MONT 105Q – Mathematical Journeys
Excel Practice on Correlation Coefficient Computations
April 13, 2016

*Directions*

Follow the steps below to create a spreadsheet illustrating the computations done to create the table on page 66 of *Naked Statistics*.

I. Open Excel and download the blank spreadsheet file from the course homepage. Note that it has column headings as Wheelan gives in his table.

II. Compute the means (AVERAGE) and standard deviations (STDEV) and put them in the appropriate spaces below the data. You should see a *difference* between your computed values and the ones that Wheelan gives at this point. Do you understand why that happened? The formula that Excel uses to compute the standard deviation is

$$s = \sqrt{\frac{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n-1}}.$$

Do you see what is happening now? How could you fix your formulas in the cells B19 and C19 to get Wheelan's numbers?

III. Compute the height in standard units (i.e. the $z$-score for the height) for Nick. Set up your formula so that you can copy and paste into the other cells in column D to repeat the calculation for the other students. (This will require some "address magic" to make sure you're always using the same numbers for the mean and the SD.) Then repeat this for the weights in column E.

IV. Now compute the product of the height in standard units and the weight in standard units for Nick, copy and paste into the other cells in column F. The *correlation coefficient*, $R$ for the height and weight data is then the total of the numbers in column F in rows 2 through 16. Compute this and adjust to 3 decimal places.

V. This computation can also be done with one Excel command, of course. (That's the way you would always do this sort of computation in the future, but it helps to know what's actually going on "under the hood.") Check your work by entering `=CORREL(B2:B16,C2:C16)` in an unused cell.

VI. Finally, insert a *scatterplot* of the data points for each student showing the height as the $x$-coordinate and the weight as the $y$-coordinate. Add a "trendline" and show the formula and the $R^2$ value on your plot. How is this $R^2$ value related to your correlation coefficient?