

Business Analytics

Data Analysis and Decision Making

5e



Albright | Winston

Overview of Applications in the Book, by Discipline

Accounting

Accounts receivable	307, 320
Auditing for price errors	355
Developing a flexible budget	587
Estimating total tax refunds	349
Estimating total taxable income	335
Overhead cost analysis	469, 481, 489, 540, 569, 573, 587

Economics/Government

Demand and cost for electricity	506
Demand for desktops and laptops	447
Demand for domestic automobiles	514
Demand for French bread	530
Demand for heating oil	586
Demand for microwaves	182
Home and condo prices	78
Housing price structure	530
Peak-load pricing	790
Presidential elections	19
Sales of new houses	619

Finance

Bond investment strategy	943
Capital budgeting	770
Cash management	901
Credit card holder sampling	312
DJIA index	58, 77
Entering risky ventures	275
Investing for college	942
Investing for retirement	531, 906
Investment strategy	759
Investor's after-tax profit	181
Land purchase decision	294
Liquidity risk management	880
Market return scenarios	152, 158
Mutual fund returns	171, 195
New car development	897
Pension fund management	764
Portfolio analysis	796
Random walk of stock prices	616
Stock hedging	810

Human Resources

Drug testing for athletes	253
Employee empowerment	433
Employee retention	397
Gender discrimination	495, 502, 510, 563, 585
Jobs in statistics and mathematics	1

Personnel testing	178
Productivity due to exercise	427

Marketing

Catalog marketing	552
Churn in cellular phone market	136
Customer complaints	397, 403
Customer valuation	913
DVD movie renters	334
Electmart sales	108
Frozen lasagna dinner buyers	125, 954, 958
Furniture pricing	530
Marketing and selling condos	921
New pizza style decisions	403, 411
New product marketing	261, 278
Olympics sponsors	401
Response to new sandwich	343, 352, 383, 385
Running shoe style decisions	294
Sales presentation ratings	370
Sales response to coupons	376
Soft-drink can style decisions	422
Student reaction to new textbook	415
Supermarket sales	197
Supermarket transactions	27
Toothpaste dispenser decisions	938
Toy sales	26
Value of free maintenance agreement	917

Miscellaneous Statistical

Anti-drunk driving campaign	457
Biotechnical engineering	295
Crime in the U.S.	54
Cruise ship entertainment	16
Election returns	200
Family income sampling	305
Forecasting U.S. population	608
IQ and bell curve	166
Predictors of successful movies	79, 532
Questionnaire responses	23
Relationship between smoking and drinking	82
Removing Vioxx from market	459
Sample size determination in legal case	301
Saving, spending, social climbing	136
Simpson's paradox	165
University admissions	396

Operations Management

Aggregate planning	750
Airline boarding strategies	812
Airline hub location decisions	783
Arrivals at bank	135

Automobile production	734
Battery lifetimes	191
Bidding for contracts	232, 882
Blending oil	727
Customer waiting at supermarket	365
Delivery times at restaurant	397
Distribution of metal strip widths	441
Expensive watch production	219
Forecasting sales	602, 605, 611, 619, 626, 630, 636, 640, 644
Inventory management	208
Learning curve for production	516
Logging operations	399
Lost baggage at airports	47
Manufacturing plastics operations	657
New drug development	253
Ordering decisions	834, 836, 848, 858, 863, 866
Out-of-spec products	387
Overbooking at airlines	198
Product mix decisions	16, 661, 688, 776
Production quantity decisions	878, 879
Production scheduling	688, 696, 891
Production, inventory, distribution decisions	718
Quality control at paper company	179
Reliability of motors	362
Site selection of motor inns	463
Supermarket checkout times	197
Timing uncertainty in construction	144
Transportation, logistics decisions	734, 743
Variability in machine parts	358
Warranty costs	378, 886
Wine production	715
Worker scheduling	750

Sports/Gaming

Baseball salaries	31, 40, 45, 50, 88
Games at McDonald's	139
Golf stats on PGA tour	95
NCAA basketball tournament simulation	930
Revenue management at casino	590
Streak shooting in basketball	201
Wheel of fortune simulation	323
Winning at craps	927
Winning the lottery	220

5TH
EDITION

Business Analytics: Data Analysis and Decision Making

S. Christian Albright

Kelly School of Business, Indiana University, Emeritus

Wayne L. Winston

Kelly School of Business, Indiana University

With cases by

Mark Broadie

Graduate School of Business, Columbia University

Peter Kolesar

Graduate School of Business, Columbia University

Lawrence L. Lapin

San Jose State University

William D. Whisler

Emeritus, California State University, Hayward



Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

**Business Analytics: Data Analysis and
Decision Making, Fifth Edition**

S. Christian Albright and Wayne L. Winston

Product Director: Joe Sabatino

Product Manager: Aaron Arnsperger

Content Developer: Krista Kellman

Product Assistant: Brad Sullender

Marketing Director: Natalie King

Assoc. Market Development Manager:
Roy Rosa

Sr. Content Project Manager: Holly Henjum

Media Developer: Chris Valentine

Manufacturing Planner: Ron Montgomery

Marketing Coordinator: Eileen Corcoran

Composition and Production Service:
diacriTech

Sr. Art Director: Stacy Jenkins Shirley

Sr. Rights Acquisitions Specialist: John Hill

Cover Designer: Joe Devine/
Red Hangar Design

Cover Image: © nadla/iStockphoto

© 2015, 2011 Cengage Learning

WCN: 02-300

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at **Cengage Learning Customer & Sales Support, 1-800-354-9706**

For permission to use material from this text or product, submit all requests online at www.cengage.com/permissions

Further permissions questions can be emailed to permissionrequest@cengage.com

Screenshots are used with permission from Microsoft Corporation, Microsoft Excel® is a registered trademark of Microsoft Corporation.

© 2013 Microsoft.

@RISK, StatTools, PrecisionTree, TopRank, RISKOptimizer, NeuralTools®, and Evolver™, are © 2013 Palisade.com.

Except where otherwise noted, all content is © Cengage Learning 2015.

Library of Congress Control Number: 2013946375

Student Edition ISBN-13: 978-1-133-58826-9

Student Edition ISBN-10: 1-133-58826-3

Package Student Edition ISBN-13: 978-1-133-62960-3

Package Student Edition ISBN-10: 1-133-62960-1

Cengage Learning

200 First Stamford Place, 4th Floor

Stamford CT 06902

USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at:

www.cengage.com/global

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**

Printed in the United States of America
1 2 3 4 5 6 7 17 16 15 14 13

To my wonderful wife Mary—my best friend and travel mate; to Sam, Lindsay, Teddy, and Archer, our second adorable grandson; and to Bryn, our ball-playing Welsh corgi! S.C.A

To my wonderful family W.L.W.

About the Authors



© Cengage Learning

S. Christian Albright got his B.S. degree in Mathematics from Stanford in 1968 and his PhD in Operations Research from Stanford in 1972. He taught in the Operations & Decision Technologies Department in the Kelley School of Business at Indiana University (IU) for close to 40 years, before retiring from teaching in 2011. While at IU, he taught courses in management science, computer simulation, statistics, and computer programming to all levels of business students, including undergraduates, MBAs, and doctoral students. In addition, he taught simulation modeling at General Motors and Whirlpool, and he taught database analysis for the Army. He published over 20 articles in leading operations research journals in the area of applied probability, and he has authored the books *Statistics for Business and Economics*, *Practical Management Science*, *Spreadsheet Modeling and Applications*, *Data Analysis for Managers*, and *VBA for Modelers*. He currently works with the Palisade Corporation developing training materials for its software products, he has developed a commercial version of his Excel® tutorial, called ExcelNow!, and he continues to revise his textbooks.

for Modelers. He currently works with the Palisade Corporation developing training materials for its software products, he has developed a commercial version of his Excel® tutorial, called ExcelNow!, and he continues to revise his textbooks.

On the personal side, Chris has been married for 42 years to his wonderful wife, Mary, who retired several years ago after teaching 7th grade English for 30 years. They have one son, Sam, who lives in Philadelphia with his wife Lindsay and their two sons, Teddy and Archer. Chris has many interests outside the academic area. They include activities with his family (especially traveling with Mary), going to cultural events at IU, power walking while listening to books on his iPod, and reading. And although he earns his livelihood from statistics and management science, his *real* passion is for playing classical piano music.



© Cengage Learning

Wayne L. Winston is Professor of Operations & Decision Technologies in the Kelley School of Business at Indiana University, where he has taught since 1975. Wayne received his B.S. degree in Mathematics from MIT and his PhD in Operations Research from Yale. He has written the successful textbooks *Operations Research: Applications and Algorithms*, *Mathematical Programming: Applications and Algorithms*, *Simulation Modeling Using @RISK*, *Practical Management Science*, *Data Analysis and Decision Making*, *Financial Models Using Simulation and Optimization*, and *Mathletics*. Wayne has published more than 20 articles in leading journals and has won many teaching awards, including the school-wide MBA award four times. He has taught classes at Microsoft, GM, Ford, Eli Lilly, Bristol-Myers Squibb, Arthur Andersen, Roche, PricewaterhouseCoopers, and NCR. His

current interest is showing how spreadsheet models can be used to solve business problems in all disciplines, particularly in finance and marketing.

Wayne enjoys swimming and basketball, and his passion for trivia won him an appearance several years ago on the television game show *Jeopardy!*, where he won two games. He is married to the lovely and talented Vivian. They have two children, Gregory and Jennifer.

Brief Contents

Preface xiii

I Introduction to Data Analysis and Decision Making I

Part 1 Exploring Data 17

2 Describing the Distribution of a Single Variable 19

3 Finding Relationships among Variables 79

Part 2 Probability and Decision Making Under Uncertainty 137

4 Probability and Probability Distributions 139

5 Normal, Binomial, Poisson, and Exponential Distributions 166

6 Decision Making under Uncertainty 222

Part 3 Statistical Inference 299

7 Sampling and Sampling Distributions 301

8 Confidence Interval Estimation 335

9 Hypothesis Testing 401

Part 4 Regression Analysis and Time Series Forecasting 461

10 Regression Analysis: Estimating Relationships 463

11 Regression Analysis: Statistical Inference 532

12 Time Series Analysis and Forecasting 590

Part 5 Optimization and Simulation Modeling 655

13 Introduction to Optimization Modeling 657

14 Optimization Models 718

15 Introduction to Simulation Modeling 812

16 Simulation Models 880

Part 6 Advanced Data Analysis 945

17 Data Mining 947

Part 7 Bonus Online Material 18-1

18 Importing Data into Excel 18-3

19 Analysis of Variance and Experimental Design 19-1

20 Statistical Process Control 20-1

Appendix A Statistical Reporting A-1

References 985

Index 987

Contents

Preface xiii

I	Introduction to Data Analysis and Decision Making	1
1-1	Introduction	2
1-2	Overview of the Book	4
1-2a	The Methods	4
1-2b	The Software	7
1-3	Modeling and Models	10
1-3a	Graphical Models	11
1-3b	Algebraic Models	11
1-3c	Spreadsheet Models	12
1-3d	A Seven-Step Modeling Process	13
1-4	Conclusion	15

2-7 Excel Tables for Filtering, Sorting, and Summarizing 63

2-8 Conclusion 71

3	Finding Relationships among Variables	79
3-1	Introduction	80
3-2	Relationships among Categorical Variables	82
3-3	Relationships among Categorical Variables and a Numerical Variable	86
3-3a	Stacked and Unstacked Formats	87
3-4	Relationships among Numerical Variables	95
3-4a	Scatterplots	95
3-4b	Correlation and Covariance	101
3-5	Pivot Tables	108
3-6	Conclusion	131

PART I **EXPLORING DATA 17**

2	Describing the Distribution of a Single Variable	19
2-1	Introduction	21
2-2	Basic Concepts	22
2-2a	Populations and Samples	22
2-2b	Data Sets, Variables, and Observations	23
2-2c	Types of Data	24
2-3	Descriptive Measures for Categorical Variables	27
2-4	Descriptive Measures for Numerical Variables	30
2-4a	Numerical Summary Measures	31
2-4b	Numerical Summary Measures with StatTools	40
2-4c	Charts for Numerical Variables	45
2-5	Time Series Data	54
2-6	Outliers and Missing Values	61
2-6a	Outliers	61
2-6b	Missing Values	61

PART 2 **PROBABILITY AND DECISION MAKING UNDER UNCERTAINTY 137**

4	Probability and Probability Distributions	139
4-1	Introduction	140
4-2	Probability Essentials	142
4-2a	Rule of Complements	142
4-2b	Addition Rule	142
4-2c	Conditional Probability and the Multiplication Rule	143
4-2d	Probabilistic Independence	146
4-2e	Equally Likely Events	147
4-2f	Subjective Versus Objective Probabilities	147
4-3	Probability Distribution of a Single Random Variable	150
4-3a	Summary Measures of a Probability Distribution	151
4-3b	Conditional Mean and Variance	154
4-4	Introduction to Simulation	156
4-5	Conclusion	160

5	Normal, Binomial, Poisson, and Exponential Distributions	166
5-1	Introduction	167
5-2	The Normal Distribution	168
5-2a	Continuous Distributions and Density Functions	168
5-2b	The Normal Density	169
5-2c	Standardizing: Z-Values	170
5-2d	Normal Tables and Z-Values	172
5-2e	Normal Calculations in Excel	174
5-2f	Empirical Rules Revisited	177
5-2g	Weighted Sums of Normal Random Variables	178
5-3	Applications of the Normal Distribution	178
5-4	The Binomial Distribution	190
5-4a	Mean and Standard Deviation of the Binomial Distribution	193
5-4b	The Binomial Distribution in the Context of Sampling	194
5-4c	The Normal Approximation to the Binomial	194
5-5	Applications of the Binomial Distribution	195
5-6	The Poisson and Exponential Distributions	207
5-6a	The Poisson Distribution	207
5-6b	The Exponential Distribution	210
5-7	Conclusion	212
6	Decision Making under Uncertainty	222
6-1	Introduction	223
6-2	Elements of Decision Analysis	225
6-2a	Payoff Tables	225
6-2b	Possible Decision Criteria	226
6-2c	Expected Monetary Value (EMV)	227
6-2d	Sensitivity Analysis	228
6-2e	Decision Trees	229
6-2f	Risk Profiles	231
6-3	The PrecisionTree Add-In	239
6-4	Bayes' Rule	252
6-5	Multistage Decision Problems and the Value of Information	256
6-5a	The Value of Information	260

6-6	Risk Aversion and Expected Utility	272
6-6a	Utility Functions	273
6-6b	Exponential Utility	273
6-6c	Certainty Equivalents	277
6-6d	Is Expected Utility Maximization Used?	279
6-7	Conclusion	280

PART 3 **STATISTICAL INFERENCE 299**

7	Sampling and Sampling Distributions	301
7-1	Introduction	302
7-2	Sampling Terminology	302
7-3	Methods for Selecting Random Samples	304
7-3a	Simple Random Sampling	304
7-3b	Systematic Sampling	309
7-3c	Stratified Sampling	310
7-3d	Cluster Sampling	313
7-3e	Multistage Sampling Schemes	314
7-4	Introduction to Estimation	315
7-4a	Sources of Estimation Error	316
7-4b	Key Terms in Sampling	317
7-4c	Sampling Distribution of the Sample Mean	318
7-4d	The Central Limit Theorem	323
7-4e	Sample Size Selection	328
7-4f	Summary of Key Ideas for Simple Random Sampling	329
7-5	Conclusion	331
8	Confidence Interval Estimation	335
8-1	Introduction	336
8-2	Sampling Distributions	338
8-2a	The t Distribution	338
8-2b	Other Sampling Distributions	341
8-3	Confidence Interval for a Mean	342
8-4	Confidence Interval for a Total	348
8-5	Confidence Interval for a Proportion	351
8-6	Confidence Interval for a Standard Deviation	357

- 8-7 Confidence Interval for the Difference between Means 360
 - 8-7a Independent Samples 361
 - 8-7b Paired Samples 369
- 8-8 Confidence Interval for the Difference between Proportions 374
- 8-9 Sample Size Selection 381
 - 8-9a Sample Size Selection for Estimation of the Mean 382
 - 8-9b Sample Size Selection for Estimation of Other Parameters 384
- 8-10 Conclusion 388
- 9 Hypothesis Testing 401
 - 9-1 Introduction 402
 - 9-2 Concepts in Hypothesis Testing 403
 - 9-2a Null and Alternative Hypotheses 404
 - 9-2b One-Tailed Versus Two-Tailed Tests 405
 - 9-2c Types of Errors 405
 - 9-2d Significance Level and Rejection Region 406
 - 9-2e Significance from p -values 406
 - 9-2f Type II Errors and Power 408
 - 9-2g Hypothesis Tests and Confidence Intervals 409
 - 9-2h Practical versus Statistical Significance 409
 - 9-3 Hypothesis Tests for a Population Mean 410
 - 9-4 Hypothesis Tests for Other Parameters 418
 - 9-4a Hypothesis Tests for a Population Proportion 418
 - 9-4b Hypothesis Tests for Differences between Population Means 421
 - 9-4c Hypothesis Test for Equal Population Variances 431
 - 9-4d Hypothesis Tests for Differences between Population Proportions 432
 - 9-5 Tests for Normality 440
 - 9-6 Chi-Square Test for Independence 446
 - 9-7 Conclusion 451
- 10 Regression Analysis: Estimating Relationships 463
 - 10-1 Introduction 464
 - 10-2 Scatterplots: Graphing Relationships 467
 - 10-2a Linear versus Nonlinear Relationships 472
 - 10-2b Outliers 472
 - 10-2c Unequal Variance 473
 - 10-2d No Relationship 473
 - 10-3 Correlations: Indicators of Linear Relationships 474
 - 10-4 Simple Linear Regression 476
 - 10-4a Least Squares Estimation 476
 - 10-4b Standard Error of Estimate 483
 - 10-4c The Percentage of Variation Explained: R-Square 484
 - 10-5 Multiple Regression 487
 - 10-5a Interpretation of Regression Coefficients 488
 - 10-5b Interpretation of Standard Error of Estimate and R-Square 491
 - 10-6 Modeling Possibilities 494
 - 10-6a Dummy Variables 494
 - 10-6b Interaction Variables 501
 - 10-6c Nonlinear Transformations 505
 - 10-7 Validation of the Fit 520
 - 10-8 Conclusion 522
- 11 Regression Analysis: Statistical Inference 532
 - 11-1 Introduction 534
 - 11-2 The Statistical Model 534
 - 11-3 Inferences About the Regression Coefficients 538
 - 11-3a Sampling Distribution of the Regression Coefficients 539

11-3b Hypothesis Tests for the Regression Coefficients and p -Values	541
11-3c A Test for the Overall Fit: The ANOVA Table	542
11-4 Multicollinearity	547
11-5 Include/Exclude Decisions	551
11-6 Stepwise Regression	556
11-7 Outliers	561
11-8 Violations of Regression Assumptions	566
11-8a Nonconstant Error Variance	566
11-8b Nonnormality of Residuals	567
11-8c Autocorrelated Residuals	568
11-9 Prediction	570
11-10 Conclusion	576
12 Time Series Analysis and Forecasting	590
12-1 Introduction	591
12-2 Forecasting Methods: An Overview	592
12-2a Extrapolation Models	592
12-2b Econometric Models	593
12-2c Combining Forecasts	594
12-2d Components of Time Series Data	594
12-2e Measures of Accuracy	597
12-3 Testing for Randomness	599
12-3a The Runs Test	601
12-3b Autocorrelation	604
12-4 Regression-Based Trend Models	607
12-4a Linear Trend	607
12-4b Exponential Trend	610
12-5 The Random Walk Model	615
12-6 Moving Averages Forecasts	619
12-7 Exponential Smoothing Forecasts	624
12-7a Simple Exponential Smoothing	624
12-7b Holt's Model for Trend	629
12-8 Seasonal Models	634
12-8a Winters' Exponential Smoothing Model	635
12-8b Deseasonalizing: The Ratio-to-Moving-Averages Method	639
12-8c Estimating Seasonality with Regression	643
12-9 Conclusion	648
13 Introduction to Optimization Modeling	657
13-1 Introduction	658
13-2 Introduction to Optimization	659
13-3 A Two-Variable Product Mix Model	660
13-4 Sensitivity Analysis	672
13-4a Solver's Sensitivity Report	673
13-4b SolverTable Add-In	676
13-4c Comparison of Solver's Sensitivity Report and SolverTable	682
13-5 Properties of Linear Models	683
13-5a Proportionality	683
13-5b Additivity	683
13-5c Divisibility	684
13-5d Discussion of Linear Properties	684
13-5e Linear Models and Scaling	685
13-6 Infeasibility and Unboundedness	685
13-6a Infeasibility	685
13-6b Unboundedness	686
13-6c Comparison of Infeasibility and Unboundedness	686
13-7 A Larger Product Mix Model	688
13-8 A Multiperiod Production Model	696
13-9 A Comparison of Algebraic and Spreadsheet Models	704
13-10 A Decision Support System	705
13-11 Conclusion	707
14 Optimization Models	718
14-1 Introduction	719
14-2 Worker Scheduling Models	720
14-3 Blending Models	727
14-4 Logistics Models	733
14-4a Transportation Models	734
14-4b Other Logistics Models	742
14-5 Aggregate Planning Models	750
14-6 Financial Models	759
14-7 Integer Optimization Models	770
14-7a Capital Budgeting Models	770

14-7b Fixed-Cost Models	775	16-3b Cash Balance Models	901
14-7c Set-Covering Models	782	16-3c Investment Models	906
14-8 Nonlinear Optimization Models	789	16-4 Marketing Models	912
14-8a Basic Ideas of Nonlinear Optimization	789	16-4a Models of Customer Loyalty	913
14-8b Managerial Economics Models	790	16-4b Marketing and Sales Models	921
14-8c Portfolio Optimization Models	793	16-5 Simulating Games of Chance	927
14-9 Conclusion	802	16-5a Simulating the Game of Craps	927
		16-5b Simulating the NCAA Basketball Tournament	930
15 Introduction to Simulation Modeling	812	16-6 An Automated Template for @Risk Models	934
15-1 Introduction	813	16-7 Conclusion	935
15-2 Probability Distributions for Input Variables	815		
15-2a Types of Probability Distributions	816		
15-2b Common Probability Distributions	819		
15-2c Using @RISK to Explore Probability Distributions	824		
15-3 Simulation and the Flaw of Averages	833		
15-4 Simulation with Built-in Excel Tools	836		
15-5 Introduction to the @RISK	847		
15-5a @RISK Features	847		
15-5b Loading @RISK	848		
15-5c @RISK Models with a Single Random Input Variable	848		
15-5d Some Limitations of @RISK	857		
15-5e @RISK Models with Several Random Input Variables	858		
15-6 The Effects of Input Distributions on Results	863		
15-6a Effect of the Shape of the Input Distribution(s)	863		
15-6b Effect of Correlated Input Variables	866		
15-7 Conclusion	871		
16 Simulation Models	880		
16-1 Introduction	882		
16-2 Operations Models	882		
16-2a Bidding for Contracts	882		
16-2b Warranty Costs	886		
16-2c Drug Production with Uncertain Yield	890		
16-3 Financial Models	896		
16-3a Financial Planning Models	897		

PART 6

ADVANCED DATA ANALYSIS 945

17 Data Mining 947

17-1 Introduction	948
17-2 Data Exploration and Visualization	949
17-2a Online Analytical Processing (OLAP)	950
17-2b PowerPivot and Power View in Excel 2013	954
17-2c Visualization Software	962
17-3 Microsoft Data Mining Add-Ins for Excel	962
17-4 Classification Methods	964
17-4a Logistic Regression	965
17-4b Neural Networks	971
17-4c Classification Trees	976
17-4d Classification and Lift	977
17-4e Classification with Rare Events	978
17-5 Clustering	980
17-6 Conclusion	982

PART 7

BONUS ONLINE MATERIAL

18 Importing Data into Excel 18-3

18-1 Introduction	18-4
18-2 Rearranging Excel Data	18-5

18-3	Importing Text Data	18-9	20-3	Introduction to Control Charts	20-7
18-4	Importing Relational Database Data	18-15	20-4	Control Charts for Variables	20-9
18-4a	Introduction to Relational Databases	18-15	20-4a	Control Charts and Hypothesis Testing	20-14
18-4b	Using Microsoft Query	18-16	20-4b	Other Out-of-Control Indications	20-16
18-4c	SQL Statements	18-28	20-4c	Rational Subsamples	20-17
18-5	Web Queries	18-30	20-4d	Deming's Funnel Experiment and Tampering	20-19
18-6	Cleansing Data	18-34	20-4e	Control Charts in the Service Industry	20-23
18-7	Conclusion	18-41	20-5	Control Charts for Attributes	20-27
19	Analysis of Variance and Experimental Design	19-1	20-5a	The p Chart	20-27
19-1	Introduction	19-2	20-5b	The Red Bead Experiment	20-31
19-2	One-Way ANOVA	19-5	20-6	Process Capability	20-34
19-2a	The Equal-Means Test	19-5	20-6a	Process Capability Indexes	20-37
19-2b	Confidence Intervals for Differences between Means	19-8	20-6b	More on Motorola and 6-sigma	20-42
19-2c	Using a Logarithmic Transformation	19-11	20-7	Conclusion	20-45
19-3	Using Regression to Perform ANOVA	19-17	Appendix A: Statistical Reporting	A-1	
19-4	The Multiple Comparison Problem	19-20	A-1	Introduction	A-1
19-5	Two-Way ANOVA	19-24	A-2	Suggestions for Good Statistical Reporting	A-2
19-5a	Confidence Intervals for Contrasts	19-31	A-2a	Planning	A-2
19-5b	Assumptions of Two-Way ANOVA	19-34	A-2b	Developing a Report	A-3
19-6	More About Experimental Design	19-35	A-2c	Be Clear	A-4
19-6a	Randomization	19-36	A-2d	Be Concise	A-5
19-6b	Blocking	19-38	A-2e	Be Precise	A-5
19-6c	Incomplete Designs	19-42	A-3	Examples of Statistical Reports	A-7
19-7	Conclusion	19-45	A-4	Conclusion	A-18
20	Statistical Process Control	20-1	References	985	
20-1	Introduction	20-3	Index	987	
20-2	Deming's 14 Points	20-4			

Preface

With today's technology, companies are able to collect tremendous amounts of data with relative ease. Indeed, many companies now have more data than they can handle. However, the data are usually meaningless until they are analyzed for trends, patterns, relationships, and other useful information. This book illustrates in a practical way a variety of methods, from simple to complex, to help you analyze data sets and uncover important information. In many business contexts, data analysis is only the first step in the solution of a problem. Acting on the solution and the information it provides to make good decisions is a critical next step. Therefore, there is a heavy emphasis throughout this book on analytical methods that are useful in decision making. Again, the methods vary considerably, but the objective is always the same—to equip you with decision-making tools that you can apply in your business careers.

We recognize that the majority of students in this type of course are *not* majoring in a quantitative area. They are typically business majors in finance, marketing, operations management, or some other business discipline who will need to analyze data and make quantitative-based decisions in their jobs. We offer a hands-on, example-based approach and introduce fundamental concepts as they are needed. Our vehicle is spreadsheet software—specifically, Microsoft Excel®. This is a package that most students already know and will almost surely use in their careers. Our MBA students at Indiana University are so turned on by the required course that is based on this book that *almost all* of them (mostly finance and marketing majors) take at least one of our follow-up elective courses in spreadsheet modeling. We are convinced that students see value in quantitative analysis when the course is taught in a practical and example-based approach.

Rationale for Writing This Book

Business Analytics: Data Analysis and Decision Making is different from the many fine textbooks written for statistics and management science. Our rationale for writing this book is based on three fundamental objectives.

- **Integrated coverage and applications.** The book provides a unified approach to

business-related problems by integrating methods and applications that have been traditionally taught in separate courses, specifically statistics and management science.

- **Practical in approach.** The book emphasizes realistic business examples and the processes managers actually use to analyze business problems. The emphasis is *not* on abstract theory or computational methods.
- **Spreadsheet-based teaching.** The book provides students with the skills to analyze business problems with tools they have access to and will use in their careers. To this end, we have adopted Excel and commercial spreadsheet add-ins.

Integrated Coverage and Applications

In the past, many business schools, including ours at Indiana University, have offered a required statistics course, a required decision-making course, and a required management science course—or some subset of these. A current trend, however, is to have only one required course that covers the basics of statistics, some regression analysis, some decision making under uncertainty, some linear programming, some simulation, and possibly others. Essentially, faculty in the quantitative area get one opportunity to teach all business students, so we attempt to cover a variety of useful quantitative methods. We are not necessarily arguing that this trend is ideal, but rather that it is a reflection of the reality at our university and, we suspect, at many others. After several years of teaching this course, we have found it to be a great opportunity to attract students to the subject and to more advanced study.

The book is also integrative in another important aspect. It not only integrates a number of analytical methods, but it also applies them to a wide variety of business problems—that is, it analyzes realistic examples from many business disciplines. We include examples, problems, and cases that deal with portfolio optimization, workforce scheduling, market share analysis, capital budgeting, new product analysis, and many others.

Practical in Approach

This book has been designed to be very example-based and practical. We strongly believe that students learn

best by working through examples, and they appreciate the material most when the examples are realistic and interesting. Therefore, our approach in the book differs in two important ways from many competitors. First, there is just enough conceptual development to give students an understanding and appreciation for the issues raised in the examples. We often introduce important concepts, such as standard deviation as a measure of variability, in the context of examples rather than discussing them in the abstract. Our experience is that students gain greater intuition and understanding of the concepts and applications through this approach.

Second, we place virtually no emphasis on hand calculations. We believe it is more important for students to understand why they are conducting an analysis and what it means than to emphasize the tedious calculations associated with many analytical techniques. Therefore, we illustrate how powerful software can be used to create graphical and numerical outputs in a matter of seconds, freeing the rest of the time for in-depth interpretation of the results, sensitivity analysis, and alternative modeling approaches. In our own courses, we move directly into a discussion of examples, where we focus almost exclusively on interpretation and modeling issues, and we let the software perform the number crunching.

Spreadsheet-based Teaching

We are strongly committed to teaching spreadsheet-based, example-driven courses, regardless of whether the basic area is data analysis or management science. We have found tremendous enthusiasm for this approach, both from students and from faculty around the world who have used our books. Students learn and remember more, and they appreciate the material more. In addition, instructors typically enjoy teaching more, and they usually receive immediate reinforcement through better teaching evaluations. We were among the first to move to spreadsheet-based teaching about two decades ago, and we have never regretted the move.

What We Hope to Accomplish in This Book

Condensing the ideas in the previous paragraphs, we hope to:

- reverse negative student attitudes about statistics and quantitative methods by making these topics real, accessible, and interesting;

- give students plenty of hands-on experience with real problems and challenge them to develop their intuition, logic, and problem-solving skills;
- expose students to real problems in many business disciplines and show them how these problems can be analyzed with quantitative methods; and
- develop spreadsheet skills, including experience with powerful spreadsheet add-ins, that add immediate value to students' other courses and for their future careers.

New in the Fifth Edition

There are several important changes in this edition.

- **Business Analytics Focus:** The term *business analytics* is now in the title of the book. This term is now ubiquitous in our area, being extremely popular with both academics and business people, and we are pleased that our methods are now so highly regarded. However, an argument can certainly be made that the methods in this book, methods we have taught for years, are essentially what the field of Business Analytics is all about. Granted, a major focus of Business Analytics is on using *large* data sets to help make good decisions. Therefore, we have added the chapter discussed in the next bullet.
- **Data Mining Coverage:** Chapter 17 on Data Mining is completely new. This topic, which we were hardly aware of just 15 years ago, is now an extremely important part of Business Analytics. Part of the reason for this is that virtually all companies have access to large data sets that they need to make sense of. Another part of the reason is that there are now powerful algorithms and accompanying software packages to perform the analyses. The topic itself is huge, with many books of various mathematical levels devoted entirely to it, and it is impossible to cover it adequately in a single chapter. However, we believe this new chapter provides an exciting, hands-on introduction to data mining.
- **Content Reductions:** Because of the combination of statistical and management science topics, the book has always been long. With the addition of the data mining chapter, it got even longer, and something had to be cut. Therefore, parts of various chapters, particularly Chapter 4 (Probability and Probability Distributions), have been deleted. Based on user surveys, we tried to

cut only the material that was used least often, but if any of your favorite sections are missing, you can contact Cengage or the authors, and we will provide you with PDF versions from the fourth edition. A complete list of deleted sections follows:¹

- Section 3.6: An Extended Example
- Section 4.5: Probability Distributions of Two Random Variables: Scenario Approach
- Section 4.6: Probability Distribution of Two Random Variables: Joint Probability Approach
- Section 4.7: Independent Random Variables (this is now discussed briefly in Section 4.2.4)
- Section 4.8: Weighted Sums of Random Variables
- Section 5.7: Fitting a Probability Distribution to Data with @RISK
- Section 9.7: One-Way ANOVA (this is now discussed extensively in the online Chapter 19: ANOVA and Experimental Design)
- Section 11.7: The Partial F Test
- Section 12.6: Autoregression Models
- **Advanced Statistical Coverage:** Three online chapters are available: Chapter 18 (Importing Data into Excel), Chapter 19 (Analysis of Variance and Experimental Design), and Chapter 20 (Statistical Process Control). The latter two of these are new online chapters, updated versions of “advanced statistical” chapters that appeared in our old *Data Analysis for Managers* book.
- **Tutorial Videos:** To help students learn, we created more than 50 videos that explain concepts and work through examples. Students can access the videos for free on the textbook companion website. A complete list of videos is available on the website and video icons appear in the margins of the textbook next to relevant topics.
- **Updated to Office 2013:** As we were creating this edition of the book, Microsoft released Office 2013. Therefore, all of the screenshots in the book are from this newer version. However, the changes from 2010 to 2013 (or even from 2007 to 2013) are not that extensive, at least for our

¹ These section numbers refer to the numbering in the previous (fourth) edition.

purposes. Therefore, if you are still using Office 2007 or 2010, this book should work fine for you.

- **Updated Problems:** As in the previous edition, we modified a number of problems, usually to update their data, and we added several brand-new problems. We also created a file, essentially a database of problems, which is available to instructors. This file, **DADM 5e Problem Database.xlsx**, indicates the context of each of the problems, and it also shows the correspondence between problems in this edition and problems in the previous edition.
- **Suggested Solutions for Conceptual Questions:** In addition to Problems, each chapter also includes Conceptual Questions to assess student comprehension of concepts on a broad level. New for this edition, suggested solutions for these Conceptual Questions have been provided.

Software

This book is based entirely on Microsoft Excel, the spreadsheet package that has become the standard analytical tool in business. Excel is an extremely powerful package, and one of our goals is to convert *casual* users into *power* users who can take full advantage of its features. If you learn no more than this, you will be acquiring a valuable skill for the business world. However, Excel has some limitations. Therefore, this book includes several Excel add-ins that greatly enhance Excel’s capabilities. As a group, these add-ins comprise what is arguably the most impressive assortment of spreadsheet-based software accompanying any book on the market.

DecisionTools® Suite Add-in

The textbook website for *Business Analytics: Data Analysis and Decision Making* provides a link to the powerful DecisionTools® Suite by Palisade Corporation. (The version available is compatible with Excel 2013 and previous versions of Excel.) This suite includes six separate add-ins:

- **@RISK**, an add-in for simulation
- **StatTools**, an add-in for statistical data analysis
- **PrecisionTree**, a graphical-based add-in for creating and analyzing decision trees
- **TopRank**, an add-in for performing what-if analyses

- **NeuralTools®**, an add-in for estimating complex, nonlinear relationships
- **Evolver™**, an add-in for performing optimization (an alternative to Excel's Solver)

We use @RISK and PrecisionTree extensively in the chapters on simulation and decision making under uncertainty, and we use StatTools extensively in the data analysis chapters.

Online access to the DecisionTools Suite, available with new copies of the book, is an academic version, slightly scaled down from the professional version that sells for hundreds of dollars and is used by many leading companies. It functions for two years when properly installed, and it puts only modest limitations on the size of data sets or models that can be analyzed.²

SolverTable Add-in

We also include SolverTable, a supplement to Excel's built-in Solver for optimization.³ If you have ever had difficulty understanding Solver's sensitivity reports, you will appreciate SolverTable. It works like Excel's data tables, except that for each input (or pair of inputs), the add-in runs Solver and reports the *optimal* output values. SolverTable is used extensively in the optimization chapters.

Windows versus Mac

In our own courses, we have seen an increasing number of students using Macintosh laptops rather than Windows laptops. Fortunately, this is *not* a problem, and our students have followed along fine with their Macs. However, these students should be advised to use a Windows emulation program (Bootcamp or Parallels are good candidates), along with Office for *Windows*. If they use Office for the Mac, they are bound to be confused, and there is no guarantee that the add-ins used throughout the book will work. In fact, the Palisade add-ins are guaranteed *not* to work.

Potential Course Structures

Although we have used the book for our own required one-semester course, there is admittedly much more material than can be covered adequately in one

² Visit www.kelley.iu.edu/albrightbooks for specific details on these limitations.

³ Although SolverTable is available on this textbook's website, it is also available for free from Albright's website, www.kelley.iu.edu/albrightbooks.

semester. We have tried to make the book as modular as possible, allowing an instructor to cover, say, simulation before optimization or vice-versa, or to omit either of these topics. The one exception is statistics. Due to the natural progression of statistical topics, the basic topics in the early chapters should be covered before the more advanced topics (regression and time series analysis) in the later chapters. With this in mind, there are several possible ways to cover the topics.

- **One-semester Required Course, with No Statistics Prerequisite** (or where MBA students need a refresher for whatever statistics they learned previously): If data analysis is the primary focus of the course, then Chapters 2–5, 7–11, and possibly Chapter 17 should be covered. Depending on the time remaining, any of the topics in Chapters 6 (decision making under uncertainty), 12 (time series analysis), 13–14 (optimization), or 15–16 (simulation) can be covered in practically any order.
- **One-semester Required Course, with a Statistics Prerequisite**: Assuming that students know the basic elements of statistics (up through hypothesis testing), the material in Chapters 2–5 and 7–9 can be reviewed quickly, primarily to illustrate how Excel and add-ins can be used to do the number crunching. The instructor can then choose among any of the topics in Chapters 6, 10–11, 12, 13–14, or 15–16 (in practically any order), or to fill the remainder of the course.
- **Two-semester Required Sequence**: Given the luxury of spreading the topics over two semesters, the entire book, or at least most of it, can be covered. The statistics topics in Chapters 2–5 and 7–9 should be covered in chronological order before other statistical topics (regression and time series analysis), but the remaining chapters can be covered in practically any order.

Custom Publishing

Cengage Learning is dedicated to making the educational experience unique for all learners by creating custom materials that best suit your course needs. With Cengage Learning you can create a custom solution where you have the ability to choose your book's content, length, sequence, even the cover design. You may combine content from multiple Cengage Learning titles and add other materials, including your own original work, to create your ideal customized

text. If you would like to learn more about our custom publishing services, please contact your Cengage Learning representative⁴ or visit us at www.cengage.com/custom.

Instructor Supplements

Textbook Website: cengage.com/login

The companion website provides immediate access to an array of teaching resources—including data and solutions files for all of the Examples, Problems, and Cases in the book, Chapters 18–20 and Appendix A, Test Bank files, PowerPoint slides, and access to the DecisionTools® Suite by Palisade Corporation and the SolverTable add-in. Also, new for this edition, the website features more than 50 videos that explain concepts and work through examples. You can easily download the instructor resources you need from the password-protected, instructor-only section of the site.

Test Bank

Cengage Learning Testing Powered by Cognero is a flexible, online system that allows you to:

- author, edit, and manage test bank content from multiple Cengage Learning solutions
- create multiple test versions in an instant
- deliver tests from your LMS, your classroom, or wherever you want

Student Supplements

Textbook Website: www.cengagebrain.com

Every new student edition of this book comes with access to the *Business Analytics: Data Analysis and Decision Making, 5e* textbook website that links to the following files and tools:

- Excel files for the examples in the chapters (usually two versions of each—a template, or data-only version, and a finished version)
- Data files required for the Problems and Cases
- **excel_tutorial.xlsx**, which contains a useful tutorial for getting up to speed in Excel
- Chapters 18–20 and Appendix A

- DecisionTools® Suite software by Palisade Corporation (described earlier)
- SolverTable add-in
- More than 50 tutorial videos that explain concepts and work through examples

To access resources, go to www.cengagebrain.com, search by ISBN 9781133629603, click on the “Free Materials” tab, and select “Access Now.” The resources you need will be listed both per chapter (by selecting a chapter from the drop-down list) and for the entire book (under Book Resources).

Student Solutions

Student Solutions to many of the problems (indicated in the text with a colored box around the problem number) are available in Excel format. You can purchase access to Student Solutions files by going to www.cengagebrain.com, searching by ISBN 9781285871332, and adding the product to your cart.

Acknowledgements

The authors would like to thank several people who helped make this book a reality. First, the authors are indebted to Peter Kolesar, Mark Broadie, Lawrence Lapin, and William Whisler for contributing some of the excellent case studies that appear throughout the book.

We are also grateful to many of the professionals who worked behind the scenes to make this book a success: Joe Sabatino, Product Director; Krista Kellman, Content Developer; Holly Henjum, Senior Content Project Manager; Roy Rosa, Associate Market Development Manager; and Product Assistant, Brad Sullender.

We also extend our sincere appreciation to the reviewers who provided feedback on the authors’ proposed changes that resulted in this fifth edition:

John Aloysius, *Walton College of Business, University of Arkansas*

Henry F. Ander, *Arizona State University*

Dr. Baabak Ashuri, *School of Building Construction, Georgia Institute of Technology*

James Behel, *Harding University*

Robert H. Burgess, *Scheller College of Business, Georgia Institute of Technology*

Paul Damien, *McCombs School of Business, University of Texas in Austin*

Parviz Ghandforoush, *Virginia Tech*

⁴ Find your Learning Consultant at sites.cengage.com/repfinder.

Betsy Greenberg, *University of Texas*
Anissa Harris, *Harding University*
Tim James, *Arizona State University*
Norman Johnson, *C.T. Bauer College of Business,
University of Houston*
Shivraj Kanungo, *The George Washington University*
Miguel Lejeune, *The George Washington University*
José Lobo, *Arizona State University*
Stuart Low, *Arizona State University*

Lance Matheson, *Virginia Tech*
Patrick R. McMullen, *Wake Forest University*
Barbara A. Price, PhD, *Georgia Southern University*
Laura Wilson-Gentry, *University of Baltimore*
Toshiyuki Yuasa, *University of Houston*

S. Christian Albright

Wayne L. Winston

August 2013

Introduction to Data Analysis and Decision Making



Konstantin Chaginy/Shutterstock.com

HOTTEST NEW JOBS: STATISTICS AND MATHEMATICS

Much of this book, as the title implies, is about data analysis. The term *data analysis* has long been synonymous with the term *statistics*, but in today's world, with massive amounts of data available in business and many other fields such as health and science, data analysis goes beyond the more narrowly focused area of traditional statistics. But regardless of what it is called, data analysis is currently a hot topic and promises to get even hotter in the future. The data analysis skills you learn here, and possibly in follow-up quantitative courses, might just land you a very interesting and lucrative job.

This is exactly the message in a recent *New York Times* article, "For Today's Graduate, Just One Word: Statistics," by Steve Lohr. (A similar article, "Math Will Rock Your World," by Stephen Baker, was the cover story for *BusinessWeek*. Both articles are available online by searching for their titles.) The statistics article begins by chronicling a Harvard anthropology and archaeology graduate, Carrie Grimes, who began her career by mapping the locations of Mayan artifacts in places like Honduras. As she states, "People think of field archaeology as Indiana Jones, but much of what you really do is data analysis." Since then, Grimes has leveraged her data analysis skills to get a job with Google, where she and many other people with a quantitative background are analyzing huge amounts of data to improve the company's search engine. As the chief economist at Google, Hal Varian, states, "I keep saying that the sexy job in the next 10 years will be statisticians. And I'm not kidding." The salaries for statisticians with doctoral degrees currently start at \$125,000, and they will probably continue to increase. (The math article indicates that mathematicians are also in great demand.)

Why is this trend occurring? The reason is the explosion of digital data—data from sensor signals, surveillance tapes, Web clicks, bar scans, public records, financial transactions, and more. In years past, statisticians typically analyzed relatively small data sets, such as opinion polls with about 1000 responses. Today's massive data sets require new statistical methods, new computer software, and most importantly for you, more young people trained in these methods and the corresponding software. Several particular areas mentioned in the articles include (1) improving Internet search and online advertising, (2) unraveling gene sequencing information for cancer research, (3) analyzing sensor and location data for optimal handling of food shipments, and (4) the recent Netflix contest for improving the company's recommendation system.

The statistics article mentions three specific organizations in need of data analysts. The first is government, where there is an increasing need to sift through mounds of data as a first step toward dealing with long-term economic needs and key policy priorities. The second is IBM, which created a Business Analytics and Optimization Services group in April 2009. This group will use the more than 200 mathematicians, statisticians, and data analysts already employed by the company, but IBM intends to retrain or hire 4000 more analysts to meet its needs. The third is Google, which needs more data analysts to improve its search engine. You may think that today's search engines are unbelievably efficient, but Google knows they can be improved. As Ms. Grimes states, "Even an improvement of a percent or two can be huge, when you do things over the millions and billions of times we do things at Google."

Of course, these three organizations are not the only organizations that need to hire more skilled people to perform data analysis and other analytical procedures. It is a need faced by *all* large organizations. Various recent technologies, the most prominent by far being the Web, have given organizations the ability to gather massive amounts of data easily. Now they need people to make sense of it all and use it to their competitive advantage. ■

1-1 INTRODUCTION

We are living in the age of technology. This has two important implications for everyone entering the business world. First, technology has made it possible to collect huge amounts of data. Retailers collect point-of-sale data on products and customers every time a transaction occurs; credit agencies have all sorts of data on people who have or would like to obtain credit; investment companies have a limitless supply of data on the historical patterns of stocks, bonds, and other securities; and government agencies have data on economic trends, the environment, social welfare, consumer product safety, and virtually everything else imaginable. It has become relatively *easy* to collect the data. As a result, data are plentiful. However, as many organizations are now beginning to discover, it is quite a challenge to make sense of all the data they have collected.

A second important implication of technology is that it has given many more people the power and responsibility to analyze data and make decisions on the basis of quantitative analysis. People entering the business world can no longer pass all of the quantitative analysis to the "quant jocks," the technical specialists who have traditionally done the number crunching. The vast majority of employees now have a desktop or laptop computer at their disposal, access to relevant data, and training in easy-to-use software, particularly spreadsheet and database software. For these employees, statistics and other quantitative methods are no longer forgotten topics they once learned in college. Quantitative analysis is now an integral part of their daily jobs.

A large amount of data already exists, and it will only increase in the future. Many companies already complain of swimming in a sea of data. However, enlightened companies are seeing this expansion as a source of competitive advantage. In fact, one of the hottest topics in today's business world is **business analytics**. This term has been created to encompass all of the types of analysis discussed in this book, so it isn't really new; we have been teaching it for years. The new aspect of business analytics is that it typically implies the analysis of very *large* data sets, the kind that companies currently encounter. By using quantitative methods to uncover the *information* in these data sets and then acting on this information—again guided by quantitative analysis—companies are able to gain advantages that their less enlightened competitors are not able to gain. Here are several pertinent examples.

- Direct marketers analyze enormous customer databases to see which customers are likely to respond to various products and types of promotions. Marketers can then target different classes of customers in different ways to maximize profits—and give their customers what they want.
- Hotels and airlines also analyze enormous customer databases to see what their customers want and are willing to pay for. By doing this, they have been able to devise very clever pricing strategies, where different customers pay different prices for the same accommodations. For example, a business traveler typically makes a plane reservation closer to the time of travel than a vacationer. The airlines know this. Therefore, they reserve seats for these business travelers and charge them a higher price for the same seats. The airlines profit from clever pricing strategies, and the customers are happy.
- Financial planning services have a virtually unlimited supply of data about security prices, and they have customers with widely differing preferences for various types of investments. Trying to find a match of investments to customers is a very challenging problem. However, customers can easily take their business elsewhere if good decisions are not made on their behalf. Therefore, financial planners are under extreme competitive pressure to analyze masses of data so that they can make informed decisions for their customers.¹
- We all know about the pressures U.S. manufacturing companies have faced from foreign competition in the past couple of decades. The automobile companies, for example, have had to change the way they produce and market automobiles to stay in business. They have had to improve quality and cut costs by orders of magnitude. Although the struggle continues, much of the success they have had can be attributed to data analysis and wise decision making. Starting on the shop floor and moving up through the organization, these companies now measure almost everything, analyze these measurements, and then act on the results of their analysis.

We talk about companies analyzing data and making decisions. However, *companies* don't really do this; *people* do it. And who will these people be in the future? They will be *you!* We know from experience that students in all areas of business, at both the undergraduate and graduate level, will be *required* to describe large complex data sets, run regression analyses, make quantitative forecasts, create optimization models, and run simulations. You are the person who will be analyzing data and making important decisions to help

¹For a great overview of how quantitative techniques have been used in the financial world, read the book *The Quants*, by Scott Patterson (Random House, 2010). It describes how quantitative models made millions for a lot of bright young analysts, but it also describes the dangers of relying totally on quantitative models, at least in the complex world of global finance.

your company gain a competitive advantage. And if you are *not* willing or able to do so, there will be plenty of other technically trained people who will be more than happy to replace you.

The goal of this book is to teach you how to use a variety of quantitative methods to analyze data and make decisions in a very hands-on way. We discuss a number of quantitative methods and illustrate their use in a large variety of realistic business situations. As you will see, this book includes many examples from finance, marketing, operations, accounting, and other areas of business. To analyze these examples, we take advantage of the Microsoft Excel[®] spreadsheet software, together with a number of powerful Excel add-ins. In each example we provide step-by-step details of the method and its implementation in Excel.

This is *not* a “theory” book. It is also not a book where you can lean comfortably back in your chair and read about how *other* people use quantitative methods. It is a “get your hands dirty” book, where you will learn best by actively following the examples throughout the book on your own PC. By the time you have finished, you will have acquired some very useful skills for today’s business world.

1-2 OVERVIEW OF THE BOOK

This book is packed with quantitative methods and examples, probably more than can be covered in any single course. Therefore, we purposely intend to keep this introductory chapter brief so that you can get on with the analysis. Nevertheless, it is useful to introduce the methods you will be learning and the tools you will be using. This section provides an overview of the methods covered in this book and the software that is used to implement them. Then the next section presents a brief discussion of models and the modeling process. The primary goal at this point is to stimulate your interest in what follows.

1-2a The Methods

This book is rather unique in that it combines topics from two separate fields: statistics and management science. Statistics is the study of data analysis, whereas management science is the study of model building, optimization, and decision making. In the academic arena these two fields have traditionally been separated, sometimes widely. Indeed, they are often housed in separate academic departments. However, from a user’s standpoint it makes little sense to separate them. Both are useful in accomplishing what the title of this book promises: data analysis and decision making.

Therefore, we do not distinguish between the statistics and the management science parts of this book. Instead, we view the entire book as a collection of useful quantitative methods that can be used to analyze data and help make business decisions. In addition, our choice of software helps to integrate the various topics. By using a single package, Excel, together with a number of add-ins, you will see that the methods of statistics and management science are similar in many important respects. Most importantly, their combination gives you the power and flexibility to solve a wide range of business problems.

Three important themes run through this book. Two of them are in the title: data analysis and decision making. The third is *dealing with uncertainty*.² Each of these themes has subthemes. Data analysis includes data description, data inference, and the search for relationships in data. Decision making includes *optimization* techniques for problems with no uncertainty, *decision analysis* for problems with uncertainty, and structured *sensitivity*

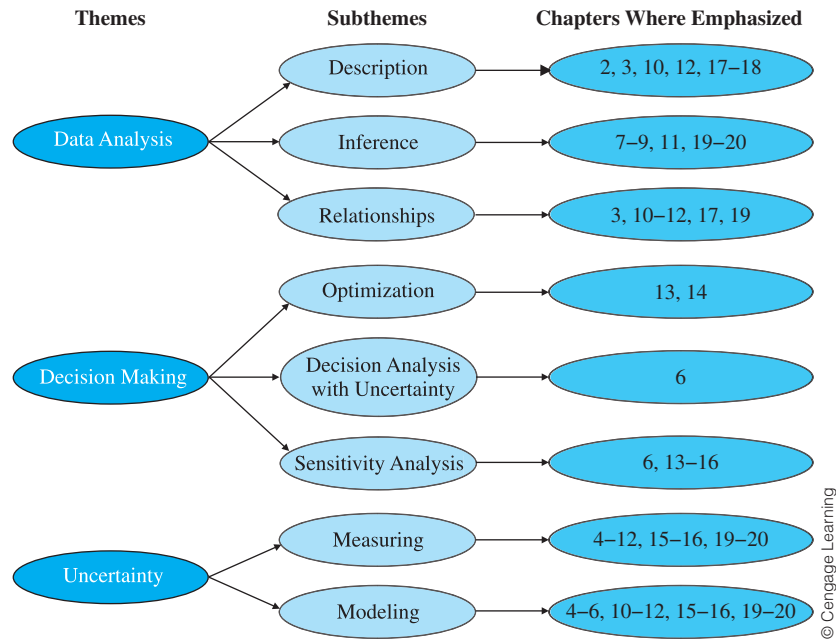
²The fact that the uncertainty theme did not find its way into the title of this book does not detract from its importance. We just wanted to keep the title reasonably short!

analysis. Dealing with uncertainty includes measuring uncertainty and modeling uncertainty explicitly. There are obvious overlaps between these themes and subthemes. When you make inferences from data and search for relationships in data, you must deal with uncertainty. When you use *decision trees* to help make decisions, you must deal with uncertainty. When you use *simulation models* to help make decisions, you must deal with uncertainty, and then you often make inferences from the simulated data.

Figure 1.1 shows where these themes and subthemes are discussed in the book. The next few paragraphs discuss the book's contents in more detail.

Figure 1.1

Themes and Subthemes



We begin in Chapters 2 and 3 by illustrating a number of ways to summarize the information in data sets. These include graphical and tabular summaries, as well as numerical summary measures such as means, medians, and standard deviations. The material in these two chapters is elementary from a mathematical point of view, but it is extremely important. As stated earlier, organizations are now able to collect huge amounts of raw data, but what does it all mean? Although there are very sophisticated methods for analyzing data, some of which are covered in later chapters, the “simple” methods in Chapters 2 and 3 are crucial for obtaining an initial understanding of the data. Fortunately, Excel and available add-ins now make this quite easy. For example, Excel’s pivot table tool for “slicing and dicing” data is an analyst’s dream come true. You will be amazed by the insights you can gain from pivot tables—with very little effort.

Uncertainty is a key aspect of most business problems. To deal with uncertainty, you need a basic understanding of probability. We discuss the key concepts in Chapters 4 and 5. Chapter 4 covers basic rules of probability and then discusses the extremely important concept of probability distributions. Chapter 5 follows up this discussion by focusing on two of the most important probability distributions, the normal and binomial distributions.

One of the best ways to make probabilistic concepts easier to understand is by using computer simulation. Therefore, simulation is a common theme that runs through this book, beginning in Chapter 4. Although the final Chapters 15 and 16 are devoted entirely to simulation, we do not hesitate to use simulation early and often to illustrate statistical concepts.

In Chapter 6 we apply probability to decision making under uncertainty. These types of problems—faced by all companies on a continual basis—are characterized by the need to make a decision *now*, even though important information, such as demand for a product or returns from investments, will not be known until later. The methods in Chapter 6 provide a rational basis for making such decisions. These methods do not guarantee perfect outcomes—the future could unluckily turn out differently than expected—but they do enable decision makers to proceed rationally and make the best of the given circumstances. Additionally, the software used to implement these methods allows decision makers, with very little extra work, to see how sensitive the optimal decisions are to inputs. This is crucial, because the inputs to many business problems are, at best, educated guesses. Finally, we examine the role of risk aversion in these types of decision problems.

In Chapters 7, 8, and 9 we discuss sampling and statistical inference. Here the basic problem is to estimate one or more characteristics of a population. If it is too expensive or time-consuming to learn about the *entire* population—and it usually is—it is instead common to select a random sample from the population and then use the information in the sample to *infer* the characteristics of the population. You see this continually on news shows that describe the results of various polls. You also see it in many business contexts. For example, auditors typically sample only a fraction of a company's records. Then they infer the characteristics of the entire population of records from the results of the sample to conclude whether the company has been following acceptable accounting standards.

In Chapters 10 and 11 we discuss the extremely important topic of regression analysis, which is used to study relationships between variables. The power of regression analysis is its generality. Every part of a business has variables that are related to one another, and regression can often be used to estimate relationships between these variables. In managerial accounting, regression is used to estimate how overhead costs depend on direct labor hours and production volume. In marketing, regression is used to estimate how sales volume depends on advertising and other marketing variables. In finance, regression is used to estimate how the return of a stock depends on the “market” return. In real estate studies, regression is used to estimate how the selling price of a house depends on the assessed valuation of the house and characteristics such as the number of bedrooms and square footage.

From regression, we move to time series analysis and forecasting in Chapter 12. This topic is particularly important for providing inputs into business decision problems. For example, manufacturing companies must forecast demand for their products to make sensible decisions about order quantities from their suppliers. Similarly, fast-food restaurants must forecast customer arrivals, sometimes down to the level of 15-minute intervals, so that they can staff their restaurants appropriately. There are many approaches to forecasting, ranging from simple to complex. Some involve regression-based methods, whereas other methods are based on extrapolation. In an extrapolation method the historical patterns of a time series variable, such as product demand or customer arrivals, are studied carefully and are then *extrapolated* into the future to obtain forecasts. In Chapter 12 we discuss both regression and extrapolation methods for forecasting.

Chapters 13 and 14 are devoted to spreadsheet optimization, with emphasis on linear programming. We assume a company must make several decisions, and there are constraints that limit the possible decisions. The job of the decision maker is to choose the decisions such that all of the constraints are satisfied and an objective, such as total profit or total cost, is optimized. The solution process consists of two steps. The first step is to build a spreadsheet model that relates the decision variables to other relevant quantities by means of logical formulas. In this first step there is no attempt to find the *optimal* solution; its only purpose is to relate all relevant quantities in a logical way. The second step is then to find the optimal solution. Fortunately, Excel contains a Solver add-in that performs the optimization. All you need to do is specify the objective, the decision variables, and the

constraints; Solver then uses powerful algorithms to find the optimal solution. As with regression, the power of this approach is its generality. An enormous variety of problems can be solved with spreadsheet optimization.

Chapters 15 and 16 illustrate a number of computer simulation models. This is not your first exposure to simulation—it is used in a number of previous chapters to illustrate statistical concepts—but here it is studied in its own right. As mentioned earlier, most business problems have some degree of uncertainty. The demand for a product is unknown, future interest rates are unknown, the delivery lead time from a supplier is unknown, and so on. Simulation allows you to build this uncertainty *explicitly* into spreadsheet models. Essentially, some cells in the model contain random values with given probability distributions. Every time the spreadsheet recalculates, these random values change, which causes “bottom-line” output cells to change as well. The trick then is to force the spreadsheet to recalculate many times and keep track of interesting outputs. In this way you can see which output values are most likely, and you can see best-case and worst-case results.

Spreadsheet simulations can be performed entirely with Excel’s built-in tools, but this is quite tedious. Therefore, we use a spreadsheet add-in to streamline the process. In particular, you will learn how the @RISK add-in from Palisade Corporation can be used to run replications of a simulation, keep track of outputs, create useful charts, and perform sensitivity analyses. With the inherent power of spreadsheets and the ease of using add-ins such as @RISK, spreadsheet simulation is becoming one of the most popular quantitative tools in the business world.

Chapter 17 is new to this edition. It provides a brief introduction to data mining, a topic of increasing importance in today’s data-driven world. Data mining is all about exploring data sets, especially large data sets, for relationships and patterns that can help companies gain a competitive advantage. It employs a number of relatively new technologies to implement various algorithms, several of which are discussed in this chapter.

Finally, there are three online chapters, 18–20, that complement topics included in the book itself. Chapter 18 discusses methods for importing the right data into Excel in the first place. Chapter 19 discusses analysis of variance (ANOVA) and experimental design. Chapter 20 discusses quality control and statistical process control. These three online chapters follow the same structure as the chapters in the book, complete with many examples and problems.

1-2b The Software

The quantitative methods in this book can be used to analyze a wide variety of business problems. However, they are not of much practical use unless you have the software to do the number crunching. Very few business problems are small enough to be solved with pencil and paper. They require powerful software.

The software included in new copies of this book, together with Microsoft Excel, provides you with a powerful combination. This software is being used—and will continue to be used—by leading companies all over the world to analyze large, complex problems. We firmly believe that the experience you obtain with this software, through working the examples and problems in this book, will give you a key competitive advantage in the marketplace.

It all begins with Excel. All of the quantitative methods that we discuss are implemented in Excel. Specifically, in this edition, we use Excel 2013.³ We cannot forecast the

³As you probably know, Microsoft made sweeping changes to Excel when it went from version 2003 to version 2007. The subsequent move to version 2010 introduced a few new changes, but nothing too dramatic, and the same is true of the most recent version, Excel 2013. If you are using version 2007 or 2010, you will see a few differences in the screenshots (which are from version 2013), but you should be able to follow along without trouble.

state of computer software in the long-term future, but Excel is currently *the* most heavily used spreadsheet package on the market, and there is every reason to believe that this state will persist for many years. Most companies use Excel, most employees and most students have been trained in Excel, and Excel is a *very* powerful, flexible, and easy-to-use package.

Built-in Excel Features

Virtually everyone in the business world knows the basic features of Excel, but relatively few know some of its more powerful features. In short, relatively few people are the “power users” we expect you to become by working through this book. To get you started, the file [excel_tutorial.xlsm](#) explains some of the “intermediate” features of Excel—features that we expect you to be able to use (see the Preface for instructions on how to access the resources that accompany this textbook). These include the **SUMPRODUCT**, **VLOOKUP**, **IF**, **NPV**, and **COUNTIF**, functions. They also include range names, data tables, Paste Special, Goal Seek, and many others. Finally, although we assume you can perform routine spreadsheet tasks such as copying and pasting, the tutorial provides many tips to help you perform these tasks more efficiently.⁴

In the body of the book, we describe several of Excel’s advanced features in more detail. For example, we introduce pivot tables in Chapter 3. This Excel tool enables you to summarize data sets in an almost endless variety of ways. As another example, we introduce Excel’s **RAND** and **RANDBETWEEN** functions for generating random numbers in Chapter 4. These functions are used in all spreadsheet simulations (at least those that do not take advantage of an add-in). In short, when an Excel tool is useful for a particular type of analysis, we usually provide step-by-step instructions on how to use it.

Solver Add-in

Chapters 13 and 14 make heavy use of Excel’s Solver add-in. This add-in, developed by Frontline Systems[®], not Microsoft, uses powerful algorithms—all behind the scenes—to perform spreadsheet optimization. Before this type of spreadsheet optimization add-in was available, specialized (nonspreadsheet) software was required to solve optimization problems. Now you can do it all within the familiar Excel environment.

SolverTable Add-in

An important theme throughout this book is sensitivity analysis: How do outputs change when inputs change? Typically these changes are made in spreadsheets with a data table, a built-in Excel tool. However, data tables don’t work in optimization models, where we want to see how the *optimal* solution changes when certain inputs change. Therefore, we include an Excel add-in called SolverTable, which works almost exactly like Excel’s data tables. (This add-in was developed by Albright.) Chapters 13 and 14 illustrate the use of SolverTable.

DecisionTools Suite

In addition to SolverTable and built-in Excel add-ins, an educational version of Palisade Corporation’s powerful DecisionTools[®] Suite is available (see the Preface for instructions on how to access it). All of the programs in this suite are Excel add-ins, so the learning curve isn’t very steep. There are six separate add-ins in this suite: @RISK, StatTools,

⁴Albright and a couple of colleagues have created a more robust commercial version of this tutorial called **ExcelNow!**. The [excel_tutorial.xlsm](#) file explains how you can upgrade to this commercial version at a very reasonable price.

PrecisionTree, NeuralTools, TopRank, and Evolver.⁵ We use only the first four in this book, but all are useful for certain tasks and are described briefly below.

@RISK

The simulation add-in @RISK enables you to run as many replications of a spreadsheet simulation as you like. As the simulation runs, @RISK automatically keeps track of the outputs you select, and it then displays the results in a number of tabular and graphical forms. @RISK also enables you to perform a sensitivity analysis, so that you can see which inputs have the most effect on the outputs. Finally, @RISK provides a number of spreadsheet functions that enable you to generate random numbers from a variety of probability distributions.

RISKOptimizer, part of @RISK, combines optimization with simulation. There are often times when you want to use simulation to model some business problem, but you also want to optimize a summary measure, such as a mean, of an output distribution. This optimization can be performed in a trial-and-error fashion, where you try a few values of the decision variable(s) and see which provides the best solution. However, RISKOptimizer provides a more automatic optimization procedure.

StatTools

Much of this book discusses basic statistical analysis. Fortunately, the Palisade suite includes a statistical add-in called StatTools. StatTools is powerful, easy to use, and capable of generating output quickly in an easily interpretable form.

PrecisionTree

The PrecisionTree add-in is used in Chapter 6 to analyze decision problems with uncertainty. The primary method for performing this type of analysis is to draw a decision tree. Decision trees are inherently graphical, and they have always been difficult to implement in spreadsheets, which are based on rows and columns. However, PrecisionTree does this in a very clever and intuitive way. Equally important, once the basic decision tree is built, you can use PrecisionTree to perform a sensitivity analysis on the model's inputs.

NeuralTools

Chapters 10 and 11 discuss how regression can be used to find a linear equation that quantifies the relationship between a dependent variable and one or more explanatory variables. Although linear regression is a powerful tool, it is not capable of quantifying all possible relationships. The NeuralTools add-in mimics the working of the human brain to find “neural networks” that quantify complex nonlinear relationships. It is used in Chapter 17 as one of several possible classification methods.

Evolver

Chapters 13 and 14 discuss how the built-in Solver add-in can optimize linear models and even some nonlinear models. But there are some “non-smooth” nonlinear models that Solver cannot handle. Fortunately, there are other optimization algorithms for such models, including “genetic” algorithms. The Evolver add-in implements these genetic algorithms. As of version 6.0, it also implements a very powerful optimization engine called OptQuest. Evolver can now solve practically any problem that Solver can solve, and it can even solve problems that are too large for Solver. (We continue to use Solver in the optimization chapters, however, because it is built into Excel.)

⁵The Palisade suite used to have another separate add-in, RISKOptimizer. Now RISKOptimizer is contained entirely in @RISK.