

BUSINESS

for CONTEMPORARY DECISION MAKING

STATISTICS

SIXTH EDITION

KEN BLACK

6TH EDITION

Business Statistics

For Contemporary Decision Making

6TH EDITION

Business Statistics

For Contemporary Decision Making

Ken Black

University of Houston—Clear Lake



WILEY

John Wiley & Sons, Inc.

Vice President & Publisher *George Hoffman*
Acquisitions Editor *Franny Kelly*
Assistant Editor *Maria Guarascio*
Executive Marketing Manager *Amy Scholz*
Editorial Assistant *Emily McGee*
Production Manager *Dorothy Sinclair*
Senior Production Editor *Sandra Dumas*
Senior Designer *Madelyn Lesure*
Executive Media Editor *Allison Morris*
Photo Department Manager *Hilary Newman*
Production Management Services *Aptara*
Associate Media Editor *Elena Santa Maria*

This book was typeset in 10/12 Minion at Aptara®, Inc. and printed and bound by R. R. Donnelley/ Jefferson City. The cover was printed by R. R. Donnelley/Jefferson City.

The paper in this book was manufactured by a mill whose forest management programs include sustained yield harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

This book is printed on acid-free paper. ∞

Copyright © 2010, 2008, 2006, 2004 by John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at www.wiley.com/go/returnlabel. Outside of the United States, please contact your local representative.

Black, Ken
Business Statistics: For Contemporary Decision Making, Sixth Edition
ISBN 13 978-0470-40901-5
ISBN 13 978-0470-55667-2

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

BRIEF CONTENTS

UNIT I

INTRODUCTION

- 1 Introduction to Statistics 2
- 2 Charts and Graphs 16
- 3 Descriptive Statistics 46
- 4 Probability 92

UNIT II

DISTRIBUTIONS AND SAMPLING

- 5 Discrete Distributions 136
- 6 Continuous Distributions 178
- 7 Sampling and Sampling Distributions 216

UNIT III

MAKING INFERENCES ABOUT POPULATION PARAMETERS

- 8 Statistical Inference: Estimation for Single Populations 250
- 9 Statistical Inference: Hypothesis Testing for Single Populations 288
- 10 Statistical Inferences About Two Populations 342
- 11 Analysis of Variance and Design of Experiments 402

UNIT IV

REGRESSION ANALYSIS AND FORECASTING

- 12 Simple Regression Analysis and Correlation 464
- 13 Multiple Regression Analysis 516
- 14 Building Multiple Regression Models 546
- 15 Time-Series Forecasting and Index Numbers 588

UNIT V

NONPARAMETRIC STATISTICS AND QUALITY

- 16 Analysis of Categorical Data 644
- 17 Nonparametric Statistics 670
- 18 Statistical Quality Control 720

APPENDICES

- A Tables 765
- B Answers to Selected Odd-Numbered Quantitative Problems 805

GLOSSARY 815

INDEX 825

The following materials are available at www.wiley.com/college/black

- 19 Decision Analysis C19-2
- Supplement 1** Summation Notation S1-1
- Supplement 2** Derivation of Simple Regression Formulas for Slope and y Intercept S2-1
- Supplement 3** Advanced Exponential Smoothing S3-1

CONTENTS

Preface	xvii
About the Author	xxvii

UNIT I

INTRODUCTION

1 Introduction to Statistics 2

Decision Dilemma: Statistics Describe the State of Business in India's Countryside 3

- 1.1 **Statistics in Business** 4
- 1.2 **Basic Statistical Concepts** 5
- 1.3 **Data Measurement** 7
 - Nominal Level 7
 - Ordinal Level 8
 - Interval Level 8
 - Ratio Level 9
 - Comparison of the Four Levels of Data 9
 - Statistical Analysis Using the Computer: Excel and Minitab 11

Summary 12

Key Terms 12

Supplementary Problems 12

Analyzing the Databases 13

Case: DiGiorno Pizza: Introducing a Frozen Pizza to Compete with Carry-Out 15

2 Charts and Graphs 16

Decision Dilemma: Energy Consumption Around the World 17

- 2.1 **Frequency Distributions** 18
 - Class Midpoint 18
 - Relative Frequency 18
 - Cumulative Frequency 19
- 2.2 **Quantitative Data Graphs** 21
 - Histograms 21
 - Using Histograms to Get an Initial Overview of the Data* 23
 - Frequency Polygons 23
 - Ogives 24
 - Dot Plots 25
 - Stem-and-Leaf Plots 25
- 2.3 **Qualitative Data Graphs** 27
 - Pie Charts 27
 - Bar Graphs 28
 - Pareto Charts 30
- 2.4 **Graphical Depiction of Two-Variable Numerical Data: Scatter Plots** 33

Summary 36

Key Terms 36

Supplementary Problems 36

Analyzing the Databases 40

Case: Soap Companies Do Battle 40

Using the Computer 41

3 Descriptive Statistics 46

Decision Dilemma: Laundry Statistics 47

3.1 Measures of Central Tendency: Ungrouped Data 47

Mode 48

Median 48

Mean 49

Percentiles 51

Steps in Determining the Location of a Percentile 51

Quartiles 52

3.2 Measures of Variability: Ungrouped Data 55

Range 55

Interquartile Range 56

Mean Absolute Deviation, Variance, and Standard Deviation 57

Mean Absolute Deviation 58

Variance 59

Standard Deviation 60

Meaning of Standard Deviation 60

Empirical Rule 60

Chebyshev's Theorem 62

Population Versus Sample Variance and Standard Deviation 63

Computational Formulas for Variance and Standard Deviation 64

z Scores 66

Coefficient of Variation 67

3.3 Measures of Central Tendency and Variability: Grouped Data 70

Measures of Central Tendency 70

Mean 70

Median 71

Mode 72

Measures of Variability 72

3.4 Measures of Shape 77

Skewness 77

Skewness and the Relationship of the Mean, Median, and Mode 77

Coefficient of Skewness 77

Kurtosis 78

Box-and-Whisker Plots 78

3.5 Descriptive Statistics on the Computer 81

Summary 83

Key Terms 84

Formulas 84

Supplementary Problems 85

Analyzing the Databases 89

Case: Coca-Cola Goes Small in Russia 89

Using the Computer 91

4 Probability 92

Decision Dilemma: Equity of the Sexes in the Workplace 93

4.1 Introduction to Probability 94

4.2 Methods of Assigning Probabilities 94

Classical Method of Assigning Probabilities 94

Relative Frequency of Occurrence 95

Subjective Probability 96

4.3 Structure of Probability 96

Experiment 96

Event 96

Elementary Events 96

Sample Space 97

Unions and Intersections 97

Mutually Exclusive Events 98

Independent Events 98

Collectively Exhaustive Events 99

Complementary Events 99

Counting the Possibilities 99

The mn Counting Rule 99

Sampling from a Population with Replacement 100

Combinations: Sampling from a Population Without Replacement 100

4.4 Marginal, Union, Joint, and Conditional Probabilities 101

4.5 Addition Laws 103

Probability Matrices 104

Complement of a Union 107

Special Law of Addition 108

4.6 Multiplication Laws 111

General Law of Multiplication 111

Special Law of Multiplication 113

4.7 Conditional Probability 116

Independent Events 119

4.8 Revision of Probabilities: Bayes' Rule 123

Summary 128

Key Terms 128

Formulas 129

Supplementary Problems 129

Analyzing the Databases 132

Case: Colgate-Palmolive Makes a "Total" Effort 133

UNIT II

DISTRIBUTIONS AND SAMPLING

5 Discrete Distributions 136

Decision Dilemma: Life with a Cell Phone 137

5.1 Discrete Versus Continuous Distributions 138

5.2 Describing a Discrete Distribution 139

Mean, Variance, and Standard Deviation of Discrete Distributions 140

Mean or Expected Value 140

Variance and Standard Deviation of a Discrete Distribution 140

5.3 Binomial Distribution 143

Solving a Binomial Problem 144

Using the Binomial Table 147

Using the Computer to Produce a Binomial Distribution 148

Mean and Standard Deviation of a Binomial Distribution 149

Graphing Binomial Distributions 150

5.4 Poisson Distribution 154

Working Poisson Problems by Formula 156

Using the Poisson Tables 157

Mean and Standard Deviation of a Poisson Distribution 158

Graphing Poisson Distributions 159

Using the Computer to Generate Poisson Distributions 159

Approximating Binomial Problems by the Poisson Distribution 160

5.5 Hypergeometric Distribution 164

Using the Computer to Solve for Hypergeometric Distribution Probabilities 166

Summary 169

Key Terms 169

Formulas 170

Supplementary Problems 170

Analyzing the Databases 175

Case: Kodak Transitions Well into the Digital Camera Market 175

Using the Computer 176

6 Continuous Distributions 178

Decision Dilemma: The Cost of Human Resources 179

6.1 The Uniform Distribution 179

Determining Probabilities in a Uniform Distribution 181

Using the Computer to Solve for Uniform Distribution Probabilities 183

6.2 Normal Distribution 184

History of the Normal Distribution 185

	Probability Density Function of the Normal Distribution	185
	Standardized Normal Distribution	186
	Solving Normal Curve Problems	187
	Using the Computer to Solve for Normal Distribution Probabilities	194
6.3	Using the Normal Curve to Approximate Binomial Distribution Problems	196
	Correcting for Continuity	198
6.4	Exponential Distribution	202
	Probabilities of the Exponential Distribution	203
	Using the Computer to Determine Exponential Distribution Probabilities	205
	Summary	207
	Key Terms	208
	Formulas	208
	Supplementary Problems	208
	Analyzing the Databases	212
	<i>Case: Mercedes Goes After Younger Buyers</i>	212
	Using the Computer	213
7	Sampling and Sampling Distributions	216
	<i>Decision Dilemma: What Is the Attitude of Maquiladora Workers?</i> 217	
7.1	Sampling	217
	Reasons for Sampling	218
	Reasons for Taking a Census	218
	Frame	219
	Random Versus Nonrandom Sampling	219
	Random Sampling Techniques	220
	Simple Random Sampling	220
	Stratified Random Sampling	221
	Systematic Sampling	222
	Cluster (or Area) Sampling	223
	Nonrandom Sampling	224
	Convenience Sampling	224
	Judgment Sampling	225
	Quota Sampling	225
	Snowball Sampling	226
	Sampling Error	226
	Nonsampling Errors	226
7.2	Sampling Distribution of \bar{x}	228
	Sampling from a Finite Population	235
7.3	Sampling Distribution of \hat{p}	237
	Summary	241
	Key Terms	242
	Formulas	242
	Supplementary Problems	242

	Analyzing the Databases	245
	<i>Case: Shell Attempts to Return to Premiere Status</i>	245
	Using the Computer	246

UNIT III

MAKING INFERENCES ABOUT POPULATION PARAMETERS

8 Statistical Inference: Estimation for Single Populations 250

Decision Dilemma: Compensation for Purchasing Managers 251

8.1	Estimating the Population Mean Using the z Statistic (σ Known)	253
	Finite Correction Factor	256
	Estimating the Population Mean Using the z Statistic when the Sample Size Is Small	257
	Using the Computer to Construct z Confidence Intervals for the Mean	258
8.2	Estimating the Population Mean Using the t Statistic (σ Unknown)	260
	The t Distribution	261
	Robustness	261
	Characteristics of the t Distribution	261
	Reading the t Distribution Table	261
	Confidence Intervals to Estimate the Population Mean Using the t Statistic	262
	Using the Computer to Construct t Confidence Intervals for the Mean	264
8.3	Estimating the Population Proportion	267
	Using the Computer to Construct Confidence Intervals of the Population Proportion	269
8.4	Estimating the Population Variance	271
8.5	Estimating Sample Size	275
	Sample Size when Estimating μ	275
	Determining Sample Size when Estimating p	277

	Summary	280
	Key Terms	280
	Formulas	280
	Supplementary Problems	281
	Analyzing the Databases	284
	<i>Case: Thermatrix</i>	284
	Using the Computer	285

9 Statistical Inference: Hypothesis Testing for Single Populations 288

Decision Dilemma: Word-of-Mouth Business Referrals and Influentials 289

9.1	Introduction to Hypothesis Testing	290
	Types of Hypotheses	291
	Research Hypotheses	291

Statistical Hypotheses 292
 Substantive Hypotheses 294
 Using the HTAB System to Test Hypotheses 295
 Rejection and Nonrejection Regions 297
 Type I and Type II Errors 298

9.2 Testing Hypotheses About a Population Mean Using the z Statistic (σ Known) 299
 Testing the Mean with a Finite Population 301
 Using the p -Value to Test Hypotheses 302
 Using the Critical Value Method to Test Hypotheses 303
 Using the Computer to Test Hypotheses About a Population Mean Using the z Statistic 306

9.3 Testing Hypotheses About a Population Mean Using the t Statistic (σ Unknown) 308
 Using the Computer to Test Hypotheses About a Population Mean Using the t Test 312

9.4 Testing Hypotheses About a Proportion 315
 Using the Computer to Test Hypotheses About a Population Proportion 319

9.5 Testing Hypotheses About a Variance 321

9.6 Solving for Type II Errors 324
 Some Observations About Type II Errors 329
 Operating Characteristic and Power Curves 329
 Effect of Increasing Sample Size on the Rejection Limits 331

Summary 334
 Key Terms 335
 Formulas 335
 Supplementary Problems 335
 Analyzing the Databases 338
Case: Frito-Lay Targets the Hispanic Market 339
 Using the Computer 340

10 Statistical Inferences About Two Populations 342

Decision Dilemma: Online Shopping 343

10.1 Hypothesis Testing and Confidence Intervals About the Difference in Two Means Using the z Statistic (Population Variances Known) 346
 Hypothesis Testing 347
 Confidence Intervals 350
 Using the Computer to Test Hypotheses About the Difference in Two Population Means Using the z Test 352

10.2 Hypothesis Testing and Confidence Intervals About the Difference in Two Means: Independent Samples and Population Variances Unknown 355
 Hypothesis Testing 355
 Using the Computer to Test Hypotheses and Construct Confidence Intervals About the

Difference in Two Population Means Using the t Test 357
 Confidence Intervals 360

10.3 Statistical Inferences for Two Related Populations 365
 Hypothesis Testing 365
 Using the Computer to Make Statistical Inferences About Two Related Populations 367
 Confidence Intervals 370

10.4 Statistical Inferences About Two Population Proportions, $p_1 - p_2$ 375
 Hypothesis Testing 375
 Confidence Intervals 379
 Using the Computer to Analyze the Difference in Two Proportions 380

10.5 Testing Hypotheses About Two Population Variances 382
 Using the Computer to Test Hypotheses About Two Population Variances 386

Summary 391
 Key Terms 391
 Formulas 391
 Supplementary Problems 392
 Analyzing the Databases 397
Case: Seitz Corporation: Producing Quality Gear-Driven and Linear-Motion Products 397
 Using the Computer 398

11 Analysis of Variance and Design of Experiments 402

Decision Dilemma: Job and Career Satisfaction of Foreign Self-Initiated Expatriates 403

11.1 Introduction to Design of Experiments 404

11.2 The Completely Randomized Design (One-Way ANOVA) 406
 One-Way Analysis of Variance 407
 Reading the F Distribution Table 411
 Using the Computer for One-Way ANOVA 411
 Comparison of F and t Values 412

11.3 Multiple Comparison Tests 418
 Tukey's Honestly Significant Difference (HSD) Test: The Case of Equal Sample Sizes 418
 Using the Computer to Do Multiple Comparisons 420
 Tukey-Kramer Procedure: The Case of Unequal Sample Sizes 422

11.4 The Randomized Block Design 426
 Using the Computer to Analyze Randomized Block Designs 430

11.5 A Factorial Design (Two-Way ANOVA) 436
 Advantages of the Factorial Design 436
 Factorial Designs with Two Treatments 437
 Applications 437

Statistically Testing the Factorial Design 438
 Interaction 439
 Using a Computer to Do a Two-Way ANOVA 444
 Summary 453
 Key Terms 454
 Formulas 454
 Supplementary Problems 455
 Analyzing the Databases 458
Case: The Clarkson Company: A Division of Tyco International 459
 Using the Computer 460

UNIT IV

REGRESSION ANALYSIS AND FORECASTING

12 Simple Regression Analysis and Correlation 464

Decision Dilemma: Predicting International Hourly Wages by the Price of a Big Mac 465

12.1 Correlation 466
12.2 Introduction to Simple Regression Analysis 469
12.3 Determining the Equation of the Regression Line 470
12.4 Residual Analysis 477
 Using Residuals to Test the Assumptions of the Regression Model 479
 Using the Computer for Residual Analysis 480
12.5 Standard Error of the Estimate 484
12.6 Coefficient of Determination 487
 Relationship Between r and r^2 489
12.7 Hypothesis Tests for the Slope of the Regression Model and Testing the Overall Model 489
 Testing the Slope 489
 Testing the Overall Model 493
12.8 Estimation 494
 Confidence Intervals to Estimate the Conditional Mean of y : $\mu_{y|x}$ 494
 Prediction Intervals to Estimate a Single Value of y 495
12.9 Using Regression to Develop a Forecasting Trend Line 498
 Determining the Equation of the Trend Line 499
 Forecasting Using the Equation of the Trend Line 500
 Alternate Coding for Time Periods 501
12.10 Interpreting the Output 504
 Summary 508
 Key Terms 509
 Formulas 509

Supplementary Problems 509
 Analyzing the Databases 513
Case: Delta Wire Uses Training as a Weapon 513
 Using the Computer 515

13 Multiple Regression Analysis 516

Decision Dilemma: Are You Going to Hate Your New Job? 517

13.1 The Multiple Regression Model 518
 Multiple Regression Model with Two Independent Variables (First Order) 519
 Determining the Multiple Regression Equation 520
 A Multiple Regression Model 520
13.2 Significance Tests of the Regression Model and Its Coefficients 525
 Testing the Overall Model 525
 Significance Tests of the Regression Coefficients 527
13.3 Residuals, Standard Error of the Estimate, and R^2 530
 Residuals 530
 SSE and Standard Error of the Estimate 531
 Coefficient of Multiple Determination (R^2) 532
 Adjusted R^2 533
13.4 Interpreting Multiple Regression Computer Output 535
 A Reexamination of the Multiple Regression Output 535

Summary 539
 Key Terms 540
 Formulas 540
 Supplementary Problems 540
 Analyzing the Databases 543
Case: Starbucks Introduces Debit Card 543
 Using the Computer 544

14 Building Multiple Regression Models 546

Decision Dilemma: Determining Compensation for CEOs 547

14.1 Nonlinear Models: Mathematical Transformation 548
 Polynomial Regression 548
 Tukey's Ladder of Transformations 551
 Regression Models with Interaction 552
 Model Transformation 554
14.2 Indicator (Dummy) Variables 560
14.3 Model-Building: Search Procedures 566
 Search Procedures 568
All Possible Regressions 568
Stepwise Regression 568

Forward Selection 572

Backward Elimination 572

14.4 Multicollinearity 576

Summary 580

Key Terms 581

Formulas 581

Supplementary Problems 581

Analyzing the Databases 584

Case: Virginia Semiconductor 585

Using the Computer 586

15 Time-Series Forecasting and Index Numbers 588

Decision Dilemma: Forecasting Air Pollution 589

15.1 Introduction to Forecasting 590

Time-Series Components 590

The Measurement of Forecasting Error 591
Error 591

Mean Absolute Deviation (MAD) 591

Mean Square Error (MSE) 592

15.2 Smoothing Techniques 594

Naïve Forecasting Models 594

Averaging Models 595

Simple Averages 595

Moving Averages 595

Weighted Moving Averages 597

Exponential Smoothing 599

15.3 Trend Analysis 604

Linear Regression Trend Analysis 604

Regression Trend Analysis Using
Quadratic Models 606

Holt's Two-Parameter Exponential Smoothing
Method 609

15.4 Seasonal Effects 611

Decomposition 611

Finding Seasonal Effects with the Computer 614

Winters' Three-Parameter Exponential Smoothing
Method 614

15.5 Autocorrelation and Autoregression 616

Autocorrelation 616

Ways to Overcome the Autocorrelation
Problem 619

Addition of Independent Variables 619

Transforming Variables 620

Autoregression 620

15.6 Index Numbers 623

Simple Index Numbers 624

Unweighted Aggregate Price Index
Numbers 624

Weighted Aggregate Price Index Numbers 625

Laspeyres Price Index 626

Paasche Price Index 627

Summary 632

Key Terms 633

Formulas 633

Supplementary Problems 633

Analyzing the Databases 638

Case: Debourgh Manufacturing Company 639

Using the Computer 640

UNIT V

NONPARAMETRIC STATISTICS AND QUALITY

16 Analysis of Categorical Data 644

Decision Dilemma: Selecting Suppliers in the Electronics Industry 645

16.1 Chi-Square Goodness-of-Fit Test 646

Testing a Population Proportion by Using the
Chi-Square Goodness-of-Fit Test as an
Alternative Technique to the z Test 652

16.2 Contingency Analysis: Chi-Square Test of Independence 656

Summary 666

Key Terms 666

Formulas 666

Supplementary Problems 666

Analyzing the Databases 668

Case: Foot Locker in the Shoe Mix 668

Using the Computer 669

17 Nonparametric Statistics 670

Decision Dilemma: How Is the Doughnut Business? 671

17.1 Runs Test 673

Small-Sample Runs Test 674

Large-Sample Runs Test 675

17.2 Mann-Whitney U Test 678

Small-Sample Case 678

Large-Sample Case 680

17.3 Wilcoxon Matched-Pairs Signed Rank Test 686

Small-Sample Case ($n \leq 15$) 686

Large-Sample Case ($n > 15$) 688

17.4 Kruskal-Wallis Test 694

17.5 Friedman Test 699

17.6 Spearman's Rank Correlation 705

Summary 710

Key Terms 711

Formulas 711

Supplementary Problems	711
Analyzing the Databases	716
<i>Case: Schwinn</i>	717
Using the Computer	718

18 Statistical Quality Control 720

Decision Dilemma: Italy's Piaggio Makes a Comeback 721

18.1 Introduction to Quality Control	722
What Is Quality Control?	722
Total Quality Management	723
<i>Deming's 14 Points</i>	724
Quality Gurus	725
Six Sigma	725
Design for Six Sigma	727
Lean Manufacturing	727
Some Important Quality Concepts	727
<i>Benchmarking</i>	728
<i>Just-in-Time Inventory Systems</i>	728
<i>Reengineering</i>	729
<i>Failure Mode and Effects Analysis</i>	730
<i>Poka-Yoke</i>	731
<i>Quality Circles and Six Sigma Teams</i>	731
18.2 Process Analysis	733
Flowcharts	733
Pareto Analysis	734
Cause-and-Effect (Fishbone) Diagrams	735
Control Charts	736
Check Sheets or Checklists	737
Histogram	738
Scatter Chart or Scatter Diagram	738
18.3 Control Charts	739
Variation	740
Types of Control Charts	740
\bar{x} Chart	740
<i>R Charts</i>	744
<i>p Charts</i>	745
<i>c Charts</i>	748
Interpreting Control Charts	750
Summary	756
Key Terms	757
Formulas	757
Supplementary Problems	758
Analyzing the Databases	761
<i>Case: Robotron-ELOTHERM</i>	762
Using the Computer	763

APPENDICES

A Tables	765
B Answers to Selected Odd-Numbered Quantitative Problems	805

GLOSSARY 815

INDEX 825

The following materials are available at www.wiley.com/college/black

19 Decision Analysis C19-2

Decision Dilemma: Decision Making at the CEO Level C19-3

19.1 The Decision Table and Decision Making Under Certainty	C19-4
Decision Table	C19-4
Decision Making Under Certainty	C19-5
19.2 Decision Making Under Uncertainty	C19-6
Maximax Criterion	C19-6
Maximin Criterion	C19-6
Hurwicz Criterion	C19-7
Minimax Regret	C19-9
19.3 Decision Making Under Risk	C19-14
Decision Trees	C19-14
Expected Monetary Value (EMV)	C19-14
Expected Value of Perfect Information	C19-18
Utility	C19-19
19.4 Revising Probabilities in Light of Sample Information	C19-22
Expected Value of Sample Information	C19-25

Summary C19-32

Key Terms C19-33

Formula C19-33

Supplementary Problems C19-33

Analyzing the Databases C19-36

Case: Fletcher-Terry: On the Cutting Edge C19-36

SUPPLEMENTS

1 Summation Notation	S1-1
2 Derivation of Simple Regression Formulas for Slope and y Intercept	S2-1
3 Advanced Exponential Smoothing	S3-1
Exponential Smoothing with Trend Effects: Holt's Method	S3-1
Exponential Smoothing with Both Trend and Seasonality: Winter's Method	S3-2
Some Practice Problems	S3-5

The sixth edition of *Business Statistics for Contemporary Decision Making* continues the rich tradition of using clear and complete, student-friendly pedagogy to present and explain business statistics topics. With the sixth edition, the author and Wiley continue to expand the vast ancillary resources available through WileyPLUS with which to complement the text in helping instructors effectively deliver this subject matter and assisting students in their learning. The resources available to both the instructor and the student through WileyPLUS have greatly expanded since the fifth edition was launched; and because of this, an effort has been made in the sixth edition to more fully integrate the text with WileyPLUS.

In the spirit of continuous quality improvement, several changes have been made in the text to help students construct their knowledge of the big picture of statistics, provide assistance as needed, and afford more opportunities to practice statistical skills. In the fifth edition, the 19 chapters were organized into four units to facilitate student understanding of the bigger view of statistics. In the sixth edition, these same 19 chapters have been organized into five units so that chapters could be grouped into smaller clusters. The nonparametric and the analysis of categorical data chapters have been moved further toward the back of the text so that the regression chapters can be presented earlier. The decision trees that were introduced in the fifth edition to provide the student with a taxonomy of inferential techniques have been improved and expanded in the sixth edition. Nonparametric inferential techniques have been separated from other inferential techniques and given their own decision tree. This has simplified the decision trees for parametric techniques and made the decision trees easier for students to decipher. Further integration of the text with WileyPLUS is addressed through icons that are used throughout the text to designate to the reader that a WileyPLUS feature is available for assistance on a particular topic. The number of databases associated with the text has been expanded from seven to nine, and one of the fifth edition databases has been replaced, thereby bringing the total of new databases in the sixth edition to three.

All of the features of the fifth edition have been retained, updated, and changed as needed to reflect today's business world in the sixth edition. One Decision Dilemma has been replaced, and nine new Statistics in Business Today features have been added. In the sixth edition, as with the fifth edition, there are 17 high-quality video tutorials with the author explaining key difficult topics and demonstrating how to work problems from challenging sections of the text.

This edition is written and designed for a two-semester introductory undergraduate business statistics course or an MBA-level introductory course. In addition, with 19 chapters, the sixth edition lends itself nicely to adaptation for a one-semester introductory business statistics course. The text is written with the assumption that the student has a college algebra mathematical background. No calculus is used in the presentation of material in the text.

An underlying philosophical approach to the text is that every statistical tool presented in the book has some business application. While the text contains statistical rigor, it is written so that the student can readily see that the proper application of statistics in the business world goes hand-in-hand with good decision making. In this edition, statistics are presented as a means for converting data into useful information that can be used to assist the business decision maker in making more thoughtful, information-based decisions. Thus, the text presents business statistics as “value added” tools in the process of converting data into useful information.



CHANGES FOR THE SIXTH EDITION

Units and Chapters

The fifth edition presented 19 chapters organized into four units. The purpose of the unit organization was to locate chapters with similar topics together, thereby increasing the likelihood that students are better able to grasp the bigger picture of statistics. As an

example, in the fifth edition, Unit II was about distributions and sampling. In this unit of four chapters, the students were introduced to eight probability distributions and to methods of sampling that are used as the basis for techniques presented later in the text.

In the sixth edition, the 18 chapters are organized into five units. The first two units of the sixth edition are the same as those used in the fifth edition. For several reasons, Unit III, Making Inferences About Population Parameters, which contained six chapters of statistical techniques for estimating population parameters and testing population parameters in the fifth edition, has been reduced from six to four chapters in the sixth edition. This makes Unit III less formidable for students to digest, simplifies tree diagrams, and moves two chapters that are less likely to be covered in many courses to later in the text. In the sixth edition, Unit IV, now named Regression Analysis and Forecasting, consists of the same four chapters as it did in the fifth edition. In addition, these four chapters have been moved up two chapters in the sixth edition. Thus, the chapter on simple regression analysis, a chapter that is covered in most courses, is now Chapter 12 instead of Chapter 14. This organization will make it easier for instructors to get to simple regression material without having to skip many chapters.

Topical Changes

Sections and topics from the fifth edition remain virtually unchanged in the sixth edition, with a few exceptions. Correlation analysis has been moved from Section 3.5 in the fifth edition to Section 12.1 in the sixth edition. With this organization, the student begins the chapter (12) on simple regression analysis by studying scatter plots and correlation. Thus, the student is able to see visually what it means for variables to be related and to begin to imagine what it would be like to fit a line through the data. In addition, students are introduced to the r statistic as a forerunner of r^2 , and they can see how the five-column analysis used to mechanically solve for r is similar to that used in solving for the equation of the regression line.

In Chapter 2, Charts and Graphs, Section 2.2 of the fifth edition, has been expanded and reorganized into two sections, Quantitative Data Graphs and Qualitative Data Graphs. In addition, a treatment of dot plots has been added to Chapter 2 as an additional quantitative data graph. Dot plots are simple to construct and easy to understand and are especially useful when analyzing small- and medium-size databases. Their importance in visually depicting business data is growing.

Upon request by text users, presentation of the median of grouped data has been added to Chapter 3, Descriptive Statistics.

Acceptance sampling, the last section of Chapter 18 of the fifth edition, has been deleted in the sixth edition. Because acceptance sampling is based on inspection and is generally only used to accept or reject a batch, it has limited usefulness in the present world of Six Sigma, lean manufacturing, and quality improvement. In place of acceptance sampling in the sixth edition, Chapter 18, Statistical Quality Control, additional information on quality gurus, quality movements, and quality concepts, has been added.

Integration of Text and WileyPLUS

WileyPLUS, with its rich resources, has been a powerful partner to this text in delivering and facilitating business statistics education for several years. Many instructors have discovered that WileyPLUS can greatly enhance the effectiveness of their business statistics course, and they use WileyPLUS hand-in-hand with the text. With this in mind, the sixth edition further integrates the text and WileyPLUS by using icons to represent such WileyPLUS features as interactive applets, videos by the author, demonstration problems, Decision Dilemma, Decision Dilemma Solved, flash cards, and databases showing exactly where each one corresponds to text topics. In this way, students are reminded in the text when there is a WileyPLUS feature available to augment their learning.

Tree Diagram of Inferential Techniques

To assist the student in sorting out the plethora of confidence intervals and hypothesis testing techniques presented in the text, tree diagrams are presented at the beginning of Unit III and Chapters 8, 9, 10, and 17. The tree diagram at the beginning of Unit III displays virtually

all of the inferential techniques presented in Chapters 8–10 so that the student can construct a view of the “forest for the trees” and determine how each technique plugs into the whole. Then at the beginning of each of these three chapters, an additional tree diagram is presented to display the branch of the tree that applies to techniques in that particular chapter. Chapter 17 includes a tree diagram for just the nonparametric statistics presented in that chapter. In the fifth edition, all of these techniques were shown on one tree diagram; and because it was determined that this made the diagram less useful and perhaps overwhelming, in the sixth edition, the nonparametric branches are placed in a separate diagram.

In determining which technique to use, there are several key questions that a student should consider. Listed here are some of the key questions (displayed in a box in the Unit III introduction) that delineate what students should ask themselves in determining the appropriate inferential technique for a particular analysis: Does the problem call for estimation (using a confidence interval) or testing (using a hypothesis test)? How many samples are being analyzed? Are you analyzing means, proportions, or variances? If means are being analyzed, is (are) the variance(s) known or not? If means from two samples are being analyzed, are the samples independent or related? If three or more samples are being analyzed, are there one or two independent variables and is there a blocking variable?

Decision Dilemma and the Decision Dilemma Solved

The popular Decision Dilemma feature included in previous editions of the text has been retained in the sixth edition along with the In Response feature, which has been renamed as Decision Dilemma Solved. The Decision Dilemmas are real business vignettes that open each chapter and set the tone for the chapter by presenting a business dilemma and asking a number of managerial or statistical questions, the solutions to which require the use of techniques presented in the chapter. The Decision Dilemma Solved feature discusses and answers the managerial and statistical questions posed in the Decision Dilemma using techniques from the chapter, thus bringing closure to the chapter. In the sixth edition, all decision dilemmas have been updated and revised. Solutions given in the Decision Dilemma Solved features have been revised for new data and for new versions of computer output. In addition, one new Decision Dilemma has been added in the sixth edition in Chapter 10. The title of this Decision Dilemma is “Online Shopping,” a current and timely topic in the business world. In this Decision Dilemma, the results of surveys by the Pew Internet/American Life Project of 2400 American adults and a Nielsen survey of over 26,000 Internet users across the globe are presented in addition to a Gallup household survey of 1043 adults and a survey of 7000 people in Europe conducted by the European Interactive Advertising Association. Some of the findings of these studies include 875 million consumers around the world have shopped online, the market for online shopping has increased by 40% in the past 2 years, and European shoppers spend an average of €750 shopping online over a 6-month period. In the Decision Dilemma, presented at the opening of the chapter, students are asked to consider some managerial and statistical questions that are later answered in the Decision Dilemma Solved feature at the end of the chapter. An example of such a question, associated with this new Decision Dilemma is this:

One study reported that the average amount spent by online American shoppers in the past 30 days is \$123 at specialty stores and \$121 at department stores. These figures are relatively close to each other and were derived from sample information. Suppose a researcher wants to test to determine if there is actually any significant difference in the average amount spent by online American shoppers in the past 30 days at specialty stores vs. department stores. How does she go about conducting such a test?

Statistics in Business Today

The sixth edition includes one or two Statistics in Business Today features in every chapter. This feature presents a real-life example of how the statistics presented in that chapter apply in the business world today. There are nine new Statistics in Business Today features in the sixth edition, which have been added for timeliness and relevance to today’s students,

and others have been revised and updated. The nine new Statistics in Business Today features are “Cellular Phone Use in Japan,” “Recycling Statistics,” “Business Travel,” “Newspaper Advertising Reading Habits of Canadians,” “Plastic Bags vs. Bringing Your Own in Japan,” “Teleworking Facts,” “Sampling Canadian Manufacturers,” “Canadian Grocery Shopping Statistics,” and “Rising Cost of Healthcare in the U.S.” As an example, from “Canadian Grocery Shopping Statistics,” Canadians take a mean of 37 stock-up trips per year, spending an average of 44 minutes in the store. They take a mean of 76 quick trips per year and average of 18 minutes in the store. On average, Canadians spend four times more money on a stock-up trip than on a quick trip. Twenty-three percent often buy items that are not on their list but catch their eye, 28% often go to a store to buy an item that is on sale, 24% often switch to another check out lane to get out faster, and 45% often bring their own bag.

New Problems

Every problem in the fifth edition has been examined for timeliness, appropriateness, and logic before inclusion in the sixth edition. Those that fell short were replaced or rewritten. While the total number of problems in the text is 950, a concerted effort has been made to include only problems that make a significant contribution to the learning process. Thirty new problems have been added to the sixth edition, replacing problems that have become less effective or relevant. Over one-third of the new problems are in Chapter 3, Descriptive Statistics, where it is especially important for the student to analyze up-to-date business situations and data. All other problems in text have been examined for currency, and many problems have revised with updated data.

All demonstration problems and example problems were thoroughly reviewed and edited for effectiveness. A demonstration problem is an extra example containing both a problem and its solution and is used as an additional pedagogical tool to supplement explanations and examples in the chapters. Virtually all example and demonstration problems in the sixth edition are business oriented and contain the most current data available.

As with the previous edition, problems are located at the end of most sections in the chapters. A significant number of additional problems are provided at the end of each chapter in the Supplementary Problems. The Supplementary Problems are “scrambled”—problems using the various techniques in the chapter are mixed—so that students can test themselves on their ability to discriminate and differentiate ideas and concepts.

New Databases

Associated with the sixth edition are nine databases, three of which are new to this edition. One new database is the 12-year Gasoline database, which includes monthly gasoline prices, the OPEC spot price each month, monthly U.S. finished motor gasoline production, and monthly U.S. natural gas well head prices over 12 years. A second new database is the Consumer Food database, which contains data on annual household income, non-mortgage household debt, geographic region, and location for 200 households. The third new database is a U.S. and International Stock Market database with 60 months of actual stock market data from the Dow Jones Industrial Average, the NASDAQ, Standard and Poor’s, Japan NIKKEI 225, Hong Kong Hang Seng, United Kingdom FTSE 100, and Mexico’s IPC. This new International Stock Market database replaced the old Stock Market database that was in the fifth edition.



VIDEOTAPE TUTORIALS BY KEN BLACK

An exciting feature of the sixth edition package that will impact the effectiveness of student learning in business statistics and significantly enhance the presentation of course material is the series of videotape tutorials by Ken Black. With the advent of online business statistics courses, increasingly large class sizes, and the number of commuter students who have

very limited access to educational resources on business statistics, it is often difficult for students to get the learning assistance that they need to bridge the gap between theory and application on their own. There are now 17 videotaped tutorial sessions on key difficult topics in business statistics delivered by Ken Black and available for all adopters on WileyPLUS. In addition, these tutorials can easily be uploaded for classroom usage to augment lectures and enrich classroom presentations. Each session is around 9 minutes in length. The 17 tutorials are:

1. Chapter 3: Computing Variance and Standard Deviation
2. Chapter 3: Understanding and Using the Empirical Rule
3. Chapter 4: Constructing and Solving Probability Matrices
4. Chapter 4: Solving Probability Word Problems
5. Chapter 5: Solving Binomial Distribution Problems, Part I
6. Chapter 5: Solving Binomial Distribution Problems, Part II
7. Chapter 6: Solving Problems Using the Normal Curve
8. Chapter 8: Confidence Intervals
9. Chapter 8: Determining Which Inferential Technique to Use, Part I, Confidence Intervals
10. Chapter 9: Hypothesis Testing Using the z Statistic
11. Chapter 9: Establishing Hypotheses
12. Chapter 9: Understanding p -Values
13. Chapter 9: Type I and Type II Errors
14. Chapter 9: Two-Tailed Tests
15. Chapter 9: Determining Which Inferential Technique to Use, Part II, Hypothesis Tests
16. Chapter 12: Testing the Regression Model I—Predicted Values, Residuals, and Sum of Squares of Error
17. Chapter 12: Testing the Regression Model II—Standard Error of the Estimate and r^2



FEATURES AND BENEFITS

Each chapter of the sixth edition contains sections called Learning Objectives, a Decision Dilemma, Demonstration Problems, Section Problems, Statistics in Business Today, Decision Dilemma Solved, Chapter Summary, Key Terms, Formulas, Ethical Considerations, Supplementary Problems, Analyzing the Databases, Case, Using the Computer, and Computer Output from both Excel 2007 and Minitab Release 15.

- **Learning Objectives.** Each chapter begins with a statement of the chapter's main learning objectives. This statement gives the reader a list of key topics that will be discussed and the goals to be achieved from studying the chapter.
- **Decision Dilemma.** At the beginning of each chapter, a short case describes a real company or business situation in which managerial and statistical questions are raised. In most Decision Dilemmas, actual data are given and the student is asked to consider how the data can be analyzed to answer the questions.
- **Demonstration Problems.** Virtually every section of every chapter in the sixth edition contains demonstration problems. A demonstration problem contains both an example problem and its solution, and is used as an additional pedagogical tool to supplement explanations and examples.
- **Section Problems.** There are over 950 problems in the text. Problems for practice are found at the end of almost every section of the text. Most problems utilize real data gathered from a plethora of sources. Included here are a few brief excerpts from some of the real-life problems in the text: “*The Wall Street Journal* reported that 40% of all workers say they would change jobs for ‘slightly higher pay.’ In

addition, 88% of companies say that there is a shortage of qualified job candidates.” “In a study by Peter D. Hart Research Associates for the Nasdaq Stock Market, it was determined that 20% of all stock investors are retired people. In addition, 40% of all U.S. adults have invested in mutual funds.” “A survey conducted for the Northwestern National Life Insurance Company revealed that 70% of American workers say job stress caused frequent health problems.” “According to Padgett Business Services, 20% of all small-business owners say the most important advice for starting a business is to prepare for long hours and hard work. Twenty-five percent say the most important advice is to have good financing ready.”

- **Statistics in Business Today.** Every chapter in the sixth edition contains at least one Statistics in Business Today feature. These focus boxes contain an interesting application of how techniques of that particular chapter are used in the business world today. They are usually based on real companies, surveys, or published research.
- **Decision Dilemma Solved.** Situated at the end of the chapter, the Decision Dilemma Solved feature addresses the managerial and statistical questions raised in the Decision Dilemma. Data given in the Decision Dilemma are analyzed computationally and by computer using techniques presented in the chapter. Answers to the managerial and statistical questions raised in the Decision Dilemma are arrived at by applying chapter concepts, thus bringing closure to the chapter.
- **Chapter Summary.** Each chapter concludes with a summary of the important concepts, ideas, and techniques of the chapter. This feature can serve as a preview of the chapter as well as a chapter review.
- **Key Terms.** Important terms are bolded and their definitions italicized throughout the text as they are discussed. At the end of the chapter, a list of the key terms from the chapter is presented. In addition, these terms appear with their definitions in an end-of-book glossary.
- **Formulas.** Important formulas in the text are highlighted to make it easy for a reader to locate them. At the end of the chapter, most of the chapter’s formulas are listed together as a handy reference.
- **Ethical Considerations.** Each chapter contains an Ethical Considerations feature that is very timely, given the serious breach of ethics and lack of moral leadership of some business executives in recent years. With the abundance of statistical data and analysis, there is considerable potential for the misuse of statistics in business dealings. The important Ethical Considerations feature underscores this potential misuse by discussing such topics as lying with statistics, failing to meet statistical assumptions, and failing to include pertinent information for decision makers. Through this feature, instructors can begin to integrate the topic of ethics with applications of business statistics. Here are a few excerpts from Ethical Considerations features: “It is unprofessional and unethical to draw cause-and-effect conclusions just because two variables are correlated.” “The business researcher needs to conduct the experiment in an environment such that as many concomitant variables are controlled as possible. To the extent that this is not done, the researcher has an ethical responsibility to report that fact in the findings.” “The reader is warned that the value λ is assumed to be constant in a Poisson distribution experiment. Business researchers may produce spurious results if the value of λ is used throughout a study; but because the study is conducted during different time periods, the value of λ is actually changing.” “In describing a body of data to an audience, it is best to use whatever statistical measures it takes to present a ‘full’ picture of the data. By limiting the descriptive measures used, the business researcher may give the audience only part of the picture and skew the way the receiver understands the data.”
- **Supplementary Problems.** At the end of each chapter is an extensive set of additional problems. The Supplementary Problems are divided into three groups: Calculating the Statistics, which are strictly computational problems; Testing Your Understanding, which are problems for application and understanding; and

Interpreting the Output, which are problems that require the interpretation and analysis of software output.

- **Analyzing the Databases.** There are nine major databases located on the student companion Web site that accompanies the sixth edition. The end-of-chapter Analyzing the Databases section contains several questions/problems that require the application of techniques from the chapter to data in the variables of the databases. It is assumed that most of these questions/problems will be solved using a computer.
- **Case.** Each end-of-chapter case is based on a real company. These cases give the student an opportunity to use statistical concepts and techniques presented in the chapter to solve a business dilemma. Some cases feature very large companies—such as Shell Oil, Coca-Cola, or Colgate Palmolive. Others pertain to small businesses—such as Thermatrix, Delta Wire, or DeBourgh—that have overcome obstacles to survive and thrive. Most cases include raw data for analysis and questions that encourage the student to use several of the techniques presented in the chapter. In many cases, the student must analyze software output in order to reach conclusions or make decisions.
- **Using the Computer.** The Using the Computer section contains directions for producing the Excel 2007 and Minitab Release 15 software output presented in the chapter. It is assumed that students have a general understanding of a Microsoft Windows environment. Directions include specifics about menu bars, drop-down menus, and dialog boxes. Not every detail of every dialog box is discussed; the intent is to provide enough information for students to produce the same statistical output analyzed and discussed in the chapter. The sixth edition has a strong focus on both Excel and Minitab software packages. More than 250 Excel 2007 or Minitab Release 15 computer-generated outputs are displayed.



WILEYPLUS

WileyPLUS is a powerful online tool that provides instructors and students with an integrated suite of teaching and learning resources, including an online version of the text, in one easy-to-use Web site. To learn more about WileyPLUS, and view a demo, please visit www.wiley.com/college/WileyPLUS.

WileyPLUS Tools for Instructors

WileyPLUS enables you to:

- Assign automatically graded homework, practice, and quizzes from the end of chapter and test bank.
- Track your students' progress in an instructor's grade book.
- Access all teaching and learning resources, including an online version of the text, and student and instructor supplements, in one easy-to-use Web site. These include full color PowerPoint slides, teaching videos, case files, and answers and animations.
- Create class presentations using Wiley-provided resources, with the ability to customize and add your own materials.

WileyPLUS Resources for Students Within WileyPLUS

In WileyPLUS, students will find various helpful tools, such as an ebook, the student study manual, videos with tutorials by the author, applets, Decision Dilemma and Decision Dilemma Solved animations, learning activities, flash cards for key terms, demonstration problems, databases in both Excel and Minitab, case data in both Excel and Minitab, and problem data in both Excel and Minitab.

- **Ebook.** The complete text is available on WileyPLUS with learning links to various features and tools to assist students in their learning.
- **Videos.** There are 17 videos of the author explaining concepts and demonstrating how to work problems for some of the more difficult topics.
- **Applets.** Statistical applets are available, affording students the opportunity to learn concepts by iteratively experimenting with various values of statistics and parameters and observing the outcomes.
- **Learning Activities.** There are numerous learning activities to help the student better understand concepts and key terms. These activities have been developed to make learning fun, enjoyable, and challenging.
- **Data Sets.** Virtually all problems in the text along with the case problems and the databases are available to students in both Excel and Minitab format.
- **Animations.** To aid students in understanding complex interactions, selected figures from the text that involve dynamic activity have been animated using Flash technology. Students can download these animated figures and run them to improve their understanding of dynamic processes.
- **Flash Cards.** Key terms will be available to students in a flash card format along with their definition.
- **Student Study Guide.** Complete answers to all odd-numbered questions.
- **Demo Problems.** Step-by-step solved problems for each chapter.



ANCILLARY TEACHING AND LEARNING MATERIALS

www.wiley.com/college/black

Students' Companion Site

The student companion Web site contains:

- All databases in both Excel and Minitab formats for easy access and use.
- Excel and Minitab files of data from all text problems and all cases. Instructors and students now have the option of analyzing any of the data sets using the computer.
- Full and complete version of Chapter 19, Decision Analysis, in PDF format. This allows an instructor the option of covering the material in this chapter in the normal manner, while keeping the text manageable in size and length.
- A section on Advanced Exponential Smoothing Techniques (from Chapter 17), which offers the instructor an opportunity to delve deeper into exponential smoothing if so desired, and derivation of the slope and intercept formulas from Chapter 12.
- A tutorial on summation theory.

Instructor's Resource Kit

All instructor ancillaries are provided on the Instructor Resource Site. Included in this convenient format are:

- **Instructor's Manual.** Prepared by Ken Black, this manual contains the worked out solutions to virtually all problems in the text. In addition, this manual contains chapter objectives, chapter outlines, chapter teaching strategies, and solutions to the cases.
- **PowerPoint Presentation Slides.** The presentation slides, prepared by Lloyd Jaisingh of Morehead State University, contain graphics to help instructors create stimulating lectures. The PowerPoint slides may be adapted using PowerPoint software to facilitate classroom use.
- **Test Bank.** Prepared by Ranga Ramasesh of Texas Christian University, the Test Bank includes multiple-choice questions for each chapter. The Test Bank is provided in Microsoft Word format.



ACKNOWLEDGMENTS

John Wiley & Sons and I would like to thank the reviewers and advisors who cared enough and took the time to provide us with their excellent insights and advice, which was used to reshape and mold the text into the sixth edition. These colleagues include: Lihui Bai, Valparaiso University; Pam Boger, Ohio University; Parag Dhumal, Winona State University; Bruce Ketler, Grove City College; Peter Lenk, University of Michigan—Ann Arbor; Robert Montague, Southern Adventist University; Robert Patterson, Penn State University—Behrend; Victor Prybutok, University of North Texas; Nikolai Pulchritudoff, California State University—Los Angeles; Ahmad Saranjam, Northeastern University; Vijay Shah, West Virginia University; Daniel Shimshak, University of Massachusetts—Boston; Cheryl Staley, Lake Land College—Mattoon; Debbie Stiver, University of Nevada—Reno; Minghe Sun, University of Texas—San Antonio.

As always, I wish to recognize my colleagues at the University of Houston—Clear Lake for their continued interest and support of this project. In particular, I want to thank William Staples, president; Carl Stockton, provost; and Ted Cummings, dean of the School of Business for their personal interest in the book and their administrative support.

There are several people within the John Wiley & Sons publishing group whom I would like to thank for their invaluable assistance on this project. These include: Franny Kelly, Maria Guarascio, Allie Morris, Lisé Johnson, and Diane Mars.

I want to express a special appreciation to my wife of 41 years, Carolyn, who is the love of my life and continues to provide both professional and personal support in my writing. Thanks also to my daughters, Wendi and Caycee, for their patience, love, and support.

—Ken Black

Ken Black is currently professor of decision sciences in the School of Business at the University of Houston–Clear Lake. Born in Cambridge, Massachusetts, and raised in Missouri, he earned a bachelor's degree in mathematics from Graceland University, a master's degree in math education from the University of Texas at El Paso, a Ph.D. in business administration in management science, and a Ph.D. in educational research from the University of North Texas.

Since joining the faculty of UHCL in 1979, Professor Black has taught all levels of statistics courses, forecasting, management science, market research, and production/operations management. In 2005, he was awarded the President's Distinguished Teaching Award for the university. He has published over 20 journal articles and 20 professional papers, as well as two textbooks: *Business Statistics: An Introductory Course* and *Business Statistics for Contemporary Decision Making*. Black has consulted for many different companies, including Aetna, the city of Houston, NYLCare, AT&T, Johnson Space Center, Southwest Information Resources, Connect Corporation, and Eagle Engineering.

Ken Black and his wife, Carolyn, have two daughters, Caycee and Wendi. His hobbies include playing the guitar, reading, traveling, and running.

UNIT I

INTRODUCTION

The study of business statistics is important, valuable, and interesting. However, because it involves a new language of terms, symbols, logic, and application of mathematics, it can be at times overwhelming. For many students, this text is their first and only introduction to business statistics, which instructors often teach as a “survey course.” That is, the student is presented with an overview of the subject, including a waterfront of techniques, concepts, and formulas. It can be overwhelming! One of the main difficulties in studying business statistics in this way is to be able to see “the forest for the trees,” that is, sorting out the myriad of topics so they make sense. With this in mind, the 18 chapters of this text have been organized into five units with each unit containing chapters that tend to present similar material. At the beginning of each unit, there is an introduction presenting the overlying themes to those chapters.

Unit I is titled Introduction because the four chapters (1–4) contained therein “introduce” the study of business statistics. In Chapter 1, students will learn what statistics are, the concepts of descriptive and inferential statistics, and levels of data measurement. In Chapter 2, students will see how raw data can be organized using various graphical and tabular techniques to facilitate their use in making better business decisions. Chapter 3 introduces some essential and basic statistics that will be used to both summarize data and as tools for techniques introduced later in the text. There will also be discussion of distribution shapes. In Chapter 4, the basic laws of probability are presented. The notion of probability underlies virtually every business statistics topic, distribution, and technique, thereby making it important to acquire an appreciation and understanding of probability. In Unit I, the first four chapters, we are developing “building blocks” that will enable students to understand and apply statistical concepts to analyze data that can assist present and future business managers in making better decisions.

Introduction to Statistics

LEARNING OBJECTIVES

The primary objective of Chapter 1 is to introduce you to the world of statistics, thereby enabling you to:

1. List quantitative and graphical examples of statistics within a business context
2. Define important statistical terms, including population, sample, and parameter, as they relate to descriptive and inferential statistics
3. Compare the four different levels of data: nominal, ordinal, interval, and ratio

Keren Su/The Image Bank/Getty Images





Statistics Describe the State of Business in India's Countryside

India is the second largest country in the world, with more than a billion people. Nearly three-quarters of the people live in rural areas scattered

about the countryside in 6,000,000 villages. In fact, it may be said that 1 in every 10 people in the world live in rural India. Presently, the population in rural India can be described as poor and semi-illiterate. With an annual per capita income of less than \$1 (U.S.) per day, rural India accounts for only about one-third of total national product sales. Less than 50% of households in rural India have electricity, and many of the roads are not paved. The annual per capita consumption for toothpaste is only 30 grams per person in rural India compared to 160 grams in urban India and 400 grams in the United States.

However, in addition to the impressive size of the population, there are other compelling reasons for companies to market their goods and services to rural India. The market of rural India has been growing at five times the rate of the urban India market. There is increasing agricultural productivity, leading to growth in disposable income, and there is a reduction in the gap between the tastes of urban and rural customers. The literacy level is increasing, and people are becoming more conscious about their lifestyles and opportunities for a better life.

Nearly two-thirds of all middle-income households in India are in rural areas, with the number of middle- and high-income households in rural India expected to grow from 80 million to 111 million over the next three years. More than one-third of all rural households now have a main source of income other than farming. Virtually every home has a radio, almost 20% have a television, and more than 30% have at least one bank account.

In the early 1990s, toothpaste consumption in rural India doubled, and the consumption of shampoo increased fourfold. Recently, other products have

done well in rural India, accounting for nearly one-half of all of the country's sales of televisions, fans, bicycles, bath soap, and other products. According to MART, a New Delhi-based research organization, rural India buys 46% of all soft drinks and 49% of motorcycles sold in India. In one year alone, the market for Coca-Cola in rural India grew by 37%, accounting for 80% of new Coke drinkers in India. Because of such factors, many U.S. and Indian firms, such as Microsoft, General Electric, Kellogg's, Colgate-Palmolive, Hindustan Lever, Godrej, Nirma Chemical Works, and Mahotra Marketing, have entered the rural Indian market with enthusiasm. Marketing to rural customers often involves building categories by persuading them to try and adopt products that they may not have used before. Rural India is a huge, relatively untapped market for businesses. However, entering such a market is not without risks and obstacles. The dilemma facing companies is whether to enter this marketplace and, if so, to what extent and how.

Managerial and Statistical Questions

1. Are the statistics presented in this report exact figures or estimates?
2. How and where could the researchers have gathered such data?
3. In measuring the potential of the rural India marketplace, what other statistics could have been gathered?
4. What levels of data measurement are represented by data on rural India?
5. How can managers use these and other statistics to make better decisions about entering this marketplace?

Source: Adapted from Raja Ramachandran, "Understanding the Market Environment of India," *Business Horizons*, January 2000; P. Balakrishna and B. Sidharth, "Selling in Rural India," *The Hindu Business Line*—Internet Edition, February 16, 2004; Rohit Bansal and Srividya Easwaran, "Creative Marketing for Rural India," research paper, <http://www.indiaonline.com>; Alex Steffen, "Rural India Ain't What It Used to Be," *WorldChanging*; <http://www.worldchanging.com/archives/001235.html>; "Corporates Turn to Rural India for Growth," BS Corporate Bureau in New Delhi, August 21, 2003, <http://www.rediff.com/money/2003/aug/21rural.htm>; Rajesh Jain, "Tech Talk: The Discovery of India: Rural India," June 20, 2003, <http://www.emergic.org/archives/indi/005721.php>. "Marketing to Rural India: Making the Ends Meet," March 8, 2007, in *India Knowledge@Wharton*. <http://knowledge.wharton.upenn.edu/india/article.cfm?articleid=4172>

Every minute of the working day, decisions are made by businesses around the world that determine whether companies will be profitable and growing or whether they will stagnate and die. Most of these decisions are made with the assistance of information gathered about the marketplace, the economic and financial environment, the workforce, the competition, and other factors. Such information usually comes in the form of data or is accompanied by data.

Business statistics provides the tool through which such data are collected, analyzed, summarized, and presented to facilitate the decision-making process, and business statistics plays an important role in the ongoing saga of decision making within the dynamic world of business.



1.1 STATISTICS IN BUSINESS

Virtually every area of business uses statistics in decision making. Here are some recent examples:

- According to a TNS Retail Forward ShopperScope survey, the average amount spent by a shopper on electronics in a three-month period is \$629 at Circuit City, \$504 at Best Buy, \$246 at Wal-Mart, \$172 at Target, and \$120 at RadioShack.
- A survey of 1465 workers by Hotjobs reports that 55% of workers believe that the quality of their work is perceived the same when they work remotely as when they are physically in the office.
- A survey of 477 executives by the Association of Executive Search Consultants determined that 48% of men and 67% of women say they are more likely to negotiate for less travel compared with five years ago.
- A survey of 1007 adults by RBC Capital Markets showed that 37% of adults would be willing to drive 5 to 10 miles to save 20 cents on a gallon of gas.
- A Deloitte Retail “Green” survey of 1080 adults revealed that 54% agreed that plastic, non-compostable shopping bags should be banned.
- A recent Household Economic Survey by Statistic New Zealand determined that the average weekly household net expenditure in New Zealand was \$956 and that households in the Wellington region averaged \$120 weekly on recreation and culture. In addition, 75% of all households were satisfied or very satisfied with their material standard of living.
- The Experience’s Life After College survey of 320 recent college graduates showed that 58% moved back home after college. Thirty-two percent then remained at home for more than a year.

You can see from these few examples that there is a wide variety of uses and applications of statistics in business. Note that in most of these examples, business researchers have conducted a study and provided us rich and interesting information.

Nick M Doi/Stock
Exclusive/Getty Images, Inc.

William King/The Image Bank/
Getty Images, Inc.



Comstock/Getty Images, Inc.

In this text we will examine several types of graphs for depicting data as we study ways to arrange or structure data into forms that are both meaningful and useful to decision makers. We will learn about techniques for sampling from a population that allow studies of the business world to be conducted more inexpensively and in a more timely manner. We will explore various ways to forecast future values and examine techniques for predicting trends. This text also includes many statistical tools for testing hypotheses and for estimating population values. These and many other exciting statistics and statistical techniques await us on this journey through business statistics. Let us begin.



BASIC STATISTICAL CONCEPTS

Business statistics, like many areas of study, has its own language. It is important to begin our study with an introduction of some basic concepts in order to understand and communicate about the subject. We begin with a discussion of the word *statistics*. The word *statistics* has many different meanings in our culture. *Webster's Third New International Dictionary* gives a comprehensive definition of **statistics** as *a science dealing with the collection, analysis, interpretation, and presentation of numerical data*. Viewed from this perspective, statistics includes all the topics presented in this text.

Photodisc/Getty Images, Inc.



The study of statistics can be organized in a variety of ways. One of the main ways is to subdivide statistics into two branches: descriptive statistics and inferential statistics. To understand the difference between descriptive and inferential statistics, definitions of *population* and *sample* are helpful. *Webster's Third New International Dictionary* defines **population** as *a collection of persons, objects, or items of interest*. The population can be a widely defined category, such as “all automobiles,” or it can be narrowly defined, such as “all Ford Mustang cars produced from 2002 to 2005.” A population can be a group of people, such as “all workers presently employed by Microsoft,” or it can be a set of objects, such as “all dishwashers produced on February 3, 2007, by the General Electric Company at the Louisville plant.” The researcher defines the population to be whatever he or she is studying. When researchers *gather data from the whole population for a given measurement of interest*, they call it a **census**. Most people are familiar with the U.S. Census. Every 10 years, the government attempts to measure all persons living in this country.

A **sample** is *a portion of the whole* and, if properly taken, is representative of the whole. For various reasons (explained in Chapter 7), researchers often prefer to work with a sample

of the population instead of the entire population. For example, in conducting quality-control experiments to determine the average life of lightbulbs, a lightbulb manufacturer might randomly sample only 75 lightbulbs during a production run. Because of time and money limitations, a human resources manager might take a random sample of 40 employees instead of using a census to measure company morale.

If a business analyst is *using data gathered on a group to describe or reach conclusions about that same group*, the statistics are called **descriptive statistics**. For example, if an instructor produces statistics to summarize a class's examination effort and uses those statistics to reach conclusions about that class only, the statistics are descriptive.

Many of the statistical data generated by businesses are descriptive. They might include number of employees on vacation during June, average salary at the Denver office, corporate sales for 2009, average managerial satisfaction score on a company-wide census of employee attitudes, and average return on investment for the Lofton Company for the years 1990 through 2008.

Another type of statistics is called **inferential statistics**. If a researcher *gathers data from a sample and uses the statistics generated to reach conclusions about the population from which the sample was taken*, the statistics are inferential statistics. The data gathered from the sample are used to infer something about a larger group. Inferential statistics are sometimes referred to as *inductive statistics*. The use and importance of inferential statistics continue to grow.

One application of inferential statistics is in pharmaceutical research. Some new drugs are expensive to produce, and therefore tests must be limited to small samples of patients. Utilizing inferential statistics, researchers can design experiments with small randomly selected samples of patients and attempt to reach conclusions and make inferences about the population.

Market researchers use inferential statistics to study the impact of advertising on various market segments. Suppose a soft drink company creates an advertisement depicting a dispensing machine that talks to the buyer, and market researchers want to measure the impact of the new advertisement on various age groups. The researcher could stratify the population into age categories ranging from young to old, randomly sample each stratum, and use inferential statistics to determine the effectiveness of the advertisement for the various age groups in the population. The advantage of using inferential statistics is that they enable the researcher to study effectively a wide range of phenomena without having to conduct a census. Most of the topics discussed in this text pertain to inferential statistics.

A *descriptive measure of the population* is called a **parameter**. Parameters are usually denoted by Greek letters. Examples of parameters are population mean (μ), population variance (σ^2), and population standard deviation (σ). A *descriptive measure of a sample* is called a **statistic**. Statistics are usually denoted by Roman letters. Examples of statistics are sample mean (\bar{x}), sample variance (s^2), and sample standard deviation (s).

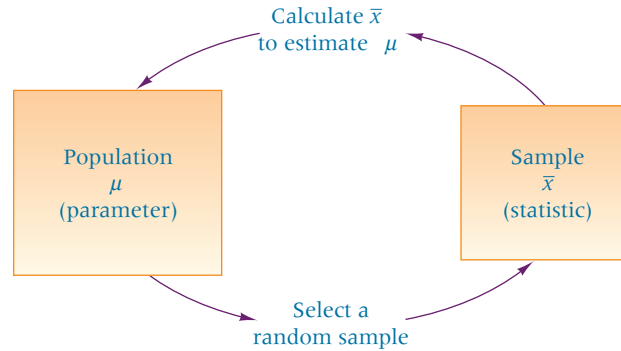
Differentiation between the terms *parameter* and *statistic* is important only in the use of inferential statistics. A business researcher often wants to estimate the value of a parameter or conduct tests about the parameter. However, the calculation of parameters is usually either impossible or infeasible because of the amount of time and money required to take a census. In such cases, the business researcher can take a random sample of the population, calculate a statistic on the sample, and infer by estimation the value of the parameter. The basis for inferential statistics, then, is the ability to make decisions about parameters without having to complete a census of the population.

For example, a manufacturer of washing machines would probably want to determine the average number of loads that a new machine can wash before it needs repairs. The parameter is the population mean or average number of washes per machine before repair. A company researcher takes a sample of machines, computes the number of washes before repair for each machine, averages the numbers, and estimates the population value or parameter by using the statistic, which in this case is the sample average. Figure 1.1 demonstrates the inferential process.

Inferences about parameters are made under uncertainty. Unless parameters are computed directly from the population, the statistician never knows with certainty whether the estimates or inferences made from samples are true. In an effort to estimate the level of confidence in the result of the process, statisticians use probability statements. For this and other reasons, part of this text is devoted to probability (Chapter 4).

FIGURE 1.1

Process of Inferential Statistics
to Estimate a Population
Mean (μ)



1.3 DATA MEASUREMENT

Millions of numerical data are gathered in businesses every day, representing myriad items. For example, numbers represent dollar costs of items produced, geographical locations of retail outlets, weights of shipments, and rankings of subordinates at yearly reviews. All such data should not be analyzed the same way statistically because the entities represented by the numbers are different. For this reason, the business researcher needs to know the *level of data measurement* represented by the numbers being analyzed.

The disparate use of numbers can be illustrated by the numbers 40 and 80, which could represent the weights of two objects being shipped, the ratings received on a consumer test by two different products, or football jersey numbers of a fullback and a wide receiver. Although 80 pounds is twice as much as 40 pounds, the wide receiver is probably not twice as big as the fullback! Averaging the two weights seems reasonable, but averaging the football jersey numbers makes no sense. The appropriateness of the data analysis depends on the level of measurement of the data gathered. The phenomenon represented by the numbers determines the level of data measurement. Four common levels of data measurement follow.

1. Nominal
2. Ordinal
3. Interval
4. Ratio

Nominal Level

The *lowest level of data measurement* is the **nominal level**. Numbers representing nominal-level data (the word *level* often is omitted) can be *used only to classify or categorize*. Employee identification numbers are an example of nominal data. The numbers are used only to differentiate employees and not to make a value statement about them. Many demographic questions in surveys result in data that are nominal because the questions are used for classification only. The following is an example of such a question that would result in nominal data:

Which of the following employment classifications best describes your area of work?

1. Educator
2. Construction worker
3. Manufacturing worker
4. Lawyer
5. Doctor
6. Other

Suppose that, for computing purposes, an educator is assigned a 1, a construction worker is assigned a 2, a manufacturing worker is assigned a 3, and so on. These numbers should be used only to classify respondents. The number 1 does not denote the top classification. It is used only to differentiate an educator (1) from a lawyer (4).

Some other types of variables that often produce nominal-level data are sex, religion, ethnicity, geographic location, and place of birth. Social Security numbers, telephone numbers, employee ID numbers, and ZIP code numbers are further examples of nominal data. Statistical techniques that are appropriate for analyzing nominal data are limited. However, some of the more widely used statistics, such as the chi-square statistic, can be applied to nominal data, often producing useful information.

Ordinal Level

Ordinal-level data measurement is higher than the nominal level. In addition to the nominal-level capabilities, ordinal-level measurement can be used to rank or order objects. For example, using ordinal data, a supervisor can evaluate three employees by ranking their productivity with the numbers 1 through 3. The supervisor could identify one employee as the most productive, one as the least productive, and one as somewhere between by using ordinal data. However, the supervisor could not use ordinal data to establish that the intervals between the employees ranked 1 and 2 and between the employees ranked 2 and 3 are equal; that is, she could not say that the differences in the amount of productivity between workers ranked 1, 2, and 3 are necessarily the same. With ordinal data, the distances or spacing represented by consecutive numbers are not always equal.

Some questionnaire Likert-type scales are considered by many researchers to be ordinal in level. The following is an example of one such scale:

This computer tutorial is	—	—	—	—	—
	not	somewhat	moderately	very	extremely
	helpful	helpful	helpful	helpful	helpful
	1	2	3	4	5

When this survey question is coded for the computer, only the numbers 1 through 5 will remain, not the adjectives. Virtually everyone would agree that a 5 is higher than a 4 on this scale and that ranking responses is possible. However, most respondents would not consider the differences between not helpful, somewhat helpful, moderately helpful, very helpful, and extremely helpful to be equal.

Mutual funds as investments are sometimes rated in terms of risk by using measures of default risk, currency risk, and interest rate risk. These three measures are applied to investments by rating them as having high, medium, and low risk. Suppose high risk is assigned a 3, medium risk a 2, and low risk a 1. If a fund is awarded a 3 rather than a 2, it carries more risk, and so on. However, the differences in risk between categories 1, 2, and 3 are not necessarily equal. Thus, these measurements of risk are only ordinal-level measurements. Another example of the use of ordinal numbers in business is the ranking of the top 50 most admired companies in *Fortune* magazine. The numbers ranking the companies are only ordinal in measurement. Certain statistical techniques are specifically suited to ordinal data, but many other techniques are not appropriate for use on ordinal data. For example, it does not make sense to say that the average of “moderately helpful” and “very helpful” is “moderately helpful and a half.”

Because nominal and ordinal data are often derived from imprecise measurements such as demographic questions, the categorization of people or objects, or the ranking of items, *nominal and ordinal data* are **nonmetric data** and are sometimes referred to as *qualitative data*.

Interval Level

Interval-level data measurement is the *next to the highest level of data in which the distances between consecutive numbers have meaning and the data are always numerical*. The distances represented by the differences between consecutive numbers are equal; that is, interval data have equal intervals. An example of interval measurement is Fahrenheit temperature. With Fahrenheit temperature numbers, the temperatures can be ranked, and the amounts of heat between consecutive readings, such as 20°, 21°, and 22°, are the same.

In addition, with interval-level data, the zero point is a matter of convention or convenience and not a natural or fixed zero point. Zero is just another point on the scale and does not mean the absence of the phenomenon. For example, zero degrees Fahrenheit is not the lowest possible temperature. Some other examples of interval level data are the percentage change in employment, the percentage return on a stock, and the dollar change in stock price.

Ratio Level

Ratio-level data measurement is *the highest level of data measurement*. Ratio data have the same properties as interval data, but ratio data have an *absolute zero*, and *the ratio of two numbers is meaningful*. The notion of absolute zero means that zero is fixed, and *the zero value in the data represents the absence of the characteristic being studied*. The value of zero cannot be arbitrarily assigned because it represents a fixed point. This definition enables the statistician to create *ratios* with the data.

Examples of ratio data are height, weight, time, volume, and Kelvin temperature. With ratio data, a researcher can state that 180 pounds of weight is twice as much as 90 pounds or, in other words, make a ratio of 180:90. Many of the data gathered by machines in industry are ratio data.

Other examples in the business world that are ratio level in measurement are production cycle time, work measurement time, passenger miles, number of trucks sold, complaints per 10,000 fliers, and number of employees. With ratio-level data, no *b* factor is required in converting units from one measurement to another—that is, $y = ax$. As an example, in converting height from yards to feet: $\text{feet} = 3 \cdot \text{yards}$.

Because interval- and ratio-level data are usually gathered by precise instruments often used in production and engineering processes, in national standardized testing, or in standardized accounting procedures, they are called **metric data** and are sometimes referred to as *quantitative data*.

Comparison of the Four Levels of Data

FIGURE 1.2

Usage Potential of Various Levels of Data

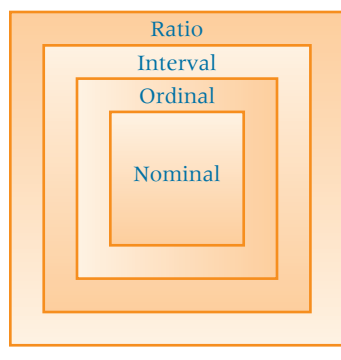


Figure 1.2 shows the relationships of the usage potential among the four levels of data measurement. The concentric squares denote that each higher level of data can be analyzed by any of the techniques used on lower levels of data but, in addition, can be used in other statistical techniques. Therefore, ratio data can be analyzed by any statistical technique applicable to the other three levels of data plus some others.

Nominal data are the most limited data in terms of the types of statistical analysis that can be used with them. Ordinal data allow the researcher to perform any analysis that can be done with nominal data and some additional analyses. With ratio data, a statistician can make ratio comparisons and appropriately do any analysis that can be performed on nominal, ordinal, or interval data. Some statistical techniques require ratio data and cannot be used to analyze other levels of data.

Statistical techniques can be separated into two categories: parametric statistics and nonparametric statistics. **Parametric statistics** require that data be interval or ratio. If the data are nominal or ordinal, **nonparametric statistics** must be used. Nonparametric statistics can also be used to analyze interval or ratio data. This text focuses largely on parametric statistics, with the exception of Chapter 16 and Chapter 17, which contain nonparametric techniques. Thus much of the material in this text requires that data be interval or ratio data.

DEMONSTRATION PROBLEM 1.1



Demonstration Problem

Many changes continue to occur in the healthcare industry. Because of increased competition for patients among providers and the need to determine how providers can better serve their clientele, hospital administrators sometimes administer a quality satisfaction survey to their patients after the patient is released. The following types of

questions are sometimes asked on such a survey. These questions will result in what level of data measurement?

1. How long ago were you released from the hospital?
2. Which type of unit were you in for most of your stay?
 - Coronary care
 - Intensive care
 - Maternity care
 - Medical unit
 - Pediatric/children's unit
 - Surgical unit
3. In choosing a hospital, how important was the hospital's location?

(circle one)

Very Important	Somewhat Important	Not Very Important	Not at All Important
-------------------	-----------------------	-----------------------	-------------------------
4. How serious was your condition when you were first admitted to the hospital?
 - Critical
 - Serious
 - Moderate
 - Minor
5. Rate the skill of your doctor:
 - Excellent
 - Very Good
 - Good
 - Fair
 - Poor

Solution

Question 1 is a time measurement with an absolute zero and is therefore ratio-level measurement. A person who has been out of the hospital for two weeks has been out twice as long as someone who has been out of the hospital for one week.

Question 2 yields nominal data because the patient is asked only to categorize the type of unit he or she was in. This question does not require a hierarchy or ranking of the type of unit. Questions 3, 4, and 5 are likely to result in ordinal-level data. Suppose a number is assigned the descriptors in each of these three questions. For question 3, "very important" might be assigned a 4, "somewhat important" a 3, "not very important" a 2, and "not at all important" a 1. Certainly, the higher the number, the more important is the hospital's location. Thus, these responses can be ranked by selection. However, the increases in importance from 1 to 2 to 3 to 4 are not necessarily equal. This same logic applies to the numeric values assigned in questions 4 and 5.

STATISTICS IN BUSINESS TODAY

Cellular Phone Use in Japan

The Communications and Information Network Association of Japan (CIAJ) conducts an annual study of cellular phone use in Japan. A recent survey was taken as part of this study using a sample of 600 cell phone users split evenly between men and women and almost equally distributed over six age brackets. The survey was administered in the greater Tokyo and Osaka metropolitan areas. The study produced several interesting findings. It was determined that 62.2% had replaced their handsets in the previous 10 months. A little more than 6% owned a second cell phone. Of these, the objective of about two-thirds was to own one for business use and a second one

for personal use. Of all those surveyed, 18.2% used their handsets to view videos, and another 17.3% were not currently using their handsets to view videos but were interested in doing so. Some of the everyday uses of cell phones included e-mailing (91.7% of respondents), camera functions (77.7%), Internet searching (46.7%), and watching TV (28.0%). In the future, respondents hoped there would be cell phones with high-speed data transmission that could be used to send and receive PC files (47.7%), for video services such as You Tube (46.9%), for downloading music albums (45.3%) and music videos (40.8%), and for downloading long videos such as movies (39.2%).

Statistical Analysis Using the Computer: Excel and Minitab

The advent of the modern computer opened many new opportunities for statistical analysis. The computer allows for storage, retrieval, and transfer of large data sets. Furthermore, computer software has been developed to analyze data by means of sophisticated statistical techniques. Some widely used statistical techniques, such as multiple regression, are so tedious and cumbersome to compute manually that they were of little practical use to researchers before computers were developed.

Business statisticians use many popular statistical software packages, including Minitab, SAS, and SPSS. Many computer spreadsheet software packages also have the capability of analyzing data statistically. In this text, the computer statistical output presented is from both the Minitab and the Microsoft Excel software.



Statistics Describe the State of Business in India's Countryside

as income, number of children, age of household heads, number of livestock, and grams of toothpaste consumed per year might be obtained. On the other hand, if researchers use a Likert scale (1-to-5 measurements) to gather responses about the interests, likes, and preferences of rural India consumers, an ordinal-level measurement would be obtained, as would the ranking of products or brands in market research studies. Other variables, such as geographic location, sex, occupation, or religion, are usually measured with nominal data.

The decision to enter the rural India market is not just a marketing decision. It involves production capacity and schedule issues, transportation challenges, financial commitments, managerial growth or reassignment, accounting issues (accounting for rural India may differ from techniques used in traditional markets), information systems, and other related areas. With so much on the line, company decision makers need as much relevant information available as possible. In this Decision Dilemma, it is obvious to the decision maker that rural India is still quite poor and illiterate. Its capacity as a market is great. The statistics on the increasing sales of a few personal-care products look promising. What are the future forecasts for the earning power of people in rural India? Will major cultural issues block the adoption of the types of products that companies want to sell there? The answers to these and many other interesting and useful questions can be obtained by the appropriate use of statistics. The 750 million people living in rural India represent the second largest group of people in the world. It certainly is a market segment worth studying further.

Several statistics were reported in the Decision Dilemma about rural India, including the average annual consumption of toothpaste per person, the percentage of households having electricity, and the percentage of households that have at least one bank account. The authors of the sources from which the Decision Dilemma was drawn never stated whether the reported statistics were based on actual data drawn from a census of rural India households or were based on estimates taken from a sample of rural households. If the data came from a census, then the totals, averages, and percentages presented in the Decision Dilemma are parameters. If, on the other hand, the data were gathered from samples, then they are statistics. Although governments especially do conduct censuses and at least some of the reported numbers could be parameters, more often than not, such data are gathered from samples of people or items. For example, in rural India, the government, academicians, or business researchers could have taken random samples of households, gathering consumer statistics that are then used to estimate population parameters, such as percentage of households with televisions, and so forth.

In conducting research on a topic like consumer consumption in rural India, there is potential for a wide variety of statistics to be gathered that represent several levels of data. For example, ratio-level measurements on items such

ETHICAL CONSIDERATIONS

With the abundance and proliferation of statistical data, potential misuse of statistics in business dealings is a concern. It is, in effect, unethical business behavior to use statistics out of context. Unethical businesspeople might use only selective data from studies to underscore their point, omitting statistics from the same studies that argue against their case. The results of statistical studies can be misstated or overstated to gain favor.

This chapter noted that if data are nominal or ordinal, then only nonparametric statistics are appropriate for

analysis. The use of parametric statistics to analyze nominal and/or ordinal data is wrong and could be considered under some circumstances to be unethical.

In this text, each chapter contains a section on ethics that discusses how businesses can misuse the techniques presented in the chapter in an unethical manner. As both users and producers, business students need to be aware of the potential ethical pitfalls that can occur with statistics.

SUMMARY

Statistics is an important decision-making tool in business and is used in virtually every area of business. In this course, the word *statistics* is defined as the science of gathering, analyzing, interpreting, and presenting data.

The study of statistics can be subdivided into two main areas: *descriptive statistics* and *inferential statistics*. Descriptive statistics result from gathering data from a body, group, or population and reaching conclusions only about that group. Inferential statistics are generated from the process of gathering sample data from a group, body, or population and reaching conclusions about the larger group from which the sample was drawn.

The appropriate type of statistical analysis depends on the level of data measurement, which can be (1) *nominal*, (2) *ordinal*, (3) *interval*, or (4) *ratio*. Nominal is the lowest level, representing classification of only such data as geographic location, sex, or Social Security number. The next level is ordi-

nal, which provides rank ordering measurements in which the intervals between consecutive numbers do not necessarily represent equal distances. Interval is the next to highest level of data measurement in which the distances represented by consecutive numbers are equal. The highest level of data measurement is ratio, which has all the qualities of interval measurement, but ratio data contain an absolute zero and ratios between numbers are meaningful. Interval and ratio data sometimes are called *metric* or *quantitative* data. Nominal and ordinal data sometimes are called *nonmetric* or *qualitative* data.

Two major types of inferential statistics are (1) *parametric statistics* and (2) *nonparametric statistics*. Use of parametric statistics requires interval or ratio data and certain assumptions about the distribution of the data. The techniques presented in this text are largely parametric. If data are only nominal or ordinal in level, nonparametric statistics must be used.

KEY TERMS



Flash Cards

census
descriptive statistics

inferential statistics
interval-level data
metric data
nominal-level data
nonmetric data
nonparametric statistics

ordinal-level data
parameter
parametric statistics
population
ratio-level data
sample
statistic
statistics

SUPPLEMENTARY PROBLEMS

- 1.1 Give a specific example of data that might be gathered from each of the following business disciplines: accounting, finance, human resources, marketing, information systems, production, and management. An example in the marketing area might be “number of sales per month by each salesperson.”
- 1.2 State examples of data that can be gathered for decision making purposes from each of the following industries: manufacturing, insurance, travel, retailing, communications, computing, agriculture, banking, and healthcare. An example in the travel industry might be the cost of business travel per day in various European cities.

- 1.3 Give an example of *descriptive* statistics in the recorded music industry. Give an example of how *inferential* statistics could be used in the recorded music industry. Compare the two examples. What makes them different?
- 1.4 Suppose you are an operations manager for a plant that manufactures batteries. Give an example of how you could use *descriptive* statistics to make better managerial decisions. Give an example of how you could use *inferential* statistics to make better managerial decisions.
- 1.5 Classify each of the following as nominal, ordinal, interval, or ratio data.
- The time required to produce each tire on an assembly line
 - The number of quarts of milk a family drinks in a month
 - The ranking of four machines in your plant after they have been designated as excellent, good, satisfactory, and poor
 - The telephone area code of clients in the United States
 - The age of each of your employees
 - The dollar sales at the local pizza shop each month
 - An employee's identification number
 - The response time of an emergency unit
- 1.6 Classify each of the following as nominal, ordinal, interval, or ratio data.
- The ranking of a company by *Fortune* 500
 - The number of tickets sold at a movie theater on any given night
 - The identification number on a questionnaire
 - Per capita income
- The trade balance in dollars
 - Profit/loss in dollars
 - A company's tax identification
 - The Standard & Poor's bond ratings of cities based on the following scales:

Rating	Grade
Highest quality	AAA
High quality	AA
Upper medium quality	A
Medium quality	BBB
Somewhat speculative	BB
Low quality, speculative	B
Low grade, default possible	CCC
Low grade, partial recovery possible	CC
Default, recovery unlikely	C

- 1.7 The Rathburn Manufacturing Company makes electric wiring, which it sells to contractors in the construction industry. Approximately 900 electric contractors purchase wire from Rathburn annually. Rathburn's director of marketing wants to determine electric contractors' satisfaction with Rathburn's wire. He developed a questionnaire that yields a satisfaction score between 10 and 50 for participant responses. A random sample of 35 of the 900 contractors is asked to complete a satisfaction survey. The satisfaction scores for the 35 participants are averaged to produce a mean satisfaction score.
- What is the population for this study?
 - What is the sample for this study?
 - What is the statistic for this study?
 - What would be a parameter for this study?

ANALYZING THE DATABASES

see www.wiley.com/college/black



Nine databases are available with this text, providing additional opportunities to apply the statistics presented in this course. These databases are located in WileyPLUS, and each is available in either Minitab or Excel format for your convenience. These nine databases represent a wide variety of business areas, such as agribusiness, consumer spending, energy, finance, healthcare, international labor, manufacturing, and the stock market. Altogether, these databases contain 61 variables and 7722 observations. The data are gathered from such reliable sources as the U.S. government's Bureau of Labor, the U.S. Department of Agriculture, the American Hospital Association, the Energy Information Administration, *Moody's Handbook of Common Stocks*, and the U.S. Census Bureau. Five of the nine databases contain time-series data. These databases are:

12-YEAR GASOLINE DATABASE

The 12-year time-series gasoline database contains monthly data for four variables: U.S. Gasoline Prices, OPEC Spot Price,

U.S. Finished Motor Gasoline Production, and U.S. Natural Gas Wellhead Price. There are 137 data entries for each variable. U.S. Gasoline Prices are given in cents, the OPEC Spot Price is given in dollars per barrel, U.S. Finished Motor Gasoline Production is given in 1000 barrels per day, and U.S. Natural Gas Wellhead Price is given in dollars per 1000 cubic feet.

CONSUMER FOOD DATABASE

The consumer food database contains five variables: Annual Food Spending per Household, Annual Household Income, Non-Mortgage Household Debt, Geographic Region of the U.S. of the Household, and Household Location. There are 200 entries for each variable in this database representing 200 different households from various regions and locations in the United States. Annual Food Spending per Household, Annual Household Income, and Non-Mortgage Household Debt are all given in dollars. The variable, Region, tells in which one of four regions the household resides. In this variable, the Northeast is coded as 1, the Midwest is coded 2, the South is coded as 3, and the West is coded as 4. The variable, Location, is coded as 1 if the household is in a metropolitan

area and 2 if the household is outside a metro area. The data in this database were randomly derived and developed based on actual national norms.

MANUFACTURING DATABASE

This database contains eight variables taken from 20 industries and 140 subindustries in the United States. Some of the industries are food products, textile mill products, furniture, chemicals, rubber products, primary metals, industrial machinery, and transportation equipment. The eight variables are Number of Employees, Number of Production Workers, Value Added by Manufacture, Cost of Materials, Value of Industry Shipments, New Capital Expenditures, End-of-Year Inventories, and Industry Group. Two variables, Number of Employees and Number of Production Workers, are in units of 1,000. Four variables, Value Added by Manufacture, Cost of Materials, New Capital Expenditures, and End-of-Year Inventories, are in million-dollar units. The Industry Group variable consists of numbers from 1 to 20 to denote the industry group to which the particular subindustry belongs. Value of Industry Shipments has been recoded to the following 1-to-4 scale.

- 1 = \$0 to \$4.9 billion
- 2 = \$5 billion to \$13.9 billion
- 3 = \$14 billion to \$28.9 billion
- 4 = \$29 billion or more

INTERNATIONAL LABOR DATABASE

This time-series database contains the civilian unemployment rates in percent from seven countries presented yearly over a 49-year period. The data are published by the Bureau of Labor Statistics of the U.S. Department of Labor. The countries are the United States, Canada, Australia, Japan, France, Germany, and Italy.

FINANCIAL DATABASE

The financial database contains observations on eight variables for 100 companies. The variables are Type of Industry, Total Revenues (\$ millions), Total Assets (\$ millions), Return on Equity (%), Earnings per Share (\$), Average Yield (%), Dividends per Share (\$), and Average Price per Earnings (P/E) ratio. The companies represent seven different types of industries. The variable “Type” displays a company’s industry type as:

- 1 = apparel
- 2 = chemical
- 3 = electric power
- 4 = grocery
- 5 = healthcare products
- 6 = insurance
- 7 = petroleum

ENERGY DATABASE

The time-series energy database consists of data on five energy variables over a period of 35 years. The five variables are U.S. Energy Consumption, World Crude Oil Production, U.S. Nuclear Electricity Generation, U.S. Coal Production, and U.S. Natural Dry Gas Production. U.S. Energy Consumption

is given in quadrillion BTUs per year, World Crude Oil Production is given in million barrels per day, U.S. Nuclear Electricity Generation is given in billion kilowatt-hours, U.S. Coal Production is given in million short tons, and U.S. Natural Dry Gas Production is given in million cubic feet.

U.S. AND INTERNATIONAL STOCK MARKET DATABASE

This database contains seven variables—three from the U.S. stock market and four from international stock markets—with data representing monthly averages of each over a period of five years resulting in 60 data points per variable. The U.S. stock market variables include the Dow Jones Industrial Average, the NASDAQ, and Standard & Poor’s 500. The four international stock market variables of Nikkei 225, Hang Seng, FTSE 100, and IPC represent Japan, Hong Kong, United Kingdom, and Mexico.

HOSPITAL DATABASE

This database contains observations for 11 variables on U.S. hospitals. These variables include Geographic Region, Control, Service, Number of Beds, Number of Admissions, Census, Number of Outpatients, Number of Births, Total Expenditures, Payroll Expenditures, and Personnel.

The region variable is coded from 1 to 7, and the numbers represent the following regions:

- 1 = South
- 2 = Northeast
- 3 = Midwest
- 4 = Southwest
- 5 = Rocky Mountain
- 6 = California
- 7 = Northwest

Control is a type of ownership. Four categories of control are included in the database:

- 1 = government, nonfederal
- 2 = nongovernment, not-for-profit
- 3 = for-profit
- 4 = federal government

Service is the type of hospital. The two types of hospitals used in this database are:

- 1 = general medical
- 2 = psychiatric

The total expenditures and payroll variables are in units of \$1,000.

AGRIBUSINESS TIME-SERIES DATABASE

The agribusiness time-series database contains the monthly weight (in 1,000 lbs.) of cold storage holdings for six different vegetables and for total frozen vegetables over a 14-year period. Each of the seven variables represents 168 months of data. The six vegetables are green beans, broccoli, carrots, sweet corn, onions, and green peas. The data are published by the National Agricultural Statistics Service of the U.S. Department of Agriculture.

ASSIGNMENT

Use the databases to answer the following questions.

1. In the manufacturing database, what is the level of data for each of the following variables?
 - a. Number of Production Workers
 - b. Cost of Materials
 - c. Value of Industry Shipments
 - d. Industry Group
2. In the hospital database, what is the level of data for each of the following variables?
 - a. Region
 - b. Control
 - c. Number of Beds
 - d. Personnel
3. In the financial database, what is the level of data for each of the following variables?
 - a. Type of Industry
 - b. Total Assets
 - c. P/E Ratio

CASE**DIGIORNO PIZZA: INTRODUCING A FROZEN PIZZA TO COMPETE WITH CARRY-OUT**

Kraft Foods successfully introduced DiGiorno Pizza into the marketplace in 1996, with first year sales of \$120 million, followed by \$200 million in sales in 1997. It was neither luck nor coincidence that DiGiorno Pizza was an instant success. Kraft conducted extensive research about the product and the marketplace before introducing this product to the public. Many questions had to be answered before Kraft began production. For example, why do people eat pizza? When do they eat pizza? Do consumers believe that carry out pizza is always more tasty?

SMI-Alcott conducted a research study for Kraft in which they sent out 1,000 surveys to pizza lovers. The results indicated that people ate pizza during fun social occasions or at home when no one wanted to cook. People used frozen pizza mostly for convenience but selected carry out pizza for a variety of other reasons, including quality and the avoidance of cooking. The Loran Marketing Group conducted focus groups for Kraft with women aged 25 to 54. Their findings showed that consumers used frozen pizza for convenience but wanted carry out pizza taste. Kraft researchers realized that if they were to launch a successful frozen pizza that could compete with carry out pizza, they had to develop a frozen pizza that (a) had restaurant take-out quality, (b) possessed flavor variety, (c) was fast and easy to prepare, and (d) had the convenience of freezer storage. To satisfy these seemingly divergent goals, Kraft developed DiGiorno Pizza, which rises in the oven as it cooks. This impressed focus group members, and in a series of blind taste tests conducted by Product Dynamics, DiGiorno Pizza beat out all frozen pizzas and finished second overall behind one carry out brand.

DiGiorno Pizza has continued to grow in sales and market share over the years. By 2005, sales had topped the \$600 million mark, and DiGiorno Pizza held nearly a quarter of the market share of frozen pizza sales. In 2004, Kraft successfully introduced DiGiorno thin, crispy-crust pizza into the market. Sales revenues continue to grow for this leading brand of frozen pizza.

Discussion

Think about the market research that was conducted by Kraft and the fact that they used several companies.

1. What are some of the populations that Kraft might have been interested in measuring for these studies? Did Kraft actually attempt to contact entire populations? What samples were taken? In light of these two questions, how was the inferential process used by Kraft in their market research? Can you think of any descriptive statistics that might have been used by Kraft in their decision-making process?
2. In the various market research efforts made by Kraft for DiGiorno, some of the possible measurements appear in the following list. Categorize these by level of data. Think of some other measurements that Kraft researchers might have made to help them in this research effort, and categorize them by level of data.
 - a. Number of pizzas consumed per week per household
 - b. Age of pizza purchaser
 - c. Zip code of the survey respondent
 - d. Dollars spent per month on pizza per person
 - e. Time in between purchases of pizza
 - f. Rating of taste of a given pizza brand on a scale from 1 to 10, where 1 is very poor tasting and 10 is excellent taste
 - g. Ranking of the taste of four pizza brands on a taste test
 - h. Number representing the geographic location of the survey respondent
 - i. Quality rating of a pizza brand as excellent, good, average, below average, poor
 - j. Number representing the pizza brand being evaluated
 - k. Sex of survey respondent

Source: Adapted from "Upper Crust," *American Demographics*, March 1999, p. 58; *Marketwatch—News That Matters* Web sites, "What's in a Name? Brand Extension Potential" and "DiGiorno Rising Crust Delivers \$200 Million," formerly at <http://www.foodexplorer.com/BUSINESS/Products/MarketAnalysis/PF02896b.htm>, last accessed in 1999. Web site for Kraft's 2005 Investor Meeting: Changing Consumers, Changing Kraft, May 25, http://media.corporateir.net/media_files/nys/kft/presentations/kft_050510c.pdf.