ECN 480/PUB 580
Assignment #3
Thursday, February 24, 2022 by end of day

Directions: Answer each question electronically in a MS Word or .pdf file. Compile your answers into a single computer file, and then upload it in Canvas under "Assignment #3." Contact me if you have any questions.

Download the dataset entitled GPA2.dta from underneath the .pdf file for this assignment in the "Assignments" section of Canvas. This is a data set containing observations on college GPAs and some other variables for over 4,000 students. We are going to try to explain what determines a college student's GPA.

1. Suppose you think that college GPA depends on a student's high school class size, his/her rank in his graduating class, his/her SAT score, whether the student is a female, and whether the student is an athlete. That is, colgpa = f(hsize, hsrank, sat, female, athlete).

Estimate the following regression in Stata and copy-and-paste your results below using Courier New, font size 8 **(3 points)**:

$$colgpa = \beta_0 + \beta_1 hsize + \beta_2 hsrank + \beta_3 sat + \beta_4 female + \beta_5 athlete + u$$

2. Suppose a student is a female. How does her GPA change as a result? **(2 points)**

3. Suppose a student is an athlete. How does his/her GPA change as a result? **(2 points)**

4. Conduct a hypothesis test for the null hypothesis that being female has no effect on GPA. Should the alternative hypothesis be a one- or two-tailed alternate hypothesis? Note that there is no correct answer here. I just want to see what you think. Report the critical value and whether you reject or fail to reject the null hypothesis. Note that since we have such a large sample size, you can use the standard normal critical value. **(3 points)**

5. Repeat #4 but for athletes. **(3 points)**

6. Conduct a F-test by writing down the null hypothesis that all $\hat{\beta}$, except for $\hat{\beta}_0$, are jointly equal to zero. Refer to Lecture #10 for a refresher on this. Recall that Stata gives you the F-statistic for this test in the upper right-hand corner of the Stata output. Compare this F-statistic to the 5% critical value from the F-table (G.3b, page 788). Do you reject or fail to reject H_0. Explain. **(3 points)**

7. Suppose you think that high school doesn't matter! That is, you think hsize and hsrank jointly have no effect on college GPA. Conduct an F-test where the null hypothesis is $\hat{\beta}_1 = \hat{\beta}_2 = 0$. Recall that Stata doesn't give you this test statistic automatically, but it is easy enough to get with the `test` command. The form for this is:

```
Test variable1 variable2
```

Where `variable1` and `variable2` are the names of the two variables you want to test. Do you reject or fail to reject H_0? Why? **(3 points)**

8. Suppose you think the effect of being female on college GPA depends on whether the female is also an athlete. Generate a new variable, named femathlete, which is an interaction term between female and athlete. Re-estimate your regression from question #1 and include this interaction term. How much does being an athlete increase or decrease a female's GPA. Is it statistically different from zero? **(4 points)**

9. Generate an interaction between the female dummy variable and the SAT dummy variable. Does a higher SAT score improve a student's college GPA more for females then males? Refer to the "different slopes" part of Lecture #11. **(3 points)**

Download the data set entitled "hprice1.dta" from underneath the .pdf file for this assignment in the "Assignments" section of Canvas. This data set consists of 88 observations for the price of a house (in thousands of dollars), along with the square footage of it, the square footage of the lot it is on, and the number of bathrooms. I wish the data set included the number of bathrooms as well, but alas it does not.

10. Suppose you think that the price of a house depends on the square footage of the house, the size of the lot the house is on, and the number of bedrooms. That is,
price = f(sqrft, lotsize, bdrms)

Estimate the following regression and report your results. Which $\hat{\beta}$ are statistically significant? **(3 points)**

11. Suppose you are a realtor trying to determine the price of house. The house is 2,500 square feet on a 10,000 square foot lot size and has 4 bedrooms. Based on your results for question #10, how much would this house sell for? **(3 points)**

12. Suppose you think the effect of the number of bedrooms depends on the size of the house. Create an interaction effect between the square footage of the house (sqrft) and the number of bedrooms (bdrms). Re-estimate the regression from #10 by including this new variable along with the other ones. Is the interaction term statistically significant? How can you tell? **(3 points)**

13. Suppose a house is 2,500 square feet. How much does an additional bedroom increase the price by? Recall that price is in thousands of dollars. How much would the price increase by if the house was 3,500 square feet instead? Refer to the section entitled "interaction between two continuous variables" in Lecture #11. **(3 points)**

14. Compare the adjusted-R^2, which is also called $\bar{R}^2$ between your regression in #12 and your regression #10. Does the $\bar{R}^2$ increase or decrease when you add the interaction term in question #12? What does this tell you about the relevance of the interaction term in question #12? Refer to the section on $\bar{R}^2$ in Lecture #6. **(2 points)**