2

Describing, Exploring, and Comparing Data

Important note: The topics of this chapter require that you use Minitab to enter data, retrieve data, save files, and print results. These functions are covered in Chapter 1 of this manual/ workbook. Be sure to understand those functions from Chapter 1 before beginning this chapter.

The main objective of Chapter 2 in *Elementary Statistics* and *Essentials of Statistics* is to introduce the tools needed to describe, explore, or compare those characteristics of a data set that are extremely important.

Important Characteristics of Data

When describing, exploring, and comparing data sets, the following characteristics are usually extremely important:

1. **Center:** Measure of center, which is a representative or average value that gives us an indication of where the middle of the data set is located

2. **Variation:** A measure of the amount that the values vary among themselves

3. **Distribution:** The nature or shape of the distribution of the data, such as bell-shaped, uniform, or skewed

4. **Outliers:** Sample values that are very far away from the vast majority of the other sample values

5. **Time:** Changing characteristics of the data over time

In this chapter, we learn how to use Minitab as a tool for investigating the above important characteristics.

Chapter 2 in *Elementary Statistics* or *Essentials of Statistics* begins with a Chapter Problem related to the issue of secondhand smoke. That Chapter Problem includes the measured levels of cotinine for three samples of people, and those measurements are included in Table 2–1, which is reproduced on the following page. One sample consists of smokers, the second sample (ETS) consists of people who do not smoke, but are exposed to environmental tobacco smoke at home or work, and the third sample (NOETS) consists of nonsmokers who are not exposed to environmental tobacco smoke. The data are included in Data Set 6 of Appendix B in *Elementary Statistics* or *Essentials of Statistics*. The data are also included in the Minitab worksheet COTININE.MTW that is included on the CD–ROM packaged with the textbook, and the data are also available at http://www.aw.com/triola. The textbook noted that cotinine is a metabolite of nicotine, meaning that cotinine is produced in the body when nicotine is absorbed. Because it is known that nicotine is absorbed through cigarette smoking, we have a way of measuring the effective presence of cigarette smoke indirectly by measuring the amount of cotinine that is present.

Chapter 2 in the textbook presents techniques for describing, exploring, and comparing data such as those included in Table 2–1. In this manual/workbook we show how Minitab can be used.

Table 2–1 Measured Cotinine Levels in Three Groups

SMOKER

1	0	131	173	265	210	44	277	32	3
35	112	477	289	227	103	222	149	313	491
130	234	164	198	17	253	87	121	266	290
123	167	250	245	48	86	284	1	208	173

ETS

384	0	69	19	1	0	178	2	13	1
4	0	543	17	1	0	51	0	197	3
0	3	1	45	13	3	1	1	1	0
0	551	2	1	1	1	0	74	1	241

NOETS

0	0	0	0	0	0	0	0	0	0
0	9	0	0	0	0	0	0	244	0
1	0	0	0	90	1	0	309	0	0
0	0	0	0	0	0	0	0	0	0

2-1 Histograms

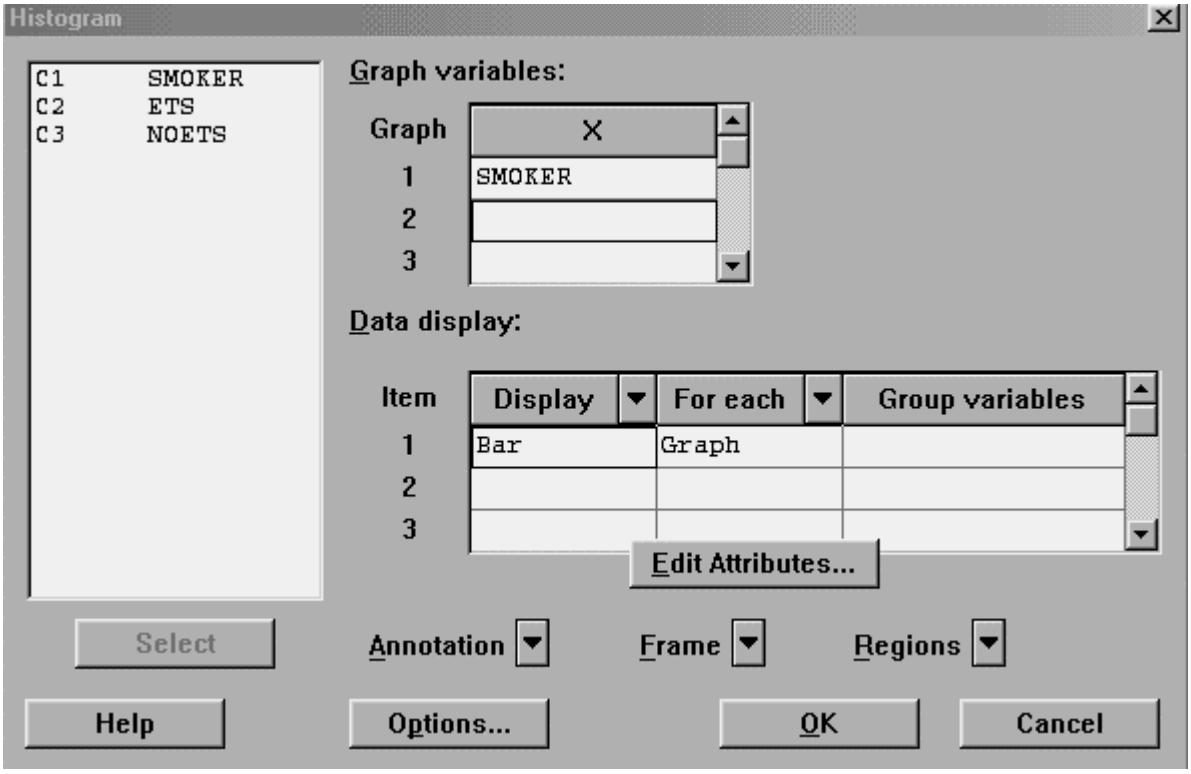
Section 2-3 of *Elementary Statistics* or *Essentials of Statistics* describes histograms and provides detailed procedures for constructing them. It is noted that a histogram is an excellent device for exploring the *distribution* of a data set. Here is the Minitab procedure for generating a histogram.

1. Enter the data in a Minitab column, such as column C1. If the data are already stored in a Minitab worksheet, retrieve the worksheet using the procedure described in Section 1-4 of this manual/workbook.

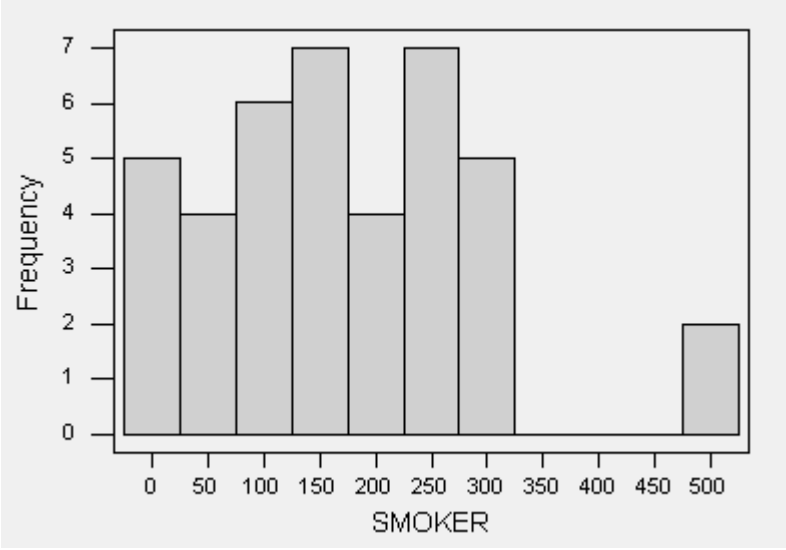
2. Click on the main menu item of **Graph**.

3. Select **Histogram** from the subdirectory.

4. In the dialog box, click on the column with the data so that the column label (such as C1) or the data name (such as SMOKER) appears in the box as shown.

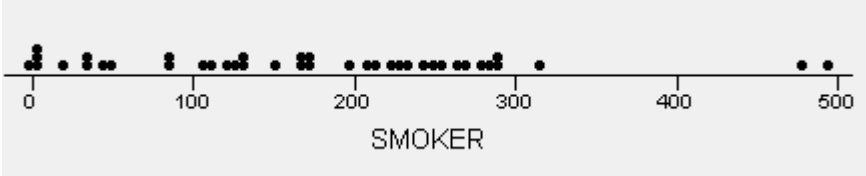


5. If you click on the **Options** bar, you can select a relative frequency histogram or a cumulative frequency histogram (or ogive). You can also change the way that the histogram is constructed. (You can specify the number of intervals to be used, or you can specify the locations of the interval midpoints or endpoints.) Using Minitab with the SMOKER data in Table 2-1 results in the following display. The initial Minitab display was improved by using the **Edit Attributes** option to shade the bars for a better visual appearance.



2-2 Dotplots

Section 2-3 of the textbook discusses dotplots. Shown below is the Minitab dotplot display of the Table 2-1 SMOKER data.



Here is the Minitab procedure for creating such a dotplot.

1. Enter or retrieve the data into a Minitab column.
2. Click on the main menu item of **Graph**.
3. Select **Dotplot**.
4. The dialog box will list the columns of data at the left. Click on the desired column.
5. Click **OK**.

2-3 Stem-and-Leaf Plots

Minitab can provide stem-and-leaf plots. The textbook notes that a **stem-and-leaf plot** represents data by separating each value into two parts: the stem (such as the leftmost digit) and the leaf (such as the rightmost digit). Here is the Minitab procedure for generating a stem-and-leaf plot.

1. Enter or retrieve the data into a Minitab column.
2. Click on the main menu item of **Graph**.
3. Select **Stem-and-Leaf**.
4. The dialog box lists columns of data at the left. Click on the desired column.
5. Click **OK**.

The textbook includes an illustration using the movie lengths (in minutes) listed in Data Set 7 of Appendix B and stored in the Minitab worksheet CHMOVIE.MTW. Here is the Minitab display that results from those values (with the actual stem–and–leaf plot highlighted with a bold font).

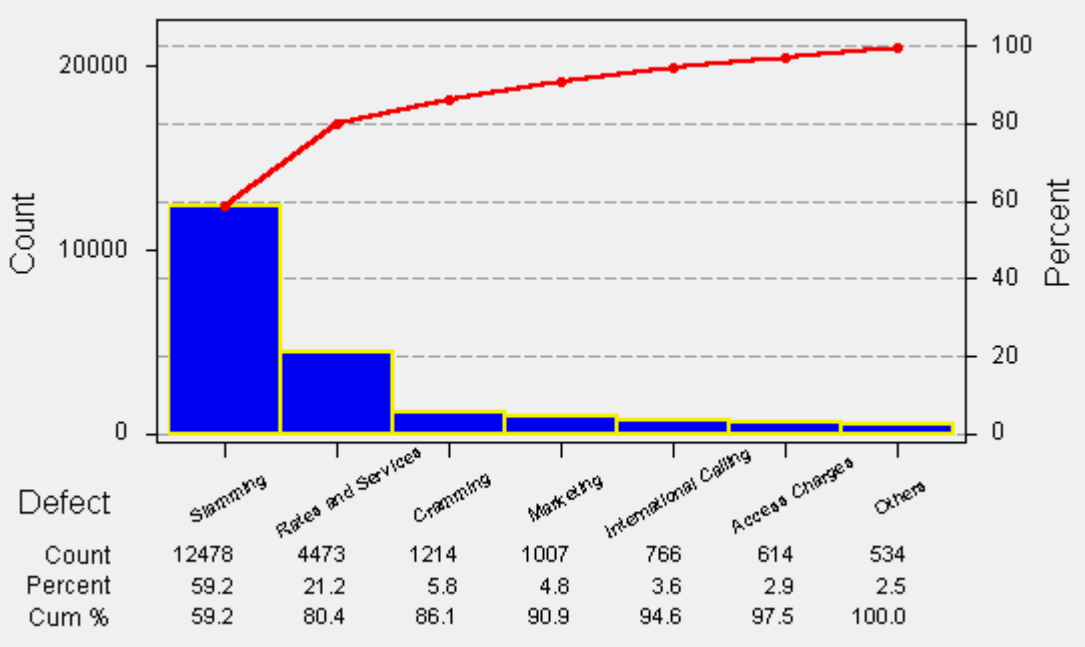
```
Stem-and-leaf of Length   N  = 50
Leaf Unit = 1.0
    2     6 44
    3     6 9
   14     7 01112334444
  (15)    7 555555666778899
   21     8 001112223334
    9     8 6899
    5     9 0024
    1     9
    1    10
    1    10
    1    11
    1    11
    1    12 0
```

Reading and Interpreting a Minitab Stem–and–Leaf Plot: You can see from the above display that Minitab has actually generated the *expanded* stem-and-leaf plot described in the textbook. In addition to the expanded stem–and–leaf plot included in the textbook, Minitab provides another column of data at the extreme left as shown above. In this Minitab display, *the leftmost column represents cumulative totals*. The left column above shows that there are 2 sample values included between 60 and 64; there are 3 values between 60 and 69; there are 14 values between 60 and 74. The left column entry of (15) indicates that there are 15 data values in the row containing the median. The left-column entries below the median row represent cumulative totals from the bottom up, so that the 1 at the bottom indicates that there is one value between 120 and 124. Similarly, the 5 in the left column indicates that there are five values between 90 and 124.

2-4 Pareto Charts

Section 2-3 of the textbook describes a Pareto chart as a bar graph for categorical data, with the bars arranged in order according to frequencies. The tallest bar is at the left, and the smaller bars are farther to the right. Figure 2–6 in the textbook is a Pareto chart based on the data in the table shown below. The Minitab Pareto chart follows the table. The line graph at the top of Minitab's Pareto chart represents cumulative totals.

Phone Complaint	Number
Rates and Services	4,473
Marketing	1,007
International Calling	766
Access Charges	614
Operator Services	534
Slamming	12,748
Cramming	1,214



Here is the procedure for using Minitab to construct a Pareto chart:

1. Enter the labels (such as Rates and Services, Marketing, etc.) in column C1.

2. Enter the corresponding frequency counts in column C2. The worksheet for the above table should appear as shown below.

↓	C1-T	C2
	Complaint	Frequency
1	Rates and Services	4473
2	Marketing	1007
3	International Calling	766
4	Access Charges	614
5	Operator Services	534
6	Slamming	12478
7	Cramming	1214

3. Select the menu items of **Stat**, **Quality Tools**, and **Pareto Chart**.

4. When the dialog box is displayed, select the "Charts defect table" option and enter C1 in the labels box and C2 in the frequency box.

5. Click **OK**.

2-5 Pie Charts

Figure 2-7 in the textbook is a pie chart representing the same data used for the Pareto chart. (See Section 2-4 of this manual/workbook.) To obtain a pie chart from Minitab, follow this procedure.

1. Enter the labels (such as Rates and Services, Marketing, etc.) in column C1.

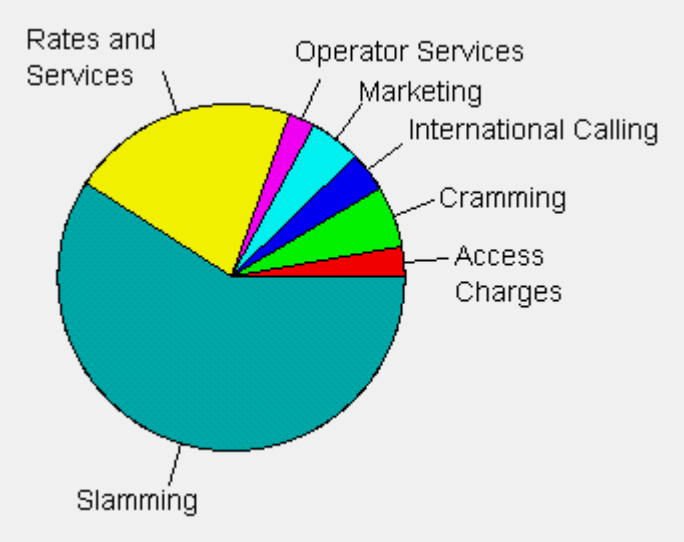
2. Enter the corresponding frequency counts in column C2. (See the Minitab screen display on the bottom of the preceding page.)

3. Select the main menu item of **Graph**.

4. Select **Pie Chart**.

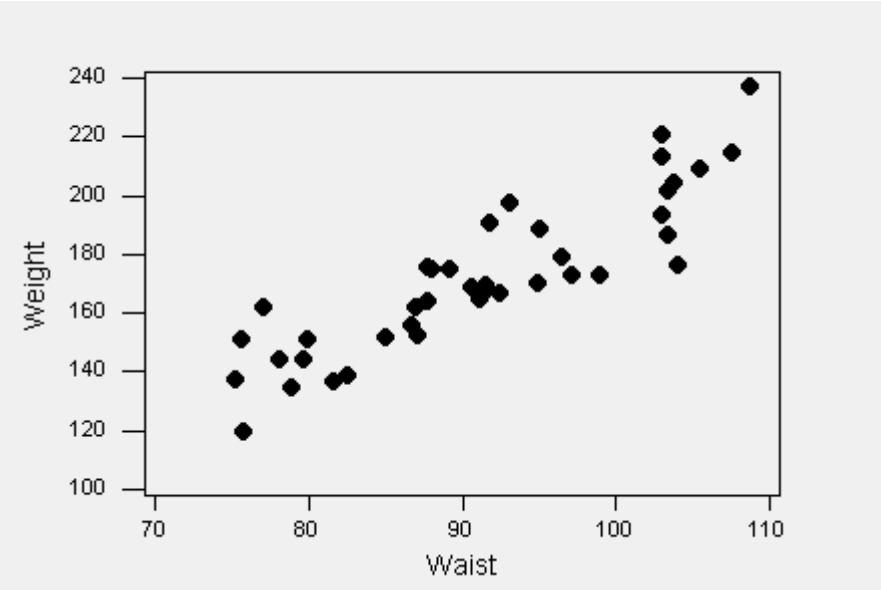
5. When the dialog box is displayed, select the "Charts defect table" option and enter C1 in the labels box and C2 in the frequency box.

6. Click **OK**. Shown below is the Minitab pie chart corresponding to Figure 2-7 in the textbook.



2-6 Scatter Diagrams

Section 2–3 of *Elementary Statistics* or *Essentials of Statistics* includes a scatter diagram (or scatterplot) generated by Minitab. A scatter diagram can be very helpful in seeing a relationship between two variables. The same Minitab display included in the textbook is reproduced below, and this graph shows that larger waist sizes of males appear to correspond to larger weights. (The sample data consist of the paired waist/weight measurements for males listed in Data Set 1 from Appendix B of the textbook.)



Here is the Minitab procedure for generating a scatter diagram.

1. Given a collection of paired data, enter the values for one of the variables in column C1, and enter the corresponding values for the second variable in column C2. (You can also use any two Minitab columns other than C1 and C2.) Be careful to enter the two columns of data so that they are matched in the same way that they are paired in the original listing.

2. Click on **Graph** from the main menu.

3. Select the subdirectory item of **Plot**.

4. In the dialog box, enter C1 in the box for the Y variable, then enter C2 in the box for the X variable (or vice versa).

5. Click **OK**.

A scatter diagram can also be plotted by selecting **Graph**, then **Character Graph**, then **Scatter Plot**, but the displayed points will be depicted using only the standard keyboard characters. A third approach is to select **Stat**, then **Regression**, then **Fitted Line Plot**. We will use this third option in Chapter 9 when we discuss paired data in much greater detail. For now, we are simply generating the scatter diagram to visually explore whether there is an obvious pattern that might reveal some relationship between the two variables.

2-7 Time–Series Graph

Time–series data are data that have been collected at different points in time, and Minitab can graph such data so that patterns become easier to recognize. Here is the Minitab procedure for generating a time–series graph.

1. Given a collection of time–series data, enter the sequential values in column C1.

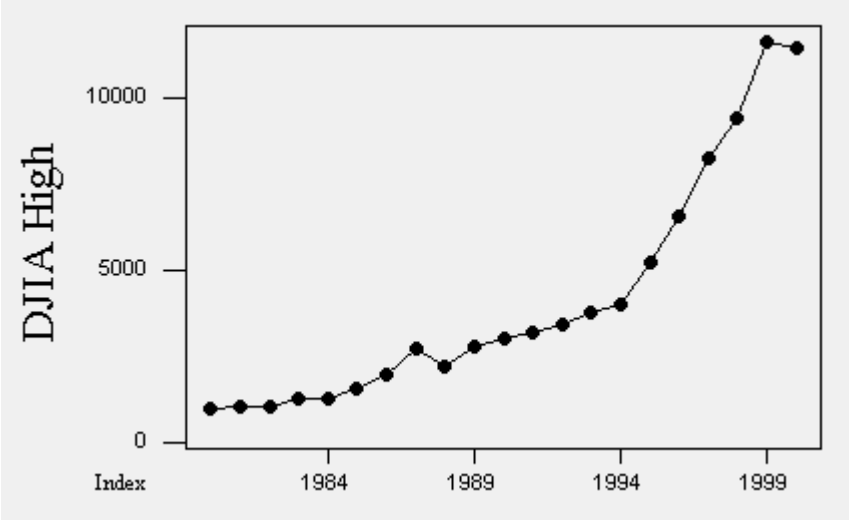
2. Select the main menu item of **Graph**.

3. Select the menu item of **Time Series Plot**.

4. A dialog box now appears. Enter C1 in the first cell under Y in the "Graph" list. Also click the **Options** button to enter a start time to be used for the horizontal axis.

5. Click **OK** twice.

Data Set 25 in Appendix B of *Elementary Statistics* or *Essentials of Statistics* includes a listing of the stock market's Dow Jones Industrial Average (DJIA) annual high values for 1980–2000. The Minitab time–series graph is shown below, and the pattern of data is clearly one of increasing values, with departures from the rising trend occurring in 1987 and 2000.



2-8 Descriptive Statistics

Given a data set, such as the 40 cotinine levels of smokers in Table 2-1 (reproduced near the beginning of this chapter) we can use Minitab to obtain descriptive statistics, including the mean, standard deviation, and quartiles. Here is the Minitab procedure.

1. Enter or retrieve the data in a Minitab column, such as C1. (See Section 1-2.)
2. Click on the main menu item of **Stat**.
3. Click on the subdirectory item of **Basic Statistics**.
4. Select **Display Descriptive Statistics**.
5. In the "Display Descriptive Statistics" dialog box that pops up, click on the column containing the data that you are investigating. (You can also click on the **Graphs** bar to generate certain graphs, including a histogram and dotplot.)
6. Click **OK**.

If you enter the 40 cotinine levels of smokers listed in Table 2-1 and follow the above procedure, you will get the Minitab display shown below.

```
Descriptive Statistics: SMOKER


Variable        N    Mean  Median  TrMean   StDev   SE Mean
SMOKER         40   172.5   170.0   164.7   119.5      18.9

Variable       Minimum    Maximum       Q1       Q3
SMOKER             0.0      491.0     86.3   252.3
```

From the above Minitab display, we see that the number of values is 40, the mean is 172.5, the median is 170.0, the standard deviation is 119.5, the minimum is 0.0, the maximum is 491.0, the first quartile is 86.3, and the third quartile is 252.3. The label of "TrMean" represents *trimmed mean*, described in Exercise 21 in Section 2-4 of the textbook. The label of "SE Mean" represents *standard error of the mean*, and is given in Section 5-5 of the textbook.

Important note: The textbook uses a simplified procedure for finding quartiles, so the quartiles Q_1 and Q_3 found with Minitab may differ slightly from those found by using the textbook method.

2-9 *z* Scores

Section 2-6 of the textbook describes *z* scores (or standard scores). For sample data with mean $\bar{x}$ and standard deviation *s*, the *z* score can be found for a sample value *x* by computing $z = (x - \bar{x})/s$. Here is the Minitab procedure for finding *z* scores that correspond to a set of *x* values. (This procedure assumes that the data set is in column C1, but any other Minitab column can be used instead.)

1. Enter a set of data in column C1.
2. Click on the main menu item of **Calc**.
3. Select **Standardize** from the subdirectory.
4. A dialog box should appear. Enter C1 for the input column and C2 for the column in which the results will be stored.
5. Click **OK**, and column C2 will magically appear with the *z* score equivalent of each of the original sample values. Because column C2 is now available as a set of data, you can explore it using Descriptive Statistics, Histogram, Dotplot, and so on.

Enter the cotinine levels of smokers from Table 2–1 into Minitab column C1 and follow the above procedure to get these *z* scores listed in column C2: −1.43496, −1.44333, . . . , 0.00439.

2-10 Boxplots

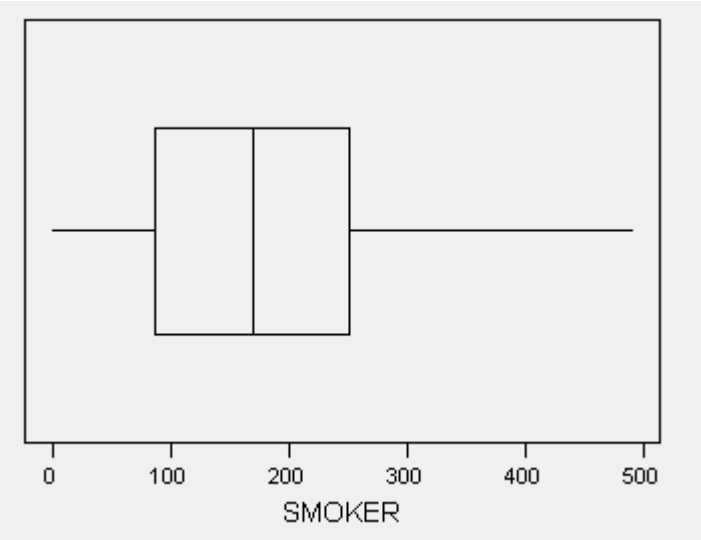
Section 2-6 of the textbook describes the construction of boxplots based on the minimum value, maximum value, median, and quartiles Q_1 and Q_3. Here is the Minitab procedure for generating a boxplot.

1. Enter or retrieve the data into column C1.
2. Click on the main menu item of **Graph**.
3. Select the subdirectory item of **Boxplot**.
4. In the dialog box, click on the column containing the data for which the boxplot will be produced.
5. Click **OK**.

Important note: The textbook and Minitab will construct boxplots using the same values for the minimum, median, and maximum, but the values of the quartiles Q_1 and Q_3 may differ, so there may be some discrepancies between Minitab boxplots and boxplots in the textbook. Such

discrepancies will usually be relatively small.

Using the 40 cotinine levels of smokers listed in Table 2-1, Minitab will display the boxplot shown below. Some other Minitab boxplots may include asterisks to identify points that appear to be outliers.



2-11 Sorting Data

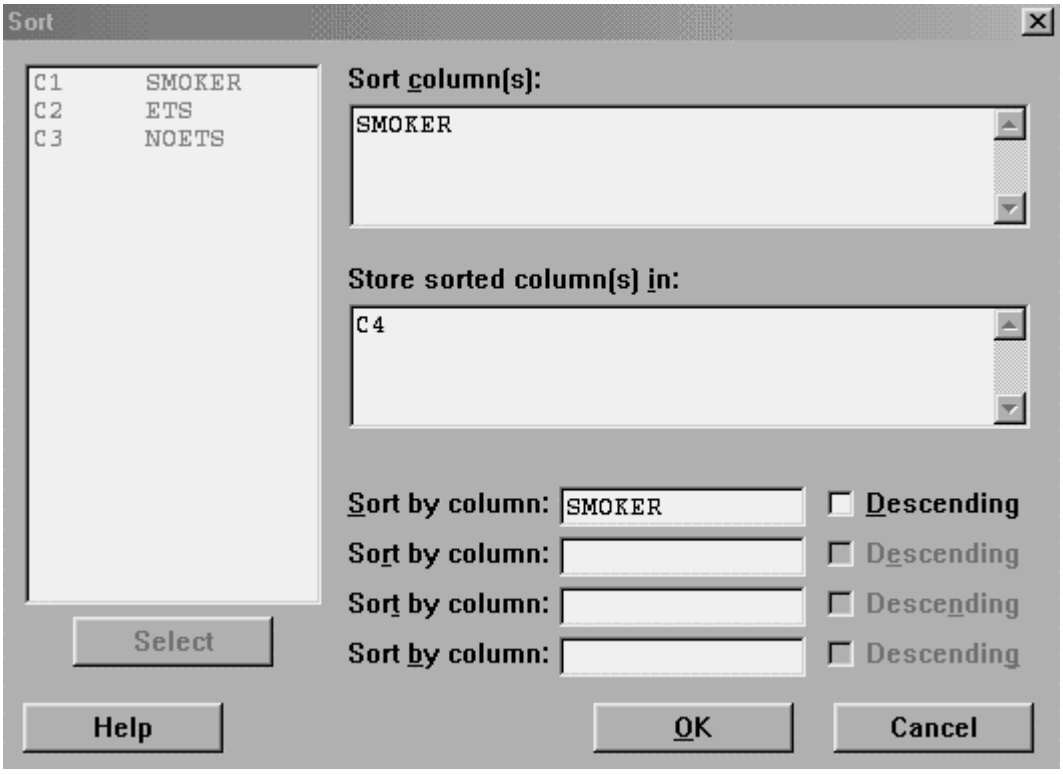
There is often a need to *sort* data. Data are sorted when they are arranged in order from low to high (or from high to low). Identification of outliers becomes easier with a sorted list of values. Here is how we can use Minitab to sort data.

1. Enter or retrieve a data set into column C1.

2. Click on the main menu item of **Manip** (for "manipulate").

3. Click on the subdirectory item of **Sort**.

4. You will now see a dialog box like the one shown below. In the "Sort column(s) box, enter the column that you want to sort. In the "Store sorted column(s) in" box, enter the column where you want the sorted data to be placed. (You can place the sorted data in the same column as the original data.) In the "Sort by column" box, enter the column to be used as the basis for sorting (usually the column containing the original unsorted data). The entries shown below tell Minitab to sort the SMOKER data (in column C1) and store the sorted list in column C4. Click **OK** after completing the dialog box.



Summary: We noted at the beginning of this chapter that the following are extremely important characteristics of data: center, variation, distribution, outliers, and pattern over time. These characteristics can be investigated using the Minitab tools described in this chapter. Here is a summary of the tools that are usually most relevant for the different characteristics:

1. **Center:** Use **Stat/Basic Statistics/Display Descriptive Statistics** to find the mean and median.

2. **Variation:** Use **Stat/Basic Statistics/Display Descriptive Statistics** to find the standard deviation. Also find the range by using the minimum and maximum values.

3. **Distribution:** Use the **Graph** menu to find a histogram, boxplot, dotplot, and stem-and-leaf plot.

4. **Outliers:** Use **Manip/Sort** to sort the data in ascending order, then examine the sample values to identify any that are very far away from almost all others.

5. **Time:** Use **Graph/Time Series Plot**. Also see Chapter 13.

The construction of *frequency tables* is the only major topic included in Chapter 2 of the textbook that is not included in this chapter of the manual/workbook. Although frequency tables are sometimes valuable in their own rite, they often play a role of being a means to achieving some greater goal, such as the construction of histograms or other important graphs.

CHAPTER 2 EXPERIMENTS: Describing, Exploring, and Comparing Data

2–1. **Comparing Heights of Men and Women** In this experiment we use two small data sets as a quick introduction to using some of the basic Minitab features. (When beginning work with new software, it is wise to first work with small data sets so that they can be entered quickly if they are lost or damaged.) The data listed below are measured heights (in inches) of random samples of men and women (taken from Data Set 1 in Appendix B of the textbook).

Men	70.8	66.2	71.7	68.7	67.6	69.2
Women	64.3	66.4	62.3	62.3	59.6	63.6

a. Find the indicated characteristics of the heights of *men* and enter the results below.

Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

b. Find the characteristics of the heights of *women* and enter the results below.

Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

c. Compare the results from parts a and b.

_____

_____

_____

2–2. **Working with Larger Data Sets** Repeat Experiment 2–1, but use the sample data for all 40 males and 40 females included in Data Set 1 that is found in Appendix B of the textbook. Instead of manually entering the 80 individual heights (which would be no fun at all), open the worksheets MHEALTH.MTW and FHEALTH.MTW that are found among the Minitab data sets on the CD–ROM that is included with the textbook.

a. Find the indicated characteristics of the heights of *men* and enter the results below.

Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

b. Find the characteristics of the heights of *women* and enter the results below.

Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

c. Compare the results from parts a and b.

_____

_____

_____

d. Are there are notable differences observed from the complete sets of sample data that could not be seen with the smaller samples listed in Experiment 2–1? If so, what are they?

_____

_____

_____

2–3. **Histogram** Use the same sets of data used in Experiment 2–2 and print histograms for the heights of the 40 men and the heights of the 40 women. Are there any notable differences in the two sets of sample data?

_____

2–4. **Boxplots** Use the same sets of data used in Experiment 2–2 and print boxplots for the heights of the 40 men and the heights of the 40 women. Do the boxplots suggest any notable differences in the two sets of sample data?

_____

How do you interpret the *asterisk* that appears in the Minitab display for the boxplot depicting the heights of the men?

_____

The two boxplots are much easier to compare if they are constructed together on the same scale. Use Minitab to generate the two boxplots together, then print the results. To accomplish this, retrieve the worksheets MHEALTH.MTW and FHEALTH.MTW, then use the Copy/Paste feature to copy the male heights in column C14 of the worksheet FHEALTH.MTW. Then select Graph/Boxplot and enter the names of the columns in the dialog box under the variable Y. Next, click on **Frame**, then **Multiple Graphs**, and click on the button for **Overlay graphs on the same page.**

2–5. **Scatterplots** Section 2–6 of this manual/workbook included an example of a scatterplot depicting the paired waist and weight measurements for a sample of 40 males. Use Minitab to print the scatterplot for the paired waist and weight measurements for the sample of 40 women. The sample values can be retrieved from the Minitab worksheet FHEALTH.MTW. How does the resulting scatterplot compare to the scatterplot shown for males? Is a similar scatterplot obtained? Does there appear to be a relationship between the waist sizes and weights of females?

_____

_____

2–6. **Scatterplots** Now it's time to be creative. When requesting a printout of a scatterplot, Experiment 2–5 specified the variables of waist size and weight for a sample of females. Using any data set from Appendix B in the textbook (other than Data Set 1), identify two paired variables that you suspect are related, then obtain a Minitab printout of the scatterplot. Does the graph support your belief that there would be a relationship? What feature of the graph suggests that there is or is not a relationship?

_____

_____

2–7. **Effect of Outlier** In this experiment we will study the effect of an *outlier*. Use the same heights of *men* used in Experiment 2-1, but change the first entry from 70.8 in. to 708 in. (This type of mistake often occurs when the key for the decimal point is not pressed with enough force.) The outlier of 708 in. is clearly a mistake, because a male with of height of 708 in. would be 59 feet tall, or about six stories tall. Although this outlier is a mistake, outliers are sometimes correct values that differ substantially from the other sample values.

Men: 708 66.2 71.7 68.7 67.6 69.2

Using this modified data set with the height of 70.8 in. changed to be the outlier of 708 in., find the following.

Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

Based on a comparison of these results to those found in Experiment 2–1, how is the mean affected by the presence of an outlier?

_____

How is the median affected by the presence of an outlier?

_____

How is the standard deviation affected by the presence of an outlier?

_____

2–8. **Effect of Outlier** In Experiment 2–3 we obtained a printout of a histogram for the heights of 40 males. Change the first height from 70.8 in. to the outlier of 708 in., then obtain the histogram. How is the histogram affected by the presence of the outlier?

_____

_____

2–9. **Time–Series Graphs** Retrieve the Minitab worksheet MISC.MTW. Print three different Minitab time–series graphs for U.S. car sales, U.S. murders and non–negligent homicides, and sunspot numbers for the years from 1980 to 2000. In each case, identify any notable patterns. Also, in each case determine whether the data values appear to come from a population that is changing over time.

_____

_____

2–10. **Sorting Data** Retrieve the Minitab worksheet SOLITAIRE.MTW. The sample data represent losing amounts (represented by negative values) and winning amounts (represented by positive values) when solitaire is played with Las Vegas rules. *Sort* the data by arranging them in order from lowest to highest.

How many of the values are negative? _____

How many of the sample values are positive? _____

What is the largest amount that was lost? _____

What is the largest amount that was won? _____

What is the mean amount won or lost per game? _____

What do the results suggest about this game of solitaire?

_____

2–11. ***z* Scores** Retrieve the Minitab worksheet CANS.MTW, then use Minitab to find the z score corresponding to each of the 175 values for the cans that are 0.0109 in. thick. [Each value is an axial load, which is the weight (in pounds) that the can supports before being crushed.] Store the 175 z scores in column C3, then find the following.

Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

What is notable about the above results? Specifically, what is notable about the value of the mean and standard deviation?

_____

_____

Will the same mean and standard deviation be obtained for *any* set of sample data? Explain how you arrived at your answer.

_____

_____

2–12. **Stem–and–Leaf Plots** Exercise 34 in Section 2–3 of the textbook lists the ages of actors and actresses when they won Oscars. Those ages are reproduced below.

Actors

32	37	36	32	51	53	33	61	35	45
55	39	76	37	42	40	32	60	38	56
48	48	40	43	62	43	42	44	41	56
39	46	31	47	45	60	46	40	36	

Actresses

50	44	35	80	26	28	41	21	61	38
49	33	74	30	33	41	31	35	41	42
37	26	34	34	35	26	61	60	34	24
30	37	31	27	39	34	26	25	33	

Use Minitab to obtain and print stem–and–leaf plots for actors and actresses, then compare them. What major difference becomes apparent?

_____

_____

Refer to the Minitab stem–and–leaf plot for the actors. Interpret the value of 14 in the column at the left.

_____

_____

Refer to the Minitab stem–and–leaf plot for the actors. Interpret the entry of (9) in the column at the left.

_____

_____

Refer to the Minitab stem–and–leaf plot for the actors. Interpret the entry of 10 in the column at the left.

_____

_____

2–13. **Pareto Chart** (This is from Exercise 21 in Section 2–3 of the textbook.)
A study was conducted to determine how people get jobs. The table lists data from 400 randomly selected subjects. The data are based on results from the National Center for Career Strategies. Use Minitab to construct a Pareto chart that corresponds to the given data. If someone would like to get a job, what seems to be the most effective approach?

Job Sources of Survey Respondents	Frequency
Help-wanted ads	56
Executive search firms	44
Networking	280
Mass mailing	20

2–14. **Pie Chart** (This is from Exercise 22 in Section 2–3 of the textbook.)
Refer to the data given in Experiment 2–13 and use Minitab to construct a pie chart. Compare the pie chart to the Pareto chart. Can you determine which graph is more effective in showing the relative importance of job sources?

_____

_____

2-15. **Combining Data** Open the Minitab worksheet M&M.MTW, which consists of six different columns of data. Use **Manip/Stack** to make a copy of the data all stacked together in column C7 of the current worksheet.

a. Find the following results for the combined data in column C7.

Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

b. Print a histogram of the data set.

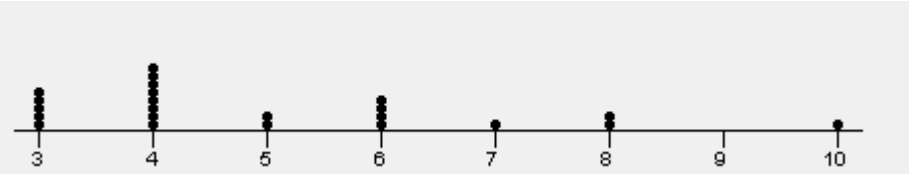
c. Print a boxplot of the data set.

d. Describe the important characteristics of the data set. Be sure to address the nature of the distribution, measures of center, measures of variation, and any other important and notable features.

_____

_____

2-16. **Interpreting Dotplot** Shown below is a Minitab dotplot. Identify the values represented in this graph, enter them in column C1, then find the indicated results that follow the dotplot.



Center: Mean: _____ Median: _____

Variation: St. Dev.:_____ Range: _____

5-Number Summary: Min.:_____ Q_1:_____ Q_2:_____ Q_3:_____ Max.:_____

Outliers: _____

2–17. **Comparing Data** Readability data were compiled from randomly selected pages of the following books:

- Tom Clancy's *The Bear and the Dragon*
- J. K. Rowling's *Harry Potter and the Sorcerer's Stone*
- Leo Tolstoy's *War and Peace*

The data are listed in Data Set 14 of Appendix B in the textbook, and they are also available as the Minitab worksheets CLANCY.MTW, ROWLING.MTW, and TOLSTOY.MTW. Use Minitab to compare the readability of the three books. Provide relevant printed displays, and write a brief report stating your conclusions. (*Hint*: It might be helpful to use Copy/Paste to combine all of the sample data in one worksheet.)

2–18. **Working with Your Own Data** Through observation or experimentation, collect your own set of sample values. Obtain at least 40 values and try to select data from an interesting population. Use Minitab to explore the data. Describe the nature of the data. That is, what do the values represent? Describe important characteristics of the data set, and include Minitab printouts to support your observations.