CHAPTER 1

Using Data to Make Better Decisions

INTRODUCTION

Every day, in every walk of life, people make decisions and choices. Some decisions are relatively unimportant—such as what to eat for breakfast, what clothes to wear, and whether to take an umbrella when we leave home. Some decisions are more important—such as what post-secondary education to pursue, where to live, and how to manage our investments. We are likely to use data to make even relatively unimportant decisions. For instance, the weather forecast helps us decide what to wear and whether to take an umbrella. We probably gather more data before we make important decisions. For example, most students carefully research the alternatives before they choose a post-secondary educational institution.

In business, important decisions have to be made every day: Which of the candidates should be hired? Should a new product be launched? Should a new branch or factory be opened? Should one be closed? While we hear stories about larger-than-life executives who make such major decisions on gut feel, most of these choices are probably based on the analysis of relevant data. For example, the Starbucks Corporation, a leading coffee retailer, uses market research data to forecast sales and choose locations for its stores (you can read more

LEARNING OBJECTIVES

After mastering the material in this chapter, you will be able to:

1. Understand the approaches to gathering data.
2. Understand why sampling is necessary.
3. Recognize that there is art and science to summarizing and analyzing data.
4. Recognize that cause-and-effect conclusions must be drawn carefully.
5. Understand that clear and honest communication of results is necessary for them to be useful.
6. Be familiar with a framework for data-based decision making.

about this at the website **www.claritas.com/target-marketing/resources/case-study/starbucks.jsp**).

Why Do We Use Data to Make Decisions?

Using data to make decisions is better than guessing.

Collecting and analyzing data usually leads to a better understanding of the problem at hand, and a better decision about the solution, as you will see in the many examples in this text.

It is likely that you are reading this book because you are enrolled in a course in "statistics." By convention and tradition, we use this term to refer to techniques for analyzing data and making decisions. However, the word "statistics" is not very illuminating to most people, and is fear-inspiring to many. What you should know is that there is nothing to be afraid of. If you can develop the habit of looking at the data carefully before you make decisions, and if you master at least some of the techniques described in this book, you will make better decisions. This approach is not only *not scary*, it can be a real plus for your life and your career. And if you are thinking "not *my* career!" you are probably wrong. There is potential for using the techniques described in this book throughout the business world. An accountant doing an audit, a financial analyst deciding which stocks to recommend, a brand manager deciding on promotional strategies, a production manager assessing quality, a human resources expert recommending compensation plans—all of these people are using the tools of statistics to do their jobs better. Rather than resist the study of "statistics" because you've heard it's hard, realize that the tools you are going to learn about are essential to success in any business career.

The techniques involved in analyzing data and making decisions can also be fun. You might think of yourself as a data detective: you have to find the right data to help with the decision, and then you have to solve the mystery of what the data are telling you (or not telling you). It can be very satisfying to solve the mystery and arrive at the right decision.

This chapter provides you with a foundation for unravelling the mysteries of using data for decision making. Section 1.1 describes methods for getting the data you need. Section 1.2 describes the importance of sampling, and some of its challenges. Section 1.3 outlines the importance of analyzing the data that are collected, and Section 1.4 describes the kinds of conclusions that can (and cannot) be made from research studies. Section 1.5 stresses the importance of clearly communicating methods and conclusions. Section 1.6 provides a framework for a data-based decision-making approach, and describes how the material in this book relates to it.

1.1 GETTING THE DATA

Once you have understood the context of the problem at hand and identified the decision that is required, you will have to decide what data to collect. In some ways, this is the most important step in the process. Faulty data leads to faulty decisions.

There are two types of data that you might use to support decision making: data collected for your specific purpose are *primary data;* and data already collected for some other purpose are *secondary data.* Suppose you wanted to know if the average mark on the last statistics test you wrote was more than 60%. One way to find out would be to ask every student what mark he or she received on the test, and then calculate the average. In this case, you would be collecting primary data. Alternatively, if you could get access to the teacher's marks database, you could use this secondary data to calculate the class average.

Primary and Secondary Data

Primary data are data that are collected for your specific purpose. The advantage of primary data collection is that you can collect exactly the data you need, and as a result you are in a much better position to make decisions. As well, you can design data collection methods so that it is possible to draw conclusions about cause-and-effect relationships. This is discussed in more detail in Section 1.4.

Primary data Data that are collected for your specific purpose.

Many manufacturing enterprises have sophisticated systems to monitor quality continually. These companies collect a range of primary data so that they can adjust production processes when products do not meet specifications. For example, the "**Six Sigma**" approach to quality control was first introduced by Motorola Inc. in the mid-1980s. It began as a process to control the number of defects in a manufacturing setting, but has since developed into a system for controlling quality in a wide range of industries, including financial services and healthcare. The Six Sigma process relies on primary data collection.

BMO Financial Group appears to know the value of using statistical methods in business. The company first adopted **Six Sigma** methods in its Product Operations Group in 2005, resulting in fewer errors, less waste, and expected savings of $86 million over a five-year period. ("Six Sigma & Business Improvement Deployment Leader of the Year: Richard Lam, BMO Financial Group." The Global Six Sigma & Business Improvement Awards. **www.tgssa.com**.)

The disadvantage of primary data is that collecting it tends to be costly or difficult, as it usually involves designing and implementing some kind of survey. For example, if you tried to collect the data about marks on the last statistics test from your classmates, you might find it very time-consuming. To start with, you might not even know who the students in your class are. Even if you could identify them accurately, some of them might refuse to tell you their test mark, and some might lie about it.

Secondary data are data that were previously collected, not for your specific purpose. The advantage of secondary data is that it is usually cheaper to obtain than primary data. It may also be sufficient for your needs.

Secondary data Data that were previously collected, not for your specific purpose.

EXAMPLE 1.1

Secondary data

Suppose you wanted to know how the gross margin for your pharmacy in Cranbrook compared with other pharmacies in British Columbia. One of the easiest ways to get this data is through Statistics Canada.

Exhibit 1.1 below shows the Statistics Canada data for pharmacies and personal care stores in British Columbia.

EXHIBIT 1.1

Statistics Canada Data for BC Pharmacies and Personal Care Stores

	2006 Operating Statistics (Annual)			
	$ thousands			%
	Operating Revenues	Cost of Goods Sold	Operating Expenses	Gross Margin
BC pharmacies and personal care stores	3,621,121	2,494,323	997,726	31.1

Source: Adapted from Statistics Canada website, **www40.statcan.ca/l01/cst01/trad38k-eng.htm**, accessed February 25, 2009.

There are many sources of secondary data about business activities in Canada. Most of these sources provide at least some of the available information on websites. Some useful sources are:

- Statistics Canada. This federal government agency collects all sorts of business information, and some of it is available for free. Libraries often house a selection of Statistics Canada publications. The website for Statistics Canada is **www.statcan.gc.ca/start-debut-eng.html**.
- Online databases, which can often be accessed through a library. There are four of particular interest:
  - Canadian Business & Current Affairs™ (CBCA) provides access to approximately 140 Canadian industry and professional periodicals and newsletters covering business and current affairs.
  - The ABI/INFORM® database provides content from business journals and international and scholarly content.
  - The Business Source® Complete online service provides access to a collection of scholarly journals and trade publications.
  - LexisNexis Academic® provides full-text documents from over 5,600 news, business, legal, medical, and reference publications.
- Industry Canada, available at **www.ic.gc.ca/eic/site/ic1.nsf/eng/h_00072.html**. The website provides access to a wide variety of economic and market research reports and statistics.
- Provincial government websites, which provide a range of business information.
- Business publications, for example
  - *Canadian Business* magazine **www.canadianbusiness.com/index.jsp**
  - *Maclean's* magazine **www2.macleans.ca/category/business**
  - *The Globe and Mail* Report on Business **www.theglobeandmail.com/report-on-business**

This is just a short list of general business information sources. Industry-specific trade publications can also provide data to help you analyze your specific industry sector. Many of these sources provide at least some data on websites. The ease of searching for information on the web, however, should not prevent you from pursuing other sources of data (including those in print form). As well, you must recognize that many web-based sources of "data" are misleading, incomplete, biased, or just plain wrong.

It takes some skill to navigate the wide array of information available (especially on the Internet), and to identify which data sources are both useful and reliable. Librarians are trained in these activities, and when you are doing research a good librarian can be your best resource.

DEVELOP YOUR SKILLS 1.1

MyStatLab

1. Suppose you wanted to collect primary data about the music preferences of students at your school. How would you go about collecting it? What problems would you anticipate, and how would you solve them?
2. If you were trying to control quality at your bicycle manufacturing plant, what primary and secondary data could you use?
3. You are working in the bicycle manufacturing industry. You are interested in trends in household spending on bicycles and accessories. Find a secondary data source for this information.
4. Do you think data provided by government sources such as Statistics Canada are completely accurate and reliable? Do some research to see if your answer is correct.
5. Choose a Canadian industry of interest to you. Try to come up with three sources of secondary data about the industry.

1.2 SAMPLING

The data you require to make a good decision are often not readily available. Even if you have the resources to do primary research, you will not likely be able to gather all the data you want.

Why Sampling Is Necessary

As a business owner, you might be interested in characteristics of your current customers but not be able to survey all of them, either because you do not know how to contact them or because you cannot afford to. As a paint manufacturer, you might want to know that all of the cans coming off your production line contain the correct amount of paint, but you cannot (or cannot afford to) measure the paint level in all cans exactly. As a brewer, you might want to be sure that all of the beer you are producing is delicious, but if you drank all of it to check, you would have none left to sell!

Fortunately, you do not need all of the data in order to make reliable decisions. Many of the techniques described in this text allow you to make decisions on the basis of sample data. For example, you might be interested in whether your customers react differently to your products according to their ages. The complete collection of data—the ages of all of your current and potential customers and their product preferences—is described as *population data*. **Population data** are the complete collection of *all* of the data of interest. Because of cost and difficulty in surveying *all* of the elements in the population, we survey only some of them. **Sample data** are a subset of population data.

Population data The complete collection of *all* of the data of interest.

Sample data A subset of population data.

Exhibit 1.2 on the next page illustrates the difference between population data and sample data. Thousands of filled paint cans are coming off a paint-production line. For quality control, a sample of these cans is examined.

The reliability of the conclusions that can be drawn about population data on the basis of sample data depends on the type of sampling used: statistical or nonstatistical sampling.

Nonstatistical Sampling

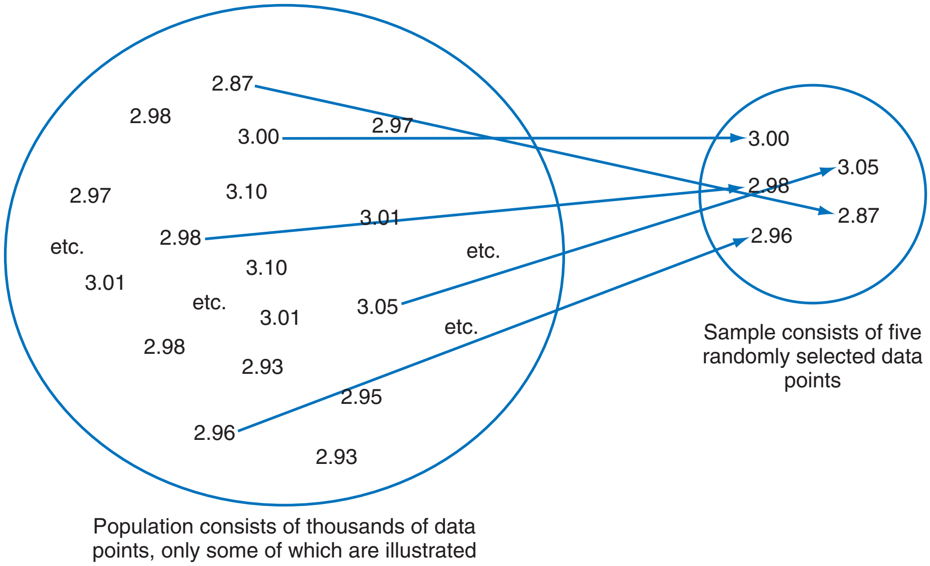
Nonstatistical sampling is sometimes referred to as *nonprobability sampling*. In **nonstatistical sampling**, the elements of the population are chosen for the sample by convenience, or according to the researcher's judgment. There is no way to estimate the probability (likelihood, or chance) that any particular element from the population will be chosen for the sample.

Nonstatistical sampling The elements of the population are chosen for the sample by convenience, or according to the researcher's judgment; there is no way to estimate the probability that any particular element from the population will be chosen for the sample.

With nonstatistical sampling, there is no way to measure the reliability of the sample results. Despite this, nonstatistical sampling can be useful. For example, although Statistics Canada uses statistical sampling for most of its data collection, it uses nonprobability sampling for testing questionnaires. Nonstatistical sampling can also legitimately be used to learn about emerging trends, or to test new product ideas. Focus groups are sometimes convened for these purposes.

EXHIBIT 1.2

Population and Sample Data, Amount of Paint in 3-Litre Cans



EXAMPLE 1.2A

Nonstatistical sampling

A retail store might take a convenience sample by surveying the first 50 customers who come into the store.

This could provide useful information, particularly if there is reason to believe that all customers of the store are essentially the same. But since customers 51 and up will not be chosen in such a sampling approach, its results should be used with caution. The customers who arrive at the store early may have very different characteristics and preferences from other customers.

EXAMPLE 1.2B

Nonstatistical sampling

The websites for some of Canada's newspapers conduct online polls and report results. For example, on February 25, 2009, *The Globe and Mail* posed this question: Has the current economic situation caused you to rethink your career goals? (*Source:* "Poll Results," www.theglobeandmail.com, accessed February 25, 2009.)

The poll results: 1,472 votes of "yes" (44%), and 1,873 votes of "no" (56%).

The sampling approach here is nonstatistical, because responses were made by volunteers. Such a sample can be described as "self-selected," because the respondents themselves choose to provide sample data. The poll results should never be interpreted as representative of the views of all Canadians. At best, such a poll indicates the views of those who visited the newspaper's website and took the time to respond to the poll. Such polls are best considered as entertainment.

Statistical Sampling

Statistical sampling is sometimes referred to as probability sampling. In **statistical sampling**, the elements of the population are chosen for the sample in a random fashion, with a known probability of inclusion. In this context, "random" does not mean "arbitrary" or "haphazard." It requires some thought and effort to select a random sample.

Statistical sampling The elements of the population are chosen for the sample in a random fashion, with a known probability of inclusion.

It is possible to make very reliable decisions about population data on the basis of data acquired through statistical sampling. The decision-making process is referred to as "inferential statistics." **Inferential statistics** is a set of techniques that allows reliable conclusions to be drawn about population data, on the basis of sample data. Usually, inferential statistics involves drawing some conclusion about, or estimating a population parameter on the basis of, a sample statistic. For example, you might want to know the average amount of paint in *all* of the paint cans you produce. This average is a population parameter. A **parameter** is a summary measure of the population data. A **sample statistic** is a summary measure of the sample data. You might take a random sample of 30 paint cans, measure the amount of paint in each can exactly, and calculate a sample average. This sample statistic will be the starting point for your conclusion about the population parameter. It would be too costly to measure the amount of paint in every can, and it isn't necessary. We can use the sample results to decide whether the paint filling line is working properly.

Inferential statistics A set of techniques that allow reliable conclusions to be drawn about population data, on the basis of sample data.

Parameter A summary measure of the population data.

Sample statistic A summary measure of the sample data.

Reliable statistical inference is possible only if the sample data are a good reflection of population data. While even the best sampling techniques cannot *guarantee* this, with statistical sampling we can estimate and control the possibility of error. It is this control that makes statistical sampling such a powerful aid to decision making.

There are many possible methods of statistical sampling. The particular sampling plan will depend on the type of data being collected, the costs involved, and the goals of the analysis or decision. The simplest of the possible methods is simple random sampling, and the inferential techniques discussed in this text all assume simple random sampling. The methods required to analyze data from more complex sampling plans (e.g., systematic sampling, stratified sampling, cluster sampling) are extensions of the foundation techniques based on simple random sampling.

Simple random sampling is a sampling process that ensures that each element of the population is equally likely to be selected. Suppose you wanted to collect some information about customer satisfaction with the delivery and installation service your store provides (for a fee) with the home theatre systems it sells. You have the budget for 10 interviews, and you want to talk with a random sample of customers. How would you select a simple random sample of 10 customers to interview?

Simple random sampling A sampling process that ensures that each element of the population is equally likely to be selected.

In order to conduct a true random sample, you must start with a list of the elements in the population. This list is called a **frame**. Because your technicians had to deliver the home theater systems, you probably have a list of the names and addresses of the customers of interest.

Frame A list of the elements in a population.

You could write all these customer names on slips of paper, put them into a container, mix them up, and then select 10 slips of paper. You would then interview the 10 randomly selected customers. However, this is a tedious process, and there is an easier way, particularly if the customer names are in a computer database. Random number generation software allows us to use computers to do the electronic equivalent of mixing up names in a hat and selecting some for a sample.

EXAMPLE 1.2C

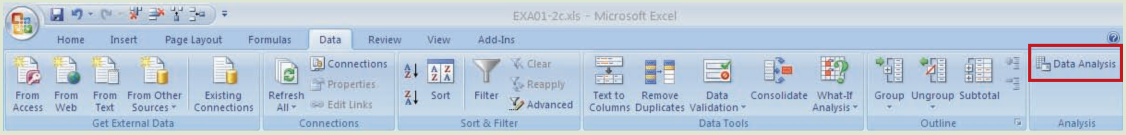
Random sampling with Excel

Suppose you have a list of the 60 customers who have bought home theatre systems in the last six months, and who also paid to have your store's technicians deliver and install the systems. You want to select a random sample of 10 of these customers for interviews about their service experience.

You can use Excel to do this. Begin by installing the **Data Analysis** tools of Excel (refer to the instructions on page xxx).

You can use Excel to generate a column of 60 random numbers beside the columns of customer names. Each name then has a randomly assigned random number tag that can be used to select a random sample. The process is as follows.

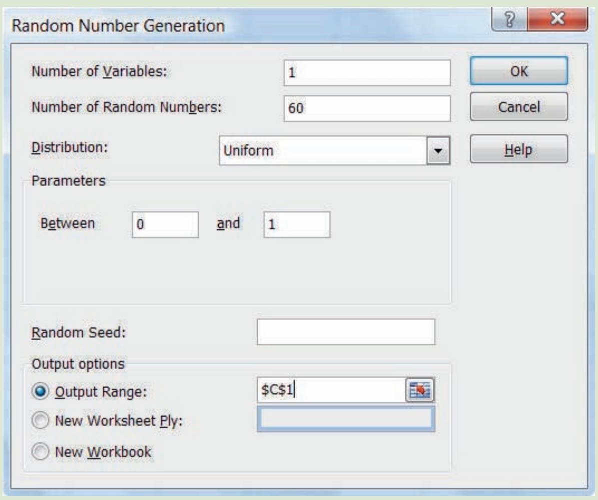
Click on the **Data** tab, then click on **Data Analysis**.



Then select **Random Number Generation**. This will activate the dialogue box illustrated in Exhibit 1.3 below.

EXHIBIT 1.3

Random Number Generation in Excel



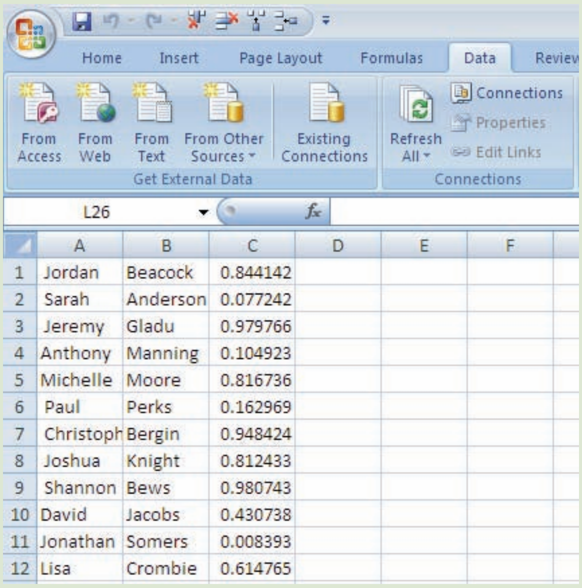
You must enter:

- **Number of Variables**, which in this case will be "1"
- **Number of Random Numbers**, which in this case will be "60." Remember, you have to create a random number for every element in the population (frame).
- **Distribution**, which should always be **Uniform** (this ensures that each digit is equally likely to appear)
- **Between** defaults to "0" and "1," which is fine
- **Output Range**, which is the top of the column next to the column containing the family names

The output you create will look something like the excerpt shown in Exhibit 1.4 below (your output may have different random numbers).

EXHIBIT 1.4

Excerpt of Excel Output for Random Number Generation (only the first 12 rows are shown)



	A	B	C	D	E	F
1	Jordan	Beacock	0.844142			
2	Sarah	Anderson	0.077242			
3	Jeremy	Gladu	0.979766			
4	Anthony	Manning	0.104923			
5	Michelle	Moore	0.816736			
6	Paul	Perks	0.162969			
7	Christoph	Bergin	0.948424			
8	Joshua	Knight	0.812433			
9	Shannon	Bews	0.980743			
10	David	Jacobs	0.430738			
11	Jonathan	Somers	0.008393			
12	Lisa	Crombie	0.614765			

⋮

(columns continue)

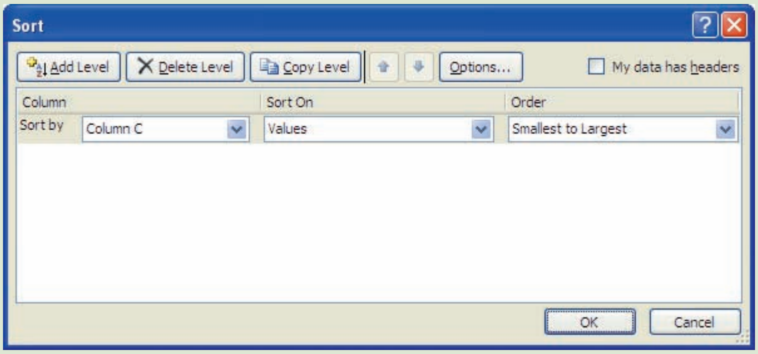
The first two columns contain the first and last names of the customers who used the delivery and installation service (your population data), and the third column contains 60 random numbers between 0 and 1. Now you can use Excel to shuffle the customer names randomly, according to the random numbers in the third column.

Place your cursor on one of the cells in the column of random numbers. Then select the **Data** tab and **Sort**, which will activate the window shown in Exhibit 1.5 on the next page.

You want to sort by column C, which is the column of random numbers, so choose that column for **Sort by**. The default sort will be based on the values of the numbers, from smallest to largest, and these defaults are fine. When you click **OK**, Excel will

EXHIBIT 1.5

Sort Dialogue Box



rearrange the population of customer names in a random way. An example of the output is shown in the excerpt below in Exhibit 1.6 (your output may have different customer names).

EXHIBIT 1.6

Excerpt of Excel Output, with Customer Names Sorted by Random Numbers

	A	B	C
1	Jonathan	Somers	0.008393
2	Andrea	Parulski	0.021546
3	Denise	Brown	0.0318
4	Tina	Brown	0.071657
5	Sarah	Anderson	0.077242
6	Colin	Sage	0.083346
7	Julie	Grainger	0.086673
8	Donnie	Schussler	0.090793
9	Vanessa	Ormsby	0.094943
10	Anthony	Manning	0.104923
11	Marc	Schalich	0.137516
12	Paul	Perks	0.162969
13	Tara	Alcott	0.165288
14	Sarah	Domchek	0.179052
15	Jennifer	Ansell	0.18955
16	Devon	Breau	0.220405
17	Shawn	Hellings	0.257149

⋮

(columns continue)

The names in the first two columns are equivalent to those you would pick out of a hat after you have electronically shuffled them around. If your sample size is 10, you would select the population elements corresponding to the first 10 names in the shuffled list, that is: Jonathan Somers, Andrea Parulski, Denise Brown, Tina Brown, Sarah Anderson, Colin Sage, Julie Grainger, Donnie Schussler, Vanessa Ormsby, Anthony Manning. (Note that these 10 names are shaded in for Exhibit 1.6, but they will not be in your Excel spreadsheet.) These are the names of a simple random sample of customers to interview about the delivery and installation service.

This discussion is designed to make you realize that taking a true random sample requires thought and effort. You may have enough information to create a true random sample if you are studying your own company. For example, you can take a random sample of employees or production output. You may also have sufficient data on hand to conduct a random sample of past customers. But you will never be able to conduct a true random sample of potential customers, because there is no way to know who they are.

Polling companies who conduct opinion polls of the population at large are also challenged to identify a true random sample. Polling companies generally gather data by telephone interviews, although online surveys are increasingly used. Of course, telephone surveys leave out anyone who does not have a phone. In Canada, over 98% of households have at least one phone number, and this small difference between the sampled population and the target population is usually not considered to be a significant problem. The telephone survey approach makes it difficult to survey those who do not answer their phones or will not respond to telephone surveys. Polling companies go to considerable trouble to overcome these difficulties so that they can make reliable statements about survey results.

Sampling and Nonsampling Error

Sampling error is the difference between the true value of the population parameter and the value of the corresponding sample statistic. Sampling error is expected. It would be unreasonable to expect the sample statistic to match the population parameter exactly. For example, if we managed to collect the test marks of a random sample of students, we would not expect the sample average mark to exactly match the true class average. Sampling error is something that we can estimate and control, as you will learn.

Sampling error The difference between the true value of the population parameter and the value of the corresponding sample statistic.

Nonsampling errors are other kinds of errors that can arise in the process of sampling a population. Suppose you plan to conduct a telephone survey of past customers about their product preferences. Here are some nonsampling errors that can occur.

Nonsampling errors Other kinds of errors that can arise in the process of sampling a population.

1. Your survey frame may contain errors. Some customers may have been missed. Their phone numbers may have been incorrectly recorded or missing. Other information may not be correct. Such **coverage errors** arise because of inaccuracy or duplication in the survey frame.
2. Some of the customers you wish to speak to may never answer the phone. This causes **nonresponse error**, which arises when data cannot be collected for some elements of the sample.
3. There can be errors in acquiring the data. If the survey questions are biased or misleading, or difficult to understand, the customer being interviewed may not give

Coverage errors Errors that arise because of inaccuracy or duplication in the survey frame.

Nonresponse error Error that arises when data cannot be collected for some elements of the sample.

Response errors Errors that arise because of problems with the survey collection instrument (e.g., the questionnaire), the interviewer (e.g., bias), the respondent (e.g., faulty memory), or the survey process (e.g., not ensuring that the respondent fits into the target group).

Processing errors Errors that occur when the data are being prepared for analysis.

Estimation errors Errors that arise because of incorrect use of statistical techniques, or calculation errors.

truthful or accurate answers. If the interviewer is not well trained, he or she may influence the survey responses. These **response errors** arise because of problems with the survey collection instrument (e.g., the questionnaire), the interviewer (e.g., bias), the respondent (e.g., faulty memory), or the survey process (e.g., taking answers from someone other than the intended respondent).

4. There can be errors in recording the data, even when data are collected with the help of a computer, and this can lead to biased results. **Processing errors** occur when the data are being prepared for analysis.
5. **Estimation errors** arise because of incorrect use of statistical techniques, or calculation errors.

It is important that you recognize that nonsampling errors can invalidate the conclusions drawn from the sample. You should be watchful for nonsampling errors when you examine research done by others, and you should take great care to avoid these errors when you are sampling. With some effort, you can greatly minimize the possibility of nonsampling error.

DEVELOP YOUR SKILLS 1.2

MyStatLab

6. Nowadays, many companies use customer relationship management (CRM) software to keep track of customer sales information, financing arrangements, and product preferences. Is this sample or population data? See if you can find an article that describes how a particular company uses CRM data.

7. A restaurant owner decides to survey diners on Friday night, because he wants to collect a lot of data and Friday is always busy. Is this a statistical or a nonstatistical sample? Why? How much should the restaurant owner rely on the sample data?

8. "A new Ipsos-Reid survey conducted on behalf of Mosaik MasterCard finds that three-quarters (77%) of Canadian postsecondary students have at least one credit card. Of these students, 72% are currently carrying a balance, and 53% plan to pay it off entirely by their next statement due date." Are these population parameters or sample statistics? (*Source:* "School Credits of a Different Kind: The Mosaik MasterCard Back to School Student Survey on Credit Card Knowledge," **www.ipsos-na.com/news/pressrelease.cfm?id=2763**, accessed June 7, 2006.)

9. As part of an employee satisfaction survey, a research team wants to conduct in-depth interviews with a random sample of 10 employees. The company's employee list is available in an Excel spreadsheet in the MyStatLab that accompanies this text. Use Excel to create a random sample of 10 employees.

DYS01-9

10. Suppose Calgary Transit wants to do a survey to see if it can improve its services for people with disabilities. The organization decides to collect information through interviews of a random sample of riders. Outline some challenges Calgary Transit might face in gathering the data.

1.3 ANALYZING THE DATA

Once you collect the data you need, the information must be organized so that you can make sense of it. Raw (that is, unorganized) data usually do not tell us much.

EXAMPLE 1.3

Analyzing the data

Suppose a Niagara region winery is interested in discovering if there is a difference between men's and women's average purchases from the winery. A random sample of 25 men and 28 women is selected, and their purchase amounts are collected. The data for purchases by men are as follows: $52.40, $20.67, $38.93, $51.32, $50.38, $46.80, $49.80, $43.19, $49.14, $22.96, $27.72, $15.71, $13.84, $24.27, $26.72, $10.58, $29.18, $31.15, $37.62, $31.61, $42.08, $31.56, $52.11, $34.98, $33.77. The data for purchases by women are: $46.32, $58.85, $47.82, $68.57, $13.80, $30.12, $37.30, $43.54, $24.73, $29.49, $67.55, $53.11, $13.17, $30.40, $49.42, $40.22, $53.99, $41.17, $28.36, $51.11, $34.76, $44.82, $58.00, $25.37, $67.25, $32.97, $31.09, $48.30.

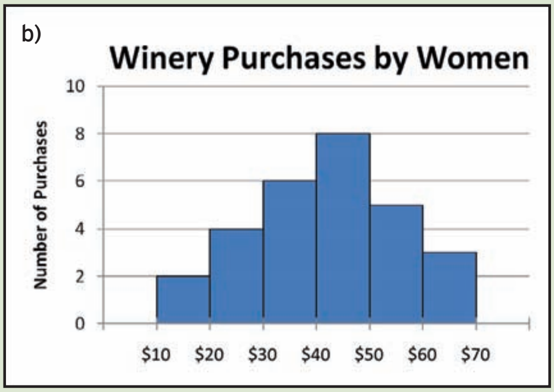
EXA01-3

Just looking at these lists of numbers does not tell us much. We cannot easily see if there is a difference in the purchases made by men and women. Probably one of your instincts is to calculate the men's and women's average purchases. The average is an example of a numerical summary measure, and is discussed along with other numerical summary measures in Chapter 3. It is often helpful to create a graph to summarize data, and such graphs are discussed in Chapter 2 of this text. The techniques in Chapters 2 and 3 are part of a branch of statistics called "descriptive statistics." Once you master these, you will be able to produce graphs and summary statistics that help you see the data more clearly. Exhibit 1.7 shows some graphs that organize and display the winery purchase data for men and women, as well as a table with some summary statistics.

These graphs and summary measures have probably helped you understand the two data sets better. You can see that the average purchase by women is higher than the average purchase by men. You can also see that there is greater variability in the purchases by women than the purchases by men.

EXHIBIT 1.7

Winery Purchase Data



c) **Summary Statistics for Winery Purchases**

Winery Purchases	Average Purchase	Lowest Purchase	Highest Purchase
By Men	$34.74	$10.58	$52.40
By Women	$41.84	$13.17	$68.57

Descriptive statistics A set of techniques to organize and summarize raw data.

Descriptive statistics are a set of techniques to organize and summarize raw data. As you explore the possibilities for summarizing and describing data, you will discover that there is both art and science involved. While there are many guidelines to help you, you will usually have to make choices about presenting data. Your goal should always be to represent the data truthfully. You should also be aware that some will choose to confuse or misrepresent data, and you should always be alert to these possibilities. Even if you do not become a statistician, you should be an informed and suitably skeptical consumer of statistical analysis done by others. You will learn more as you explore Chapters 2 and 3.

DEVELOP YOUR SKILLS 1.3

MyStatLab

11. An advertisement claims that prices for electronic organizers have decreased by 125%. What do you think of this claim?

12. In the latter part of 2008, stock markets around the world declined significantly. Exhibit 1.8 shows two graphs depicting the Standard and Poor's Toronto Stock

EXHIBIT 1.8

S&P TSX Composite Index, January 2, 2007, to March 27, 2009

Source: "S&P/TSX Composite Index (Interi (^GSPTSE)," Yahoo Canada Finance, **http://ca.finance.yahoo.com/q/hp?s=%5EGSPTSE**, accessed March 28, 2009.

Exchange Composite Index for the period from January 2, 2007 to March 27, 2009. Which graph is the better representaion of the index during this time period, and why?

13. "[Student] Jane Woodsman has shown astonishing progress in academic achievement over the last semester. Jane's average grade has increased by a substantial 20%." Based on this statement, what is your impression of Jane Woodsman's academic success? (State your conclusion before you read any further.)

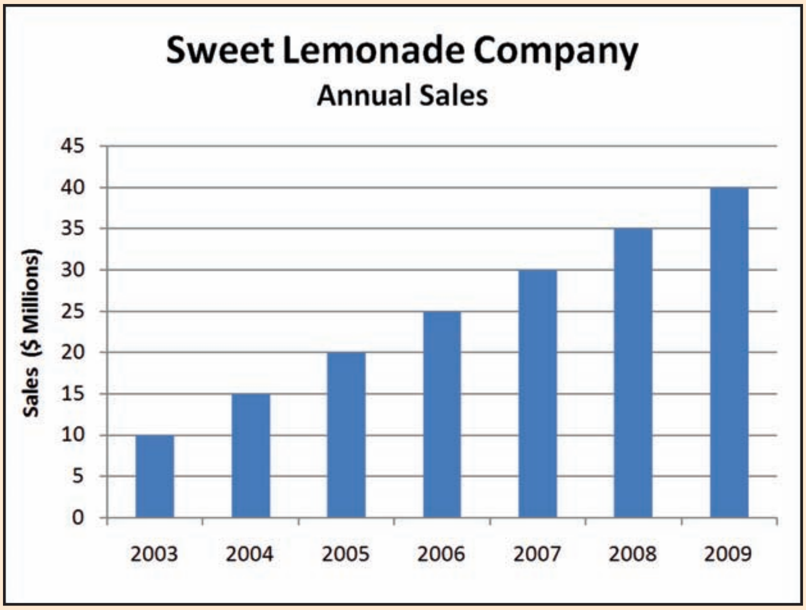
 Now, here are the facts. Last semester, Jane's average grade was 13.8%. This semester, Jane's average grade is 20% higher, that is, 1.20(13.8%) = 16.6%. While Jane's grades have improved, she is still a long way from academic success. How would you rewrite the initial statement in this question so that it was more honest?

14. You are the national manager for a company that provides express oil changes at a number of locations, and you are concerned that the service level at one of these locations is not up to your standard. The location keeps records of its service completion times (from customer arrival to departure). You've asked to see these records, but the manager says he's too busy to send them to you. He also claims that the maximum wait time during the recent quarter has decreased by 50%, compared with the previous quarter. Would you be satisfied, or would you insist on seeing the records?

15. Consider the graph shown in Exhibit 1.9 below. Do you think it is a fair representation of sales over the period?

EXHIBIT 1.9

Annual Sales Graph



1.4 MAKING DECISIONS

After you have collected and analyzed the data, you will be in a position to make some decisions. The type of conclusion that you can draw depends on how the data were gathered.

Suppose a newspaper wants to know if its readers' average income has increased, compared with five years ago. This information could be very useful, because an attractive reader income profile helps to sell advertising. In this case, the decision to be made is a straightforward one: has readers' average income increased? By looking at the sample evidence, we can decide whether or not this appears to be the case, and you will learn reliable techniques to do this as you progress through this text. It is unlikely that the newspaper is particularly interested in *why* readers' average income has increased, if in fact it has.

Observational study The researcher observes what is already taking place.

This is an example of an observational study. In an **observational study**, the researcher observes what is already taking place. Many marketing studies are observational in nature. Companies collect data about their customers in order to understand them better.

Sometimes research is aimed at understanding cause and effect, that is, pinpointing *why* an observed change or difference has occurred. Example 1.4a illustrates.

EXAMPLE 1.4A

Cause and effect cannot be concluded from observational studies

Suppose that the vice-president of operations is interested in how training methods affect the number of worker errors. One way to study this would be to record the number of errors for a random sample of workers at two different plants—one where training has taken place, and one where it has not.

The first decision that must be made is whether the sample data provide evidence to suggest that the error rates differ between the two plants. Once you have mastered the techniques in this text, you will be able to make this decision.

But there is also a further question. If the error rates differ, is it because of the training program? The vice-president might be tempted to conclude that it is. But this further conclusion cannot be made on the basis of an observational study such as the one described here. Factors other than the training program could have caused the difference in the error rates. For example, the skills and experience of the workers could affect the number of errors. Choosing a random sample at each plant may result in these factors being about equal across the two samples. But there are other factors that will not be randomized. For example, the supervisor at one plant might be better at motivating workers to make fewer errors. Traffic conditions may be better at one plant, so that the workers arrive less tired and frustrated, and less prone to error. Because of these other possibilities, it is not possible to make a strong conclusion about the cause of the differences in error rates on the basis of this observational study.

Experimental study The researcher designs the study so that conclusions about causation can be drawn.

To make a stronger conclusion about the cause of any difference in error rates, the vice-president could conduct an experimental study. An **experimental study** is designed so that conclusions about causation can be drawn.

EXAMPLE 1.4B

Cause and effect may be concluded from experimental studies

Suppose an experimental study is designed as follows. A group of workers is randomly selected at a particular plant, and the numbers of errors made by the group are recorded over a period of time. The researcher then intervenes by putting the workers through a training program. The numbers of worker errors are recorded again after the training.

In this case, if the numbers of worker errors after the training are lower, it is easier to conclude that the training is the cause. However, even in this case, there could be other explanations. For example, the workers who took the training might have been motivated to pay particular attention and work carefully, simply because they were selected for the study. The results of the study might not generalize to the entire population of workers.

You can see from Example 1.4b that even experimental studies can have limitations, although generally they lead to stronger conclusions than observational studies. It is very important that you recognize what conclusions are justified—and what conclusions are *not* justified—from any particular research study. While statistical analysis can add to our understanding of the world, it cannot tell us everything.

This may surprise you. Many people expect that since statistical analysis involves data and mathematics (or at least arithmetic) and computers, the results should be very clear. This can sometimes be the case, but not always. The claim can be made that "using data to make decisions is better than guessing." This is true, but it does not imply that statistical techniques always lead to clear answers. Judgment is required to interpret the results of statistical analyses. In fact, judgment is required throughout the analysis, as you will see during your introduction to the techniques in this text. The ability to make good judgments develops with practice, experience, and reflection. This does not mean that anything goes, or that you can "prove" anything you want with statistics. Some conclusions and some choices are clearly unacceptable. In cases where judgment is required, you must be prepared to defend your choices objectively. And you should always critically evaluate the judgments made by others in their statistical analyses.

DEVELOP YOUR SKILLS 1.4

MyStatLab

16. A study published in 2006 indicated that higher family incomes are associated with better physical, social/emotional, cognitive, and behavioural well-being among children (*Source:* "Study: Family Income and the Well-Being of Children," **www.statcan.ca/Daily/English/060511/d060511c.htm**, accessed June 8, 2006). Barbara Arneil, a political science professor at the University of British Columbia, was quoted in a related article in *The Globe and Mail* as saying that the findings suggest the following: "If the outcomes for our kids are better with higher incomes, the goal should be to get both parents into the work force." (*Source:* "Well-Off Children Do Better on Tests, Study Finds," **www.theglobeandmail.com**, accessed May 15, 2006). Should you agree with Professor Arneil's conclusion?
17. A study of 10,000 people born in the UK in 1958 revealed that a man 6 feet in height was more likely to have a partner and children than was a man of average height (5 feet, 10 inches). Does this mean that women prefer to marry tall men?
18. An insurance company with a number of branch offices across the country has designed a diary system for salespeople, which is designed to increase their productivity. The diary system is colour-coded: green is for time spent with clients making sales calls, blue is for time spent supporting existing clients, and red is downtime (lunch, travelling, etc.). The idea is that the visual cue of the coded diary will focus the salesperson's attention on sales. A number of poor performers were selected and introduced to the diary system. Average sales increased by 11%. Can we conclude that the diary system caused the increase?
19. If you wanted to test the effectiveness of the diary colour-coding system described in the previous exercise, how would you design your study?
20. An ice cream store hired a mascot to entice people into the store to buy ice cream. During the period when the mascot was working, sales increased by 15%. Can we conclude that the mascot caused the increase in sales?

1.5 COMMUNICATION

This text focuses on introductory techniques for analyzing data and making decisions about population data on the basis of sample data. Once you master these techniques, you will be on your way to becoming a successful data detective. However, your detective work will be useful only if you learn to communicate your methods and conclusions clearly.

When reporting on your statistical analysis, you should completely describe the problem at hand, and the goal of any decisions to be made. You should also provide information about how the data were collected, organized, and summarized, usually showing both graphs and numbers. Describe the statistical techniques you used, as well as any judgments about areas of uncertainty. Finally, make a clear statement of conclusions or decisions, with justification. You may provide the technical detail in the body of the report or in an appendix, but it must always be presented so that an informed reader can assess the methods used and whether the conclusions are justified. How much of the technical detail is included in the body of the report depends on its intended reader: a report written for colleagues who are familiar with statistics would be written differently than a report for a supervisor with little understanding of statistics.

It is also important that you state your conclusions carefully, as Example 1.5 illustrates.

EXAMPLE 1.5

State conclusions carefully

Consider a study of a newspaper's readers' incomes. Here are three different ways the conclusion of the study might be expressed:

1. With a p-value of 0.0123, H_0 is rejected.
2. The average income of this newspaper's readers has increased over the last five years.
3. Sample evidence suggests that the average income of this newspaper's readers has increased over the last five years.

While statement 1 might be technically correct, it cannot be understood without reference to the study. But this is a minor point. The more important problem is that most people would have no idea what this statement means, even after referring to the study! It is important that you learn to state the results of your statistical analysis in language that can be easily understood by almost anyone. While you need to understand the specialized language of statistics, you should also learn to translate it for an audience without statistical knowledge. Bosses and clients are not impressed by fancy jargon; they are frustrated by it.

Statement 2 is much more promising, because anybody could read it and understand it. However, this statement has a major problem: it does not even hint at the fact that this conclusion is made on the basis of sample results. Even if the sampling and the analysis were done properly, there is still a chance that a rare and unrepresentative sample was taken. The conclusion could be completely wrong. However, the statement does not even hint at this possibility, and so it is simply not correct, unless you have somehow managed to collect all the relevant population data.

Statement 3 is the best of the three, by far. First of all, it is understandable. It also indicates that a sample has been taken, and it is not nearly as definitive as statement 2.

You might think that the differences between statement 2 and statement 3 are minor. The newspaper's employees are probably going to act and speak as if statement 2 were true, as they try to sell more newspaper ads, so why go to the bother of stating conclusions so carefully? The answer is quite simple: statement 2 is not truthful.

DEVELOP YOUR SKILLS 1.5

MyStatLab

21. Think about the likely statistical knowledge of the intended reader of your statistical analysis in the following situations, and describe how much technical detail you would plan to include in the body of the report.

a. The national manager of quality control wants a report on quality control measures at the plant where you work.

b. The human resources department has asked you to analyze the educational levels of middle managers in the organization.

c. Your boss has asked you to prepare a report on a new product's characteristics, which will be used in a consumer magazine article about the product.

22. Comment on the following statement about a study based on a random sample of past customers of your firm: "Study results prove that our products are more attractive to younger buyers, so we should target the under-35 market."

23. One of the mistakes that beginning statisticians often make is to focus their reports on themselves, instead of on the data being analyzed. Instead of talking about what "I" did, you should be talking about what the data show. Rewrite the following statement with focus on the data, not the analyst.

"I took a random sample of 30 paint cans off the line, and carefully measured the exact amount of paint in each can. The sample average was 3.012 litres, just a bit above the desired level of 3 litres. I plotted this on the control chart, and I noticed that this number was within control limits. So I concluded that there was no problem with the paint filling line."

24. Another mistake that beginning statisticians make is to use subjective language to describe results, thus influencing the reader's interpretation. The goal should be to let the data speak for themselves. Rewrite the following statement without the subjective flavour, and with the proper focus on the data. (Use your imagination to make the report more specific, if you wish.)

"After extensive and painstaking analysis, I am delighted to conclude that our study shows that the new training program dramatically increased worker productivity. You should not hesitate to expand the training program across the nation, as the analysis is crystal-clear."

25. "The study concludes that being tall increases your chances of being a CEO." Do you think that this conclusion could be valid? How would you rewrite this to be more truthful?

1.6 A FRAMEWORK FOR DATA-BASED DECISION MAKING

A general approach to data-based decision making is outlined below. It summarizes the discussion in this chapter.

Steps in Data-Based Decision Making

1. Understand the problem and its context as thoroughly as possible. Be clear about the goal of a good decision.

2. Think about what kind of data would help you make a better decision. See if helpful data are already available somewhere. Decide how to collect data if necessary, keeping in mind the kind of conclusion you want to be able to make.
3. Collect the data, if the benefit of making a better decision justifies the cost of collecting and analyzing the data.
4. Examine and summarize the data.
5. Analyze the data in the context of the decision that is required. This may require using the sample data to:
   - estimate some unknown quantity
   - test if a claim (hypothesis) seems to be true
   - build a model
6. Communicate the decision-making process. This requires:
   - a clear statement of the problem at hand, and the goal of the decision
   - a description of how the data were collected or located
   - a summary of the data
   - a description of the estimation, hypothesis test, or model-building process(es)
   - a statement of what decision should be made, with justification

All of the steps in this process are important. Below are some comments about how these steps relate to the material in this text.

Step 1 In the pages of this book, you will find regular encouragement to *think* carefully about the problem at hand and how to approach it. Learning to do this successfully takes practice.

Steps 2 and 3 Generally, in this book, the data will be provided for you and simple random sampling will be assumed. If you continue your study of statistics beyond your introductory course, you will at some point have to gain knowledge and experience concerning primary research and design of experiments.

Step 4 Part II of the book (Chapters 2 and 3) provides an introduction to the methods of descriptive statistics.

Step 5 Most of this book is focused on inferential statistics. In Part III (Chapters 4–6), the building blocks for inferential statistics are presented. Part IV (Chapters 7–12) introduces a range of techniques to use sample evidence to estimate unknown quantities or decide if a claim appears to be true. Part V (Chapters 13 and 14) introduces techniques to model the relationships between a quantitative variable and one or more explanatory variables.

Step 6 Throughout this text, there are Guides to Decision Making (see the detailed table of contents), which will help you with the technical side of the decision-making process. While this text does not focus on report writing, you may have an opportunity to develop your communication skills as you complete assignments associated with the course you are probably taking.

Chapter 1 has given you an overview of data-based decision making. The techniques you are about to learn are very powerful, but only if they are used correctly and wisely. You will be able to use this chapter as a reference when you are thinking about a business decision (or any other kind of decision) that you need to make.

Chapter Summary

1

Getting the Data

Primary data are data that are collected for your specific purpose. Secondary data are data that were previously collected not for your specific purpose. Some useful sources of secondary data about business activities in Canada are listed on page 4.

Sampling

Population data are the complete collection of *all* of the data of interest. Sample data are a subset of population data. The reliability of the conclusions that can be drawn about a population on the basis of sample data depends on the type of sampling used: statistical or nonstatistical sampling.

In nonstatistical sampling, the elements of the population are chosen for the sample by convenience, or according to the researcher's judgment. With nonstatistical sampling, there is no way to measure the reliability of the sample results. Despite this, nonstatistical sampling can be useful for testing questionnaires, learning about emerging trends, or testing new product ideas.

Statistical sampling is sometimes referred to as probability sampling. In statistical sampling, the elements of the population are chosen for the sample in a random fashion, with a known probability of inclusion. It is possible to make very reliable decisions about population data on the basis of data acquired through statistical sampling, using a set of techniques called inferential statistics. Usually, inferential statistics involves drawing some conclusion or estimating a population parameter on the basis of a sample statistic. A parameter is a summary measure of the population data. A statistic is a summary measure of the sample data.

The simplest method of statistical sampling is simple random sampling. Simple random sampling is a sampling process that ensures that each element of the population is equally likely to be selected. In order to conduct a true random sample, you must start with a list of the elements in a population, which is called a frame. It is possible to use random number generation software to randomly select elements from the population, as Example 1.2c on pages 8–11 illustrates.

Sampling error is the difference between the true value of the population parameter and the value of the corresponding sample statistic, which arises because we are examining only a subset of the population. Sampling error is expected, and is something that we can estimate and control. Nonsampling errors are other kinds of errors that can arise in the process of sampling a population. Nonsampling errors can invalidate the conclusions drawn from the sample.

Analyzing the Data

Once you collect the data you need, the information must be organized so that you can make sense of it. Raw (that is, unorganized) data usually do not tell us much. Descriptive statistics are techniques to organize and summarize raw data.

Making Decisions

In an observational study, the researcher observes what is already taking place. For example, many marketing studies are observational in nature. Companies collect data about their customers in order to better understand them.

Sometimes research is aimed at understanding cause and effect, that is, *why* an observed change or difference has occurred. Cause and effect cannot be concluded from observational studies. An experimental study is designed so that conclusions about causation can be drawn. Even experimental studies have their limitations in terms of establishing cause-and-effect relationships, particularly when human behaviour is being studied.

Judgment is required to interpret the results of statistical analyses. In cases where judgment is required, you must be prepared to defend your choices objectively. You should always critically evaluate the judgments made by others in their statistical analyses.

Communication

A report of any statistical analysis you do should completely describe the problem at hand, and the goal of any decisions to be made. It should also provide information about how the data were collected, organized, and summarized (usually with both graphs and numbers). The statistical techniques used and any judgments about areas of uncertainty should also be described. Finally, a clear statement of conclusions or decisions should be made, with justification.

The technical detail may be provided in the body of the report or in an appendix, but it must always be presented so that an informed reader can assess the methods used and whether the conclusions are justified. How much of the technical detail is included in the body of the report depends on its intended reader.

A Framework for Data-Based Decision Making

A general approach to data-based decision making is outlined in the box on pages 19–20.

MyStatLab

Go to MyStatLab at www.mathxl.com. You can practise the exercises indicated with red in the Develop Your Skills and Chapter Review Exercises as often as you want, and guided solutions will help you find answers step by step. You'll find a personalized study plan available to you too!

CHAPTER REVIEW EXERCISES

WARM-UP EXERCISES

1. Why do businesses collect data before making major business decisions?

2. Why are business decisions made on the basis of only a limited amount of data? Can such decisions be reliable?

3. Is it valid to draw conclusions about population data on the basis of a convenience sample? Why or why not?

4. What numerical measure do students often rely on to summarize their performance in a particular semester?

5. Why is it important to communicate the results of your statistical analysis so that non-statisticians can understand?

THINK AND DECIDE

6. What are some of the difficulties in gathering data through personal interviews? As food for thought, watch some clips from a TV show segment called Talking to Americans (from *This Hour Has 22 Minutes*), in which Rick Mercer travelled to American cities and interviewed Americans about Canadian politics, geography, and weather. You will find some of these clips at **www.youtube.com.**

7. A poll conducted for Canwest News Service and Global Nation reported that 26% of Canadians either have made or plan to make changes in how their retirement funds are invested this year. (*Source:* "Most Canadians Holding Steady on Retirement Investments: Poll," *Canada.com*, **www.canada.com/Most+Canadians+holding+steady+retirement+investments+Poll/1324814/story.html**, accessed February 25, 2009). The poll was conducted through an online survey of 1,003 Canadians. Why were only 1,003 Canadians surveyed? Are you surprised at the relatively low sample size, given the size of the Canadian population?

8. The Top 100 Employers list is compiled annually by Mediacorp Canada Inc. Mediacorp invites several thousand Canadian employers to complete an extensive application, and about a thousand apply. On the basis of the applications Mediacorp rates the employers, and publishes a book profiling the top 100. Do you think that this process is an effective way to identify Canada's "top 100" employers? You can find more details at **www.canadastop100.com/research.html.**

9. A BBC News article is titled "Short Workers Lose Small Fortune." The article says that each inch of height added USD $789 to annual pay. Professor Tim Judge from the University of Florida, who led the study described in the article, suggested that the reason for this was evolution. He is quoted as saying: "When humans evolved as a species and still lived in the jungles or the plain, they ascribed leader-like qualities to tall people because they thought they would be better able to protect them." What conclusions can be justified from this kind of a study, assuming it was done in a statistically correct manner? (*Source:* "Short Workers Lose Small Fortune," *BBC News,* **http://news.bbc.co.uk/2/hi/health/3200296.stm**, accessed February 25, 2009).

10. Your school has just approved a new policy that will make the campus entirely smoke-free. This means no smoking on the property at any time, starting at the beginning of the next semester. Because you smoke a pack a day, you are alarmed. You collect a sample of opinions about the new policy from students and staff near one of the designated smoking areas (where you spend a lot of time). You find that there are many people who share your alarm. You produce some graphs and tables, and write a report about your findings, which you take to the president of the school. She listens to your tale of woe but then tells you that your study is not representative of the views of all of the members of the school community. Where did you go wrong?

11. Many companies are building customer databases through loyalty programs. One example is the Optimum program of Shoppers Drug Mart Corporation. Customers collect Optimum points when they spend money at the stores. Each purchase is recorded, and so the company collects information about customer demographics and shopping patterns. These data can be used in a variety of ways to target marketing efforts. For example, flyer distribution can be limited to neighbourhoods where the flyers are most likely to have an impact. Do some research to identify other customer loyalty programs that allow companies to collect customer information.

12. An article published in September 2006 was called "No Booze? You May Lose: Why Drinkers Earn More Money Than Nondrinkers." (*Source:* "No Booze? You May Lose," Reason Foundation, **www.reason.org/pb44.pdf**, accessed February 25, 2009). Researchers reported that drinkers earn 10–14% more than people who do not drink. In an attempt to explain why this might be the case, the study suggests that drinkers are more social than non-drinkers, and that greater social networks lead to higher earnings. What kind of a study was this? Is it possible to draw a strong conclusion that drinking causes higher earnings on the basis of such a study?

13. In the early 1980s, Coca-Cola spent millions of dollars on taste tests and interviews with consumers to test a new flavour of Coke. "The data was unequivocal: Consumers preferred the new formula 8% more than Pepsi and an astonishing 20% more than the original Coca-Cola recipe." (*Source:* Ken Hunt, "Brand Surgery," *Report on Business,* November 2008, p. 36). Despite this, when the New Coke was launched, it was a disaster, because people were very loyal to the original Coke. Describe a question that researchers could have asked that would have indicated this potential problem.

14. In a study produced by the Conference Board of Canada about the impact of overtime on employee health, there is a section called "About the Study Data." This section describes in some detail the database on which the study relied. Why would the Conference Board include this kind of detail in their report?

15. A Statistics Canada study on commuting patterns in Canada reveals that the proportion of workers using a car to get to work fell from 80.7% in 2001 to 80% in 2006 (*Source:* Statistics

Canada, "Commuting Patterns and Places of Work of Canadians, 2006 Census," Catalogue no. 97-561, p. 10, **www12.statcan.gc.ca/english/census06/analysis/pow/pdf/97-561-XIE2006001.pdf**, accessed February 25, 2009). The report also noted than in 2006, more than 10 million Canadian workers drove their car to work, which is 714,900 more drivers than in 2001. These data were presented in a section of the report titled "The proportion of drivers is decreasing." Do you think this title accurately reflects the change in commuting patterns from 2001 to 2006?

THINK AND DECIDE USING EXCEL

16. Lotto 6/49 is a lottery in which players select six numbers from 1 to 49. Prizes vary according to how closely players' selections match the winning set of numbers. Many players do not actually select their six numbers; they rely instead on a "quick pick," where the numbers are selected (presumably randomly) for them. Use Excel to generate a quick pick for Lotto 6/49. Check to see if the numbers you selected would have won in the most recent draw (you can check at **www.olg.ca/lotteries/games/howtoplay.do?game=lotto649**). It is not recommended that you actually play the lottery. The probability of winning is quite low.

CRE01-17

17. A data set of marks is available in the file called CRE01-17. Select 10 random samples of size 10 from this data set. Using the **=AVERAGE** function in Excel, calculate the average mark from the population data, and calculate the sample average from each sample. Are the sample averages fairly close to the true population average?

18. Repeat exercise 17, but this time select 10 random samples of size 15 from the marks data set. Compare the sample averages you get from these 10 samples with the sample averages you got in exercise 17. Which averages are closer to the true population average? What do you expect the answer should be, and why?

19. Calculate the average of the sample averages from exercise 17. Is this average close to the population average? Calculate the average of the sample averages from exercise 18. Is this average close to the population average? Is it closer?

TEST YOUR KNOWLEDGE

20. The government of Ontario began collecting data on "key performance indicators" (KPI) for colleges in 1998. Data on graduate employment rates, graduate satisfaction, and employer satisfaction are collected by external service providers through telephone surveys (some 40,000 graduates, 10,000 employers, and 90,000 current students are surveyed). The data are used as a basis for distributing some of the colleges' funding.

a. What are some of the difficulties that would be involved in collecting and interpreting these data?

b. For the 1999–2000 year, 59% of the students surveyed (overall) indicated that they were very satisfied or satisfied with college services. In the 2000–2001 year, 66.8% of the students surveyed indicated that they were very satisfied or satisfied. This percentage has been above 66% for the entire period from 2000 to 2007 (the most recent data available at the time of publication). Do you think this means that the colleges all improved their services significantly between 1999–2000 and the following school year? (*Source:* Colleges Ontario, "Key Performance Indicators," **www.collegesontario.org/outcomes/key-performance-indicators.html**, accessed February 25, 2009).

c. Go to the website mentioned in part b, and download the document called "2005 KPI Trends." What techniques are used to summarize the trends in key performance indicators?

d. In the section on the "KPI Overview," in the document you downloaded for part c, read the section on the student satisfaction rate. What difficulties are there in comparing the data from 1999–2000 and the data from the subsequent years?