Business Analytics
Principles, Concepts, and
Applications with SAS
What, Why, and How

Marc J. Schniederjans
Dara G. Schniederjans
Christopher M. Starkey
This book is dedicated to Miles Starkey. He is what brings purpose to our lives and gives us a future.
This page intentionally left blank
## Contents-at-a-Glance

Preface ................................................................. xvi

**PART I:** What Is Business Analytics? .............................. 1

Chapter 1: What Is Business Analytics? ......................... 3

**PART II:** Why Is Business Analytics Important? ............ 15

Chapter 2: Why Is Business Analytics Important? ............ 17

Chapter 3: What Resource Considerations Are Important to Support Business Analytics? .................. 29

**PART III:** How Can Business Analytics Be Applied? ....... 43

Chapter 4: How Do We Align Resources to Support Business Analytics within an Organization? ........... 45

Chapter 5: What Is Descriptive Analytics? ...................... 63

Chapter 6: What Is Predictive Analytics? ......................... 95

Chapter 7: What Is Prescriptive Analytics? ...................... 117

Chapter 8: A Final Business Analytics Case Problem .......... 137

**PART IV:** Appendixes ............................................. 161

Appendix A: Statistical Tools .................................... 163

Appendix B: Linear Programming ................................. 195

Appendix C: Duality and Sensitivity Analysis in Linear Programming ................................................. 229
Appendix D: Integer Programming .......................... 249
Appendix E: Forecasting .................................... 257
Appendix F: Simulation ...................................... 281
Appendix G: Decision Theory ............................... 289
Index ....................................................... 321
# Table of Contents

Preface .......................................................... xvi
Conceptual Content .............................................. xvi
Software ............................................................ xvii
Analytic Tools ..................................................... xvii

## PART I: What Is Business Analytics? .................. 1

### Chapter 1: What Is Business Analytics? .......... 3

1.1 Terminology .............................................. 3
1.2 Business Analytics Process .......................... 7
1.3 Relationship of BA Process and Organization Decision-Making Process ........................................... 10
1.4 Organization of This Book ............................ 12
Summary .......................................................... 13
Discussion Questions ......................................... 13
References ...................................................... 14

## PART II: Why Is Business Analytics Important? .. 15

### Chapter 2: Why Is Business Analytics Important? .. 17

2.1 Introduction .............................................. 17
2.2 Why BA Is Important: Providing Answers to Questions ........................................... 18
2.3 Why BA Is Important: Strategy for Competitive Advantage ........................................... 20
2.4 Other Reasons Why BA Is Important ............... 23

#### 2.4.1 Applied Reasons Why BA Is Important ........ 23

2.4.2 The Importance of BA with New Sources of Data ........................................... 24
Summary .......................................................... 26
Discussion Questions ......................................... 26
References ...................................................... 26

### Chapter 3: What Resource Considerations Are Important to Support Business Analytics? .... 29

3.1 Introduction .............................................. 29
3.2 Business Analytics Personnel ....................... 30
<table>
<thead>
<tr>
<th>Chapter 6: What Is Predictive Analytics?</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>95</td>
</tr>
<tr>
<td>6.2 Predictive Modeling</td>
<td>96</td>
</tr>
<tr>
<td>6.2.1 Logic-Driven Models</td>
<td>96</td>
</tr>
<tr>
<td>6.2.2 Data-Driven Models</td>
<td>98</td>
</tr>
<tr>
<td>6.3 Data Mining</td>
<td>99</td>
</tr>
<tr>
<td>6.3.1 A Simple Illustration of Data Mining</td>
<td>100</td>
</tr>
<tr>
<td>6.3.2 Data Mining Methodologies</td>
<td>101</td>
</tr>
<tr>
<td>6.4 Continuation of Marketing/Planning Case Study Example: Prescriptive Analytics Step in the BA Process</td>
<td>104</td>
</tr>
<tr>
<td>6.4.1 Case Study Background Review</td>
<td>104</td>
</tr>
<tr>
<td>6.4.2 Predictive Analytics Analysis</td>
<td>105</td>
</tr>
<tr>
<td>Summary</td>
<td>113</td>
</tr>
<tr>
<td>Discussion Questions</td>
<td>113</td>
</tr>
<tr>
<td>Problems</td>
<td>114</td>
</tr>
<tr>
<td>References</td>
<td>115</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 7: What Is Prescriptive Analytics?</th>
<th>117</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>117</td>
</tr>
<tr>
<td>7.2 Prescriptive Modeling</td>
<td>118</td>
</tr>
<tr>
<td>7.3 Nonlinear Optimization</td>
<td>120</td>
</tr>
<tr>
<td>7.4 Continuation of Marketing/Planning Case Study Example: Prescriptive Step in the BA Analysis</td>
<td>127</td>
</tr>
<tr>
<td>7.4.1 Case Background Review</td>
<td>127</td>
</tr>
<tr>
<td>7.4.2 Prescriptive Analysis</td>
<td>127</td>
</tr>
<tr>
<td>Summary</td>
<td>132</td>
</tr>
<tr>
<td>Addendum</td>
<td>132</td>
</tr>
<tr>
<td>Discussion Questions</td>
<td>133</td>
</tr>
<tr>
<td>Problems</td>
<td>133</td>
</tr>
<tr>
<td>References</td>
<td>134</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 8: A Final Business Analytics Case Problem</th>
<th>137</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Introduction</td>
<td>137</td>
</tr>
<tr>
<td>8.2 Case Study: Problem Background and Data</td>
<td>138</td>
</tr>
<tr>
<td>8.3 Descriptive Analytics Analysis</td>
<td>139</td>
</tr>
</tbody>
</table>
Appendix C: Duality and Sensitivity Analysis in Linear Programming
C.3.3 A Primal Minimization Problem ......................... 238
C.3.4 A Second Primal Minimization Problem ............... 242
C.4 Determining the Economic Value of a Resource with Duality ................................................. 244
C.5 Duality Practice Problems .................................. 245

Appendix D:  Integer Programming ............................... 249
D.1 Introduction .................................................. 249
   D.1.1 What Is Integer Programming? ...................... 249
   D.1.2 Zero-One IP Problems/Models .......................... 250
D.2 Solving IP Problems/Models .................................. 250
   D.2.1 Introduction ............................................ 250
   D.2.2 A Maximization IP Problem ........................... 251
   D.2.3 A Minimization IP Problem ........................... 252
D.3 Solving Zero-One Programming Problems/Models .......... 253
D.4 Integer Programming Practice Problems .................... 254

Appendix E:  Forecasting ........................................ 257
E.1 Introduction .................................................. 257
E.2 Types of Variation in Time Series Data .................... 258
   E.2.1 Trend Variation ......................................... 260
   E.2.2 Seasonal Variation ...................................... 260
   E.2.3 Cyclical Variation ...................................... 261
   E.2.4 Random Variation ...................................... 261
   E.2.5 Forecasting Methods .................................... 261
E.3 Simple Regression Model .................................... 262
   E.3.1 Model for Trend ........................................ 262
   E.3.2 Computer-Based Solution .............................. 263
   E.3.3 Interpreting the Computer-Based Solution and Forecasting Statistics .......................... 266
E.4 Multiple Regression Models .................................. 267
   E.4.1 Introduction ............................................ 267
   E.4.2 Application ............................................. 268
   E.4.3 Limitations on the Use of Multiple Regression Models in Forecasting Time Series Data ........ 269
E.5 Simple Exponential Smoothing ............................... 270
   E.5.1 Introduction ............................................ 270
   E.5.2 An Example of Exponential Smoothing .................. 271
About the Authors

Marc J. Schniederjans is the C. Wheaton Battey Distinguished Professor of Business in the College of Business Administration at the University of Nebraska-Lincoln and has served on the faculty of three other universities. Professor Schniederjans is a Fellow of the Decision Sciences Institute (DSI) and in 2014–2015 will serve as DSI’s president. His prior experience includes owning and operating his own truck leasing business. He is currently a member of the Institute of Supply Management (ISM), the Production and Operations Management Society (POMS), and Decision Sciences Institute (DSI). Professor Schniederjans has taught extensively in operations management and management science. He has won numerous teaching awards and is an honorary member of the Golden Key honor society and the Alpha Kappa Psi business honor society. He has published more than a hundred journal articles and has authored or coauthored twenty books in the field of management. The title of his most recent book is Reinventing the Supply Chain Life Cycle, and his research has encompassed a wide range of operations management and decision science topics. He has also presented more than one hundred research papers at academic meetings. Professor Schniederjans is serving on five journal editorial review boards, including Computers & Operations Research, International Journal of Information & Decision Sciences, International Journal of Information Systems in the Service Sector, Journal of Operations Management, and Production and Operations Management. He is also serving as an area editor for the journal Operations Management Research and as an associate editor for the International Journal of Strategic Decision Sciences and International Journal of the Society Systems Science and Management Review: An International Journal (Korea). In addition, Professor Schniederjans has served as a consultant and trainer to various business and government agencies.
Dara G. Schniederjans is an assistant professor of Supply Chain Management at the University of Rhode Island, College of Business Administration. She has published articles in journals such as *Decision Support Systems*, *Journal of the Operational Research Society*, and *Business Process Management Journal*. She has also coauthored two text books and coedited a readings book. She has contributed chapters to readings utilizing quantitative and statistical methods. Dara has served as a guest coeditor for a special issue on *Business Ethics in Social Sciences* in the *International Journal of Society Systems Science*. She has also served as a website coordinator for Decisions Sciences Institute. She currently teaches courses in Supplier Relationship Management and Operations Management.

Christopher M. Starkey is an economics student at the University of Connecticut-Storrs. He has presented papers at the Academy of Management and Production and Operations Management Society meetings. He currently teaches courses in Principles of Microeconomics and has taught Principles of Macroeconomics. His current research interests include macroeconomic and monetary policy, as well as other decision-making methodologies.
Preface

Like the face on the cover of this book, we are bombarded by information every day. We do our best to sort out and use the information to help us get by, but sometimes we are overwhelmed by the abundance of data. This can lead us to draw wrong conclusions and make bad decisions. When you are a global firm collecting millions of transactions and customer behavior data from all over the world, the size of the data alone can make the task of finding useful information about customers almost impossible. For that firm and even smaller businesses, the solution is to apply business analytics (BA). BA helps sort out large data files (called “big data”), find patterns of behavior useful in predicting the future, and allocate resources to optimize decision-making. BA involves a step-wise process that aids firms in managing big data in a systematic procedure to glean useful information, which can solve problems and pinpoint opportunities for enhanced business performance.

This book has been written to provide a basic education in BA that can serve both academic and practitioner markets. In addition to bringing BA up-to-date with literature and research, this book explains the BA process in simple terms and supporting methodologies useful in its application. Collectively, the statistical and quantitative tools presented in this book do not need substantial prerequisites other than basic high school algebra. To support both markets, a substantial number of solved problems are presented along with some case study applications to train readers in the use of common BA tools and software. Practitioners will find the treatment of BA methodologies useful review topics. Academic users will find chapter objectives and discussion questions helpful for serving their needs while also having an opportunity to obtain an Instructor’s Guide with chapter-end problem solutions and exam questions.

The purpose of this book is to explain what BA is, why it is important to know, and how to do it. To achieve this purpose, the book presents conceptual content, software familiarity, and some analytic tools.

Conceptual Content

The conceptual material is presented in the first eight chapters of the book. (See Section 1.4 in Chapter 1 for an explanation of the book’s organization.) The conceptual content covers much more than what BA is about. It explains why BA is important in terms of providing answers to questions, how it can be used to achieve competitive
advantage, and how to align an organization to make best use of it. The book explains the managerial aspects of creating a BA presence in an organization and the skills BA personnel are expected to possess. The book also describes data management issues such as data collection, outsourcing, data quality, and change management as they relate to BA.

Having created a managerial foundation explaining “what” and “why” BA is important, the remaining chapters focus on “how” to do it. Embodied in a three-step process, BA is explained to have descriptive, predictive, and prescriptive analytic steps. For each of these steps, this book presents a series of strategies and best practice guides to aid in the BA process.

Software

Much of what BA is about involves the use of software. Unfortunately, no single software covers all aspects of BA. Many institutions prefer one type of software over others. To provide flexibility, this book’s use of software provides some options and can be used by readers who are not even interested in running computer software. In this book, SAS® and Lingo® software are utilized to model and solve problems. The software treatment is mainly the output of these software systems, although some input and instructions on their use are provided. For those not interested in running software applications, the exposure to the printouts provides insight into their informational value. This book recognizes that academic curriculums prefer to uniquely train students in the use of software and does not duplicate basic software usage. As a prerequisite to using this book, it is recommended that those interested in running software applications for BA become familiar with and are instructed on the use of whatever software is desired.

Analytic Tools

The analytic tool materials are chiefly contained in this book’s appendixes. BA is a statistical, management information system (MIS) and quantitative methods tools-oriented subject. Although the conceptual content in the book overviews how to undertake the BA process, the implementation of how to actually do BA requires quantitative tools. Because some practitioners and academic programs are less interested in the technical aspects of BA, the bulk of the quantitative material is presented
in the appendixes. These appendixes provide an explanation and illustration of a substantial body of BA tools to support a variety of analyses. Some of the statistical tools that are explained and illustrated in this book include statistical counting (permutations, combinations, repetitions), probability concepts (approaches to probability, rules of addition, rules of multiplication, Bayes’s theorem), probability distributions (binomial, Poisson, normal, exponential), confidence intervals, sampling methods, simple and multiple regression, charting, and hypothesis testing. Although management information systems are beyond the scope of this book, the software applications previously mentioned are utilized to illustrate search, clustering, and typical data mining applications of MIS technology. In addition, quantitative methods and tools explained and illustrated in this book include linear programming, duality and sensitivity analysis, integer programming, zero-one programming, forecasting modeling, nonlinear optimization, simulation analysis, breakeven analysis, and decision theory (certainty, risk, uncertainty, expected value opportunity loss analysis, expected value of perfect information, expected value of imperfect information).

We want to acknowledge the help of individuals who provided needed support for the creation of this book. First, we really appreciate the support of our editor, Jeanne Glasser Levine, and the outstanding staff at Pearson. They made creating this book a pleasure and worked with us to improve the final product. Decades of writing books with other publishers permitted us to recognize how using a top-tier publisher like we did makes a difference. We thank Alan McHugh, who developed the image on our book cover. His constant willingness to explore and be innovative with ideas made a significant contribution to our book. We also want to acknowledge the great editing help we received from Jill Schniederjans. Her skill has reduced the wordiness and enhanced the content (making parts less boring to read). Finally, we would like to acknowledge the help of Miles Starkey, whose presence and charm have lifted our spirits and kept us on track to meet completion deadlines.

Although many people have assisted in preparing this book, its accuracy and completeness are our responsibility. For all errors that this book may contain, we apologize in advance.

Marc J. Schniederjans
Dara G. Schniederjans
Christopher M. Starkey
This page intentionally left blank
What Is Business Analytics?

Chapter objectives:

• Define business analytics.
• Explain the relationship of analytics and business intelligence to the subject of business analytics.
• Describe the three steps of the business analytics process.
• Describe four data classification measurement scales.
• Explain the relationship of the business analytics process with the organization decision-making process.

1.1 Terminology

Business analytics begins with a *data set* (a simple collection of data or a data file) or commonly with a *database* (a collection of data files that contain information on people, locations, and so on). As databases grow, they need to be stored somewhere. Technologies such as *computer clouds* (hardware and software used for data remote storage, retrieval, and computational functions) and *data warehousing* (a collection of databases used for reporting and data analysis) store data. Database storage areas have become so large that a new term was devised to describe them. *Big data* describes the collection of data sets that are so large and complex that software systems are hardly able to process them (Isson and Harriott, 2013, pp. 57–61). Isson and Harriott (2013, p. 61) define *little data* as anything that is not big data. Little data describes the smaller data segments or files that help individual businesses keep track of customers. As a means of sorting through data to find useful information, the application of analytics has found new purpose.
Three terms in business literature are often related to one another: analytics, business analytics, and business intelligence. Analytics can be defined as a process that involves the use of statistical techniques (measures of central tendency, graphs, and so on), information system software (data mining, sorting routines), and operations research methodologies (linear programming) to explore, visualize, discover, and communicate patterns or trends in data. Simply, analytics converts data into useful information. Analytics is an older term commonly applied to all disciplines, not just business. A typical example of the use of analytics is the weather measurements collected and converted into statistics, which in turn predict weather patterns.

There are many types of analytics, and there is a need to organize these types to understand their uses. We will adopt the three categories (*descriptive*, *predictive*, and *prescriptive*) that the Institute of Operations Research and Management Sciences (INFORMS) organization (www.informs.org) suggests for grouping the types of analytics (see Table 1.1). These types of analytics can be viewed independently. For example, some firms may only use descriptive analytics to provide information on decisions they face. Others may use a combination of analytic types to glean insightful information needed to plan and make decisions.

<table>
<thead>
<tr>
<th>Type of Analytics</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive</td>
<td>The application of simple statistical techniques that describe what is contained in a data set or database. Example: An age bar chart is used to depict retail shoppers for a department store that wants to target advertising to customers by age.</td>
</tr>
<tr>
<td>Predictive</td>
<td>An application of advanced statistical, information software, or operations research methods to identify predictive variables and build predictive models to identify trends and relationships not readily observed in a descriptive analysis. Example: Multiple regression is used to show the relationship (or lack of relationship) between age, weight, and exercise on diet food sales. Knowing that relationships exist helps explain why one set of independent variables influences dependent variables such as business performance.</td>
</tr>
<tr>
<td>Prescriptive</td>
<td>An application of decision science, management science, and operations research methodologies (applied mathematical techniques) to make best use of allocable resources. Example: A department store has a limited advertising budget to target customers. Linear programming models can be used to optimally allocate the budget to various advertising media.</td>
</tr>
</tbody>
</table>

The purposes and methodologies used for each of the three types of analytics differ, as can be seen in Table 1.2. These differences distinguish *analytics* from *business analytics*. Whereas analytics is focused on generating insightful information from
data sources, business analytics goes the extra step to leverage analytics to create an improvement in measurable business performance. Whereas the process of analytics can involve any one of the three types of analytics, the major components of business analytics include all three used in combination to generate new, unique, and valuable information that can aid business organization decision-making. In addition, the three types of analytics are applied sequentially (descriptive, then predictive, then prescriptive). Therefore, *business analytics* (BA) can be defined as a process beginning with business-related data collection and consisting of sequential application of descriptive, predictive, and prescriptive major analytic components, the outcome of which supports and demonstrates business decision-making and organizational performance. Stubbs (2011, p. 11) believes that BA goes beyond plain analytics, requiring a clear relevancy to business, a resulting insight that will be implementable, and performance and value measurement to ensure a successful business result.

**Table 1.2 Analytic Purposes and Tools**

<table>
<thead>
<tr>
<th>Type of Analytics</th>
<th>Purpose</th>
<th>Examples of Methodologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive</td>
<td>To identify possible trends in large data sets or databases. The purpose is to get a rough picture of what generally the data looks like and what criteria might have potential for identifying trends or future business behavior.</td>
<td>Descriptive statistics, including measures of central tendency (mean, median, mode), measures of dispersion (standard deviation), charts, graphs, sorting methods, frequency distributions, probability distributions, and sampling methods.</td>
</tr>
<tr>
<td>Predictive</td>
<td>To build predictive models designed to identify and predict future trends.</td>
<td>Statistical methods like multiple regression and ANOVA. Information system methods like data mining and sorting. Operations research methods like forecasting models.</td>
</tr>
<tr>
<td>Prescriptive</td>
<td>To allocate resources optimally to take advantage of predicted trends or future opportunities.</td>
<td>Operations research methodologies like linear programming and decision theory.</td>
</tr>
</tbody>
</table>

*Business intelligence* (BI) can be defined as a set of processes and technologies that convert data into meaningful and useful information for business purposes. Although some believe that BI is a broad subject that encompasses analytics, business analytics, and information systems (Bartlett, 2013, p.4), others believe it is mainly focused on collecting, storing, and exploring large database organizations for information useful to decision-making and planning (Negash, 2004). One function that is generally accepted as a major component of BI involves storing an organization’s data in computer cloud storage or in data warehouses. Data warehousing is not an analytics or business analytics function, although the data can be used for analysis. In application,
BI is focused on querying and reporting, but it can include reported information from a BA analysis. BI seeks to answer questions such as what is happening now and where, and also what business actions are needed based on prior experience. BA, on the other hand, can answer questions like why something is happening, what new trends may exist, what will happen next, and what is the best course for the future.

In summary, BA includes the same procedures as plain analytics but has the additional requirement that the outcome of the analytic analysis must make a measurable impact on business performance. BA includes reporting results like BI but seeks to explain why the results occur based on the analysis rather than just reporting and storing the results, as is the case with BI. Analytics, BA, and BI will be mentioned throughout this book. A review of characteristics to help differentiate these terms is presented in Table 1.3.

Table 1.3 Characteristics of Analytics, Business Analytics, and Business Intelligence

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Analytics</th>
<th>Business Analytics (BA)</th>
<th>Business Intelligence (BI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business performance planning role</td>
<td>What is happening, and what will be happening?</td>
<td>What is happening now, what will be happening, and what is the best strategy to deal with it?</td>
<td>What is happening now, and what have we done in the past to deal with it?</td>
</tr>
<tr>
<td>Use of descriptive analytics as a major component of analysis</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Use of predictive analytics as a major component of analysis</td>
<td>Yes</td>
<td>Yes</td>
<td>No (only historically)</td>
</tr>
<tr>
<td>Use of prescriptive analytics as a major component of analysis</td>
<td>Yes</td>
<td>Yes</td>
<td>No (only historically)</td>
</tr>
<tr>
<td>Use of all three in combination</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Business focus</td>
<td>Maybe</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Focus of storing and maintaining data</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Required focus of improving business value and performance</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
1.2 Business Analytics Process

The complete business analytics process involves the three major component steps applied sequentially to a source of data (see Figure 1.1). The outcome of the business analytics process must relate to business and seek to improve business performance in some way.

![Business analytics process diagram](image)

**Figure 1.1 Business analytics process**

The logic of the BA process in Figure 1.1 is initially based on a question: What valuable or problem-solving information is locked up in the sources of data that an organization has available? At each of the three steps that make up the BA process, additional questions need to be answered, as shown in Figure 1.1. Answering all these questions requires mining the information out of the data via the three steps of analysis that comprise the BA process. The analogy of digging in a mine is appropriate for the BA process because finding new, unique, and valuable information that can lead to a successful strategy is just as good as finding gold in a mine. SAS, a major
analytic corporation (www.sas.com), actually has a step in its BA process, *Query Drill-down*, which refers to the mining effort of questioning and finding answers to pull up useful information in the BA analysis. Many firms routinely undertake BA to solve specific problems, whereas other firms undertake BA to explore and discover new knowledge to guide organizational planning and decision-making to improve business performance.

The size of some data sources can be unmanageable, overly complex, and generally confusing. Sorting out data and trying to make sense of its informational value requires the application of descriptive analytics as a first step in the BA process. One might begin simply by sorting the data into groups using the four possible classifications presented in Table 1.4. Also, incorporating some of the data into spreadsheets like Excel and preparing cross tabulations and contingency tables are means of restricting the data into a more manageable data structure. Simple measures of central tendency and dispersion might be computed to try to capture possible opportunities for business improvement. Other descriptive analytic summarization methods, including charting, plotting, and graphing, can help decision makers visualize the data to better understand content opportunities.

<table>
<thead>
<tr>
<th>Type of Data Measurement Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical Data</td>
<td>Data that is grouped by one or more characteristics. Categorical data usually involves cardinal numbers counted or expressed as percentages. Example 1: Product markets that can be characterized by categories of “high-end” products or “low-income” products, based on dollar sales. It is common to use this term to apply to data sets that contain items identified by categories as well as observations summarized in cross-tabulations or contingency tables.</td>
</tr>
<tr>
<td>Ordinal Data</td>
<td>Data that is ranked or ordered to show relational preference. Example 1: Football team rankings not based on points scored but on wins. Example 2: Ranking of business firms based on product quality.</td>
</tr>
<tr>
<td>Interval Data</td>
<td>Data that is arranged along a scale, in which each value is equally distant from others. It is ordinal data. Example 1: A temperature gauge. Example 2: A survey instrument using a Likert scale (that is, 1, 2, 3, 4, 5, 6, 7), where 1 to 2 is perceived as equidistant to the interval from 2 to 3, and so on. Note: In ordinal data, the ranking of firms might vary greatly from first place to second, but in interval data, they would have to be relationally proportional.</td>
</tr>
<tr>
<td>Ratio Data</td>
<td>Data expressed as a ratio on a continuous scale. Example 1: The ratio of firms with green manufacturing programs is twice that of firms without such a program.</td>
</tr>
</tbody>
</table>
From Step 1 in the Descriptive Analytic analysis (see Figure 1.1), some patterns or variables of business behavior should be identified representing targets of business opportunities and possible (but not yet defined) future trend behavior. Additional effort (more mining) might be required, such as the generation of detailed statistical reports narrowly focused on the data related to targets of business opportunities to explain what is taking place in the data (what happened in the past). This is like a statistical search for predictive variables in data that may lead to patterns of behavior a firm might take advantage of if the patterns of behavior occur in the future. For example, a firm might find in its general sales information that during economic downtimes, certain products are sold to customers of a particular income level if certain advertising is undertaken. The sales, customers, and advertising variables may be in the form of any of the measurable scales for data in Table 1.4, but they have to meet the three conditions of BA previously mentioned: clear relevancy to business, an implementable resulting insight, and performance and value measurement capabilities.

To determine whether observed trends and behavior found in the relationships of the descriptive analysis of Step 1 actually exist or hold true and can be used to forecast or predict the future, more advanced analysis is undertaken in Step 2, Predictive Analytic analysis, of the BA process. There are many methods that can be used in this step of the BA process. A commonly used methodology is multiple regression. (See Appendix A, “Statistical Tools,” and Appendix E, “Forecasting,” for a discussion on multiple regression and ANOVA testing.) This methodology is ideal for establishing whether a statistical relationship exists between the predictive variables found in the descriptive analysis. The relationship might be to show that a dependent variable is predictively associated with business value or performance of some kind. For example, a firm might want to determine which of several promotion efforts (independent variables measured and represented in the model by dollars in TV ads, radio ads, personal selling, or magazine ads) is most efficient in generating customer sales dollars (the dependent variable and a measure of business performance). Care would have to be taken to ensure the multiple regression model was used in a valid and reliable way, which is why ANOVA and other statistical confirmatory analyses support the model development. Exploring a database using advanced statistical procedures to verify and confirm the best predictive variables is an important part of this step in the BA process. This answers the questions of what is currently happening and why it happened between the variables in the model.

A single or multiple regression model can often forecast a trend line into the future. When regression is not practical, other forecasting methods (exponential smoothing, smoothing averages) can be applied as predictive analytics to develop needed forecasts of business trends. (See Appendix E.) The identification of future
trends is the main output of Step 2 and the predictive analytics used to find them. This helps answer the question of what will happen.

If a firm knows where the future lies by forecasting trends as they would in Step 2 of the BA process, it can then take advantage of any possible opportunities predicted in that future state. In Step 3, *Prescriptive Analytics analysis*, operations research methodologies can be used to optimally allocate a firm’s limited resources to take best advantage of the opportunities it found in the predicted future trends. Limits on human, technology, and financial resources prevent any firm from going after all opportunities it may have available at any one time. Using prescriptive analytics allows the firm to allocate limited resources to optimally achieve objectives as fully as possible. For example, *linear programming* (a constrained optimization methodology) has been used to maximize the profit in the design of supply chains (Paksoy et al., 2013). (Note: Linear programming and other optimization methods are presented in Appendixes B, “Linear Programming,” C, “Duality and Sensitivity Analysis in Linear Programming,” and D, “Integer Programming.”) This third step in the BA process answers the question of how best to allocate and manage decision-making in the future.

In summary, the three major components of descriptive, predictive, and prescriptive analytics arranged as steps in the BA process can help a firm find opportunities in data, predict trends that forecast future opportunities, and aid in selecting a course of action that optimizes the firm’s allocation of resources to maximize value and performance. The BA process, along with various methodologies, will be detailed in Chapters 5 through 10.

### 1.3 Relationship of BA Process and Organization Decision-Making Process

The BA process can solve problems and identify opportunities to improve business performance. In the process, organizations may also determine strategies to guide operations and help achieve competitive advantages. Typically, solving problems and identifying strategic opportunities to follow are organization decision-making tasks. The latter, identifying opportunities, can be viewed as a problem of strategy choice requiring a solution. It should come as no surprise that the BA process described in Section 1.2 closely parallels classic organization decision-making processes. As depicted in Figure 1.2, the business analytics process has an inherent relationship to the steps in typical organization decision-making processes.
Chapter 1 • What Is Business Analytics?

1. Descriptive analytic analysis
2. Predictive analytic analysis
3. Prescriptive analytic analysis

Source of data

Outcome of both of these processes: Measurable increase in business value and performance

Figure 1.2 Comparison of business analytics and organization decision-making processes

*Source: Adapted from Figure 1 in Elbing (1970), pp. 12–13.

The organization decision-making process (ODMP) developed by Elbing (1970) and presented in Figure 1.2 is focused on decision-making to solve problems but could also be applied to finding opportunities in data and deciding what is the best course of action to take advantage of them. The five-step ODMP begins with the perception of disequilibrium, or the awareness that a problem exists that needs a decision. Similarly, in the BA process, the first step is to recognize that databases may contain information that could both solve problems and find opportunities to improve business performance. Then in Step 2 of the ODMP, an exploration of the problem to determine its size, impact, and other factors is undertaken to diagnose what the problem is. Likewise, the BA descriptive analytic analysis explores factors that might prove useful in solving problems and offering opportunities. The ODMP problem statement step is similarly structured to the BA predictive analysis to find strategies, paths, or trends that clearly define a problem or opportunity for an organization to solve problems. Finally, the ODMP’s last steps of strategy selection and implementation involve the same kinds of tasks that the BA process requires in the final prescriptive step.
(make an optimal selection of resource allocations that can be implemented for the betterment of the organization).

The decision-making foundation that has served ODMP for many decades parallels the BA process. The same logic serves both processes and supports organization decision-making skills and capacities.

1.4 Organization of This Book

This book is designed to answer three questions about BA:

- What is it?
- Why is it important?
- How do you do it?


Knowing the importance of explaining how BA is undertaken, the rest of the book’s chapters and appendixes are devoted to answering that question. Chapter 4, “How Do We Align Resources to Support Business Analytics within an Organization?” explains how an organization needs to support BA. Chapter 5, “What Is Descriptive Analytics?” Chapter 6, “What Is Predictive Analytics?” and Chapter 7, “What Is Prescriptive Analytics?” detail and illustrate the three respective steps in the BA process. To further illustrate how to conduct a BA analysis, Chapter 8, “A Final Business Analytics Case Problem,” provides an example of BA. Supporting the analytic discussions is a series of analytically oriented appendixes that follow Chapter 8.

Part III, “How Can Business Analytics Be Applied?” includes quantitative analyses utilizing computer software. In an effort to provide some diversity of software usage, SAS and LINGO software output are presented. Because of the changing nature of software and differing educational backgrounds, this book does not provide extensive software explanation.

In addition to the basic content that makes up the body of the chapters, there are pedagogy enhancements that can aid learning. All chapters begin with chapter objectives and end with a summary, discussion questions, and, where needed, references. In addition, Chapters 5 through 8 have sample problems with solutions, as well as additional assignment problems.
Some of the more detailed explanations of methodologies are presented in the appendixes. Their positioning in the appendixes is designed to enhance content flow and permit more experienced readers a flexible way to select only the technical content they might want to use. An extensive index allows quick access to terminology.

**Summary**

This chapter has introduced important terminology and defined business analytics in terms of a unique process useful in securing information on which decisions can be made and business opportunities seized. Data classification measurement scales were also briefly introduced to aid in understanding the types of measures that can be employed in BA. The relationship of the BA process and the organization decision-making process was explained in terms of how they complement each other. This chapter ended with a brief overview of this book’s organization and how it is structured to aid learning.

Knowing *what* business analytics is about is important, but equally important is knowing *why* it is important. Chapter 2 begins to answer the question.

**Discussion Questions**

1. What is the difference between analytics and business analytics?
2. What is the difference between business analytics and business intelligence?
3. Why are the steps in the business analytics process sequential?
4. How is the business analytics process similar to the organization decision-making process?
5. Why does interval data have to be relationally proportional?
References


This page intentionally left blank
Index

a priori probabilities, 168
accuracy statistics
- MAD (mean absolute deviation), 278
- MAPE (mean absolute percentage error), 279
- MSE (mean square error), 278
addition, rules of, 169-170
additivity in LP (Linear Programming) models, 224
administrators, 31
aligning business analytics, 45-46
- management issues, 54
  - change management, 58-59
  - ensuring data quality, 56-57
  - establishing information policy, 54
  - measuring business analytics contribution, 58
  - outsourcing business analytics, 55-56
organization structures, 46-51
  - centralized BA organization structure, 49-50
  - functional organization structure, 48
  - hierarchical relationships, 46
  - matrix organization structure, 48
  - project structure, 47-48
  - reasons for BA initiative and organization failure, 50-51
teams, 51-53
  - collaboration, 52-53
  - participant roles, 51-52
  - reasons for team failures, 53
alternative hypothesis, 189
alternatives (DT), 290
Analysis ToolPak, 39
analytics. See also DT (decision theory)
- alignment. See business analytics alignment
analytic purposes and tools, 5
business analytics personnel, 30-33
  - administrators, 31
  - BAP (Business Analytics Professional) exam, 30-31
  - designers, 31
  - developers, 31
  - skills and competency requirements, 32-33
  - solution experts, 31
  - technical specialists, 31
business analytics process
  - data measurement scales, 8
  - explained, 7-10
  - relationship with organization decision-making process (ODMP), 10-12
characteristics of, 6
correlation analysis, 98
decision analysis. See DT (decision theory)
definition of, 3-4
descriptive analytics
  - analytic purposes and tools, 5
  - definition of, 4
  - descriptive statistics, 74-79
  - explained, 63-68
  - illustrative sales data sets, 64
  - marketing/planning case study, 87
  - probability distributions, 84-86
  - sampling estimation, 82-84
  - sampling methods, 79-81
  - supply chain shipping case study, 139-145
discriminant analysis, 102
forecasting. See forecasting
predictive analytics, 98
  - analytic purposes and tools, 5
  - data mining, 99-104
  - data-driven models, 98
  - definition of, 4
  - explained, 95-96
  - logic-driven models, 96-98
  - marketing/planning case study, 104-113
  - supply chain shipping case study, 146-153
prescriptive analytics
  - analytic purposes and tools, 5
  - definition of, 4
  - explained, 117-118
  - integer programming. See IP (integer programming)
  - marketing/planning case study, 127-131
  - methodologies, 118
  - nonlinear optimization, 119-126
  - prescriptive modeling, 118
regression analysis, 98
sensitivity analysis
  - economic value of resources, determining, 244
  - overview, 230-231
primal maximization problems, 231-238
primal minimization problems, 238-243
simulation, 98, 120, 281
computer simulation methods, 288
deterministic simulation, 125-126, 281-282
practice problems, 288
probabilistic simulation, 282-288
analytics analysts, 51
analytics modelers, 51
analytics process designers, 51
ANOVA testing, 9, 266
applications of business analytics to enhance decision-making, 23-24
applied LP (Linear Programming) model, 196
artificial variables, 214
assessing probability
  Frequency Theory, 167-168
  Principle of Insufficient Reason, 168
  rules of addition, 169-170
  rules of multiplication, 170-173
associations, 39, 101
assumptions for simple regression model, 266
averages, smoothing
  example of, 274
  explained, 273-274
  formula, 273

B
BA team heads, 51
backward decision method, 303-306
backward step-wise regression, 107
BAP (Business Analytics Professional) exam, 30-31
bar charts, 69
Bayes's theorem, 307-314
belief of physical proximity, 50
BI (business intelligence), 5-6
big data
data mining, 38-40
text mining, 39
types of information obtainable, 39
web mining, 39
definition of, 3
descriptive analytics
descriptive statistics, 74-79
need for, 73-74
probability distributions, 84-86
sampling estimation, 82-84
sampling methods, 79-81
supply chain shipping case study, 139-145
importing into SAS, 68
and need for BA (business analytics), 17
and need for DBMS systems, 36-37
predictive analytics
data mining, 99-104
data-driven models, 98
explained, 95-96
logic-driven models, 96-98
marketing/planning case study, 104-113
supply chain shipping case study, 146-153
prescriptive analytics
explained, 117-118
marketing/planning case study, 127-131
methodologies, 118
nonlinear optimization, 119-126
prescriptive modeling, 118
supply chain shipping case study, 153-159
problems with, 17-18
SAS simulation, 288
billing and reminder systems, 34
binding constraints, 218
binomial probability distribution, 175-177
binomial tests, 193
blending formulations, 221-222
branch-and-bound method, 250-252
business analytics alignment, 45-46
management issues, 54
  change management, 58-59
  ensuring data quality, 56-57
  establishing information policy, 54
  measuring business analytics contribution, 55
  outsourcing business analytics, 55-56
organization structures, 46-51
centralized BA organization structure, 49-50
  functional organization structure, 48
  hierarchical relationships, 46
  matrix organization structure, 48
  project structure, 47-48
  reasons for BA initiative and organization failure, 50-51
teams, 51-53
  collaboration, 52-53
  participant roles, 51-52
  reasons for team failures, 53
business analytics personnel, 30-33
administrators, 31
BAP (Business Analytics Professional) exam, 30-31
designers, 31
developers, 31
skills and competency requirements, 32-33
solution experts, 31
technical specialists, 31
business analytics process
data measurement scales, 8
explained, 7-10
relationship with organization decision-making process (ODMP), 10-12
Business Analytics Professional (BAP) exam, 30-31
business domain experts, 51-52
business intelligence (BI), 5-6
business performance tracking, 24
business process improvement, demonstrating, 158-159
butcher problem example (LP), 202-204

C
CAP (Certified Analytic Professional), 30
carried inventory units, 285
case studies
explained, 119
marketing/planning case study, 87
case study background, 87-88, 104-105, 127
descriptive analytics, 88-92
predictive analytics, 104-113
prescriptive analytics, 127-131
supply chain shipping case study
descriptive analytics, 139-145
predictive analytics, 146-153
prescriptive analytics, 153-159
problem background and data, 137-138
categorical data, 8
categorizing data, 33-35
cause-and-effect diagrams, 97
central limit theorem, 84
centralized BA organization structure, 49-50
certainty
decision-making under certainty, 292
maximax criterion, 292
maximin criterion, 293
explained, 290
in LP (Linear Programming) models, 224
certifications
BAP (Business Analytics Professional) exam, 30-31
CAP (Certified Analytic Professional), 30
IBM, 31
Certified Analytic Professional (CAP), 30
championing change, 60
change management, 58-60
best practices, 60
targets, 59
charts, 69-74. See also diagrams
bar charts, 69
column charts, 69
histograms, 69
line charts, 69
pie charts, 69
scatter charts, 69
Chi-Square tests, 193
Claritas, 35
Clarke Special Parts problem example, 208-209
classification, 39, 101
clearly stated goals, 60
cluster random sampling, 80
clustering, 101
data mining, 39
hierarchical clustering, 102-103
K-mean clustering, 103
coding, checking for, 57
coefficients
confidence coefficient, 84-85
contribution coefficients, 201
correlation analysis, 105-106
kurtosis, 74
skewedness, 74
technology coefficients, 198
Z values, 84-85
Cognizure BAP (Business Analytics Professional) exam, 30-31
collaboration
lack of, 50
in teams, 52-53
collectively exhaustive set of events, 169
column charts, 70
combinations, 165
communication
good communication, 60
lack of, 53
competency requirements for business analytics personnel, 32-33
competition data sources, 34
competitive advantage
achieving with business analytics, 20-22
innovation, 22
operations efficiency, 22
price leadership, 22
product differentiation, 22
service effectiveness, 22
sustainability, 22
completeness, checking for, 57
compound events, 169
computer simulation methods, 288
conditional probabilities, 172
confidence coefficient, 84-85
confidence intervals, 82-84
costants
binding constraints, 218
formulating, 128-129, 202
LP (Linear Programming), 198-200
nonbinding constraints, 218
redundant constraints, 218
continuous probability distributions, 174, 181-188
exponential probability distribution, 186-188
normal probability distribution, 181-186
standard normal probability distribution, 183-184
continuous random variables, 173-174
correlation analysis, 98
correlation coefficients
marketing/planning case study, 105-106
for multiple regression model, 268-269

counting, 74, 163
combinations, 165
permutations, 163-164
repetitions, 166

credit union example of business analysis, 19
CRM (customer relationship management) systems, 34
cubic model forecasts
formula for cubic regression models, 150
supply chain shipping case study, 149-151
culture as target of change management, 59
current data, checking for, 57
curve fitting, 276
for nonlinear optimization, 121-125
supply chain shipping case study, 146-150
customer demographics, 34
customer internal data, 34
customer profitability, increasing, 23
customer relationship management (CRM) systems, 34
customer satisfaction, 34
customer service problem example (LP), 207-208
cyclical variation, 261

D
data definition, 37
data inspection items, 57
data management technology, 36
data managers, 52
data manipulation language tools, 37
data marts, 38
data measurement scales, 8
data mining, 38-40, 99-104
methodologies, 101-104
discriminant analysis, 102
hierarchical clustering, 102-103
K-mean clustering, 103
neural networks, 101-102
table of, 101
simple illustration of, 100-101
text mining, 39
types of information obtainable, 39
web mining, 39
data privacy, 35-36
data quality
ensuring, 56-57
overview, 35-36
data sets, 3
data sources
categorizing data, 33-35
data privacy, 35-36
data quality, 35-36
external sources, 35
internal sources, 34
new sources of data, applying business analytics to, 24-25
data visualization
charts, 69-74
bar charts, 69
column charts, 69
histograms, 69
line charts, 69
pie charts, 69
scatter charts, 69

diagrams
cause-and-effect diagrams, 97
influence diagrams, 97
data warehouses, 38
database management systems (DBMS), 36-37
databases, 3
database encyclopedia content, 38
DBMS (database management systems), 36-37
data-driven models, 98
DBMS (database management systems), 36-37
decision analysis, 119
decision environments. See also DT (decision theory)
certainty
decision-making under certainty, 292-293
explained, 290
risk
decision-making under risk, 293-297
explained, 290
uncertainty
decision-making under uncertainty, 297-301
explained, 291
decision theory. See DT (decision theory)
decision trees, 303-306
decision variables, defining, 128, 201
delegation of responsibility, 51
dependent event outcomes, 171
descriptive analytics
analytic purposes and tools, 5
data visualization, 69-74
definition of, 4
descriptive statistics, 74-79
explained, 63-68
illustrative sales data sets, 64
marketing/planning case study, 87
probability distributions, 84-86
sampling estimation, 82-84
sampling methods, 79-81
supply chain shipping case study, 139-145
actual monthly customer demand in motors, 140-142
Chicago customer demand (graph), 143
estimated shipping costs per motor, 139-140
Houston customer demand (graph), 144
INDEX

Kansas City customer demand (graph), 143
Little Rock customer demand (graph), 145
Oklahoma City customer demand (graph), 144
Omaha customer demand (graph), 145

problem background and data, 138-139
descriptive statistics, 74-79
designers, 31
deterministic simulation, 125-126, 281-282
developers, 31
diagrams. See also charts
cause-and-effect diagrams, 97
influence diagrams, 97
diet problem example (LP), 204-206
digital analytics, 24-25
discrete probability distributions, 174-181
binomial probability distribution, 175-177
geometric probability distribution, 180
hypergeometric probability distribution, 181
Poisson probability distribution, 178-180
discrete random variables, 173
discriminant analysis, 102
distribution free tests, 193
divisibility in LP (Linear Programming) models, 224
downloading LINGO, 214
DT (decision theory)
Bayes's theorem, 307-314
decision-making under certainty, 292
maximax criterion, 292
maximin criterion, 293
decision-making under risk, 293
EV (expected value) criterion, 294-295
expected opportunity loss criterion, 295-297
origin of probabilities, 294
decision-making under uncertainty, 297
Hurwicz criterion, 298-299
Laplace criterion, 297-298
maximax criterion, 298
maximin criterion, 298
minimax criterion, 299-301
enhancing decision-making with business analytics, 23-24
EVPI (expected value of perfect information), 301-302
model elements, 290
model formulation, 291-292
overview, 289
practice problems, 314-319
sequential decisions and decision trees, 303-306
types of decision environments, 290-291
duality
dual problems, 229
dual solutions, 229
duality practice problems, 245-247
economic value of resources, determining, 244
informational value of, 230
overview, 229
primal maximization problems, 231-238
primal minimization problems, 238-243
Dun & Bradstreet, 35
duplication, checking for, 57
Durbin-Watson Autocorrelation Test, 269
E
economic data sources, 34
economic value of resources, determining, 244
ensuring data quality, 56-57
enterprise resource planning (ERP) systems, 34
equations. See formulas
Equifax, 35
ERP (enterprise resource planning) systems, 34
errors
confidence intervals, 82-84
standard error, 74
establishing information policy, 54
estimation, sampling, 82-84, 98
EV (expected value) criterion, 294-295
events
collectively exhaustive set of events, 169
compound events, 169
dependent event outcomes, 171
independent event outcomes, 170
mutually exclusive events, 169
EVPI (expected value of perfect information), 301-302
executive sponsorship, lack of, 50
expected opportunity loss criterion, 295-297
expected value (EV) criterion, 294-295
expected value of perfect information (EVPI), 301-302
experiments, 173
EXPO function, 188
exponential probability distribution, 186-188
exponential smoothing
example of, 271-272
explained, 270-271
formula, 270
external data sources, 35
F
factorials, 164
failures
failure to deliver, 53
failure to provide value, 53
reasons for BA initiative and organization failure, 50-51
reasons for team failures, 53
farming problem example (LP), 206-207
Federal Division problem example (LP), 209-211
files (data set), creating, 64-68
finiteness in LP (Linear Programming) models, 224
fitting models to data, 275-276
forecasting, 98, 101
data mining, 39
fitting models to data, 275-276
forecast accuracy statistics
  MAD (mean absolute deviation), 278
  MAPE (mean absolute percentage error), 279
  MSE (mean square error), 278
forecasting methods, 261-262
forecasting model formula, 111
model selection, 277-279
multiple regression models
  application, 268-269
  explained, 267
  formula, 267
  limitations in forecasting time series data, 269
overview, 257
parameters for models, selecting, 277-279
practice problems, 279
simple exponential smoothing
  example of, 271-272
  explained, 270-271
  formula, 270
simple regression model, 262
  assumptions, 266
  computer-based solution, 263-266
  model for trend, 262-263
  statistical values and tests, 266-267
smoothing averages
  example of, 274
  explained, 273-274
  formula, 273
supply chain shipping case study
  developing forecasting models, 146-150
  resulting warehouse customer demand forecasts, 152-153
  validating forecasting models, 150
types of variation in time series data
cyclical variation, 261
forecasting methods, 261-262
overview, 258-260
random variation, 261
seasonal variation, 260
trend variation, 260
formulas
  confidence intervals, 82-84
  constraints, 128-129
  cubic regression models, 150
  DT (decision theory) models, 291-292
  forecasting model, 111
  linear regression model, 151
  MAD (mean absolute deviation), 278
  MAPE (mean absolute percentage error), 279
  MSE (mean square error), 278
  multiple regression model, 267
  objective function, 128
  quadratic regression model, 151, 276
  simple exponential smoothing, 270
  simple regression model, 262
  smoothing averages, 273
forward step-wise regression, 107
F-ratio statistic, 109-110
frequency functions, 174
Frequency Theory, 167-168
F-Test Two-Sample for Variances tool, 191
functional organization structure, 48
functions
  EXPO, 188
  frequency functions, 174
  NORMAL, 185
  objective, 197-198, 201-202
G
generalized LP (Linear Programming) model, 196
geometric probability distribution, 180
given requirements, stating, 200
goals, 60
Google Insights for Search, 39
Google Trends, 39
H
hardware, 36
hierarchical clustering, 102-103
hierarchical relationships, 46
histograms, 73
human resources
decisions, 23
human resources data, 34
lack of, 51
Hurwicz criterion, 298-299
hypergeometric probability distribution, 181
hypothesis testing, 189-194
I
IBM's SPSS software, 40
IMF (International Monetary Fund), 35
implementation specialists, 52
importance of business analytics
  applications to enhance decision-making, 23-24
  new sources of data, 24-25
  overview, 17-18
  providing answers to questions, 18-20
  strategy for competitive advantage, 20-22
inability to delegate responsibility, 51
inability to prove success, 53
inconsistent values, checking for, 57
increasing customer profitability, 23
independent event outcomes, 170
infeasible solutions, 220-221
influence diagrams, 97
information policy, establishing, 54
information technology (IT)
computer hardware, 36
computer software, 36
data management technology, 37
data marts, 38
data mining, 38-40
data warehouses, 38
database encyclopedia content, 37
DBMS (database management systems), 36-37
infrastructure, 36
networking and telecommunications technology, 37
INFORMS, 30
Initial Cluster Centers table, 103-104
innovation, achieving with business analytics, 22
Insufficient Reason, Principle of, 168
integer linear programming. See IP (integer programming)
integer programming. See IP (integer programming)
integrated processes, lack of, 51
internal data sources, 34
International Monetary Fund (IMF), 35
interval data, 8
intervals (confidence), 82-84
inventory shortage, 285
IP (integer programming), 119
explained, 249-250
IP problems/models, solving, 250
maximization IP problem, 251
minimization IP problem, 252
mixed integer programming problem/model, 249-250
practice problems, 254-255
supply chain shipping case study, 153-155
ZOP (zero-one programming) explained, 250
problems/models solving, 253-254
IT (information technology)
computer hardware, 36
computer software, 36
data management technology, 37
data marts, 38
data mining, 38-40
data warehouses, 38
database encyclopedia content, 38
DBMS (database management systems), 36-37
infrastructure, 36
networking and telecommunications technology, 37
K-mean clustering, 103-104
Kolmogorov-Smirnov (One-Way) tests, 193
kurtosis, 74
L
Laplace criterion, 297-298
leadership, lack of, 50
level of significance, 189
limited context perception, 50
Lindo Systems LINGO. See LINGO
line charts
explained, 70
marketing/planning case study, 90
linear forecasts, 151
Linear Programming. See LP (Linear Programming)
linear regression model, formulating, 151
linearity in LP (Linear Programming) models, 223
LINGO, 40
downloading, 214
IP problems/models, solving
maximization IP problem, 251
minimization IP problem, 252
LP (Linear Programming) solutions
computer-based solution with simplex method, 211-218
infeasible solutions, 220-221
marketing/planning case study, 129-131
practice problems, 224-228
supply chain shipping case study, 155-157
unbounded solutions, 220
overview, 40
primal maximization problems, 231-238
primal minimization problems, 238-243
trial versions, 214
ZOP (zero-one programming) problems/models, solving, 253-254
little data, 3, 17-18
logic-driven models, 96-98
cause-and-effect diagrams, 97
influence diagrams, 97
loss values, expected opportunity loss criterion, 295-297
LP (Linear Programming), 119
applied LP model, 196
blending formulations, 221-222
computer-based solutions with simplex method, 211-212
LINGO solution, 214-218
simplex variables, 212-214
constraints, 198-200
duality
duality practice problems, 245-247
economic value of resources, determining, 244
informational value of, 230
J-K
joint probability, 169
judgment sampling, 80
justification, lack of, 53
overview, 229
primal maximization problems, 231-238
primal minimization problems, 238-243
sensitivity analysis, 230-231
generalized LP model, 196
infeasible solutions, 220-221
maximization models, 195-196
minimization models, 195-196
multidimensional decision variable formulations, 222-223
necessary assumptions, 223-224
nonnegativity and given requirements, 200
objective function, 197-198
overview, 195-196
practice problems, 224-228
problem/model formulation
butcher problem example, 202-204
Clarke Special Parts problem example, 208-209
customer service problem example, 207-208
diet problem example, 204-206
farming problem example, 206-207
Federal Division problem example, 209-211
marketing/planning case study, 127-129
stepwise procedure, 201-202
unbounded solutions, 220

M

MAD (mean absolute deviation), 150, 278
management issues, 54
change management, 58-60
  best practices, 60
targets, 59
ensuring data quality, 56-57
establishing information policy, 54
measuring business analytics contribution, 58
outsourcing business analytics, 55-56
advantages of, 55
disadvantages of, 55-56
MAPE (mean absolute percentage error), 279
margin probability, 307, 311
marketing/planning case study, 87
case study background, 87-88, 104-105, 127
descriptive analytics, 88-92
predictive analytics
  forecasting model formula, 111
  F-ratio statistic, 109-110
  R-Square statistics, 109
  step-wise multiple regression, 105-106
predictive analytics analysis, 104-113
prescriptive analytics
  formulation of LP marketing/planning
  model, 127-129
  predictive validity, 131
  solution for LP marketing/planning model,
  129-131
matrix organization structure, 48
maximax criterion, 292, 298
maximin criterion, 293, 298
maximization IP problem, solving, 251
maximization models
  LP (Linear Programming), 195-196
    primal maximization problems, 231-238
maximum/minimum, 74
mean, 74
mean absolute deviation (MAD), 150, 278
mean absolute percentage error (MAPE), 279
mean square error (MSE), 278
measured performance, 60
measuring business analytics contribution, 58
median, 74
merchandize strategy optimization, 23
methods (sampling), 79-81
Microsoft Excel, 39
minimax criterion, 299-301
minimization models
  LP (Linear Programming), 195-196
    minimization IP problem, solving, 252
    primal minimization problems, 238-243
minimum/maximum, 74
mining data. See data mining
mixed integer programming problem/model, 249-250
MLP (Multilayer Perception), 102
mobile analytics, 25
mode, 74
model fitting, 275-276
modeling
  cubic model forecasts
    formula for cubic regression models, 150
    supply chain shipping case study, 149-151
 DT (decision theory)
    decision environments, 290-291
    model elements, 290
    model formulation, 291-292
    overview, 289
fitting models to data, 275-276
forecasting models. See forecasting
linear regression model, 151
LP (Linear Programming)
  applied LP model, 196
  blending formulations, 221-222
  computer-based solutions with simplex
    method, 211-219
  constraints, 198-200
  generalized LP model, 196
  infeasible solutions, 220-221
  maximization models, 195-196
  minimization models, 195-196
  multidimensional decision variable
    formulations, 222-223
  necessary assumptions, 223-224
nonnegativity and given requirements, 200
objective function, 197-198
problem/model formulation, 201-211
unbounded solutions, 220
model selection, 277-279
multiple regression models, 267
application, 268-269
limitations in forecasting time series data, 269
parameters for models, selecting, 277-279
predictive modeling
data-driven models, 98
logic-driven models, 96-98
prescriptive modeling, 119
case studies, 119
decision analysis, 119
integer programming. See IP (integer programming)
linear programming. See LP (Linear Programming)
nonlinear optimization, 119-126
other methodologies, 120
simulation, 120
quadratic regression model, 151, 276
simple regression model, 262
assumptions, 266
computer-based solution, 263-266
formula, 262
model for trend, 262-263
statistical values and tests, 266-267
simulation, 281
deterministic simulation, 281-282
practice problems, 288
probabilistic simulation, 282-288
monitoring analysts, 52
Monte Carlo simulation method
application, 284-288
procedure, 282-284
MSE (mean square error), 278
multidimensional decision variable formulations, 222-223
Multilayer Perception (MLP), 102
multiple regression models, 9
application, 268-269
explained, 267
formula, 267
limitations in forecasting time series data, 269
multiplication, rules of, 170-173
mutually exclusive events, 169

N
N function, 74
need for business analytics
applications to enhance decision-making, 23-24
new sources of data, 24-25
overview, 17-18

O
objective approach to probability, 167
objective function, 125, 197-198, 201-202
ODMP (organization decision-making process), 10-12
operations efficiency, achieving with business analytics, 22
optimal shipping schedule, determining, 155-157
optimization, nonlinear, 119-126
deterministic simulation, 125-126
other methodologies, 126
SAS curve fitting, 121-125

nonnegativity, 129, 200
nonparametric hypothesis testing, 189, 193-194
nonparametric tests, 193
NORMAL function, 185
normal probability distribution, 181-186
null hypothesis, 189

organization decision-making process (ODMP), 10-12
organization structures, 46-51
centralized BA organization structure, 49-50
functional organization structure, 48
hierarchical relationships, 46
matrix organization structure, 48
project structure, 47-48
reasons for BA initiative and organization failure, 50-51
as target of change management, 59
organizational planning, 20
origin of probabilities, 294
outcomes, 173
outliers, checking for, 57
outsourcing business analytics, 55-56
advantages of, 55
disadvantages of, 55-56
P
parameter behavior, 283
parameters for models, selecting, 277-279
parametric hypothesis testing, 191-194
payoffs (DT), 290
period sampling, 79
permutations, 163-164
personnel, 30-33
administrators, 31
BAP (Business Analytics Professional) exam, 30-31
designers, 31
developers, 31
skills and competency requirements, 32-33
solution experts, 31
as target of change management, 59
technical specialists, 31
physical proximity, belief of, 50
pie charts, 71
planning (organizational), 20
Poisson probability distribution, 178-180
policy (information), 54
posterior probabilities, 312
practice problems
DT (decision theory), 314-319
duality, 245-247
forecasting, 279
IP (integer programming), 254-255
LP (Linear Programming), 224-228
simulation, 288
predictive analytics, 98
analytic purposes and tools, 5
data mining, 99-104
methodologies, 101-104
simple illustration of, 100-101
data-driven models, 98
definition of, 4
explained, 95-96
logic-driven models, 96-98
cause-and-effect diagrams, 97
influence diagrams, 97
marketing/planning case study, 104-113
case study background, 104-105
forecasting model formula, 111
F-ratio statistic, 109-110
R-Square statistics, 109
step-wise multiple regression, 105-106
supply chain shipping case study, 146
cubic model forecasts, 149-150
developing forecasting models, 146-150
problem background and data, 138-139
resulting warehouse customer demand forecasts, 152-153
SAS curve-fitting analysis, 146-150
validating forecasting models, 150
predictive modeling, 98
data-driven models, 98
logic-driven models, 96-98
cause-and-effect diagrams, 97
influence diagrams, 97
prescriptive analytics
analytic purposes and tools, 5
definition of, 4
explained, 117-118
marketing/planning case study
case study background, 127
formulation of LP marketing/planning model, 127-129
predictive validity, 131
solution for LP marketing/planning model, 129-131
methodologies, 118
prescriptive modeling, 118-119
case studies, 119
decision analysis, 119
integer programming. See IP (integer programming)
linear programming. See LP (Linear Programming)
nonlinear optimization, 118-126
other methodologies, 120
simulation, 120
supply chain shipping case study, 153
demonstrating business performance improvement, 158-159
determining optimal shipping schedule, 155-157
problem background and data, 138-139
selecting and developing optimization shipping model, 153-155
summary of BA procedure for manufacturer, 157
prescriptive modeling, 118
case studies, 119
decision analysis, 119
IP (integer programming), 119
explained, 249-250
IP problems/models, solving, 250-252
practice problems, 254-255
ZOP (zero-one programming) problems/models, solving, 250, 253-254
linear programming. See LP (Linear Programming)
nonlinear optimization, 119-126
deterministic simulation, 125-126
other methodologies, 126
SAS curve fitting, 121-125
other methodologies, 120
simulation, 120
price leadership, achieving with business analytics, 22
primal maximization problems, 231-238
primal minimization problems, 238-243
primal problem, 229
Principle of Insufficient Reason, 168
privacy (data), 35-36
probabilistic simulation
Monte Carlo simulation method
application, 284-288
procedure, 282-284
overview, 282
probability. See also DT (decision theory)
Bayes’s theorem, 307-314
marginal probability, 307
Monte Carlo simulation method, 284-288
origin of probabilities, 294
probabilistic simulation, 282
Monte Carlo simulation method procedure,
282-284
overview, 282
probability concepts, 167
a priori probabilities, 168
Frequency Theory, 167-168
objective approach to probability, 167
Principle of Insufficient Reason, 168
rules of addition, 169-170
rules of multiplication, 170-173
subjective approach to probability, 168
probability distributions, 173-174
binomial probability distribution, 175-177
exponential probability distribution,
186-188
geometric probability distribution, 180
hypergeometric probability distribution,
181
normal probability distribution, 181-186
Poisson probability distribution, 178-180
random variables, 173
probability density functions, 174
probability distributions, 84-86, 98, 173-174
continuous probability distributions, 181-188
exponential probability distribution,
186-188
normal probability distribution, 181-186
standard normal probability distribution,
183-184
discrete probability distributions, 174-180
binomial probability distribution, 175-177
geometric probability distribution, 180
hypergeometric probability distribution,
181
Poisson probability distribution, 178-180
posterior probabilities, 312
random variables, 173
Proc Cluster, 103-104
PROC REG DATA function, 264
process of business analytics
data measurement scales, 8
explained, 7-10
integrated processes, lack of, 51
relationship with organization decision-making
process (ODMP), 10-12
product data, 34
product differentiation, achieving with business
analytics, 22
production data, 34
project structure, 47-48
providing answers to questions, 18-20
Q
quadratic forecasts, 151
quadratic regression model, 151, 276
quality of data
ensuring, 56-57
overview, 35-36
Query Drilldown, 7
questionnaires, 34
questions business analytics seeks to answer, 18
quota sampling, 80
R
Radial Basis Function (RBF), 102
random variables, 173
continuous random variables, 173-174
discrete random variables, 173
random variation, 261
range, 74
ratio data, 8
RBF (Radial Basis Function), 102
reducing risk, 23
redundant constraints, 218
regression analysis, 98
fitting models to data, 275-276
multiple regression models
application, 268-269
explained, 267
formula, 267
limitations in forecasting time
series data, 269
quadratic regression model, 276
simple exponential smoothing
example of, 271-272
explained, 270-271
formula, 270
simple regression model, 262
assumptions, 266
computer-based solution, 263-266
model for trend, 262-263
statistical values and tests, 266-267
smoothing averages
  example of, 274
  explained, 273-274
  formula, 273
step-wise multiple regression, 107-109
relevance, checking for, 57
relevant ranges, 230
reordering data, 89
repetitions, 166
replacement, sampling without, 171
responsibility, inability to delegate, 51
revised probabilities, 312
risk
decision-making under risk, 293
  EV (expected value) criterion, 294-295
  expected opportunity loss criterion, 295-297
  origin of probabilities, 294
  explained, 290
  risk reduction, 23
roles (team), 51-52
R-Square statistics, 109
R-Squared (Adjusted) statistic, 110-111
rules
  of addition, 169-170
  of multiplication, 170-173
run testing, 193

S
SALES_DATA SAS file, creating, 64-68
sample variance, 74
sampling, 98
  sampling estimation, 82-84
  sampling methods, 79-81
  sampling without replacement, 171
SAS Analytics Pro, 7, 40
  big data simulation, 288
  case studies. See case studies
charts
  bar charts, 69
  column charts, 70
  histograms, 73
  line charts, 70
  marketing/planning case study, 88-92
  pie charts, 71
  scatter charts, 72
curve fitting
  for nonlinear optimization, 121-125
  supply chain shipping case study, 146-150
descriptive analytics
charts, 69-74
  confidence intervals, 82-84
  data set files, creating, 64-68
  descriptive statistics, 74-79
  marketing/planning case study, 88-92
  sampling analysis, 81
  supply chain shipping case study, 139-145
deterministic simulation, 125-126
EXPO function, 188
fitting models to data, 275-276
NORMAL function, 185
other methodologies, 126
predictive analytics
  association network capabilities, 102
  forecasting model formula, 111
  F-ratio statistic, 109-110
  K-Means cluster software, 103-104
  marketing/planning case study, 104-113
  R-Square statistics, 109
  R-Squared (Adjusted) statistic, 110-111
  step-wise multiple regression, 107-109
  supply chain shipping case study, 146-153
prescriptive analytics, 153-159
PROC REG DATA function, 264
  simple regression model function, 265
t-test statistics, 193
scatter charts, 72
seasonal variation, 260
selecting
  models, 277-279
  parameters for models, 277-279
senior management support, 60
sensitivity analysis
  economic value of resources, determining, 244
  overview, 230-231
  primal maximization problems, 231-238
  primal minimization problems, 238-243
sequences, 101
  data mining, 39
  sequential decisions and decision trees, 303-306
sequential decisions, 303-306
service effectiveness, achieving with business analytics, 22
significance, level of, 189
simple exponential smoothing
  example of, 271-272
  explained, 270-271
  formula, 270
simple random sampling, 80
simple regression model, 262
  assumptions, 266
  computer-based solution, 263-266
  model for trend, 262-263
  statistical values and tests, 266-267
simplex method, 211-212
  LINGO, 214-218
  simplex variables, 212-214
    artificial variables, 214
    slack variables, 212-213
    surplus variables, 213
simplex variables, 212-214
  artificial variables, 214
  slack variables, 212-213
  surplus variables, 213
simulation, 98, 120, 281
  computer simulation methods, 288
  deterministic simulation, 125-126, 281-282
  practice problems, 288
  probabilistic simulation, 282
    Monte Carlo simulation method, 282-284
    Monte Carlo simulation method application, 284-288
skewedness, 74
skill requirements for business analytics personnel, 32-33
slack variables, 212-213
smoothing
  simple exponential smoothing
    equation, 270
    example of, 271-272
    explained, 270-271
  smoothing averages
    example of, 274
    explained, 273-274
    formula, 273
smoothing averages
  example of, 274
  explained, 273-274
  formula, 273
  supply chain shipping case study, 152
social media analytics, 24-25
software, 36. See also specific software
solution experts, 31
Solver, 39
SPSS software, 40, 102
standard deviation, 74
standard error, 74
standard normal probability distribution, 84-85, 183-184
states of nature (DT), 290
stating nonnegativity and given requirements, 129
statistical testing, 189-194
statistical tools
  charts, 69-74
    bar charts, 69
    column charts, 70
    histograms, 73
    line charts, 70
    pie charts, 71
    scatter charts, 72
  confidence intervals, 82-84
  counting, 163
    combinations, 165
    permutations, 163-164
    repetitions, 166
descriptive statistics, 74-79
Durbin-Watson Autocorrelation Test, 269
forecast accuracy statistics
  MAD (mean absolute deviation), 278
  MAPE (mean absolute percentage error), 279
  MSE (mean square error), 278
F-ratio statistic, 109-110
probability concepts, 167
  a priori probabilities, 168
  conditional probabilities, 172
  Frequency Theory, 167-168
  objective approach to probability, 167
  Principle of Insufficient Reason, 168
  rules of addition, 169-170
  rules of multiplication, 170-173
  subjective approach to probability, 168
probability distributions, 173-174
  binomial probability distribution, 175-177
  exponential probability distribution, 186-188
  geometric probability distribution, 180
  hypergeometric probability distribution, 181
  normal probability distribution, 181-186
  Poisson probability distribution, 178-180
  random variables, 173
R-Square statistics, 109
R-Squared (Adjusted) statistic, 110-111
statistical testing, 189-194
step-wise multiple regression, 107-109
strategy for competitive advantage, 20-22
stratified random sampling, 80
structured data analytics, 25
subjective approach to probability, 168
success, proving, 53
sum, 74
supply chain shipping case study
  descriptive analytics analysis, 139-145
    actual monthly customer demand in motors, 140-142
    Chicago customer demand (graph), 143
    estimated shipping costs per motor, 139-140
    Houston customer demand (graph), 144
    Kansas City customer demand (graph), 143
    Little Rock customer demand (graph), 145
    Oklahoma City customer demand (graph), 144
    Omaha customer demand (graph), 145
  predictive analytics analysis, 146
    cubic model forecasts, 149-150
    developing forecasting models, 146-150
    resulting warehouse customer demand forecasts, 152-153
    SAS curve-fitting analysis, 146-150
    validating forecasting models, 150
prescriptive analysis, 153

demonstrating business performance improvement, 158-159
determining optimal shipping schedule, 155-157
selecting and developing optimization shipping model, 153-155
summary of BA procedure for manufacturer, 157

problem background and data, 137-138

support, lack of, 50

surplus variables, 213

sustainability, achieving with business analytics, 22

systematic random sampling, 80

T

targets of change management, 59
tasks as target of change management, 59
teams, 51-53

  collaboration, 52-53
  participant roles, 51-52
  reasons for team failures, 53
technical specialists, 31
technology as target of change management, 59
technology coefficients, 198
testing (statistical), 189-194
text analytics, 24-25
text mining, 39
time series data, variation in
cyclical variation, 261
forecasting methods, 261-262
overview, 258-260
random variation, 261
seasonal variation, 260
trend variation, 260
trend variation, 260
trends, predicting with simple regression model, 262-263
trials, 173
t-t-test: Paired Two Sample Means, 191
type of problem, determining, 128

U

unbounded solutions, 220

uncertainty

  decision-making under uncertainty, 297
  Hurwicz criterion, 298-299
  Laplace criterion, 297-298
  maximax criterion, 298
  maximin criterion, 298
  minimax criterion, 299-301
  explained, 291

  U.S. Census, 35

V

validating forecasting models, 150

value

  EV (expected value) criterion, 294-295
  EVPI (expected value of perfect information), 301-302
  expected opportunity loss criterion, 295-297
  failure to provide value, 53
  inconsistent values, checking for, 57

variables

  slack variables, 212-213
  surplus variables, 213

variance, 74, 214

variation in time series data
cyclical variation, 261
forecasting methods, 261-262
overview, 258-260
random variation, 261
seasonal variation, 260
trend variation, 260

visualizing data

  charts, 69-74
    bar charts, 69
    column charts, 70
    histograms, 73
    line charts, 70
    pie charts, 71
    scatter charts, 72

  diagrams
    cause-and-effect diagrams, 97
    influence diagrams, 97

W

warehouses (data), 38
web logs, 34
web mining, 39
Wilcoxon Signed-Rank tests, 193

X-Y-Z

Z values, 84-85, 184

zero-one programming (ZOP) model

explained, 250
problems/models, solving, 253-254