To restate the premise of this section, statistics, when used properly, can be beneficial in obtaining much information, but when used improperly, can lead to much misinformation. It is like your automobile. If you use your automobile to get to school or work or to go on a vacation, that's good. But if you use it to run over your neighbor's dog because it barks all night long and tears up your flower garden, that's not so good!

1–6 Computers and Calculators

Objective 8

Explain the importance of computers and calculators in statistics.

In the past, statistical calculations were done with pencil and paper. However, with the advent of calculators, numerical computations became much easier. Computers do all the numerical calculation. All one does is to enter the data into the computer and use the appropriate command; the computer will print the answer or display it on the screen. Now the TI-83 Plus or TI-84 Plus graphing calculator accomplishes the same thing.

There are many statistical packages available; this book uses MINITAB and Microsoft Excel. Instructions for using MINITAB, the TI-83 Plus or TI-84 Plus graphing calculator, and Excel have been placed at the end of each relevant section, in subsections entitled Technology Step by Step.

You should realize that the computer and calculator merely give numerical answers and save the time and effort of doing calculations by hand. You are still responsible for understanding and interpreting each statistical concept. In addition, you should realize that the results come from the data and do not appear magically on the computer. Doing calculations using the procedure tables will help you reinforce this idea.

The author has left it up to instructors to choose how much technology they will incorporate into the course.

Technology *Step by Step*

MINITAB
Step by Step

General Information

MINITAB statistical software provides a wide range of statistical analysis and graphing capabilities.

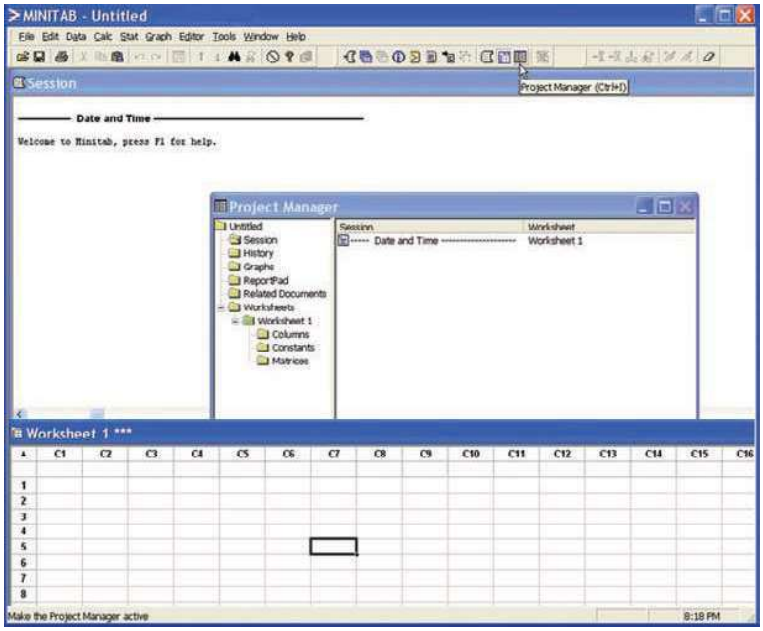
Take Note

In this text you will see captured screen images from computers running MINITAB Release 14. If you are using an earlier or later release of MINITAB, the screens you see on your computer may bear slight visual differences from the screens pictured in this text. **But don't be alarmed!** All the Step by Step operations described in this text, including the commands, the menu options, and the functionality, will work just fine on your computer.

Start the Program

1. Click the Windows Start Menu, then All Programs.
2. Click the MINITAB folder and then click  MINITAB, the program icon. The program screen will look similar to the one shown here. You will see the Session Window, the Worksheet Window, and perhaps the Project Manager Window.
3. Click the Project Manager icon on the toolbar to bring the project manager to the front.

For Vista, click the Start button, then "All Programs." Next click "MINITAB Solutions" and then "MINITAB Statistical Software English."

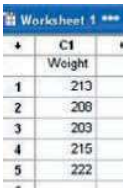


To use the program, data must be entered from the keyboard or from a file.

Entering Data in MINITAB

In MINITAB, all the data for one variable are stored in a column. Step by step instructions for entering these data follow.

Data
213 208 203 215 222



1. Click in row 1 of Worksheet 1***. This makes the worksheet the active window and puts the cursor in the first cell. The small data entry arrow in the upper left-hand corner of the worksheet should be pointing down. If it is not, click it to change the direction in which the cursor will move when you press the [Enter] key.
2. Type in each number, pressing [Enter] after each entry, including the last number typed.
3. *Optional:* Click in the space above row 1 to type in **Weight,** the column label.

Save a Worksheet File

4. Click on the **File Menu.** *Note:* This is *not* the same as clicking the disk icon  .
5. Click **Save Current Worksheet As . . .**
6. In the dialog box you will need to verify three items:
   a) Save in: Click on or type in the disk drive and directory where you will store your data. For a CD this might be **A:.**
   b) File Name: Type in the name of the file, such as **MyData.**
   c) Save as Type: The default here is MINITAB. An extension of mtw is added to the name.

   Click [Save]. The name of the worksheet will change from Worksheet 1*** to MyData.MTW.

Open the Databank File

The raw data are shown in Appendix D. There is a row for each person's data and a column for each variable. MINITAB data files comprised of data sets used in this book, including the

Databank, are available on the accompanying CD-ROM or at the Online Learning Center (www.mhhe.com/bluman). Here is how to get the data from a file into a worksheet.

1. Click **File>Open Worksheet.** A sequence of menu instructions will be shown this way.

   Note: This is *not* the same as clicking the file icon  . If the dialog box says Open Project instead of Open Worksheet, click [Cancel] and use the correct menu item. The Open Worksheet dialog box will be displayed.

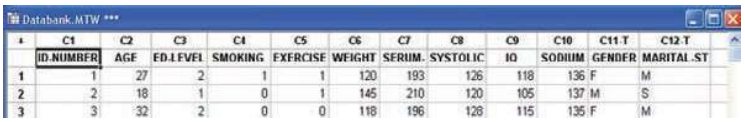
2. You must check three items in this dialog box.

   a) The Look In: dialog box should show the directory where the file is located.

   b) Make sure the Files of Type: shows the correct type, MINITAB [*.mtw].

   c) Double-click the file name in the list box Databank.mtw. A dialog box may inform you that a copy of this file is about to be added to the project. Click on the checkbox if you do not want to see this warning again.

3. Click the [OK] button. The data will be copied into a second worksheet. Part of the worksheet is shown here.



Databank.MTW ***

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11-T	C12-T
	ID.NUMBER	AGE	ED.LEVEL	SMOKING	EXERCISE	WEIGHT	SERUM.	SYSTOLIC	IQ	SODIUM	GENDER	MARITAL.ST
1	1	27	2	1	1	120	193	126	118	136	F	M
2	2	18	1	0	1	145	210	120	105	137	M	S
3	3	32	2	0	0	118	196	128	115	135	F	M

   a) You may maximize the window and scroll if desired.

   b) C12-T Marital Status has a T appended to the label to indicate alphanumeric data. MyData.MTW is not erased or overwritten. Multiple worksheets can be available; however, only the active worksheet is available for analysis.

4. To switch between the worksheets, select **Window>MyData.MTW.**

5. Select **File>Exit** to quit. To save the project, click [Yes].

6. Type in the name of the file, **Chapter01.** The Data Window, the Session Window, and settings are all in one file called a project. Projects have an extension of mpj instead of mtw.

   Clicking the disk icon  on the menu bar is the same as selecting **File>Save Project.**

   Clicking the file icon  is the same as selecting **File>Open Project.**

7. Click [Save]. The mpj extension will be added to the name. The computer will return to the Windows desktop. The two worksheets, the Session Window results, and settings are saved in this project file. When a project file is opened, the program will start up right where you left off.

TI-83 Plus or TI-84 Plus
Step by Step

The TI-83 Plus or TI-84 Plus graphing calculator can be used for a variety of statistical graphs and tests.

General Information

To turn calculator on:
Press **ON** key.
To turn calculator off:
Press **2nd [OFF].**

To reset defaults only:

1. Press **2nd,** then **[MEM].**
2. Select **7,** then **2,** then **2.**

Optional. To reset settings on calculator and clear memory: (*Note:* This will clear all settings and programs in the calculator's memory.)
Press **2nd,** then **[MEM].** Then press **7,** then **1,** then **2.**
(Also, the contrast may need to be adjusted after this.)

To adjust contrast (if necessary):
Press **2nd.** Then press and hold ▲ to darken or ▼ to lighten contrast.
To clear screen:
Press **CLEAR.**
(*Note:* This will return you to the screen you were using.)
To display a menu:
Press appropriate menu key. Example: **STAT.**
To return to home screen:
Press **2nd,** then **[QUIT].**
To move around on the screens:
Use the arrow keys.
To select items on the menu:
Press the corresponding number or move the cursor to the item, using the arrow keys. Then press **ENTER.**
(*Note:* In some cases, you do not have to press **ENTER,** and in other cases you may need to press **ENTER** twice.)

Entering Data

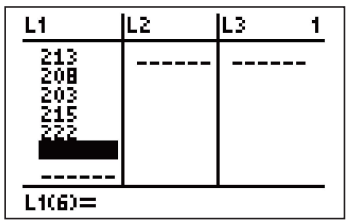
To enter single-variable data (if necessary, clear the old list):

1. Press **STAT** to display the Edit menu.
2. Press **ENTER** to select 1:Edit.
3. Enter the data in L_1 and press **ENTER** after each value.
4. After all data values are entered, press **STAT** to get back to the Edit menu or **2nd [QUIT]** to end.

Example TI1–1

Enter the following data values in L_1: **213, 208, 203, 215, 222.**

Output



To enter multiple-variable data:
The TI-83 Plus or TI-84 Plus will take up to six lists designated L_1, L_2, L_3, L_4, L_5, and L_6.

1. To enter more than one set of data values, complete the preceding steps. Then move the cursor to L_2 by pressing the ▶ key.
2. Repeat the steps in the preceding part.

Editing Data

To correct a data value before pressing **ENTER,** use ◀ and retype the value and press **ENTER.**
To correct a data value in a list after pressing **ENTER,** move cursor to incorrect value in list and type in the correct value. Then press **ENTER.**
To delete a data value in a list:
Move cursor to value and press **DEL.**
To insert a data value in a list:

1. Move cursor to position where data value is to be inserted, then press **2nd [INS].**
2. Type data value; then press **ENTER.**

To clear a list:

1. Press **STAT,** then **4.**
2. Enter list to be cleared. Example: To clear L_1, press **2nd [L_1].** Then press **ENTER.**

   (*Note:* To clear several lists, follow STEP 1, but enter each list to be cleared, separating them with commas. To clear all lists at once, follow STEP 1; then press **ENTER.**)

Sorting Data

To sort the data in a list:

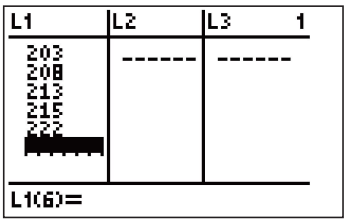
1. Enter the data in L_1.
2. Press **STAT 2** to get SortA to sort the list in ascending order.
3. Then press **2nd [L_1] ENTER.**

The calculator will display Done.

4. Press **STAT ENTER** to display sorted list.

   (*Note:* The SortD or **3** sorts the list in descending order.)

Output



Example TI1–2

Sort in ascending order the data values entered in Example TI1–1.

Excel
Step by Step

General Information

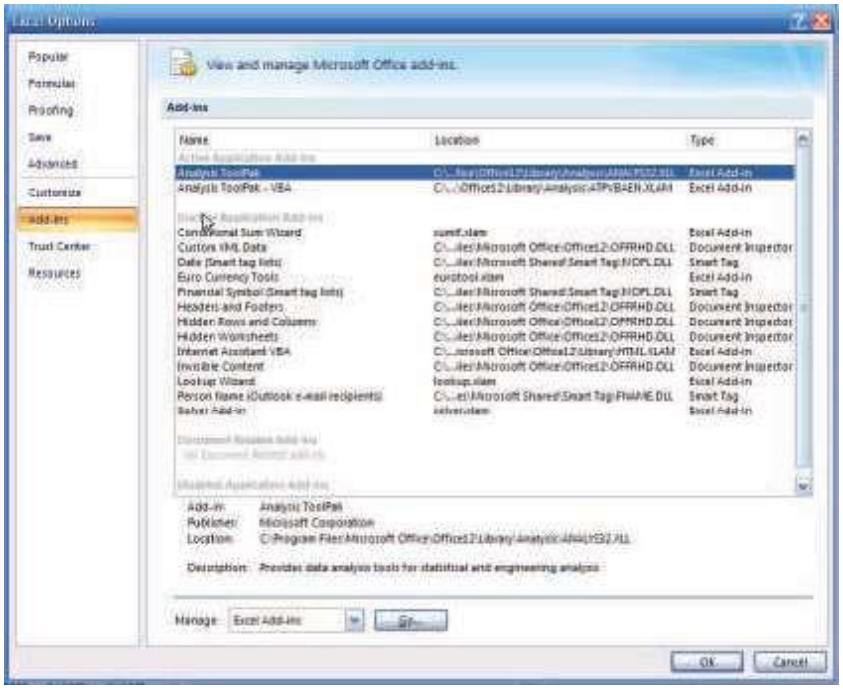
Microsoft Excel 2007 has two different ways to solve statistical problems. First, there are built-in functions, such as STDEV and CHITEST, available from the standard toolbar by clicking Formulas, then selecting the Insert Function icon. Another feature of Excel that is useful for calculating multiple statistical measures and performing statistical tests for a set of data is the Data Analysis command found in the Analysis Tool-Pak Add-in.

To load the Analysis Tool-Pak:

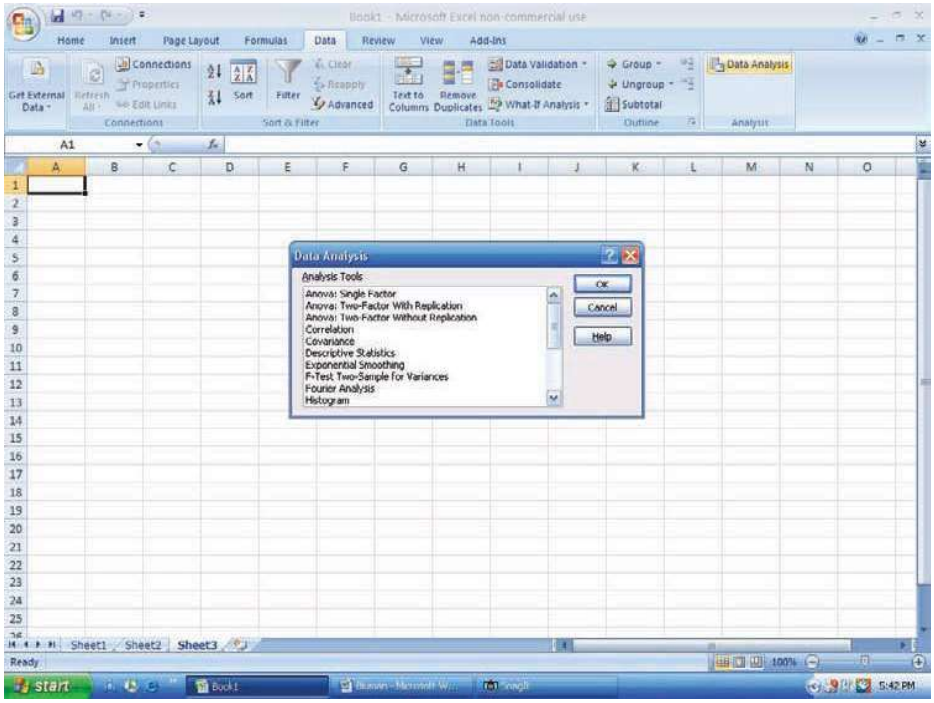
Click the Microsoft Office button, then select Excel Options.

Excel's Analysis ToolPak Add-In

1. Click **Add-Ins,** and select **Add-ins** from the list of options on the left side of the options box.
2. Select the Analysis Tool-Pak, then click the **Go** button at the bottom of the options box.



3. After loading the Analysis Tool-Pak, the Data Analysis command is available in the Analysis group on the Data tab.



MegaStat

Later in this text you will encounter a few Excel Technology Step by Step operations that will require the use of the MegaStat Add-in for Excel. MegaStat can be downloaded from the CD that came with your textbook as well as from the text's Online Learning Center at www.mhhe.com/bluman.

1. Save the Zip file containing the MegaStat Excel Add-in file (MegaStat.xls) and the associated help file on your computer's hard drive.
2. After opening the Zip file, double-click the MegaStat Add-in file, then Extract the MegaStat program to your computer's hard drive. After extracting the file, you can load the MegaStat Add-in to Excel by double-clicking the MegaStat.xls file. When the Excel program opens to load the Add-in, choose the Enable Macros option.
3. After installation of the add-in, you will be able to access MegaStat by selecting the Add-ins tab on the Excel toolbar.
4. If MegaStat is not listed under Add-ins when you reopen the Excel program, then you can access MegaStat by double-clicking the MegaStat.xls file at any time.

Entering Data

1. Select a cell at the top of a column on an Excel worksheet where you want to enter data. When working with data values for a single variable, you will usually want to enter the values into a single column.
2. Type each data value and press **[Enter]** or **[Tab]** on your keyboard.

You can also add more worksheets to an Excel workbook by clicking the Insert Worksheet icon located at the bottom of an open workbook.

Example XL1–1: Opening an existing Excel workbook/worksheet

1. Open the Microsoft Office Excel 2007 program.

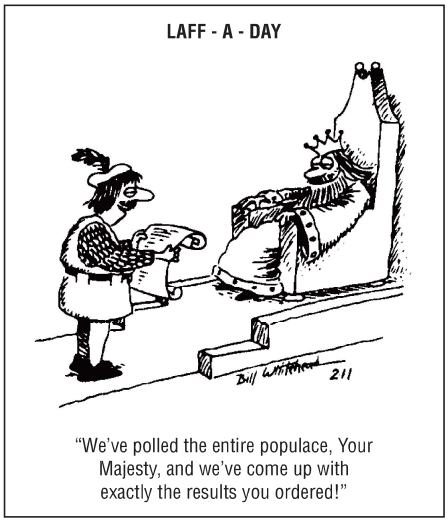
2. Click the Microsoft Click Office button  , then click the Open file function. The Open dialog box will be displayed.
3. In the Look in box, click the folder where the Excel workbook file is located.
4. Double-click the file name in the list box. The selected workbook file will be opened in Excel for editing.

Summary*

- The two major areas of statistics are descriptive and inferential. Descriptive statistics includes the collection, organization, summarization, and presentation of data. Inferential statistics includes making inferences from samples to populations, estimations and hypothesis testing, determining relationships, and making predictions. Inferential statistics is based on *probability theory.* (1–1)
- Data can be classified as qualitative or quantitative. Quantitative data can be either discrete or continuous, depending on the values they can assume. Data can also be measured by various scales. The four basic levels of measurement are nominal, ordinal, interval, and ratio. (1–2)
- Since in most cases the populations under study are large, statisticians use subgroups called samples to get the necessary data for their studies. There are four basic methods used to obtain samples: random, systematic, stratified, and cluster. (1–3)
- There are two basic types of statistical studies: observational studies and experimental studies. When conducting observational studies, researchers observe what is happening or what has happened and then draw conclusions based on these observations. They do not attempt to manipulate the variables in any way. (1–4)
- When conducting an experimental study, researchers manipulate one or more of the independent or explanatory variables and see how this manipulation influences the dependent or outcome variable. (1–4)
- Finally, the applications of statistics are many and varied. People encounter them in everyday life, such as in reading newspapers or magazines, listening to the radio, or watching television. Since statistics is used in almost every field of endeavor, the educated individual should be knowledgeable about the vocabulary, concepts, and procedures of statistics. Also, everyone should be aware that statistics can be misused. (1–5)
- Today, computers and calculators are used extensively in statistics to facilitate the computations. (1–6)

Unusual Stat

The chance that someone will attempt to burglarize your home in any given year is 1 in 20.



"We've polled the entire populace, Your Majesty, and we've come up with exactly the results you ordered!"

*The numbers in parentheses indicate the chapter section where the material is explained.

Important Terms

cluster sample 12
confounding variable 15
continuous variables 6
control group 14
convenience sample 12
data 3
data set 3
data value or datum 3
dependent variable 14
descriptive statistics 4
discrete variables 6
experimental study 14
explanatory variable 14
Hawthorne effect 15
hypothesis testing 4
independent variable 14
inferential statistics 4
interval level of measurement 8
measurement scales 7
nominal level of measurement 7
observational study 13
ordinal level of measurement 8
outcome variable 14
population 4
probability 4
qualitative variables 6
quantitative variables 6
quasi-experimental study 14
random sample 10
random variable 3
ratio level of measurement 8
sample 4
statistics 3
stratified sample 12
systematic sample 11
treatment group 14
variable 3

Answers not appearing on the page can be found in the answers appendix.

Review Exercises

***Note:* All odd-numbered problems and even-numbered problems marked with "ans" are included in the answer section at the end of this book. The numbers in parentheses indicate the chapter section where the process to arrive at a solution is explained.**

1. Name and define the two areas of statistics. (1–1)

2. What is probability? Name two areas where probability is used. (1–1) Probability deals with events that occur by chance. It is used in gambling and insurance.

3. Suggest some ways statistics can be used in everyday life. (1–1) Answers will vary.

4. Explain the differences between a sample and a population. (1–1) A population is the totality of all subjects possessing certain common characteristics that are being studied.

5. Why are samples used in statistics? (1–1)

6. **(ans)** In each of these statements, tell whether descriptive or inferential statistics have been used.

a. By 2040 at least 3.5 billion people will run short of water (World Future Society). Inferential
b. Nine out of ten on-the-job fatalities are men (Source: *USA TODAY Weekend*). Descriptive
c. Expenditures for the cable industry were $5.66 billion in 1996 (Source: *USA TODAY*). Descriptive
d. The median household income for people aged 25–34 is $35,888 (Source: *USA TODAY*). Descriptive
e. Allergy therapy makes bees go away (Source: *Prevention*). Inferential
f. Drinking decaffeinated coffee can raise cholesterol levels by 7% (Source: American Heart Association).
g. The national average annual medicine expenditure per person is $1052 (Source: *The Greensburg Tribune Review*). Descriptive
h. Experts say that mortgage rates may soon hit bottom (Source: *USA TODAY*). (1–1) Inferential

7. Classify each as nominal-level, ordinal-level, interval-level, or ratio-level measurement.

a. Pages in the 25 best-selling mystery novels. Ratio
b. Rankings of golfers in a tournament. Ordinal
c. Temperatures inside 10 pizza ovens. Interval
d. Weights of selected cell phones. Ratio
e. Salaries of the coaches in the NFL. Ratio
f. Times required to complete a chess game. Ratio
g. Ratings of textbooks (poor, fair, good, excellent). Ordinal
h. Number of amps delivered by battery chargers. Ratio
i. Ages of childern in a day care center. Ratio
j. Categories of magazines in a physician's office (sports, women's, health, men's, news). (1–2) Normal

8. Classify each variable as qualitative or quantitative.

a. Marital status of nurses in a hospital. Qualitative
b. Time it takes to run a marathon. Quantitative
c. Weights of lobsters in a tank in a restaurant. Quantitative
d. Colors of automobiles in a shopping center parking lot. Qualitative
e. Ounces of ice cream in a large milkshake. Quantitative
f. Capacity of the NFL football stadiums. Quantitative
g. Ages of people living in a personal care home. (1–2) Quantitative

9. Classify each variable as discrete or continuous.

a. Number of pizzas sold by Pizza Express each day. Discrete
b. Relative humidity levels in operating rooms at local hospitals. Continuous
c. Number of bananas in a bunch at several local supermarkets. Discrete
d. Lifetimes (in hours) of 15 iPod batteries. Continuous
e. Weights of the backpacks of first graders on a school bus. Continuous
f. Number of students each day who make appointments with a math tutor at a local college. Discrete
g. Blood pressures of runners in a marathon. (1–2) Continuous

10. Give the boundaries of each value.

a. 36 inches. 35.5–36.5
b. 105.4 miles. 105.35–105.45
c. 72.6 tons. 72.55–72.65
d. 5.27 centimeters. 5.265–5.275
e. 5 ounces. (1–2) 4.5–5.5

11. Name and define the four basic sampling methods. (1–3) Random, systematic, stratified, cluster

12. **(ans)** Classify each sample as random, systematic, stratified, or cluster.

a. In a large school district, all teachers from two buildings are interviewed to determine whether they believe the students have less homework to do now than in previous years. Cluster
b. Every seventh customer entering a shopping mall is asked to select her or his favorite store. Systematic
c. Nursing supervisors are selected using random numbers to determine annual salaries. Random
d. Every 100th hamburger manufactured is checked to determine its fat content. Systematic
e. Mail carriers of a large city are divided into four groups according to gender (male or female) and according to whether they walk or ride on their routes. Then 10 are selected from each group and interviewed to determine whether they have been bitten by a dog in the last year. (1–3) Stratified

13. Give three examples each of nominal, ordinal, interval, and ratio data. (1–2) Answers will vary.

14. For each of these statements, define a population and state how a sample might be obtained. Answers will vary.

a. The average cost of an airline meal is $4.55 (Source: *Everything Has Its Price,* Richard E. Donley, Simon and Schuster).
b. More than 1 in 4 United States children have cholesterol levels of 180 milligrams or higher (Source: The American Health Foundation).
c. Every 10 minutes, 2 people die in car crashes and 170 are injured (Source: National Safety Council estimates).
d. When older people with mild to moderate hypertension were given mineral salt for 6 months, the average blood pressure reading dropped by 8 points systolic and 3 points diastolic (Source: *Prevention*).
e. The average amount spent per gift for Mom on Mother's Day is $25.95 (Source: The Gallup Organization). (1–3)

15. Select a newspaper or magazine article that involves a statistical study, and write a paper answering these questions. Answers will vary.

a. Is this study descriptive or inferential? Explain your answer.
b. What are the variables used in the study? In your opinion, what level of measurement was used to obtain the data from the variables?
c. Does the article define the population? If so, how is it defined? If not, how could it be defined?
d. Does the article state the sample size and how the sample was obtained? If so, determine the size of the sample and explain how it was selected. If not, suggest a way it could have been obtained.
e. Explain *in your own words* what procedure (survey, comparison of groups, etc.) might have been used to determine the study's conclusions.
f. Do you agree or disagree with the conclusions? State your reasons.

16. Information from research studies is sometimes taken out of context. Explain why the claims of these studies might be suspect. Answers will vary.

a. Based on a recent telephone survey, 72% of those contacted shop online.
b. In Greenville County there are 8324 deer.
c. Nursing school graduates from Fairview University earn on average $33,456.
d. Only 5% of the men surveyed wash the dishes after dinner.
e. A recent study shows that high school dropouts spend less time on the Internet than those who graduated; therefore, the Internet raises your IQ.
f. Most shark attacks occur in ocean water that is 3 feet deep; therefore, it is safer to swim in deep water. (1–5)

17. Identify each study as being either observational or experimental.

a. Subjects were randomly assigned to two groups, and one group was given an herb and the other group a placebo. After 6 months, the numbers of respiratory tract infections each group had were compared. Experimental
b. A researcher stood at a busy intersection to see if the color of the automobile that a person drives is related to running red lights. Observational

c. A researcher finds that people who are more hostile have higher total cholesterol levels than those who are less hostile. Observational

d. Subjects are randomly assigned to four groups. Each group is placed on one of four special diets—a low-fat diet, a high-fish diet, a combination of low-fat diet and high-fish diet, and a regular diet. After 6 months, the blood pressures of the groups are compared to see if diet has any effect on blood pressure. (1–4) Experimental

18. Identify the independent variable(s) and the dependent variable for each of the studies in Exercise 17. (1–4)

19. For each of the studies in Exercise 17, suggest possible confounding variables. (1–4)

20. Beneficial Bacteria According to a pilot study of 20 people conducted at the University of Minnesota, daily doses of a compound called arabinogalactan over a period of 6 months resulted in a significant increase in the beneficial lactobacillus species of bacteria. Why can't it be concluded that the compound is beneficial for the majority of people? (1–5) Only 20 people were used in the study.

21. Comment on the following statement, taken from a magazine advertisement: "In a recent clinical study, Brand ABC (actual brand will not be named) was proved to be 1950% better than creatine!" (1–5) The only time claims can be proved is when the entire population is used.

22. In an ad for women, the following statement was made: "For every 100 women, 91 have taken the road less traveled." Comment on this statement. (1–5)

23. In many ads for weight loss products, under the product claims and in small print, the following statement is made: "These results are not typical." What does this say about the product being advertised? (1–5)

24. In an ad for moisturizing lotion, the following claim is made: ". . . it's the number 1 dermatologist-recommended brand." What is misleading about this claim? (1–5) There is no mention of how this conclusion was obtained.

25. An ad for an exercise product stated: "Using this product will burn 74% more calories." What is misleading about this statement? (1–5) "74% more calories" than what? No comparison group is stated.

26. "Vitamin E is a proven antioxidant and may help in fighting cancer and heart disease." Is there anything ambiguous about this claim? Explain. (1–5) Since the word *may* is used, there is no guarantee that the product will help fight cancer.

27. "Just 1 capsule of Brand X can provide 24 hours of acid control." (Actual brand will not be named.) What needs to be more clearly defined in this statement? (1–5) What is meant by "24 hours of acid control"?

28. ". . . Male children born to women who smoke during pregnancy run a risk of violent and criminal behavior that lasts well into adulthood." Can we infer that smoking during pregnancy is responsible for criminal behavior in people? (1–5) No. There are many other factors that contribute to criminal behavior.

29. Caffeine and Health In the 1980s, a study linked coffee to a higher risk of heart disease and pancreatic cancer. In the early 1990s, studies showed that drinking coffee posed minimal health threats. However, in 1994, a study showed that pregnant women who drank 3 or more cups of tea daily may be at risk for spontaneous abortion. In 1998, a study claimed that women who drank more than a half-cup of caffeinated tea every day may actually increase their fertility. In 1998, a study showed that over a lifetime, a few extra cups of coffee a day can raise blood pressure, heart rate, and stress (Source: "Bottom Line: Is It Good for You? Or Bad?" by Monika Guttman, *USA TODAY Weekend*). Suggest some reasons why these studies appear to be conflicting. (1–5) Possible answer: It could be the amount of caffeine in the coffee or tea. It could have been the brewing method.

Extending the Concepts

30. Find an article that describes a statistical study, and identify the study as observational or experimental. Answers will vary.

31. For the article that you used in Exercise 30, identify the independent variable(s) and dependent variable for the study. Answers will vary.

32. For the article that you selected in Exercise 30, suggest some confounding variables that may have an effect on the results of the study. Answers will vary.

Statistics Today

Are We Improving Our Diet?–Revisited

Researchers selected a *sample* of 23,699 adults in the United States, using phone numbers selected at *random,* and conducted a *telephone survey.* All respondents were asked six questions:

1. How often do you drink juices such as orange, grapefruit, or tomato?
2. Not counting juice, how often do you eat fruit?
3. How often do you eat green salad?
4. How often do you eat potatoes (not including french fries, fried potatoes, or potato chips)?
5. How often do you eat carrots?
6. Not counting carrots, potatoes, or salad, how many servings of vegetables do you usually eat?

Researchers found that men consumed fewer servings of fruits and vegetables per day (3.3) than women (3.7). Only 20% of the population consumed the recommended 5 or more daily servings. In addition, they found that youths and less-educated people consumed an even lower amount than the average.

Based on this study, they recommend that greater educational efforts be undertaken to improve fruit and vegetable consumption by Americans and to provide environmental and institutional support to encourage increased consumption.

Source: Mary K. Serdula, M.D., et al., "Fruit and Vegetable Intake Among Adults in 16 States: Results of a Brief Telephone Survey," *American Journal of Public Health* 85, no. 2. Copyright by the American Public Health Association.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. Probability is used as a basis for inferential statistics. True
2. The heights of the mountains in the state of Alaska are an example of a variable. True
3. The lowest level of measurement is the nominal level. True
4. When the population of college professors is divided into groups according to their rank (instructor, assistant professor, etc.) and then several are selected from each group to make up a sample, the sample is called a cluster sample. False
5. The variable temperature is an example of a quantitative variable. True
6. The height of basketball players is considered a continuous variable. True
7. The boundary of a value such as 6 inches would be 5.9–6.1 inches. False

Select the best answer.

8. The number of ads on a one-hour television show is what type of data?
   a. Nominal
   b. Qualitative
   c. Discrete
   d. Continuous
9. What are the boundaries of 25.6 ounces?
   a. 25–26 ounces
   b. 25.55–25.65 ounces
   c. 25.5–25.7 ounces
   d. 20–39 ounces
10. A researcher divided subjects into two groups according to gender and then selected members from each group for her sample. What sampling method was the researcher using?
    a. Cluster
    b. Random
    c. Systematic
    d. Stratified

11. Data that can be classified according to color are measured on what scale?

(a.) Nominal
b. Ratio
c. Ordinal
d. Interval

12. A study that involves no researcher intervention is called

a. An experimental study.
b. A noninvolvement study.
(c.) An observational study.
d. A quasi-experimental study.

13. A variable that interferes with other variables in the study is called

(a.) A confounding variable.
b. An explanatory variable.
c. An outcome variable.
d. An interfering variable.

Use the best answer to complete these statements.

14. Two major branches of statistics are _____ and _____. Descriptive, inferential

15. Two uses of probability are _____ and _____. Gambling, insurance

16. The group of all subjects under study is called a(n) _____. Population

17. A group of subjects selected from the group of all subjects under study is called a(n) _____. Sample

18. Three reasons why samples are used in statistics are
a. _____ *b.* _____ *c.* _____.
a. Saves time *b.* Saves money *c.* Use when population is infinite

19. The four basic sampling methods are
a. _____ *b.* _____ *c.* _____ *d.* _____.
a. Random *b.* Systematic *c.* Cluster *d.* Stratified

20. A study that uses intact groups when it is not possible to randomly assign participants to the groups is called a(n) _____ study. Quasi-experimental

21. In a research study, participants should be assigned to groups using _____ methods, if possible. Random

22. For each statement, decide whether descriptive or inferential statistics is used.

a. The average life expectancy in New Zealand is 78.49 years (Source: *World Factbook*). Descriptive
b. A diet high in fruits and vegetables will lower blood pressure (Source: Institute of Medicine). Inferential
c. The total amount of estimated losses for Hurricane Katrina was $125 billion (Source: *The World Almanac and Book of Facts*). Descriptive
d. Researchers stated that the shape of a person's ears is relative to the person's aggression (Source: *American Journal of Human Biology*). Inferential
e. In 2013, the number of high school graduates will be 3.2 million students (Source: National Center for Education). Inferential

23. Classify each as nominal level, ordinal level, interval level, or ratio level of measurement.

a. Rating of movies as G, PG, and R Nominal
b. Number of candy bars sold on a fund drive Ratio
c. Classification of automobiles as subcompact, compact, standard, and luxury Ordinal
d. Temperatures of hair dryers Interval
e. Weights of suitcases on a commercial airliner Ratio

24. Classify each variable as discrete or continuous.

a. Ages of people working in a large factory Continuous
b. Number of cups of coffee served at a restaurant Discrete
c. The amount of drug injections into a guinea pig Continuous
d. The time it takes a student to drive to school Continuous
e. The number of gallons of milk sold each day at a grocery store Discrete

25. Give the boundaries of each.

a. 32 minutes 31.5–32.5 minutes
b. 0.48 millimeter 0.475–0.485 millimeter
c. 6.2 inches 6.15–6.25 inches
d. 19 pounds 18.5–19.5 pounds
e. 12.1 quarts 12.05–12.15 quarts

Critical Thinking Challenges

1. **World's Busiest Airports** A study of the world's busiest airports was conducted by *Airports Council International*. Describe three variables that one could use to determine which airports are the busiest. What *units* would one use to measure these variables? Are these variables categorical, discrete, or continuous?
2. **Smoking and Criminal Behavior** The results of a study published in *Archives of General Psychiatry* stated that male children born to women who smoke during pregnancy run a risk of violent and criminal behavior that lasts into adulthood. The results of this study were challenged by some people in the media. Give several reasons why the results of this study would be challenged.
3. **Piano Lessons Improve Math Ability** The results of a study published in *Neurological Research* stated that second-graders who took piano lessons and played a computer math game more readily grasped math problems in fractions and proportions than a similar group who took an English class and played the same math game. What type of inferential study was this? Give several reasons why the piano lessons could improve a student's math ability.
4. **ACL Tears in Collegiate Soccer Players** A study of 2958 collegiate soccer players showed that in 46 anterior cruciate ligament (ACL) tears, 36 were in women. Calculate the percentages of tears for each gender.
   a. Can it be concluded that female athletes tear their knees more often than male athletes?
   b. Comment on how this study's conclusion might have been reached.
5. **Anger and Snap Judgments** Read the article entitled "Anger Can Cause Snap Judgments" and answer the following questions.
   a. Is the study experimental or observational?
   b. What is the independent variable?
   c. What is the dependent variable?
   d. Do you think the sample sizes are large enough to merit the conclusion?
   e. Based on the results of the study, what changes would you recommend to persons to help them reduce their anger?
6. **Hostile Children Fight Unemployment** Read the article entitled "Hostile Children Fight Unemployment" and answer the following questions.
   a. Is the study experimental or observational?
   b. What is the independent variable?
   c. What is the dependent variable?
   d. Suggest some confounding variables that may have influenced the results of the study.
   e. Identify the three groups of subjects used in the study.

ANGER CAN CAUSE SNAP JUDGMENTS

Anger can make a normally unbiased person act with prejudice, according to a forthcoming study in the journal *Psychological Science*.

Assistant psychology professors David DeSteno at Northeastern University in Boston and Nilanjana Dasgupta at the University of Massachusetts, Amherst, randomly divided 81 study participants into two groups and assigned them a writing task designed to induce angry, sad or neutral feelings. In a subsequent test to uncover nonconscious associations, angry subjects were quicker to connect negatively charged words—like war, death and vomit—with members of the opposite group—even though the groupings were completely arbitrary.

"These automatic responses guide our behavior when we're not paying attention," says DeSteno, and they can lead to discriminatory acts when there is pressure to make a quick decision. "If you're aware that your emotions might be coloring these gut reactions," he says, "you should take time to consider that possibility and adjust your actions accordingly."

—*Eric Strand*

Source: Reprinted with permission from *Psychology Today*, Copyright © (2004) Sussex Publishers, Inc.

UNEMPLOYMENT

Hostile Children Fight Unemployment

Aggressive children may be destined for later long-term unemployment. In a study that began in 1968, researchers at the University of Jyvaskyla in Finland examined about 300 participants at ages 8, 14, 27, and 36. They looked for aggressive behaviors like hurting other children, kicking objects when angry, or attacking others without reason.

Their results, published recently in the *International Journal of Behavioral Development*, suggest that children with low self-control of emotion —especially aggression—were significantly more prone to long-term unemployment. Children with behavioral inhibitions—such as passive and anxious behaviors—were also indirectly linked to unemployment as they lacked the preliminary initiative needed for school success. And while unemployment rates were high in Finland during the last data collection, jobless participants who were aggressive as children were less likely to have a job two years later than their nonaggressive counterparts.

Ongoing unemployment can have serious psychological consequences, including depression, anxiety and stress. But lead researcher Lea Pulkkinen, Ph.D., a Jyvaskyla psychology professor, does have encouraging news for parents: Aggressive children with good social skills and child-centered parents were significantly less likely to be unemployed for more than two years as adults.

—*Tanya Zimbardo*

Source: Reprinted with permission from *Psychology Today,* Copyright © (2001) Sussex Publishers, Inc.

Data Projects

1. **Business and Finance** Investigate the types of data that are collected regarding stock and bonds, for example, price, earnings ratios, and bond ratings. Find as many types of data as possible. For each, identify the level of measure as nominal, ordinal, interval, or ratio. For any quantitative data, also note if they are discrete or continuous.

2. **Sports and Leisure** Select a professional sport. Investigate the types of data that are collected about that sport, for example, in baseball, the level of play (A, AA, AAA, Major League), batting average, and home-run hits. For each, identify the level of measure as nominal, ordinal, interval, or ratio. For any quantitative data, also note if they are discrete or continuous.

3. **Technology** Music organization programs on computers and music players maintain information about a song, such as the writer, song length, genre, and your personal rating. Investigate the types of data collected about a song. For each, identify the level of measure as nominal, ordinal, interval, or ratio. For any quantitative data, also note if they are discrete or continuous.

4. **Health and Wellness** Think about the types of data that can be collected about your health and wellness, things such as blood type, cholesterol level, smoking status, and BMI. Find as many data items as you can. For each, identify the level of measure as nominal, ordinal, interval, or ratio. For any quantitative data, also note if they are discrete or continuous.

5. **Politics and Economics** Every 10 years since 1790, the federal government has conducted a census of U.S. residents. Investigate the types of data that were collected in the 2010 census. For each, identify the level of measure as nominal, ordinal, interval, or ratio. For any quantitative data, also note if they are discrete or continuous. Use the library or a genealogy website to find a census form from 1860. What types of data were collected? How do the types of data differ?

6. **Your Class** Your school probably has a database that contains information about each student, such as age, county of residence, credits earned, and ethnicity. Investigate the types of student data that your college collects and reports. For each, identify the level of measure as nominal, ordinal, interval, or ratio. For any quantitative data, also note if they are discrete or continuous.

Answers to Applying the Concepts

Section 1–1 Attendance and Grades

1. The variables are grades and attendance.
2. The data consist of specific grades and attendance numbers.
3. These are descriptive statistics; however, if an inference were made to all students, then that would be inferential statistics.
4. The population under study is students at Manatee Community College (MCC).
5. While not specified, we probably have data from a sample of MCC students.
6. Based on the data, it appears that, in general, the better your attendance, the higher your grade.

Section 1–2 Safe Travel

1. The variables are industry and number of job-related injuries.
2. The type of industry is a qualitative variable, while the number of job-related injuries is quantitative.
3. The number of job-related injuries is discrete.
4. Type of industry is nominal, and the number of job-related injuries is ratio.
5. The railroads do show fewer job-related injuries; however, there may be other things to consider. For example, railroads employ fewer people than the other transportation industries in the study.
6. A person's choice of transportation might also be affected by convenience issues, cost, service, etc.
7. Answers will vary. One possible answer is that the railroads have the fewest job-related injuries, while the airline industry has the most job-related injuries (more than twice those of the railroad industry). The numbers of job-related injuries in the subway and trucking industries are fairly comparable.

Section 1–3 American Culture and Drug Abuse

Answers will vary, so this is one possible answer.

1. I used a telephone survey. The advantage to my survey method is that this was a relatively inexpensive survey method (although more expensive than using the mail) that could get a fairly sizable response. The disadvantage to my survey method is that I have not included anyone without a telephone. (*Note:* My survey used a random dialing method to include unlisted numbers and cell phone exchanges.)
2. A mail survey also would have been fairly inexpensive, but my response rate may have been much lower than what I got with my telephone survey. Interviewing would have allowed me to use follow-up questions and to clarify any questions of the respondents at the time of the interview. However, interviewing is very labor- and cost-intensive.
3. I used ordinal data on a scale of 1 to 5. The scores were 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree.
4. The random method that I used was a random dialing method.
5. To include people from each state, I used a stratified random sample, collecting data randomly from each of the area codes and telephone exchanges available.
6. This method allowed me to make sure that I had representation from each area of the United States.
7. Convenience samples may not be representative of the population, and a convenience sample of adolescents would probably differ greatly from the general population with regard to the influence of American culture on illegal drug use.

Section 1–4 Just a Pinch Between Your Cheek and Gum

1. This was an experiment, since the researchers imposed a treatment on each of the two groups involved in the study.
2. The independent variable is whether the participant chewed tobacco or not. The dependent variables are the students' blood pressures and heart rates.
3. The treatment group is the tobacco group—the other group was used as a control.
4. A student's blood pressure might not be affected by knowing that he or she was part of a study. However, if the student's blood pressure were affected by this knowledge, all the students (in both groups) would be affected similarly. This might be an example of the placebo effect.
5. Answers will vary. One possible answer is that confounding variables might include the way that the students chewed the tobacco, whether or not the students smoked (although this would hopefully have been evened out with the randomization), and that all the participants were university students.
6. Answers will vary. One possible answer is that the study design was fine, but that it cannot be generalized beyond the population of university students (or people around that age).

This page intentionally left blank

CHAPTER 2

Frequency Distributions and Graphs

Objectives

After completing this chapter, you should be able to

1. Organize data using a frequency distribution.
2. Represent data in frequency distributions graphically using histograms, frequency polygons, and ogives.
3. Represent data using bar graphs, Pareto charts, time series graphs, and pie graphs.
4. Draw and interpret a stem and leaf plot.

Outline

Statistics Today

How Your Identity Can Be Stolen

Identity fraud is a big business today. The total amount of the fraud in 2006 was $56.6 billion. The average amount of the fraud for a victim is $6383, and the average time to correct the problem is 40 hours. The ways in which a person's identity can be stolen are presented in the following table:

Lost or stolen wallet, checkbook, or credit card	38%
Friends, acquaintances	15
Corrupt business employees	15
Computer viruses and hackers	9
Stolen mail or fraudulent change of address	8
Online purchases or transactions	4
Other methods	11

Source: Javelin Strategy & Research; Council of Better Business Bureau, Inc.

Looking at the numbers presented in a table does not have the same impact as presenting numbers in a well-drawn chart or graph. The article did not include any graphs. This chapter will show you how to construct appropriate graphs to represent data and help you to get your point across to your audience.

See Statistics Today—Revisited at the end of the chapter for some suggestions on how to represent the data graphically.

Introduction

When conducting a statistical study, the researcher must gather data for the particular variable under study. For example, if a researcher wishes to study the number of people who were bitten by poisonous snakes in a specific geographic area over the past several years, he or she has to gather the data from various doctors, hospitals, or health departments.

To describe situations, draw conclusions, or make inferences about events, the researcher must organize the data in some meaningful way. The most convenient method of organizing data is to construct a *frequency distribution.*

After organizing the data, the researcher must present them so they can be understood by those who will benefit from reading the study. The most useful method of presenting the data is by constructing *statistical charts* and *graphs.* There are many different types of charts and graphs, and each one has a specific purpose.

This chapter explains how to organize data by constructing frequency distributions and how to present the data by constructing charts and graphs. The charts and graphs illustrated here are histograms, frequency polygons, ogives, pie graphs, Pareto charts, and time series graphs. A graph that combines the characteristics of a frequency distribution and a histogram, called a stem and leaf plot, is also explained.

2–1

Organizing Data

Objective 1

Organize data using a frequency distribution.

Wealthy People

Suppose a researcher wished to do a study on the ages of the top 50 wealthiest people in the world. The researcher first would have to get the data on the ages of the people. In this case, these ages are listed in *Forbes Magazine*. When the data are in original form, they are called **raw data** and are listed next.

49	57	38	73	81
74	59	76	65	69
54	56	69	68	78
65	85	49	69	61
48	81	68	37	43
78	82	43	64	67
52	56	81	77	79
85	40	85	59	80
60	71	57	61	69
61	83	90	87	74

Since little information can be obtained from looking at raw data, the researcher organizes the data into what is called a *frequency distribution.* A frequency distribution consists of *classes* and their corresponding *frequencies.* Each raw data value is placed into a quantitative or qualitative category called a **class.** The **frequency** of a class then is the number of data values contained in a specific class. A frequency distribution is shown for the preceding data set.

Class limits	Tally	Frequency
35–41	///	3
42–48	///	3
49–55	////	4
56–62	𝍸 𝍸	10
63–69	𝍸 𝍸	10
70–76	𝍸	5
77–83	𝍸 𝍸	10
84–90	𝍸	5
	Total	50

Now some general observations can be made from looking at the frequency distribution. For example, it can be stated that the majority of the wealthy people in the study are over 55 years old.

Unusual Stat

Of Americans 50 years old and over, 23% think their greatest achievements are still ahead of them.

A **frequency distribution** is the organization of raw data in table form, using classes and frequencies.

The classes in this distribution are 35–41, 42–48, etc. These values are called *class limits.* The data values 35, 36, 37, 38, 39, 40, 41 can be tallied in the first class; 42, 43, 44, 45, 46, 47, 48 in the second class; and so on.