
CHAPTER 1

# Introduction to Business Analytics

## Learning Objectives

After studying this chapter, you will be able to:

- Define business analytics.
- Explain why analytics is important in today's business environment.
- State some typical examples of business applications in which analytics would be beneficial.
- Summarize the evolution of business analytics and explain the concepts of business intelligence, operations research and management science, and decision support systems.
- Explain and provide examples of descriptive, predictive, and prescriptive analytics.
- State examples of how data are used in business.
- Explain the difference between a data set and a database.
- Define a metric and explain the concepts of measurement and measures.
- Explain the difference between a discrete metric and continuous metric, and provide examples of each.
- Describe the four groups of data classification, categorical, ordinal, interval, and ratio, and provide examples of each.
- Explain the concept of a model and various ways a model can be characterized.
- Define and list the elements of a decision model.
- Define and provide an example of an influence diagram.
- Use influence diagrams to build simple mathematical models.
- Use predictive models to compute model outputs.
- Explain the difference between uncertainty and risk.
- Define the terms *optimization*, *objective function*, and *optimal solution*.
- Explain the difference between a deterministic and stochastic decision model.
- List and explain the steps in the problem-solving process.

**Most of** you have likely been to a zoo, seen the animals, had something to eat, and bought some souvenirs. You probably wouldn't think that managing a zoo is very difficult; after all, it's just feeding and taking care of the animals, right? A zoo might be the last place that you would expect to find business analytics being used, but not anymore. The Cincinnati Zoo & Botanical Garden has been an "early adopter" and one of the first organizations of its kind to exploit business analytics.[1]

Despite generating more than two-thirds of its budget through its own fund-raising efforts, the zoo wanted to reduce its reliance on local tax subsidies even further by increasing visitor attendance and revenues from secondary sources such as membership, food and retail outlets. The zoo's senior management surmised that the best way to realize more value from each visit was to offer visitors a truly transformed customer experience. By using business analytics to gain greater insight into visitors' behavior and tailoring operations to their preferences, the zoo expected to increase attendance, boost membership, and maximize sales.

The project team—which consisted of consultants from IBM and BrightStar Partners, as well as senior executives from the zoo—began translating the organization's goals into technical solutions. The zoo worked to create a business analytics platform that was capable of delivering the desired goals by combining data from ticketing and point-of-sale systems throughout the zoo with membership information and geographical data gathered from the ZIP codes of all visitors. This enabled the creation of reports and dashboards that give everyone from senior managers to zoo staff access to real-time information that helps them optimize operational management and transform the customer experience.

By integrating weather forecast data, the zoo is able to compare current forecasts with historic attendance and sales data, supporting better decision-making for labor scheduling and inventory planning. Another area where the solution delivers new insight is food service. By opening food outlets at specific times of day when demand is highest (for example, keeping ice cream kiosks open in the final hour before the zoo closes), the zoo has been able to increase sales significantly. The zoo has been able to increase attendance and revenues dramatically, resulting in annual ROI of 411%. The business

[1]Source: IBM Software Business Analtyics, "Cincinnati Zoo transforms customer experience and boosts profits," © IBM Corporation 2012.

analytics initiative paid for itself within three months, and delivers, on average, benefits of $738,212 per year. Specifically,

- The zoo has seen a 4.2% rise in ticket sales by targeting potential visitors who live in specific ZIP codes.
- Food revenues increased by 25% by optimizing the mix of products on sale and adapting selling practices to match peak purchase times.
- Eliminating slow-selling products and targeting visitors with specific promotions enabled an 18% increase in merchandise sales.
- Cut marketing expenditure, saving $40,000 in the first year, and reduced advertising expenditure by 43% by eliminating ineffective campaigns and segmenting customers for more targeted marketing.

Because of the zoo's success, other organizations such as Point Defiance Zoo & Aquarium, in Washington state, and History Colorado, a museum in Denver, have embarked on similar initiatives.

In recent years, analytics has become increasingly important in the world of business, particularly as organizations have access to more and more data. Managers today no longer make decisions based on pure judgment and experience; they rely on factual data and the ability to manipulate and analyze data to support their decisions. As a result, many companies have recently established analytics departments; for instance, IBM reorganized its consulting business and established a new 4,000-person organization focusing on analytics.[2] Companies are increasingly seeking business graduates with the ability to understand and use analytics. In fact, in 2011, the U.S. Bureau of Labor Statistics predicted a 24% increase in demand for professionals with analytics expertise.

No matter what your academic business concentration is, you will most likely be a future user of analytics to some extent and work with analytics professionals. The purpose of this book is to provide you with a basic introduction to the concepts, methods, and models used in business analytics so that you will develop not only an appreciation for its capabilities to support and enhance business decisions, but also the ability to use business analytics at an elementary level in your work. In this chapter, we introduce you to the field of business analytics, and set the foundation for many of the concepts and techniques that you will learn.

[2]Matthew J. Liberatore and Wenhong Luo, "The Analytics Movement: Implications for Operations Research," *Interfaces*, 40, 4 (July–August 2010): 313–324.

## What Is Business Analytics?

Everyone makes decisions. Individuals face personal decisions such as choosing a college or graduate program, making product purchases, selecting a mortgage instrument, and investing for retirement. Managers in business organizations make numerous decisions every day. Some of these decisions include what products to make and how to price them, where to locate facilities, how many people to hire, where to allocate advertising budgets, whether or not to outsource a business function or make a capital investment, and how to schedule production. Many of these decisions have significant economic consequences; moreover, they are difficult to make because of uncertain data and imperfect information about the future. Thus, managers need good information and assistance to make such critical decisions that will impact not only their companies but also their careers. What makes business decisions complicated today is the overwhelming amount of available data and information. Data to support business decisions—including those specifically collected by firms as well as through the Internet and social media such as Facebook—are growing exponentially and becoming increasingly difficult to understand and use. This is one of the reasons why analytics is important in today's business environment.

**Business analytics**, or simply **analytics**, is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions. Business analytics is "a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving."[3] Business analytics is supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software packages such as SAS or Minitab, and more-complex business intelligence suites that integrate data with analytical software.

Tools and techniques of business analytics are used across many areas in a wide variety of organizations to improve the management of customer relationships, financial and marketing activities, human capital, supply chains, and many other areas. Leading banks use analytics to predict and prevent credit fraud. Manufacturers use analytics for production planning, purchasing, and inventory management. Retailers use analytics to recommend products to customers and optimize marketing promotions. Pharmaceutical firms use it to get life-saving drugs to market more quickly. The leisure and vacation industries use analytics to analyze historical sales data, understand customer behavior, improve Web site design, and optimize schedules and bookings. Airlines and hotels use analytics to dynamically set prices over time to maximize revenue. Even sports teams are using business analytics to determine both game strategy and optimal ticket prices.[4] Among the many organizations that use analytics to make strategic decisions and manage day-to-day operations are Harrah's Entertainment, the Oakland Athletics baseball and New England Patriots football teams, Amazon.com, Procter & Gamble, United Parcel Service (UPS), and Capital One bank. It was reported that nearly all firms with revenues of more than $100 million are using some form of business analytics.

Some common types of decisions that can be enhanced by using analytics include

- pricing (for example, setting prices for consumer and industrial goods, government contracts, and maintenance contracts),
- customer segmentation (for example, identifying and targeting key customer groups in retail, insurance, and credit card industries),

---

[3]Liberatore and Luo, "The Analytics Movement."
[4]Jim Davis, "8 Essentials of Business Analytics," in "Brain Trust—Enabling the Confident Enterprise with Business Analytics" (Cary, NC: SAS Institute, Inc., 2010): 27–29. www.sas.com/bareport

- merchandising (for example, determining brands to buy, quantities, and allocations),
- location (for example, finding the best location for bank branches and ATMs, or where to service industrial equipment),

and many others in operations and supply chains, finance, marketing, and human resources—in fact, in every discipline of business.[5]

Various research studies have discovered strong relationships between a company's performance in terms of profitability, revenue, and shareholder return and its use of analytics. Top-performing organizations (those that outperform their competitors) are three times more likely to be sophisticated in their use of analytics than lower performers and are more likely to state that their use of analytics differentiates them from competitors.[6] However, research has also suggested that organizations are overwhelmed by data and struggle to understand how to use data to achieve business results and that most organizations simply don't understand how to use analytics to improve their businesses. Thus, understanding the capabilities and techniques of analytics is vital to managing in today's business environment.

One of the emerging applications of analytics is helping businesses learn from social media and exploit social media data for strategic advantage.[7] Using analytics, firms can integrate social media data with traditional data sources such as customer surveys, focus groups, and sales data; understand trends and customer perceptions of their products; and create informative reports to assist marketing managers and product designers.

## Evolution of Business Analytics

Analytical methods, in one form or another, have been used in business for more than a century. However, the modern evolution of analytics began with the introduction of computers in the late 1940s and their development through the 1960s and beyond. Early computers provided the ability to store and analyze data in ways that were either very difficult or impossible to do so manually. This facilitated the collection, management, analysis, and reporting of data, which is often called **business intelligence (BI)**, a term that was coined in 1958 by an IBM researcher, Hans Peter Luhn.[8] Business intelligence software can answer basic questions such as "How many units did we sell last month?" "What products did customers buy and how much did they spend?" "How many credit card transactions were completed yesterday?" Using BI, we can create simple rules to flag exceptions automatically, for example, a bank can easily identify transactions greater than $10,000 to report to the Internal Revenue Service.[9] BI has evolved into the modern discipline we now call **information systems (IS)**.

---

[5]Thomas H. Davenport, "How Organizations Make Better Decisions," edited excerpt of an article distributed by the International Institute for Analytics published in "Brain Trust—Enabling the Confident Enterprise with Business Analytics" (Cary, NC: SAS Institute, Inc., 2010): 8–11. www.sas.com/bareport
[6]Thomas H. Davenport and Jeanne G. Harris, *Competing on Analytics* (Boston: Harvard Business School Press, 2007): 46; Michael S. Hopkins, Steve LaValle, Fred Balboni, Nina Kruschwitz, and Rebecca Shockley, "10 Data Points: Information and Analytics at Work," *MIT Sloan Management Review*, 52, 1 (Fall 2010): 27–31.
[7]Jim Davis, "Convergence—Taking Social Media from Talk to Action," *SASCOM* (First Quarter 2011): 17.
[8]H. P. Luhn, "A Business Intelligence System." *IBM Journal* (October 1958).
[9]Jim Davis, "Business Analytics: Helping You Put an Informed Foot Forward," in "Brain Trust—Enabling the Confident Enterprise with Business Analytics," (Cary, NC: SAS Institute, Inc., 2010): 4–7. www.sas.com/bareport

**Statistics** has a long and rich history, yet only rather recently has it been recognized as an important element of business, driven to a large extent by the massive growth of data in today's world. Google's chief economist stated that statisticians surely have the "really sexy job" for the next decade.[10] Statistical methods allow us to gain a richer understanding of data that goes beyond business intelligence reporting by not only summarizing data succinctly but also finding unknown and interesting relationships among the data. Statistical methods include the basic tools of description, exploration, estimation, and inference, as well as more advanced techniques like regression, forecasting, and data mining.

Much of modern business analytics stems from the analysis and solution of complex decision problems using mathematical or computer-based models—a discipline known as operations research, or management science. *Operations research (OR)* was born from efforts to improve military operations prior to and during World War II. After the war, scientists recognized that the mathematical tools and techniques developed for military applications could be applied successfully to problems in business and industry. A significant amount of research was carried on in public and private think tanks during the late 1940s and through the 1950s. As the focus on business applications expanded, the term *management science (MS)* became more prevalent. Many people use the terms *operations research* and *management science* interchangeably, and the field became known as **Operations Research/Management Science (OR/MS)**. Many OR/MS applications use **modeling and optimization**—techniques for translating real problems into mathematics, spreadsheets, or other computer languages, and using them to find the best ("optimal") solutions and decisions. INFORMS, the Institute for Operations Research and the Management Sciences, is the leading professional society devoted to OR/MS and analytics, and publishes a bimonthly magazine called *Analytics* (http://analytics-magazine.com/). Digital subscriptions may be obtained free of charge at the Web site.

**Decision support systems (DSS)** began to evolve in the 1960s by combining business intelligence concepts with OR/MS models to create analytical-based computer systems to support decision making. DSSs include three components:

1. *Data management*. The data management component includes databases for storing data and allows the user to input, retrieve, update, and manipulate data.
2. *Model management*. The model management component consists of various statistical tools and management science models and allows the user to easily build, manipulate, analyze, and solve models.
3. *Communication system*. The communication system component provides the interface necessary for the user to interact with the data and model management components.[11]

DSSs have been used for many applications, including pension fund management, portfolio management, work-shift scheduling, global manufacturing and facility location, advertising-budget allocation, media planning, distribution planning, airline operations planning, inventory control, library management, classroom assignment, nurse scheduling, blood distribution, water pollution control, ski-area design, police-beat design, and energy planning.[12]

---

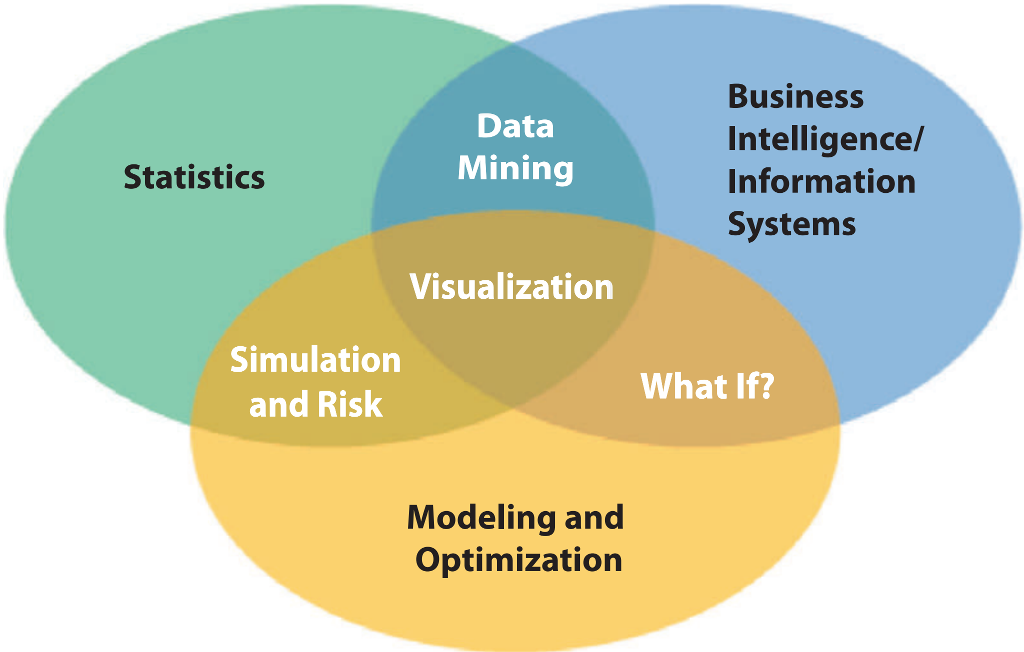
[10] James J. Swain, "Statistical Software in the Age of the Geek," Analytics-magazine.org, March/April 2013, pp. 48–55. www.informs.org

[11] William E. Leigh and Michael E. Doherty, *Decision Support and Expert Systems* (Cincinnati, OH: South-Western Publishing Co., 1986).

[12] H. B. Eom and S. M. Lee, "A Survey of Decision Support System Applications (1971–April 1988)," *Interfaces*, 20, 3 (May–June 1990): 65–79.

Figure 1.1

A Visual Perspective of Business Analytics




Modern business analytics can be viewed as an integration of BI/IS, statistics, and modeling and optimization as illustrated in Figure 1.1. While the core topics are traditional and have been used for decades, the uniqueness lies in their intersections. For example, **data mining** is focused on better understanding characteristics and patterns among variables in large databases using a variety of statistical and analytical tools. Many standard statistical tools as well as more advanced ones are used extensively in data mining. **Simulation and risk analysis** relies on spreadsheet models and statistical analysis to examine the impacts of uncertainty in the estimates and their potential interaction with one another on the output variable of interest. Spreadsheets and formal models allow one to manipulate data to perform **what-if analysis**—how specific combinations of inputs that reflect key assumptions will affect model outputs. What-if analysis is also used to assess the sensitivity of optimization models to changes in data inputs and provide better insight for making good decisions.

Perhaps the most useful component of business analytics, which makes it truly unique, is the center of Figure 1.1—**visualization**. Visualizing data and results of analyses provide a way of easily communicating data at all levels of a business and can reveal surprising patterns and relationships. Software such as IBM's Cognos system exploits data visualization for query and reporting, data analysis, dashboard presentations, and scorecards linking strategy to operations. The Cincinnati Zoo, for example, has used this on an iPad to display hourly, daily, and monthly reports of attendance, food and retail location revenues and sales, and other metrics for prediction and marketing strategies. UPS uses telematics to capture vehicle data and display them to help make decisions to improve efficiency and performance. You may have seen a **tag cloud** (see the graphic at the beginning of this chapter), which is a visualization of text that shows words that appear more frequently using larger fonts.

The most influential developments that propelled the use of business analytics have been the personal computer and spreadsheet technology. Personal computers and spreadsheets provide a convenient way to manage data, calculations, and visual graphics simultaneously, using intuitive representations instead of abstract mathematical notation. Although the early

## Analytics in Practice: Harrah's Entertainment[13]

One of the most cited examples of the use of analytics in business is Harrah's Entertainment. Harrah's owns numerous hotels and casinos and uses analytics to support revenue management activities, which involve selling the right resources to the right customer at the right price to maximize revenue and profit. The gaming industry views hotel rooms as incentives or rewards to support casino gaming activities and revenues, not as revenue-maximizing assets. Therefore, Harrah's objective is to set room rates and accept reservations to maximize the expected gaming profits from customers. They begin with collecting and tracking of customers' gaming activities (playing slot machines and casino games) using Harrah's "Total Rewards" card program, a customer loyalty program that provides rewards such as meals, discounted rooms, and other perks to customers based on the amount of money and time they spend at Harrah's. The data collected are used to segment customers into more than 20 groups based on their expected gaming activities. For each customer segment, analytics forecasts demand for hotel rooms by arrival date and length of stay. Then Harrah's uses a prescriptive model to set prices and allocate rooms to these customer segments. For example, the system might offer complimentary rooms to customers who are expected to generate a gaming profit of at least $400 but charge $325 for a room if the profit is expected to be only $100. Marketing can use the information to send promotional offers to targeted customer segments if it identifies low-occupancy rates for specific dates.

applications of spreadsheets were primarily in accounting and finance, spreadsheets have developed into powerful general-purpose managerial tools for applying techniques of business analytics. The power of analytics in a personal computing environment was noted some 20 years ago by business consultants Michael Hammer and James Champy, who said, "When accessible data is combined with easy-to-use analysis and modeling tools, frontline workers —when properly trained—suddenly have sophisticated decision-making capabilities."[14] Although many good analytics software packages are available to professionals, we use Microsoft Excel and a powerful add-in called *Analytic Solver Platform* throughout this book.

## Impacts and Challenges

The impact of applying business analytics can be significant. Companies report reduced costs, better risk management, faster decisions, better productivity, and enhanced bottom-line performance such as profitability and customer satisfaction. For example, 1-800-flowers.com uses analytic software to target print and online promotions with greater accuracy; change prices and offerings on its Web site (sometimes hourly); and optimize its marketing, shipping, distribution, and manufacturing operations, resulting in a $50 million cost savings in one year.[15]

Business analytics is changing how managers make decisions.[16] To thrive in today's business world, organizations must continually innovate to differentiate themselves from competitors, seek ways to grow revenue and market share, reduce costs, retain existing customers and acquire new ones, and become faster and leaner. IBM suggests that

[13]Based on Liberatore and Luo, "The Analytics Movement"; and Richard Metters et al., "The 'Killer Application' of Revenue Management: Harrah's Cherokee Casino & Hotel," *Interfaces*, 38, 3 (May–June 2008): 161–175.

[14]Michael Hammer and James Champy, *Reengineering the Corporation* (New York: HarperBusiness, 1993): 96.

[15]Jim Goodnight, "The Impact of Business Analytics on Performance and Profitability," in "Brain Trust—Enabling the Confident Enterprise with Business Analytics" (Cary, NC: SAS Institute, Inc., 2010): 4–7. www.sas.com/bareport

[16]*Analytics: The New Path to Value*, a joint MIT Sloan Management Review and IBM Institute for Business Value study.

traditional management approaches are evolving in today's analytics-driven environment to include more fact-based decisions as opposed to judgment and intuition, more prediction rather than reactive decisions, and the use of analytics by everyone at the point where decisions are made rather than relying on skilled experts in a consulting group.[17] Nevertheless, organizations face many challenges in developing analytics capabilities, including lack of understanding of how to use analytics, competing business priorities, insufficient analytical skills, difficulty in getting good data and sharing information, and not understanding the benefits versus perceived costs of analytics studies. Successful application of analytics requires more than just knowing the tools; it requires a high-level understanding of how analytics supports an organization's competitive strategy and effective execution that crosses multiple disciplines and managerial levels.

A 2011 survey by Bloomberg Businessweek Research Services and SAS concluded that business analytics is still in the "emerging stage" and is used only narrowly within business units, not across entire organizations. The study also noted that many organizations lack analytical talent, and those that do have analytical talent often don't know how to apply the results properly. While analytics is used as part of the decision-making process in many organizations, most business decisions are still based on intuition.[18] Therefore, while many challenges are apparent, many more opportunities exist. These opportunities are reflected in the job market for analytics professionals, or "data scientists," as some call them. The *Harvard Business Review* called data scientist "the sexiest job of the 21st century," and McKinsey & Company predicted a 50 to 60% shortfall in data scientists in the United States by 2018.[19]

## Scope of Business Analytics

Business analytics begins with the collection, organization, and manipulation of data and is supported by three major components:[20]

1. *Descriptive analytics*. Most businesses start with **descriptive analytics**—the use of data to understand past and current business performance and make informed decisions. Descriptive analytics is the most commonly used and most well-understood type of analytics. These techniques categorize, characterize, consolidate, and classify data to convert it into useful information for the purposes of understanding and analyzing business performance. Descriptive analytics summarizes data into meaningful charts and reports, for example, about budgets, sales, revenues, or cost. This process allows managers to obtain standard and customized reports and then drill down into the data and make queries to understand the impact of an advertising campaign, for example, review business performance to find problems or areas of opportunity, and identify patterns and trends in data. Typical questions that descriptive analytics helps answer are "How much did we sell in each region?" "What was our revenue and profit last quarter?" "How many and what types of complaints did we

[17]"Business Analytics and Optimization for the Intelligent Enterprise" (April 2009). www.ibm.com/qbs/intelligent-enterprise
[18]Bloomberg Businessweek Research Services and SAS, "The Current State of Business Analytics: Where Do We Go From Here?" (2011).
[19]Andrew Jennings, "What Makes a Good Data Scientist?" *Analytics Magazine* (July–August 2013): 8–13. www.analytics-magazine.org
[20]Parts of this section are adapted from Irv Lustig, Brenda Dietric, Christer Johnson, and Christopher Dziekan, "The Analytics Journey," *Analytics* (November/December 2010). www.analytics-magazine.org

resolve?" "Which factory has the lowest productivity?" Descriptive analytics also helps companies to classify customers into different segments, which enables them to develop specific marketing campaigns and advertising strategies.

2. *Predictive analytics*. **Predictive analytics** seeks to predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time. For example, a marketer might wish to predict the response of different customer segments to an advertising campaign, a commodities trader might wish to predict short-term movements in commodities prices, or a skiwear manufacturer might want to predict next season's demand for skiwear of a specific color and size. Predictive analytics can predict risk and find relationships in data not readily apparent with traditional analyses. Using advanced techniques, predictive analytics can help to detect hidden patterns in large quantities of data to segment and group data into coherent sets to predict behavior and detect trends. For instance, a bank manager might want to identify the most profitable customers or predict the chances that a loan applicant will default, or alert a credit-card customer to a potential fraudulent charge. Predictive analytics helps to answer questions such as "What will happen if demand falls by 10% or if supplier prices go up 5%?" "What do we expect to pay for fuel over the next several months?" "What is the risk of losing money in a new business venture?"
3. *Prescriptive analytics.* Many problems, such as aircraft or employee scheduling and supply chain design, for example, simply involve too many choices or alternatives for a human decision maker to effectively consider. **Prescriptive analytics** uses optimization to identify the best alternatives to minimize or maximize some objective. Prescriptive analytics is used in many areas of business, including operations, marketing, and finance. For example, we may determine the best pricing and advertising strategy to maximize revenue, the optimal amount of cash to store in ATMs, or the best mix of investments in a retirement portfolio to manage risk. The mathematical and statistical techniques of predictive analytics can also be combined with optimization to make decisions that take into account the uncertainty in the data. Prescriptive analytics addresses questions such as "How much should we produce to maximize profit?" "What is the best way of shipping goods from our factories to minimize costs?" "Should we change our plans if a natural disaster closes a supplier's factory: if so, by how much?"

## Analytics in Practice: Analytics in the Home Lending and Mortgage Industry[21]

Sometime during their lives, most Americans will receive a mortgage loan for a house or condominium. The process starts with an application. The application contains all pertinent information about the borrower that the lender will need. The bank or mortgage company then initiates a process that leads to a loan decision. It is here that key information about the borrower is provided by third-party providers. This information includes a credit report, verification of income, verification of assets, verification of employment, and an appraisal of the property among others. The result of the processing function is a complete loan file that contains all the information and documents needed to underwrite the loan, which is the next step in the process. Underwriting is where the loan application is evaluated for its risk. Underwriters evaluate whether the borrower can make payments on time, can afford to pay back the loan, and has sufficient collateral in the property to back up the

*(continued)*

[21]Contributed by Craig Zielazny, BlueNote Analytics, LLC.

loan. In the event the borrower defaults on their loan, the lender can sell the property to recover the amount of the loan. But, if the amount of the loan is greater than the value of the property, then the lender cannot recoup their money. If the underwriting process indicates that the borrower is creditworthy, has the capacity to repay the loan, and the value of the property in question is greater than the loan amount, then the loan is approved and will move to closing. Closing is the step where the borrower signs all the appropriate papers agreeing to the terms of the loan.

In reality, lenders have a lot of other work to do. First, they must perform a quality control review on a sample of the loan files that involves a manual examination of all the documents and information gathered. This process is designed to identify any mistakes that may have been made or information that is missing from the loan file. Because lenders do not have unlimited money to lend to borrowers, they frequently sell the loan to a third party so that they have fresh capital to lend to others. This occurs in what is called the secondary market. Freddie Mac and Fannie Mae are the two largest purchasers of mortgages in the secondary market. The final step in the process is servicing. Servicing includes all the activities associated with providing the customer service on the loan like processing payments, managing property taxes held in escrow, and answering questions about the loan.

In addition, the institution collects various operational data on the process to track its performance and efficiency, including the number of applications, loan types and amounts, cycle times (time to close the loan), bottlenecks in the process, and so on. Many different types of analytics are used:

Descriptive Analytics—This focuses on historical reporting, addressing such questions as:

- How many loan apps were taken each of the past 12 months?
- What was the total cycle time from app to close?
- What was the distribution of loan profitability by credit score and loan-to-value (LTV), which is the mortgage amount divided by the appraised value of the property.

Predictive Analytics—Predictive modeling use mathematical, spreadsheet, and statistical models, and address questions such as:

- What impact on loan volume will a given marketing program have?
- How many processors or underwriters are needed for a given loan volume?
- Will a given process change reduce cycle time?

Prescriptive Analytics—This involves the use of simulation or optimization to drive decisions. Typical questions include:

- What is the optimal staffing to achieve a given profitability constrained by a fixed cycle time?
- What is the optimal product mix to maximize profit constrained by fixed staffing?

The mortgage market has become much more dynamic in recent years due to rising home values, falling interest rates, new loan products, and an increased desire by home owners to utilize the equity in their homes as a financial resource. This has increased the complexity and variability of the mortgage process and created an opportunity for lenders to proactively use the data that are available to them as a tool for managing their business. To ensure that the process is efficient, effective and performed with quality, data and analytics are used every day to track what is done, who is doing it, and how long it takes.

A wide variety of tools are used to support business analytics. These include:

- Database queries and analysis
- “Dashboards” to report key performance measures
- Data visualization
- Statistical methods
- Spreadsheets and predictive models
- Scenario and “what-if” analyses
- Simulation

- Forecasting
- Data and text mining
- Optimization
- Social media, Web, and text analytics

Although the tools used in descriptive, predictive, and prescriptive analytics are different, many applications involve all three. Here is a typical example in retail operations.

## EXAMPLE 1.1 Retail Markdown Decisions[22]

**As you probably know from your shopping experiences, most department stores and fashion retailers clear their seasonal inventory by reducing prices. The key question they face is what prices should they set—and when should they set them—to meet inventory goals and maximize revenue? For example, suppose that a store has 100 bathing suits of a certain style that go on sale from April 1 and wants to sell all of them by the end of June. Over each week of the 12-week selling season, they can make a decision to discount the price. They face two decisions: When to reduce the price and by how much? This results in 24 decisions to make. For a major national chain that may carry thousands of products, this can easily result in millions of decisions that store managers have to make. Descriptive analytics can be used to examine historical data for similar products, such as the number of units sold, price at each point of sale, starting and ending inventories, and special promotions, newspaper ads, direct marketing ads, and so on, to understand what the results of past decisions achieved. Predictive analytics can be used to predict sales based on pricing decisions. Finally, prescriptive analytics can be applied to find the best set of pricing decisions to maximize the total revenue.**

## Software Support

Many companies, such as IBM, SAS, and Tableau have developed a variety of software and hardware solutions to support business analytics. For example, IBM's Cognos Express, an integrated business intelligence and planning solution designed to meet the needs of midsize companies, provides reporting, analysis, dashboard, scorecard, planning, budgeting, and forecasting capabilities. It's made up of several modules, including Cognos Express Reporter, for self-service reporting and ad hoc query; Cognos Express Advisor, for analysis and visualization; and Cognos Express Xcelerator, for Excel-based planning and business analysis. Information is presented to the business user in a business context that makes it easy to understand, with an easy to use interface they can quickly gain the insight they need from their data to make the right decisions and then take action for effective and efficient business optimization and outcome. SAS provides a variety of software that integrate data management, business intelligence, and analytics tools. SAS Analytics covers a wide range of capabilities, including predictive modeling and data mining, visualization, forecasting, optimization and model management, statistical analysis, text analytics, and more. Tableau Software provides simple drag and drop tools for visualizing data from spreadsheets and other databases. We encourage you to explore many of these products as you learn the basic principles of business analytics in this book.

[22]Inspired by a presentation by Radhika Kulkarni, SAS Institute, "Data-Driven Decisions: Role of Operations Research in Business Analytics," INFORMS Conference on Business Analytics and Operations Research, April 10–12, 2011.

## Data for Business Analytics

Since the dawn of the electronic age and the Internet, both individuals and organizations have had access to an enormous wealth of data and information. *Data* are numerical facts and figures that are collected through some type of measurement process. *Information* comes from analyzing data—that is, extracting meaning from data to support evaluation and decision making.

Data are used in virtually every major function in a business. Modern organizations—which include not only for-profit businesses but also nonprofit organizations—need good data to support a variety of company purposes, such as planning, reviewing company performance, improving operations, and comparing company performance with competitors' or best-practice benchmarks. Some examples of how data are used in business include the following:

- Annual reports summarize data about companies' profitability and market share both in numerical form and in charts and graphs to communicate with shareholders.
- Accountants conduct audits to determine whether figures reported on a firm's balance sheet fairly represent the actual data by examining samples (that is, subsets) of accounting data, such as accounts receivable.
- Financial analysts collect and analyze a variety of data to understand the contribution that a business provides to its shareholders. These typically include profitability, revenue growth, return on investment, asset utilization, operating margins, earnings per share, economic value added (EVA), shareholder value, and other relevant measures.
- Economists use data to help companies understand and predict population trends, interest rates, industry performance, consumer spending, and international trade. Such data are often obtained from external sources such as Standard & Poor's Compustat data sets, industry trade associations, or government databases.
- Marketing researchers collect and analyze extensive customer data. These data often consist of demographics, preferences and opinions, transaction and payment history, shopping behavior, and a lot more. Such data may be collected by surveys, personal interviews, focus groups, or from shopper loyalty cards.
- Operations managers use data on production performance, manufacturing quality, delivery times, order accuracy, supplier performance, productivity, costs, and environmental compliance to manage their operations.
- Human resource managers measure employee satisfaction, training costs, turnover, market innovation, training effectiveness, and skills development.

Such data may be gathered from primary sources such as internal company records and business transactions, automated data-capturing equipment, or customer market surveys and from secondary sources such as government and commercial data sources, custom research providers, and online research.

Perhaps the most important source of data today is data obtained from the Web. With today's technology, marketers collect extensive information about Web behaviors, such as the number of page views, visitor's country, time of view, length of time, origin and destination paths, products they searched for and viewed, products purchased, what reviews they read, and many others. Using analytics, marketers can learn what content is being viewed most often, what ads were clicked on, who the most frequent visitors are, and what types of visitors browse but don't buy. Not only can marketers understand what customers have done, but they can better predict what they intend to do in the future. For example, if a bank knows that a customer has browsed for mortgage rates and homeowner's insurance, they can target the customer with homeowner loans rather than credit cards or automobile loans. Traditional Web data are now being enhanced with social media data from Facebook, cell phones, and even Internet-connected gaming devices.

As one example, a home furnishings retailer wanted to increase the rate of sales for customers who browsed their Web site. They developed a large data set that covered more than 7,000 demographic, Web, catalog, and retail behavioral attributes for each customer. They used predictive analytics to determine how well a customer would respond to different e-mail marketing offers and customized promotions to individual customers. This not only helped them to determine where to most effectively spend marketing resources but doubled the response rate compared to previous marketing campaigns, with a projected multimillion dollar increase in sales.[23]

## Data Sets and Databases

A **data set** is simply a collection of data. Marketing survey responses, a table of historical stock prices, and a collection of measurements of dimensions of a manufactured item are examples of data sets. A **database** is a collection of related files containing records on people, places, or things. The people, places, or things for which we store and maintain information are called *entities*.[24] A database for an online retailer that sells instructional fitness books and DVDs, for instance, might consist of a file for three entities: publishers from which goods are purchased, customer sales transactions, and product inventory. A database file is usually organized in a two-dimensional table, where the columns correspond to each individual element of data (called *fields*, or *attributes*), and the rows represent records of related data elements. A key feature of computerized databases is the ability to quickly relate one set of files to another.

Databases are important in business analytics for accessing data, making queries, and other data and information management activities. Software such as Microsoft Access provides powerful analytical database capabilities. However, in this book, we won't be delving deeply into databases or database management systems but will work with individual database files or simple data sets. Because spreadsheets are convenient tools for storing and manipulating data sets and database files, we will use them for all examples and problems.

## EXAMPLE 1.2 A Sales Transaction Database File[25]

**Figure 1.2 shows a portion of sales transactions on an Excel worksheet for a particular day for an online seller of instructional fitness books and DVDs. The fields are shown in row 3 of the spreadsheet and consist of the customer ID, region, payment type, transaction code, source of the sale, amount, product purchased, and time of day. Each record (starting in row 4) has a value for each of these fields.**

[23]Based on a presentation by Bill Franks of Teradata, "Optimizing Customer Analytics: How Customer Level Web Data Can Help," INFORMS Conference on Business Analytics and Operations Research, April 10–12, 2011.

[24]Kenneth C. Laudon and Jane P. Laudon, *Essentials of Management Information Systems*, 9th ed. (Upper Saddle River, NJ: Prentice Hall, 2011): 159.

[25]Adapted and modified from Kenneth C. Laudon and Jane P. Laudon, *Essentials of Management Information Systems*.

**Figure 1.2**

A Portion of Excel File *Sales Transactions* Database

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | **Sales Transactions: July 14** | | | | | | | |
| 2 | | | | | | | | |
| 3 | **Cust ID** | **Region** | **Payment** | **Transaction Code** | **Source** | **Amount** | **Product** | **Time Of Day** |
| 4 | 10001 | East | Paypal | 93816545 | Web | $20.19 | DVD | 22:19 |
| 5 | 10002 | West | Credit | 74083490 | Web | $17.85 | DVD | 13:27 |
| 6 | 10003 | North | Credit | 64942368 | Web | $23.98 | DVD | 14:27 |
| 7 | 10004 | West | Paypal | 70560957 | Email | $23.51 | Book | 15:38 |
| 8 | 10005 | South | Credit | 35208817 | Web | $15.33 | Book | 15:21 |
| 9 | 10006 | West | Paypal | 20978903 | Email | $17.30 | DVD | 13:11 |
| 10 | 10007 | East | Credit | 80103311 | Web | $177.72 | Book | 21:59 |
| 11 | 10008 | West | Credit | 14132683 | Web | $21.76 | Book | 4:04 |
| 12 | 10009 | West | Paypal | 40128225 | Web | $15.92 | DVD | 19:35 |
| 13 | 10010 | South | Paypal | 49073721 | Web | $23.39 | DVD | 13:26 |

## Big Data

Today, nearly all data are captured digitally. As a result, data have been growing at an overwhelming rate, being measured by terabytes ($10^{12}$ bytes), petabytes ($10^{15}$ bytes), exabytes ($10^{18}$ bytes), and even by higher-dimensional terms. Just think of the amount of data stored on Facebook, Twitter, or Amazon servers, or the amount of data acquired daily from scanning items at a national grocery chain such as Kroger and its affiliates. Walmart, for instance, has over one million transactions each hour, yielding more than 2.5 petabytes of data. Analytics professionals have coined the term **big data** to refer to massive amounts of business data from a wide variety of sources, much of which is available in real time, and much of which is uncertain or unpredictable. IBM calls these characteristics *volume*, *variety*, *velocity*, and *veracity*. Most often, big data revolves around customer behavior and customer experiences. Big data provides an opportunity for organizations to gain a competitive advantage—if the data can be understood and analyzed effectively to make better business decisions.

The volume of data continue to increase; what is considered "big" today will be even bigger tomorrow. In one study of information technology (IT) professionals in 2010, nearly half of survey respondents ranked data growth among their top three challenges. Big data come from many sources, and can be numerical, textual, and even audio and video data. Big data are captured using sensors (for example, supermarket scanners), click streams from the Web, customer transactions, e-mails, tweets and social media, and other ways. Big data sets are unstructured and messy, requiring sophisticated analytics to integrate and process the data, and understand the information contained in them. Not only are big data being captured in real time, but they must be incorporated into business decisions at a faster rate. Processes such as fraud detection must be analyzed quickly to have value. IBM has added a fourth dimension: veracity—the level of reliability associated with data. Having high-quality data and understanding the uncertainty in data are essential for good decision making. Data veracity is an important role for statistical methods.

Big data can help organizations better understand and predict customer behavior and improve customer service. A study by the McKinsey Global Institute noted that "The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. Using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world."[26] However, understanding big

[26]James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey & Company May 2011.

data requires advanced analytics tools such as data mining and text analytics, and new technologies such as cloud computing, faster multi-core processors, large memory spaces, and solid-state drives.

## Metrics and Data Classification

A **metric** is a unit of measurement that provides a way to objectively quantify performance. For example, senior managers might assess overall business performance using such metrics as net profit, return on investment, market share, and customer satisfaction. A plant manager might monitor such metrics as the proportion of defective parts produced or the number of inventory turns each month. For a Web-based retailer, some useful metrics are the percentage of orders filled accurately and the time taken to fill a customer's order. **Measurement** is the act of obtaining data associated with a metric. **Measures** are numerical values associated with a metric.

Metrics can be either discrete or continuous. A **discrete metric** is one that is derived from counting something. For example, a delivery is either on time or not; an order is complete or incomplete; or an invoice can have one, two, three, or any number of errors. Some discrete metrics associated with these examples would be the proportion of on-time deliveries; the number of incomplete orders each day, and the number of errors per invoice. **Continuous metrics** are based on a continuous scale of measurement. Any metrics involving dollars, length, time, volume, or weight, for example, are continuous.

Another classification of data is by the type of measurement scale. Data may be classified into four groups:

1. **Categorical (nominal) data**, which are sorted into categories according to specified characteristics. For example, a firm's customers might be classified by their geographical region (North America, South America, Europe, and Pacific); employees might be classified as managers, supervisors, and associates. The categories bear no quantitative relationship to one another, but we usually assign an arbitrary number to each category to ease the process of managing the data and computing statistics. Categorical data are usually counted or expressed as proportions or percentages.
2. **Ordinal data**, which can be ordered or ranked according to some relationship to one another. College football or basketball rankings are ordinal; a higher ranking signifies a stronger team but does not specify any numerical measure of strength. Ordinal data are more meaningful than categorical data because data can be compared to one another. A common example in business is data from survey scales—for example, rating a service as poor, average, good, very good, or excellent. Such data are categorical but also have a natural order (excellent is better than very good) and, consequently, are ordinal. However, ordinal data have no fixed units of measurement, so we cannot make meaningful numerical statements about differences between categories. Thus, we cannot say that the difference between excellent and very good is the same as between good and average, for example. Similarly, a team ranked number 1 may be far superior to the number 2 team, whereas there may be little difference between teams ranked 9th and 10th.
3. **Interval data**, which are ordinal but have constant differences between observations and have arbitrary zero points. Common examples are time and temperature. Time is relative to global location, and calendars have arbitrary starting dates (compare, for example, the standard Gregorian calendar with the Chinese

calendar). Both the Fahrenheit and Celsius scales represent a specified measure of distance—degrees—but have arbitrary zero points. Thus we cannot take meaningful ratios; for example, we cannot say that 50 degrees is twice as hot as 25 degrees. However, we can compare differences. Another example is SAT or GMAT scores. The scores can be used to rank students, but only differences between scores provide information on how much better one student performed over another; ratios make little sense. In contrast to ordinal data, interval data allow meaningful comparison of ranges, averages, and other statistics.

In business, data from survey scales, while technically ordinal, are often treated as interval data when numerical scales are associated with the categories (for instance, 1 = poor, 2 = average, 3 = good, 4 = very good, 5 = excellent). Strictly speaking, this is not correct because the "distance" between categories may not be perceived as the same (respondents might perceive a larger gap between poor and average than between good and very good, for example). Nevertheless, many users of survey data treat them as interval when analyzing the data, particularly when only a numerical scale is used without descriptive labels.

4. **Ratio data**, which are continuous and have a natural zero. Most business and economic data, such as dollars and time, fall into this category. For example, the measure dollars has an absolute zero. Ratios of dollar figures are meaningful. For example, knowing that the Seattle region sold $12 million in March whereas the Tampa region sold $6 million means that Seattle sold twice as much as Tampa.

This classification is hierarchical in that each level includes all the information content of the one preceding it. For example, ordinal data are also categorical, and ratio information can be converted to any of the other types of data. Interval information can be converted to ordinal or categorical data but cannot be converted to ratio data without the knowledge of the absolute zero point. Thus, a ratio scale is the strongest form of measurement.

## EXAMPLE 1.3 Classifying Data Elements in a Purchasing Database[27]

Figure 1.3 shows a portion of a data set containing all items that an aircraft component manufacturing company has purchased over the past 3 months. The data provide the supplier; order number; item number, description, and cost; quantity ordered; cost per order, the suppliers' accounts payable (A/P) terms; and the order and arrival dates. We may classify each of these types of data as follows:

- Supplier—categorical
- Order Number—ordinal
- Item Number—categorical
- Item Description—categorical
- Item Cost—ratio
- Quantity—ratio
- Cost per Order—ratio
- A/P Terms—ratio
- Order Date—interval
- Arrival Date—interval

We might use these data to evaluate the average speed of delivery and rank the suppliers (thus creating ordinal data) by this metric. (We see how to do this in the next chapter).

[27]Based on Laudon and Laudon, *Essentials of Management Information Systems*.

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **Purchase Orders** | | | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | **Supplier** | **Order No.** | **Item No.** | **Item Description** | **Item Cost** | **Quantity** | **Cost per order** | **A/P Terms (Months)** | **Order Date** | **Arrival Date** |
| 4 | Hulkey Fasteners | Aug11001 | 1122 | Airframe fasteners | $ 4.25 | 19,500 | $ 82,875.00 | 30 | 08/05/11 | 08/13/11 |
| 5 | Alum Sheeting | Aug11002 | 1243 | Airframe fasteners | $ 4.25 | 10,000 | $ 42,500.00 | 30 | 08/08/11 | 08/14/11 |
| 6 | Fast-Tie Aerospace | Aug11003 | 5462 | Shielded Cable/ft. | $ 1.05 | 23,000 | $ 24,150.00 | 30 | 08/10/11 | 08/15/11 |
| 7 | Fast-Tie Aerospace | Aug11004 | 5462 | Shielded Cable/ft. | $ 1.05 | 21,500 | $ 22,575.00 | 30 | 08/15/11 | 08/22/11 |
| 8 | Steelpin Inc. | Aug11005 | 5319 | Shielded Cable/ft. | $ 1.10 | 17,500 | $ 19,250.00 | 30 | 08/20/11 | 08/31/11 |
| 9 | Fast-Tie Aerospace | Aug11006 | 5462 | Shielded Cable/ft. | $ 1.05 | 22,500 | $ 23,625.00 | 30 | 08/20/11 | 08/26/11 |
| 10 | Steelpin Inc. | Aug11007 | 4312 | Bolt-nut package | $ 3.75 | 4,250 | $ 15,937.50 | 30 | 08/25/11 | 09/01/11 |
| 11 | Durrable Products | Aug11008 | 7258 | Pressure Gauge | $ 90.00 | 100 | $ 9,000.00 | 45 | 08/25/11 | 08/28/11 |
| 12 | Fast-Tie Aerospace | Aug11009 | 6321 | O-Ring | $ 2.45 | 1,300 | $ 3,185.00 | 30 | 08/25/11 | 09/04/11 |
| 13 | Fast-Tie Aerospace | Aug11010 | 5462 | Shielded Cable/ft. | $ 1.05 | 22,500 | $ 23,625.00 | 30 | 08/25/11 | 09/02/11 |
| 14 | Steelpin Inc. | Aug11011 | 5319 | Shielded Cable/ft. | $ 1.10 | 18,100 | $ 19,910.00 | 30 | 08/25/11 | 09/05/11 |
| 15 | Hulkey Fasteners | Aug11012 | 3166 | Electrical Connector | $ 1.25 | 5,600 | $ 7,000.00 | 30 | 08/25/11 | 08/29/11 |

**Figure 1.3**

Portion of Excel File *Purchase Orders* Data

## Data Reliability and Validity

Poor data can result in poor decisions. In one situation, a distribution system design model relied on data obtained from the corporate finance department. Transportation costs were determined using a formula based on the latitude and longitude of the locations of plants and customers. But when the solution was represented on a geographic information system (GIS) mapping program, one of the customers was in the Atlantic Ocean.

Thus, data used in business decisions need to be reliable and valid. **Reliability** means that data are accurate and consistent. **Validity** means that data correctly measure what they are supposed to measure. For example, a tire pressure gauge that consistently reads several pounds of pressure below the true value is not reliable, although it is valid because it does measure tire pressure. The number of calls to a customer service desk might be counted correctly each day (and thus is a reliable measure), but not valid if it is used to assess customer dissatisfaction, as many calls may be simple queries. Finally, a survey question that asks a customer to rate the quality of the food in a restaurant may be neither reliable (because different customers may have conflicting perceptions) nor valid (if the intent is to measure customer satisfaction, as satisfaction generally includes other elements of service besides food).

# Models in Business Analytics

To make a decision, we must be able to specify the decision alternatives that represent the choices that can be made and criteria for evaluating the alternatives. Specifying decision alternatives might be very simple; for example, you might need to choose one of three corporate health plan options. Other situations can be more complex; for example, in locating a new distribution center, it might not be possible to list just a small number of alternatives. The set of potential locations might be anywhere in the United States or even within a large geographical region such as Asia. Decision criteria might be to maximize discounted net profits, customer satisfaction, or social benefits or to minimize costs, environmental impact, or some measure of loss.

Many decision problems can be formalized using a model. A **model** is an abstraction or representation of a real system, idea, or object. Models capture the most important features of a problem and present them in a form that is easy to interpret. A model can be as simple as a written or verbal description of some phenomenon, a visual representation such as a graph or a flowchart, or a mathematical or spreadsheet representation (see Example 1.4).

Models can be descriptive, predictive, or prescriptive, and therefore are used in a wide variety of business analytics applications. In Example 1.4, note that the first two

## EXAMPLE 1.4 Three Forms of a Model

The sales of a new product, such as a first-generation iPad, Android phone, or 3-D television, often follow a common pattern. We might represent this in one of three following ways:

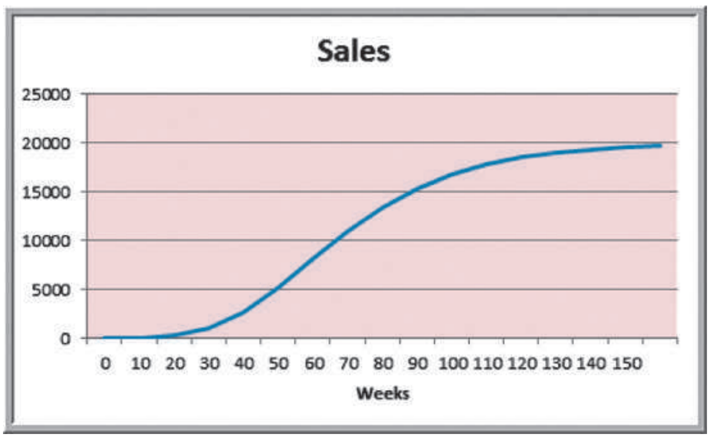
1. A simple verbal description of sales might be: The rate of sales starts small as early adopters begin to evaluate a new product and then begins to grow at an increasing rate over time as positive customer feedback spreads. Eventually, the market begins to become saturated and the rate of sales begins to decrease.
2. A sketch of sales as an S-shaped curve over time, as shown in Figure 1.4, is a visual model that conveys this phenomenon.
3. Finally, analysts might identify a mathematical model that characterizes this curve. Several different mathematical functions do this; one is called a *Gompertz curve* and has the formula: $S = ae^{be^{ct}}$, where $S$ = sales, $t$ = time, $e$ is the base of natural logarithms, and $a$, $b$, and $c$ are constants. Of course, you would not be expected to know this; that's what analytics professionals do. Such a mathematical model provides the ability to predict sales quantitatively, and to analyze potential decisions by asking "what if?" questions.

forms of the model are purely descriptive; they simply explain the phenomenon. While the mathematical model also describes the phenomenon, it can be used to predict sales at a future time. Models are usually developed from theory or observation and establish relationships between actions that decision makers might take and results that they might expect, thereby allowing the decision makers to predict what might happen based on the model.

Models complement decision makers' intuition and often provide insights that intuition cannot. For example, one early application of analytics in marketing involved a study of sales operations. Sales representatives had to divide their time between large and small customers and between acquiring new customers and keeping old ones. The problem was to determine how the representatives should best allocate their time. Intuition suggested that they should concentrate on large customers and that it was much harder to acquire a new customer than to keep an old one. However, intuition could not tell whether they should concentrate on the 100 largest or the 1,000 largest customers, or how much effort to spend on acquiring new customers. Models of sales force effectiveness and customer response patterns provided the insight to make these decisions. However, it is important to understand that all models are only representations of the real world and, as such, cannot capture every nuance that decision makers face in reality. Decision makers must often

**Figure 1.4**

New Product Sales Over Time



modify the policies that models suggest to account for intangible factors that they might not have been able to incorporate into the model.

A simple descriptive model is a visual representation called an **influence diagram** because it describes how various elements of the model influence, or relate to, others. An influence diagram is a useful approach for conceptualizing the structure of a model and can assist in building a mathematical or spreadsheet model. The elements of the model are represented by circular symbols called *nodes*. Arrows called *branches* connect the nodes and show which elements influence others. Influence diagrams are quite useful in the early stages of model building when we need to understand and characterize key relationships. Example 1.5 shows how to construct simple influence diagrams, and Example 1.6 shows how to build a mathematical model, drawing upon the influence diagram.

## EXAMPLE 1.5 An Influence Diagram for Total Cost

**From basic business principles, we know that the total cost of producing a fixed volume of a product is comprised of fixed costs and variable costs. Thus, a simple influence diagram that shows these relationships is given in Figure 1.5.**

**We can develop a more detailed model by noting that the variable cost depends on the unit variable cost as well as the quantity produced. The expanded model is shown in Figure 1.6. In this figure, all the nodes that have no branches pointing into them are inputs to the model. We can see that the unit variable cost and fixed costs are data inputs in the model. The quantity produced, however, is a decision variable because it can be controlled by the manager of the operation. The total cost is the output (note that it has no branches pointing out of it) that we would be interested in calculating. The variable cost node links some of the inputs with the output and can be considered as a "building block" of the model for total cost.**

**Figure 1.5**

**An Influence Diagram Relating Total Cost to Its Key Components**

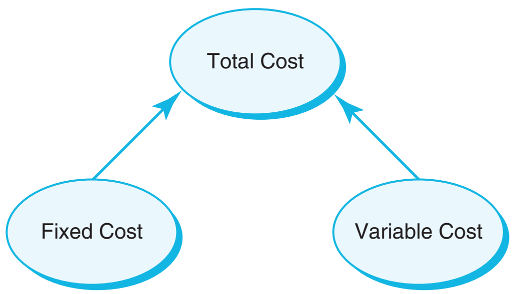
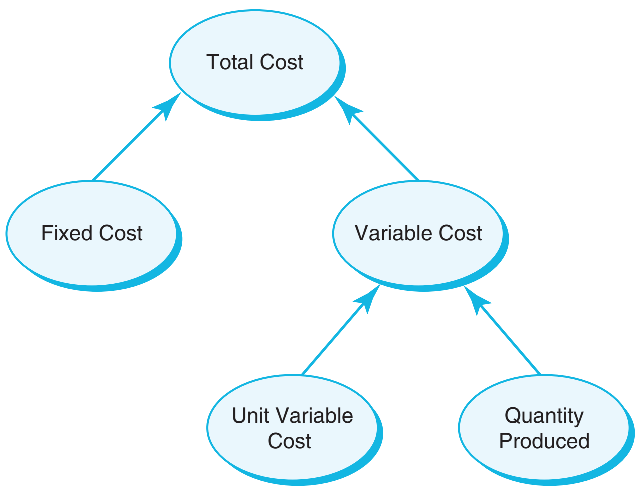



**Figure 1.6**

**An Expanded Influence Diagram for Total Cost**



## EXAMPLE 1.6 Building a Mathematical Model from an Influence Diagram

We can develop a mathematical model from the influence diagram in Figure 1.6. First, we need to specify the precise nature of the relationships among the various quantities. For example, we can easily state that

$$\text{Total Cost} = \text{Fixed Cost} + \text{Variable Cost} \tag{1.1}$$

Logic also suggests that the variable cost is the unit variable cost times the quantity produced. Thus,

$$\text{Variable Cost} = \text{Unit Variable Cost} \times \text{Quantity Produced} \tag{1.2}$$

By substituting this into equation (1.1), we have

$$\begin{aligned} \text{Total Cost} &= \text{Fixed Cost} + \text{Variable Cost} \\ &= \text{Fixed Cost} + \text{Unit Variable Cost} \times \text{Quantity Produced} \end{aligned} \tag{1.3}$$

Using these relationships, we may develop a mathematical representation by defining symbols for each of these quantities:

$TC$ = total cost
$V$ = unit variable cost
$F$ = fixed cost
$Q$ = quantity produced

This results in the model

$$TC = F + VQ \tag{1.4}$$

## Decision Models

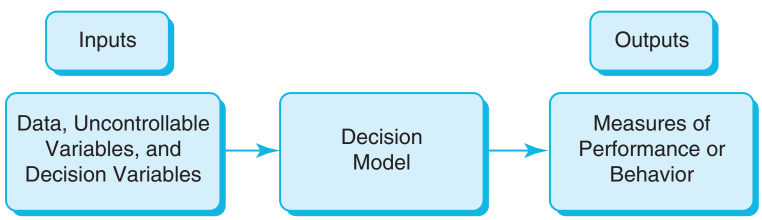
A **decision model** is a logical or mathematical representation of a problem or business situation that can be used to understand, analyze, or facilitate making a decision. Most decision models have three types of input:

1. *Data*, which are assumed to be constant for purposes of the model. Some examples would be costs, machine capacities, and intercity distances.
2. *Uncontrollable variables*, which are quantities that can change but cannot be directly controlled by the decision maker. Some examples would be customer demand, inflation rates, and investment returns. Often, these variables are uncertain.
3. *Decision variables*, which are controllable and can be selected at the discretion of the decision maker. Some examples would be production quantities (see Example 1.5), staffing levels, and investment allocations.

Decision models characterize the relationships among the data, uncontrollable variables, and decision variables, and the outputs of interest to the decision maker (see Figure 1.7). Decision models can be represented in various ways, most typically with mathematical functions and spreadsheets. Spreadsheets are ideal vehicles for implementing decision models because of their versatility in managing data, evaluating different scenarios, and presenting results in a meaningful fashion.

**Figure 1.7**

Nature of Decision Models



How might we use the model in Example 1.6 to help make a decision? Suppose that a manufacturer has the option of producing a part in-house or outsourcing it from a supplier (the decision variables). Should the firm produce the part or outsource it? The decision depends on the anticipated volume of demand (an uncontrollable variable); for high volumes, the cost to manufacture in-house will be lower than outsourcing, because the fixed costs can be spread over a large number of units. For small volumes, it would be more economical to outsource. Knowing the total cost of both alternatives (based on data for fixed and variable manufacturing costs and purchasing costs) and the break-even point would facilitate the decision. A numerical example is provided in Example 1.7.

## EXAMPLE 1.7 A Break-Even Decision Model

Suppose that a manufacturer can produce a part for \$125/unit with a fixed cost of \$50,000. The alternative is to outsource production to a supplier at a unit cost of \$175. The total manufacturing cost is expressed by using equation (1.5):

$$TC \text{ (manufacturing)} = \$50{,}000 + \$125 \times Q$$

and the total outsourcing cost can be written as

$$TC \text{ (outsourcing)} = \$175 \times Q$$

Mathematical models are easy to manipulate; for example, it is easy to find the break-even volume by setting $TC$ (manufacturing) $=$ $TC$ (outsourcing) and solving for $Q$:

$$\$50{,}000 + \$125 \times Q = \$175 \times Q$$
$$\$50{,}000 = 50 \times Q$$
$$Q = 1{,}000$$

Thus, if the anticipated production volume is greater than 1,000, it is more economical to manufacture the part; if it is less than 1,000, then it should be outsourced. This is shown graphically in Figure 1.8.

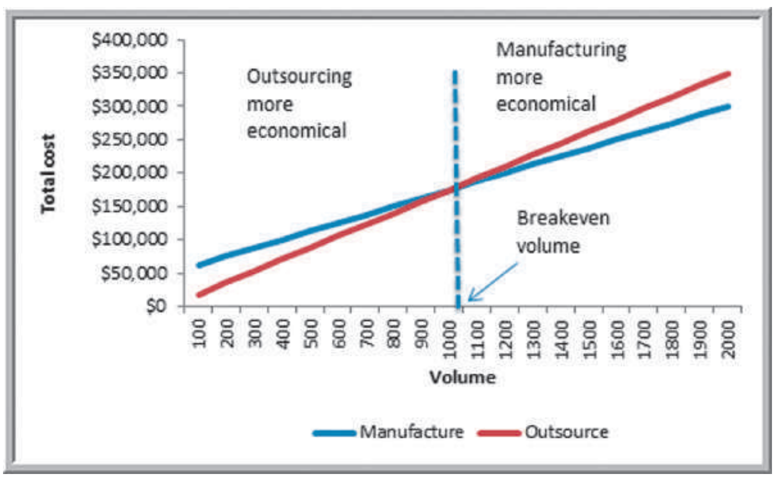
We may also develop a general formula for the break-even point by letting $C$ be the unit cost of outsourcing the part and setting $TC$ (manufacturing) $=$ $TC$ (outsourcing) using the formulas:

$$F + VQ = CQ$$
$$Q = \frac{F}{C - V} \qquad (1.5)$$

Many models are developed by analyzing historical data. Example 1.8 shows how historical data might be used to develop a decision model that can be used to predict the impact of pricing and promotional strategies in the grocery industry.

**Figure 1.8**

Graphical Illustration of Break-Even Analysis



## EXAMPLE 1.8 A Sales-Promotion Decision Model

In the grocery industry, managers typically need to know how best to use pricing, coupons, and advertising strategies to influence sales. Grocers often study the relationship of sales volume to these strategies by conducting controlled experiments to identify the relationship between them and sales volumes.[28] That is, they implement different combinations of pricing, coupons, and advertising, observe the sales that result, and use analytics to develop a predictive model of sales as a function of these decision strategies.

For example, suppose that a grocer who operates three stores in a small city varied the price, coupons (yes = 1, no = 0), and advertising expenditures in a local newspaper over a 16-week period and observed the following sales:

| Week | Price ($) | Coupon (0,1) | Advertising ($) | Store 1 Sales (Units) | Store 2 Sales (Units) | Store 3 Sales (Units) |
|---|---|---|---|---|---|---|
| 1 | 6.99 | 0 | 0 | 501 | 510 | 481 |
| 2 | 6.99 | 0 | 150 | 772 | 748 | 775 |
| 3 | 6.99 | 1 | 0 | 554 | 528 | 506 |
| 4 | 6.99 | 1 | 150 | 838 | 785 | 834 |
| 5 | 6.49 | 0 | 0 | 521 | 519 | 500 |
| 6 | 6.49 | 0 | 150 | 723 | 790 | 723 |
| 7 | 6.49 | 1 | 0 | 510 | 556 | 520 |
| 8 | 6.49 | 1 | 150 | 818 | 773 | 800 |
| 9 | 7.59 | 0 | 0 | 479 | 491 | 486 |
| 10 | 7.59 | 0 | 150 | 825 | 822 | 757 |
| 11 | 7.59 | 1 | 0 | 533 | 513 | 540 |
| 12 | 7.59 | 1 | 150 | 839 | 791 | 832 |
| 13 | 5.49 | 0 | 0 | 484 | 480 | 508 |
| 14 | 5.49 | 0 | 150 | 686 | 683 | 708 |
| 15 | 5.49 | 1 | 0 | 543 | 531 | 530 |
| 16 | 5.49 | 1 | 150 | 767 | 743 | 779 |

To better understand the relationships among price, coupons, and advertising, the grocer might have developed the following model using business analytics tools:

$$\text{sales} = 500 - 0.05 \times \text{price} + 30 \times \text{coupons} + 0.08 \times \text{advertising} + 0.25 \times \text{price} \times \text{advertising}$$

In this model, the decision variables are price, coupons, and advertising. The values 500, −0.05, 30, 0.08, and 0.25 are effects of the input data to the model that are estimated from the data obtained from the experiment. They reflect the impact on sales of changing the decision variables. For example, an increase in price of $1 results in a 0.05-unit decrease in weekly sales; using coupons results in a 30-unit increase in weekly sales. In this example, there are no uncontrollable input variables. The output of the model is the sales units of the product. For example, if the price is $6.99, no coupons are offered and no advertising is done (the experiment corresponding to week 1), the model estimates sales as

$$\text{sales} = 500 - 0.05 \times \$6.99 + 30 \times 0 + 0.08 \times 0 + 0.25 \times \$6.99 \times 0 = 500 \text{ units}$$

We see that the actual sales in week 1 varied between 481 and 510 in the three stores. Thus, this model predicts a good estimate for sales; however, it does not tell us anything about the potential variability or prediction error. Nevertheless, the manager can use this model to evaluate different pricing, promotion, and advertising strategies, and help choose the best strategy to maximize sales or profitability.

[28]Roger J. Calantone, Cornelia Droge, David S. Litvack, and C. Anthony di Benedetto. "Flanking in a Price War," *Interfaces*, 19, 2 (1989): 1–12.

## Model Assumptions

All models are based on assumptions that reflect the modeler's view of the "real world." Some assumptions are made to simplify the model and make it more tractable; that is, able to be easily analyzed or solved. Other assumptions might be made to better characterize historical data or past observations. The task of the modeler is to select or build an appropriate model that best represents the behavior of the real situation. For example, economic theory tells us that demand for a product is negatively related to its price. Thus, as prices increase, demand falls, and vice versa (a phenomenon that you may recognize as **price elasticity**—the ratio of the percentage change in demand to the percentage change in price). Different mathematical models can describe this phenomenon. In the following examples, we illustrate two of them. (Both of these examples can be found in the Excel file *Demand Prediction Models*. We introduce the use of spreadsheets in analytics in the next chapter.)

### EXAMPLE 1.9 A Linear Demand Prediction Model

A simple model to predict demand as a function of price is the linear model

$$D = a - bP \tag{1.6}$$

where $D$ is the demand rate, $P$ is the unit price, $a$ is a constant that estimates the demand when the price is zero, and $b$ is the slope of the demand function. This model is most applicable when we want to predict the effect of small changes around the current price. For example, suppose we know that when the price is \$100, demand is 19,000 units and that demand falls by 10 for each dollar of price increase. Using simple algebra, we can determine that $a = 20{,}000$ and $b = 10$. Thus, if the price is \$80, the predicted demand is

$$D = 20{,}000 - 10(80) = 19{,}200 \text{ units}$$

If the price increases to \$90, the model predicts demand as

$$D = 20{,}000 - 10(90) = 19{,}100 \text{ units}$$

If the price is \$100, demand would be

$$D = 20{,}000 - 10(100) = 19{,}000 \text{ units}$$

and so on. A chart of demand as a function of price is shown in Figure 1.9 as price varies between \$80 and \$120. We see that there is a constant decrease in demand for each \$10 increase in price, a characteristic of a linear model.

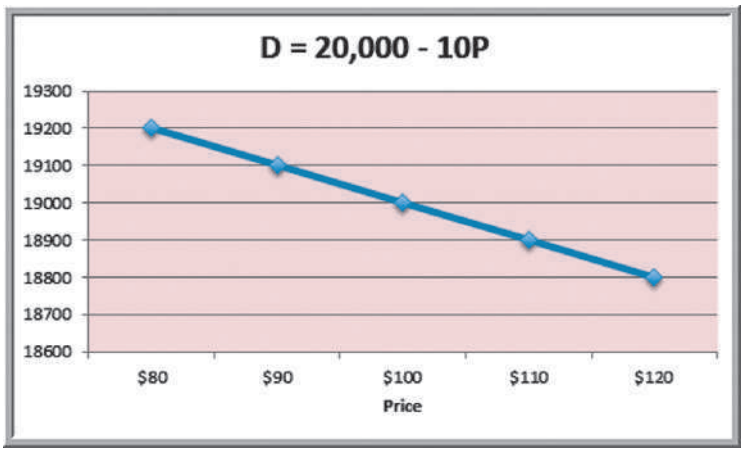



**Figure 1.9** Graph of Linear Demand Model $D = a - bP$

## EXAMPLE 1.10 A Nonlinear Demand Prediction Model

An alternative model assumes that price elasticity is constant. In this case, the appropriate model is

$$D = cP^{-d} \tag{1.7}$$

where, $c$ is the demand when the price is 0 and $d > 0$ is the price elasticity. To be consistent with Example 1.9, we assume that when the price is zero, demand is 20,000. Therefore, $c = 20{,}000$. We will also, as in Example 1.9, assume that when the price is \$100, $D = 19{,}000$. Using these values in equation (1.7), we can determine the value for $d$ (we can do this mathematically using logarithms, but we'll see how to do this very easily using Excel in Chapter 11); this is $d = -0.0111382$. Thus, if the price is \$80, then the predicted demand is

$$D = 20{,}000(80)^{-0.0111382} = 19{,}047.$$

If the price is 90, the demand would be

$$D = 20{,}000(90)^{-0.0111382} = 19022.$$

If the price is 100, demand is

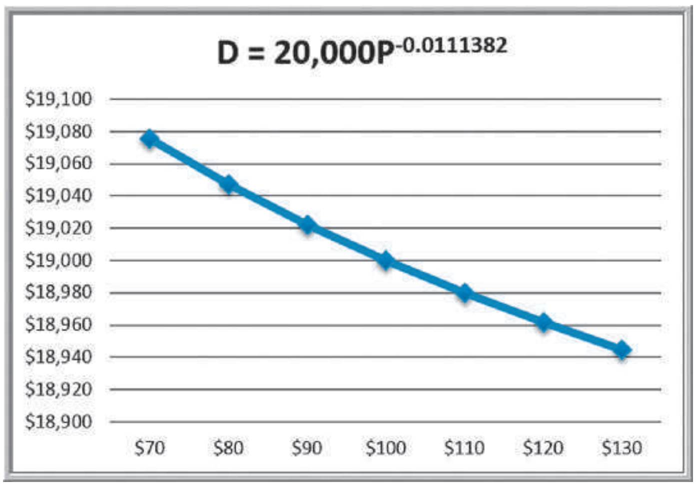
$$D = 20{,}000(100)^{-0.0111382} = 19{,}000.$$

A graph of demand as a function of price is shown in Figure 1.10. The predicted demand falls in a slight nonlinear fashion as price increases. For example, demand decreases by 25 units when the price increases from \$80 to \$90, but only by 22 units when the price increases from \$90 to \$100. If the price increases to \$100, you would see a smaller decrease in demand. Therefore, we see a nonlinear relationship in contrast to Example 1.9.

Both models in Examples 1.9 and 1.10 make different predictions of demand for different prices (other than \$90). Which model is best? The answer may be neither. First of all, the development of realistic models requires many price point changes within a carefully designed experiment. Secondly, it should also include data on competition and customer disposable income, both of which are hard to determine. Nevertheless, it is possible to develop price elasticity models with limited price ranges and narrow customer segments. A good starting point would be to create a historical database with detailed information on all past pricing actions. Unfortunately, practitioners have observed that such models are not widely used in retail marketing, suggesting a lot of opportunity to apply business analytics.[29]

**Figure 1.10**

Graph of Nonlinear Demand Model $D = cP^{-d}$




[29] Ming Zhang, Clay Duan, and Arun Muthupalaniappan, "Analytics Applications in Consumer Credit and Retail Marketing," analytics-magazine.org, November/December 2011, pp. 27–33.

## Uncertainty and Risk

As we all know, the future is always uncertain. Thus, many predictive models incorporate uncertainty and help decision makers analyze the risks associated with their decisions. **Uncertainty** is imperfect knowledge of what will happen; **risk** is associated with the consequences and likelihood of what might happen. For example, the change in the stock price of Apple on the next day of trading is uncertain. However, if you own Apple stock, then you face the risk of losing money if the stock price falls. If you don't own any stock, the price is still uncertain although you would not have any risk. Risk is evaluated by the magnitude of the consequences and the likelihood that they would occur. For example, a 10% drop in the stock price would incur a higher risk if you own $1 million than if you only owned $1,000. Similarly, if the chances of a 10% drop were 1 in 5, the risk would be higher than if the chances were only 1 in 100.

The importance of risk in business has long been recognized. The renowned management writer, Peter Drucker, observed in 1974:

> To try to eliminate risk in business enterprise is futile. Risk is inherent in the commitment of present resources to future expectations. Indeed, economic progress can be defined as the ability to take greater risks. The attempt to eliminate risks, even the attempt to minimize them, can only make them irrational and unbearable. It can only result in the greatest risk of all: rigidity.[30]

Consideration of risk is a vital element of decision making. For instance, you would probably not choose an investment simply on the basis of the return you might expect because, typically, higher returns are associated with higher risk. Therefore, you have to make a trade-off between the benefits of greater rewards and the risks of potential losses. Analytic models can help assess this. We will address this in later chapters.

## Prescriptive Decision Models

A prescriptive decision model helps decision makers to identify the best solution to a decision problem. **Optimization** is the process of finding a set of values for decision variables that minimize or maximize some quantity of interest—profit, revenue, cost, time, and so on—called the **objective function**. Any set of decision variables that optimizes the objective function is called an **optimal solution**. In a highly competitive world where one percentage point can mean a difference of hundreds of thousands of dollars or more, knowing the best solution can mean the difference between success and failure.

### EXAMPLE 1.11 A Prescriptive Model for Pricing

**To illustrate an example of a prescriptive model, suppose that a firm wishes to determine the best pricing for one of its products to maximize revenue over the next year. A market research study has collected data that estimate the expected annual sales for different levels of pricing. Analysts determined that sales can be expressed by the following model:**

$$\text{sales} = -2.9485 \times \text{price} + 3{,}240.9$$

**Because revenue equals price × sales, a model for total revenue is**

$$\begin{aligned}\text{total revenue} &= \text{price} \times \text{sales}\\ &= \text{price} \times (-2.9485 \times \text{price} + 3240.9)\\ &= 22.9485 \times \text{price}^2 + 3240.9 \times \text{price}\end{aligned}$$

**The firm would like to identify the price that maximizes the total revenue. One way to do this would be to try different prices and search for the one that yields the highest total revenue. This would be quite tedious to do by hand or even with a calculator. We will see how to do this easily on a spreadsheet in Chapter 11.**

[30]P. F. Drucker, *The Manager and the Management Sciences in Management: Tasks, Responsibilities, Practices* (London: Harper and Row, 1974).

Although the pricing model did not, most optimization models have **constraints**—limitations, requirements, or other restrictions that are imposed on any solution, such as "do not exceed the allowable budget" or "ensure that all demand is met." For instance, a consumer products company manager would probably want to ensure that a specified level of customer service is achieved with the redesign of the distribution system. The presence of constraints makes modeling and solving optimization problems more challenging; we address constrained optimization problems later in this book, starting in Chapter 13.

For some prescriptive models, analytical solutions—closed-form mathematical expressions or simple formulas—can be obtained using such techniques as calculus or other types of mathematical analyses. In most cases, however, some type of computer-based procedure is needed to find an optimal solution. An **algorithm** is a systematic procedure that finds a solution to a problem. Researchers have developed effective algorithms to solve many types of optimization problems. For example, Microsoft Excel has a built-in add-in called *Solver* that allows you to find optimal solutions to optimization problems formulated as spreadsheet models. We use *Solver* in later chapters. However, we will not be concerned with the detailed mechanics of these algorithms; our focus will be on the use of the algorithms to solve and analyze the models we develop.

If possible, we would like to ensure that an algorithm such as the one *Solver* uses finds the best solution. However, some models are so complex that it is impossible to solve them optimally in a reasonable amount of computer time because of the extremely large number of computations that may be required or because they are so complex that finding the best solution cannot be guaranteed. In these cases, analysts use **search algorithms**—solution procedures that generally find good solutions without guarantees of finding the best one. Powerful search algorithms exist to obtain good solutions to extremely difficult optimization problems. These are discussed in the supplementary online Chapter A.

Prescriptive decision models can be either *deterministic* or *stochastic*. A **deterministic model** is one in which all model input information is either known or assumed to be known with certainty. A **stochastic model** is one in which some of the model input information is uncertain. For instance, suppose that customer demand is an important element of some model. We can make the assumption that the demand is known with certainty; say, 5,000 units per month. In this case we would be dealing with a deterministic model. On the other hand, suppose we have evidence to indicate that demand is uncertain, with an average value of 5,000 units per month, but which typically varies between 3,200 and 6,800 units. If we make this assumption, we would be dealing with a stochastic model. These situations are discussed in the supplementary online Chapter B.

## Problem Solving with Analytics

The fundamental purpose of analytics is to help managers solve problems and make decisions. The techniques of analytics represent only a portion of the overall problem-solving and decision-making process. **Problem solving** is the activity associated with defining, analyzing, and solving a problem and selecting an appropriate solution that solves a problem. Problem solving consists of several phases:

1. recognizing a problem
2. defining the problem
3. structuring the problem
4. analyzing the problem
5. interpreting results and making a decision
6. implementing the solution

## Recognizing a Problem

Managers at different organizational levels face different types of problems. In a manufacturing firm, for instance, top managers face decisions of allocating financial resources, building or expanding facilities, determining product mix, and strategically sourcing production. Middle managers in operations develop distribution plans, production and inventory schedules, and staffing plans. Finance managers analyze risks, determine investment strategies, and make pricing decisions. Marketing managers develop advertising plans and make sales force allocation decisions. In manufacturing operations, problems involve the size of daily production runs, individual machine schedules, and worker assignments. Whatever the problem, the first step is to realize that it exists.

How are problems recognized? Problems exist when there is a gap between what is happening and what we think should be happening. For example, a consumer products manager might feel that distribution costs are too high. This recognition might result from comparing performance with a competitor, observing an increasing trend compared to previous years.

## Defining the Problem

The second step in the problem-solving process is to clearly define the problem. Finding the real problem and distinguishing it from symptoms that are observed is a critical step. For example, high distribution costs might stem from inefficiencies in routing trucks, poor location of distribution centers, or external factors such as increasing fuel costs. The problem might be defined as improving the routing process, redesigning the entire distribution system, or optimally hedging fuel purchases.

Defining problems is not a trivial task. The complexity of a problem increases when the following occur:

- The number of potential courses of action is large.
- The problem belongs to a group rather than to an individual.
- The problem solver has several competing objectives.
- External groups or individuals are affected by the problem.
- The problem solver and the true owner of the problem—the person who experiences the problem and is responsible for getting it solved—are not the same.
- Time limitations are important.

These factors make it difficult to develop meaningful objectives and characterize the range of potential decisions. In defining problems, it is important to involve all people who make the decisions or who may be affected by them.

## Structuring the Problem

This usually involves stating goals and objectives, characterizing the possible decisions, and identifying any constraints or restrictions. For example, if the problem is to redesign a distribution system, decisions might involve new locations for manufacturing plants and warehouses (where?), new assignments of products to plants (which ones?), and the amount of each product to ship from different warehouses to customers (how much?). The goal of cost reduction might be measured by the total delivered cost of the product. The manager would probably want to ensure that a specified level of customer service—for instance, being able to deliver orders within 48 hours—is achieved with the redesign. This is an example of a constraint. Structuring a problem often involves developing a formal model.

## Analyzing the Problem

Here is where analytics plays a major role. Analysis involves some sort of experimentation or solution process, such as evaluating different scenarios, analyzing risks associated with various decision alternatives, finding a solution that meets certain goals, or determining an optimal solution. Analytics professionals have spent decades developing and refining a variety of approaches to address different types of problems. Much of this book is devoted to helping you understand these techniques and gain a basic facility in using them.

## Interpreting Results and Making a Decision

Interpreting the results from the analysis phase is crucial in making good decisions. Models cannot capture every detail of the real problem, and managers must understand the limitations of models and their underlying assumptions and often incorporate judgment into making a decision. For example, in locating a facility, we might use an analytical procedure to find a "central" location; however, many other considerations must be included in the decision, such as highway access, labor supply, and facility cost. Thus, the location specified by an analytical solution might not be the exact location the company actually chooses.

## Implementing the Solution

This simply means making it work in the organization, or translating the results of a model back to the real world. This generally requires providing adequate resources, motivating employees, eliminating resistance to change, modifying organizational policies, and developing trust. Problems and their solutions affect people: customers, suppliers, and employees. All must be an important part of the problem-solving process. Sensitivity to political and organizational issues is an important skill that managers and analytical professionals alike must possess when solving problems.

In each of these steps, good communication is vital. Analytics professionals need to be able to communicate with managers and clients to understand the business context of the problem and be able to explain results clearly and effectively. Such skills as constructing good visual charts and spreadsheets that are easy to understand are vital to users of analytics. We emphasize these skills throughout this book.

### Analytics in Practice: Developing Effective Analytical Tools at Hewlett-Packard[31]

Hewlett-Packard (HP) uses analytics extensively. Many applications are used by managers with little knowledge of analytics. These require that analytical tools be easily understood. Based on years of experience, HP analysts compiled some key lessons. Before creating an analytical decision tool, HP asks three questions:

1. *Will analytics solve the problem?* Will the tool enable a better solution? Should other non analytical solutions be used? Are there organizational or other issues that must be resolved? Often, what may appear to be an analytical problem may actually be rooted in problems of incentive misalignment, unclear ownership and accountability, or business strategy.
2. *Can we leverage an existing solution?* Before "reinventing the wheel," can existing solutions address the problem? What are the costs and benefits?
3. *Is a decision model really needed?* Can simple decision guidelines be used instead of a formal decision tool?

(*continued*)

[31]Based on Thomas Olavson and Chris Fry, "Spreadsheet Decision-Support Tools: Lessons Learned at Hewlett-Packard," *Interfaces*, 38, 4, July–August 2008: 300–310.

Once a decision is made to develop an analytical tool, they use several guidelines to increase the chances of successful implementation:

- *Use prototyping*–a quick working version of the tool designed to test its features and gather feedback;
- *Build insight, not black boxes.* A "black box" tool is one that generates an answer, but may not provide confidence to the user. Interactive tools that creates insights to support a decision provide better information.
- *Remove unneeded complexity.* Simpler is better. A good tool can be used without expert support.
- *Partner with end users in discovery and design.* Decision makers who will actually use the tool should be involved in its development.
- *Develop an analytic champion.* Someone (ideally, the actual decision maker) who is knowledgeable about the solution and close to it must champion the process.

## Key Terms

Algorithm
Big data
Business analytics (analytics)
Business intelligence (BI)
Categorical (nominal) data
Constraint
Continuous metric
Data mining
Data set
Database
Decision model
Decision support systems (DSS)
Descriptive analytics
Deterministic model
Discrete metric
Influence diagram
Information systems (IS)
Interval data
Measure
Measurement
Metric
Model
Modeling and optimization
Objective function
Operations Research/Management Science (OR/MS)
Optimal solution
Optimization
Ordinal data
Predictive analytics
Prescriptive analytics
Price elasticity
Problem solving
Ratio data
Reliability
Risk
Search algorithm
Simulation and risk analysis
Statistics
Stochastic model
Tag cloud
Uncertainty
Validity
Visualization
What-if analysis

## Fun with Analytics

Mr. John Toczek, an analytics manager at ARAMARK Corporation, maintains a Web site called the PuzzlOR (OR being "Operations Research") at www.puzzlor.com. Each month he posts a new puzzle. Many of these can be solved using techniques in this book; however, even if you cannot develop a formal model, the puzzles can be fun and competitive challenges for students. We encourage you to explore these, in addition to the formal problems, exercises, and cases in this book. A good one to start with is "SurvivOR" from June 2010. Have fun!

## Problems and Exercises

**1.** Discuss how business analytics can be used in sports, such as tennis, cricket, football, and so on. Identify as many opportunities as you can for each.

**2.** A multinational hotel chain has been implementing analytics digital marketing to its customers. However, the responses to the digital campaigns have not been favorable, and the revenue generation has not been as expected. Currently, they are trying to solve this problem by focusing on similar campaigns that use the same promotional content, and changing these campaigns to suit the specific tastes of the consumers in each nation. Discuss how business analytics can be utilized by the hotel management in this scenario. What is the data required to facilitate good decisions?

**3.** Suggest some metrics that a hotel might want to collect about their guests. How might these metrics be used with business analytics to support decisions at the hotel?

**4.** Suggest some metrics that a railway or bus ticketing agency might want to collect. Describe how a manager might utilize this data to facilitate better decisions.

**5.** Classify each of the data elements in the *Sales Transactions* database (Figure 1.1) as categorical, ordinal, interval, or ratio data and explain why.

**6.** Identify each of the variables in the Excel file *Credit Approval Decisions* as categorical, ordinal, interval, or ratio and explain why.

**7.** Classify each of the variables in the Excel file *Weddings* as categorical, ordinal, interval, or ratio and explain why.

**8.** A survey handed out to individuals at a major shopping mall in a small Florida city in July asked the following:

- gender
- age
- ethnicity
- length of residency
- overall satisfaction with city services (using a scale of 1–5, going from poor to excellent)
- quality of schools (using a scale of 1–5, going from poor to excellent)

What types of data (categorical, ordinal, interval, or ratio) would each of the survey items represent and why?

**9.** A bank developed a model for predicting the average checking and savings account balance as balance $= -17{,}732 + 367 \times$ age $+ 1{,}300 \times$ years education $+ 0.116 \times$ household wealth.

**a.** Explain how to interpret the numbers in this model.

**b.** Suppose that a customer is 32 years old, is a college graduate (so that years education $= 16$), and has a household wealth of \$150,000. What is the predicted bank balance?

**10.** Four key marketing decision variables are price ($P$), advertising ($A$), transportation ($T$), and product quality ($Q$). Consumer demand ($D$) is influenced by these variables. The simplest model for describing demand in terms of these variables is

$$D = k - pP + aA + tT + qQ$$

where $k$, $p$, $a$, $t$, and $q$ are positive constants.

**a.** How does a change in each variable affect demand?

**b.** How do the variables influence each other?

**c.** What limitations might this model have? Can you think of how this model might be made more realistic?

**11.** A firm installs 1500 air conditioners which need to be serviced every six months. The firm can hire a team from its logistics department at a fixed cost of $6,000. Each unit will be serviced by the team at $15.00. The firm can also outsource this at a cost of $17.00 inclusive of all charges.

**a.** For the given number of units, compute the total cost of servicing for both options. Which is a better decision?

**b.** Find the break-even volume and characterize the range of volumes for which it is more economical to outsource.

**12.** Return on investment (ROI) is computed in the following manner: ROI is equal to turnover multiplied by earnings as a percent of sales. Turnover is sales divided by total investment. Total investment is current assets (inventories, accounts receivable, and cash) plus fixed assets. Earnings equal sales minus the cost of sales. The cost of sales consists of variable production costs, selling expenses, freight and delivery, and administrative costs.

**a.** Construct an influence diagram that relates these variables.

**b.** Define symbols and develop a mathematical model.

**13.** Total marketing effort is a term used to describe the critical decision factors that affect demand: price, advertising, distribution, and product quality. Let the variable $x$ represent total marketing effort. A typical model that is used to predict demand as a function of total marketing effort is

$$D = ax^b$$

Suppose that $a$ is a positive number. Different model forms result from varying the constant $b$. Sketch the graphs of this model for $b = 0$, $b = 1$, $0 < b < 1$, $b < 0$, and $b > 1$. What does each model tell you about the relationship between demand and marketing effort? What assumptions are implied? Are they reasonable? How would you go about selecting the appropriate model?

**14.** Automobiles have different fuel economies (mpg), and commuters drive different distances to work or school. Suppose that a state Department of Transportation (DOT) is interested in measuring the average monthly fuel consumption of commuters in a certain city. The DOT might sample a group of commuters and collect information on the number of miles driven per day, number of driving days per month, and the fuel economy of their cars. Develop a predictive model for calculating the amount of gasoline consumed, using the following symbols for the data.

$G$ = gallons of fuel consumed per month
$m$ = miles driven per day to and from work or school
$d$ = number of driving days per month
$f$ = fuel economy in miles per gallon

Suppose that a commuter drives 30 miles round trip to work 20 days each month and achieves a fuel economy of 34 mpg. How many gallons of gasoline are used?

**15.** A manufacturer of mp3 players is preparing to set the price on a new model. Demand is thought to depend on the price and is represented by the model

$$D = 2{,}500 - 3P$$

The accounting department estimates that the total costs can be represented by

$$C = 5{,}000 + 5D$$

Develop a model for the total profit in terms of the price, $P$.

**16.** The demand for airline travel is quite sensitive to price. Typically, there is an inverse relationship between demand and price; when price decreases, demand increases and vice versa. One major airline has found that when the price ($P$) for a round trip between Chicago and Los Angeles is $600, the demand ($D$) is 500 passengers per day. When the price is reduced to $400, demand is 1,200 passengers per day.

**a.** Plot these points on a coordinate system and develop a linear model that relates demand to price.

**b.** Develop a prescriptive model that will determine what price to charge to maximize the total revenue.

**c.** By trial and error, can you find the optimal solution that maximizes total revenue?

## Case: Drout Advertising Research Project[32]

Jamie Drout is interested in perceptions of gender stereotypes within beauty product advertising, which includes soap, deodorant, shampoo, conditioner, lotion, perfume, cologne, makeup, chemical hair color, razors, skin care, feminine care, and salon services; as well as the perceived benefits of empowerment advertising. Gender stereotypes specifically use cultural perceptions of what constitutes an attractive, acceptable, and desirable man or woman, frequently exploiting specific gender roles, and are commonly employed in advertisements for beauty products. Women are represented as delicately feminine, strikingly beautiful, and physically flawless, occupying small amounts of physical space that generally exploit their sexuality; men as strong and masculine with chiseled physical bodies, occupying large amounts of physical space to maintain their masculinity and power. In contrast, empowerment advertising strategies negate gender stereotypes and visually communicate the unique differences in each individual. In empowerment advertising, men and women are to represent the diversity in beauty, body type, and levels of perceived femininity and masculinity. Her project is focused on understanding consumer perceptions of these advertising strategies.

Jamie conducted a survey using the following questionnaire:

1. What is your gender?
   Male
   Female
2. What is your age?
3. What is the highest level of education you have completed?
   Some High School Classes
   High School Diploma
   Some Undergraduate Courses
   Associate Degree
   Bachelor Degree
   Master Degree
   J.D.
   M.D.
   Doctorate Degree
4. What is your annual income?
   $0 to <$10,000
   $10,000 to <$20,000
   $20,000 to <$30,000
   $30,000 to <$40,000
   $40,000 to <$50,000
   $50,000 to <$60,000
   $60,000 to <$70,000
   $70,000 to <$80,000
   $80,000 to <$90,000
   $90,000 to <$110,000
   $110,000 to <$130,000
   $130,000 to <$150,000
   $150,000 or More
5. On average, how much do you pay for beauty and hygiene products or services per year? Include references to the following products: soap, deodorant, shampoo, conditioner, lotion, perfume, cologne, makeup, chemical hair color, razors, skin care, feminine care, and salon services.
6. On average, how many beauty and hygiene advertisements, if at all, do you think you view or hear per day? Include references to the following advertisements: television, billboard, Internet, radio, newspaper, magazine, and direct mail.
7. On average, how many of those advertisements, if at all, specifically subscribe to gender roles and stereotypes?
8. On the following scale, what role, if any, do these advertisements have in reinforcing specific gender stereotypes?
   Drastic
   Influential
   Limited
   Trivial
   None
9. To what extent do you agree that empowerment advertising, which explicitly communicates the unique differences in each individual, would help transform cultural gender stereotypes?
   Strongly agree
   Agree
   Somewhat agree
   Neutral
   Somewhat disagree
   Disagree
   Strongly disagree
10. On average, what percentage of advertisements that you view or hear per day currently utilize empowerment advertising?

[32]I express my appreciation to Jamie Drout for providing this original material from her class project as the basis for this case.

**Assignment:** Jamie received 105 responses, which are given in the Excel file *Drout Advertising Survey*. Review the questionnaire and classify the data collected from each question as categorical, ordinal, interval, or ratio. Next, explain how the data and subsequent analysis using business analytics might lead to a better understanding of stereotype versus empowerment advertising. Specifically, state some of the key insights that you would hope to answer by analyzing the data.

An important aspect of business analytics is good communication. Write up your answers to this case formally in a well-written report as if you were a consultant to Ms. Drout. This case will continue in Chapters 3, 4, 6, and 7, and you will be asked to use a variety of descriptive analytics tools to analyze the data and interpret the results. As you do this, add your insights to the report, culminating in a complete project report that fully analyzes the data and draws appropriate conclusions.

## Case: Performance Lawn Equipment

In each chapter of this book, we use a database for a fictitious company, Performance Lawn Equipment (PLE), within a case exercise for applying the tools and techniques introduced in the chapter.[33] To put the database in perspective, we first provide some background about the company, so that the applications of business analytic tools will be more meaningful.

PLE, headquartered in St. Louis, Missouri, is a privately owned designer and producer of traditional lawn mowers used by homeowners. In the past 10 years, PLE has added another key product, a medium-size diesel power lawn tractor with front and rear power takeoffs, Class I three-point hitches, four-wheel drive, power steering, and full hydraulics. This equipment is built primarily for a niche market consisting of large estates, including golf and country clubs, resorts, private estates, city parks, large commercial complexes, lawn care service providers, private homeowners with five or more acres, and government (federal, state, and local) parks, building complexes, and military bases. PLE provides most of the products to dealerships, which, in turn, sell directly to end users. PLE employs 1,660 people worldwide. About half the workforce is based in St. Louis; the remainder is split among their manufacturing plants.

In the United States, the focus of sales is on the eastern seaboard, California, the Southeast, and the south central states, which have the greatest concentration of customers. Outside the United States, PLE's sales include a European market, a growing South American market, and developing markets in the Pacific Rim and China. The market is cyclical, but the different products and regions balance some of this, with just less than 30% of total sales in the spring and summer (in the United States), about 25% in the fall, and about 20% in the winter. Annual sales are approximately $180 million.

Both end users and dealers have been established as important customers for PLE. Collection and analysis of end-user data showed that satisfaction with the products depends on high quality, easy attachment/dismount of implements, low maintenance, price value, and service. For dealers, key requirements are high quality, parts and feature availability, rapid restock, discounts, and timeliness of support.

PLE has several key suppliers: Mitsitsiu, Inc., the sole source of all diesel engines; LANTO Axles, Inc., which provides tractor axles; Schorst Fabrication, which provides subassemblies; Cuberillo, Inc, supplier of transmissions; and Specialty Machining, Inc., a supplier of precision machine parts.

To help manage the company, PLE managers have developed a "balanced scorecard" of measures. These data, which are summarized shortly, are stored in the form of a Microsoft Excel workbook (*Performance Lawn Equipment*) accompanying this book. The database contains various measures captured on a monthly or quarterly basis and used by various managers to evaluate business performance. Data for each of the key measures are stored in a separate worksheet. A summary of these worksheets is given next:

- *Dealer Satisfaction*, measured on a scale of 1–5 (1 = poor, 2 = less than average, 3 = average, 4 = above average, and 5 = excellent). Each year, dealers in each region are surveyed about their overall satisfaction with PLE. The worksheet contains summary data from surveys for the past 5 years.
- *End-User Satisfaction*, measured on the same scale as dealers. Each year, 100 users from each region are surveyed. The worksheet contains summary data for the past 5 years.

[33]The case scenario was based on *Gateway Estate Lawn Equipment Co. Case Study*, used for the 1997 Malcolm Baldrige National Quality Award Examiner Training course. This material is in the public domain. The database, however, was developed by the author.

- *2014 Customer Survey*, results from a survey for customer ratings of specific attributes of PLE tractors: quality, ease of use, price, and service on the same 1–5 scale. This sheet contains 200 observations of customer ratings.
- *Complaints*, which shows the number of complaints registered by all customers each month in each of PLE's five regions (North America, South America, Europe, the Pacific, and China).
- *Mower Unit Sales* and *Tractor Unit Sales*, which provide sales by product by region on a monthly basis. Unit sales for each region are aggregated to obtain world sales figures.
- *Industry Mower Total Sales* and *Industry Tractor Total Sales*, which list the number of units sold by all producers by region.
- *Unit Production Costs*, which provides monthly accounting estimates of the variable cost per unit for manufacturing tractors and mowers over the past 5 years.
- *Operating and Interest Expenses*, which provides monthly administrative, depreciation, and interest expenses at the corporate level.
- *On-Time Delivery*, which provides the number of deliveries made each month from each of PLE's major suppliers, number on time, and the percent on time.
- *Defects After Delivery*, which shows the number of defects in supplier-provided material found in all shipments received from suppliers.
- *Time to Pay Suppliers*, which provides measurements in days from the time the invoice is received until payment is sent.
- *Response Time*, which gives samples of the times taken by PLE customer-service personnel to respond to service calls by quarter over the past 2 years.
- *Employee Satisfaction*, which provides data for the past 4 years of internal surveys of employees to determine their overall satisfaction with their jobs, using the same scale used for customers. Employees are surveyed quarterly, and results are stratified by employee category: design and production, managerial, and sales/administrative support.

In addition to these business measures, the PLE database contains worksheets with data from special studies:

- *Engines*, which lists 50 samples of the time required to produce a lawn-mower blade using a new technology.
- *Transmission Costs*, which provides the results of 30 samples each for the current process used to produce tractor transmissions and two proposed new processes.
- *Blade Weight*, which provides samples of mower-blade weights to evaluate the consistency of the production process.
- *Mower Test*, which lists test results of mower functional performance after assembly for 30 samples of 100 units each.
- *Employee Retention*, data from a study of employee duration (length of hire) with PLE. The 40 subjects were identified by reviewing hires from 10 years prior and identifying those who were involved in managerial positions (either hired into management or promoted into management) at some time in this 10-year period.
- *Shipping Cost,* which gives the unit shipping cost for mowers and tractors from existing and proposed plants for a supply-chain-design study.
- *Fixed Cost,* which lists the fixed cost to expand existing plants or build new facilities, also as part of the supply-chain-design study.
- *Purchasing Survey*, which provides data obtained from a third-party survey of purchasing managers of customers of Performance Lawn Care.

Elizabeth Burke has recently joined the PLE management team to oversee production operations. She has reviewed the types of data that the company collects and has assigned you the responsibility to be her chief analyst in the coming weeks. To prepare for this task, you have decided to review each worksheet and determine whether the data were gathered from internal sources, external sources, or have been generated from special studies. Also, you need to know whether the measures are categorical, ordinal, interval, or ratio. Prepare a report summarizing the characteristics of the metrics used in each worksheet.