Problem Set 2
MFE 402: Econometrics
Professor Rossi

This problem set is designed to review material on the sampling distribution of least squares.

Question 1

a. Use the formulas for the least squares estimators to express the least squares intercept as a weighted sum (i.e., a linear combination) of the Y values in a similar way as is done in the lecture notes for the slope (see Ch1, pg 70–72).
b. Use the formula in part (a) to show that b_0 is an unbiased estimator for β_0. That is, show $\mathbb{E}[b_0] = \beta_0$. You may not use the fact $\mathbb{E}[b_1] = \beta_1$ unless you explicitly prove it.
c. Use the formula in part (a) to show that $\text{Var}(b_0) = \sigma^2 \left[\frac{1}{N} + \frac{\bar{X}^2}{(N-1)s_X^2}\right]$.

Note that parts (b) and (c) are somewhat challenging.

Question 2

a. Write a function in `R` (using `function()`) to simulate from a simple regression model. This function should accept as inputs: β_0 (intercept), β_1 (slope), X (a vector of values), and σ (error standard deviation). You will need to use `rnorm()` to simulate from the normal distribution. The function should return a vector of Y values.
b. Simulate Y values from your function and make a scatterplot of X versus simulated Y. When simulating, use the `vwretd` data from the `marketRf` dataset as the X vector, and choose $\beta_0 = 1$, $\beta_1 = 20$, and $\sigma = 1$. Then add the fitted regression line to the plot as well as the true conditional mean line (the function `abline()` may be helpful).

Question 3

Assume $Y = \beta_0 + \beta_1 X + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Let $\beta_0 = 2$, $\beta_1 = 0.6$, and $\sigma^2 = 2$.

a. Use your `R` function from question 1 to simulate the sampling distribution of the intercept. Use a sample size of 300 and calculate b_0 for 10,000 samples. Plot a histogram of the sampling distribution of b_0. You may find slide 75 of Chapter 1 of the course notes to be helpful.
b. Calculate the empirical value for $\mathbb{E}[b_0]$ from your simulation and provide the theoretical value for $\mathbb{E}[b_0]$ (you might find question 1b to be helpful here). Compare the simulated and theoretical values.
c. Calculate the empirical value for $\text{Var}(b_0)$ from your simulation and provide the theoretical value for $\text{Var}(b_0)$ (you might find question 1c to be helpful here). Compare the simulated and theoretical values.

Question 4

Fit a regression of the Vanguard 500 Index Fund returns (VFIAX in the `Vanguard` dataset from the `DataAnalytics` package) on the `vwretd` series (from the `marketRF` dataset in the `DataAnalytics` package).

a. Test the hypothesis $H_0^a : \beta_1 = 1$ at the 0.05 level of significance using t-statistics. Report your decision (accept or reject the null hypothesis).
b. Test the hypothesis $H_0^b : \beta_0 = 0$ at the 0.01 level of significance using p-values. Report your decision (accept or reject the null hypothesis).

You may **not** use the `summary()` command or a similar command that "automatically" computes t and p values. You must compute the t and p values "by hand". You may, however, use `qt()`, `pt()`, or similar commands.

Question 5

Standard errors and p-values.

a. What is a standard error of a test statistic? How is a standard error different from a standard deviation?
b. What is sampling error? How does the standard error capture sampling error?
c. Your friend Steven is working as an investment analyst and comes to you with some output from some statistical method that you've never heard of. Steven tells you that the output has both parameter estimates and standard errors. He then asks, "how do I interpret and use the standard errors?" What do you say to Steven to help him even though you don't know what model is involved?
d. Your friend Xingua works with Steven. She also needs help with her statistical model. Her output reports a test statistic and the p-value. Xingua has a Null Hypothesis and a significance level in mind, but she asks "how do I interpret and use this output?" What do you say to Xingua to help her even though you don't know what model is involved?

Question 6

Use the fitted regression of `VGHCX` (in the `Vanguard` dataset from the `DataAnalytics` package) on `vwretd` (from the `marketRF` dataset in the `DataAnalytics` package) to answer the following questions. You may not use the `predict()` command. You must perform the calculations "by hand". Note that the data has values like 0.003, which is a positive return of 0.3%.

a. Compute an estimate of the conditional mean of the Vanguard HCX fund's return given that the market is up by 5%.
b. Compute an estimate of the conditional standard deviation of the Vanguard HCX fund's return given that the market is up by 10%.
c. Compute an estimate of the prediction error (s_{pred}) for a prediction of the Vanguard HCX fund's return given that the market is up by 15%.