|  |
| --- |
| By Cliff Rodriguez |
| Problem Set 1 |
| Due: January 31, 2018 |

|  |
| --- |
| Cliff Rodriguez |

Contents

[Question 1 2](#_Toc504487748)

[Question 2 2](#_Toc504487749)

[Question 3 6](#_Toc504487750)

[Question 4 7](#_Toc504487751)

[Figure 1: mean and standard deviation for each variable in dataset ps1q2.dta 2](#_Toc504487752)

[Figure 2: covariance and correlation between years of education and hourly wages in dataset ps1q2.dta 3](#_Toc504487753)

# Question 1

Question 1 uses dataset ps1q1.dta

1. 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.
2. Three examples of things that could be included in for this model are
   1. Height of father
   2. Heightof Mother
   3. Were prenatal vitamins used by the mother
3. Give and intuitive argument for why:
   1. might be positive in this model if an additional cigarette increases birthweight.
   2. might be negative in this model if an additional cigarette decreases birthweight.
4. Simple regression does or does not solve the problem of omitted variable bias and reverse causality
5. *Using the output from STATA below:*

**

*Estimated Equation: bwghti* *119.77* *-.5137721 cigsi* *ui*

*119.77*

*Interpretation: If no cigarettes are smoked per day during pregnancy the estimated birthweight is 7.34 ounces.*

*-.5137721*

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

1. The predictedbirthweight when *cigsi =0 is 119.77 ounces.*

The predictedbirthweight when *cigsi =20 is 109.5 ounces.*

Comparing the values of *119.77* ounces and 109.5 ounces, which are the expected values for birthweight corresponding to a, expecting mother smoking 0 and 20 cigarettes per day while pregnant indicates that smoking a pack a day will decrease the expected birthweight of the baby by roughly 10 ounces.

1. *I am having issues getting this question to make sense, unless a positive Ui is needed using the equation: bwghti* *119.77* *-.5137721 cigsi* *ui* 
   1. *The strong assumption made in this model includes … because it is linear.*
   2. *How would I model this using the natural log?*
   3. *Concerns using log*

# Question 2

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

*Calculated using the command sum*

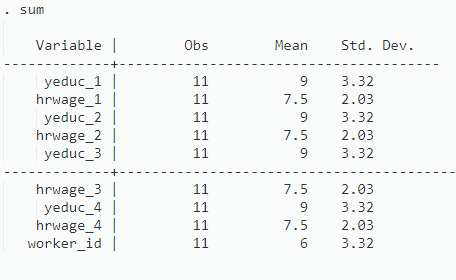


Figure : Mean and standard deviation for each variable in dataset ps1q2.dta

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

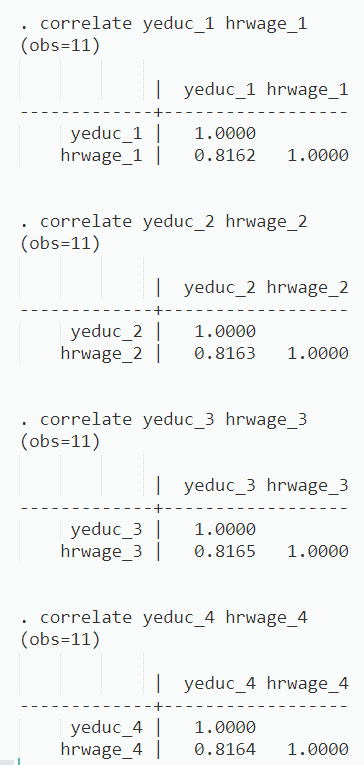


Figure : correlation between years of education and hourly wages in dataset ps1q2.dta



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

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

Firm 1 = .5

Firm 2 = .5

Firm 3 = .5

Firm 4 = .5

**

**

1. 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:
   * 1. Firm 1: 50 cents per hour
     2. Firm 2: 50 cents per hour
     3. Firm 3: 50 cents per hour
     4. Firm 4: 50 cents per hour
2. The relationship between hourly wages and years of education for each firm is graphed below.

Firm 1



Firm 2



Firm 3



Firm 4



1. Reviewing the graphs the prediction is equally good for each firm. This is because the value for each is .5, meaning income increases by .50 cents per hour per each additional year of education.
2. 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.
3. Based on points *v* and *vi* above it is suggested that ……. About the choice of model in estimation

# Question 3

1. See appendix I for work

*=*

*=*

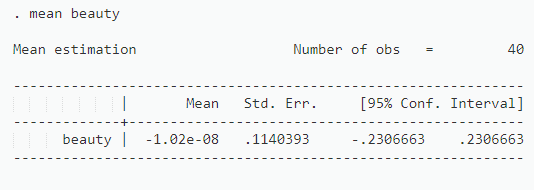
1. The term in this model is/is not useful because…
2. The GPA score is predicted to be …. if the ACT score increases by 5 points?
3. The fitted values and residuals for each observation are presented in appendix II.

Calculation showing residuals approximately sum to zero:

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

# Question 4

1. *= -1.02 e-08, or near zero*



1. *=*
2. The covariance between course evaluation and beauty is …

**

**

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

1. The correlation between course evaluations and beauty is shown below using STATA and the

**

**

Which one should I use? Which part of the chart should hold x versus y variable?

1. The data with beauty plotted on the x-axis is below.



1. Using to calculate the regression slope coefficient the value is …
2. For the data in ps1q4.dta the value of is

This does or does not relate to *course\_eval* because

**

1. Using the regression

*The OLS estimates are:*

*= 4.53*

*= .19*

*compare in step vi with found in step vi:*

1. 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, measures how much of the variation in course\_evaluation score is explained by beauty. 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)

I prefer R2 or RMSE? WHY?

1. *in step vii is .0508.*

*of the variation in course\_eval score is predicted by beauty.*