Tech Support: Analysis of Variance	25-22
	Brief Case: Analyze your Multifactor Experiment	25-24
	Appendixes	811
	A. Answers	811
	B. Tables and Selected Formulas	867
	C. Credits	887
	Index	889