Homework 3

Due: Thursday 2/20/20 by 8:30am

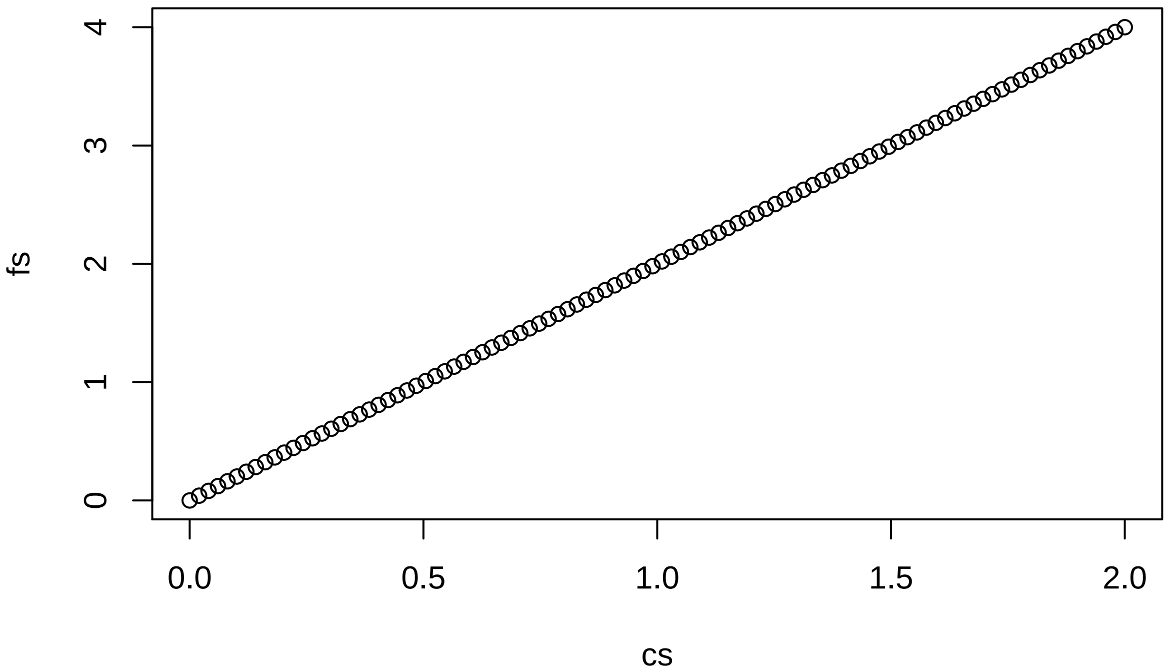
This homework assignment focuses on material covered in Chapter 1 of the textbook. You may find the following three `R` hints useful:

- All text on a line that follows the `#` sign is treated as a comment, i.e. ignored by `R` when you run the line of code.
- Suppose you are asked to plot a function of some variable, let's call it $f(c) = 2c$. Then you will need to choose the values of c to consider, and then evaluate the function at each of those values.

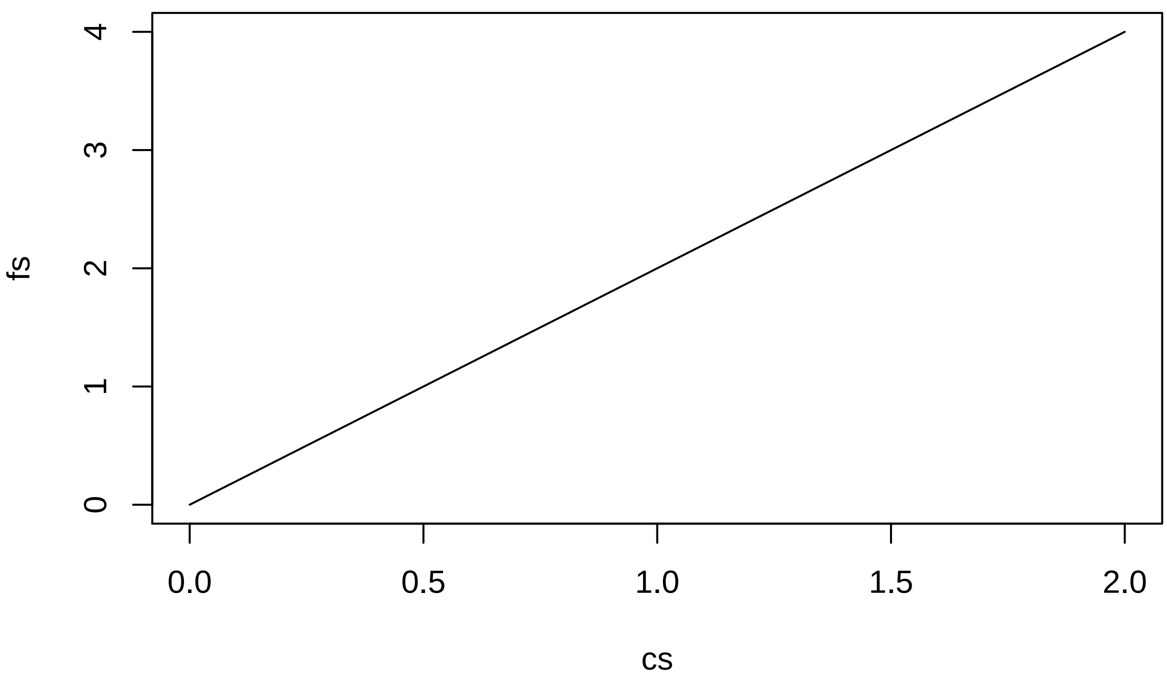
```
# First, choose values of c to consider. Let's look at values of c between 0 and 2
cs <- seq(0, 2, length.out = 100) # This returns 100 values from 0 to 2
# Alternatively: cs <- seq(0, 2, by = 0.01) # This returns values from 0 to 2,
# each 0.01 apart

# Now, we need to create an empty vector to store the function values in
# Note: We want to have the same number of function values as values of c
fs <- rep(NA, length(cs)) # This makes a vector with the same length as cs,
# Each element fs[i] for i = 1,...,100 is defined as NA, which is R's way
# of indicating a missing value

# Now we're going to go one-by-one through the values in cs and record the
# function value in fs
for (i in 1:length(cs)) {
  fs[i] <- cs[i]*2
}
# Ok! Now we can make our plot by just plotting cs and fs against each other
plot(cs, fs) # The default is to make a plot of points
```




```
plot(cs, fs, type = "l") # Sometimes we might prefer to make a line plot by connecting the points
```



- Suppose you are asked to simulate n independent normal random variables $x_1, \ldots, x_n$ with different means m_i, but the same variance $\sigma^2 = 1$. Suppose that you already have a variable `n` that gives the number of random variables you want to simulate n, and an $n \times 1$ vector `m` that gives the means for each of the n random variables. You can do this as follows:

```
x <- rnorm(n, m, 1) # Simulates the n random variables
x[1] # Extracts and prints the first value
x[2] # Extracts and prints the second value
```

1. Suppose Instagram magically knew that every time the number of times user i purchases a product,

denoted by Y_i, is related to the number of times the product has been advertised to user i, denoted by X_i, as follows:

$$Y_i = 1 + 2X_i + \epsilon_i$$

where ϵ_i is a *normal* random error term with mean $E\{\epsilon_i\} = 0$ and variance $\sigma^2\{\epsilon_i\} = 0.1$; ϵ_i and ϵ_j are uncorrelated so that their covariance is zero (i.e., $\sigma\{\epsilon_i, \epsilon_j\} = 0$ for all $i \neq j$) for $i = 1, \ldots n$.

(a) Using `R`, make a plot with three panels. You can make a single plot with three panels by typing `par(mfrow = c(1, 3))` before running any lines of code that create plots. Plot the density of the errors ϵ_i for $X_i = 0$, $X_i = 1$, and $X_i = 10$, using a separate panel for each value of X_i. Ensure that the axes are the same across all three plots.

(b) Based on the assumed model and the information provided, can we conclude that the number of times user i purchases a product Y_i is independent of the number of times user j purchases a product Y_j?

(c) Based on the assumed model and the information provided, can we state the exact probability that a single value Y_i will be greater than 4 given that $X_i = 1$?

(d) Simulate $n = 100$ observations from the model, with $X_i = i$. Using `R`, make a scatter plot of the data and overlay the regression function on the scatterplot.

(e) Repeat (d), but instead of assuming that $\sigma^2\{\epsilon_i\} = 0.1$, assume that $\sigma^2\{\epsilon_i\} = 10$. In at most one sentence, describe how increasing $\sigma^2\{\epsilon_i\}$ changes how the regression function relates to the scatter plot.

2. Problem 1.27 from the `.pdf` version of the textbook. Requires use of the `muscle` data that has been posted on the Homework page.

3. Problem 1.34 from the `.pdf` version of the textbook.

4. Problem 1.42 from the `.pdf` version of the textbook.

5. Integrative Experience Step 2, as described in `ieproject.pdf` on the Project page.