

Project 3

Math 121

*Please type your solutions to the questions below in Microsoft Word or Google Docs (or other document editor), then print your document. It is due in class on **Thursday, December 1st**. Use complete sentences to write your answers. It is okay to discuss the problems with other students, but all of your solutions must be explained in your own words.*

A PE teacher at one middle school wanted to know if there was a correlation between the number of push-ups seventh-graders could do and their mile-run times. She collected the data in the following file:

<http://people.hsc.edu/faculty-staff/blins/StatsExamples/PEclass.xlsx>

1. Open the file above in a spreadsheet (either Excel or Google Sheets will work). Use the spreadsheet to make a scatterplot that shows the relation between push ups and times on the mile run. You should include a (clearly labeled) copy of the scatterplot in your document.
2. Use the function =CORREL() to find the correlation coefficient for the scatterplot.
3. What is the average and the standard deviation of the number of push-ups that a middle school student can do? Use =STDEV() to find the standard deviation.
4. What is the average and the standard deviation for the mile run times?
5. Find the formula for the least squares regression line.
6. Use the least squares regression line to predict the average mile run times of students who can do 40 push ups and also for students who can't do any push ups.
7. Explain the meaning of the slope of the least squares regression line (including its units) in words.
8. Because the slope is a statistic, not a parameter, it might not accurately represent the population due to random error. You can make a confidence interval for what the slope is in the population of all 7th graders by using the following formula:

$$m \pm t^* \frac{m\sqrt{1-R^2}}{R\sqrt{n-2}}$$

where the critical t^* value has $n - 2$ degrees of freedom. Here m is the slope based on the sample, R is the correlation coefficient, and n is the combined sample size which counts all of the students since we aren't comparing boys versus girls. Find a 95% confidence interval for the slope.