Edmond Ho-Yin Lau
Aaron Meininger
Summer Session 1
Econ 113
7/8/2019

Homework 2

1. The five classic Gauss-Markov assumptions for simple linear regression are: (1) linearity - the parameters being estimated using OLS method have to be linear themselves, (2) random - the data must be randomly sampled from the population, (3) non-collinearity - the regressors being calculated aren't perfectly correlated with each other, (4) exogeneity - regressors aren't correlated with errors terms, (5) homoscedasticity - no matter what the values of our regressors might be, error of variance is constant if all 5 assumptions hold, this OLS is the best linear unbiased estimator

2. There are two reasons why the sample mean may deviate from the null hypothesis, (1) is the choice of null where $\mu_0 \neq \mu_y$ and (2) random sampling where $\overline{Y} \neq \mu_0 = \mu_y$. The steps for testing a hypothesis are: (1) making an initial assumption about the population parameter, (2) gather data and calculate the sample mean and std. error, (3) compute "t-stat", (4) choose "significance level α: the probability of rejecting H_0 when it is true, (5) look at the t-table and compare t-stat to the critical level at the chosen significance level, (6) if the absolute value of the t-stat exceeds the t-critical value when we reject that $\overline{x} =$ c, if it does not, then we do not reject that they are equal.

3. Given: n = 100, $\overline{x} = 45221$, standard deviation = 30,450

a. Standard Deviation of mean annual income: $\frac{S}{\sqrt{n}} = \frac{30{,}450}{\sqrt{100}} = \mathbf{3045}$

b. Hypothesis: H_0: M = 40000, H_A: M < 40000

T-test = $\frac{\overline{x}-M}{\frac{S}{\sqrt{n}}} = \frac{45221-40000}{\frac{30450}{\sqrt{100}}} = 1.714614 \approx 1.71$

Critical value at dF = $100 - 1 = 99$, $\alpha = 0.10$ Crit = $t_{0.10,\,99} = -1.29$

Since T-stat > -1.29, we can not reject H_0

Conclusion: Fail to reject H_0, there is not enough sufficient evidence to support the claim and we can conclude that annual income is not less than 40,000

c. Hypothesis: H_0: M = 42000, H_A: M ≠ 42000

T-test = $\frac{\overline{x}-M}{\frac{S}{\sqrt{n}}} = \frac{45221-42000}{\frac{30450}{\sqrt{100}}} = 1.0578 \approx 1.06$

Critical value at dF = 100 - 1 = 99, $\alpha = 0.05$ Crit = $t_{0.05/2,99} = \pm 1.984$

Since T-stat $1.06 < \pm 1.984$ we can not reject H_0

Conclusion: Fail to reject H_0, there is not enough sufficient evidence of different mean than 42000

d. Confidence interval for part (c):

$\alpha = 0.05$, $df = 100 - 1 = 99$

$T_{crit} = t_{0,05/2,99} = 1.984$

95% Confidence interval: $\overline{x} \pm t_{crit} \bullet \frac{S}{\sqrt{n}} = 45221 \pm 1.984 \bullet \frac{30450}{\sqrt{100}} = 45221 \pm 6041.28 =$ **39179.72, 51262.28**

e. P-value at t = 1.71 and dF = 99

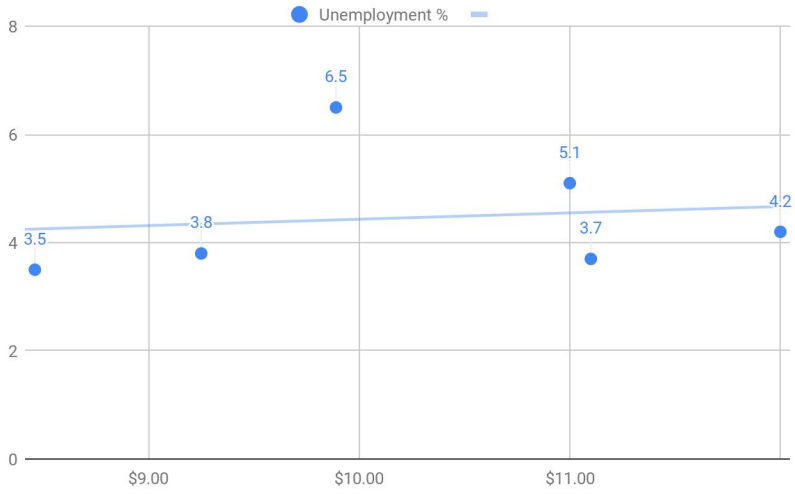
P-value = Tdist (1.71, 99, 1) = 0.0452

f. In part c, we fail to reject H_0

Conclusion: Failed to reject H_0, there is not enough sufficient evidence of different mean than 42,000

4.

State	Unemployment %	Minimum Wage
Alaska	6.5	$9.89
Arizona	5.1	$11.00
Arkansas	3.8	$9.25
California	4.2	$12.00
Colorado	3.7	$11.10
Florida	3.5	$8.46



a. With intercept at (0, 4.1) and a slope of ½

b. $\text{Mean}_{\text{Minimum Wage}}\ \overline{w} = \frac{1}{n}\sum w_i = \frac{9.89+11.00+9.25+12.00+11.10+8.46}{6} = \mathbf{10.28}$

$\text{Variance}_{\text{Minimum Wage}} = \frac{1}{n-1}\sum_{i=1}^{n}(w_i - \overline{w})^2 = \frac{1}{5}[(9.89-10.28)^2 + (11.00-10.28)^2 + (9.25-10.28)^2 + (12.00-10.28)^2 + (11.10-10.28)^2 + (8.46-10.28)^2] = \mathbf{1.74}$

$\text{Mean}_{\text{Unemployment}}\ \overline{u} = \frac{1}{n}\sum u_i = \frac{6.5+5.1+3.8+4.2+3.7+3.5}{6} = \mathbf{4.47}$

$\text{Variance}_{\text{Unemployment}} = \frac{1}{n-1}\sum_{i=1}^{n}(u_i - \overline{u})^2 = \frac{1}{5}[(6.5-4.47)^2 + (5.1-4.47)^2 + (3.8-4.47)^2 + (4.2-4.47)^2 + (3.7-4.47)^2 + (3.5-4.47)^2] = \mathbf{1.32}$

Covariance(Wage, Unemployment) = $\frac{1}{n-1}\sum_{i=1}^{n}(w_i - \overline{w})(u_i - \overline{u})$ =

$\frac{1}{5}$ ((9.89 − 10.28)(6.5 − 4.47) + (11.00 − 10.28)(5.1 − 4.47) + (9.25 − 10.28)(3.8 − 4.47) + (12.00 − 10.28)(4.2 − 4.47) + (11.10 − 10.28)(3.7 − 4.47) + (8.46 − 10.28)(3.5 − 4.47) = **0.21**

c. $\beta_1 = \frac{Cov(X,Y)}{Var(X)} = \frac{0.21}{1.74} =$ **0.12**

$\beta_0 = \overline{Y} - \beta_1 \overline{X} = 4.47 - 0.12 \times 10.28 =$ **3.23**

$\widehat{unemp}_i =$ **3.23+0.11777*minwage***

For one more dollar increase in minimum wage, the predicted unemployment rate increases by 0.11777.

d. The predicted unemployment rate for a state with a minimum wage of 9 dollars is 3.23 + 0.11777(9) = **4.2899%**

e. The predicted change in unemployment rate is: (0.11777)(-3)= **-0.353331**

f. SSR = $\sum_{i=1}^{n}(y_i - \hat{y})^2 = (6.5 - 4.394)^2 + (5.1 - 4.525)^2 + (3.8 - 4.319)^2 + (4.2 - 4.643)^2 +$

$(3.7 - 4.537)^2 + (3.5 - 4.226)^2 = 6.459$

SST = $\sum_{i=1}^{n}(y_i - \overline{y})^2 = (6.5 - 4.4667)^2 + (5.1 - 4.46667)^2 + (3.8 - 4.4667)^2 + (4.2 - 4.4667)^2$

$+ (3.7 - 4.4667)^2 + (3.5 - 4.4667)^2 = 6.57$

R^2 = 1 -(SSR/SST) = 1-(6.459/6.57) = 0.01689497717

g. The assumption required in order to use regression through the origin when regressing unemployment on wage is that if minimum wage is 0, unemployment must also be 0. Regression through the origin only makes sense if we assume that means if individuals are paid nothing then there'd be 0 possibility of there being a job opening for every individual.

h. $\beta_1 = \frac{\sum_{i=1}^{N} x_i y_i}{x_i^2} = \frac{(9.89*6.5)+(11.00*5.1)+(9.25*3.8)+(12.00*4.2)+(11.10*3.7)+(8.46*3.5)}{9.89^2+11.00^2+9.25^2+12^2+11.1^2+8.46^2} = 0.4300899222$

$\widehat{unemp} = 0.4300899222 minwage$

i. Minimum wage is 9

$\widehat{unemp} = 0.4300899222\,(9) = 3.8708$

Predicted unemployment for a state with a minimum wage of 9 is 3.8708%, as compared to our previous answer in part (d) with 4.2899%

5.

a. For 1 increase in number of parks, there could be R.S. 9000 increase in prices of houses

b. For 30 number of parks, price = $600,000 + 9,000 * 30 = 870,000$

For 50 number of parks, price = $600,000 + 9,000 * 50 = 1,050,000$

c. $1,280,300 = 600,000 + 9,000 * n$

$680,300 = 9,000 * n$

Parks (n) = 76

d. 600,000 + 9,000*41 = 969,000

969,000-710,000 = **259,000 Error**

e. Predicted house price 600,000+9,000(1,700)=15,900,000

This equation is only linear, it needs to consider other facts like disposable income, purchasing power, and in terms of price can't keep increasing with the increasing in prices.

f. For every single unit increase in number of parks per 1000HH leads to an increase of houses by 53,100

g. HousePrice = 510,000+51300(30*1000/20,000) = **$589,650**

h. HousePrice = 510,000+51300(1700*1000/8,000,000) = **$521,283.75**

i. For every single unit increase in number of parks per 1000HH leads to an increase of houses by 70%

j. Expected change in price of house = 0.7 * 100(1.5-1)= **35% increase in price of house**

6.

a**. regress colGPA hsGPA**

```
      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(1, 139)       =     28.85
       Model |  3.33506006         1  3.33506006   Prob > F        =    0.0000
    Residual |  16.0710394       139  .115618988   R-squared       =    0.1719
-------------+----------------------------------   Adj R-squared   =    0.1659
       Total |  19.4060994       140  .138614996   Root MSE        =    .34003

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsGPA |   .4824346   .0898258     5.37   0.000      .304833    .6600362
       _cons |   1.415434   .3069376     4.61   0.000     .8085635    2.022304
------------------------------------------------------------------------------
```

colGPA = 1.415+0.482hsGPA

b. The interpretation of my regression is that if person has a high school GPA that is 0, then their expected college GPA will be 1.415. The coefficient on high school GPA that is an increase unit in hsGPA increases colGPA by 0.482. I believe this makes sense, hsGPA can correlate with how you study, learn, etc. towards college.

c. For example, we will use 2.5 as our hsGPA for student B, meaning student A has 4.0 hsGPA

$colGPA_{Studnet\ B}$ = 1.415+0.482(2.5) = 2.62

$colGPA_{Studnet\ A}$ = 1.415+0.482(4.0) = 3.34

Difference = 0.72

d. From part a) the regression result is 0.1719, which means 17.19% variation of college gpa is predicted by high school gpa.

7.

a. **tab lecturesskipped**

```
lecturesski |
       pped |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         44       31.21       31.21
        7.5 |          1        0.71       31.91
         15 |          9        6.38       38.30
         30 |         48       34.04       72.34
```

```
         60 |         25       17.73       90.07
         90 |          9        6.38       96.45
        120 |          3        2.13       98.58
        150 |          2        1.42      100.00
------------+-----------------------------------
      Total |        141      100.00
```

In lecturesskipped the modal number is 30, as it is the most frequent of the sample with 48 out of 141.

b. **regress colGPA lecturesskipped**

```
      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(1, 139)       =     10.23
       Model |  1.33028272         1  1.33028272   Prob > F        =    0.0017
    Residual |  18.0758167       139  .130041847   R-squared       =    0.0685
-------------+----------------------------------   Adj R-squared   =    0.0618
       Total |  19.4060994       140  .138614996   Root MSE        =    .36061

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lecturessk~d |   -.002984    .000933    -3.20   0.002    -.0048287   -.0011394
       _cons |   3.153084   .0427751    73.71   0.000      3.06851    3.237658
------------------------------------------------------------------------------
```

colGPA = 3.153 - 0.003lecturesskipped
The coefficient means for every unit of lecturesskipped (or for every lecture skipped), the expected colGPA decreases by 0.003

c. colGPA = 3.153 - 0.003(20) = 3.09
The expected gpa of someone who skips 20 lectures is 3.09

d. **generate SkipPerWeek = lecturesskipped/30**
regress colGPA SkipPerWeek

```
      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(1, 139)       =     10.23
       Model |  1.33028272         1  1.33028272   Prob > F        =    0.0017
    Residual |  18.0758167       139  .130041847   R-squared       =    0.0685
-------------+----------------------------------   Adj R-squared   =    0.0618
       Total |  19.4060994       140  .138614996   Root MSE        =    .36061

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 SkipPerWeek |  -.0895215   .0279896    -3.20   0.002    -.1448619    -.034181
       _cons |   3.153084   .0427751    73.71   0.000      3.06851    3.237658
------------------------------------------------------------------------------
```

colGPA = 3.153-0.0895*SkipPerWeek

For every unit of SkipPerWeek (how many lectures skipped per week), the expected college gpa can decrease by 0.0895

The relationship between the slopes is that we scaled the variable lecturesskipped by 30 in part b and d.

e. **gen lnColGPA = ln(colGPA)**
regress lnColGPA SkipPerWeek

```
      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(1, 139)       =     10.54
       Model |  .145014145         1  .145014145   Prob > F        =    0.0015
    Residual |  1.91318221       139  .013763901   R-squared       =    0.0705
-------------+----------------------------------   Adj R-squared   =    0.0638
       Total |  2.05819635       140  .014701403   Root MSE        =    .11732

------------------------------------------------------------------------------
    lnColGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 SkipPerWeek |   -.029557    .009106    -3.25   0.001    -.0475611   -.0115529
       _cons |    1.14185   .0139162    82.05   0.000     1.114336    1.169365
------------------------------------------------------------------------------
```

ln(colGPA) = 1.142 - 0.0296*SkipPerWeek

Slope coefficient = -0.0296, if a student skips another lecture a week, college gpa can be expected to drop by 2.96%.