

Numerical Descriptive Measures

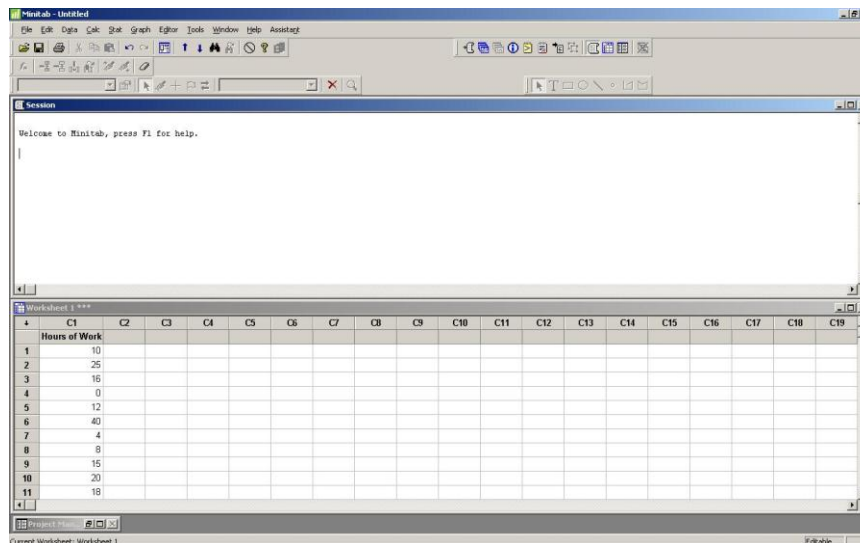
Chapter **3**

Measures of Central Tendency and Measures of Dispersion

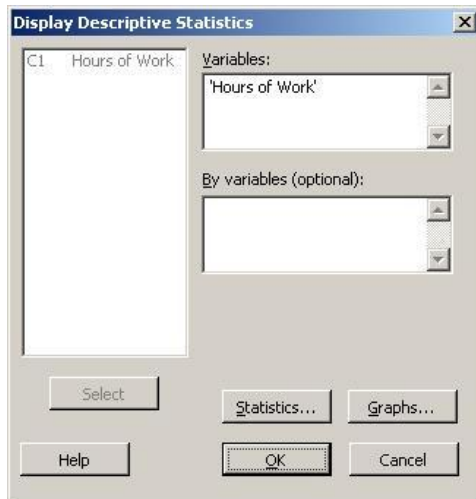
A sample of 40 students at a university was randomly selected, and each student was asked the number of hours spent working each week. The raw data is shown in the table below.

10	12	15	0	5	42	35	32	10	16
25	40	20	16	24	9	14	15	21	13
16	4	18	22	10	6	30	25	15	10
0	8	12	15	23	36	25	16	24	12

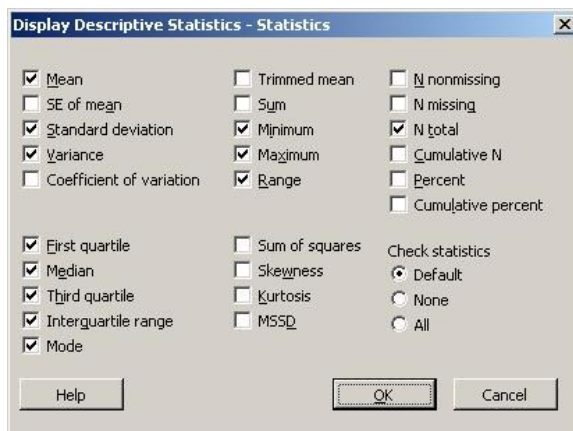
Enter the data into a Minitab worksheet. Enter the number of hours of work into column **C1** and label the column “Hours of Work” in the gray cell.



To find the descriptive statistics of this data, click on **Stat → Basic Statistics → Display Descriptive Statistics**. Either double-click the Hours of Work column name from the left side of the screen or type in **C1** into the **Variables** field.



Click on the **Statistics** button. A list of each statistic that Minitab will calculate will open. Select the checkbox next to the statistics desired, or deselect the checkbox next to the ones that are not desired. For the measures of central tendency, the Mean, Median and Mode are available options described in the textbook. For the measures of dispersion, the Standard Deviation, Variance, First and Third Quartiles, Interquartile Range, Minimum, Maximum, and Range are available options described in the textbook. There are numerous other statistics available to choose, if desired. When finished selecting, click **OK**. Click **OK** on the **Display Descriptive Statistics** window.



The selected statistics will be displayed in the Session Window.

The Total Count of 40 shows that there are 40 students in the sample.

The Mean of the sample is 17.52 hours of work per week.

The StDev of 10.22 is the standard deviation of the sample.

The Variance of the sample is 104.41.

The five number summary of the data set (the Minimum, Q1, Median, Q3, and Maximum) are 0, 10, 15.5, 24, and 42.

The Range of the data is 42.

The IQR of 14 is the Interquartile Range.

There are 3 Modes of the data: 10, 15, and 16, each with 4 observations.

The screenshot shows the Minitab interface. The 'Session' window displays the following descriptive statistics for 'Hours of Work':

Variable	Total Count	Mean	StDev	Variance	Minimum	Q1	Median	Q3
Hours of Work	40	17.52	10.22	104.41	0.00	10.00	15.50	24.00

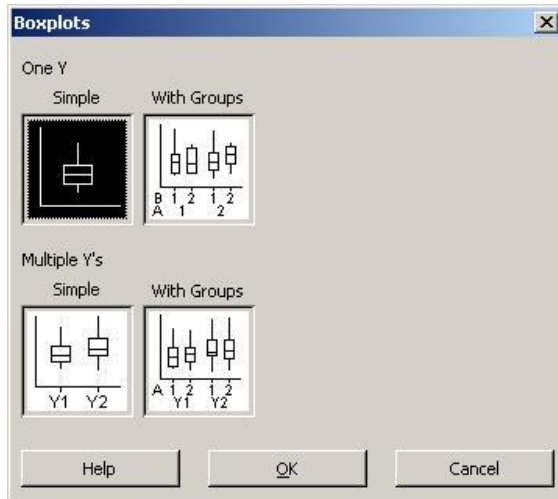
Variable	Maximum	Range	IQR	Mode	N for Mode
Hours of Work	42.00	42.00	14.00	10, 15, 16	4

The 'Worksheet' window shows the following data for 'Hours of Work':

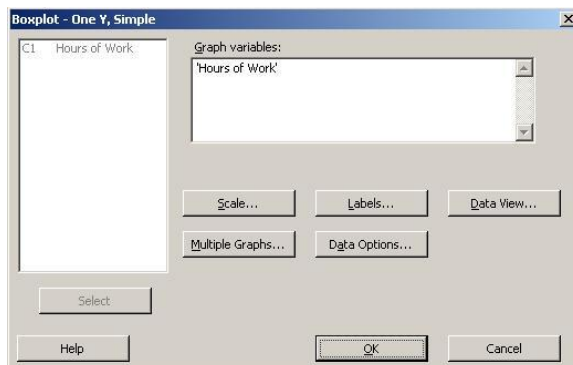
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
1	10																		
2	25																		
3	16																		
4	0																		
5	12																		
6	40																		
7	4																		
8	8																		
9	15																		
10	20																		
11	18																		

Boxplots

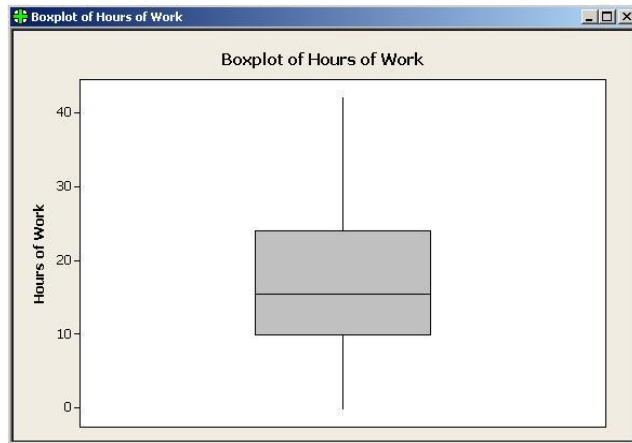
To demonstrate Boxplots, the student hours of work data from above will be used. To create the Boxplot, click on **Graph → Boxplot**. Click on and highlight the **One Y, Simple** Boxplot on the top left of the window, then click **OK**.



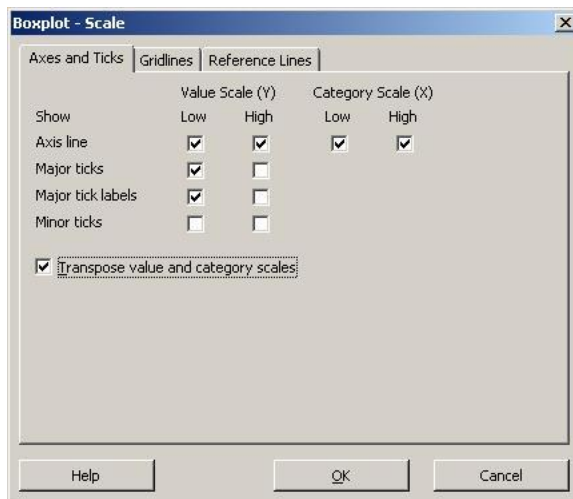
On the screen that appears, either double-click the Hours of Work column name from the left side of the screen or type in **C1**.



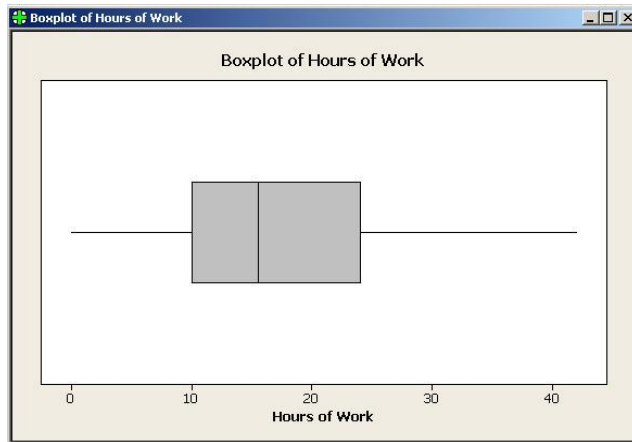
The default boxplot will name the chart “Boxplot of *Graph Variable*.” To give the chart a more meaningful name, click on the **LABELS** button. In the window, enter your new name for the chart in the **Title** field, adding any **Subtitles** if necessary. When finished click **OK**. Click **OK** on the **Boxplot, One-Y, Simple** window. The Boxplot will be displayed.



Notice that Minitab defaults a vertical boxplot while the textbook uses a horizontal boxplot. To construct a horizontal Boxplot, click on **Graph → Boxplot**, choose the **One Y, Simple** Boxplot on the top left of the window, then click **OK**. After selecting the data for the **Graph Variables** field, click on the **SCALE** button. Select the checkbox next to **Transpose value and category scales**, then click **OK**.



The horizontal Boxplot will be displayed.

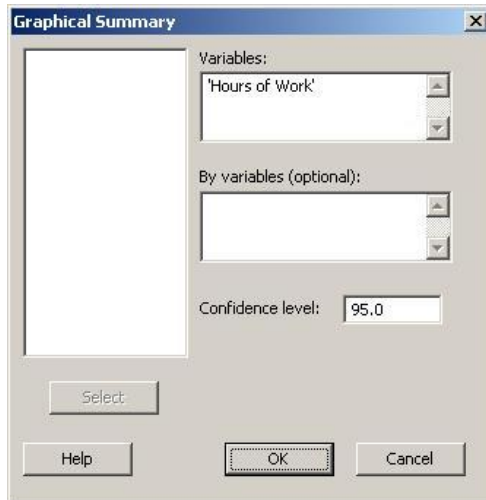


If you want to print the graph, click on **File → Print Graph** while the graph is open. Once the graph prints, you can close the Graph Window by clicking on the **X** in the upper right corner of the window of the graph window. If you want to save the graph, you may select **Save** to do so, or you may close without saving.

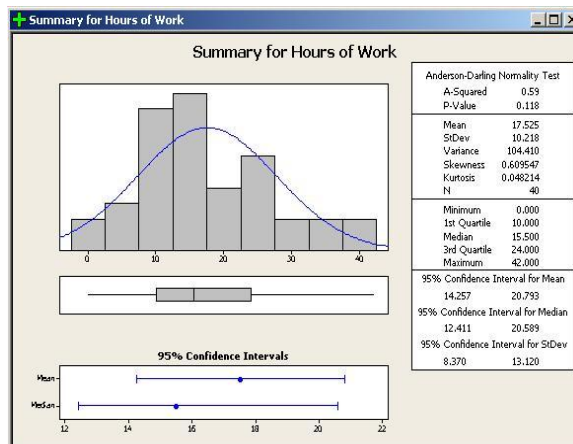
You can edit many options on the graph by right clicking on the part that you want to change. A window of options will appear from which you can choose. To see the actual data used to draw the Boxplot, hold the cursor over each bar and the Q1, Median, Q3, IQR, and endpoints of the whiskers will be displayed. If there are any outliers, hold the cursor over the outlier to see the data value of each.

Graphical Summary

Another method of calculating the default descriptive statistics, as well as viewing the boxplot and histogram of the data is available. Using the same data, click on **Stat** → **Basic Statistics** → **Graphical Summary**. Either double-click the Hours of Work column name from the left side of the screen or type in **C1** into the **Variables** field. When finished, click **OK**.



The resulting summary contains a histogram of the data, a boxplot of the data, the Mean, Standard Deviation, Variance, and N (sample size), as well as the Five Number Summary. The graphs of the histogram and boxplot may be edited by right-clicking on the part of the graph to be changed and selecting the options.



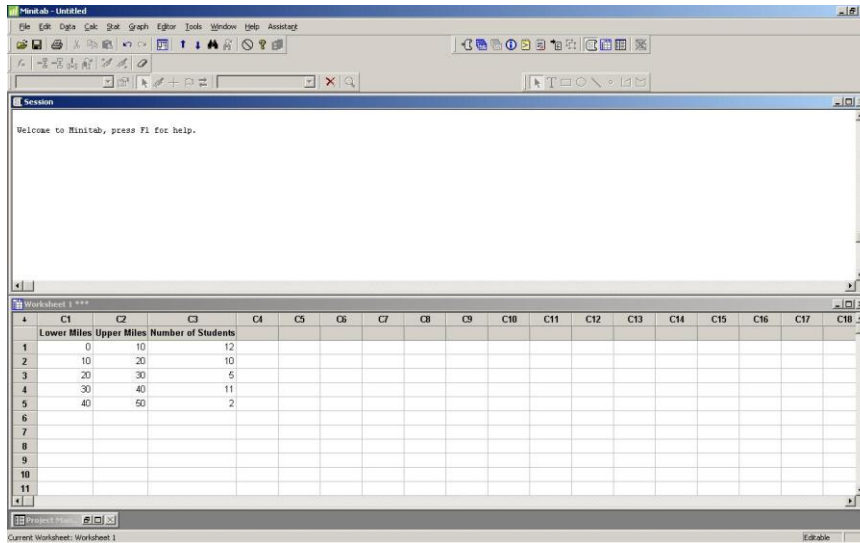
If you want to print the graph, click on **File → Print Graph** while the graph is open. Once the graph prints, you can close the Graph Window by clicking on the **X** in the upper right corner of the window of the Graph Window. If you want to save the graph, you may select **Save** to do so, or you may close without saving.

Computing the Mean for Grouped Data

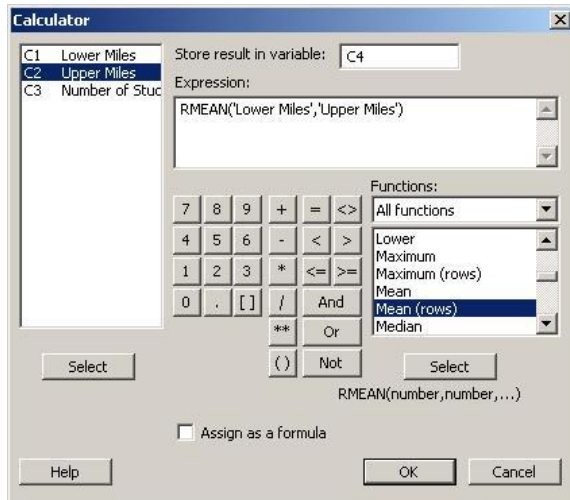
The following table gives the frequency distribution of the daily commuting distance (in miles) from home to school for 40 students at a university:

Daily Commute (miles)	Number of Students
0 to less than 10	12
10 to less than 20	10
20 to less than 30	5
30 to less than 40	11
40 to less than 50	2

Enter the lower class limits into column **C1**, the upper class limits into column **C2**, and the Frequency into column **C3** of a Minitab worksheet.



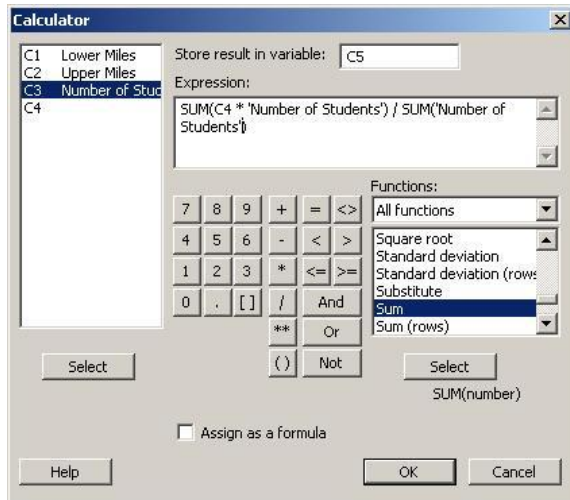
To calculate the midpoints of each interval, click on **Calc** → **Calculator**. **Store result in variable C4** and enter the **Expression** `RMEAN('Lower Miles', 'Upper Miles')`. This can be performed through entering the information by hand, or by highlighting and selecting the function for Mean (rows), and then highlighting and selecting column **C1** (Lower Miles) and **C2** (Upper Miles). This will calculate the midpoints of each interval. Click **OK**.



Column **C4** will contain the midpoints.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
1	0	10	12	5														
2	10	20	10	15														
3	20	30	5	25														
4	30	40	11	35														
5	40	50	2	45														
6																		
7																		
8																		
9																		
10																		
11																		

To calculate the mean of the grouped data, click on **Calc** → **Calculator**. **Store result in variable C5** and enter the **Expression** $\text{Sum}(\text{C4} * \text{C3}) / \text{Sum}(\text{C3})$. This multiplies the midpoint of each interval by the frequency of each interval, sums them, and then divides by the total frequency. This can be performed through entering the information by hand, or by highlighting and selecting the function for Sum and the columns C4 and C3 (Number of Students). When finished, click **OK**.



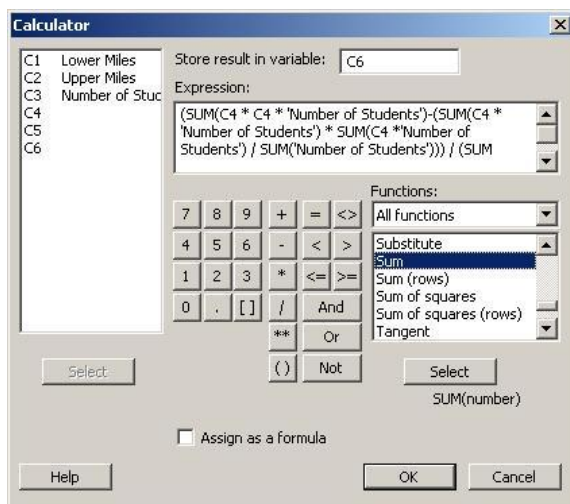
The mean will be displayed in column **C5**. The mean for this dataset is 20.25 miles.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
1	0	10	12	5	20.25													
2	10	20	10	15														
3	20	30	5	25														
4	30	40	11	35														
5	40	50	2	45														
6																		
7																		
8																		
9																		
10																		
11																		

Computing the Sample Variance and Standard Deviation for Grouped Data

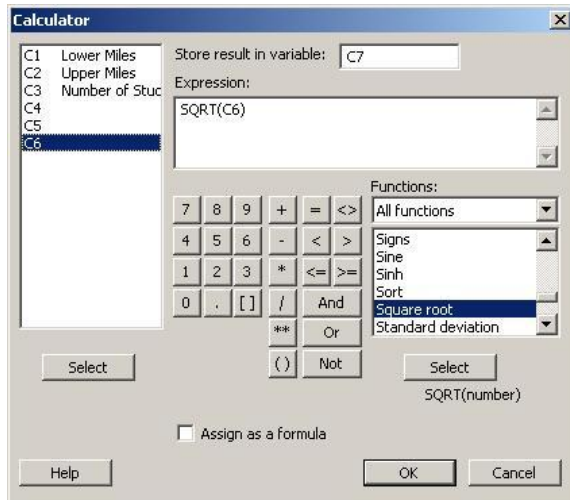
The above example of the commuting distances for students will be used to calculate the variance and standard deviation for grouped data. The shortcut formulas will be used in the calculations for Minitab. Click on **Calc** → **Calculator**. **Store result in variable C6** and enter the **Expression**

$(\text{SUM}(\text{C4} * \text{C4} * \text{'Number of Students'}) - (\text{SUM}(\text{C4} * \text{'Number of Students'}) * \text{SUM}(\text{C4} * \text{'Number of Students'}) / \text{SUM}(\text{'Number of Students'})) / (\text{SUM}(\text{'Number of Students'}) - 1)$. Click **OK**.



The sample Variance will be displayed in column **C6**. The sample variance of this data set is 174.295.

To calculate the sample Standard Deviation, **Calc** → **Calculator**. **Store result in variable C6** and enter the **Expression** $\text{SQRT}(\text{C6})$. Click **OK**.



The sample Standard Deviation will be displayed in column **C7**. The sample Standard Deviation of this data set is 13.2021.

The screenshot shows the Minitab Worksheet with the following data:

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
1	0	10	12	5	20.25	174.296	13.2021											
2	10	20	10	15														
3	20	30	5	25														
4	30	40	11	35														
5	40	50	2	45														
6																		
7																		
8																		
9																		
10																		
11																		

Suggested Exercises

Section 3.1

3.11, 3.12, 3.15, 3.23

Section 3.2

3.46, 3.48, 3.53

Section 3.3

3.64, 3.66

Section 3.5

3.93, 3.96, 3.97

Section 3.6

3.101, 3.102, 3.105

Supplementary Exercises

3.111, 3.112, 3.113, 3.120, 3.136, 3.140

Technology Assignments

TA 3.1, TA 3.11