This document was edited on December 12 at 12:00
This is an optional assignment. The credit will be awarded in one of two ways:

- **Homework Replacement** Of the four homeworks assigned this semester, the lowest is already being dropped. This assignment will replace the next lowest but only if it improves your grade. Essentially, this would mean you had five homeworks and get to drop two.

- **Extra Credit** The assignment will be worth up to 5% extra credit in the course. Your score on the assignment will determine how much of the 5% it is worth. So if you get a 90% on this assignment you will have earned $0.9 * 5\% = 4.5\%$ extra credit. If you get a 20% on this assignment you will have earned $0.2 * 5\% = 1\%$ extra credit.

There is no need to choose which option you prefer - I will assign the credit based on which method improves your grade the most. However, I will not be assigning partial credit on this assignment. Each problem will marked either correct or incorrect.

Show **all** of your work on this assignment and answer each question fully in the given context.

A Bayesian model starts a lot like the statistical models we have discussed in class up to this point. As in the case of a statistical model, we treat the observable results as random variables whose distributions may depend on values that are known to us (which we call covariates) as well as parameters which are unknown to us. Assuming that we have these distributions have the capacity to reasonably describe the actual real-world mechanisms producing the results that we will observe, then our interest mainly lies in estimating the values of the unknown parameters. So for instance, a linear model with additive errors could be written as:

Linear Regression Model
With $i = 1, 2, \ldots, n$ observed values y_i based on known covariates x_i we define the relationship:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

where ϵ_i are all independent of each other.

The parameters β_0, β_1, and σ^2 in this case are unknown to us but we assume that they actually exist and are constants. We then use some method (maximum likelihood for instance) to estimate them. This is where the Bayesian approach changes. In a Bayesian analysis, we assume that the parameters are also random variables and have their own distributions. A Bayesian version of this model could be described as follows:

Bayesian Linear Regression Model
With $i = 1, 2, \ldots, n$ observed values y_i based on known covariates x_i we define the relationship:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

and prior distributions

$$\beta_0 | \sigma^2 \sim N(0, \sigma^2)$$

$$\beta_1 | \sigma^2 \sim N(0, \sigma^2)$$

$$1/\sigma^2 = \tau^2 \sim Uniform(0, M)$$

where M is largeand where ϵ_i are all independent of each other.

This assignment explores the differences between Bayesian analysis of a linear modeal and the so-called frequentist approach.

Part I: Frequentist Approach

Consider the first model above (the non-Bayesian one) with responses $y_1, y_2, \ldots, y_n$ which correspond to covariates $x_1, x_2, \ldots, x_n$.

a.) Show that the maximum likelihood estimator of β_1 is

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

b.) Show that the maximum likelihood estimator of β_0 is

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

c.) Show that the maximum likelihood estimator of β_0 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

d.) In `R` build a function that will simulate a sample that follows where the linear model where $\beta_0 = 200$ and $\beta_1 = -120$ and $^2 = 10$ for $x_i = 0, 1, 2, \ldots, 100$. Use this function to generate a single sample. Report the value of $\bar{y}$, $\bar{x}$ and the maximum likelihood estimators of the parameters β_0, β_1 and σ^2.

e.) In part d.), you generated a single sample from a linear model and estimated the MLEs of each paramter. Now generate 10,000 such samples and calculate the MLEs for the paramters from each of the samples. Plot a histogram of the samples.

f.) In part e.), you generated 10,000 simulated samples and caluclated 10,000 values for each MLE estimator. What is the mean of the MLE you generated from the samples? What two values capture the middle 95% of the sampled values?

Part II: Bayesian Approach

Consider the second model (the Bayesian one) with responses $y_1, y_2, \ldots, y_n$ which correspond to covariates $x_1, x_2, \ldots, x_n$.

a.) Show that the posterior distribution of σ^2 is $\text{Gamma}(a + n/2,\, b + 0.5\sum_{i=1}^{n} y_i^2)$.

b.) Show that the joint posterior probability distribution function for $\beta_0, \beta_1 | \sigma^2$ is bivariate normal with variances n, $\sum_{i=1}^{n} x_i^2$ and $\phi = \frac{\sum_{i=1}^{n} x_i}{\sqrt{n \sum_{i=1}^{n} x_i^2}}$

d.) In `R` build a function that will simulate a sample that follows where the linear model where $\beta_0 = 200$ and $\beta_1 = -120$ and $^2 = 10$ for $x_i = 0, 1, 2, \ldots, 100$. With this sample, generate a 1,000 values from the posterior distribution for β_0, β_1, and σ^2. Find the average of the samples for each of these posterior samples.

e.) Repeat part d.) 1000 times to generate 1000 averages from the posterior. How do these samples compare to the samples of the maximum likelohood estimates from Part I.