A

Name:

Student Number:

Midterm 1

ECO3151 C

Christopher Ksoll

Winter 2014

Instructions:

1. Print your name and student number at the top of this midterm
2. No programmable calculators
3. You can answer in pencil or pen
4. This midterm consists of 7 short answer questions
5. Total marks: 60

Question 1

a) Show that $(\hat{\beta}_0 - \beta_0) = \bar{u} + (\beta_1 - \hat{\beta}_1)\bar{x}$, where$\bar{u}$ is the average error terms in the sample you are looking at. (5 marks)

b) If a simple linear regression yields $\hat{\beta}_1 = 0$, show that its R^2 is equal to zero. (5 marks)

Question 2

You are interested in measuring the impact of government job training programs on subsequent wages. Your co-author tells you that he can access a dataset where job training was randomly assignment, i.e. the decision on who would (or would not) receive the training program was determined by a lottery.

a) Why would that be useful? (5 marks)

b) Outline the regression equation you would estimate (2 points)

c) What conditions would you need for your estimates to have a causal interpretation (3 points)

Question 3

The data used in the following question come from a study of an adult literacy program that I am studying in India. *Literacytestscore* is the test score on a literacy test, age is the age of the learner and age2 is the quadratic in age. The regression output from stata is attached at the end of the exam.

a) Write the normal form of the regression results (2 points)

b) What is the impact of age on the test score? What is the impact of IQ on the test score? (6 points)

c) What is the interpretation of the coefficient on the constant (2 points)

Question 4
Is $E(z(m - E(m)) = E(m(z - E(z))$? Prove your conclusion (5 marks)

Question 5
Does the OLS approach imply that $\sum_{i=1}^{n} u_i/n = 0$. Justify your answer. (5 marks)

Question 6

A sample consists of 506 communities in the Boston area where $logprice_i$ is the logarithm of the median housing price in community i, and $dist_i$ is the average distance of the community from five employment centers. Given the information found in the log file (which is on the last page of the midterm)

a) Write the population model. (2 marks)

b) Interpret the estimate of $\hat{\beta}_1$. (3 marks)

c) In this course, the professor has stressed the importance of the causality assumption. What is it? and do you think it holds in this case? Justify your answer. (5 marks)

d) Provide two additional variables that should be included in the model. Write out the new population model. Make sure to clearly define your variables, and explain why they should be included. (5 marks)

Question 7

Discuss the validity of the following statement: Many factors that affect children's grades, other than family income, are not observable to the empirical researcher. As such, one will never be able to reasonably estimate the causal impact of family income on grades. (5 marks)

Equations

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad SSR = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

$$VAR(x) = E(x - E(x))(x - E(x)) = E(x^2) - [E(x)]^2$$

$$y_i = \hat{y}_i + \hat{u}_i$$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

$$R^2 = SSE/SST$$

$$var(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$$

$$sd(\hat{\beta}_1) = \frac{\sigma}{\sqrt{SST_x}}$$

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SST_x}}$$

$$\sum_{i=1}^{n} x_i + y_i = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i, \quad \sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i, \quad \sum_{i=1}^{n} a = na$$

Question 3

```
. reg literacytestscore age age2 iq

      Source |       SS       df       MS              Number of obs =     121
-------------+------------------------------           F(  3,   117) =   13.38
       Model |   5318.8316     3  1772.94387            Prob > F      =  0.0000
    Residual |  15503.0362   117  132.504583            R-squared     =  0.2554
-------------+------------------------------           Adj R-squared =  0.2364
       Total |  20821.8678   120  173.515565            Root MSE      =  11.511

------------------------------------------------------------------------------
literacyte~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -2.099088    .610936    -3.44   0.001    -3.309014    -.889161
        age2 |   .0211823   .0087927     2.41   0.018     .0037689    .0385958
          iq |   .7035932   .2632154     2.67   0.009     .1823088    1.224878
       _cons |   57.30414   10.06078     5.70   0.000      37.3793    77.22898
------------------------------------------------------------------------------
```

Question 6

```
. reg lprice dist

      Source |       SS       df       MS              Number of obs =     506
-------------+------------------------------           F(  1,   504) =   66.76
       Model |  9.89355935     1  9.89355935            Prob > F      =  0.0000
    Residual |  74.6886656   504  .148191797            R-squared     =  0.1170
-------------+------------------------------           Adj R-squared =  0.1152
       Total |   84.582225   505  .167489554            Root MSE      =  .38496

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dist |   .0664575   .0081335     8.17   0.000     .0504777    .0824373
       _cons |   9.688801   .0352988   274.48   0.000      9.61945    9.758152
------------------------------------------------------------------------------
```