ECN 480 Assign 3, Winter 2022

Tuesday, February 22, 2022 12:30 AM

ECN 480/PUB 580 Cason Konzer
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)**:

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

```
. reg colgpa hsize hsrank sat female athlete

      Source |       SS           df       MS      Number of obs   =     4,137
-------------+----------------------------------   F(5, 4131)      =    315.77
       Model |  496.115255         5  99.2230511   Prob > F        =    0.0000
    Residual |  1298.08042     4,131  .314229102   R-squared       =    0.2765
-------------+----------------------------------   Adj R-squared   =    0.2756
       Total |  1794.19567     4,136  .433799728   Root MSE        =    .56056

------------------------------------------------------------------------------
      colgpa | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       hsize |   .0655423   .0065315    10.03   0.000     .0527371    .0783476
      hsrank |  -.0037877   .0001811   -20.92   0.000    -.0041427   -.0034327
         sat |   .0017035   .0000673    25.30   0.000     .0015715    .0018355
      female |    .170757   .0181267     9.42   0.000      .135219     .206295
     athlete |   .1536686   .0427887     3.59   0.000     .0697798    .2375575
       _cons |    .830143   .0722384    11.49   0.000     .6885169     .971769
------------------------------------------------------------------------------
```

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

Female Students are Predicted to have a 0.171 GPA point increase compared to non-Females.

3. Suppose a student is an athlete. How does his/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)**

Athletes are Predicted to have a 0.154 GPA point increase compared to non-athletes.

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)**

$\boxed{H_0: \hat{\beta}_4 = 0}$ $\boxed{H_A: \hat{\beta}_4 \neq 0}$ This is a 2-tailed test.

t-statistic: 9.42 95% Critical Value: 1.96

As $|9.42| > 1.96$; we reject the null hypothesis.

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

$\boxed{H_0: \hat{\beta}_5 = 0}$ $\boxed{H_A: \hat{\beta}_5 \neq 0}$ This is a 2-tailed test

t-statistic: 3.59 95% Critical Value: 1.96

As $|3.59| > 1.96$; we reject the null hypothesis.

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)**

$\boxed{H_0: \hat{\beta}_1 = \hat{\beta}_2 = \hat{\beta}_3 = \hat{\beta}_4 = \hat{\beta}_5 = 0}$ $\boxed{H_A: H_0 \text{ is False}}$

$q = 5$

$N - K - 1 \approx \infty$

Critical Value ($F_{5,\infty}$) @ 95 % Confidence: 2.21

F-statistic: 315.77

As $|315.77| > 2.21$; We reject the null hypothesis.

---

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)**

```
. test hsize hsrank

 ( 1)  hsize = 0
 ( 2)  hsrank = 0

       F(  2,  4131) =  227.77
            Prob > F =    0.0000
```

$\boxed{H_0: \hat{\beta}_1 = \hat{\beta}_2 = 0}$ $\boxed{H_A: H_0 \text{ is False}}$

$q = 2$; $N-K-1 \approx \infty$

Critical Value ($F_{2,\infty}$) @ 95% Confidence : 3.0

F-statistic: 227.77

As $|227.77| > 3.0$; We reject H_0.

We are not confident in H_0

---

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)**

```
. gen femathlete = female*athlete

. reg colgpa hsize hsrank sat female athlete femathlete

      Source |       SS           df       MS      Number of obs   =     4,137
-------------+----------------------------------   F(6, 4130)      =    263.13
       Model |  496.184089         6  82.6973482   Prob > F        =    0.0000
    Residual |  1298.01158     4,130  .314288519   R-squared       =    0.2765
-------------+----------------------------------   Adj R-squared   =    0.2755
```

given you are a female, if you are also an athlete, your GPA is

```
       Model |  496.184089         6  82.6973482   Prob > F        =    0.0000
    Residual |  1298.01158     4,130  .314288519   R-squared       =    0.2765
-------------+----------------------------------   Adj R-squared   =    0.2755
       Total |  1794.19567     4,136  .433799728   Root MSE        =    .56061

------------------------------------------------------------------------------
      colgpa | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       hsize |   .0655476   .0065321    10.03   0.000     .0527411    .0783541
      hsrank |  -.0037919   .0001813   -20.91   0.000    -.0041474   -.0034364
         sat |   .0017044   .0000674    25.30   0.000     .0015723    .0018365
      female |    .172304   .0184273     9.35   0.000     .1361766    .2084315
     athlete |   .1649329   .0490974     3.36   0.001     .0686756    .2611902
  femathlete |  -.0455352   .0972993    -0.47   0.640    -.2362942    .1452237
       _cons |   .8287026   .0723107    11.46   0.000     .6869346    .9704705
------------------------------------------------------------------------------
```

athlete, your GPA is Expected to decrease by 0.0455 points.

A p-value of 0.64 shows that this difference is not statistically different from 0.

---

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)**

```
. gen femsat = female*sat

. reg colgpa hsize hsrank sat female athlete femsat

      Source |       SS           df       MS      Number of obs   =     4,137
-------------+----------------------------------   F(6, 4130)      =    263.08
       Model |  496.115333         6  82.6858888   Prob > F        =    0.0000
    Residual |  1298.08034     4,130  .314305167   R-squared       =    0.2765
-------------+----------------------------------   Adj R-squared   =    0.2755
       Total |  1794.19567     4,136  .433799728   Root MSE        =    .56063

------------------------------------------------------------------------------
      colgpa | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       hsize |   .0655459   .0065363    10.03   0.000     .0527314    .0783605
      hsrank |  -.0037878   .0001813   -20.90   0.000    -.0041432   -.0034325
         sat |   .0017027   .0000857    19.86   0.000     .0015346    .0018707
      female |   .1686497   .1351357     1.25   0.212    -.0962891    .4335885
     athlete |   .1536053   .0429826     3.57   0.000     .0693363    .2378743
      femsat |   2.05e-06   .0001305     0.02   0.987    -.0002538    .0002579
       _cons |   .8310206   .0912701     9.11   0.000     .6520821    1.009959
------------------------------------------------------------------------------
```

A p-value of 0.987 says that We do not see Statistical significance & thus being a female does not change the expectation of college gpa with respect to sat score.

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)**

We make the assumption that all of

(3 points)

```
. reg price sqrft lotsize bdrms

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(3, 84)        =     57.46
       Model |  617130.701         3  205710.234   Prob > F        =    0.0000
    Residual |  300723.805        84  3580.0453    R-squared       =    0.6724
-------------+----------------------------------   Adj R-squared   =    0.6607
       Total |  917854.506        87  10550.0518   Root MSE        =    59.833

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |   .1227782   .0132374     9.28   0.000     .0964541    .1491022
     lotsize |   .0020677   .0006421     3.22   0.002     .0007908    .0033446
       bdrms |   13.85252   9.010145     1.54   0.128    -4.065141    31.77018
       _cons |  -21.77031   29.47504    -0.74   0.462    -80.38466    36.84405
------------------------------------------------------------------------------
```

We make the assumption that all of our X's are one-tailed... e.g. increasing square footage of lot, housing, or increasing the # of bedrooms can only increase the price of the house.

∴ all $\hat{\beta}$'s are of statistical significance

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)**

Based on the regression...

$$\widehat{price} = 0.1228(sqrft) + 0.0021(lotsize) + 13.8525(bdrms)$$

$$= 0.1228(2500) + 0.0021(10000) + 13.8525(4)$$

$$= 307 + 21 + 55.1 = 308.41$$

$$\widehat{price} = \$308{,}410$$

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)**

```
. reg price sqrft lotsize bdrms sqrbeds

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(4, 83)        =     46.48
       Model |  634546.045         4  158636.511   Prob > F        =    0.0000
    Residual |  283308.461        83  3413.35495   R-squared       =    0.6913
-------------+----------------------------------   Adj R-squared   =    0.6765
       Total |  917854.506        87  10550.0518   Root MSE        =    58.424

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |   .0337926   .0414616     0.82   0.417    -.0486728    .1162579
     lotsize |   .0019927   .0006279     3.17   0.002     .0007439    .0032416
       bdrms |  -33.71534   22.82291    -1.48   0.143    -79.10919    11.67852
     sqrbeds |   .0218268   .0096631     2.26   0.027     .0026074    .0410462
       _cons |   165.4265   87.73015     1.89   0.063    -9.065246    339.9182
------------------------------------------------------------------------------
```

The interaction term IS statistically significant

We can tell by examination of the p-value

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)**

2500 sq ft: $\dfrac{\Delta \text{price}}{1 \text{ bedroom}} = -33.715 + 0.0218(2500) = -33.715 + 54.5 = 20.785$

\$20,785

3500 sq ft: $\dfrac{\Delta \text{price}}{1 \text{ bedroom}} = -33.715 + 0.0218(3500) = -33.715 + 76.3 = 42.585$

\$42,585

---

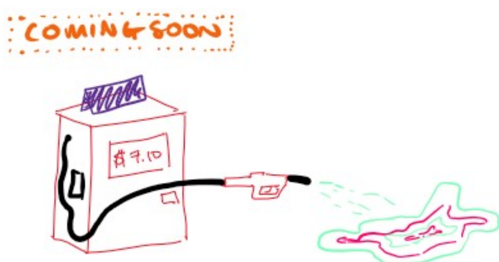
14. Compare the adjusted-R^2, which is also called $\bar{R}^2$ between your regression in #12 and your regression #10. Does the $\bar{R}^2$ increase or decrease when you add the interaction term in question #12? What does this tell you about the relevance of the interaction term in question #12? Refer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

$\bar{R}^2_{\#10} = 0.6607$

$\bar{R}^2_{\#12} = 0.6765$

$\bar{R}^2$ increases when adding in the interaction term

This tells us that the interaction term is relevant & improves our model's predictive power!)



He Who Hedges gas prices by buying oil stocks

Oooooo