**1.** Consider the data set given below

x <- c(0.18, -1.54, 0.42, 0.95)

And weights given by

w <- c(2, 1, 3, 1)

Give the value (a, b, c, or d) that minimizes the least squares equation

a: 0.1471  
b: 0.300  
c: 0.0025  
d: 1.077

**2.** Consider the following data set  
x <- c(0.8, 0.47, 0.51, 0.73, 0.36, 0.58, 0.57, 0.85, 0.44, 0.42)  
y <- c(1.39, 0.72, 1.55, 0.48, 1.19, -1.59, 1.23, -0.65, 1.49, 0.05)  
Fit the regression through the origin and get the slope treating y as the outcome and x as the regressor.

a: 0.59915  
b: -0.04462  
c: -1.713  
d: 0.8263

**3.** From the datasets package ‘mtcars’, fit the regression model with mpg as the outcome and weight (wt) and horsepower (hp) as the predictor. Use the matrix notation  , give the slope coefficient and adjusted as R output

**4.** Consider the following data set (used above as well).

x <- c(0.8, 0.47, 0.51, 0.73, 0.36, 0.58, 0.57, 0.85, 0.44, 0.42)  
y <- c(1.39, 0.72, 1.55, 0.48, 1.19, -1.59, 1.23, -0.65, 1.49, 0.05)

What is the intercept for fitting the model with x as the predictor and y as the outcome?

**5.** You know that both the predictor and response have mean 0. What can be said about the intercept when you fit a linear regression?

a: It must be identically 0.  
b: It is undefined as you have to divide by zero.  
c: It must be exactly one.  
d: Nothing about the intercept can be said from the information given.

**6.** Consider the data given by

x <- c(0.8, 0.47, 0.51, 0.73, 0.36, 0.58, 0.57, 0.85, 0.44, 0.42)

What value minimizes the sum of the squared distances between these points and itself?

a: 0.36  
b: 0.573  
c: 0.8  
d: 0.44

 calculate the mean?

**7.** Let the slope having fit Y as the outcome and X as the predictor be denoted as . Let the slope from fitting X as the outcome and Y as the predictor be denoted as . Suppose that you divide by ; in other words consider .

What is this ratio always equal to?

a: Cor(Y,X)  
b: Var(Y)/Var(X)   
c: 1  
d: 2SD(Y)/SD(X)

Show how you obtained this result (hint, check out slide 76).

**Bonus 1**: Demonstration of the Central Limit Theorem: let , the sum of 20 independent Uniform(0,1) random variables. In R, create 1000 simulations of x and plot their histogram. On the histogram, overlay a graph of the normal density function. Comment on any differences between the histogram and the curve.

**Bonus 2**: Distribution of averages and differences: the heights of men in the United States are approximately normally distributed with mean 69.1 inches and standard deviation 2.9 inches. The heights of women are approximately normally distributed with mean 63.7 inches and standard deviation 2.7 inches. Let x be the average height of 100 randomly sampled men, and y be the average height of 100 randomly sampled women. In R, create 1000 simulations of x − y and plot their histogram. Using the simulations, compute the mean and standard deviation of the distribution of x − y and compare to their exact values.