CS2065 Spring 2022
**Discussion: Linear Regression (Lab 09)**

In the previous worksheet, we introduced correlation as a way of quantifying the strength and direction of a linear relationship between two variables. However, the correlation coefficient can do more than just tell us about how clustered the points in a scatter plot are about a straight line. It can also help us define the straight line about which the points (in original units) are clustered, also known as the *regression line*.

The formula for the *slope* and *intercept* for the regression line are shown below. In fact, by a remarkable fact of mathematics, the line uniquely defined by the slope and intercept below is *always* the best straight line for prediction.

$$\textbf{slope of the regression line} \ = \ r \cdot \frac{\text{SD of } y}{\text{SD of } x}$$

$$\textbf{intercept of the regression line} \ = \ \text{average of } y \ - \ \text{slope} \cdot \text{average of } x$$

Question 1. Suppose you are given the scatter diagram shown below that shows the relationship between the height of fathers and the height of children. You have calculated the line of best fit (shown in yellow). Suppose you encounter a new family where the father has a height of 70 inches. How would you predict the height of the children in that family?

Heights of Fathers vs Heights of Children
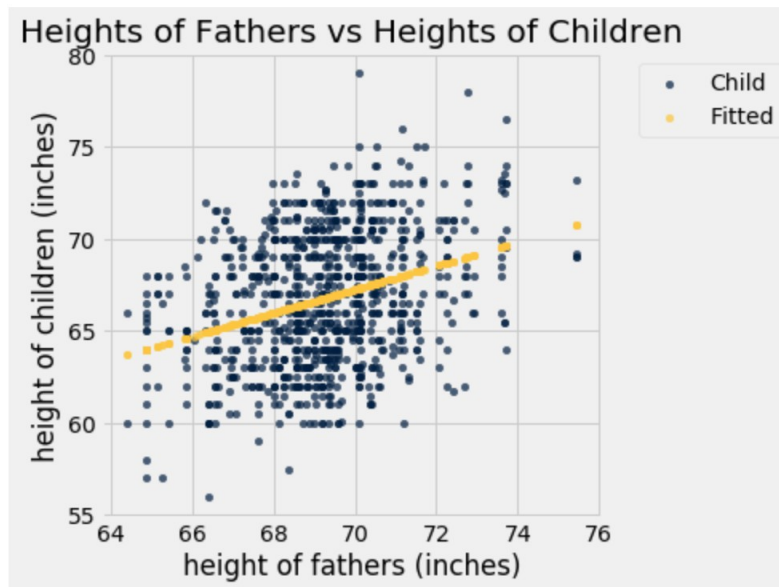
Question 2. We want to investigate the correlation between the daily ounces of coffee consumed by an individual and the number of hours the individual stayed awake on that day. It is our intention to use the ounces of coffee consumed to predict the number of hours the

individual stayed awake. The data from our sample of 500 people has the following characteristics:

- The number of ounces of coffee consumed has a mean of 12 ounces and SD of 4
- The number of hours stayed awake has a mean of 16 and an SD of 2
- The correlation between the number of ounces of coffee consumed and number of hours spent awake is 0.5.
- Suppose the scatter plot is roughly linear.

a) What is the slope of the line of best fit?

b) What is the intercept of the line of best fit?

c) Suppose your friend is in this population. She told you that she consumed 24 ounces of coffee that morning. Use your line of best fit to predict how many hours she will stay awake today.