Name:

Student Number:

Midterm 1

ECO3151

Winter 2018

VERSION 1

Instructions:

1. Print your name and student number at the top of this midterm
2. (a) No programmable calculators

   (b) No textbooks, notes, or scrap paper.

   (c) Close all phones
3. You can answer in pencil or pen
4. This midterm consists of 9 short answer questions
5. Total marks: 50

Question 1

A sample consists of 709 individuals where $sleep_i$ is the number of minutes individual i spent sleeping per week, and age_i is his/her age in years. 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) Provide two variables that could possibly affect your dependent variable, but was not accounted for in the simple linear regression model. Make sure to clearly define your variables and briefly explain why you believe they would affect your dependent variable. (5 marks)

Question 2

Explain in words the importance of having a random sample in empirical work. (5 marks)

Question 3

a) A clear distinction was made between u_i and $\hat{u}_i$. What is that distinction? You cannot exclusively rely on equations to answer the question. You must explain your answer in words (3 marks)

b) Under what conditions would $u_i = \hat{u}_i$ for $i = 1, \ldots, n$ hold? (3 marks)

Question 4

It has been argued in class that the linear projection plays a critical role if one is to rely on the method of moments estimator. Discuss. (5 marks)

Question 5

Show that $\hat{y}_i - \bar{y} = \hat{\beta}_1^{ols}(x_i - \bar{x})$ (5 marks)

Question 7
Discuss the validity of the following statement (5 marks):

Many factors, other than family income, affect children's grades. As such, one will never be able to reasonably estimate the causal impact of family income on grades using a simple linear regression approach.

Question 8
Show that $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})x_i$. (5 marks)

Question 9
Given the following regression output

$$\hat{y} = \underset{(3.2)}{400} - \underset{(2.0)}{150x}$$

$$R^2 = 0.50 \qquad n = 200,231$$

where the standard errors are in brackets. Discuss the validity of the following statements. Said differently, can one draw the following conclusion based on the above information. Explain your answer.

a) x causes y (3 marks)

b) The effect of x is economically significant. (3 marks)

c) $\hat{\beta}_1$ is an unbiased estimator of β_1. (3 marks)

Equations

$$E(x+y) = E(x) + E(y)$$
$$E(ax) = aE(x) \quad \text{(where } a \text{ is a constant)}$$
$$E(a) = a \quad \text{(where } a \text{ is a constant)}$$

$$Var(x) = E[(x - E(x))(x - E(x))] = E(x^2) - [E(x)]^2$$
$$Cov(x,y) = E[(x - E(x))(y - E(y))] = E(xy) - E(x)E(y)$$
$$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$$

$$y = \beta_0 + \beta_1 x + u$$
$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$\sum_{i=1}^{n} x_i + y_i = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$
$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i \quad \text{(where } a \text{ is a constant)}$$
$$\sum_{i=1}^{n} a = na \quad \text{(where } a \text{ is a constant)}$$

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$
$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} (x_i - \bar{x}) y_i$$
$$\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x}) x_i$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$
$$\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$$

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

$$\hat{\beta}_1^{ols} = \hat{\beta}_1^{mm} = \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$\hat{\beta}_0^{ols} = \hat{\beta}_0^{mm} = \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\sum_{i=1}^{n} \hat{u}_i = 0$$
$$\sum_{i=1}^{n} \hat{u} x_i = 0$$

$$var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$sd(\hat{\beta}_1) = \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$
$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}$$
$$var(\hat{\beta}_0) = \frac{\sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$sd(\hat{\beta}_0) = \frac{\sigma \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$
$$se(\hat{\beta}_0) = \frac{\hat{\sigma} \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

SLR.1 Linear in parameters
SLR.2 Random sampling
SLR.3 Sample variation in the explanatory variable
SLR.4 Zero conditional mean (i.e. $E(u|x) = 0$)
SLR.5 Homoskedasticity (i.e. $Var(u|x) = \sigma^2$)

Note:
$E(u|x) = 0 \quad \Leftrightarrow \quad [Corr(u, x) = 0$ and "the model has the correct functional form"$]$

```
  ___  ____  ____  ____  ____ tm
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/
 Statistics/Data Analysis
```

```
  ___  ____  ____  ____  ____ tm
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   10.1  Copyright 1984-2009
  Statistics/Data Analysis           StataCorp
                                     4905 Lakeway Drive
     Special Edition                 College Station, Texas 77845 USA
                                     800-STATA-PC        http://www.stata.com
                                     979-696-4600        stata@stata.com
                                     979-696-4601 (fax)

Single-user Stata for Windows perpetual license:
       Serial number:  81910047170
         Licensed to:  Pierre Brochu
                       University of Ottawa

Notes:
      1.  (/m# option or -set memory-) 10.00 MB allocated to data
      2.  (/v# option or -set maxvar-) 5000 maximum variables

1 . use "C:\Data_Wooldridge\SLEEP75.DTA", clear

2 . keep sleep age

3 . sum sleep age

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       sleep |       706    3266.356    444.4134        755       4695
         age |       706    38.81586    11.34264         23         65

4 . gen lsleep=ln(sleep)

5 . gen lage=ln(age)

6 . reg sleep age

      Source |       SS       df       MS              Number of obs =     706
-------------+------------------------------           F(  1,   704) =    5.80
       Model |  1137207.85     1  1137207.85           Prob > F      =  0.0163
    Residual |   138102628   704  196168.506           R-squared     =  0.0082
-------------+------------------------------           Adj R-squared =  0.0068
       Total |   139239836   705  197503.313           Root MSE      =  442.91

------------------------------------------------------------------------------
       sleep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   3.540881   1.470639     2.41   0.016     .6535177    6.428244
       _cons |   3128.913   59.46811    52.61   0.000     3012.157    3245.669
------------------------------------------------------------------------------

7 .
```