Chapter 2
Methods for Describing Sets of Data

2.1 Introduction

Chapter 1 served to introduce many of the basic statistical concepts employed in all types of data analysis problems. Two main areas of statistics emerge from Chapter 1 - descriptive and inferential statistics. Chapter 2 focuses on the descriptive area and looks at both graphical and numerical techniques that allow statisticians to summarize data that has been collected. Many of the techniques used to summarize data discussed in *Statistics* can easily be performed with **XLSTAT** and **Excel**. Our purpose is to explain these techniques and to illustrate them using the examples presented in the text as well as additional examples provided here. Listed below are the various techniques that **XLSTAT** and **Excel** offer that can be used to generate the graphical and numerical topics presented in Chapter 2.

XLSTAT and **Excel** offer a wide array of graphing options to the statistician. When working with qualitative data, **XLSTAT** allows the statistician to create customized **pie charts** and **bar graphs**. For quantitative data, **histograms, dot plots**, **box plots**, **stem-and-leaf displays**, and **scatterplots** are easy to create. As with most database and statistical software programs, **XLSTAT** and **Excel** provide a wide array of numerical description of data. The three measures of central tendency (**mean**, **median**, and **mode**) and the three measures of variability (**range**, **variance**, and **standard deviation**) are all available in the descriptive statistics menu of both **XLSTAT** and **Excel**. Measures of relative standing (**percentiles**, **quartiles**, and **z-scores**) are available as well, but not as easy to access as the measures of central tendency and measures of spread.

When both **XLSTAT** and **Excel** provide methods of generating the same output, we will show the **XLSTAT** technique, as we believe **XLSTAT** provides the easiest path for the student to generate the desired output. The following examples from *Statistics* are solved with **Excel** and **XLSTAT** in this chapter:

Excel Companion Exercise	Page	Statistics Text	Data File Name
2.1	32	Table 2.1	APHASIA
2.2	35	Example 2.2	WLTASK
2.3	38	Example 2.2	WLTASK
2.4	42	Example 2.19	MEDFACTORS
2.5	44	Table 2.2	EPAGAS
2.6	46	Table 2.2	EPAGAS
2.7	48	Example 2.14	WLTASK
2.8	50	Example 2.15	

2.2 Graphical Techniques in Excel

2.2.1 Graphical Techniques for Qualitative Data

Bar graphs, pie charts, and frequency displays are all useful to create when working with qualitative data. The **XLSTAT Add-In** offers an easy method of creating these displays within the **Visualizing data** menu. We illustrate with the following example.

Exercise 2.1: As an example we turn to Table 2.1 from the *Statistics* text:

Consider a study on aphasia published in the Journal of Communication Disorders (Mar. 1995). Aphasia is the "impairment or loss of the faculty of using or understanding spoken or written language." Three types of aphasia have been identified by researchers: Broca's, conduction, and anomic. The researchers wanted to determine whether one type of aphasia occurs more often than any other and, if so, how often. Consequently, they measured the type of aphasia for a sample of 22 adult aphasics. Table 2.1 (from the data file **APHASIA**) gives the type of aphasia diagnosed for each aphasic in the sample.

Table 2.1

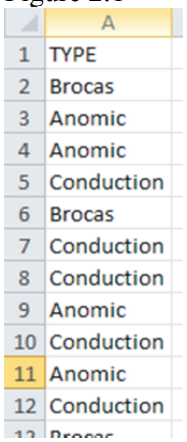
Brocas	Brocas	Conduction	Anomic	Anomic	Conduction
Anomic	Conduction	Anomic	Brocas	Conduction	
Anomic	Conduction	Conduction	Anomic	Brocas	
Conduction	Anomic	Brocas	Anomic	Anomic	

a. Use a statistical software package to create a frequency display for these data.
b. Use a statistical software package to create a frequency histogram for these data.
c. Use a statistical software package to create a frequency pie chart for these data.

Solution:

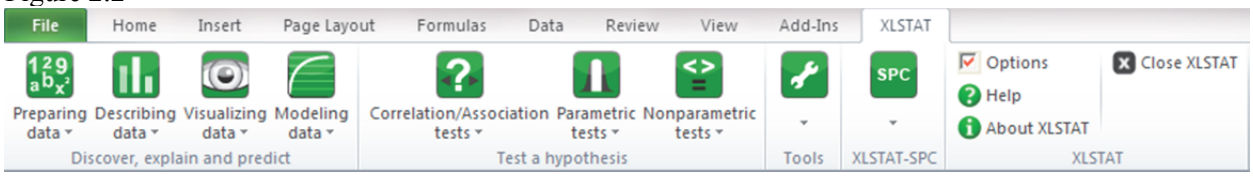
We solve this problem by using the **Visualizing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the Data File **APHASIA** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 2.1. Use the mouse to select the data shown in the workbook.

Figure 2.1



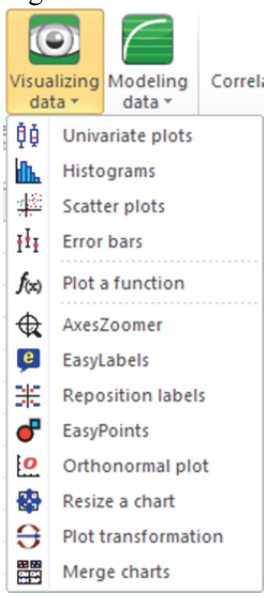
To create the desired displays, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 2.2.

Figure 2.2



To generate the desired plots, we click on the **Visualizing data** menu and select the **Univariate plots** option shown in Figure 2.3.

Figure 2.3



This opens the **Univariate plots** menu shown in Figures 2.4 - 2.6. In order to work with qualitative data, we need to make sure that the **Qualitative data** box is checked in the **General** tab on the **Univariate plots** menu. In addition, we need to specify the location of the data that is to be analyzed. In our data set, the data is located in Column A, rows 2 – 23, with row 1 being the variable label. We specify the location in the **Qualitative data** box and check the **Sample labels** box to indicate the first row of data represents the variable name, **TYPE**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described.

Click on the **Outputs** tab (shown in Figure 2.5) to indicate the type of information you would like displayed in the frequency display. In our example, we have checked the **All** box to display all information. Note here that the only side of this tab that is active is the **Qualitative data** side. We only have indicated that we are working with qualitative data in this example, so the **Qualitative data** is the only active **Output** allowed.

Figure 2.4

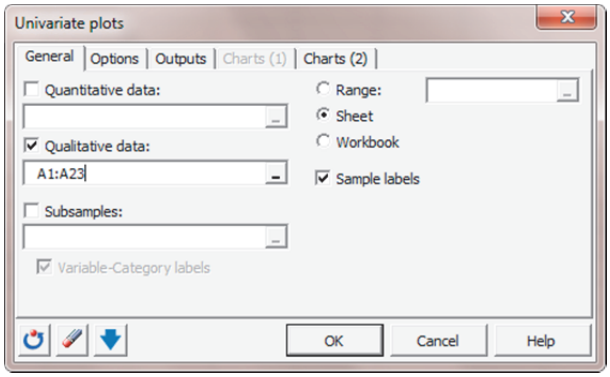
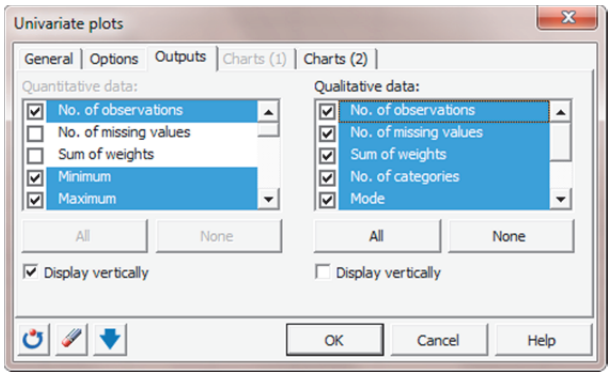
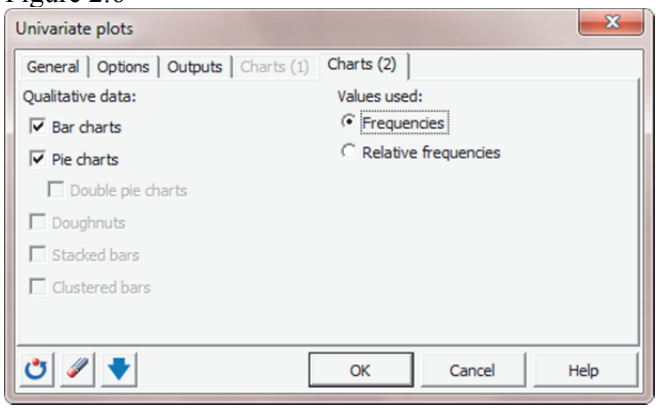


Figure 2.5



We click on the **Charts (2)** tab shown in Figure 2.6 to select the type of charts to create. We have selected both the **Bar charts** and **Pie charts** options by checking the appropriate boxes. In addition, we must specify the type of charts to create by selecting either the **Frequencies** or **Relative frequencies** option. We click **OK** to generate the desired output.

Figure 2.6



The Frequency and Relative Frequency output for the data set shown in Table 2.2. Please note that the output created by **XLSTAT** contains additional information that has been deleted to produce this table.

Table 2.2

Category	Frequency per category	Rel. frequency per category (%)
Anomic	10.0000	45.4545
Brocas	5.0000	22.7273
Conduction	7.0000	31.8182

In addition, the bar graph and pie chart for the data set are shown in Figures 2.7 and 2.8.

Figure 2.7

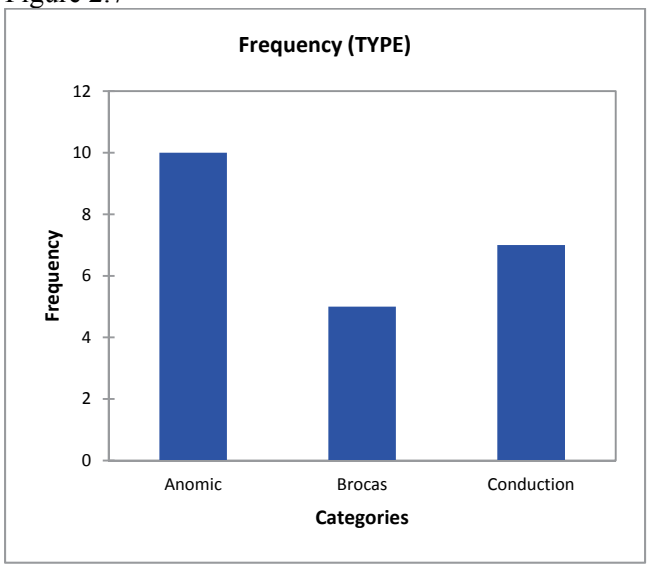
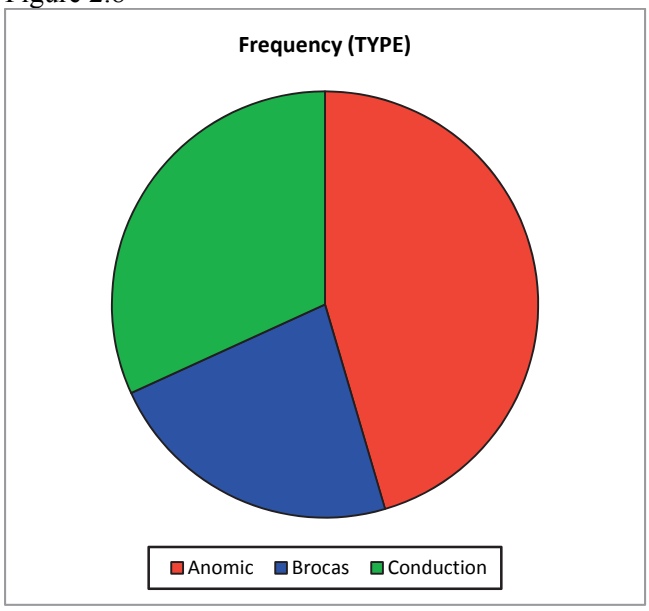


Figure 2.8



Compare these output to the ones created in the *Statistics* text. We see that they look very similar to the output created by other software illustrated in the text.

2.2.2 Graphical Techniques for Quantitative Data

Histograms, dot plots, box plots, stem-and-leaf displays, and scatterplots are all useful to create when working with quantitative data. The **XLSTAT Add-In** offers an easy method of creating these displays within the **Visualizing data** menu. We begin with illustrating how to create a histogram using the following example.

Exercise 2.2: As an example we turn to Example 2.2 from the *Statistics* text:

Over 60 years ago, famous child psychologist Jean Piaget devised a test of basic perceptual and conceptual skills dubbed the "water-level task". Subjects were shown a drawing of a glass being held at a 45° angle and asked to draw a line representing the true surface of the water. Today, research psychologists continue to use the task to test the perception of both adults and children. In one study, the water-level task was given to several groups which included 20 male bartenders and 20 female waitresses (Psychological Science, March 1995). For each participant, the researchers measured the deviation (in angle degrees) of the judged line from the true line. These deviations (simulated on the basis of summary results presented in the journal article are shown in Table 2.3 (from the data file **WLTASK**). [Note: Deviations can be negative if the judged angle is smaller than the angle of the true line.

Table 2.3

Bartender	-9	6	10	6	10	-3	7	8	6	14
	7	8	-5	2	-1	0	2	3	0	2
Waitress	7	10	25	8	10	8	12	9	35	10
	12	11	7	10	21	-1	4	0	16	-1

Use a statistical software package to create a frequency histogram for these data.

Solution:

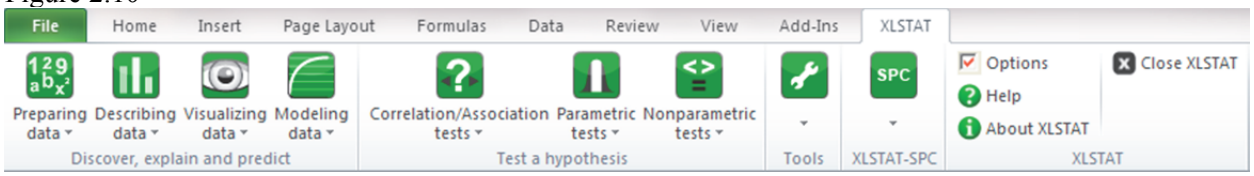
We solve this problem by using the **Visualizing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the Data File **WLTASK** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 2.9. Use the mouse to select the data shown in the workbook.

Figure 2.9

	A	B	C	D
1	Gender	Group	Deviation	Judge
2	Female	Waitres	7	More5Above
3	Female	Waitres	10	More5Above
4	Female	Waitres	25	More5Above
5	Female	Waitres	8	More5Above
6	Female	Waitres	10	More5Above
7	Female	Waitres	8	More5Above
8	Female	Waitres	12	More5Above
9	Female	Waitres	9	More5Above
10	Female	Waitres	35	More5Above
11	Female	Waitres	10	More5Above
12	Female	Waitres	12	More5Above
13	Female	Waitres	11	More5Above
14	Female	Waitres	7	More5Above

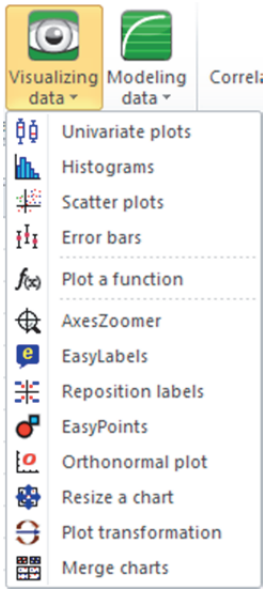
To create the histogram, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 2.10.

Figure 2.10



To generate the desired plots, we click on **Visualizing data** menu and select the **Histograms** option shown in Figure 2.11.

Figure 2.11



This opens the **Histogram** menu shown in Figures 2.12 and 2.13. The **Histograms** menu gives a wide variety of options to the user to customize the type of histogram desired. We will focus on the simplest path to create a histogram. First, the user must specify the data that is to be analyzed in the **General** tab of the **Histograms** menu (Figure 2.12). In our data set, the data is located in Column C, rows 1 - 41, with row 1 being the variable label. We specify the location in the **Data** box and check the **Sample labels** box to indicate the first row of data represents the variable name, **Deviation**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described.

Click on the **Charts** tab (shown in Figure 2.13) to indicate the type of histogram you would like displayed. In our example, we have indicated a frequency bar graph by clicking the **Histograms** box, clicking on the **Bars** button, and selecting a **Frequency** bar graph from the pull-down menu. We click **OK** to create the histogram shown in Figure 2.14.

Figure 2.12

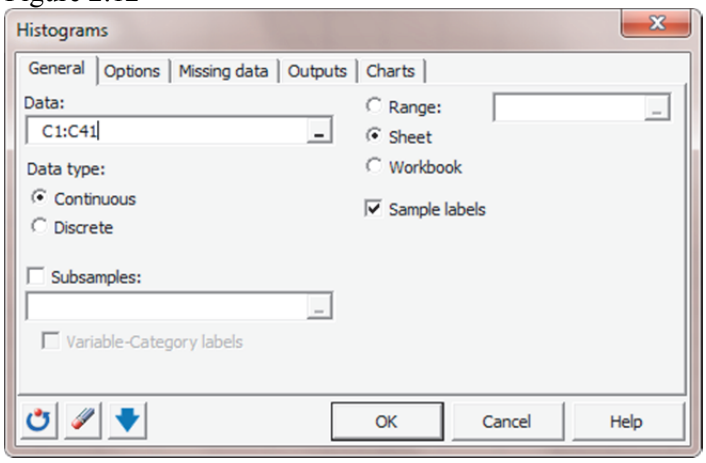


Figure 2.13

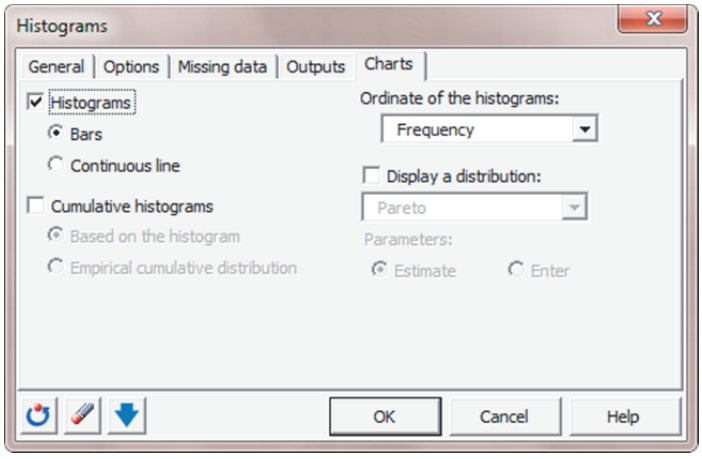
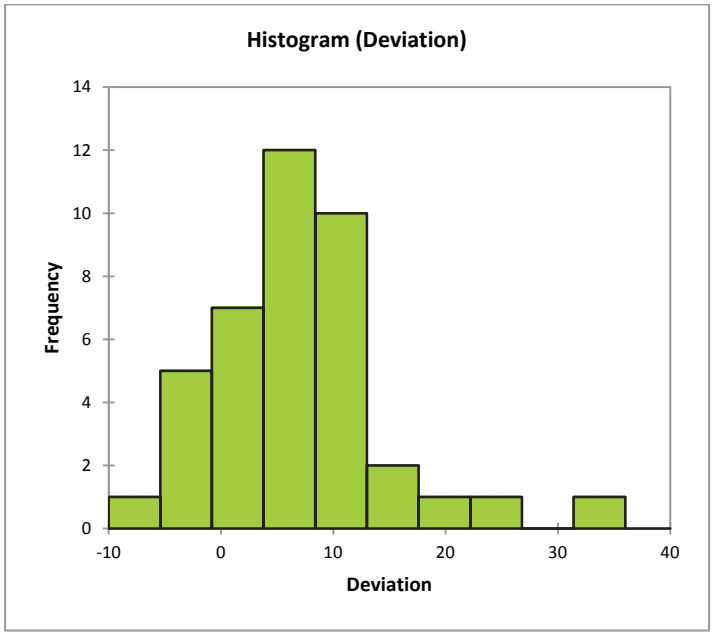


Figure 2.14



Compare this histogram to the one created in the Statistics text. Please note that by making different selections in the various options within the Histograms menu, the look and style of the histogram you create can change dramatically. **XLSTAT** allows the user to customize the histogram to provide the desired information.

We now illustrate how to create dot plots, stem-and-leaf displays and box plots with **XLSTAT**. The **XLSTAT Add-In** offers an easy method of creating these displays within the **Visualizing data** menu. We illustrate with the following example.

Exercise 2.3: As an example we turn to Example 2.2 from the *Statistics* text:

Over 60 years ago, famous child psychologist Jean Piaget devised a test of basic perceptual and conceptual skills dubbed the "water-level task". Subjects were shown a drawing of a glass being held at a 45° angle and asked to draw a line representing the true surface of the water. Today, research psychologists continue to use the task to test the perception of both adults and children. In one study, the water-level task was given to several groups which included 20 male bartenders and 20 female waitresses (Psychological Science, March 1995). For each participant, the researchers measured the deviation (in angle degrees) of the judged line from the true line. These deviations (simulated on the basis of summary results presented in the journal article are shown in Table 2.4 (from the data file **WLTASK**). [Note: Deviations can be negative if the judged angle is smaller than the angle of the true line.

Table 2.4

Bartender	-9	6	10	6	10	-3	7	8	6	14
	7	8	-5	2	-1	0	2	3	0	2
Waitress	7	10	25	8	10	8	12	9	35	10
	12	11	7	10	21	-1	4	0	16	-1

Use a statistical software package to create a dot plot, stem-and-leaf display, and a box plot for these data.

Solution:

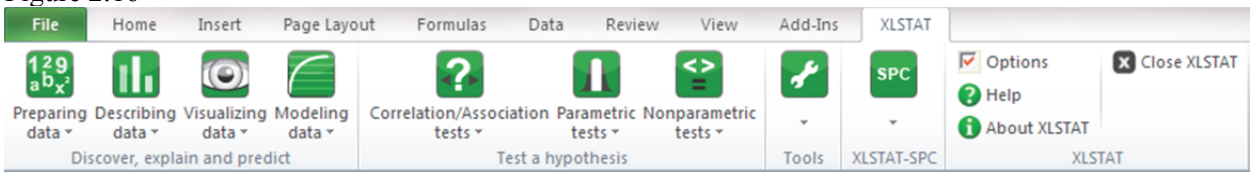
We solve this problem by using the **Visualizing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the Data File **WLTASK** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 2.15. Use the mouse to select the data shown in the workbook.

To create the desired displays, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 2.16.

Figure 2.15

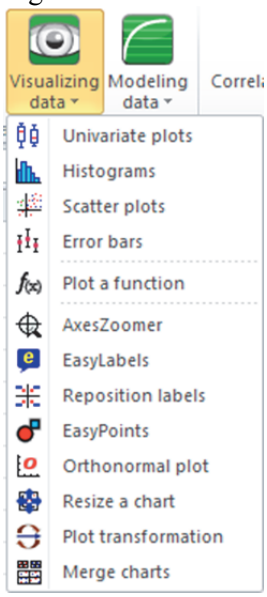
	A	B	C	D
1	Gender	Group	Deviation	Judge
2	Female	Waitres	7	More5Above
3	Female	Waitres	10	More5Above
4	Female	Waitres	25	More5Above
5	Female	Waitres	8	More5Above
6	Female	Waitres	10	More5Above
7	Female	Waitres	8	More5Above
8	Female	Waitres	12	More5Above
9	Female	Waitres	9	More5Above
10	Female	Waitres	35	More5Above
11	Female	Waitres	10	More5Above
12	Female	Waitres	12	More5Above
13	Female	Waitres	11	More5Above
14	Female	Waitres	7	More5Above

Figure 2.16



To generate the desired plots, we click on the **Visualizing data** menu and select the **Univariate plots** option shown in Figure 2.17.

Figure 2.17



This opens the **Univariate plots** menu shown in Figures 2.18 and 2.19. In order to work with quantitative data, we need to make sure that the **Quantitative data** box is checked in the **General** tab on the **Univariate plots** menu. In addition, we need to specify the location of the data that is to be analyzed. In our data set, the data is located in Column A, rows 2 – 23, with row 1 being the variable label. We specify the location in the **Quantitative data** box and check the **Sample labels** box to indicate the first row of data represents the variable name, **TYPE**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described.

We click on the **Charts (1)** tab shown in Figure 2.19 to select the type of charts to create. We have selected the **Box plots, Scattergrams** (to create the dot plot) and **Stem-and-leaf plots** options by checking the appropriate boxes. We check the Horizontal button to create horizontal dot plots and box plots. We click **OK** to generate the desired output shown in Figures 2.20 – 2.22.

Figure 2.18

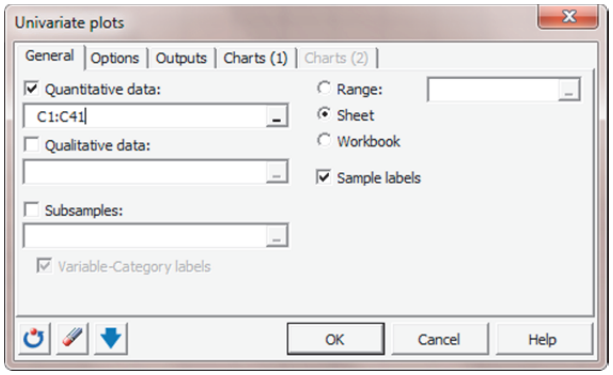


Figure 2.19

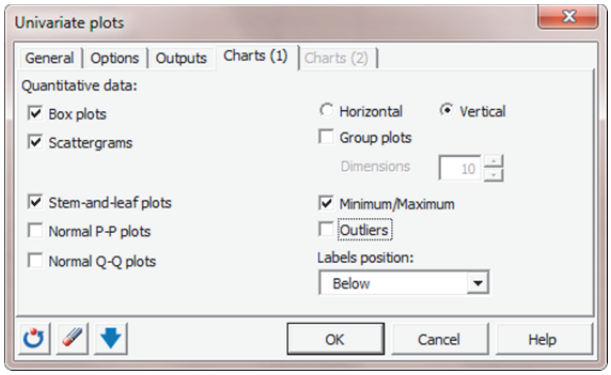


Figure 2.20

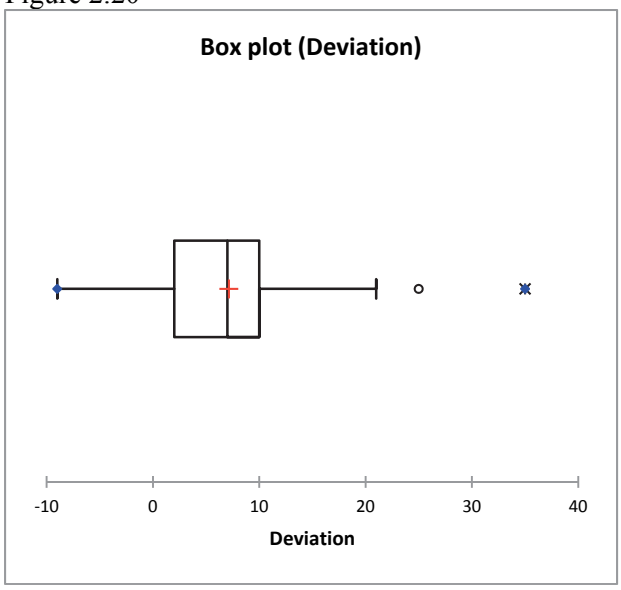


Figure 2.21

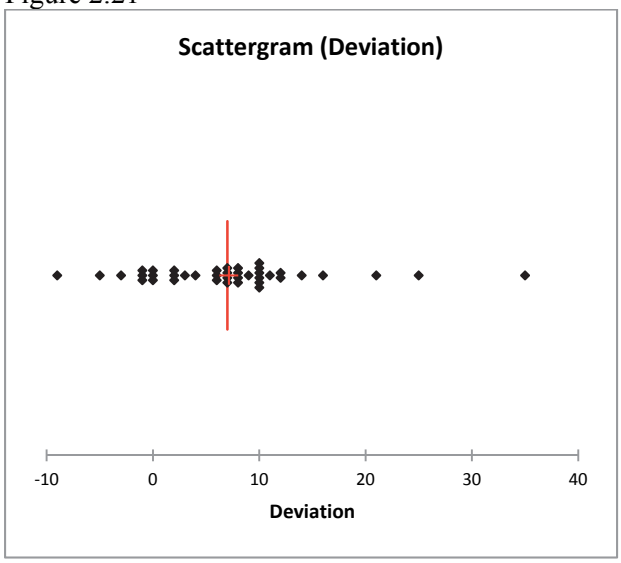


Figure 2.22

Stem-and-leaf plot (Deviation):

Unit:	1
-9	X
-8	
-7	
-6	
-5	X
-4	
-3	X
-2	
-1	X X X
0	X X X
1	
2	X X X
3	X
4	X
5	
6	X X X
7	X X X X
8	X X X X
9	X
10	X X X X X X
11	X
12	X X
13	
14	X
15	
16	X
17	
18	
19	
20	
21	X
22	
23	
24	
25	X
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	X

Compare these plots to the examples of similar plots shown in the *Statistics* text. We find that the box plots and dot plots generated by **XLSTAT** compare favorably with the ones shown in the text. The stem-and-leaf display, however, looks different. This is because **XLSTAT** organizes its stems differently than the examples shown in the text. This results in a stem-and-leaf display that appears more spread out. Currently, there are no options in **XLSTAT** that allow the user to specify different stem organization.

The final plot that we can create within **XLSTAT** is the scatterplot. We illustrate with the following example.

Exercise 2.4: As an example we turn to Example 2.19 from the *Statistics* text:.

A medical item used to administer to a hospital patient is called a factor. For example, factors can be intravenous (IV) tubing, IV fluid, needles, shave kits, bedpans, diapers, dressings, medications, and even code carts. The coronary care unit at Bayonet Point Hospital (St. Petersburg, Florida) recently investigated the relationship between the number of factors administered per patient and the patient's length of stay (in days). Data on these two variables for a sample of 50 coronary care patients are given in Table 2.3. Use a scattergram to describe the relationship between the two variables of interest, number of factors, and length of stay.

Table 2.5

Number of Factors	Length of Stay (in Days)	Number of Factors	Length of Stay (in Days)	Number of Factors	Length of Stay (in Days)
231	9	233	8	115	4
323	7	260	4	202	6
113	8	224	7	206	5
208	5	472	12	360	6
162	4	220	8	84	3
117	4	383	6	331	9
159	6	301	9	302	7
169	9	262	7	60	2
55	6	354	11	110	2
77	3	142	7	131	5
103	4	286	9	364	4
147	6	341	10	180	7
230	6	201	5	134	6
78	3	158	11	401	15
525	9	243	6	155	4
121	7	156	6	338	8
248	5	184	7		

Solution:

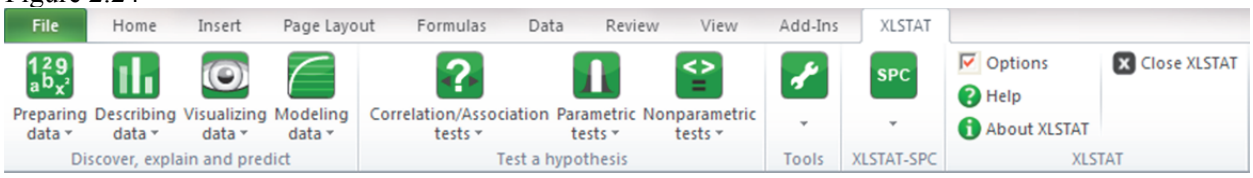
We solve this problem by using the **Visualizing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the Data File **MEDFACTORS** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 2.23. Use the mouse to select the data shown in the workbook.

To create the desired displays, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 2.24.

Figure 2.23

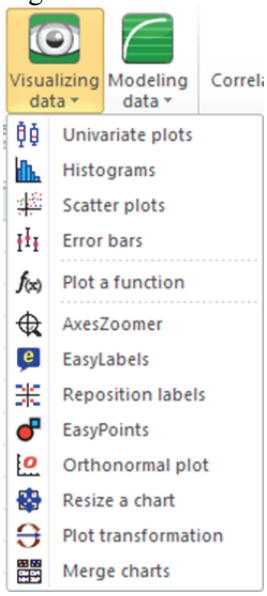
	A	B
1	LOS	FACTORS
2	9	231
3	7	323
4	8	113
5	5	208
6	4	162
7	4	117
8	6	159
9	9	169
10	6	55
11	3	77
12	4	103
13	6	147
14	6	230

Figure 2.24



To generate the desired plots, we click on the **Visualizing data** menu and select the **Scatter plots** option shown in Figure 2.25.

Figure 2.25



This opens the **Scatter plots** menu shown in Figures 2.26. We need to specify the location of the data that is to be analyzed. In our data set, the **LOS** data located in column A is to be plotted along the y-axis and the **FACTORS** data located in column B is to be plotted along the x-axis. We specify the location of these variables in the appropriate box in the **General** tab and check the **Variable labels** box to indicate the first row of data represents the variable names, **LOS** and **FACTORS**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data boxes as previously described. We click **OK** to generate the scatterplot shown in Figure 2.27.

Figure 2.26

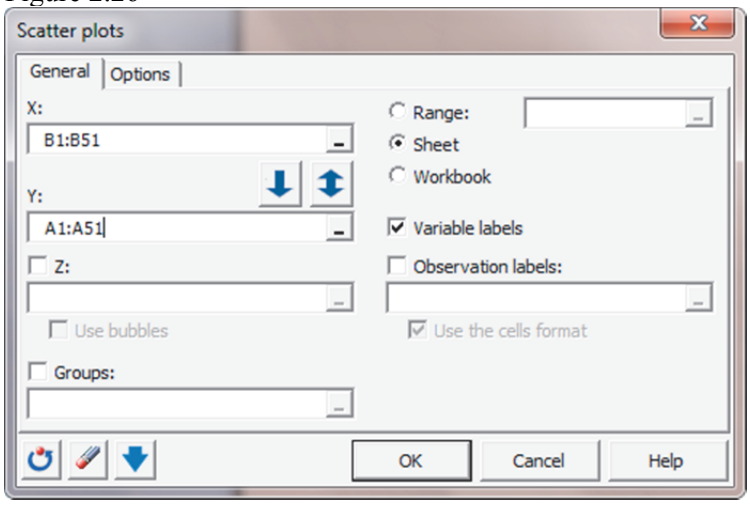
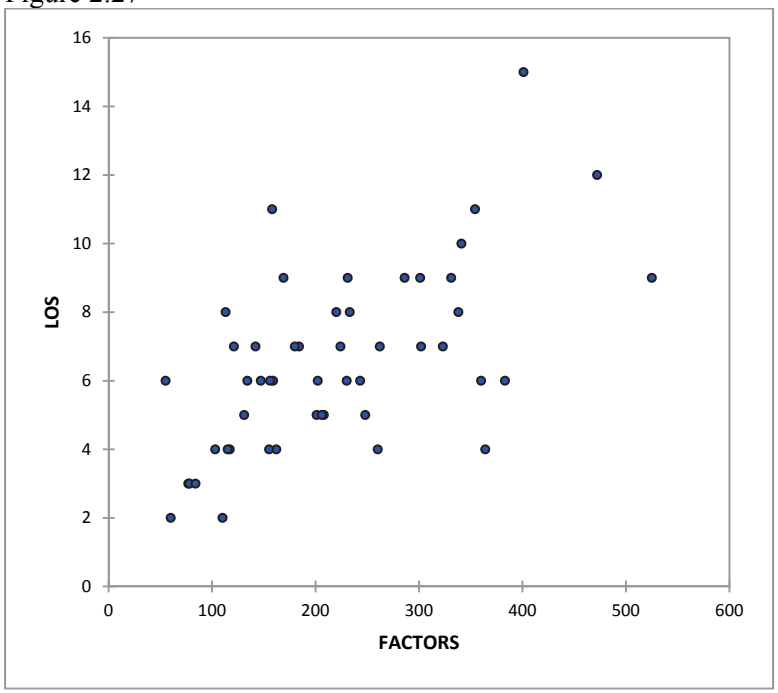


Figure 2.27



Compare this scatterplot to the one created in the Statistics text. You will find that the **XLSTAT** plot compares very favorably.

2.3 Numerical Techniques in Excel

2.3.1 Measures of Center

XLSTAT allows the user to create many descriptive measures of data through the use of the Descriptive Statistics data analysis. While **XLSTAT** doesn't distinguish between the different types of numerical measures, we choose to follow the *Statistics* text and look at the measures of center, spread, and relative standing one at a time. We begin with measures of center.

Exercise 2.5 We use *Statistics* Table 2.2 to illustrate the measures of center.

Calculate the measures of center for the 100 EPA mileages shown in Table 2.6 (from the data file **EPAGAS).**

Table 2.6

36.3	32.7	40.5	36.2	38.5	36.3	41.0	37.0	37.1	39.9
41.0	37.3	36.5	37.9	39.0	36.8	31.8	37.2	40.3	36.9
36.9	41.2	37.6	36.0	35.5	32.5	37.3	40.7	36.7	32.9
37.1	36.6	33.9	37.9	34.8	36.4	33.1	37.4	37.0	33.8
44.9	32.9	40.2	35.9	38.6	40.5	37.0	37.1	33.9	39.8
36.8	36.5	36.4	38.2	39.4	36.6	37.6	37.8	40.1	34.0
30.0	33.2	37.7	38.3	35.3	36.1	37.0	35.9	38.0	36.8
37.2	37.4	37.7	35.7	34.4	38.2	38.7	35.6	35.2	35.0
42.1	37.5	40.0	35.6	38.8	38.4	39.0	36.7	34.8	38.1
36.7	33.6	34.2	35.1	39.7	39.3	35.8	34.5	39.5	36.9

Solution:

We solve this problem by using the **Describing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the Data File **EPAGAS** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 2.27. Use the mouse to select the data shown in the workbook.

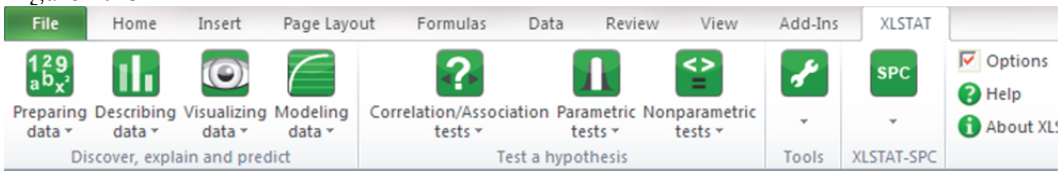
Figure 2.27

EPAGAS

	A
1	MPG
2	36.3
3	41
4	36.9
5	37.1
6	44.9
7	36.8
8	30
9	37.2
10	42.1
11	36.7
12	32.7
13	37.3
14	41.2

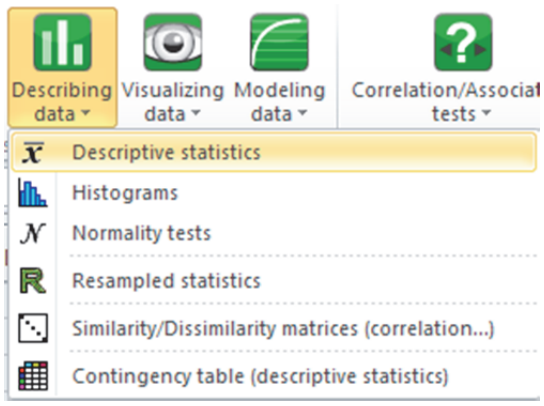
To create the desired displays, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 2.28.

Figure 2.28



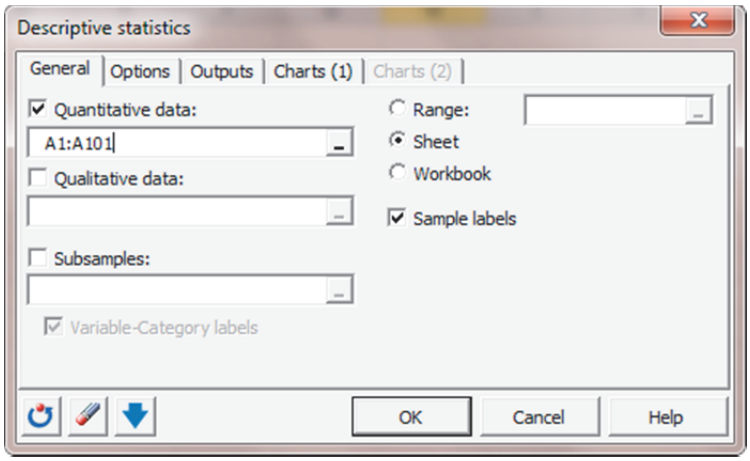
To generate the desired plots, we click on the **Describing data** menu and select the **Descriptive statistics** option shown in Figure 2.29.

Figure 2.29



This opens the **Descriptive statistics** menu shown in Figures 2.30. In order to work with quantitative data, we need to make sure that the **Quantitative data** box is checked in the **General** tab on the **Descriptive statistics** menu. In addition, we need to specify the location of the data that is to be analyzed. In our data set, the data is located in Column A, rows 1 - 101, with row 1 being the variable label. We specify the location in the **Quantitative data** box and check the **Sample labels** box to indicate the first row of data represents the variable name, **MPG**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described.

Figure 2.30



The **Descriptive statistics** menu offers many different options within the other tabs shown in the menu, but they are not needed to generate the basic descriptive measures that we require here. There are also some charting options that are available that we discussed earlier in this chapter. To create the descriptive statistics, we click **OK**.

Figure 2.31

Statistic	MPG
No. of observations	100
Minimum	30.0000
Maximum	44.9000
Range	14.9000
1st Quartile	35.6750
Median	37.0000
3rd Quartile	38.3250
Mean	36.9940
Variance (n-1)	5.8462
Standard deviation (n-1)	2.4179

We see the mean and median for the 100 MPG's shown in the table above. Compare these values to the calculations shown in the text.

2.3.2 Measures of Variation

We now illustrate how to generate measures of variation with **XLSTAT** with the following example.

Exercise 2.6 We use *Statistics* Table 2.2 to illustrate the measures of variation.

Calculate the measures of variation for the 100 EPA mileages shown in Table 2.7 (from the data file **EPAGAS).**

Table 2.7

36.3	32.7	40.5	36.2	38.5	36.3	41.0	37.0	37.1	39.9
41.0	37.3	36.5	37.9	39.0	36.8	31.8	37.2	40.3	36.9
36.9	41.2	37.6	36.0	35.5	32.5	37.3	40.7	36.7	32.9
37.1	36.6	33.9	37.9	34.8	36.4	33.1	37.4	37.0	33.8
44.9	32.9	40.2	35.9	38.6	40.5	37.0	37.1	33.9	39.8
36.8	36.5	36.4	38.2	39.4	36.6	37.6	37.8	40.1	34.0
30.0	33.2	37.7	38.3	35.3	36.1	37.0	35.9	38.0	36.8
37.2	37.4	37.7	35.7	34.4	38.2	38.7	35.6	35.2	35.0
42.1	37.5	40.0	35.6	38.8	38.4	39.0	36.7	34.8	38.1
36.7	33.6	34.2	35.1	39.7	39.3	35.8	34.5	39.5	36.9

Solution:

We solve this problem by using the **Describing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the Data File **EPAGAS** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 2.32. Use the mouse to select the data shown in the workbook.

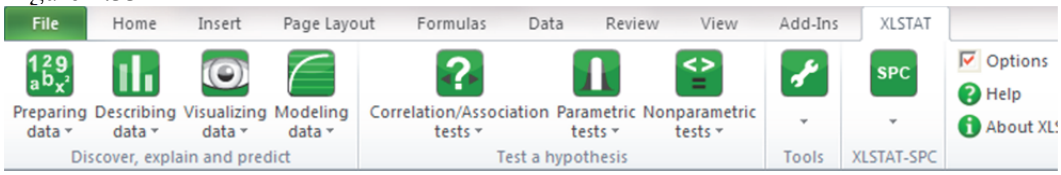
Figure 2.32

EPAGAS

	A
1	MPG
2	36.3
3	41
4	36.9
5	37.1
6	44.9
7	36.8
8	30
9	37.2
10	42.1
11	36.7
12	32.7
13	37.3
14	41.2

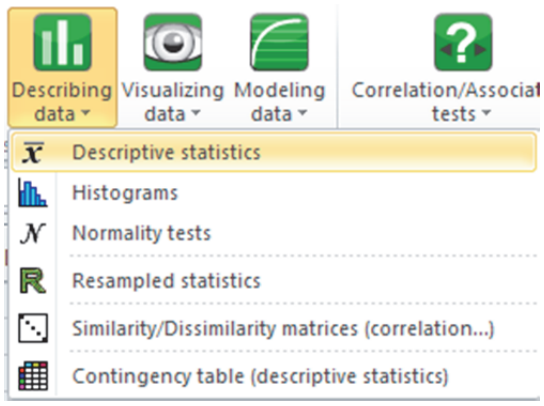
To create the desired displays, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 2.33.

Figure 2.33



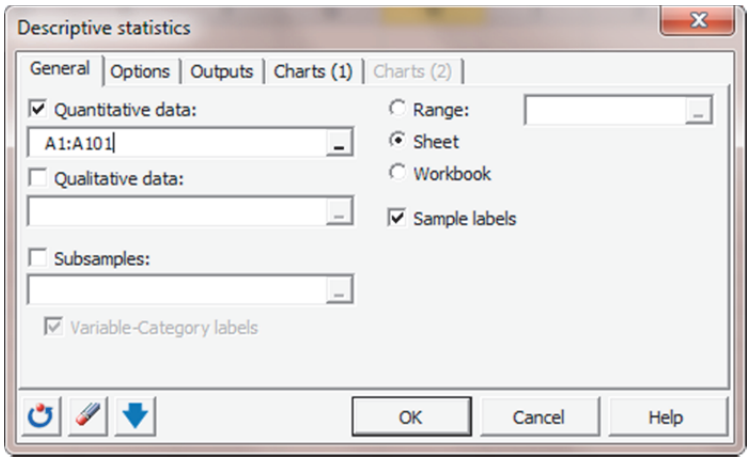
To generate the desired plots, we click on the **Describing data** menu and select the **Descriptive statistics** option shown in Figure 2.34.

Figure 2.34



This opens the **Descriptive statistics** menu shown in Figures 2.35. In order to work with quantitative data, we need to make sure that the **Quantitative data** box is checked in the **General** tab on the **Descriptive statistics** menu. In addition, we need to specify the location of the data that is to be analyzed. In our data set, the data is located in Column A, rows 1 - 101, with row 1 being the variable label. We specify the location in the **Quantitative data** box and check the **Sample labels** box to indicate the first row of data represents the variable name, **MPG**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described.

Figure 2.35



The **Descriptive statistics** menu offers many different options within the other tabs shown above in the menu, but they are not needed to generate the basic descriptive measures that we require here. There are also some charting options that are available that we discussed earlier in this chapter. To create the descriptive statistics, we click **OK**.

Figure 2.36

Statistic	MPG
No. of observations	100
Minimum	30.0000
Maximum	44.9000
Range	14.9000
1st Quartile	35.6750
Median	37.0000
3rd Quartile	38.3250
Mean	36.9940
Variance (n-1)	5.8462
Standard deviation (n-1)	2.4179

We see the range, standard deviation, and variance for the 100 MPG's shown in the table above. Compare these values to the calculations shown in the text.

2.3.3 Measure of Relative Standing

XLSTAT does not offer a way to easily calculate the two measures of relative standing, percentiles or z-scores, for an observation. Fortunately, **Excel** allows the user to calculate these measures utilizing two of its many functions. We first look at how **Excel** calculates percentiles.

Exercise 2.7: We use Example 2.14 from the *Statistics* text.

Refer to the water-level task deviation for the 40 subjects in Table 2.4. Locate the 25^{th} and 95^{th} percentile of the observations.

Solution:

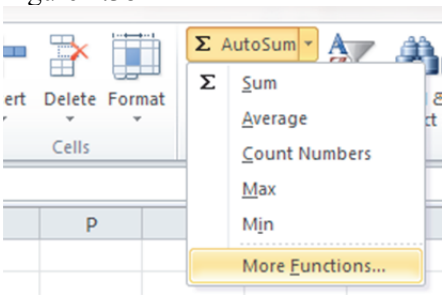
Before we begin, we must access the data set for this example. **Open** the Data File **WLTASK** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 2.37. Use the mouse to select the data shown in the workbook.

Figure 2.37

	A	B	C	D
1	Gender	Group	Deviation	Judge
2	Female	Waitres	7	More5Above
3	Female	Waitres	10	More5Above
4	Female	Waitres	25	More5Above
5	Female	Waitres	8	More5Above
6	Female	Waitres	10	More5Above
7	Female	Waitres	8	More5Above
8	Female	Waitres	12	More5Above
9	Female	Waitres	9	More5Above
10	Female	Waitres	35	More5Above
11	Female	Waitres	10	More5Above
12	Female	Waitres	12	More5Above
13	Female	Waitres	11	More5Above
14	Female	Waitres	7	More5Above

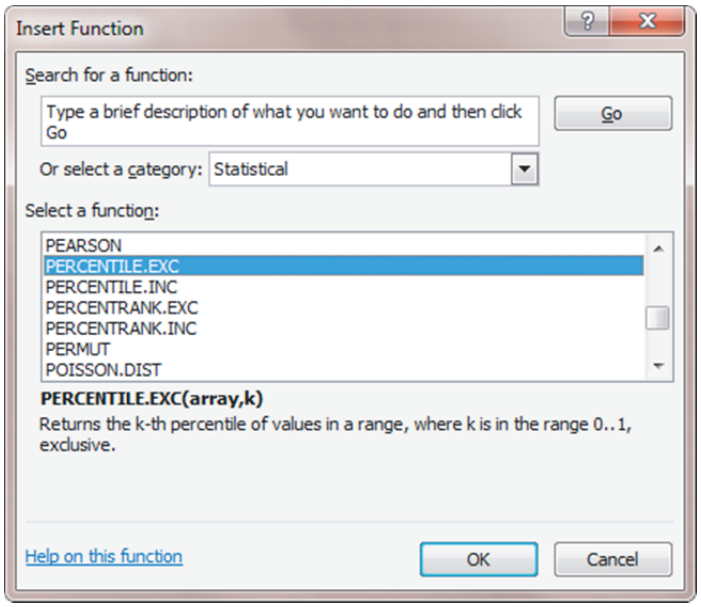
Once the data is available, click the Σ icon located in the **Editing** group of the **Home** tab.

Figure 2.38



By clicking on the down arrow, you can access the **More Functions** menu. Choose the **Statistical Function Category** and cursor down until you reach the function name **PERCENTILE.EXC.**

Figure 2.39



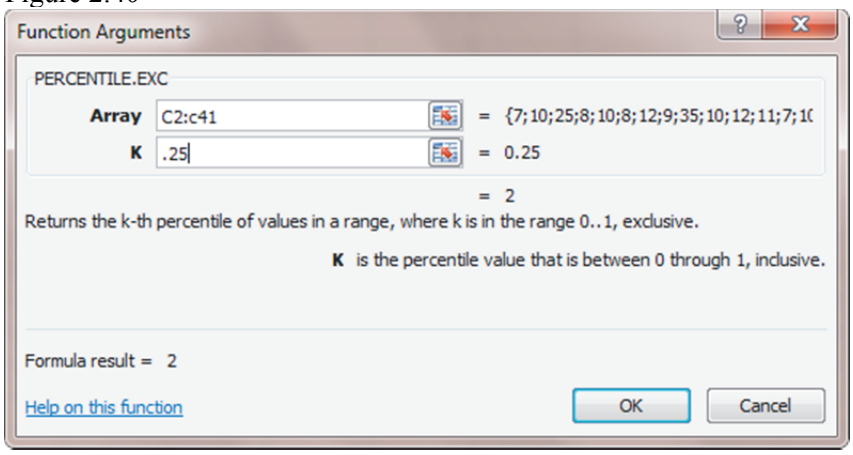
The **PERCENTILE.EXC** function has the form:

PERCENTILE.EXC(array,k)

where **array** represents the location of the data set that you want to find the percentile for, and **k** is a number between 0 and 1 that represents the percentile that is desired.

For this example, the 40 water-level task deviations are located in column C in rows 2 through 41. We assign the **Array** location to be **C2:C41** (see Figure 2.40). We also assign the value of **K** to be .25 representing the 25th percentile. Click **OK**.

Figure 2.40



Excel returns a value of 2. We interpret that a water-task deviation of 2 represents the 25th percentile of the 40 water-task level deviations in our data set. By changing the K to .95, **Excel** tells us that the 95th percentile is the value 24.8 (not shown). By changing the value of K, we can find any percentile we want.

The second measure of relative standing is the z-score. Again, we turn in **Excel** to a function that will allow the user to calculate values for a z-score. For purposes of illustration, we will again use the data from Example 2.10 to find a z-value.

Exercise 2.8: We use Example 2.15 from the *Statistics* text.

Suppose a sample of 2,000 high school seniors' verbal SAT scores is selected. The mean and standard deviation are

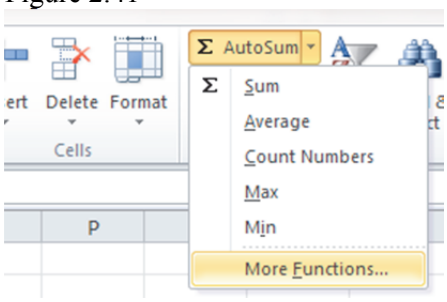
$$\bar{x} = 550 \quad s = 75$$

Suppose Joe Smith's score is 475. What is his sample z-score?

Solution:

Before we begin, we must access the Z-score function within **Excel** that will allow us to compute the z-score. We click the Σ icon located in the **Editing** group of the **Home** tab.

Figure 2.41

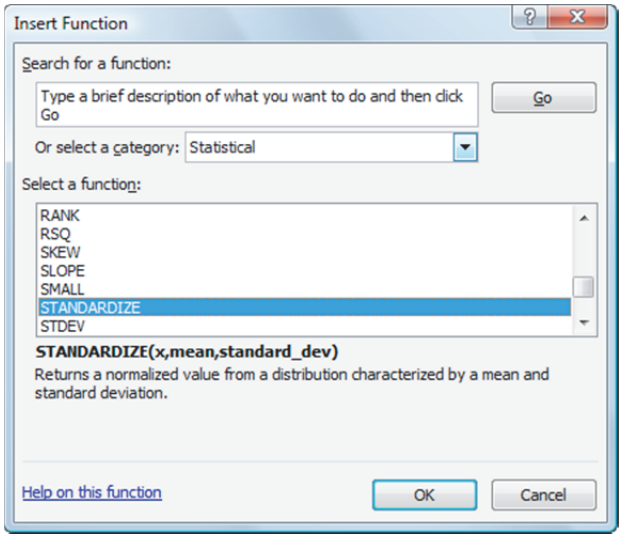


By clicking on the down arrow, you can access the **More Functions** menu. Choose the **Statistical Function Category** and cursor down until you reach the function name **STANDARDIZE** (see Figure 2.42). The PERCENTILE function has the form:

STANDARDIZE (x, mean, standard_dev)

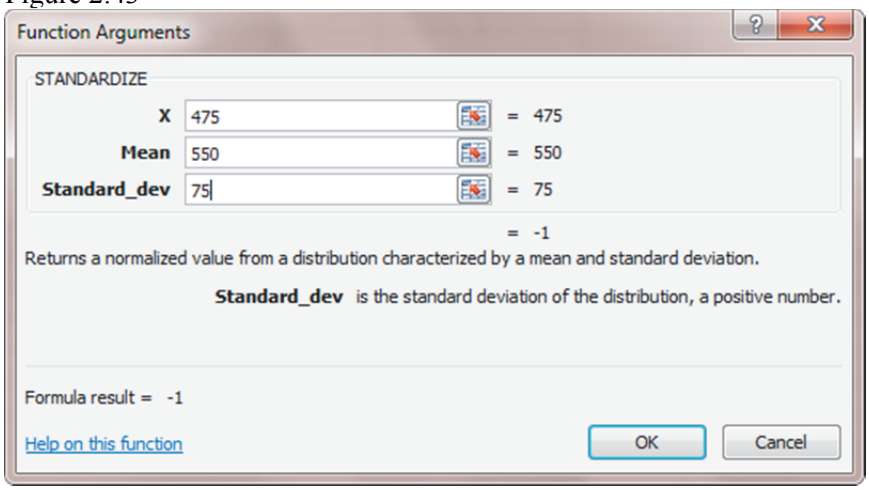
where **x** represents the value that you wish to determine the z-score for,
mean represents the mean of the data set that you want to find the z-score for, and
standard deviation represents the standard deviation of the data set that you want to find the z-score for.

Figure 2.42



For this example, we use the value of **475** as our choice for **X** in the STANDARDIZE function (see Figure 2.43). From the data provided in the example, we know to use a value of **550** for the **mean** and a value of **75** for the **standard deviation**. Click **OK.**

Figure 2.43



Excel returns a value -1. We make the interpretation that a verbal SAT score of 475 would fall one standard deviation below the mean verbal SAT score of all 2,000 students. By changing the values of X, Mean, and Standard Deviation, we can find z-scores for a wide variety of situations.

2.4 Technology Lab

The Technology Lab consists of problems for the student to practice the techniques presented in each lesson. Each problem is taken from the homework exercises within the *Statistics* text and includes an **Excel** data set that should be used to create the desired output. The completed output has been included with each problem so that the student can verify that he/she is generating the correct output.

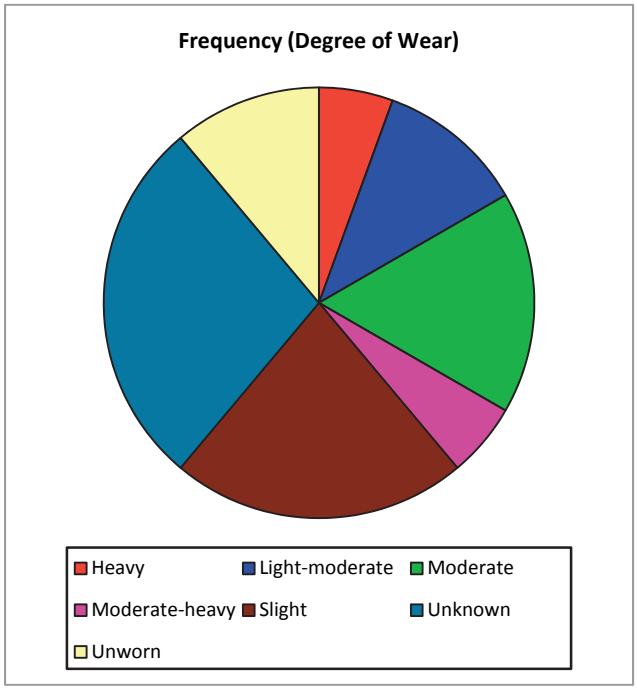
Cheek teeth of extinct primates. The characteristics of cheek teeth (e.g. molars) can provide anthropologists with information on the dietary habits of extinct mammals. The cheek teeth of an extinct primate species was the subject of research reported in the American Journal of Physical Anthropology (Vol. 142, 2010). A total of 18 cheek teeth extracted from skulls discovered in western Wyoming were analyzed. Each tooth was classified according to degree of wear (unworn, slight, light-moderate, moderate, moderate-heavy, or heavy. In addition to degree of wear, the researchers recorded the dentary depth of molars (in millimeters) for the teeth. The data is shown in the accompanying table.

Degree of Wear	Dentary Depth	Degree of Wear	Dentary Depth	Degree of Wear	Dentary Depth
Unknown	18.12	Moderate-heavy	15.76	Unworn	13.25
Unknown	19.48	Moderate	17	Light-moderate	16.12
Unknown	19.36	Slight	13.96	Light-moderate	18.13
Moderate	15.94	Slight	16.55	Moderate	14.02
Slight	15.83	Slight	15.7	Unworn	14.04
Unknown	19.7	Heavy	17.83	Unknown	16.2

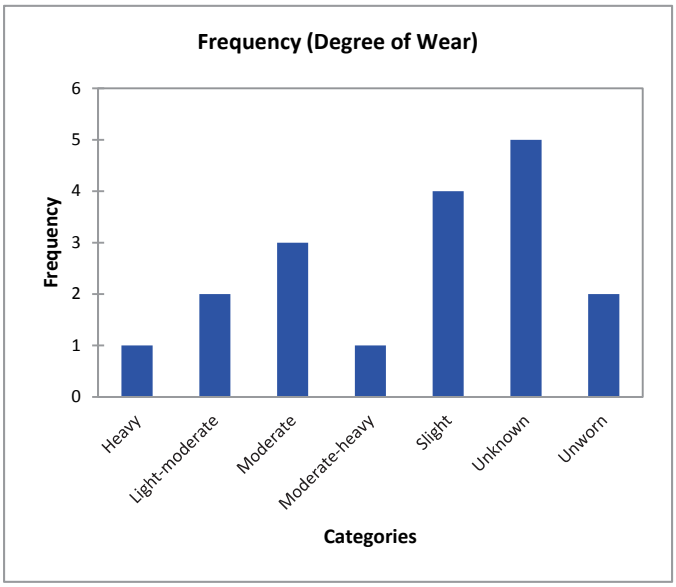
a. Construct a pie chart for the Degree of Wear categories.
b. Construct a bar graph for the Degree of Wear categories
c. Construct a stem-and-leaf display for the Dentary Depth measurements.
d. Construct a histogram for the Dentary Depth measurements.
e. Construct a boxplot for the Dentary Depth measurements to identify any outliers in the data set.
f. Calculate the mean, median, standard deviation, and variance for the 18 dentary depth measurements.
g. Find the 70^{th} percentile of the Dentary Depth measurements.
h. Find the z-score for a tooth with a Dentary Depth measurement of 15.7 millimeters.

XLSTAT Output

a.



b.



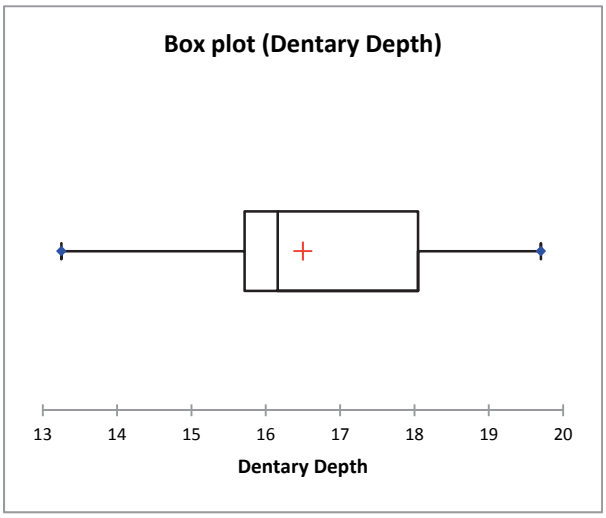
c.

Stem-and-leaf plot (Dentary Depth):

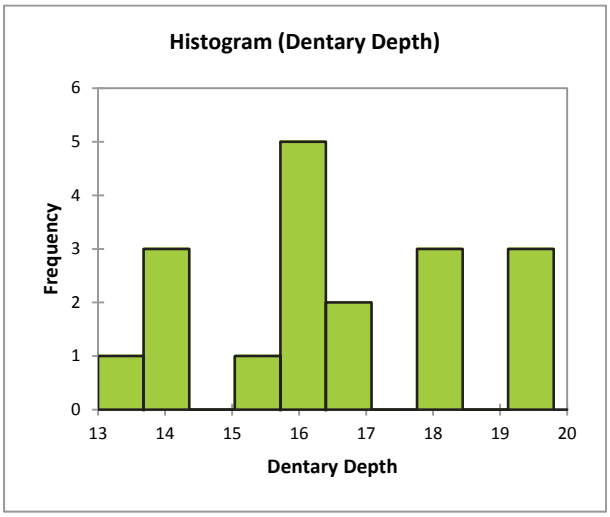
Unit: 1

13	2 0
14	0 0
15	7 8 8 9
16	1 2 5
17	0 8
18	1 1
19	4 5 7

d.



e.

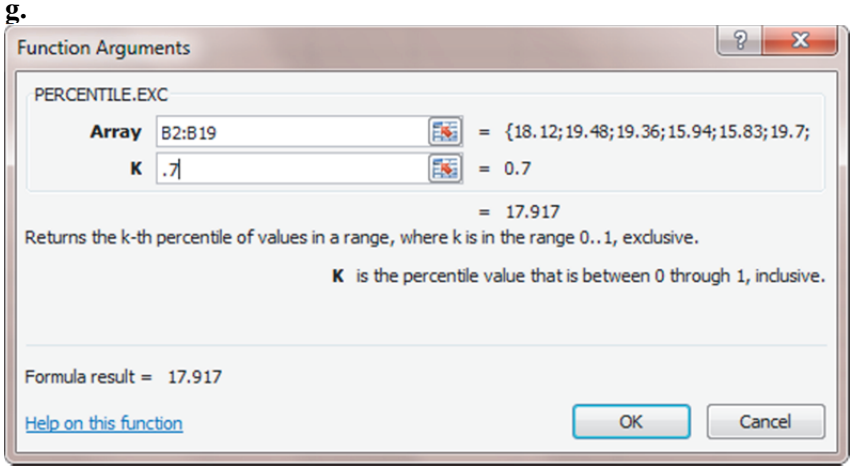


f.

Descriptive statistics (Quantitative data):

Statistic	Dentary Depth
No. of observations	18
Minimum	13.2500
Maximum	19.7000
1st Quartile	15.7150
Median	16.1600
3rd Quartile	18.0475
Mean	16.4994
Variance (n-1)	3.8825
Standard deviation (n-1)	1.9704

g.



h.