CHAPTER 1

An Introduction to Business Statistics

Learning Objectives

When you have mastered the material in this chapter, you will be able to:

- **LO1-1** Define a variable.
- **LO1-2** Describe the difference between a quantitative variable and a qualitative variable.
- **LO1-3** Describe the difference between cross-sectional data and time series data.
- **LO1-4** Construct and interpret a time series (runs) plot.
- **LO1-5** Identify the different types of data sources: existing data sources, experimental studies, and observational studies.
- **LO1-6** Describe the difference between a population and a sample.
- **LO1-7** Distinguish between descriptive statistics and statistical inference.
- **LO1-8** Explain the importance of random sampling.
- **LO1-9** Identify the ratio, interval, ordinal, and nominative scales of measurement (Optional).

Chapter Outline

1.1 Data

1.2 Data Sources

1.3 Populations and Samples

1.4 Three Case Studies That Illustrate Sampling and Statistical Inference

1.5 Ratio, Interval, Ordinal, and Nominative Scales of Measurement (Optional)

The subject of **statistics** involves the study of how to collect, analyze, and interpret data. **Data** are facts and figures from which conclusions can be drawn. Such conclusions are important to the decision making of many professions and organizations. For example, **economists** use conclusions drawn from the latest data on unemployment and inflation to help the government make policy decisions. **Financial planners** use recent trends in stock market prices and economic conditions to make investment decisions. **Accountants** use **sample data** concerning a company's *actual sales revenues* to assess whether the company's *claimed sales revenues* are valid. **Marketing professionals** help businesses decide which products to develop and market by using data that reveal consumer preferences. **Production supervisors** use manufacturing data to evaluate, control, and improve product quality. **Politicians** rely on data from public opinion polls to formulate legislation and to devise campaign strategies. **Physicians and hospitals** use data on the effectiveness of drugs and surgical procedures to provide patients with the best possible treatment.

In this chapter we begin to see how we collect and analyze data. As we proceed through the chapter, we introduce several case studies. These case studies (and others to be introduced later) are revisited throughout later chapters as we learn the statistical methods needed to analyze them. Briefly, we will begin to study three cases:

C

The Cell Phone Case. A bank estimates its cellular phone costs and decides whether to outsource management of its wireless resources by studying the calling patterns of its employees.

The Marketing Research Case. A bottling company investigates consumer reaction to a new bottle design for one of its popular soft drinks.

The Car Mileage Case. To determine if it qualifies for a federal tax credit based on fuel economy, an automaker studies the gas mileage of its new midsize model.

1.1 Data

LO1-1 Define a variable.

Data sets, elements, and variables We have said that data are facts and figures from which conclusions can be drawn. Together, the data that are collected for a particular study are referred to as a **data set.** For example, Table 1.1 is a data set that gives information about the new homes sold in a Florida luxury home development over a recent three-month period. Potential buyers in this housing community could choose either the "Diamond" or the "Ruby" home model design and could have the home built on either a lake lot or a treed lot (with no water access).

In order to understand the data in Table 1.1, note that any data set provides information about some group of individual **elements,** which may be people, objects, events, or other entities. The information that a data set provides about its elements usually describes one or more characteristics of these elements.

Any characteristic of an element is called a **variable.**

For the data set in Table 1.1, each sold home is an element, and four variables are used to describe the homes. These variables are (1) the home model design, (2) the type of lot on which the home was built, (3) the list (asking) price, and (4) the (actual) selling price. Moreover, each home model design came with "everything included"—specifically, a complete, luxury interior package and a choice of one of three different architectural exteriors. Therefore, because there were no interior or exterior options to purchase, the (actual) selling price of a home depended solely on the home model design and whatever price reduction (based partially on the lot type) that the community developer (builder) was willing to give.

TABLE 1.1 A Data Set Describing Five Home Sales DS HomeSales

Home	Model Design	Lot Type	List Price	Selling Price
1	Diamond	Lake	$494,000	$494,000
2	Ruby	Treed	$447,000	$398,000
3	Diamond	Treed	$494,000	$440,000
4	Diamond	Treed	$494,000	$469,000
5	Ruby	Lake	$447,000	$447,000

The data in Table 1.1 are real (with some minor modifications to protect privacy) and were provided by a business executive—a friend of the authors—who recently received a promotion and needed to move to central Florida. While searching for a new home, the executive and his family visited the luxury home community and decided they wanted to purchase a Diamond model on a treed lot. The list price of this home was $494,000, but the developer offered to sell it for an "incentive" price of $469,000. Intuitively, the incentive price's $25,000 savings off list price seemed like a good deal. However, the executive resisted making an immediate decision. Instead, he decided to collect data on the selling prices of new homes recently sold in the community and use the data to assess whether the developer might be amenable to a lower offer. In order to collect "relevant data," the executive talked to local real estate professionals and learned that new homes sold in the community during the previous three months were a good indicator of current home value. Using real estate sales records, the executive also learned that five of the community's new homes had sold in the previous three months. The data given in Table 1.1 are the data that the executive collected about these five homes.

LO1-2 Describe the difference between a quantitative variable and a qualitative variable.

In order to understand the conclusions the business executive reached using the data in Table 1.1, we need to further discuss variables. For any variable describing an element in a data set, we carry out a **measurement** to assign a value of the variable to the element. For example, in the real estate example, real estate sales records gave the actual selling price of each home to the nearest dollar. In another example, a credit card company might measure the time it takes for a cardholder's bill to be paid to the nearest day. Or, in a third example, an automaker might measure the gasoline mileage obtained by a car in city driving to the nearest one-tenth of a mile per gallon by conducting a mileage test on a driving course prescribed by the Environmental Protection Agency (EPA). If the possible measurements of the values of a variable are numbers that represent quantities (that is, "how much" or "how many"), then the variable is said to be **quantitative.** For example, the actual selling price of a home, the payment time of a bill, and the gasoline mileage of a car are all quantitative. However, if we simply record into which of several categories an element falls, then the variable is said to be **qualitative** or **categorical.** Examples of categorical variables include (1) a person's gender, (2) the make of an automobile, (3) whether a person who purchases a product is satisfied with the product, and (4) the type of lot on which a home is built.[1]

Of the four variables in Table 1.1, two variables—list price and selling price—are quantitative, and two variables—model design and lot type—are qualitative. Furthermore, when the business executive examined Table 1.1, he noted that homes on lake lots had sold at their list price, but homes on treed lots had not. Because the executive and his family wished to purchase a Diamond model on a treed lot, the executive also noted that two Diamond models on treed lots had sold in the previous three months. One of these Diamond models had sold for the incentive price of $469,000, but the other had sold for a lower price of $440,000. Hoping to pay the lower price for his family's new home, the executive offered $440,000 for the Diamond model on the treed lot. Initially, the home builder turned down this offer, but two days later the builder called back and accepted the offer. The executive had used data to buy the new home for $54,000 less than the list price and $29,000 less than the incentive price!

LO1-3 Describe the difference between cross-sectional data and time series data.

LO1-4 Construct and interpret a time series (runs) plot.

Cross-sectional and time series data Some statistical techniques are used to analyze *cross-sectional data,* while others are used to analyze *time series data.* **Cross-sectional data** are data collected at the same or approximately the same point in time. For example, suppose that a bank wishes to analyze last month's cell phone bills for its employees. Then, because the cell phone costs given by these bills are for different employees in the same month, the cell phone costs are cross-sectional data. **Time series data** are data collected over different time periods. For example, Table 1.2 presents the average basic cable television rate in the United States for each of the years 1999 to 2009. Figure 1.1 is a **time series plot**—also called a **runs plot**—of these data. Here we plot each television rate on the vertical scale versus its corresponding time index on the horizontal scale. For instance, the first cable rate ($28.92) is plotted versus 1999, the second cable rate ($30.37) is plotted versus 2000, and so forth. Examining the time series plot, we see that the cable rates increased substantially from 1999 to 2009. Finally, because the five homes in Table 1.1 were sold over a three-month period that represented a relatively stable real estate market, we can consider the data in Table 1.1 to essentially be cross-sectional data.

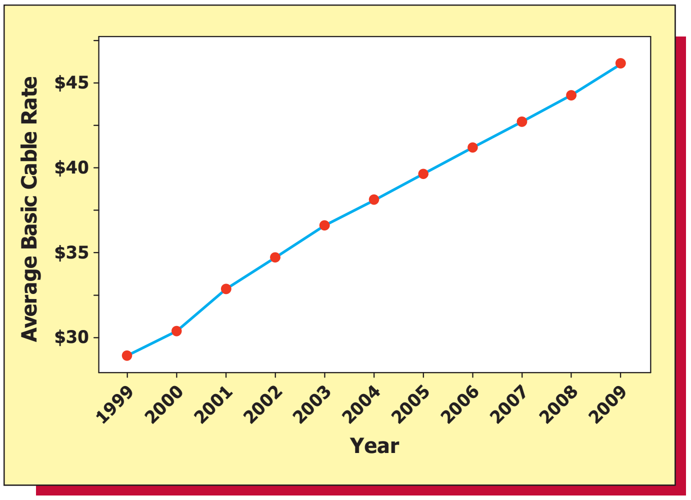
[1]Optional Section 1.5 discusses two types of quantitative variables (ratio and interval) and two types of qualitative variables (ordinal and nominative).

TABLE 1.2 The Average Basic Cable Rates in the U.S. from 1999 to 2009
DS BasicCable

Year	Average Basic Cable Rate
1999	$ 28.92
2000	30.37
2001	32.87
2002	34.71
2003	36.59
2004	38.14
2005	39.63
2006	41.17
2007	42.72
2008	44.28
2009	46.13

Source: U.S. Energy Information Administration, http://www.eia.gov/

FIGURE 1.1 Time Series Plot of the Average Basic Cable Rates in the U.S. from 1999 to 2009
DS BasicCable



1.2 Data Sources

Data can be obtained from *existing sources* or from **experimental and observational studies.**

LO1-5 Identify the different types of data sources: existing data sources, experimental studies, and observational studies.

Existing sources Sometimes we can use data *already gathered* by public or private sources. The Internet is an obvious place to search for electronic versions of government publications, company reports, and business journals, but there is also a wealth of information available in the reference section of a good library or in county courthouse records.

If a business needs information about incomes in the Northeastern states, a natural source is the US Census Bureau's website at http://www.census.gov. By following various links posted on the homepage, you can find income and demographic data for specific regions of the country. Other useful websites for economic and financial data are listed in Table 1.3. All of these are trustworthy sources.

TABLE 1.3 Examples of Public Economic and Financial Data Sites

Title	Website	Data Type
Global Financial Data	https://www.globalfinancialdata.com/index.html	Annual data on stock markets, inflation rates, interest rates, exchange rates, etc.
National Bureau of Economic Research Macrohistory Database	http://www.nber.org/databases/macrohistory/contents/index.html	Historic data on production, construction, employment, money, prices, asset market transactions, foreign trade, and government activity
Federal Reserve Economic Data	http://research.stlouisfed.org/fred2/	Historical U.S. economic and financial data, including daily U.S. interest rates, monetary and business indicators, exchange rate, balance of payments, and regional economic data
Bureau of Labor Statistics	http://stats.bls.gov/	Data concerning employment, inflation, consumer spending, productivity, safety, labor demographics, and the like.
WebEc Economics Data	http://netec.wustl.edu/WebEc/	One of the best complete economics data links including both international and domestic data categorized by area and country
Economic Statistics Briefing Room	http://clinton2.nara.gov/fsbr/esbr.htm	Links to the most current available values of federal economic indicators in 8 categories

Source: Prepared by Lan Ma and Jeffrey S. Simonoff. The authors provide no warranty as to the accuracy of the information provided.

However, given the ease with which anyone can post documents, pictures, weblogs, and videos on the World Wide Web, not all sites are equally reliable. If we were to use a search engine from Google, Netscape, Yahoo, Ask.com, or AltaVista (just to name a few) to find information about the price of a two-bedroom apartment in Manhattan, we would be inundated by millions of "hits." (In fact, a recent search on Google using the keywords "price 2 bedroom apartments Manhattan" yielded 1,040,000 sites.) Some of the sources will be more useful, exhaustive, and error-free than others. Fortunately, the search engines prioritize the lists and provide the most relevant and highly used sites first.

Obviously, performing such web searches costs next to nothing and takes relatively little time, but the tradeoff is that we are also limited in terms of the type of information we are able to find. Another option may be to use a private data source. Most companies keep employee records, for example, and retail establishments retain information about their customers, products, and advertising results. Manufacturing companies may collect information about their processes and defect propagation in order to monitor quality. If we have no affiliation with these companies, however, these data may be more difficult to obtain.

Another alternative would be to contact a data collection agency, which typically incurs some kind of cost. You can either buy subscriptions or purchase individual company financial reports from agencies like Dun & Bradstreet, Bloomberg, Dow Jones & Company, Travel Industry of America, Graduate Management Admission Council, and the Educational Testing Service. If you need to collect specific information, some companies, such as ACNielsen and Information Resources, Inc., can be hired to collect the information for a fee.

Experimental and observational studies There are many instances when the data we need are not readily available from a public or private source. The data might not have been collected at all or they may have been collected in a statistically unsound manner. In cases like these, we need to collect the data ourselves. Suppose we work for a soft drink producer and want to assess consumer reactions to a new bottled water. Since the water has not been marketed yet, we may choose to conduct taste tests, focus groups, or some other market research. Projecting political election results also requires information that is not readily available. In this case, exit polls and telephone surveys are commonly used to obtain the information needed to predict voting trends. New drugs for fighting disease are tested by collecting data under carefully controlled and monitored experimental conditions. In many marketing, political, and medical situations of these sorts, companies hire outside consultants or statisticians to help them obtain appropriate data. Regardless of whether newly minted data are gathered in-house or by paid outsiders, this type of data collection requires much more time, effort, and expense than are needed when data can be found from public or private sources.

When initiating a study, we first define our variable of interest, or **response variable.** Other variables, typically called **factors,** that may be related to the response variable of interest will also be measured. When we are able to set or manipulate the values of these factors, we have an **experimental study.** For example, a pharmaceutical company might wish to determine the most appropriate daily dose of a cholesterol-lowering drug for patients having cholesterol levels over 240 mg/dL, a level associated with a high risk of coronary disease. (http://www.heart.org/HEARTORG/Conditions/Cholesterol/AboutCholesterol/What-Your-Cholesterol-Levels-Mean_UCM_305562_Article.jsp) The company can perform an experiment in which one sample of patients receives a placebo; a second sample receives some low dose; a third a higher dose; and so forth. This is an experiment because the company controls the amount of drug each group receives. The optimal daily dose can be determined by analyzing the patients' responses to the different dosage levels given.

When analysts are unable to control the factors of interest, the study is **observational.** In studies of diet and cholesterol, patients' diets are not under the analyst's control. Patients are often unwilling or unable to follow prescribed diets; doctors might simply ask patients what they eat and then look for associations between the factor *diet* and the response variable *cholesterol*.

Asking people what they eat is an example of performing a **survey.** In general, people in a survey are asked questions about their behaviors, opinions, beliefs, and other characteristics. For instance, shoppers at a mall might be asked to fill out a short questionnaire which seeks their opinions about a new bottled water. In other observational studies, we might simply observe the behavior of people. For example, we might observe the behavior of shoppers as they look at a store display, or we might observe the interactions between students and teachers.

Exercises for Sections 1.1 and 1.2

CONCEPTS

connect

1.1 Define what we mean by a *variable,* and explain the difference between a quantitative variable and a qualitative (categorical) variable.

1.2 Below we list several variables. Which of these variables are quantitative and which are qualitative? Explain.
- **a** The dollar amount on an accounts receivable invoice.
- **b** The net profit for a company in 2009.
- **c** The stock exchange on which a company's stock is traded.
- **d** The national debt of the United States in 2009.
- **e** The advertising medium (radio, television, or print) used to promote a product.

1.3 Discuss the difference between cross-sectional data and time series data. If we record the total number of cars sold in 2011 by each of 10 car salespeople, are the data cross-sectional or time series data? If we record the total number of cars sold by a particular car salesperson in each of the years 2007, 2008, 2009, 2010, and 2011, are the data cross-sectional or time series data?

1.4 Consider a medical study that is being performed to test the effect of smoking on lung cancer. Two groups of subjects are identified; one group has lung cancer and the other one doesn't. Both are asked to fill out a questionnaire containing questions about their age, sex, occupation, and number of cigarettes smoked per day. What is the response variable? Which are the factors? What type of study is this (experimental or observational)?

METHODS AND APPLICATIONS

1.5 Consider the five homes in Table 1.1 (page 3). What do you think you would have to pay for a Ruby model on a treed lot?

1.6 Consider the five homes in Table 1.1 (page 3). What do you think you would have to pay for a Diamond model on a lake lot? For a Ruby model on a lake lot?

1.7 The number of Bismark X-12 electronic calculators sold at Smith's Department Stores over the past 24 months have been: 197, 211, 203, 247, 239, 269, 308, 262, 258, 256, 261, 288, 296, 276, 305, 308, 356, 393, 363, 386, 443, 308, 358, and 384. Make a time series plot of these data. That is, plot 197 versus month 1, 211 versus month 2, and so forth. What does the time series plot tell you? DS CalcSale

1.3 Populations and Samples

LO1-6 Describe the difference between a population and a sample.

We often collect data in order to study a population.

A **population** is the set of all elements about which we wish to draw conclusions.

Examples of populations include (1) all of last year's graduates of Dartmouth College's Master of Business Administration program, (2) all current MasterCard cardholders, and (3) all Buick LaCrosses that have been or will be produced this year.

We usually focus on studying one or more variables describing the population elements. If we carry out a measurement to assign a value of a variable to each and every population element, we have a *population of measurements* (sometimes called *observations*). If the population is small, it is reasonable to do this. For instance, if 150 students graduated last year from the Dartmouth College MBA program, it might be feasible to survey the graduates and to record all of their starting salaries. In general:

If we examine all of the population measurements, we say that we are conducting a **census** of the population.

Often the population that we wish to study is very large, and it is too time-consuming or costly to conduct a census. In such a situation, we select and analyze a subset (or portion) of the population elements.

A **sample** is a subset of the elements of a population.

For example, suppose that 8,742 students graduated last year from a large state university. It would probably be too time-consuming to take a census of the population of all of their starting salaries. Therefore, we would select a sample of graduates, and we would obtain and record their starting salaries. When we measure a characteristic of the elements in a sample, we have a **sample of measurements.**

We often wish to describe a population or sample.

LO1-7 Distinguish between descriptive statistics and statistical inference.

Descriptive statistics is the science of describing the important aspects of a set of measurements.

As an example, if we are studying a set of starting salaries, we might wish to describe (1) how large or small they tend to be, (2) what a typical salary might be, and (3) how much the salaries differ from each other.

When the population of interest is small and we can conduct a census of the population, we will be able to directly describe the important aspects of the population measurements. However, if the population is large and we need to select a sample from it, then we use what we call **statistical inference.**

Statistical inference is the science of using a sample of measurements to make generalizations about the important aspects of a population of measurements.

For instance, we might use a sample of starting salaries to **estimate** the important aspects of a population of starting salaries. In the next section, we begin to look at how statistical inference is carried out.

LO1-8 Explain the importance of random sampling.

1.4 Three Case Studies That Illustrate Sampling and Statistical Inference

Random samples When we select a sample from a population, we hope that the information contained in the sample reflects what is true about the population. One of the best ways to achieve this goal is to select a *random sample*. In Section 7.1 we will precisely define a random sample.[2] For now, it suffices to know that one intuitive way to select a random sample would begin by placing numbered slips of paper representing the population elements in a suitable container. We would thoroughly mix the slips of paper and (blindfolded) choose slips of paper from the container. The numbers on the chosen slips of paper would identify the randomly selected population elements that make up the random sample. In Section 7.1 we will discuss more practical methods for selecting a random sample. We will also see that, although in many situations it is not possible to select a sample that is exactly random, we can sometimes select a sample that is approximately random.

We now introduce three case studies that illustrate the need for a random (or approximately random) sample and the use of such a sample in making statistical inferences. After studying these cases, the reader has the option of studying Section 7.1 (see page 267) and learning practical ways to select random and approximately random samples.

EXAMPLE 1.1 The Cell Phone Case: Reducing Cellular Phone Costs C

Part 1: The cost of company cell phone use Rising cell phone costs have forced companies having large numbers of cellular users to hire services to manage their cellular and other wireless resources. These cellular management services use sophisticated software and mathematical models to choose cost-efficient cell phone plans for their clients. One such firm, mindWireless of Austin, Texas, specializes in automated wireless cost management. According to Kevin Whitehurst, co-founder of mindWireless, cell phone carriers count on *overage*—using more minutes than one's plan allows—and *underage*—using fewer minutes than those already paid for—to deliver almost half of their revenues.[3] As a result, a company's typical cost of cell phone use can be excessive—18 cents per minute or more. However, Mr. Whitehurst explains that by using mindWireless automated cost management to select calling plans, this cost can be reduced to 12 cents per minute or less.

In this case we consider a bank that wishes to decide whether to hire a cellular management service to choose its employees' calling plans. While the bank has over 10,000 employees on

[2] Actually, there are several different kinds of random samples. The type we will define is sometimes called a *simple random sample*. For brevity's sake, however, we will use the term *random sample*.

[3] The authors would like to thank Kevin Whitehurst for help in developing this case.

TABLE 1.4 A Sample of Cellular Usages (in minutes) for 100 Randomly Selected Employees
DS CellUse

75	485	37	547	753	93	897	694	797	477
654	578	504	670	490	225	509	247	597	173
496	553	0	198	507	157	672	296	774	479
0	822	705	814	20	513	546	801	721	273
879	433	420	521	648	41	528	359	367	948
511	704	535	585	341	530	216	512	491	0
542	562	49	505	461	496	241	624	885	259
571	338	503	529	737	444	372	555	290	830
719	120	468	730	853	18	479	144	24	513
482	683	212	418	399	376	323	173	669	611

many different types of calling plans, a cellular management service suggests that by studying the calling patterns of cellular users on 500-minute-per-month plans, the bank can accurately assess whether its cell phone costs can be substantially reduced. The bank has 2,136 employees on a variety of 500-minute-per-month plans with different basic monthly rates, different overage charges, and different additional charges for long distance and roaming. It would be extremely time consuming to analyze in detail the cell phone bills of all 2,136 employees. Therefore, the bank will estimate its cellular costs for the 500-minute plans by analyzing last month's cell phone bills for a *random sample* of 100 employees on these plans.[4]

Part 2: A random sample When the random sample of 100 employees is chosen, the number of cellular minutes used by each sampled employee during last month (the employee's *cellular usage*) is found and recorded. The 100 cellular-usage figures are given in Table 1.4. Looking at this table, we can see that there is substantial overage and underage—many employees used far more than 500 minutes, while many others failed to use all of the 500 minutes allowed by their plan. In Chapter 3 we will use these 100 usage figures to estimate the bank's cellular costs and decide whether the bank should hire a cellular management service.

EXAMPLE 1.2 The Marketing Research Case: Rating a Bottle Design

C

Part 1: Rating a bottle design The design of a package or bottle can have an important effect on a company's bottom line. In this case a brand group wishes to research consumer reaction to a new bottle design for a popular soft drink. To do this, the brand group will show consumers the new bottle and ask them to rate the bottle image. For each consumer interviewed, a bottle image **composite score** will be found by adding the consumer's numerical responses to the five questions shown in Figure 1.2. It follows that the minimum possible bottle image composite

FIGURE 1.2 The Bottle Design Survey Instrument

Please circle the response that most accurately describes whether you agree or disagree with each statement about the bottle you have examined.

Statement	Strongly Disagree						Strongly Agree
The size of this bottle is convenient.	1	2	3	4	5	6	7
The contoured shape of this bottle is easy to handle.	1	2	3	4	5	6	7
The label on this bottle is easy to read.	1	2	3	4	5	6	7
This bottle is easy to open.	1	2	3	4	5	6	7
Based on its overall appeal, I like this bottle design.	1	2	3	4	5	6	7

[4]In Chapter 8 we will discuss how to plan the *sample size*—the number of elements (for example, 100) that should be included in a sample. Throughout this book we will take large enough samples to allow us to make reasonably accurate statistical inferences.

TABLE 1.5 A Sample of Bottle Design Ratings (Composite Scores for a Sample of 60 Shoppers)
DS Design

34	33	33	29	26	33	28	25	32	33
32	25	27	33	22	27	32	33	32	29
24	30	20	34	31	32	30	35	33	31
32	28	30	31	31	33	29	27	34	31
31	28	33	31	32	28	26	29	32	34
32	30	34	32	30	30	32	31	29	33

score is 5 (resulting from a response of 1 on all five questions) and the maximum possible bottle image composite score is 35 (resulting from a response of 7 on all five questions). Furthermore, experience has shown that the smallest acceptable bottle image composite score for a successful bottle design is 25.

Part 2: An approximately random sample Because it is impossible to show the new bottle to "all consumers," the brand group will use the *mall intercept method* to select a sample of consumers. This method chooses a mall and a sampling time so that shoppers at the mall during the sampling time are a representative cross-section of all consumers. Then, shoppers are intercepted as they walk past a designated location in such a way that an approximately random sample of shoppers at the mall is selected. When the brand group uses this mall intercept method to interview a sample of 60 shoppers at a mall on a particular Saturday, the 60 bottle image composite scores in Table 1.5 are obtained. Because these scores vary from a minimum of 20 to a maximum of 35, we might infer that *most* consumers would rate the new bottle design between 20 and 35. Furthermore, 57 of the 60 composite scores are at least 25. Therefore, we might estimate that a proportion of $57/60 = .95$ (that is, 95 percent) of all consumers would give the bottle design a composite score of at least 25. In future chapters we will further analyze the composite scores.

Processes Sometimes we are interested in studying the population of all of the elements that will be or could potentially be produced by a *process*.

A **process** is a sequence of operations that takes inputs (labor, materials, methods, machines, and so on) and turns them into outputs (products, services, and the like).

Processes produce output *over time*. For example, this year's Buick LaCrosse manufacturing process produces LaCrosses over time. Early in the model year, General Motors might wish to study the population of the city driving mileages of all Buick LaCrosses that will be produced during the model year. Or, even more hypothetically, General Motors might wish to study the population of the city driving mileages of all LaCrosses that could *potentially* be produced by this model year's manufacturing process. The first population is called a **finite population** because only a finite number of cars will be produced during the year. The second population is called an **infinite population** because the manufacturing process that produces this year's model could in theory always be used to build "one more car." That is, theoretically there is no limit to the number of cars that could be produced by this year's process. There are a multitude of other examples of finite or infinite hypothetical populations. For instance, we might study the population of all waiting times that will or could potentially be experienced by patients of a hospital emergency room. Or we might study the population of all the amounts of grape jelly that will be or could potentially be dispensed into 16-ounce jars by an automated filling machine. To study a population of potential process observations, we sample the process—often at equally spaced time points—over time.

EXAMPLE 1.3 The Car Mileage Case: Estimating Mileage

C

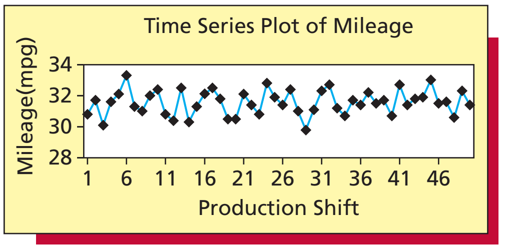
Part 1: Auto fuel economy Personal budgets, national energy security, and the global environment are all affected by our gasoline consumption. Hybrid and electric cars are a vital part of a long-term strategy to reduce our nation's gasoline consumption. However, until use of these cars is

TABLE 1.6 A Sample of 50 Mileages DS GasMiles

30.8	30.8	32.1	32.3	32.7
31.7	30.4	31.4	32.7	31.4
30.1	32.5	30.8	31.2	31.8
31.6	30.3	32.8	30.7	31.9
32.1	31.3	31.9	31.7	33.0
33.3	32.1	31.4	31.4	31.5
31.3	32.5	32.4	32.2	31.6
31.0	31.8	31.0	31.5	30.6
32.0	30.5	29.8	31.7	32.3
32.4	30.5	31.1	30.7	31.4

Note: Time order is given by reading down the columns from left to right.

FIGURE 1.3 A Time Series Plot of the 50 Mileages



more widespread and affordable, the most effective way to conserve gasoline is to design gasoline powered cars that are more fuel efficient.[5] In the short term, "that will give you the biggest bang for your buck," says David Friedman, research director of the Union of Concerned Scientists' Clean Vehicle Program.[6]

In this case study we consider a tax credit offered by the federal government to automakers for improving the fuel economy of gasoline powered midsize cars. According to *The Fuel Economy Guide—2012 Model Year,* virtually every gasoline powered midsize car equipped with an automatic transmission has an EPA combined city and highway mileage estimate of 26 miles per gallon (mpg) or less.[7] Furthermore, the EPA has concluded that a 5 mpg increase in fuel economy is significant and feasible.[8] Therefore, suppose that the government has decided to offer the tax credit to any automaker selling a midsize model with an automatic transmission that achieves an EPA combined city and highway mileage estimate of at least 31 mpg.

Part 2: Sampling a process Consider an automaker that has recently introduced a new midsize model with an automatic transmission and wishes to demonstrate that this new model qualifies for the tax credit. In order to study the population of all cars of this type that will or could potentially be produced, the automaker will choose a sample of 50 of these cars. The manufacturer's production operation runs 8 hour shifts, with 100 midsize cars produced on each shift. When the production process has been fine tuned and all start-up problems have been identified and corrected, the automaker will select one car at random from each of 50 consecutive production shifts. Once selected, each car is to be subjected to an EPA test that determines the EPA combined city and highway mileage of the car.

Suppose that when the 50 cars are selected and tested, the sample of 50 EPA combined mileages shown in Table 1.6 is obtained. A time series plot of the mileages is given in Figure 1.3. Examining this plot, we see that, although the mileages vary over time, they do not seem to vary in any unusual way. For example, the mileages do not tend to either decrease or increase (as did the basic cable rates in Figure 1.1) over time. This intuitively verifies that the midsize car manufacturing process is producing consistent car mileages over time, and thus we can regard the 50 mileages as an approximately random sample that can be used to make statistical inferences about the population of all possible midsize car mileages. Therefore, because the 50 mileages vary from a minimum of 29.8 mpg to a maximum of 33.3 mpg, we might conclude that most midsize cars produced by the manufacturing process will obtain between 29.8 mpg and 33.3 mpg. Moreover, because 38 out of the 50 mileages—or 76 percent of the mileages—are greater than or equal to the tax credit standard of 31 mpg, we have some evidence that the "typical car" produced by the process will meet or exceed the tax credit standard. We will further evaluate this evidence in later chapters.

[5, 6]Bryan Walsh, "Plugged In," *Time*, September 29, 2008 (see page 56).

[7]The "26 miles per gallon (mpg) or less" figure relates to midsize cars with an automatic transmission *and* at least a 4 cylinder, 2.4 liter engine (such cars are the most popular midsize models). Therefore, when we refer to a midsize car with an automatic transmission in future discussions, we are assuming that the midsize car also has at least a 4 cylinder, 2.4 liter engine.

[8]The authors wish to thank Jeff Alson of the EPA for this information.

Ethical guidelines for statistical practice The American Statistical Association, the leading U.S. professional statistical association, has developed the report "Ethical Guidelines for Statistical Practice."[9] This report provides information that helps statistical practitioners to consistently use ethical statistical practices and that helps users of statistical information avoid being misled by unethical statistical practices. Unethical statistical practices can take a variety of forms, including:

- **Improper sampling** Purposely selecting a biased sample—for example, using a nonrandom sampling procedure that overrepresents population elements supporting a desired conclusion or that underrepresents population elements not supporting the desired conclusion—is unethical. In addition, discarding already sampled population elements that do not support the desired conclusion is unethical. More will be said about proper and improper sampling in Chapter 7.
- **Misleading charts, graphs, and descriptive measures** In Section 2.7, we will present an example of how misleading charts and graphs can distort the perception of changes in salaries over time. Using misleading charts or graphs to make the salary changes seem much larger or much smaller than they really are is unethical. In Section 3.1, we will present an example illustrating that many populations of individual or household incomes contain a small percentage of very high incomes. These very high incomes make the *population mean income* substantially larger than the *population median income*. In this situation we will see that the population median income is a better measure of the typical income in the population. Using the population mean income to give an inflated perception of the typical income in the population is unethical.
- **Inappropriate statistical analysis or inappropriate interpretation of statistical results** The American Statistical Association report emphasizes that selecting many different samples and running many different tests can eventually (by random chance alone) produce a result that makes a desired conclusion seem to be true, when the conclusion really isn't true. Therefore, continuing to sample and run tests until a desired conclusion is obtained and not reporting previously obtained results that do not support the desired conclusion is unethical. Furthermore, we should always report our sampling procedure and sample size and give an estimate of the reliability of our statistical results. Estimating this reliability will be discussed in Chapter 7 and beyond.

The above examples are just an introduction to the important topic of unethical statistical practices. The American Statistical Association report contains 67 guidelines organized into eight areas involving general professionalism and ethical responsibilities. These include responsibilities to clients, to research team colleagues, to research subjects, and to other statisticians, as well as responsibilities in publications and testimony and responsibilities of those who employ statistical practitioners.

Exercises for Sections 1.3 and 1.4

CONCEPTS

1.8 Define a *population.* Give an example of a population.

1.9 Explain the difference between a census and a sample.

1.10 Explain the term *descriptive statistics*. Explain the term *statistical inference.*

1.11 Define a process.

METHODS AND APPLICATIONS

1.12 THE VIDEO GAME SATISFACTION RATING CASE DS VideoGame

A company that produces and markets video game systems wishes to assess its customers' level of satisfaction with a relatively new model, the XYZ-Box. In the six months since the introduction of the model, the company has received 73,219 warranty registrations from purchasers. The company

[9] American Statistical Association, "Ethical Guidelines for Statistical Practice," 1999.

will select a random sample of 65 of these registrations and will conduct telephone interviews with the purchasers. Specifically, each purchaser will be asked to state his or her level of agreement with each of the seven statements listed on the survey instrument given in Figure 1.4. Here, the level of agreement for each statement is measured on a 7-point Likert scale. Purchaser satisfaction will be measured by adding the purchaser's responses to the seven statements. It follows that for each consumer the minimum composite score possible is 7 and the maximum is 49. Furthermore, experience has shown that a purchaser of a video game system is "very satisfied" if his or her composite score is at least 42. Suppose that when the 65 customers are interviewed, their composite scores are as given in Table 1.7. Using the data, estimate limits between which most of the 73,219 composite scores would fall. Also, estimate the proportion of the 73,219 composite scores that would be at least 42.

1.13 THE BANK CUSTOMER WAITING TIME CASE DS WaitTime

A bank manager has developed a new system to reduce the time customers spend waiting to be served by tellers during peak business hours. Typical waiting times during peak business hours under the current system are roughly 9 to 10 minutes. The bank manager hopes that the new system will lower typical waiting times to less than six minutes and wishes to evaluate the new system. When the new system is operating consistently over time, the bank manager decides to select a sample of 100 customers that need teller service during peak business hours. Specifically, for each of 100 peak business hours, the first customer that starts waiting for teller service at or after a randomly selected time during the hour will be chosen. In Exercise 7.5 (see page 271) we will discuss how to obtain a randomly selected time during an hour. When each customer is chosen, the number of minutes the customer spends waiting for teller service is recorded. The 100 waiting times that are observed are given in Table 1.8. Using the data, estimate limits between which the waiting times of most of the customers arriving during peak business hours would be. Also, estimate the proportion of waiting times of customers arriving during peak business hours that are less than six minutes.

FIGURE 1.4 The Video Game Satisfaction Survey Instrument

Statement	Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
The game console of the XYZ-Box is well designed.	1	2	3	4	5	6	7
The game controller of the XYZ-Box is easy to handle.	1	2	3	4	5	6	7
The XYZ-Box has high quality graphics capabilities.	1	2	3	4	5	6	7
The XYZ-Box has high quality audio capabilities.	1	2	3	4	5	6	7
The XYZ-Box serves as a complete entertainment center.	1	2	3	4	5	6	7
There is a large selection of XYZ-Box games to choose from.	1	2	3	4	5	6	7
I am totally satisfied with my XYZ-Box game system.	1	2	3	4	5	6	7

TABLE 1.7 Composite Scores for the Video Game Satisfaction Rating Case DS VideoGame

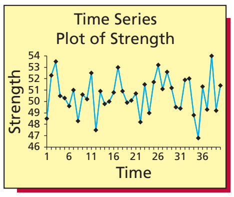
39	44	46	44	44
45	42	45	44	42
38	46	45	45	47
42	40	46	44	43
42	47	43	46	45
41	44	47	48	
38	43	43	44	
42	45	41	41	
46	45	40	45	
44	40	43	44	
40	46	44	44	
39	41	41	44	
40	43	38	46	
42	39	43	39	
45	43	36	41	

TABLE 1.8 Waiting Times (in Minutes) for the Bank Customer Waiting Time Case DS WaitTime

1.6	6.2	3.2	5.6	7.9	6.1	7.2
6.6	5.4	6.5	4.4	1.1	3.8	7.3
5.6	4.9	2.3	4.5	7.2	10.7	4.1
5.1	5.4	8.7	6.7	2.9	7.5	6.7
3.9	.8	4.7	8.1	9.1	7.0	3.5
4.6	2.5	3.6	4.3	7.7	5.3	6.3
6.5	8.3	2.7	2.2	4.0	4.5	4.3
6.4	6.1	3.7	5.8	1.4	4.5	3.8
8.6	6.3	.4	8.6	7.8	1.8	5.1
4.2	6.8	10.2	2.0	5.2	3.7	5.5
5.8	9.8	2.8	8.0	8.4	4.0	
3.4	2.9	11.6	9.5	6.3	5.7	
9.3	10.9	4.3	1.3	4.4	2.4	
7.4	4.7	3.1	4.8	5.2	9.2	
1.8	3.9	5.8	9.9	7.4	5.0	

TABLE 1.9 Trash Bag Breaking Strengths DS TrashBag

48.5	50.7
52.3	48.2
53.5	51.5
50.5	49.0
50.3	51.7
49.6	53.2
51.0	51.1
48.3	52.6
50.6	51.2
50.2	49.5
52.5	49.4
47.5	51.9
50.9	52.0
49.8	48.8
50.0	46.8
50.8	51.3
53.0	49.3
50.9	54.0
49.9	49.2
50.1	51.4



1.14 THE TRASH BAG CASE[10] DS TrashBag

A company that produces and markets trash bags has developed an improved 30-gallon bag. The new bag is produced using a specially formulated plastic that is both stronger and more biodegradable than previously used plastics, and the company wishes to evaluate the strength of this bag. The *breaking strength* of a trash bag is considered to be the amount (in pounds) of a representative trash mix that when loaded into a bag suspended in the air will cause the bag to sustain significant damage (such as ripping or tearing). The company has decided to select a sample of 40 of the new trash bags. For each of 40 consecutive hours, the first trash bag produced at or after a randomly selected time during the hour is chosen. The bag is then subjected to a *breaking strength test*. The 40 breaking strengths obtained are given in Table 1.9. Estimate limits between which the breaking strengths of most trash bags would fall. Assume that the trash bag manufacturing process is operating consistently over time.

1.5 Ratio, Interval, Ordinal, and Nominative Scales of Measurement (Optional)

LO1-9 Identify the ratio, interval, ordinal, and nominative scales of measurement (Optional).

In Section 1.1 we said that a variable is **quantitative** if its possible values are *numbers that represent quantities* (that is, "how much" or "how many"). In general, a quantitative variable is measured on a scale having a *fixed unit of measurement* between its possible values. For example, if we measure employees' salaries to the nearest dollar, then one dollar is the fixed unit of measurement between different employees' salaries. There are two types of quantitative variables: **ratio** and **interval.** A **ratio variable** is a quantitative variable measured on a scale such that ratios of its values are meaningful and there is an inherently defined zero value. Variables such as salary, height, weight, time, and distance are ratio variables. For example, a distance of zero miles is "no distance at all," and a town that is 30 miles away is "twice as far" as a town that is 15 miles away.

An **interval variable** is a quantitative variable where ratios of its values are not meaningful and there is not an inherently defined zero value. Temperature (on the Fahrenheit scale) is an interval variable. For example, zero degrees Fahrenheit does not represent "no heat at all," just that it is very cold. Thus, there is no inherently defined zero value. Furthermore, ratios of temperatures are not meaningful. For example, it makes no sense to say that 60° is twice as warm as 30°. In practice, there are very few interval variables other than temperature. Almost all quantitative variables are ratio variables.

In Section 1.1 we also said that if we simply record into which of several categories a population (or sample) unit falls, then the variable is **qualitative** (or **categorical**). There are two types of qualitative variables: **ordinal** and **nominative.** An **ordinal variable** is a qualitative variable for which there is a meaningful *ordering,* or *ranking,* of the categories. The measurements of an ordinal variable may be nonnumerical or numerical. For example, a student may be asked to rate the teaching effectiveness of a college professor as excellent, good, average, poor, or unsatisfactory. Here, one category is higher than the next one; that is, "excellent" is a higher rating than "good," "good" is a higher rating than "average," and so on. Therefore, teaching effectiveness is an ordinal variable having nonnumerical measurements. On the other hand, if (as is often done) we substitute the numbers 4, 3, 2, 1, and 0 for the ratings excellent through unsatisfactory, then teaching effectiveness is an ordinal variable having numerical measurements.

In practice, both numbers and associated words are often presented to respondents asked to rate a person or item. When numbers are used, statisticians debate whether the ordinal variable is "somewhat quantitative." For example, statisticians who claim that teaching effectiveness rated as 4, 3, 2, 1, or 0 is *not* somewhat quantitative argue that the difference between 4 (excellent) and 3 (good) may not be the same as the difference between 3 (good) and 2 (average). Other statisticians argue that as soon as respondents (students) see equally spaced numbers (even though the numbers are described by words), their responses are affected enough to make the variable (teaching effectiveness) somewhat quantitative. Generally speaking, the specific words associated with the numbers probably substantially affect whether an ordinal variable may be

[10]This case is based on conversations by the authors with several employees working for a leading producer of trash bags. For purposes of confidentiality, we have withheld the company's name.

considered somewhat quantitative. It is important to note, however, that in practice numerical ordinal ratings are often analyzed as though they are quantitative. Specifically, various arithmetic operations (as discussed in Chapters 2 through 14) are often performed on numerical ordinal ratings. For example, a professor's teaching effectiveness average and a student's grade point average are calculated.

To conclude this section, we consider the second type of qualitative variable. A **nominative variable** is a qualitative variable for which there is no meaningful ordering, or ranking, of the categories. A person's gender, the color of a car, and an employee's state of residence are nominative variables.

Exercises for Section 1.5

CONCEPTS

1.15 Discuss the difference between a ratio variable and an interval variable.

1.16 Discuss the difference between an ordinal variable and a nominative variable.

connect

METHODS AND APPLICATIONS

1.17 Classify each of the following qualitative variables as ordinal or nominative. Explain your answers.

Qualitative Variable	Categories
Statistics course letter grade	A B C D F
Door choice on *Let's Make A Deal*	Door #1 Door #2
Television show classifications	TV-G TV-PG TV-14 TV-MA
Personal computer ownership	Yes No
Restaurant rating	***** **** *** ** *
Income tax filing status	Married filing jointly Married filing separately Single Head of household Qualifying widow(er)

1.18 Classify each of the following qualitative variables as ordinal or nominative. Explain your answers.

Qualitative Variable	Categories
Personal computer operating system	DOS Windows XP Windows Vista Windows 7
Motion picture classifications	G PG PG-13 R NC-17 X
Level of education	Elementary Middle school High school College Graduate school
Rankings of the top 10 college football teams	1 2 3 4 5 6 7 8 9 10
Exchange on which a stock is traded	AMEX NYSE NASDAQ Other
Zip code	45056 90015 etc.

Chapter Summary

We began this chapter by discussing **data.** We learned that the data that are collected for a particular study are referred to as a **data set,** and we learned that **elements** are the entities described by a data set. In order to determine what information we need about a group of elements, we define important **variables,** or characteristics, describing the elements. **Quantitative variables** are variables that use numbers to measure quantities (that is, "how much" or "how many") and **qualitative, or categorical, variables** simply record into which of several categories an element falls.

We next discussed the difference between cross-sectional data and time series data. **Cross-sectional data** are data collected at the same or approximately the same point in time. **Time series data** are data collected over different time periods. There are various **sources of data.** Specifically, we can obtain data from **existing sources** or from **experimental or observational studies** done in-house or by paid outsiders.

We often collect data to study a **population,** which is the set of all elements about which we wish to draw conclusions. We saw that, since many populations are too large to examine in their entirety, we frequently study a population by selecting a **sample,** which is a subset of the population elements. Next we learned that, if the information contained in a sample is to accurately represent the population, then the sample should be **randomly selected** from the population.

We concluded this chapter with optional Section 1.5, which considered different types of quantitative and qualitative variables. We learned that there are two types of **quantitative variables—ratio variables,** which are measured on a scale such that ratios of its values are meaningful and there is an inherently defined zero value, and **interval variables,** for which ratios are not meaningful and there is no inherently defined zero value. We also saw that there are two types of **qualitative variables—ordinal variables,** for which there is a meaningful ordering of the categories, and **nominative variables,** for which there is no meaningful ordering of the categories.

Glossary of Terms

categorical (qualitative) variable: A variable having values that indicate into which of several categories a population element belongs. (pages 4, 14)
census: An examination of all the elements in a population. (page 7)
cross-sectional data: Data collected at the same or approximately the same point in time. (page 4)
data: Facts and figures from which conclusions can be drawn. (page 3)
data set: Facts and figures, taken together, that are collected for a statistical study. (page 3)
descriptive statistics: The science of describing the important aspects of a set of measurements. (page 8)
element: A person, object, or other entity about which we wish to draw a conclusion. (page 3)
experimental study: A statistical study in which the analyst is able to set or manipulate the values of the factors. (page 6)
factor: A variable that may be related to the response variable. (page 6)
finite population: A population that contains a finite number of elements. (page 10)
infinite population: A population that is defined so that there is no limit to the number of elements that could potentially belong to the population. (page 10)
interval variable: A quantitative variable such that ratios of its values are not meaningful and for which there is not an inherently defined zero value. (page 14)
measurement: The process of assigning a value of a variable to an element in a population or sample. (page 4)
nominative variable: A qualitative variable for which there is no meaningful ordering, or ranking, of the categories. (page 15)
observational study: A statistical study in which the analyst is not able to control the values of the factors. (page 6)
ordinal variable: A qualitative variable for which there is a meaningful ordering or ranking of the categories. (page 14)
population: The set of all elements about which we wish to draw conclusions. (page 7)
process: A sequence of operations that takes inputs and turns them into outputs. (page 10)
qualitative (categorical) variable: A variable having values that indicate into which of several categories a population element belongs. (pages 4, 14)
quantitative variable: A variable having values that are numbers representing quantities. (pages 4, 14)
ratio variable: A quantitative variable such that ratios of its values are meaningful and for which there is an inherently defined zero value. (page 14)
response variable: A variable of interest that we wish to study. (page 6)
sample: A subset of the elements in a population. (page 7)
statistical inference: The science of using a sample of measurements to make generalizations about the important aspects of a population. (page 8)
survey: An instrument employed to collect data. (page 6)
time series data: Data collected over different time periods. (page 4)
time series plot (runs plot): A plot of time series data versus time. (page 4)
variable: A characteristic of a population or sample element. (page 3)

Supplementary Exercises

1.19 THE COFFEE TEMPERATURE CASE DS Coffee

According to the website of the American Association for Justice[11] Stella Liebeck of Albuquerque, New Mexico, was severely burned by McDonald's coffee in February 1992. Liebeck, who received third-degree burns over 6 percent of her body, was awarded $160,000 in compensatory damages and $480,000 in punitive damages. A postverdict investigation revealed that the coffee temperature at the local Albuquerque McDonald's had dropped from about 185°F before the trial to about 158° after the trial.

This case concerns coffee temperatures at a fast-food restaurant. Because of the possibility of future litigation and to possibly improve the coffee's taste, the restaurant wishes to study the temperature of the coffee it serves. To do this, the restaurant personnel measure the temperature of the coffee being dispensed (in degrees Fahrenheit) at a randomly selected time during each of the 24 half-hour periods from 8 A.M. to 7:30 P.M. on a given day. This is then repeated on a second day, giving the 48 coffee temperatures in Table 1.10. Make a time series plot of the coffee temperatures, and assuming process consistency, estimate limits between which most of the coffee temperatures at the restaurant would fall.

1.20 In the article "Accelerating Improvement" published in *Quality Progress,* Gaudard, Coates, and Freeman describe a restaurant that caters to business travelers and has a self-service breakfast buffet. Interested in customer satisfaction, the manager conducts a survey over a three-week period and finds that the main customer complaint is having to wait too long to be seated. On each day from September 11 to October 1, a problem-solving team records the percentage of patrons who must wait more than one minute to be seated. A time series plot of the daily percentages is shown in Figure 1.5.[12] What does the time series plot tell us about how to improve the waiting time situation?

[11] American Association for Justice, June 16, 2006.

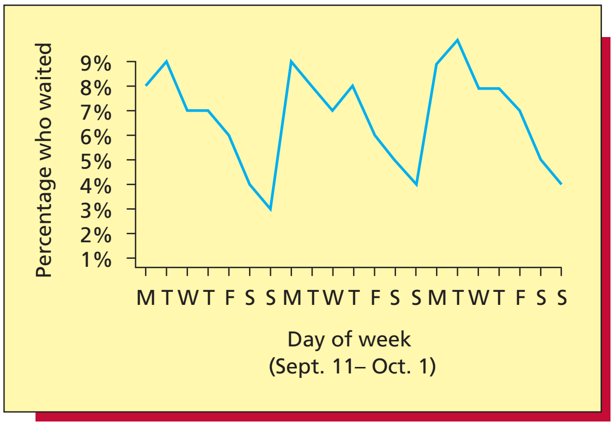
[12] The source of Figure 1.5 is M. Gaudard, R. Coates, and L. Freeman, "Accelerating Improvement," *Quality Progress,* October 1991, pp. 81–88. © 1991 American Society for Quality Control. Used with permission.

TABLE 1.10 The Coffee Temperatures for Exercise 1.19 DS Coffee

154°F	156	158	166
165	151	160	158
148	161	153	173
157	157	161	162
160	154	160	155
157	159	158	150
152	155	169	165
149	153	163	154
171	173	146	160
168	164	167	162
165	161	162	159
164	151	159	166

Note: **Time order is given by reading down the columns from left to right.**

FIGURE 1.5 Runs Plot of Daily Percentages of Customers Waiting More Than One Minute to Be Seated (for Exercise 1.20)



1.21 Internet Exercise

The website maintained by the U.S. Census Bureau provides a multitude of social, economic, and government data. In particular, this website houses selected data from the most recent *Statistical Abstract of the United States* (http://www.census.gov/compendia/statab/). Among these selected features are "Frequently Requested Tables" that can be accessed simply by clicking on the label. Go to the U.S. Census Bureau website and open the "Frequently requested tables" from the *Statistical Abstract.* Find the table of "Consumer Price Indexes by Major Groups." Construct runs plots of (1) the price index for all items over time (years), (2) the price index for food over time, (3) the price index for fuel oil over time, and (4) the price index for electricity over time. For each runs plot, describe apparent trends in the price index.

Excel, MegaStat, and MINITAB for Statistics

In this book we use three types of software to carry out statistical analysis—Excel 2010, MegaStat, and MINITAB 16. **Excel** is, of course, a general purpose electronic spreadsheet program and analytical tool. The analysis ToolPak in Excel includes many procedures for performing various kinds of basic statistical analyses. **MegaStat** is an add-in package that is specifically designed for performing statistical analysis in the Excel spreadsheet environment. **MINITAB** is a computer package designed expressly for conducting statistical analysis. It is widely used at many colleges and universities and in a large number of business organizations. The principal advantage of Excel is that, because of its broad acceptance among students and professionals as a multipurpose analytical tool, it is both well-known and widely available. The advantages of a special-purpose statistical software package like MINITAB are that it provides a far wider range of statistical procedures and it offers the experienced analyst a range of options to better control the analysis. The advantages of MegaStat include (1) its ability to perform a number of statistical calculations that are not automatically done by the procedures in the Excel ToolPak and (2) features that make it easier to use than Excel for a wide variety of statistical analyses. In addition, the output obtained by using MegaStat is automatically placed in a standard Excel spreadsheet and can be edited by using any of the features in Excel. MegaStat can be copied from the book's website. Excel, MegaStat, and MINITAB, through built-in functions, programming languages, and macros, offer almost limitless power. Here, we will limit our attention to procedures that are easily accessible via menus without resort to any special programming or advanced features.

Commonly used features of Excel 2010, MegaStat, and MINITAB 16 are presented in this chapter along with an initial application—the construction of a time series plot of the gas mileages in Table 1.6. You will find that the limited instructions included here, along with the built-in help features of all three software packages, will serve as a starting point from which you can discover a variety of other procedures and options. Much more detailed descriptions of MINITAB 16 can be found in other sources, in particular in the manual *Meet MINITAB 16 for Windows.* This manual is available in print and as a pdf file, viewable using Adobe Acrobat Reader, on the MINITAB Inc. website—go to http://www.minitab.com/uploadedFiles/Shared_Resources/Documents/MeetMinitab/EN16_MeetMinitab.pdf. Similarly, there are a number of alternative reference materials for Microsoft Excel 2010. Of course, an understanding of the related statistical concepts is essential to the effective use of any statistical software package.

Appendix 1.1 ■ Getting Started with Excel

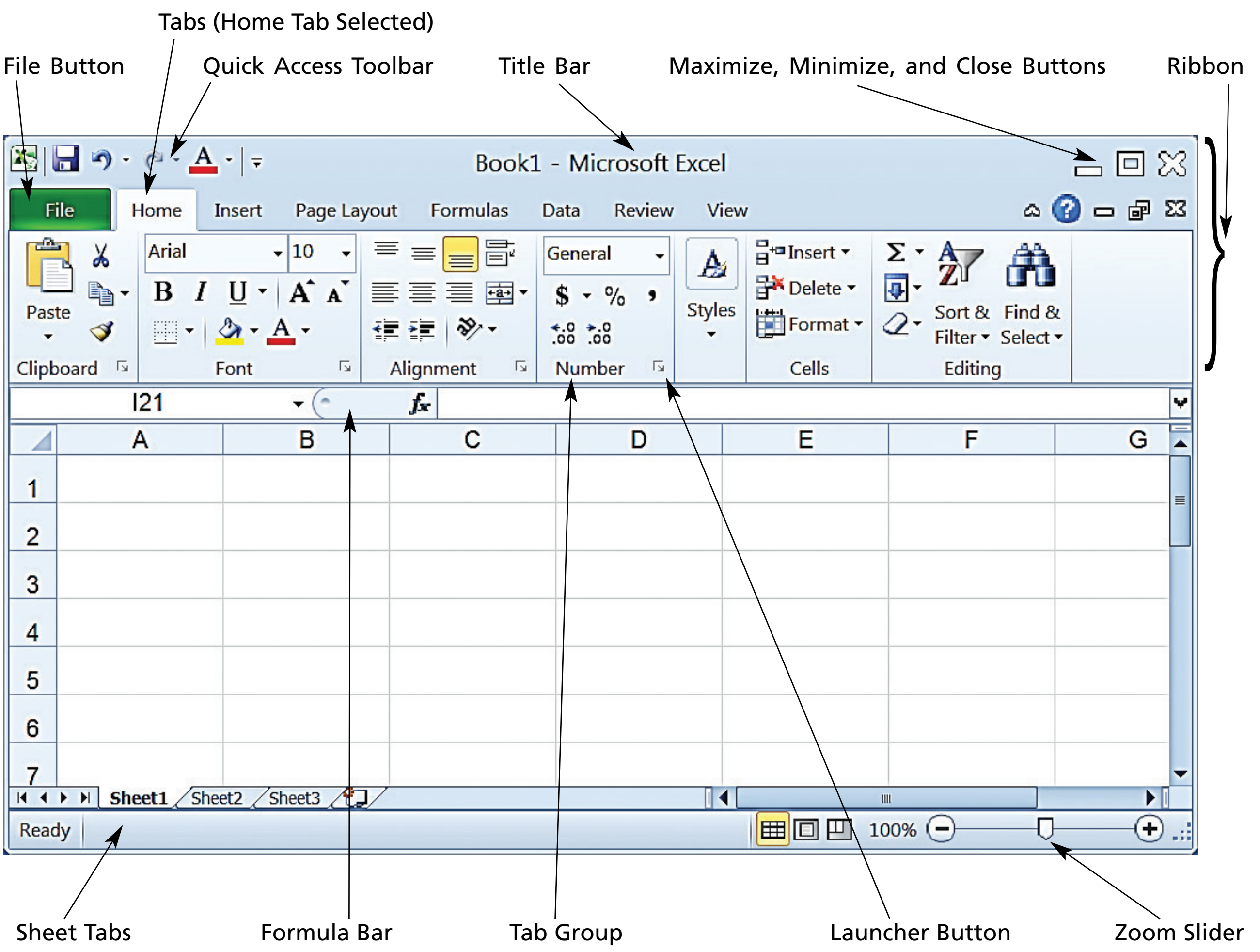
Because Excel 2010 may be new to some readers, and because the Excel 2010 window looks somewhat different from previous versions of Excel, we will begin by describing some characteristics of the Excel 2010 window. Versions of Excel prior to 2007 employed many drop-down menus. This meant that many features were "hidden" from the user, which resulted in a steep learning curve for beginners. In Excel 2007 and 2010, Microsoft tried to reduce the number of features that are hidden in drop-down menus. Therefore, Excel 2010 displays all of the applicable commands needed for a particular type of task at the top of the Excel window. These commands are represented by a tab-and-group arrangement called the **ribbon**—see the right side of the illustration of an Excel 2010 window below. The commands displayed in the ribbon are regulated by a series of **tabs** located near the top of the ribbon. For example, in the illustration below, the **Home tab** is selected. If we selected a different tab, say, for example, the **Page Layout tab,** the commands displayed by the ribbon would be different.

We now briefly describe some basic features of the Excel 2010 window:

1 **File button:** By clicking on this button, the user obtains a menu of often used commands—for example, Open, Save, Print, and so forth. This menu also provides access to a large number of Excel options settings.

2 **Tabs:** Clicking on a tab results in a ribbon display of features, commands, and options related to a particular type of task. For example, when the ***Home tab*** is selected (as in the figure below), the features, commands, and options displayed by the ribbon are all related to making entries into the Excel worksheet. As another example, if the ***Formulas tab*** is selected, all of the features, commands, and options displayed in the ribbon relate to using formulas in the Excel worksheet.

3 **Quick access toolbar:** This toolbar displays buttons that provide shortcuts to often used commands. Initially, this toolbar displays Save, Undo, and Redo buttons. The user can customize this toolbar by adding shortcut buttons for other commands (such as, New, Open, Quick Print, and so forth). This can be done by clicking on the arrow button directly to the right of the Quick Access toolbar and by making selections from the "Customize" drop-down menu that appears.



4 **Title bar:** This bar shows the name of the currently active workbook and contains the Quick Access Toolbar as well as the Maximize, Minimize, and Close buttons.

5 **Ribbon:** A grouping of toolbars, tabs, commands, and features related to performing a particular kind of task—for example, making entries into the Excel spreadsheet. The particular features displayed in the ribbon are controlled by selecting a *Tab.* If the user is working in the spreadsheet workspace and wishes to reduce the number of features displayed by the ribbon, this can be done by right-clicking on the ribbon and by selecting "Minimize the Ribbon." We will often Minimize the Ribbon in the Excel appendices of this book in order to focus attention on operations being performed and results being displayed in the Excel spreadsheet.

6 **Sheet tabs:** These tabs show the name of each sheet in the Excel workbook. When the user clicks a sheet tab, the selected sheet becomes active and is displayed in the Excel spreadsheet. The name of a sheet can be changed by double-clicking on the appropriate sheet tab and by entering the new name.

7 **Formula bar:** When a worksheet cell is selected, the formula bar displays the current content of the cell. If the cell content is defined by a formula, the defining formula is displayed in the formula bar.

8 **Tab group:** This is a labeled grouping of commands and features related to performing a particular type of task.

9 **Launcher button:** Some of the tab groups have a launcher button—for example, the Clipboard, Font, Alignment, and Number tab groups each have such a button. Clicking on the launcher button opens a dialog box or task pane related to performing operations in the tab group.

10 **Zoom slider:** By moving this slider right or left, the cells in the Excel spreadsheet can be enlarged or reduced in size.

We now take a look at some features of Excel that are common to many analyses. When the instructions call for a sequence of selections, the sequence will be presented in the following form:

Select **Home : Format : Row Height**

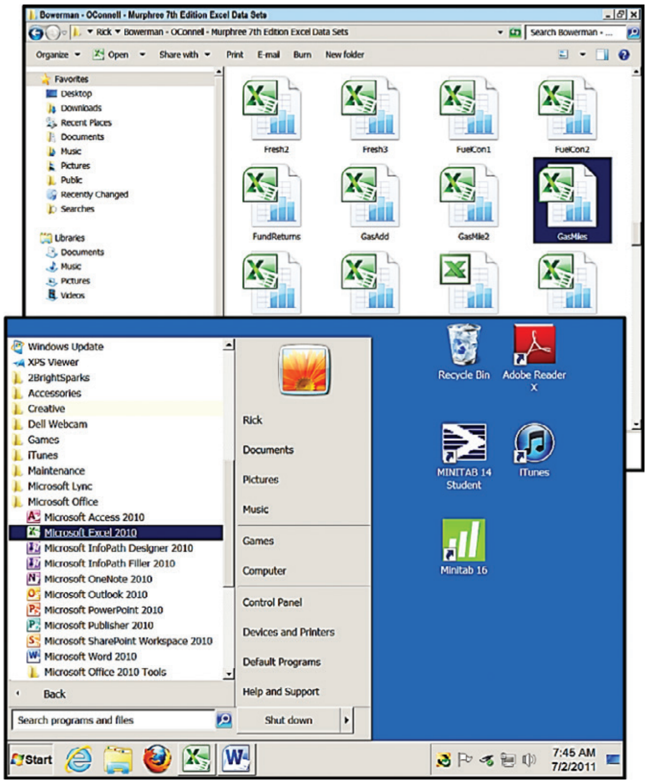
This notation indicates that we first select the Home tab on the ribbon, then we select Format from the Cells Group on the ribbon, and finally we select Row Height from the Format drop-down menu.

For many of the statistical and graphical procedures in Excel, it is necessary to provide a range of cells to specify the location of data in the spreadsheet. Generally, the range may be specified either by typing the cell locations directly into a dialog box or by dragging the selected range with the mouse. Although for the experienced user, it is usually easier to use the mouse to select a range, the instructions that follow will, for precision and clarity, specify ranges by typing in cell locations. The selected range may include column or variable labels—labels at the tops of columns that serve to identify variables. When the selected range includes such labels, it is important to select the "Labels check box" in the analysis dialog box.

Starting Excel Procedures for starting Excel may vary from one installation to the next. If you are using a public computing laboratory, you may wish to consult local documentation. For typical Excel installations, you will generally be able to start Excel with a sequence of selections from the Microsoft Windows start menu something like the following:

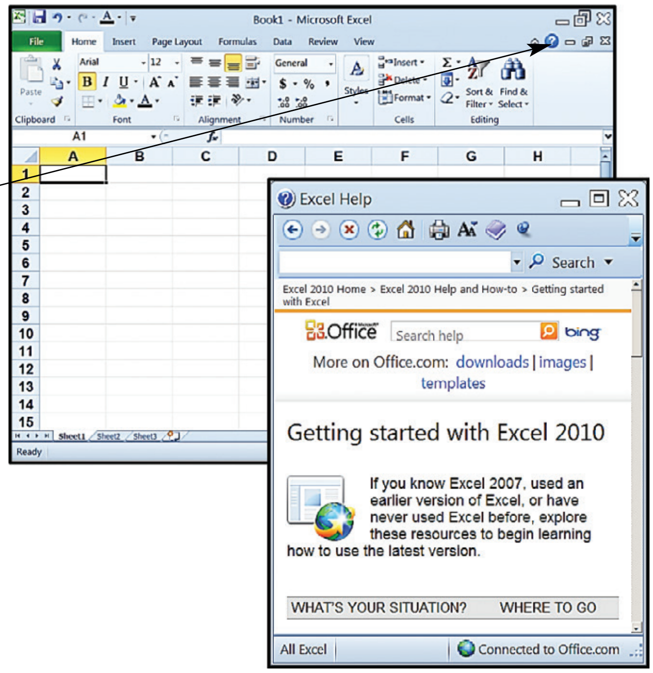
Start : All Programs : Microsoft Office : Microsoft Excel 2010

You can also start Excel with a previously saved Excel spreadsheet (like GasMiles.xlsx or one of the other data files that can be downloaded from this book's website) by double-clicking on the spreadsheet file's icon in Windows Explorer.



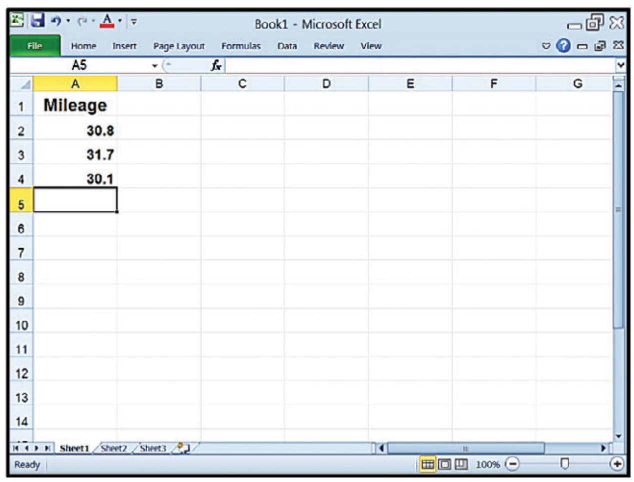
After starting Excel, the display will generally show a blank Excel workbook.

Help resources Like most Windows programs, Excel includes on-line help via a Help Menu that includes search capability as well as a table of contents. To display the Help Menu, click on the "Question Mark" button in the upper-right corner of the ribbon.



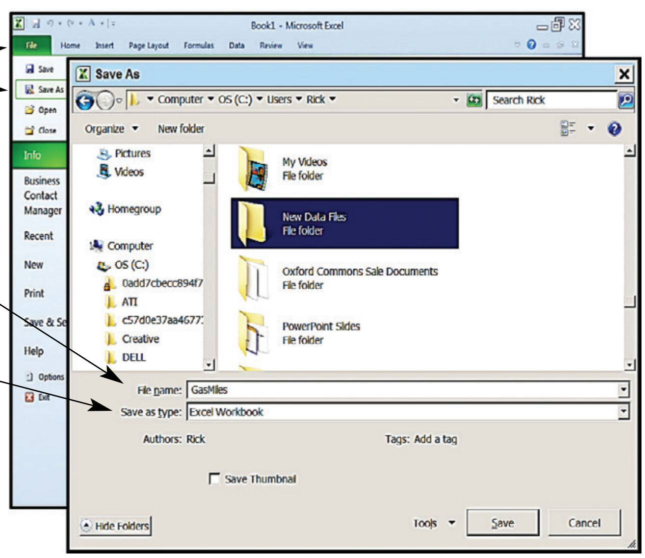
Entering data (entering the gas mileages in Table 1.6 on page 11) from the keyboard (data file: GasMiles.xlsx):

- In a new Excel workbook, click on cell A1 in Sheet1 and type a label—that is, a variable name—say, Mileage, for the gasoline mileages.
- Beginning in cell A2 (directly under the column label Mileage) type the mileages from Table 1.6 on page 11 down the column, pressing the Enter key following each entry.



Saving data (saving the gasoline mileage data):

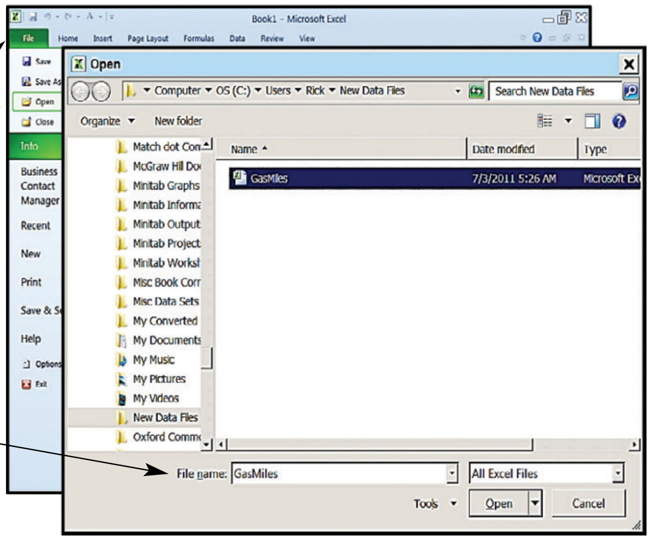
- To begin, click on the **File** button and select **Save As**.
- In the "Save As" dialog box, select the destination drive and folder. Here we have selected a file folder called New Data Files in Rick's System folder on the local C drive.
- Enter the desired file name in the "File name" window. In this case we have chosen the name GasMiles.
- Select Excel Workbook in the "Save as type" window.
- Click the **Save** button in the "Save As" dialog box.



Retrieving an Excel spreadsheet containing the gasoline mileages in Table 1.6 on page 11 (data file: GasMiles.xlsx):

- Select **File : Open**

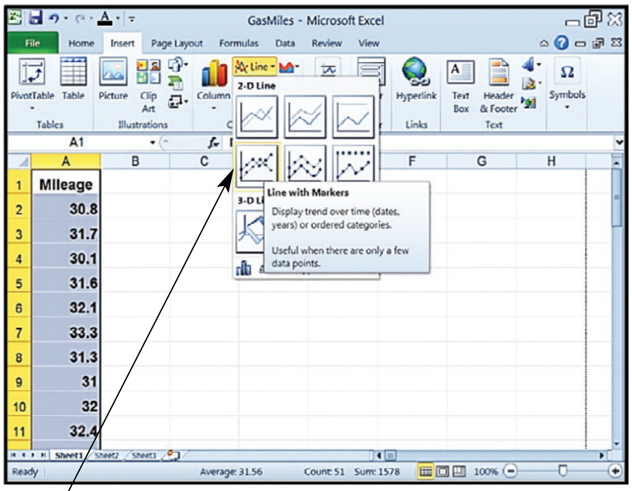
  That is, click on the File button and then select Open.
- In the Open dialog box, select the desired source drive, folder, and file. Here we have selected the GasMiles file in a folder named New Data Files in Rick's System folder on the local C drive. When you select the desired file by clicking on it, the file name will be shown in the "File name" window.
- Click the Open button in the Open dialog box.

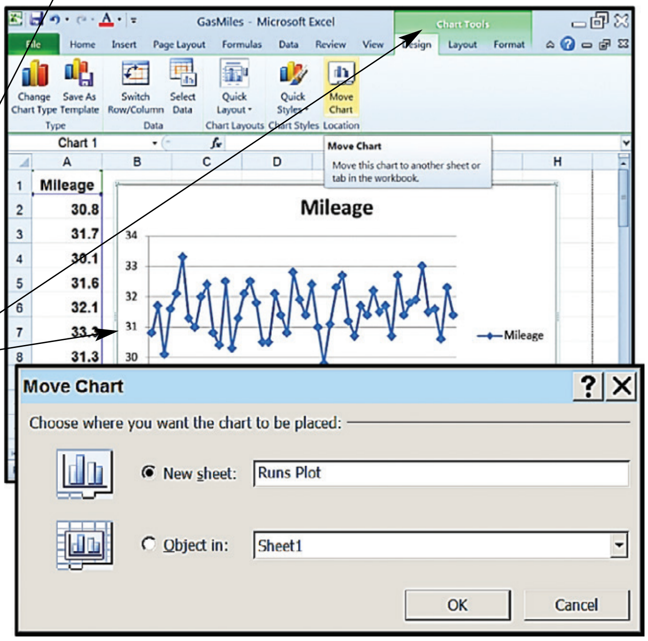


Creating a time series (runs) plot similar to Figure 1.3 on page 11 (data file: GasMiles.xlsx):

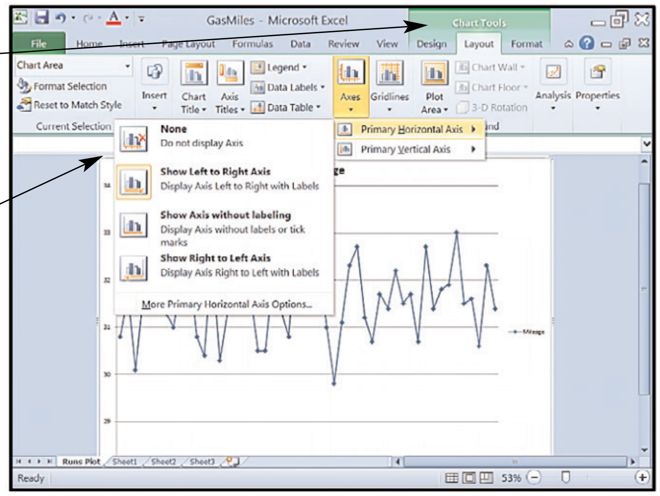
- Enter the gasoline mileage data into column A of the worksheet with label Mileage in cell A1.
- Click on any cell in the column of mileages, or select the range of the data to be charted by dragging with the mouse. Selecting the range of the data is good practice because—if this is not done—Excel will sometimes try to construct a chart using all of the data in your worksheet. The result of such a chart is often nonsensical. Here, of course, we only have one column of data—so there would be no problem. But, in general, it is a good idea to select the data before constructing a graph.
- Select **Insert : Line : 2-D Line : Line with Markers**

  Here select the Insert tab and then select Line from the Charts group. When Line is selected, a gallery of line charts will be displayed. From the gallery, select the desired chart—in this case a 2-D Line chart with markers. The proper chart can be selected by looking at the sample pictures. As an alternative, if the cursor is hovered over a picture, a descriptive "tool tip" of the chart type will be displayed. In this case, the "Line with Markers" tool tip was obtained by hovering the cursor over the highlighted picture.
- When you click on the "2-D Line with Markers" icon, the chart will appear in a graphics window and the Chart Tools ribbon will be displayed.
- To prepare the chart for editing, it is best to move the chart to a new worksheet—called a "chart sheet." To do this, click on the **Design tab** and select **Move Chart.**
- In the Move Chart dialog box, select the "New sheet" option, enter a name for the new sheet—here, "Runs Plot"—into the "New sheet" window, and click OK.





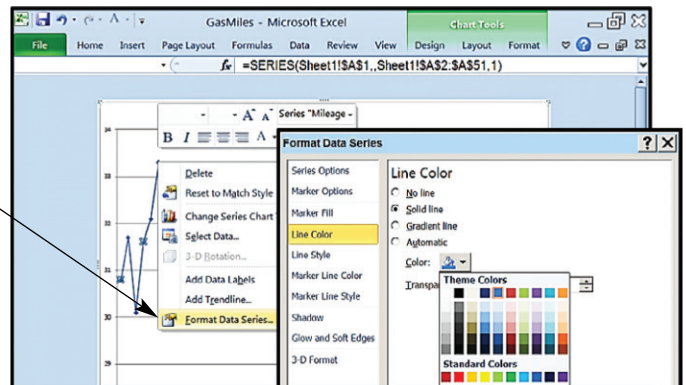
- The Chart Tools ribbon will be displayed and the chart will be placed in a chart sheet in a larger format that is more convenient for editing.
- In order to edit the chart, select the **Layout** tab from the Chart Tools ribbon. By making selections from the ribbon, many chart attributes can be edited. For instance, when you click on Axes as shown, various options for formatting the horizontal and vertical axes can be selected.



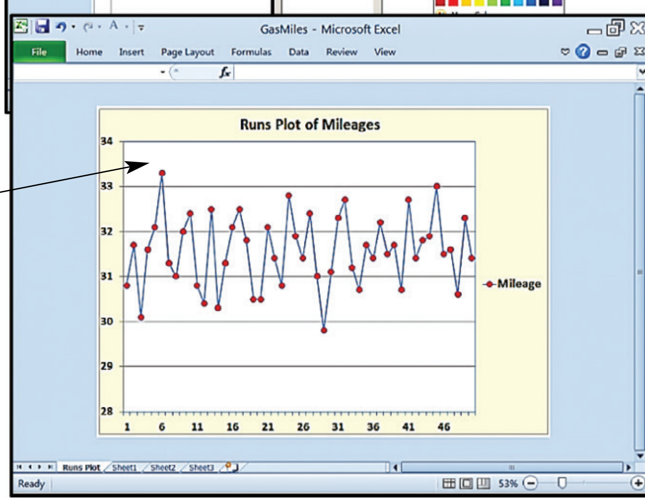
- A chart can also be edited by right-clicking on the portion of the chart that we wish to revise. For instance, in the screen shown, we have right-clicked on one of the plotted data points. When this is done, we obtain a menu as shown. If we select "Format Data Series," we obtain a dialog box that provides many options for editing the data series (the plotted points and their connecting lines). For example, if (as shown) we select

Line Color : Solid Line

and then click on the Color arrow button, we obtain a drop-down menu that allows us to select a desired color for the connecting lines between the plotted points. We can edit other portions of the chart in the same way.



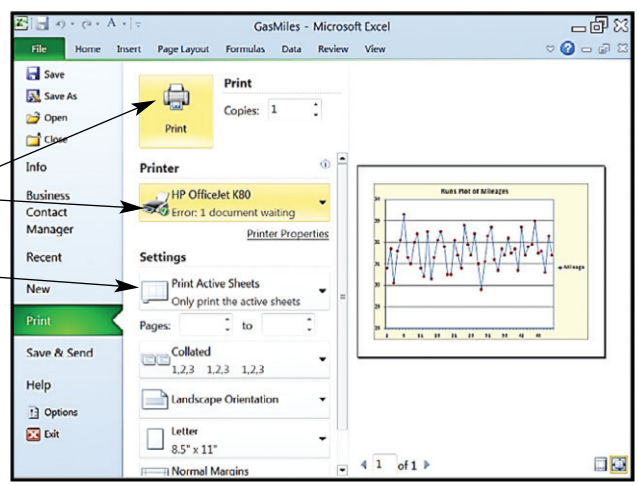
- Here we show an edited runs plot. This revised chart was constructed from the original runs plot created by Excel using various options like those illustrated above. This chart can be copied directly from the worksheet (simply right-click on the graph and select Copy from the pop-up menu) and can then be pasted into a word-processing document.



The chart can be printed from this worksheet as follows:

- Select **File : Print**
  That is, click on the File button and then select Print.
- Select the desired printer in the Printer Name window and then click Print.

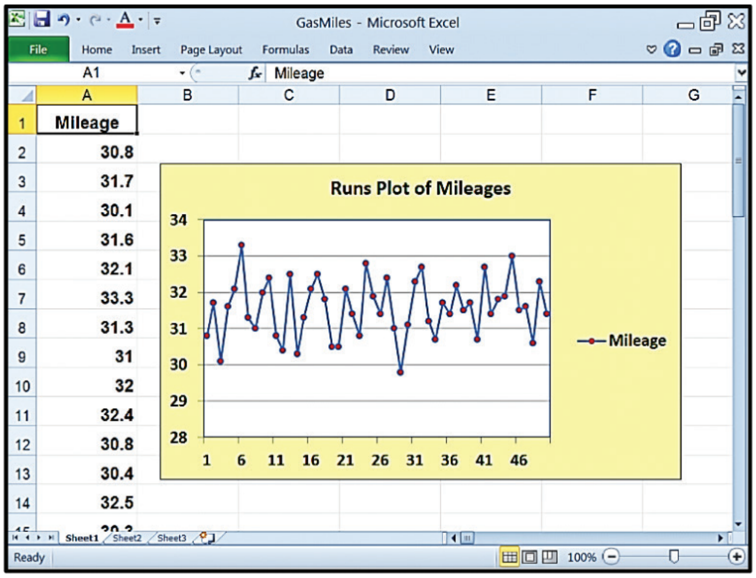
There are many print options available in Excel—for printing a selected range, selected sheets, or an entire workbook—making it possible to build and print fairly sophisticated reports directly from Excel.



Printing a spreadsheet with an embedded graph:

- Click outside the graph to print both the worksheet contents (here the mileage data) and the graph. Click on the graph to print only the graph.
- Select **File : Print**

  That is, click on the File button and then select Print.
- Select the desired printer in the Printer Name window and click OK in the Print dialog box.



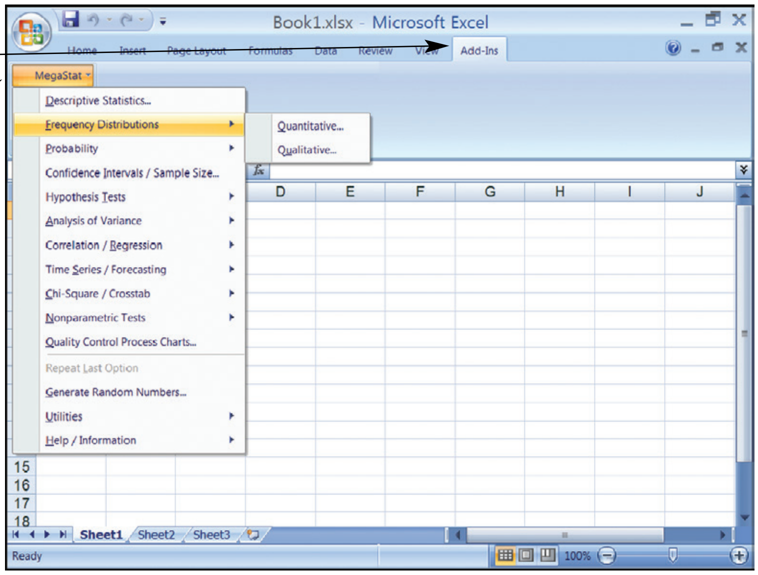
Including Excel output in reports You can easily copy Excel results—selected spreadsheet ranges and graphs—to the Windows clipboard. Then paste them into an open word processing document. Once copied to a word-processing document, Excel results can be documented, edited, resized, and rearranged as desired into a cohesive record of your analysis. The cut and paste process is quite similar to the MINITAB examples at the end of Appendix 1.3.

Calculated Results As we proceed through this book, you will see that Excel often expresses calculated results that are fractions in **scientific notation**. For example, Excel might express the results of a calculation as 7.77 E-6. To get the decimal point equivalent, the "E-6" says we must move the decimal point 6 places to the left. This would give us the fraction .00000777.

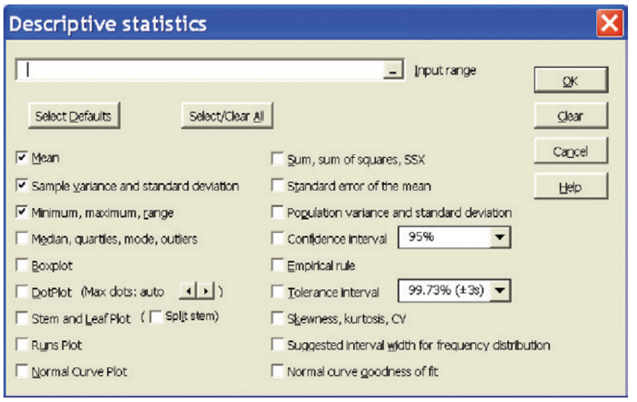
Appendix 1.2 ■ Getting Started with MegaStat

MegaStat, which was developed by Professor J. B. Orris of Butler University, is an Excel add-in that performs statistical analyses within an Excel workbook. Instructions for installing MegaStat can be found on this book's website.

- After installation, you can access MegaStat by clicking on the Add-Ins tab (on the ribbon) and by then selecting MegaStat from the Add-Ins group of Menu Commands. When you select MegaStat, the MegaStat menu appears as shown in the screen. Most of the menu options display submenus. If a menu item is followed by an ellipsis (...), clicking it will display a dialog box for that option.



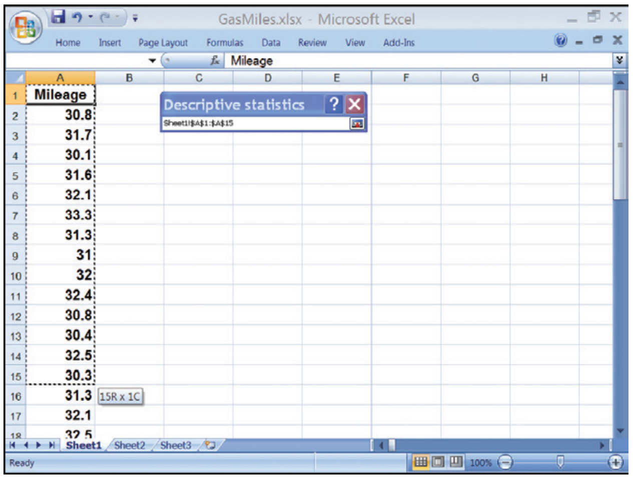
- A dialog box allows you to specify the data to be used and other inputs and options. A typical dialog box is shown in the screen to the right.
- After you have selected the needed data and options, you click OK. The dialog box then disappears and MegaStat performs the analysis.



Before we look at specific dialog boxes, we will describe some features that are common to all of the options. MegaStat use is intuitive and very much like other Excel operations; however, there are some features unique to MegaStat.

Data selection Most MegaStat dialog boxes have fields where you select input ranges that contain the data to be used. Such a field is shown in the dialog box illustrated above—it is the long horizontal window with the label "Input range" to its right. Input ranges can be selected using four methods:

1 **Pointing and dragging with the mouse.** Simply select the desired data by pointing to the data, by left-clicking on the first data item, and dragging the cursor to select the rest of the data as illustrated below.



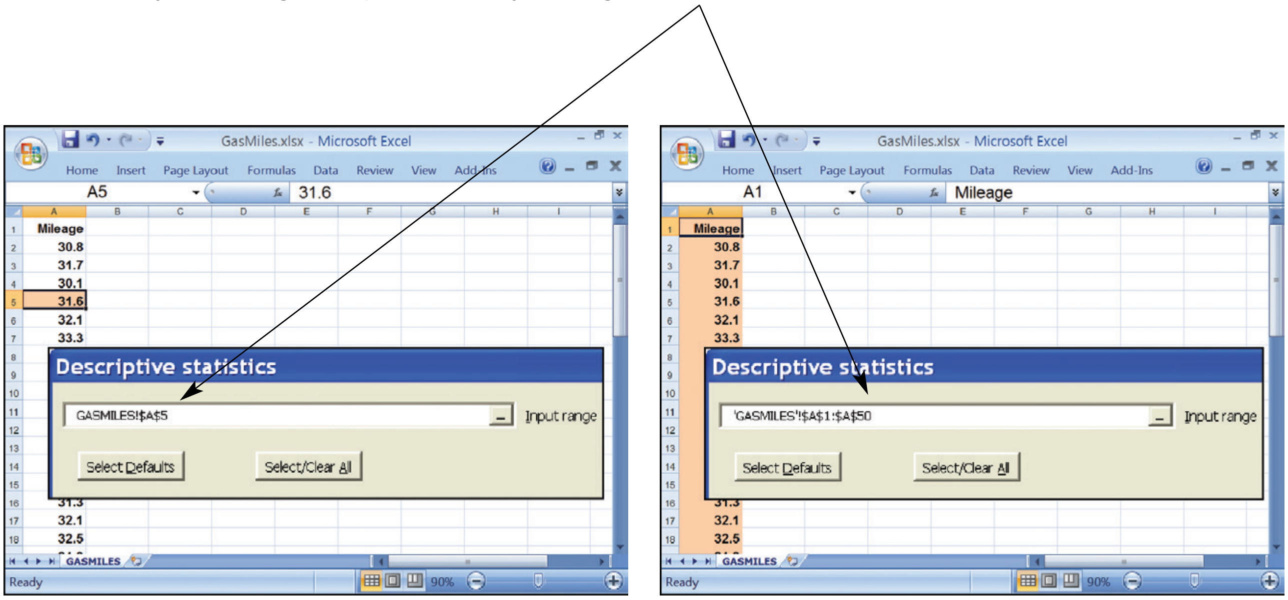
Because the dialog box "pops-up" on the screen, it may block some of your data. You can move a dialog box around on the screen by placing the mouse pointer over the title bar (colored area at the top), and by then clicking and holding the left mouse button while dragging the dialog box to a new location. You can even drag it partially off the screen.

You will also notice that when you start selecting data by dragging the mouse pointer, the dialog box will collapse to a smaller size to help you see the underlying data. It will automatically return to full size when you release the mouse button. You can also collapse and uncollapse the dialog box manually by clicking the collapse (-) button at the right end of the field. Clicking the button again will uncollapse the dialog box. (Never use the X button to try to collapse or uncollapse a dialog box.)

2 **Using MegaStat's AutoExpand feature.** Pointing and dragging to select data can be tedious if you have a lot of data. When you drag the mouse down it is easy to overshoot the selection and then you have to drag the mouse back until you get the area correctly selected. AutoExpand allows rapid data selection without having to drag through the entire column of data. Here's how it works:

- Make sure the input box has the focus (that is, click in it to make the input box active). An input box has the focus when the insertion pointer is blinking in it.
- Click in one cell of the column you want. If more than one column is being selected, drag the mouse across the columns.
- Right-click over the input field or left-click the label "Input Range" to the right of the input box. The data range will expand to include all of the rows in the region where you selected one row.

This procedure is illustrated below. In the left screen, we have left-clicked on one cell in the column of data labeled Mileage. In the right screen, we see the result after we right-click over the input field or left-click on the label "Input range." Notice that the entire column of data has been selected in the right screen. This can be seen by examining the input field or by looking at the column of data.



With a little practice you will find this is a very efficient way to select data. The only time you cannot use it is when you want to use a partial column of data. You should also be aware that the autoexpand stops when it finds a blank cell; thus any summations or other calculations at the bottom of a column would be selected.

Note: When using the above methods of data selection you may select variables in an alternating sequence by holding the CTRL key while making multiple selections.

3 **Typing the name of a named range.** If you have previously identified a range of cells using Excel's name box, you may use that name to specify a data range in a MegaStat dialog box. This method can be very useful if you are using the same data for several different statistical procedures.

4 **Typing a range address.** You may type any valid Excel range address, for example, $A$1:$A$101, into the input field. This is the most cumbersome way to specify data ranges, but it certainly works.

Data labels For most procedures, the first cell in each input range can be a label. If the first cell in a range is text, it is considered a label; if the first cell is a numeric value, it is considered data. If you want to use numbers as variable labels, you must enter the numbers as text by preceding them with a single quote mark—for instance, '2. Even though Excel stores times and dates as numbers, MegaStat will recognize them as labels if they are formatted as time/date values. If data labels are not part of the input range, the program automatically uses the cell immediately above the data range as a label if it contains a text value. If an option can consider the entire first row (or column) of an input range as labels, any numeric value in the row will cause the entire row to be treated as data. Finally, if the program detects sequential integers (1,2,3...) in a location where you might want labels, it will display a warning message. Otherwise, the rule is: **text cells are labels, numeric cells are data.**

Output When you click OK on a MegaStat dialog box, it performs some statistical analysis and needs a place to put its output. It looks for a worksheet named Output. If it finds one, it goes to the end of it and appends its output; if it doesn't find an Output worksheet, it creates one. MegaStat will never make any changes to the user's worksheets; it only sends output to its Output sheet.

MegaStat makes a good attempt at formatting the output, but **it is important to remember that the Output sheet is just a standard Excel worksheet and can be modified in any way by the user.** You can adjust column widths and change any formatting that you think needs improvement. You can insert, delete, and modify cells. You can copy all or part of the output to another worksheet or to another application such as a word processor.

When the program generates output, it adjusts column widths for the current output. If you have previous output from a different option already in the Output sheet, the column widths for the previous output may be altered. You can attempt to fix this by manually adjusting the column widths. Alternatively, you can make it a practice to always start a new output sheet. The **Utilities menu** has options for **deleting the Output sheet, for making a copy of it, and for starting a new one.**

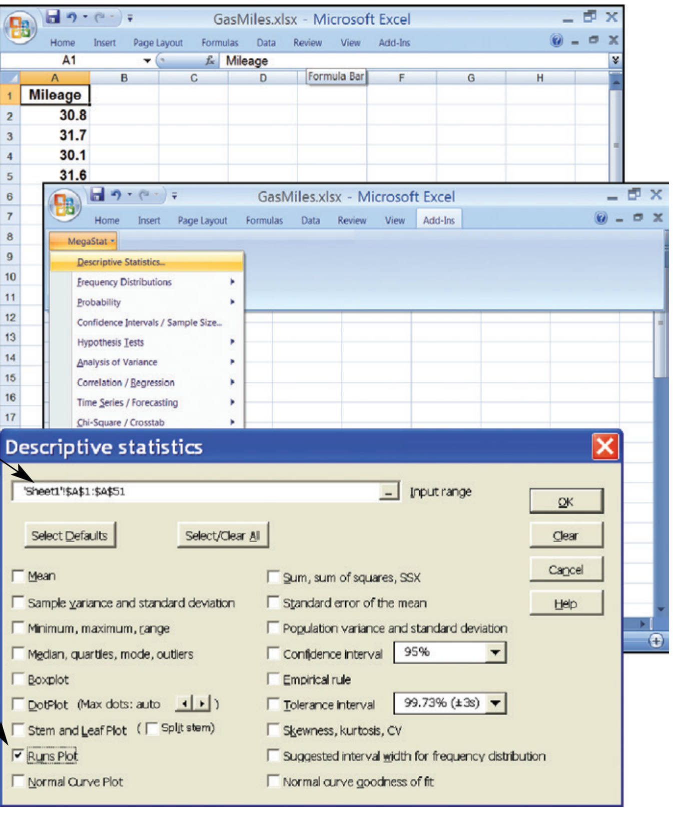
An example We now give an example of using MegaStat to carry out statistical analysis. When the instructions call for a sequence of selections, the sequence will be presented in the following form:

Add-Ins : MegaStat : Probability : Counting Rules

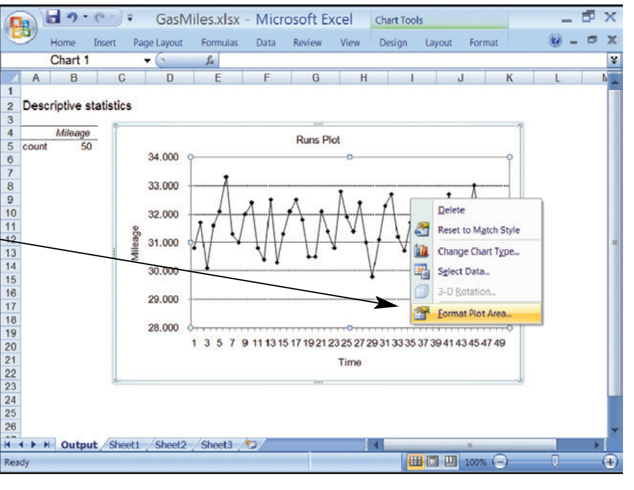
This notation says that **Add-Ins** is the first selection (from the ribbon), **MegaStat** is the second selection from the Add-Ins group of Menu Commands; next **Probability** is selected from the MegaStat drop-down menu; and finally **Counting Rules** is selected from the Probability submenu.

Creating a time series (runs) plot of gasoline mileages similar to Figure 1.3 on page 11 (data file: GasMiles.xlsx):

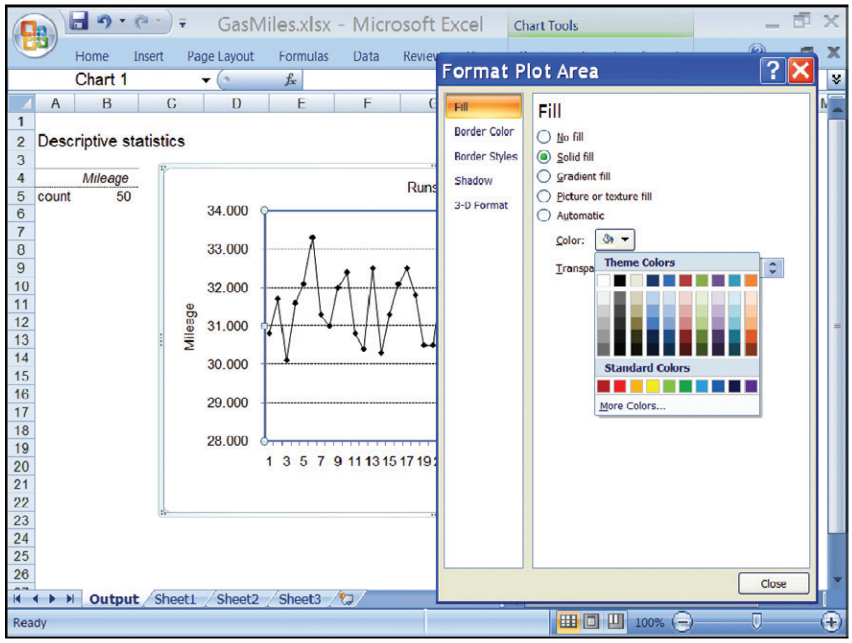
- Enter the mileage data in Table 1.6 on page 11 into column A with the label Mileage in cell A1 and with the 50 mileages in cells A2 through A51.
- Select **Add-Ins : MegaStat : Descriptive Statistics**
- In the Descriptive Statistics dialog box, enter the range $A$1:$A$51 into the Input range box. The easiest way to do this is to use the MegaStat AutoExpand feature. Simply select one cell in column A (say, cell A4, for instance) by clicking on the cell. Then, either right-click in the Input range box or left-click on the label "Input range" to the right of the Input range box.
- Place a checkmark in the Runs Plot checkbox.
- Click OK in the Descriptive Statistics dialog box.



MegaStat places the resulting analysis (in this case the runs plot) in an output worksheet. This is a standard Excel worksheet, which can be edited using any of the usual Excel features. For instance, by right-clicking on various portions of the runs plot graphic, the plot can be edited in many ways. Here we have right-clicked on the plot area. By selecting **Format Plot Area,** we are able to edit the graphic in a variety of ways.



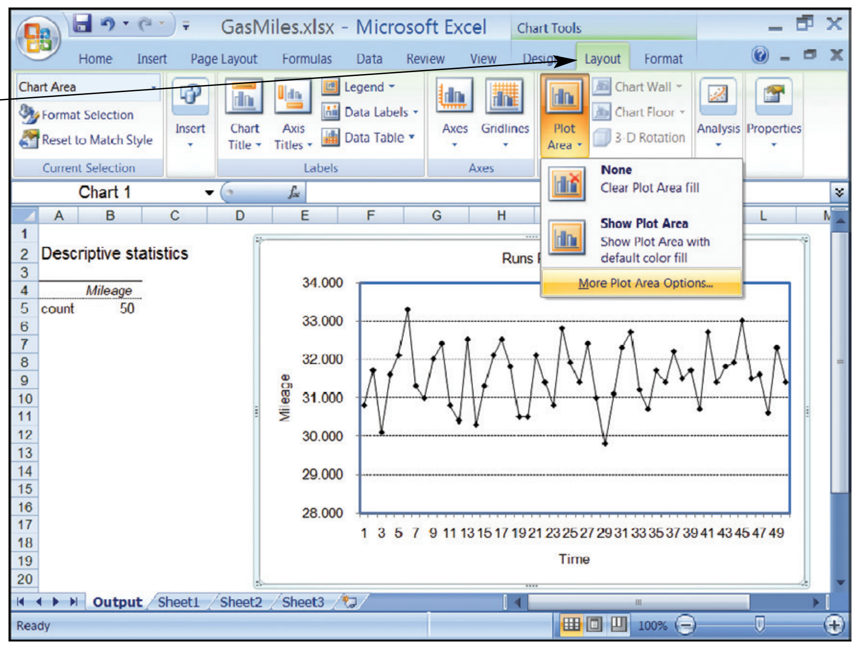
In the Format Plot Area dialog box, we can add color to the runs plot and edit the plot in many other ways.



Alternatively, we can edit the runs plot by selecting

Chart Tools : Layout

By making selections from the Labels, Axes, and Background groups, the plot can be edited in a variety of ways. For example, in the screen shown we have selected the Plot Area button in the Background group. This gives us many options for editing the plot area of the graphic.



Appendix 1.3 ■ Getting Started with MINITAB

We begin with a look at some features of MINITAB that are common to most analyses. When the instructions call for a sequence of selections from a series of menus, the sequence will be presented in the following form:

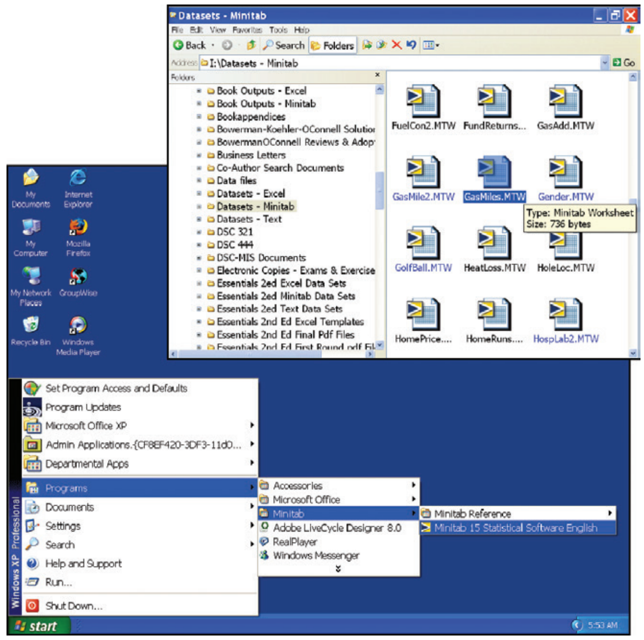
Stat : Basic Statistics : Descriptive Statistics

This notation indicates that Stat is the first selection from the Minitab menu bar, next Basic Statistics is selected from the Stat pull-down menu, and finally Descriptive Statistics is selected from the Basic Statistics pull-down menu.

Starting MINITAB Procedures for starting MINITAB may vary from one installation to the next. If you are using a public computing laboratory, you may have to consult local documentation. For typical MINITAB installations, you will generally be able to start MINITAB with a sequence of selections from the Microsoft Windows Start menu something like the following:

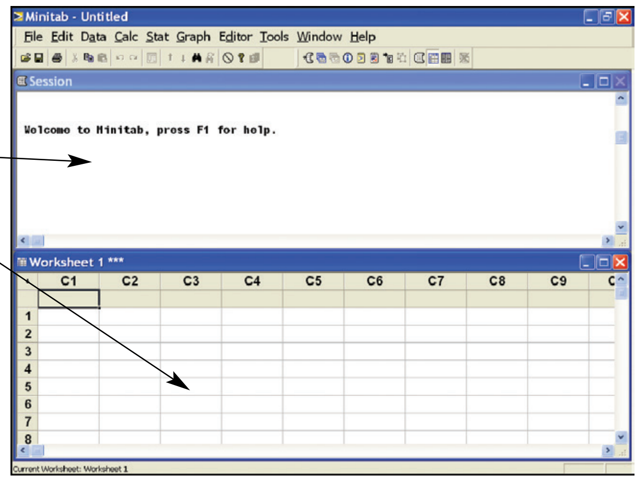
- Select **Start : Programs : Minitab : Minitab 15 Statistical Software English**

You can also start MINITAB with a previously saved MINITAB worksheet (like GasMiles.MTW or one of the many other data files that can be downloaded from this book's website) by double-clicking on the worksheet's icon in the Windows Explorer.



After you start MINITAB, the display is partitioned into two working windows. These windows serve the following functions:

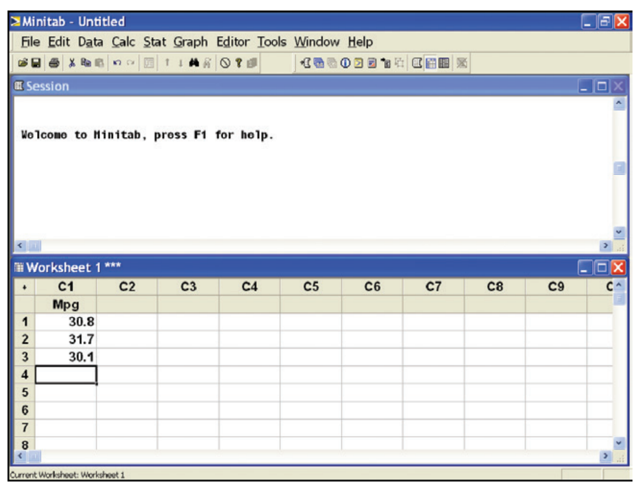
- The "Session window" is the area where MINITAB commands and basic output are displayed.
- The "Data window" is an Excel-like worksheet where data can be entered and edited.



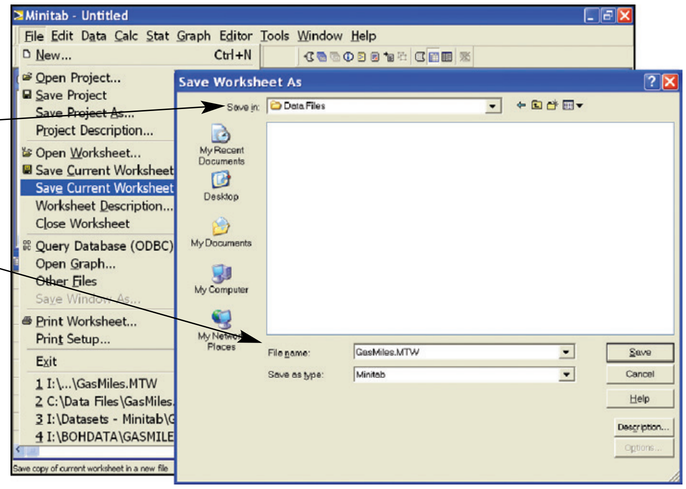
Help resources Like most Windows programs, MINITAB includes online help via a Help Menu. The Help feature includes standard Contents and Search entries as well as Tutorials that introduce MINITAB concepts and walk through some typical MINITAB sessions. Also included is a StatGuide that provides guidance for interpreting statistical tables and graphs in a practical, easy-to-understand way.

Entering data (entering the gasoline mileage data in Table 1.6 on page 11) from the keyboard:

- In the Data window, click on the cell directly below C1 and type a name for the variable—say, Mpg—and press the Enter key.
- Starting in row 1 under column C1, type the values for the variable (gasoline mileages from Table 1.6 on page 11) down the column, pressing the Enter key after each number is typed.

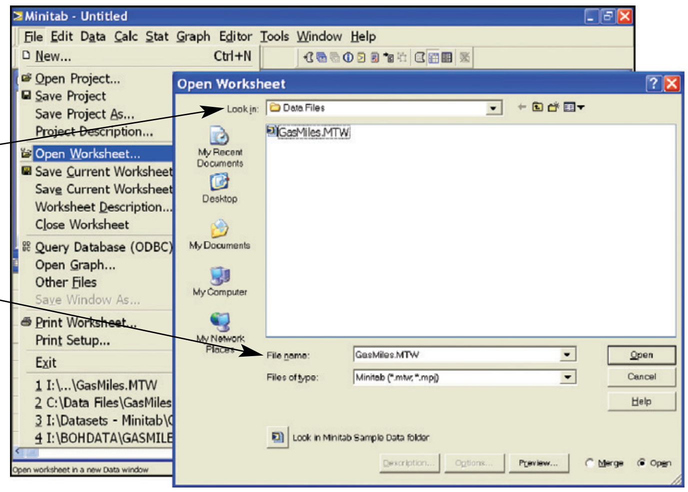


Saving data (saving the gasoline mileage data):



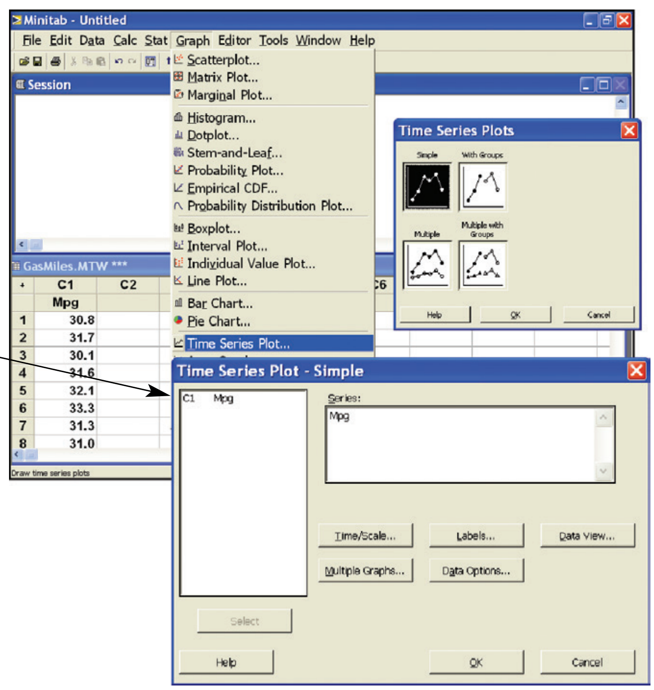
- Select **File : Save Current Worksheet As**
- In the "Save Worksheet As" dialog box, use the "Save in" drop-down menu to select the destination drive and folder. (Here we have selected a folder named Data Files on the Local C drive.)
- Enter the desired file name in the File name box. Here we have chosen the name GasMiles. MINITAB will automatically add the extension .MTW.
- Click the Save button in the "Save Worksheet As" dialog box.

Retrieving a MINITAB worksheet containing the gasoline mileage data in Table 1.6 (data file: GasMiles .MTW):



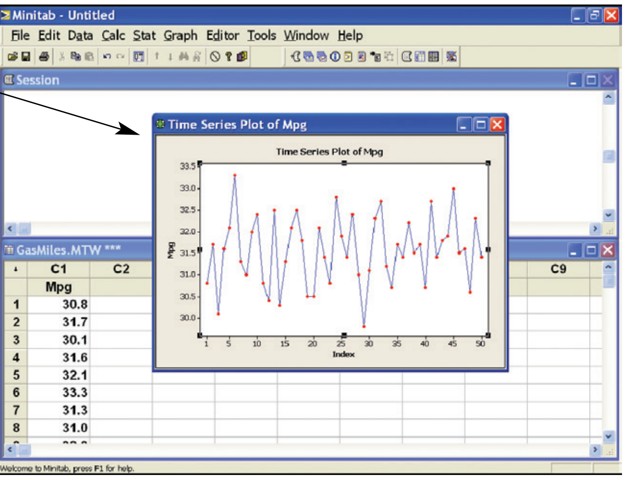
- Select **File : Open Worksheet**
- In the Open Worksheet dialog box, use the "Look in" drop-down menu to select the source drive and folder. (Here we have selected a folder named Data Files on the Local C drive.)
- Enter the desired file name in the File name box. (Here we have chosen the MINITAB worksheet GasMiles.MTW.)
- Click the Open button in the Open Worksheet dialog box.
- MINITAB may display a dialog box with the message, "A copy of the content of this file will be added to the current project." If so, click OK.

Creating a time series (or runs) plot similar to Figure 1.3 on page 11 (data file: GasMiles.MTW):

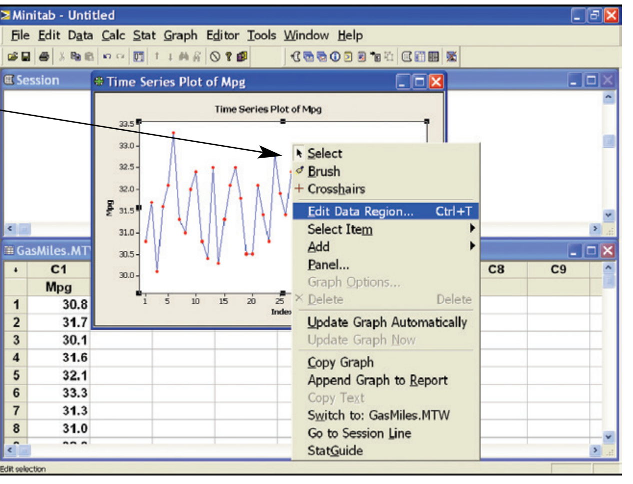


- Select **Graph : Time Series Plot**
- In the Time Series Plots dialog box, select Simple, which produces a time series plot of data that is stored in a single column, and click OK.
- In the "Time Series Plot—Simple" dialog box, enter the name of the variable, Mpg, into the Series window. Do this either (1) by typing its name, or (2) by double-clicking on its name in the list of variables on the left side of the dialog box. Here, this list consists of the single variable Mpg in column C1.
- Click OK in the "Time Series Plot—Simple" dialog box.

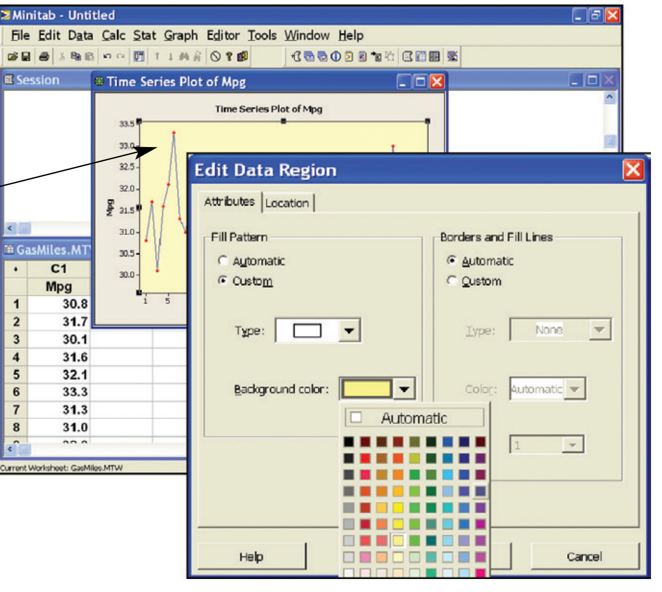
- **The time series plot** will appear in a graphics window.



- **The graph can be edited** by right-clicking on the portion you wish to edit. For instance, here we have right-clicked on the data region.
- Selecting "Edit Data Region" from the pop-up window gives a dialog box that allows you to edit this region. The *x* and *y* scales, *x* and *y* axis labels, title, plot symbols, connecting lines, data region, figure region, and so forth can all be edited by right-clicking on that particular portion of the graph.

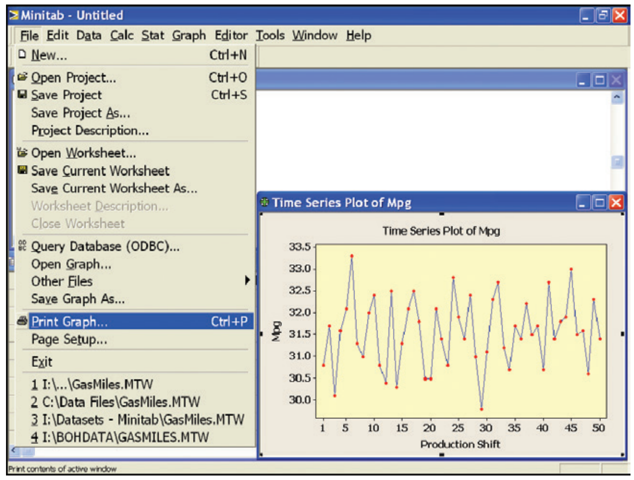


- For instance, after right-clicking on the data region and then selecting "Edit Data Region" from the pop-up menu, the Edit Data Region dialog box allows us to edit various attributes of this region. As shown, selecting Custom and clicking on the Background Color arrow allows us to change the background color of the data region.



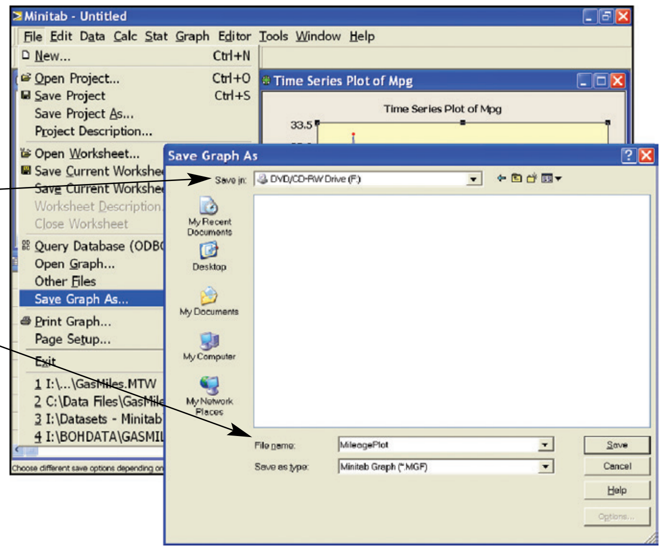
Printing a high-resolution graph similar to Figure 1.3 on page 11 (data file: GasMiles.MTW):

- Click in the graphics window to select it as the active window.
- Select **File : Print Graph** to print the graph.
- Select the appropriate printer and click OK in the Print dialog box.



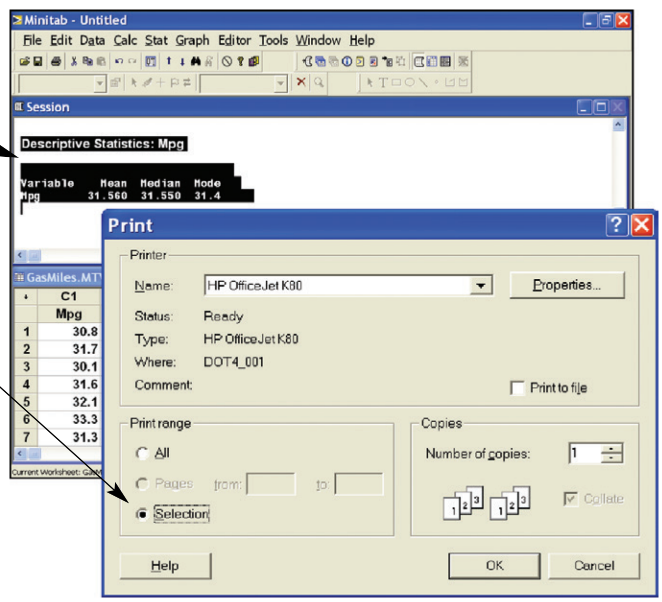
Saving the high-resolution graph:

- Click on the graph to make the graphics window the active window.
- Select **File : Save Graph As**
- In the "Save Graph As" dialog box, use the "Save in" drop-down menu to select the destination drive and folder (here we have selected the DVD/CD-RW drive).
- Enter the desired file name in the File name box (here we have chosen the name MileagePlot). MINITAB will automatically add the file extension .MGF.
- Click the Save button in the "Save Graph As" dialog box.



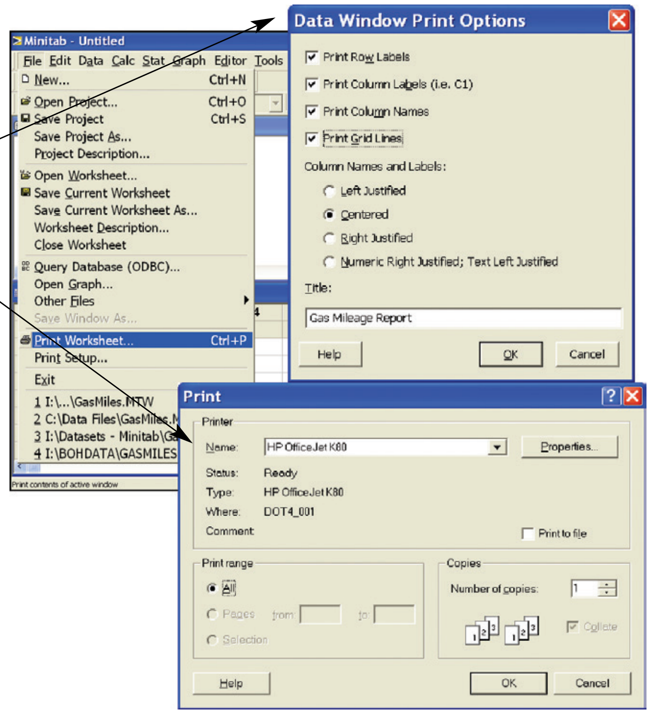
Printing data from the Session window (shown) or Data window (data file: GasMiles.MTW):
To print selected output from the Session window:

- Use the mouse to select the desired output or text (selected output will be reverse-highlighted in black).
- Select **File : Print Session Window**
- In the Print dialog box, the Print range will be the "Selection" option. To print the entire session window, select the Print range to be "All."
- Select the desired printer from the Printer Name drop-down menu.
- Click OK in Print dialog box.



To print the contents of the Data window (that is, to print the MINITAB worksheet):

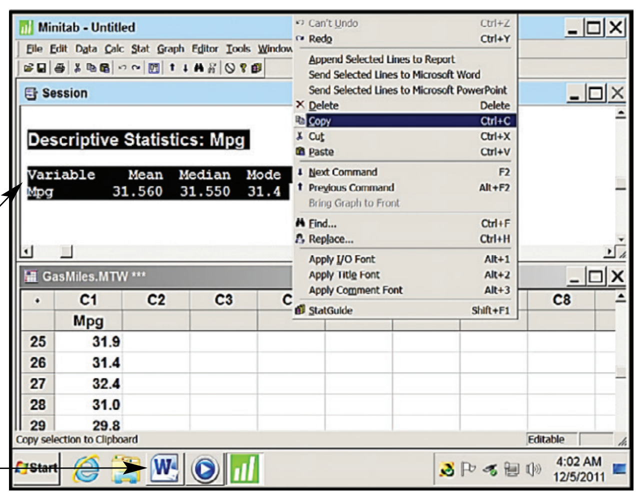
- Click in the Data window to select it as active.
- Select **File : Print Worksheet**
- Make selections as desired in the Data Window Print Options dialog box, add a title in the Title window if desired, and click OK.
- Select the desired printer from the Printer Name drop-down menu and click OK in the Print dialog box.

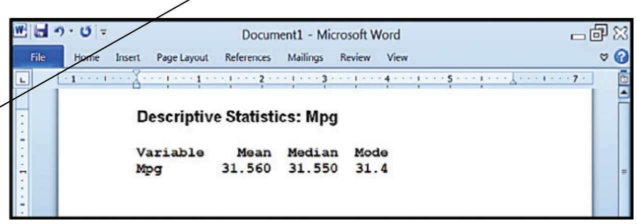


Including MINITAB output in reports The immediately preceding examples show how to print various types of output directly from MINITAB. Printing is a useful way to capture a quick hard-copy record of an analysis result. However, you may prefer at times to collect selected analysis results and arrange them with related narrative documentation in a report that can be saved and printed as a unit. This is easily accomplished by copying selected MINITAB results to the Windows clipboard and by pasting them into your favorite word processor. Once copied to a word processor document, MINITAB results can be documented, edited, resized, and rearranged as desired into a cohesive record of your analysis. The following sequence of screens illustrates the process of copying MINITAB output into a Microsoft Word document.

Copying session window output to a word processing document:

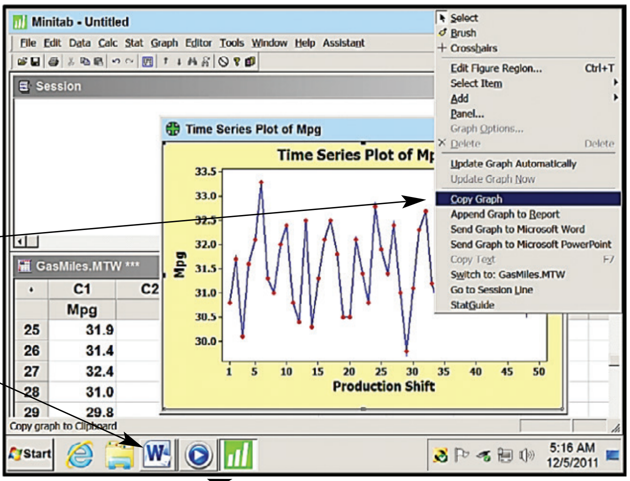
- Be sure to have a word processing document open to receive the results.
- Use the scroll bar on the right side of the Session window to locate the results you wish to copy and drag the mouse to select the desired output (selected output will be reverse-highlighted in black).
- Copy the selected output to the Windows clipboard by clicking the Copy icon on the MINITAB toolbar or by right-clicking on the selected text and then selecting Copy from the pop-up menu.
- Switch to your word processing document by clicking the Microsoft Word button on the Windows task bar.
- Click in your word processing document to position the cursor at the desired insertion point.
- Click the Paste button on the word processing power bar or right-click at the insertion point and select Paste from the pop-up menu.
- Return to your MINITAB session by clicking the MINITAB button on the Windows task bar.





Copying high-resolution graphics output to a word processing document:

- Be sure to have a word processing document open to receive the results.
- Copy the selected contents of the high-resolution graphics window to the Windows clipboard by right-clicking in the graphics window and by then clicking Copy Graph on the pop-up menu.
- Switch to your word processing document by clicking the Microsoft Word button on the Windows task bar.
- Click in your word processing document to position the cursor at the desired insertion point.
- Click the Paste button on the word processor power bar or right-click at the insertion point and select Paste from the pop-up menu.
- Return to your MINITAB session by clicking the MINITAB button on the Windows task bar.



Results Here is how the copied results might appear in Microsoft Word. These results can be edited, resized, repositioned, and combined with your own additional documentation to create a cohesive record of your analysis.