

## Minitab Exercise 2 – Describing Variation

Rev 2/16 Minitab 17

In this session you will learn to

- Generate random data
- Create a stem and leaf diagram
- Perform basic statistics
- Create a histogram
- Create a histogram under charts with modified attributes
- Edit a chart

Note: You may type into this document for submission of the assignment.

### Step 1. Start Minitab

Start **Minitab 17**. If you are already in Minitab, choose **File>New>Minitab Project** to get an empty project.

### Step 2. Generating Random data

To get data for this exercise we are going to generate some random distributions within Minitab.

To generate the red data

**Calc>Random data>Normal**

Generate: **500** rows of data

Store in column: **Red**

Mean: **100**

Standard Deviation: **5**

For the Green data, generate 500 rows with mean 100 and standard deviation 10.

### Step 3. Generate Stem and Leaf diagram

Let's generate a stem and leaf diagram for the red data.

**Graph>Stem and Leaf**

Variable: **Red**

**OK**

Notice that the “stem” values do not have to be unique numbers, but there will be different “leaf” numbers in them where the stem number is the same.

What is the minimum value: 84 maximum: 113

What is the interval for the leaf third from the top? (Numbers in what range are in it?) 88-89

How many bins (intervals) are on your stem and leaf diagram? 15

What do the numbers on the far left tell you? (Hint: Go back into the Stem and Leaf dialog box and click on Help.) Counts - row with ( ) has median in it

Observations below median - cumulative counts up to median row starting with lowest observation.

Above median - cumulative counts from largest observation

Example Solution

*Because you are working with random data, your answers will vary from this solution.*

#### Step 4. Compute Basic Statistics and create a histogram.

We want to compute statistics and display our data as a histogram. We can run both distributions at the same time.

**Stat>Basic Statistics>Display Descriptive Statistics**

Highlight **C1 Red** and click **Select** to make **Red** your variable column. (do the same for Green)

Click on **Graphs**.

Select **Histogram of data**. Click **OK**

Click on **Statistics**. Verify that the statistics in the table below are checked.. Click **OK**

Click **OK**

The histogram of your data will appear on your screen.

In the Session window, you will see the descriptive statistics displayed. To move the session window in front of the graph window, select **Window>Session**. (You could also resize the graphic and move it to see the session window. Or if you can see part of the Session window just click on it and it will bring it forward.)

Enter the data from the session window in the appropriate column.

		Red	Green
N total	Number of data observations	500	500
Mean	Average of data observations	99.628	100.88
Median	Value for which there is an equal number of observations above and below	99.605	100.94
St Dev	Standard deviation = $\sqrt{\sum (x - \bar{x})^2 / (n-1)}$	4.940	9.51
Minimum	Smallest value	84.099	75.21
Maximum	Largest value	113.761	137.32
Q1, Q3	Quartiles. To calculate quartiles, the data is ordered from smallest to largest. Q1 is the observation at position (N+1)/4. Q3 is the observation at 3(N+1)/4. (Note: Minitab interpolates)	Q1 96.465	95.17
		Q3 102.40	137.32

Do the min and max for red shown here agree with your answer in Step 3? yes - but to more decimal places

Compare the red histogram to the red stem & leaf diagram. How are they similar? Different? (Consider numbers of bins and centering of data)

very similar: stem+leaf has 15 bins, histogram has 16  
"bell shaped" distribution with center close to 100

Compare the red and green distribution histograms. Notice that they have a different scale which makes it difficult to directly compare them. Creating graphics under Descriptive Statistics is fast but rather limited. While you could do considerable editing to make the comparable, it is really easier to use Graphics.

## Step 5. Using Graphics to create Histograms

Let's create graphs with the same scale for both distributions so that we can more easily compare them.

Choose **Graph>Histogram>Simple**

We can graph both histograms on the same chart, so for the **graph variables** select both Red and Green.

Click **Multiple Graphs**.

Select **On separate Graphs**

For the scales, Select **Same Y** and **Same X, including same bins**.

Click **OK**.

With **Labels**, you can add titles, footnotes and other text and data labels. These things you can also add later using the graph editing tools. For now, don't use this.

Click **OK** to generate the graph.

*Discuss/compare the two graphs now. Why would you want to use the same scales/bins?*

*Mucheaster to compare both centering and spread with same scale. Can easily see that green has higher standard deviation (spread),*

## Step 6 Editing the graph

We will just edit the green data histogram.

Notice that the X axis is labeled with the name of your column. We need to rename this.

Double click on the X axis label. Enter a new label such as Observation Values. OK.

You can resize the text box or move it around by dragging the edges or the box.

Edit the title the same way to something appropriate. Increase the text size to make it more prominent.

Double click on the bars to get the **Edit bars** dialog box. Under **Attributes**, give your bars a custom fill color.

While you are in the **Edit bars** dialog box, click on **Binning**. This is where you could change the number of bins to see if changing the bins will give you a better depiction of the distribution. Change the **Interval Definition** from **Automatic** to **Number of Intervals**. Enter a number of intervals that is double what is currently in the box.

*How did changing the number of bins help/hurt in evaluating if the distribution is symmetric or not?*

*Did not change much in this case. But spike at 105- much more dominant in 2nd graph.*

Back on your graph, double-click on the numbers of the Y scale to get the **Edit Scale** dialog box. Click on each of the tabs to see what you can do. Under the **Alignment** tab, change the text angle to 15.

Stem-and-leaf of Red N = 500  
Leaf Unit = 1.0

Stem-and-leaf of Red N = 500  
Leaf Unit = 1.0

### Descriptive Statistics: Red, Green

Variable	Total Count	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Red	500	99.628	4.940	84.099	96.465	99.605	102.968	113.761
Green	500	100.88	9.51	75.21	95.17	100.94	106.40	137.32



