

Business Statistics

3rd Edition
Global Edition

Norean R. Sharpe
Georgetown University

Richard D. De Veaux
Williams College

Paul F. Velleman
Cornell University

With Contributions by David Bock

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Contents

	Preface	11
	Index of Applications	22
Part I	Exploring and Collecting Data	
Chapter 1	Data and Decisions (E-Commerce)	29
	1.1 What Are Data? 30 • 1.2 Variable Types, 34 • 1.3 Data Sources: Where, How, and When, 37	
	Ethics in Action	39
	Technology Help: Data	41
	Brief Case: Credit Card Bank	42
Chapter 2	Displaying and Describing Categorical Data (KEEN, Inc.)	47
	2.1 Summarizing a Categorical Variable, 48 • 2.2 Displaying a Categorical Variable, 49	
	2.3 Exploring Two Categorical Variables: Contingency Tables, 53 • 2.4 Segmented Bar Charts and Mosaic Plots, 57 • 2.5 Simpson's Paradox, 61	
	Ethics in Action	64
	Technology Help: Displaying Categorical Data	65
	Brief Case: Credit Card Bank	67
Chapter 3	Displaying and Describing Quantitative Data (AIG)	77
	3.1 Displaying Quantitative Variables, 78 • 3.2 Shape, 81 • 3.3 Center, 84	
	3.4 Spread of the Distribution, 86 • 3.5 Shape, Center, and Spread—A Summary, 88	
	3.6 Standardizing Variables, 88 • 3.7 Five-Number Summary and Boxplots, 90	
	3.8 Comparing Groups, 93 • 3.9 Identifying Outliers, 95 • 3.10 Time Series Plots, 97	
	*3.11 Transforming Skewed Data, 100	
	Ethics in Action	105
	Technology Help: Displaying and Summarizing Quantitative Variables	108
	Brief Cases: Detecting the Housing Bubble and Socio-Economic Data on States	110
Chapter 4	Correlation and Linear Regression (Amazon.com)	125
	4.1 Looking at Scatterplots, 126 • 4.2 Assigning Roles to Variables in Scatterplots, 129	
	4.3 Understanding Correlation, 130 • 4.4 Lurking Variables and Causation, 134	
	4.5 The Linear Model, 136 • 4.6 Correlation and the Line, 137 • 4.7 Regression to the Mean, 140 • 4.8 Checking the Model, 142 • 4.9 Variation in the Model and R^2 , 145	
	4.10 Reality Check: Is the Regression Reasonable? 147 • 4.11 Nonlinear Relationships, 149	
	Ethics in Action	154
	Technology Help: Correlation and Regression	157
	Brief Cases: Fuel Efficiency, Cost of Living, and Mutual Funds	159
	Case Study I: Paralyzed Veterans of America	172
Part II	Modeling with Probability	
Chapter 5	Randomness and Probability (Credit Reports and the Fair Isaacs Corporation)	175
	5.1 Random Phenomena and Probability, 176 • 5.2 The Nonexistent Law of Averages, 178 • 5.3 Different Types of Probability, 179 • 5.4 Probability Rules, 181 • 5.5 Joint Probability and Contingency Tables, 186 • 5.6 Conditional Probability, 187 • 5.7 Constructing Contingency Tables, 190 • 5.8 Probability Trees, 191 • *5.9 Reversing the Conditioning: Bayes' Rule, 193	
	Ethics in Action	195
	Technology Help: Generating Random Numbers	197
	Brief Case: Global Markets	198
Chapter 6	Random Variables and Probability Models (Metropolitan Life Insurance Company)	209
	6.1 Expected Value of a Random Variable, 210 • 6.2 Standard Deviation of a Random Variable, 212 • 6.3 Properties of Expected Values and Variances, 215 • 6.4 Bernoulli Trials, 219 • 6.5 Discrete Probability Models, 220	

	Ethics in Action	227
	Technology Help: Random Variables and Probability Models	229
	Brief Case: Investment Options	230
Chapter 7	The Normal and Other Continuous Distributions (The NYSE)	237
	7.1 The Standard Deviation as a Ruler, 238 • 7.2 The Normal Distribution, 240 • 7.3 Normal Probability Plots, 248 • 7.4 The Distribution of Sums of Normals, 249 • 7.5 The Normal Approximation for the Binomial, 253 • 7.6 Other Continuous Random Variables, 255	
	Ethics in Action	259
	Technology Help: Probability Calculations and Plots	260
	Brief Case: Price/Earnings and Stock Value	261
Chapter 8	Surveys and Sampling (Roper Polls)	271
	8.1 Three Ideas of Sampling, 272 • 8.2 Populations and Parameters, 276 • 8.3 Common Sampling Designs, 276 • 8.4 The Valid Survey, 282 • 8.5 How to Sample Badly, 284	
	Ethics in Action	287
	Technology Help: Random Sampling	289
	Brief Cases: Market Survey Research and The GfK Roper Reports Worldwide Survey	290
Chapter 9	Sampling Distributions and Confidence Intervals for Proportions (Marketing Credit Cards: The MBNA Story)	299
	9.1 The Distribution of Sample Proportions, 300 • 9.2 A Confidence Interval for a Proportion, 305 • 9.3 Margin of Error: Certainty vs. Precision, 310 • 9.4 Choosing the Sample Size, 314	
	Ethics in Action	319
	Technology Help: Confidence Intervals for Proportions	321
	Brief Cases: Has Gold Lost Its Luster? and Forecasting Demand	322
	Case Study II: Real Estate Simulation	332
Part III	Inference for Decision Making	
Chapter 10	Testing Hypotheses about Proportions (Dow Jones Industrial Average)	333
	10.1 Hypotheses, 334 • 10.2 A Trial as a Hypothesis Test, 336 • 10.3 P-Values, 337	
	10.4 The Reasoning of Hypothesis Testing, 339 • 10.5 Alternative Hypotheses, 341	
	10.6 P-Values and Decisions: What to Tell about a Hypothesis Test, 344	
	Ethics in Action	348
	Technology Help: Hypothesis Tests	350
	Brief Cases: Metal Production and Loyalty Program	351
Chapter 11	Confidence Intervals and Hypothesis Tests for Means (Guinness & Co.)	359
	11.1 The Central Limit Theorem, 360 • 11.2 The Sampling Distribution of the Mean, 363	
	11.3 How Sampling Distribution Models Work, 365 • 11.4 Gosset and the t -Distribution, 366 • 11.5 A Confidence Interval for Means, 368 • 11.6 Assumptions and Conditions, 370 • 11.7 Testing Hypotheses about Means—the One-Sample t -Test, 376	
	Ethics in Action	381
	Technology Help: Inference for Means	383
	Brief Cases: Real Estate and Donor Profiles	385
Chapter 12	More about Tests and Intervals (Traveler's Insurance)	395
	12.1 How to Think about P-Values, 397 • 12.2 Alpha Levels and Significance, 402	
	12.3 Critical Values, 404 • 12.4 Confidence Intervals and Hypothesis Tests, 405	
	12.5 Two Types of Errors, 408 • 12.6 Power, 410	
	Ethics in Action	414
	Brief Case: Confidence Intervals and Hypothesis Tests	415
Chapter 13	Comparing Two Means (Visa Global Organization)	423
	13.1 Comparing Two Means, 424 • 13.2 The Two-Sample t -Test, 427 • 13.3 Assumptions and Conditions, 427 • 13.4 A Confidence Interval for the Difference Between Two Means, 431 • 13.5 The Pooled t -Test, 434 • 13.6 Paired Data, 439 • 13.7 Paired t -Methods, 440	

	Ethics in Action	446
	Technology Help: Comparing Two Groups	448
	Brief Cases: Real Estate and Consumer Spending Patterns (Data Analysis)	451
Chapter 14	Inference for Counts: Chi-Square Tests (SAC Capital)	469
	14.1 Goodness-of-Fit Tests, 471 • 14.2 Interpreting Chi-Square Values, 476	
	14.3 Examining the Residuals, 477 • 14.4 The Chi-Square Test of Homogeneity, 478	
	14.5 Comparing Two Proportions, 482 • 14.6 Chi-Square Test of Independence, 484	
	Ethics in Action	490
	Technology Help: Chi-Square	492
	Brief Cases: Health Insurance and Loyalty Program	494
	Case Study III: Investment Strategy Segmentation	506
Part IV	Models for Decision Making	
Chapter 15	Inference for Regression (Nambé Mills)	507
	15.1 A Hypothesis Test and Confidence Interval for the Slope, 508 • 15.2 Assumptions and Conditions, 512 • 15.3 Standard Errors for Predicted Values, 518 • 15.4 Using Confidence and Prediction Intervals, 520	
	Ethics in Action	523
	Technology Help: Regression Analysis	525
	Brief Cases: Frozen Pizza and Global Warming?	526
Chapter 16	Understanding Residuals (Kellogg's)	541
	16.1 Examining Residuals for Groups, 542 • 16.2 Extrapolation and Prediction, 545	
	16.3 Unusual and Extraordinary Observations, 548 • 16.4 Working with Summary Values, 551 • 16.5 Autocorrelation, 552 • 16.6 Transforming (Re-expressing) Data, 554	
	16.7 The Ladder of Powers, 558	
	Ethics in Action	565
	Technology Help: Examining Residuals	566
	Brief Cases: Gross Domestic Product and Energy Sources	567
Chapter 17	Multiple Regression (Zillow.com)	583
	17.1 The Multiple Regression Model, 585 • 17.2 Interpreting Multiple Regression Coefficients, 587 • 17.3 Assumptions and Conditions for the Multiple Regression Model, 589 • 17.4 Testing the Multiple Regression Model, 597 • 17.5 Adjusted R^2 and the F-statistic, 599 • *17.6 The Logistic Regression Model, 601	
	Ethics in Action	608
	Technology Help: Regression Analysis	610
	Brief Case: Golf Success	612
Chapter 18	Building Multiple Regression Models (Bolliger and Mabillard)	625
	18.1 Indicator (or Dummy) Variables, 628 • 18.2 Adjusting for Different Slopes—Interaction Terms, 632 • 18.3 Multiple Regression Diagnostics, 634 • 18.4 Building Regression Models, 640 • 18.5 Collinearity, 648 • 18.6 Quadratic Terms, 651	
	Ethics in Action	657
	Technology Help: Building Multiple Regression Models	659
	Brief Case: Building Models	660
Chapter 19	Time Series Analysis (Whole Foods Market®)	671
	19.1 What Is a Time Series? 672 • 19.2 Components of a Time Series, 673	
	19.3 Smoothing Methods, 676 • 19.4 Summarizing Forecast Error, 681	
	19.5 Autoregressive Models, 683 • 19.6 Multiple Regression-based Models, 689	
	19.7 Choosing a Time Series Forecasting Method, 700 • 19.8 Interpreting Time Series Models: The Whole Foods Data Revisited, 701	
	Ethics in Action	702
	Technology Help: Time Series	705
	Brief Cases: U.S. Trade with the European Union and Tiffany & Co.	705
	Case Study IV: Health Care Costs	718

Part V	Selected Topics in Decision Making	
Chapter 20	Design and Analysis of Experiments and Observational Studies (Capital One)	721
	20.1 Observational Studies, 722 • 20.2 Randomized, Comparative Experiments, 724	
	20.3 The Four Principles of Experimental Design, 725 • 20.4 Experimental Designs, 727	
	20.5 Issues in Experimental Design, 731 • 20.6 Analyzing a Design in One Factor—	
	The One-Way Analysis of Variance, 733 • 20.7 Assumptions and Conditions for ANOVA, 737	
	*20.8 Multiple Comparisons, 740 • 20.9 ANOVA on Observational Data, 742	
	20.10 Analysis of Multifactor Designs, 743	
	Ethics in Action	753
	Technology Help: Analysis of Variance	757
	Brief Case: Design a Multifactor Experiment	758
Chapter 21	Quality Control (Sony)	771
	21.1 A Short History of Quality Control, 772 • 21.2 Control Charts for Individual	
	Observations (Run Charts), 776 • 21.3 Control Charts for Measurements: \bar{X} and R	
	Charts, 780 • 21.4 Actions for Out-of-Control Processes, 786 • 21.5 Control Charts for	
	Attributes: p Charts and c Charts, 792 • 21.6 Philosophies of Quality Control, 795	
	Ethics in Action	796
	Technology Help: Quality Control Charts	798
	Brief Case: Laptop Touchpad Quality	799
Chapter 22	Nonparametric Methods (i4cp)	807
	22.1 Ranks, 808 • 22.2 The Wilcoxon Rank-Sum/Mann-Whitney Statistic, 809	
	22.3 Kruskal-Wallis Test, 813 • 22.4 Paired Data: The Wilcoxon Signed-Rank Test, 816	
	*22.5 Friedman Test for a Randomized Block Design, 819 • 22.6 Kendall's Tau:	
	Measuring Monotonicity, 820 • 22.7 Spearman's Rho, 821 • 22.8 When Should You Use	
	Nonparametric Methods? 822	
	Ethics in Action	823
	Technology Help: Nonparametric Methods	825
	Brief Case: Real Estate Reconsidered	826
Chapter 23	Decision Making and Risk (Data Description, Inc.)	835
	23.1 Actions, States of Nature, and Outcomes, 836 • 23.2 Payoff Tables and Decision Trees,	
	837 • 23.3 Minimizing Loss and Maximizing Gain, 838 • 23.4 The Expected Value of an	
	Action, 839 • 23.5 Expected Value with Perfect Information, 840 • 23.6 Decisions Made	
	with Sample Information, 841 • 23.7 Estimating Variation, 843 • 23.8 Sensitivity, 845	
	23.9 Simulation, 846 • 23.10 More Complex Decisions, 848	
	Ethics in Action	848
	Brief Cases: Texaco-Pennzoil and Insurance Services, Revisited	850
Chapter 24	Introduction to Data Mining (Paralyzed Veterans of America)	857
	24.1 The Big Data Revolution, 858 • 24.2 Direct Marketing, 859 • 24.3 The Goals of	
	Data Mining, 861 • 24.4 Data Mining Myths, 862 • 24.5 Successful Data Mining, 863	
	24.6 Data Mining Problems, 865 • 24.7 Data Mining Algorithms, 865 • 24.8 The Data	
	Mining Process, 869 • 24.9 Summary, 871	
	Ethics in Action	872
	Case Study V: Marketing Experiment	874
	Appendixes	A-1
	A. Answers	A-1
	B. Tables and Selected Formulas	A-55
	C. Photo Acknowledgments	A-74
	Subject Index	I-1