# Problem Set 5: Probability and Uncertainty

## Part 1: Simulation

Create a simulated data set with a dependent variable that is a linear function of a treatment variable and a confounding variable. Fit a linear model for the true data generating process and print the summary table. Then, do the following:

a. Using the true model, demonstrate that the coefficient for your treatment variable follows the central limit theorem. That is, demonstrate that the coefficient's sampling distribution is approximately normal.
b. Compute the bootstrapped standard error for the coefficient of the treatment variable.
c. Fit a model that omits the confounding variable. Repeat part (a) for this new model and plot the sampling distribution of the treatment variable's coefficient. How do your results differ? What does this imply about statistical tests based on a coefficient's sampling distribution?

## Part 2: Data Analysis

For this part of the assignment, use any data set you like. It can be from the course materials, simulated, or drawn from another source.

a. Conduct a hypothesis test for a difference in means. You decide what the hypotheses are, whether you use a t-test or a z-test, and what the level of significance is. Explain your decisions, and interpret your results both substantively and statistically.
b. Using the same data, fit a linear model. Interpret the coefficient, standard error, t-value, and p-value.