GX5004: HW 2

To the extent possible, please work individually. R and Python have extensive libraries online that can guide you on this assignment. Please feel free to work collectively on the challenging questions at the end of the assignment.

1. In words, the "standard normal distribution" is a normal distribution with mean zero and variance one, often denoted N(0,1). For a random variable X that is distributed as a standard normal, mathematically we write $X \sim N(0,1)$.
   a. Using R or Python, write code to draw at random 100 observations from a N(0,1).
   b. Submit your code, together with summary statistics about your draw, such as the mean, variance and standard deviation.
2. Using R or Python, draw at random 10 observations from a N(0,1).
   a. Following the steps of hypothesis testing, setting α = 0.05, test whether the sample mean is different than zero using the sample standard deviation for your test.
   b. Repeat this exercise using 100 observations drawn at random from a N(0,1).
   c. Repeat this exercise with 1,000 observations drawn at random from a N(0,1).
   d. What conclusions, if any, do you draw from increasing the sample size?
   e. Submit your code and results.
3. We discussed at some length the bivariate linear regression model, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Go to http://www.random.org/integers/ and generate two series of 10 random integers with values between 0 and 9. Call one series x and the other y. Using Python or R, calculate the best fit line using only sample averages, sample variances, and sample covariance. Submit your series, your code, and your calculations.
4. Go to Yahoo! Finance and download a daily price series for a particular publicly traded stock of your choice for a ten-year time period (don't use Apple), as well as the daily price series on the exchange on which it trades. Go to http://www.federalreserve.gov/releases/h15/data.htm and grab a rate on short-term U.S. Treasurys. Generate the log returns for each of these series, and develop risk-free returns for the publicly traded stock as well as the market on which it trades. Using R or Python, examine the risk-free log returns of the publicly traded stock. Examine the average as well as the variance. Using a histogram with ample bins, plot these returns. Generate a scatterplot similar to the scatterplot shown in class. Finally, fit a linear model to obtain estimates of what finance folks call the "alpha" and the "beta". Is "alpha" significantly different than zero at a 95% level? Does a 95% confidence level for "beta" include one? Submit all code and results. (NB: It is custom to spell the plural of Treasury as Treasurys.)
5. The phrase "data generating process" (or DGP) is often used to describe hypothetically the process by which observations arise in the real world. In this problem, we will work with a specific DGP and evaluate features of $\widehat{\beta_1}$.
   a. Suppose your DGP is $y_i = 1 + 2x_i + \epsilon_i$, where $x \sim N(0,1)$ and $\epsilon \sim N(0,1)$.
   b. Using R or Python, write code to generate 1,000 draws for x and ϵ. Use these to generate y in accordance with the DGP.

c. Using R or Python, write code to estimate the bivariate model, $y_i = \beta_0 + \beta_1 x_i$ and summarize the findings. Submit code and results.

d. Repeat b. and c. above five times (for a new set of random draws for each replication). Submit code and results.

e. Challenging question: Suppose you repeated b. and c. above 1,000 times, each time recording the estimated value of β_1. What do you think a histogram of these 1,000 replications of the estimate value of β_1 would show? If you'd like, give this a shot using R or Python.

f. Very challenging question: Suppose that you were not interested in the estimate of β_1 per se, but instead in some functional transformation, such as the estimate of $\exp(\beta_1)$. What might you do with your 1,000 replications from e. above to inform you about the distribution of the estimate of $\exp(\beta_1)$? If you'd like, give this a shot using R or Python.