**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ AP Stats – Lab 5 – Instructions**

**Is It Normal? Assessing the Normality of a Data Set**

In order to analyze a data set using the methods discussed in Chapter 2, we must deem it appropriate to model the data set with a Normal curve. Your textbook proposes two methods for doing so (paraphrased):

*(1) Construct a frequency histogram or stemplot. See if the graph is approximately bell-shaped and symmetric about the mean. Then compare the count of observations in each interval with the 68-95-99.7 rule. Note that smaller data sets rarely fit the 68-95-99.7 rule well, and this is even true of observations taken from a larger population that really has a Normal distribution, due to variability.*

*(2) Construct a normal probability plot and assess the linearity of the plot.*

In this lab, you will use Fathom to assess normality with the both methods.

1. Open Fathom and open the file, “Lab5\_IsItNormal” from your student drive (Chapter 2 folder).
2. Click File > Show Page Breaks. Change the page orientation to landscape under File > Print Setup.
3. I have provided a data set with different bear “attributes” (length, head length, weight, etc.). Choose your favorite of the quantitative measurement attributes to analyze; you will only work with one attribute.
4. I have also provided a histogram with some values/formulas shown for the ID attribute (just as a place-holder). Drag your attribute of choice into the x-axis of the graph to replace the ID attribute.
5. I have added plotted values on the histogram for , as well as , , and . These values should have updated automatically for your new attribute; you should see the numeric values below the histogram and a line on your histogram representing each value.
6. Now to the analysis: first, is your graph is *approximately* bell-shaped and symmetric about the mean? You can modify bin widths and scale to further investigate. Comment on the shape of the distribution and then continue either way. You will eventually print out your Fathom output to support your answers.
7. Next, compare the count of observations in each interval to the 68-95-99.7 rule. I have included a “count()” value that tells you the sample size, as well as another “count()” value that you can manipulate to give you precise counts in the desired intervals. You will need to double click the formula and change the ID attributes to your attribute of choice, and you will need to change the other numeric values (include units!) to reflect the bounds of your desired intervals.
   1. How many observations are in the interval ? What percent of the sample is this?
   2. How many observations are in the interval ? What percent of the sample is this?
   3. How many observations are in the interval ? What percent of the sample is this?
8. I have included a function to plot a normal density curve against your data for further comparison (look for “frequency of…”). You will need to double click the function and change the ID attributes to your attribute of choice. (Note that the number in front of the normalDensity function is just a scaling factor for the height of the Normal function to appear on the same graph as the histogram. Trial and error will allow you to pick a value that appears about right.)
9. Finally, construct a Normal Probability Plot. I have added an attribute to your data set called Expected ZScores. Right click on this attribute and select “Edit Formula.” You will need to replace the ID attribute with your attribute of choice. Now, drag in a new “Graph,” and plot your measurement attribute of choice on the x-axis. Drag the Expected ZScores attribute to the y-axis. Lastly, drag in a new “Summary” table from your menu and drag the same two attributes into the table to display the correlation coefficient.
10. Taking all of the above into account, do you think it is appropriate to model your chosen data set (attribute) with a Normal curve? Explain your reasoning.