Edmond Ho-Yin Lau
Aaron Meininger
Summer Session 1
Econ 113
7/15/2019

Homework 3

1a. The form $\widehat{wage} = \widehat{\beta}_0 + \widehat{\beta}_1 YrsEdu.$, we want to change wages from dollars to thousands of dollars. For us to change wages from thousands of dollars to dollars, we must divide by 1000, causing $\widehat{\beta}_0$ to become $1.5*10^{-3}$ and $\widehat{\beta}_1$ to become $0.28*10^{-3}$.

b. New YrsEdu = YrsEdu/7

$\widehat{\beta}_{1New} = 7\widehat{\beta}$

$\widehat{\beta}_{0New} = \widehat{wage} - 7\widehat{\beta} * \frac{\overline{YrsEdu}}{7}$

$\widehat{\beta}_{0New} = \widehat{\beta}_0$

Hence $\widehat{\beta}_0 = 15.00$, $\widehat{\beta}_1 = 0.04$

c. Here $YrsEdu_{New} = \frac{YrsEdu}{180}$ and $wage_{new} = wage*10^{-2}$

$\widehat{\beta}_{1New} = \widehat{\beta}_1 * 10^{-2} * 180 = \widehat{\beta}_1 * 1.8$

$\widehat{\beta}_{0New} = 10^{-2}\overline{wage} - \widehat{\beta}_1 * 1.8 * \frac{\overline{YrsEdu}}{180} = 10^{-2}\widehat{\beta}_0$

Hence $\widehat{\beta}_{0New} = 0.15 \; and \; \widehat{\beta}_{1New} = 0.504$

2a. $\widehat{Pass113} = 0.50 + 0.01 * Attendance$

A one unit change in attendance, increases the expectation of passing 113 by 0.01. If a student does not attend 113 at all, the expectation of passing 113 is 0.50.

b. $\widehat{Pass113} = 0.50 + 0.01\,(10) = 0.60$

c. The number of classes done by student A be x and number of classes done by B be x+10

(0.5+0.01x)-(0.5+0.01(x+10))

(0.5+0.01x)-(0.5+0.01x+0.1)

0.5-0.01x-0.5-0.01-0.1

The difference between the 2 would be 0.1.

d. $\widehat{pass113} = 0.5 + 0.01 * attendance$

$0.5 + 0.01 * (attendanceInWeeks) * 7$

$0.5 + 0.07 * attendanceInWeeks$

3a. A one unit change in experience has 9% effect on expected wage. If there was no change in experience, there would be a 0% effect on expected wage, and our expected base wage would be 3.31.

b. $ln\,\widehat{wage} = 3.31 + 0.0075exp$

One unit change in experience per month has 0.0075% effect on expected wage, if there was no change in experience, there would be 0% effect on expected wage, and our expected bsae wage would be 3.31.

c. If we increase experience by 1% we expect wage to increase by 0.2 units. If there was no experience, our expected base wage would be 10.20.
d. If we change experience by 1%, we expect wage to change by 1%. If there was no experience, our expected base wage would be 2.67.

4a. Spurious correlation is when 2 data information are not exactly correlated logically, but statistically correlated, for example if weight gain was correlated with wage increase.
b. A spurious correlation is a relationship between two variables that appear to have interdependence or association with each other but actually do not. (Source: https://www.investopedia.com/terms/s/spurious_correlation.asp)
c. Reverse Causality is believing that A causes B to change when it is really B that causes A to change, i.e. people going to the gym causes them to eat unhealthy, when really eating unhealthy causes people going to the gym.
d. Reverse causality means that X and Y are associated, but not in the way you would expect. Instead of X causing a change in Y, it is really the other way around: Y is causing changes in X. (Source:https://www.statisticshowto.datasciencecentral.com/reverse-causality/)
e. R^2 is the proportion of variation in an outcome that can be explained by a dependent variable, i.e. variation of wage that can be explained by experience.
f. R-Squared(R^2) – measures how much better the regression line predicts the true values of y than just the mean, i.e. what proportion of the variation in Y can be explained by X. $0<R^2<1$
g. The six comments of R^2 are (1) it can be a low percentage of variation even though the model is correct, (2) it can be a high percentage of variation even if the model is incorrect, (3) does not predict errors, (4) R^2 can not be compared for each data set given, (5) it is more useful to compare models with similar variation, (6) useless trying to figure out which causes which.

5a. The intercept in linear regression is value of regression when predictor value is 0. If we run regression through origin then we are excluding the intercept term. Estimate lesser number of parameters which is an advantage.

b. Case 1: regressing weight on height for people, intercept term is meaningless, height is 0 then weight is also 0.
Case 2: regressing price of flats on area, area is 0 then flat price is 0.

c. A safe approach to determine whether or not to include an intercept term is to estimate with an intercept then test if $\widehat{\beta}_0$ is 0 through a hypothesis test of H_0: $\widehat{\beta}_0 = 0$.

6a. **reg wage south**

```
      Source |       SS           df       MS      Number of obs   =       519
-------------+----------------------------------   F(1, 517)       =      5.38
       Model |  72.3235133         1  72.3235133   Prob > F        =    0.0208
```

```
    Residual |  6951.29038       517  13.4454359   R-squared       =    0.0103
-------------+----------------------------------   Adj R-squared   =    0.0084
       Total |  7023.61389       518  13.5591002   Root MSE        =    3.6668

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       south |  -.7775594   .3352596    -2.32   0.021    -1.436198   -.1189207
       _cons |   6.164458   .2012419    30.63   0.000     5.769105     6.55981
------------------------------------------------------------------------------
```

Each additional increase in wage, wage of being in the South decreases by 0.78. It is statistically significant at the 95% confidence level because our p-value = 0.021 which is less than 0.05.

b. **reg wage south educ**

```
      Source |       SS           df       MS      Number of obs   =       519
-------------+----------------------------------   F(2, 516)       =     50.82
       Model |  1155.84794         2  577.923971   Prob > F        =    0.0000
    Residual |  5867.76595       516  11.3716394   R-squared       =    0.1646
-------------+----------------------------------   Adj R-squared   =    0.1613
       Total |  7023.61389       518  13.5591002   Root MSE        =    3.3722

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       south |  -.4339759   .3103253    -1.40   0.163    -1.043632    .1756805
        educ |    .524723   .0537554     9.76   0.000     .4191166    .6303295
       _cons |  -.5431464   .7116492    -0.76   0.446    -1.941232    .8549397
------------------------------------------------------------------------------
```

Each additional increase in wage, wage of being in South decreases by 0.43 which is increased compared to A. Although it is not statistically significant at 95% confidence level because p > 0.05, this does suggest that the education in the South may not be good.

c. gen exper2 = exper^2
gen exper3 = exper^3
reg wage south educ exper exper2 exper3

```
      Source |       SS           df       MS      Number of obs   =       519
-------------+----------------------------------   F(5, 513)       =     38.96
       Model |  1932.99045         5   386.59809   Prob > F        =    0.0000
    Residual |  5090.62344       513  9.92324258   R-squared       =    0.2752
-------------+----------------------------------   Adj R-squared   =    0.2681
       Total |  7023.61389       518  13.5591002   Root MSE        =    3.1501

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       south |  -.6659115   .2923572    -2.28   0.023    -1.240276   -.0915468
        educ |   .5762782   .0534839    10.77   0.000     .4712038    .6813526
       exper |   .3405903   .0861538     3.95   0.000     .1713327    .5098479
      exper2 |   -.008365   .0044502    -1.88   0.061    -.0171078    .0003778
      exper3 |   .0000517   .0000632     0.82   0.414    -.0000725    .0001758
       _cons |  -3.782425   .8175734    -4.63   0.000    -5.388629   -2.176221
------------------------------------------------------------------------------
```

d. Additional variables that could be added would be regression include comparing wages in South to North, the gender in South, and years in college.