# Linear Regression
# HW 1
# Due 8/25 at 11:59pm

**Directions:** Submit a .pdf file containing your responses for the homework. The .pdf can be converted from Latex file, pictures of your handwriting solutions, word files, markdown files, etc (anything that can be converted into .pdf). If there are coding problems, only include the theory responses question in the .pdf, and upload a separate notebook for Python code.

1. Consider the following statement: "For the ordinary least squares method to be fully valid, it is required that the distribution of Y be normal." Is this statement true or false, and why?

2. Read section 1.8 of the textbook. When there is a Normal distribution assumption on the error terms, we can also formulate Maximum Likelihood Estimators for $\beta_0,\ \beta_1,\ \&\ \sigma^2$ . Use the likelihood function (1.26) to find the estimators $\hat{\beta}_0,\ \hat{\beta}_1,\ \&\ \hat{\sigma}^2$, and show that the estimators $\beta_0,\ \beta_1$, are the same as the Least Square estimators. You do not need to check second derivatives to prove the maximum values.

3. The solution for the LS estimator of the slope in simple linear regression is:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{2} x_i^2 - n\bar{x}^2}.$$

   Prove that this expression is equivalent to the alternate formulation:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{2}(x_i - \bar{x})^2} = \frac{SSXY}{SSX}.$$

4. Recall that the residual for the $i^{th}$ observation is defined as $e_i = y_i - \hat{y}_i$. Prove that $E(e_i) = 0$.

5. (a) When asked to state the simple linear regression model, a student wrote:

$$E(Y_i) = \beta_0 + \beta_1 x_i + \epsilon_i.$$

   Do you agree? Why or why not?

   (b) Consider the classical simple linear regression model. Suppose that the true parameter values are $\beta_0 = 2$, $\beta_1 = 4$, and $\sigma^2 = 9$. State the distributions of $Y$ at $x = 1, 2$, and 4, and explain how you found them.

6. Consider the Rotten Tomatoes movie rating example.

   (a) Interpret the slope and the intercept in the real-life context of the problem.

   (b) Suppose the Borderlands movie is about to be released and critics have given it a score of 8 (out of 100) on Rotten Tomatoes. Using the fitted simple linear regression line, what do we predict the audience rating will be?

   (c) What does the SLR prediction in (b) suggest about who the regression line thinks is the harsher judge: audiences or critics?

   (d) Suppose "The Quiet Place: Day One" movie is about to be released and critics have given it a score of 86 (out of 100) on Rotten Tomatoes. Using the fitted simple linear regression line, what do we predict the audience rating will be?

(e) What does the SLR prediction (d) suggest about who the regression line thinks is the harsher judge: audiences or critics?

(f) Consider your findings in (c) and (e). Provide a reasonable explanation as to how one can reconcile these two results.

(g) What value of critic ratings will the SLR model predict the exact same score for audience ratings? Derive a general formula for this value in terms of $\hat{\beta}_0$ & $\hat{\beta}_1$.