**Math 155 – Minitab instructions for correlation and regression**

**10.2: Correlation** – Finding the linear correlation coefficient

Enter the data set into the first two columns of Minitab – “x” in column C1, “y” in column “C2.” Then, select Stat > Basic Statistics > Correlation…



After you click here, a “Correlation” window will pop up. In the example below (see next page), we will be looking at the BEARS.MTW from the Minitab Sample Data folder; based on the data, we will determine whether there is a statistically significant correlation between weight (in pounds), column C10, and neck girth (in inches), column C7.



Select “Pearson correlation” from the drop-down menu. Leave “Display p-values” checked. Click “OK.”

Now go to the Session Window, and you’ll see the following:

**Correlation: Weight, Neck.G**

Pearson correlation of Weight and Neck.G = 0.943

P-Value = 0.000

The result 0.943 is the correlation coefficient. The way we interpret this is by its absolute value (sometimes it is negative) – the closer the absolute value is to 1, the stronger the evidence of correlation. The reported P-Value, 0.000, confirms this as well. The P-Value refers to the test of H0: no correlation, vs. H1: correlation; with such a low P-Value, we would certainly reject the null hypothesis, and thus accept the conclusion that weight and neck girth are correlated.

**10.3: Regression** - Finding a linear relationship between two variables

Continuing the preceding example – now that we’ve established a correlation between weight and neck girth, let’s now look for a way to predict weight based on neck girth. We’ll do this by looking for a linear relationship between the two: that is, an equation of the form $y=b\_{0}+b\_{1}x$, where y=weight and x=neck girth. With this setup, neck girth would be the “independent variable,” x, and weight would be the “dependent variable,” y. (Note: you may be more familiar with the equation “$y=mx+b$” – the above equation is the same, except the “slope” is called $b\_{1}$ rather than $m$ in this context.)

First, select Stat > Regression > Fitted Line Plot…



This will take you to the “Fitted Line Plot” window. Here, select your “X” and “Y” columns as appropriate – again, the idea is that you will be looking for an equation that predicts Y, given a value of X. Following the setup described at the top of this page, we’d select these as follows:



Note: leave the “Linear” circle selected. (The other regression options shown here – “quadratic” and “cubic” – are interesting, but beyond the scope of this course!) Click “OK”...

…and you should see this:



What we are seeing is a scatterplot of Neck Girth (X) vs. Weight (Y). (You could also get this, without the fitted line, by selecting “scatterplot” from the “graphs” menu.) The red line is the “best fit” line for the given points.

As is always the case for a linear relationship, one should interpret “slope” as the rate of change of Y with respect to X. In this case, the slope of 20.57 tells us that we expect a bear’s weight to increase by 20.57 pounds for every 1 inch increase in a bear’s neck girth.

The “regression equation,” $Y=-246.5+20.57X$, may be used to predict the weight, Y, of other bears (from the population of bears from which the original sample was selected) using neck girth, X. For example, if another bear in the population is found to have a neck girth of 25 inches, then we may expect that bear’s weight to be about $-246.5+20.57\left(25\right)=-246.5+514.25=267.75$ pounds.

(NOTE: As discussed in Section 10.3, regression equation’s predictions are only appropriate if correlation has been shown. In an example where correlation is not shown (that is, if we had not gotten a low enough P-Value from the test on the second page of this handout), then the best prediction we could come up with for a value of Y would be the sample mean for Y.)