Project 2 **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 ***Monday, Nov 17****. 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.*

The spreadsheet linked below contains data from the 2018 High Bridge half marathon. It includes the race times (in minutes), as well as the age and gender of the runners.

`http://people.hsc.edu/faculty-staff/blins/classes/spring23/math121/halfmarathon.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 age and time in minutes to finish the half marathon. 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.

> **Solution:** The correlation coefficient is $R = 0.364$.

3. Find the formula for the least squares regression line. You can use Excel or Google Sheets to get the formula for the regression line from the scatterplot.

> **Solution:** The least squares regression line is $y = 0.862x + 100.6$ where y is time to finish in minutes and x is the runners age.

4. Use the least squares regression line to predict the average half marathon times of runners who are 20 years old and also for runners who are 70 years old.

> **Solution:** The predicted y-values are 117.836 minutes for a 20 year old and 160.926 minutes for a 70 year old.

5. Explain the meaning of the slope of the least squares regression line (including its units) in words.

> **Solution:** The slope is 0.862 minutes per year. It means that runners take about 0.862 minutes longer to finish on average, for every year older they get.

6. 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 half marathon runners 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 sample size. Find a 95% confidence interval for the slope.

Solution: Based on this sample, we can be 95% confident that the slope in the population is between 0.454 and 1.27 minutes per year.