By Cliff Rodriguez

Intro to Econometrics - Problem Set 1

Due: January 31, 2018

Cliff Rodriguez

Intro to Econometrics - Problem Set 1
Cliff Rodriguez

Contents

Question 1

Question 1 uses dataset ps1q1.dta

(i) U_i is a constant that represents a value for the impact of non-internalized variables that are relevant to the system – it is an error term or random disturbance and is thought of as the unobserved factors that impact Y.

(ii) Three examples of things that could be included in U_i for this model are

a. Height of father

b. Height of Mother

c. Were prenatal vitamins used by the mother

(iii) Give and intuitive argument for why:

a. α_i might be positive in this model if an additional cigarette increases birthweight.

b. α_i might be negative in this model if an additional cigarette decreases birthweight.

(iv) Simple regression does not solve the problem of omitted variable bias and reverse causality because in the least there are omitted variables. Examples of omitted variables are listed in ii, and include the height of each parent and whether or not prenatal vitamins were taken during pregnancy. Reverse causality is not applicable to this problem because the unborn baby can't cause the parent to smoke more cigarettes.

(v) Using the output from STATA below:

```
. reg bwght cigs

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(1, 1386)      =     32.24
       Model |  13060.4194         1  13060.4194   Prob > F        =    0.0000
    Residual |    561551.3     1,386  405.159668   R-squared       =    0.0227
-------------+----------------------------------   Adj R-squared   =    0.0220
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.129

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.5137721   .0904909    -5.68   0.000    -.6912861   -.3362581
       _cons |   119.7719   .5723407   209.27   0.000     118.6492    120.8946
------------------------------------------------------------------------------

.
```

Estimated Equation: $bwght_i = 119.77 + -.5137721\ cigs_i$

$\alpha_0 = 119.77$

Interpretation: If no cigarettes are smoked per day during pregnancy the estimated birthweight is 119.77 ounces.

$\alpha_1 = -.5137721$

Interpretation: For each cigarette smoked per day is expected to decreases birthweight by .51 ounces.

(vi) The predicted birthweight when $cigs_i = 0$ *is 119.77 ounces.*
The predicted birthweight when $cigs_i = 20$ *is 109.5 ounces.*

Comparing the values of *119.77* ounces, the expected birthweight of a baby when the mother smoked zero cigarettes per day, and 109.5 ounces the expected birthweight for a baby when mother smokes 20 cigarettes per day indicates that smoking a pack a day will decrease the expected birthweight of the baby by roughly 10 ounces.

(vii) An expected birthweight of 125 ounces is not possible with this model

(viii)

a. *The strong assumption made in this model includes … because it is linear.*
b. *This model could be graphed using the equation below, is using a log scale for the data was desirable.*

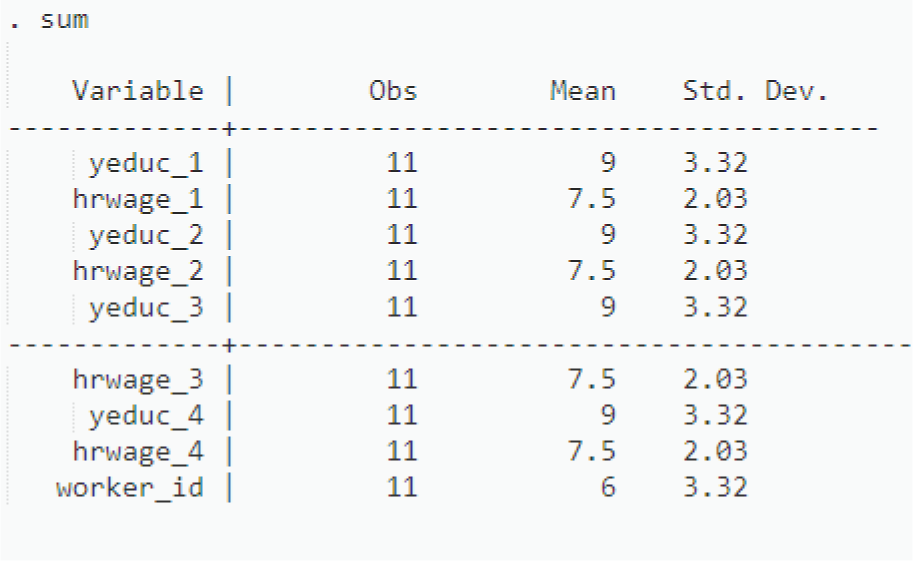
$$bwght_i = 119.77 - \ln(.5137721\ cigs_i)$$

c. *Working log data would be useful in this model if bwght (yi) has a constant level of change for a percentage change in cigs(xi).*

Question 2

(i)

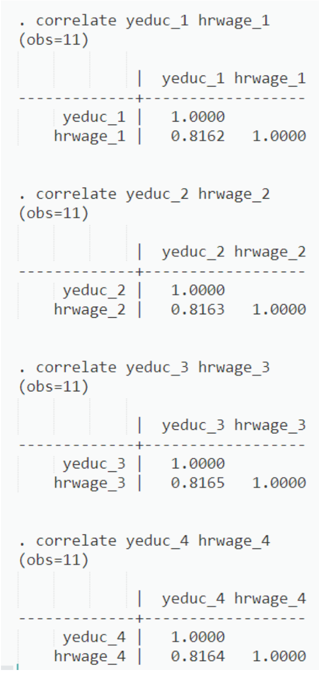
a. The mean and standard deviation for each variable in dataset ps1q2.dta are presented below.



```
. sum

    Variable |        Obs        Mean    Std. Dev.
-------------+---------------------------------------
     yeduc_1 |         11           9     3.32
    hrwage_1 |         11         7.5     2.03
     yeduc_2 |         11           9     3.32
    hrwage_2 |         11         7.5     2.03
     yeduc_3 |         11           9     3.32
-------------+-----------------------------------------
    hrwage_3 |         11         7.5     2.03
     yeduc_4 |         11           9     3.32
    hrwage_4 |         11         7.5     2.03
   worker_id |         11           6     3.32
```

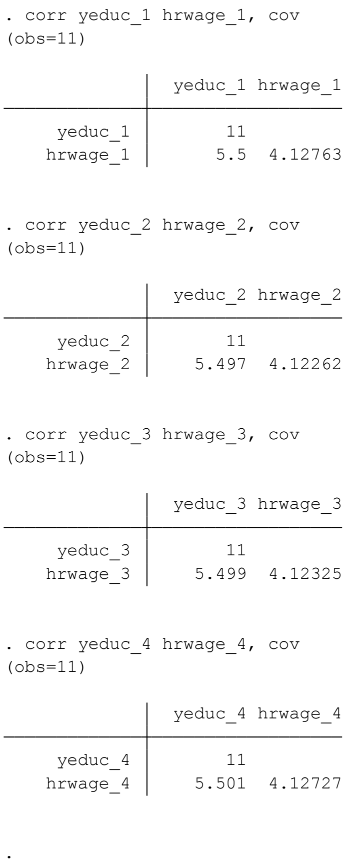
Figure 1: Mean and standard deviation for each variable in dataset ps1q2.dta

b. The covariance and correlation between years of education and hourly wages in dataset ps1q2.dta are presented below.



```
. correlate yeduc_1 hrwage_1
(obs=11)

             |  yeduc_1 hrwage_1
-------------+------------------
     yeduc_1 |   1.0000
    hrwage_1 |   0.8162   1.0000


. correlate yeduc_2 hrwage_2
(obs=11)

             |  yeduc_2 hrwage_2
-------------+------------------
     yeduc_2 |   1.0000
    hrwage_2 |   0.8163   1.0000


. correlate yeduc_3 hrwage_3
(obs=11)

             |  yeduc_3 hrwage_3
-------------+------------------
     yeduc_3 |   1.0000
    hrwage_3 |   0.8165   1.0000


. correlate yeduc_4 hrwage_4
(obs=11)

             |  yeduc_4 hrwage_4
-------------+------------------
     yeduc_4 |   1.0000
    hrwage_4 |   0.8164   1.0000
```

Figure 2: Correlation between years of education and hourly wages in dataset ps1q2.dta



```
. corr yeduc_1 hrwage_1, cov
(obs=11)

             |  yeduc_1 hrwage_1
-------------+------------------
     yeduc_1 |       11
    hrwage_1 |      5.5  4.12763


. corr yeduc_2 hrwage_2, cov
(obs=11)

             |  yeduc_2 hrwage_2
-------------+------------------
     yeduc_2 |       11
    hrwage_2 |    5.497  4.12262


. corr yeduc_3 hrwage_3, cov
(obs=11)

             |  yeduc_3 hrwage_3
-------------+------------------
     yeduc_3 |       11
    hrwage_3 |    5.499  4.12325


. corr yeduc_4 hrwage_4, cov
(obs=11)

             |  yeduc_4 hrwage_4
-------------+------------------
     yeduc_4 |       11
    hrwage_4 |    5.501  4.12727


.
```

Figure 3: Covariance between years of education and hourly wages in dataset ps1q2.dta

(ii) For each firm the OLS regress for hourly wages on years of education is below.

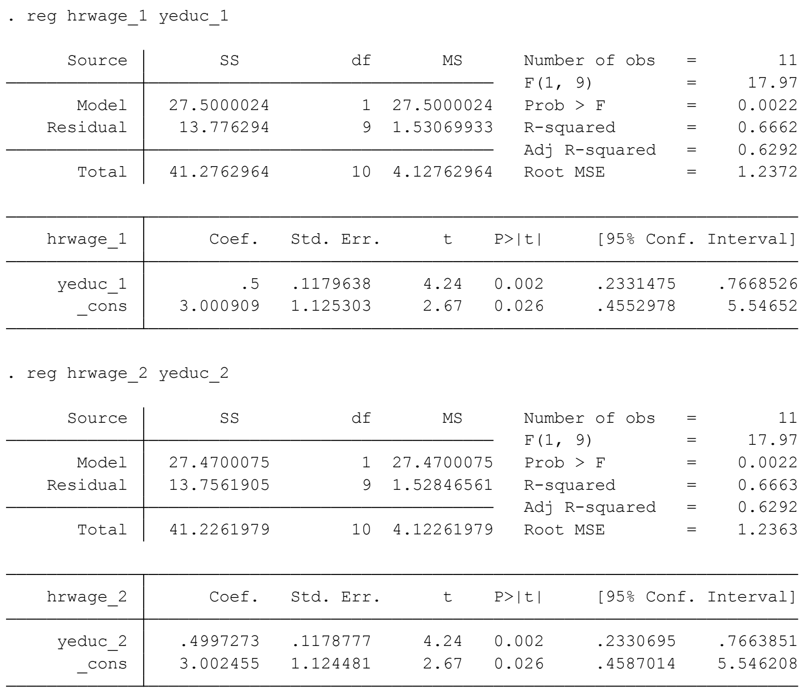
Interpretation: For each firm, the predicted increase on hourly wages for each additional year of education is 50 cents.

Firm 1 α_1 = .5

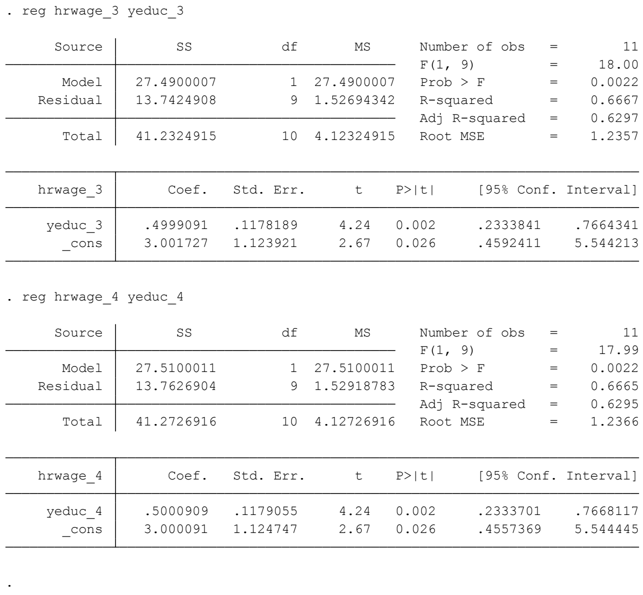
Firm 2 α_1 = .5

Firm 3 α_1 = .5

Firm 4 α_1 = .5



```
. reg hrwage_1 yeduc_1

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =     17.97
       Model |  27.5000024         1  27.5000024   Prob > F        =    0.0022
    Residual |   13.776294         9  1.53069933   R-squared       =    0.6662
-------------+----------------------------------   Adj R-squared   =    0.6292
       Total |  41.2762964        10  4.12762964   Root MSE        =    1.2372

------------------------------------------------------------------------------
    hrwage_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yeduc_1 |         .5   .1179638     4.24   0.002     .2331475    .7668526
       _cons |   3.000909   1.125303     2.67   0.026     .4552978     5.54652
------------------------------------------------------------------------------

. reg hrwage_2 yeduc_2

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =     17.97
       Model |  27.4700075         1  27.4700075   Prob > F        =    0.0022
    Residual |  13.7561905         9  1.52846561   R-squared       =    0.6663
-------------+----------------------------------   Adj R-squared   =    0.6292
       Total |  41.2261979        10  4.12261979   Root MSE        =    1.2363

------------------------------------------------------------------------------
    hrwage_2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yeduc_2 |   .4997273   .1178777     4.24   0.002     .2330695    .7663851
       _cons |   3.002455   1.124481     2.67   0.026     .4587014    5.546208
------------------------------------------------------------------------------
```



```
. reg hrwage_3 yeduc_3

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =     18.00
       Model |  27.4900007         1  27.4900007   Prob > F        =    0.0022
    Residual |  13.7424908         9  1.52694342   R-squared       =    0.6667
-------------+----------------------------------   Adj R-squared   =    0.6297
       Total |  41.2324915        10  4.12324915   Root MSE        =    1.2357

------------------------------------------------------------------------------
    hrwage_3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yeduc_3 |   .4999091   .1178189     4.24   0.002     .2333841    .7664341
       _cons |   3.001727   1.123921     2.67   0.026     .4592411    5.544213
------------------------------------------------------------------------------

. reg hrwage_4 yeduc_4

      Source |       SS           df       MS      Number of obs   =        11
-------------+----------------------------------   F(1, 9)         =     17.99
       Model |  27.5100011         1  27.5100011   Prob > F        =    0.0022
    Residual |  13.7626904         9  1.52918783   R-squared       =    0.6665
-------------+----------------------------------   Adj R-squared   =    0.6295
       Total |  41.2726916        10  4.12726916   Root MSE        =    1.2366

------------------------------------------------------------------------------
    hrwage_4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yeduc_4 |   .5000909   .1179055     4.24   0.002     .2333701    .7668117
       _cons |   3.000091   1.124747     2.67   0.026     .4557369    5.544445
------------------------------------------------------------------------------

.
```

(iii) Based on calculations made using the OLS method, for each firm the added value from one year of education is listed below for each firm:

i. Firm 1: 50 cents per hour

ii. Firm 2: 50 cents per hour

iii. Firm 3: 50 cents per hour

iv. Firm 4: 50 cents per hour

(iv) The relationship between hourly wages and years of education for each firm is graphed below.

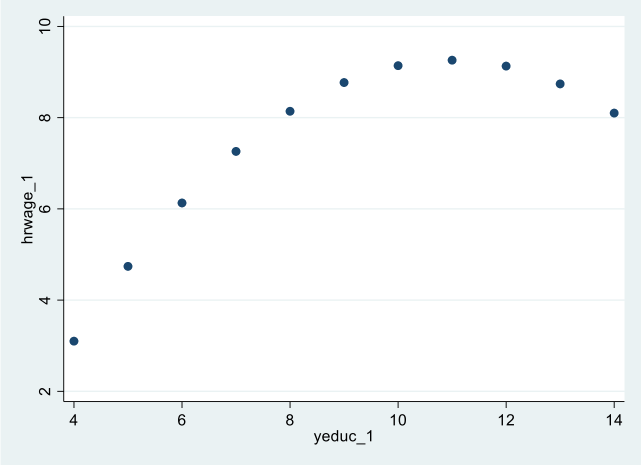


Figure 4: Firm 1

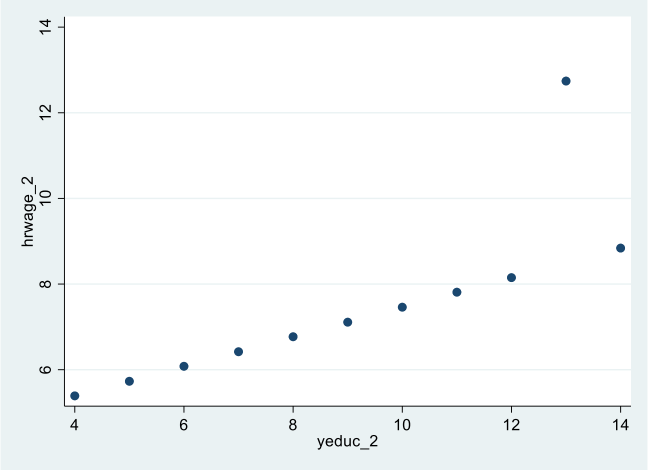


Figure 5: Firm 2

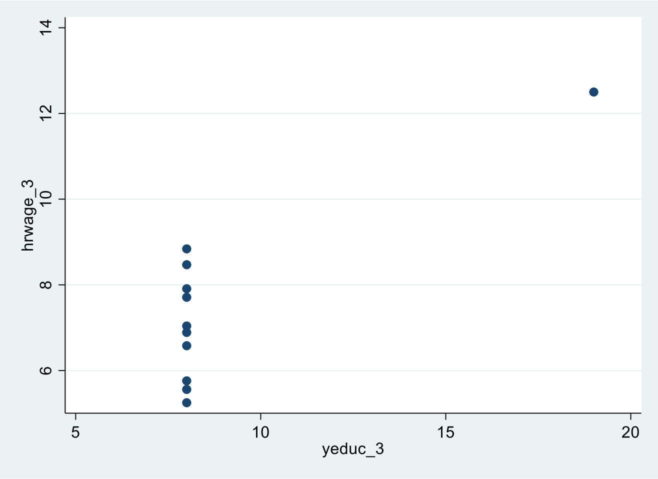


Figure 6: Firm 3

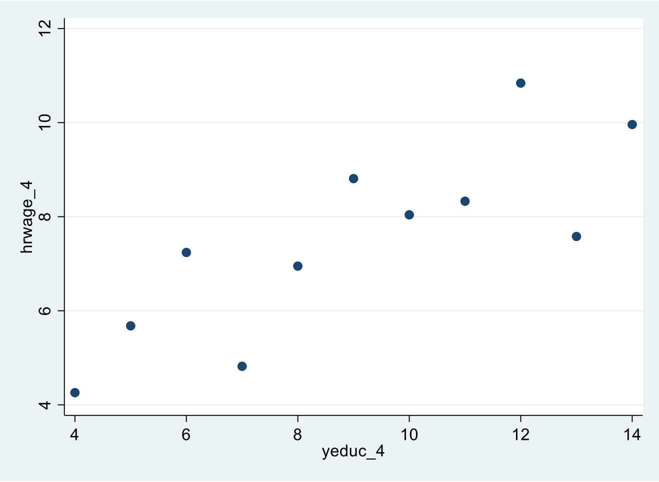


Figure 7: Firm 4

(v) Reviewing the graphs the prediction is equally good for each firm. This is because the $\alpha_1$value for each is .5, meaning income increases by .50 cents per hour per each additional year of education.

(vi) Reviewing the graphs the relationship between hourly wages and years of education is/is not the same. This is evident because the graphs all have a unique pattern.

(vii) Based on points *v* and *vi* above it is suggested that this model is not very useful in predicting wages based on years of education.

Question 3

(i) See appendix I for work

$\hat{\beta}_0$ =

$\hat{\beta}_1$ =

(ii) The $\hat{\beta}_0$ term in this model is/is not useful because any student enrolled has a GPA

(iii) The GPA score is predicted to be points higher if the ACT score increases by 5 points?

(iv) The fitted values and residuals for each observation are presented in appendix I.

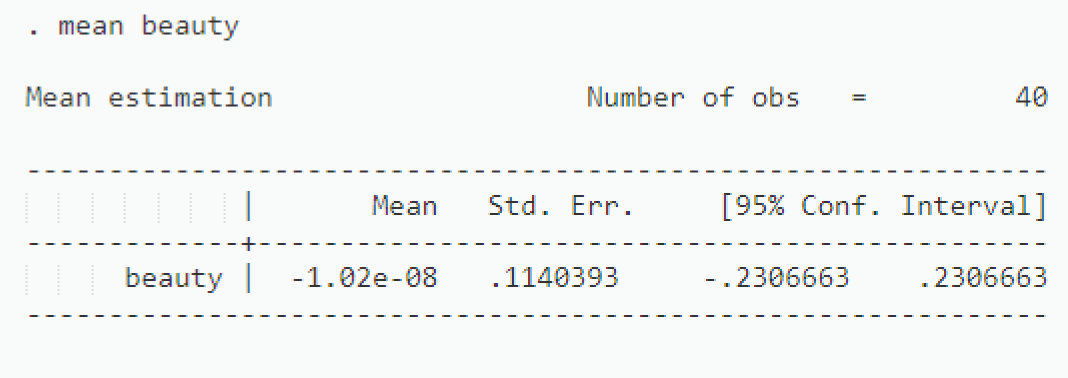
$$\text{GPA} - \overline{\text{GPA}} = .02$$

$$\text{ACT} - \overline{\text{ACT}} = 0$$

(v) Blank of the variation in GPA for the eight students is explained by the ACT. This is because…

Question 4

(i) $\overline{beauty}$ *= -1.02 e-08, or near zero*

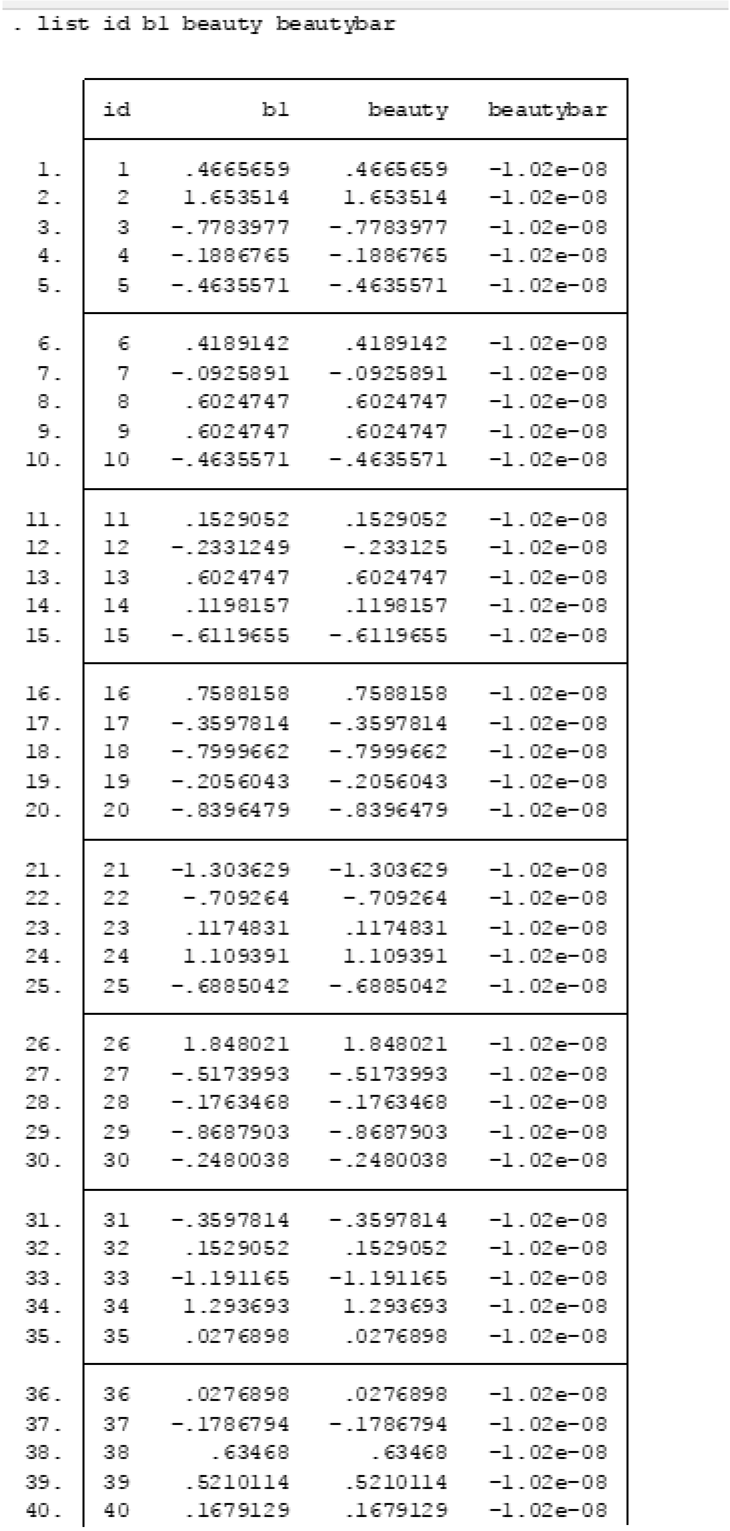


```
. mean beauty

Mean estimation                         Number of obs    =         40

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
      beauty |  -1.02e-08   .1140393     -.2306663    .2306663
--------------------------------------------------------------
```

(ii) $beauty_1 - \overline{beauty}$ =

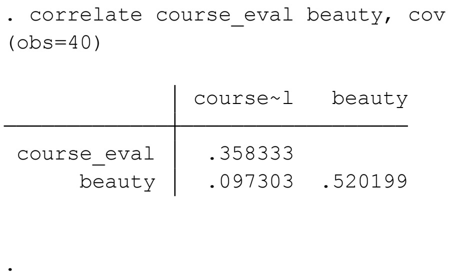
STATA CODE:

$$beautybar = 1.02E8$$
$$generate\ b1 \ = \ beauty - \ beautybar$$



```
. list id bl beauty beautybar

     +------------------------------------------+
     | id          bl      beauty     beautybar |
     |------------------------------------------|
  1. |  1    .4665659    .4665659     -1.02e-08 |
  2. |  2    1.653514    1.653514     -1.02e-08 |
  3. |  3   -.7783977   -.7783977     -1.02e-08 |
  4. |  4   -.1886765   -.1886765     -1.02e-08 |
  5. |  5   -.4635571   -.4635571     -1.02e-08 |
     |------------------------------------------|
  6. |  6    .4189142    .4189142     -1.02e-08 |
  7. |  7   -.0925891   -.0925891     -1.02e-08 |
  8. |  8    .6024747    .6024747     -1.02e-08 |
  9. |  9    .6024747    .6024747     -1.02e-08 |
 10. | 10   -.4635571   -.4635571     -1.02e-08 |
     |------------------------------------------|
 11. | 11    .1529052    .1529052     -1.02e-08 |
 12. | 12   -.2331249    -.233125     -1.02e-08 |
 13. | 13    .6024747    .6024747     -1.02e-08 |
 14. | 14    .1198157    .1198157     -1.02e-08 |
 15. | 15   -.6119655   -.6119655     -1.02e-08 |
     |------------------------------------------|
 16. | 16    .7588158    .7588158     -1.02e-08 |
 17. | 17   -.3597814   -.3597814     -1.02e-08 |
 18. | 18   -.7999662   -.7999662     -1.02e-08 |
 19. | 19   -.2056043   -.2056043     -1.02e-08 |
 20. | 20   -.8396479   -.8396479     -1.02e-08 |
     |------------------------------------------|
 21. | 21   -1.303629   -1.303629     -1.02e-08 |
 22. | 22    -.709264    -.709264     -1.02e-08 |
 23. | 23    .1174831    .1174831     -1.02e-08 |
 24. | 24    1.109391    1.109391     -1.02e-08 |
 25. | 25   -.6885042   -.6885042     -1.02e-08 |
     |------------------------------------------|
 26. | 26    1.848021    1.848021     -1.02e-08 |
 27. | 27   -.5173993   -.5173993     -1.02e-08 |
 28. | 28   -.1763468   -.1763468     -1.02e-08 |
 29. | 29   -.8687903   -.8687903     -1.02e-08 |
 30. | 30   -.2480038   -.2480038     -1.02e-08 |
     |------------------------------------------|
 31. | 31   -.3597814   -.3597814     -1.02e-08 |
 32. | 32    .1529052    .1529052     -1.02e-08 |
 33. | 33   -1.191165   -1.191165     -1.02e-08 |
 34. | 34    1.293693    1.293693     -1.02e-08 |
 35. | 35    .0276898    .0276898     -1.02e-08 |
     |------------------------------------------|
 36. | 36    .0276898    .0276898     -1.02e-08 |
 37. | 37   -.1786794   -.1786794     -1.02e-08 |
 38. | 38      .63468      .63468     -1.02e-08 |
 39. | 39    .5210114    .5210114     -1.02e-08 |
 40. | 40    .1679129    .1679129     -1.02e-08 |
```

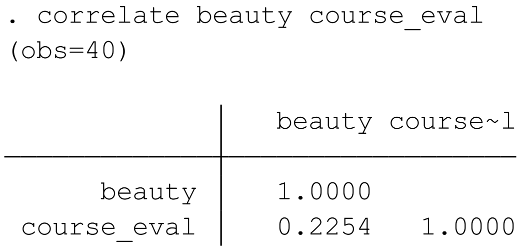
(iii) The covariance between course evaluation and beauty is .097 rating of instructor/units



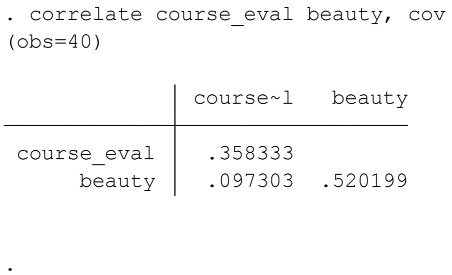
```
. correlate course_eval beauty, cov
(obs=40)

             |  course~l   beauty
-------------+-------------------
 course_eval |   .358333
      beauty |   .097303  .520199


.
```

The units of measure for the covariance between course evaluation and beauty is rating of instructor over units and this does not have a real world interpretation.

(iv) The correlation between course evaluations and beauty is shown below using STATA and the $P_x = \frac{cov(X,Y)}{sd(x)sd(y)}$



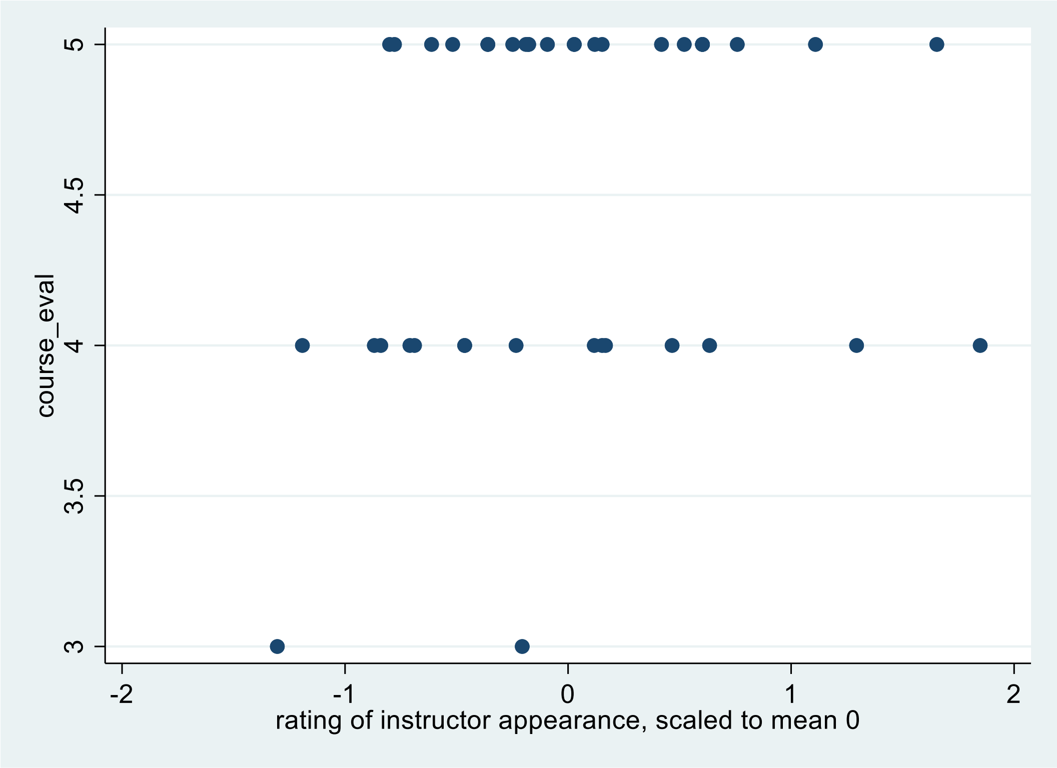
```
. correlate beauty course_eval
(obs=40)

             |   beauty course~l
-------------+------------------
      beauty |   1.0000
 course_eval |   0.2254   1.0000
```



```
. correlate course_eval beauty, cov
(obs=40)

             |  course~l   beauty
-------------+-------------------
 course_eval |   .358333
      beauty |   .097303  .520199


.
```

$$p_x = \frac{cov(X,Y)}{sd(x)sd(y} = \frac{.097303}{(.721248) * (.5986095)} = .225402$$

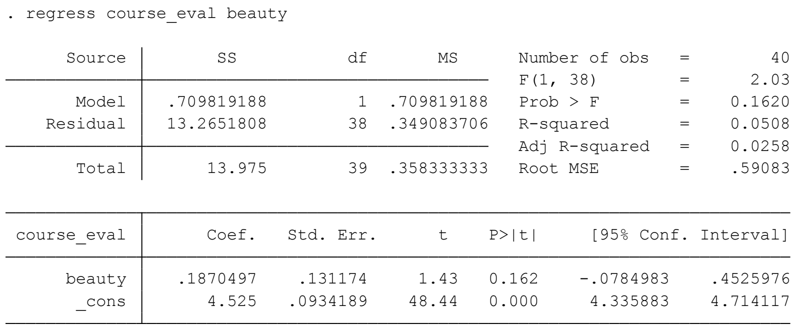
(v) The data with beauty plotted on the x-axis is below.



(vi) Using $\hat{\beta}_1 = \frac{cov(X,Y)}{var(X)}$ to calculate the regression slope coefficient the value is .279

```
. correlate beauty course_eval, cov
(obs=40)

             |   beauty course~l
-------------+------------------
      beauty |  .520199
 course_eval |  .097303   .358333


. sum beauty course_eval

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      beauty |         40   -1.02e-08     .721248  -1.303629   1.848021
 course_eval |         40       4.525    .5986095          3          5

. generate beta1hat = .097303/(.5896095)^2

. display beta1hat
.27989638

.
```

(vii) For the data in ps1q4.dta the value of $\hat{\beta}_0$ is 4.525 and this does relate to course_eval because it indicates a course evaluation of 4.525 if beauty is zero.



```
. regress course_eval beauty

      Source |       SS           df       MS      Number of obs   =        40
-------------+----------------------------------   F(1, 38)        =      2.03
       Model |  .709819188         1  .709819188   Prob > F        =    0.1620
    Residual |  13.2651808        38  .349083706   R-squared       =    0.0508
-------------+----------------------------------   Adj R-squared   =    0.0258
       Total |      13.975        39  .358333333   Root MSE        =    .59083

------------------------------------------------------------------------------
 course_eval |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      beauty |   .1870497     .131174     1.43   0.162    -.0784983    .4525976
       _cons |      4.525    .0934189    48.44   0.000     4.335883    4.714117
------------------------------------------------------------------------------
```

(viii) Using the regression $course_{eval} = \beta_0 + \beta_1 * beauty + u_i$

The OLS estimates is:

$\hat{\beta}_1$ *= .187*

compare β_1 in step vi with β_1 found in step vi:

The step vi estimates is:

$\hat{\beta}_1$ *= .225*

(ix) R^2 is the ratio of the explained variation compared to the total variation and is interpreted as the fraction of the sample variation in Y (dependent variable) that is explained by X (independent variable). For this exact model, in question 4, R^2 measures how much of the variation in course_evaluation score is explained by beauty. R^2 is generally multiplied by 100 to change it to a percent.

The standard error (RMSE) of a regression measures the sample standard deviation of the forecast of errors (without any degrees of freedom adjustment)

R^2 is preferred because it is unitless.

(x) R^2*in step vii is .0508.*
5% *of the variation in course_eval score is predicted by beauty.*

Appendix I