Queen's University
Department of Economics

ECON 351* -- Introductory Econometrics

ASSIGNMENT 1: ANSWERS

Winter Term 2009 M.G. Abbott

TOPIC: **Interpreting OLS Coefficient Estimates in Simple Linear Regression Models**

INSTRUCTIONS:

- Answer all questions on standard-sized 8.5 x 11-inch paper.
- Answers need not be typewritten (document processed), but if hand-written must be legible. Illegible assignments will be returned unmarked.
- Please label clearly each answer with the appropriate question number and letter. Securely staple all answer sheets together, and make certain that your ***name(s)*** and ***student number(s)*** are printed clearly at the top of each answer sheet.
- Students submitting joint assignments with one other student must ensure that the name and student number of each student are printed clearly at the top of each answer sheet. Submit only one copy of the assignment.

MARKING: Marks for each question are indicated in parentheses. Total marks for the assignment equal 70. Marks are given for both content and presentation.

SOFT DUE DATE: **Monday February 9, 2009 by 4:00 pm**.

HARD DUE DATE: **Thursday February 12, 2009 by 4:00 pm**.

- Assignments submitted **on or before** the soft due date will receive a bonus of 2 points to a maximum total mark of 70.
- Assignments submitted **after** the hard due date will be penalized 20 points per day.
- Please submit your assignments either to me in class, or by depositing them in the ECON 351 slot of the **Assignment Collection Box** located immediately **inside the double doors** on the **second floor of Dunning Hall** (opposite the elevator).

DATA FILE: **351assn1w09.raw** (a text-format, or ASCII-format, data file)

- ***Data Description:*** A random sample of 472 employees drawn from the 1976 U.S. population of all employed paid workers.
- ***Variable Definitions:***

  $WAGE_i \equiv$ average hourly earnings of worker i in 1976, in *dollars per hour*.
  $ED_i \equiv$ years of formal education completed by worker i, in *years*.
  $FEMALE_i \equiv$ an indicator variable equal to 1 if worker i is female, and 0 if worker i is male.

- ***Stata Infile Statement:*** Use the following *Stata* **infile** statement to read the text-format data file **351assn1w09.raw**:

  infile wage ed female using 351assn1w09.raw

QUESTIONS and ANSWERS:

(5 marks)

1. Compile a table of descriptive summary statistics for the sample data. The table should include for each of the variables in the dataset: the sample mean, the sample standard deviation, the minimum sample value, and the maximum sample value. How many females and how many males are there in the sample?

ANSWER Question 1 **(1 mark) per column in table**

Variable Name	Mean	Std. Deviation	Minimum	Maximum
wage	5.95917	3.77002	0.53	25.00
ed	12.6419	2.69169	0	18
female	0.480932	0.500166	0	1

Number of *females* in the sample = **227** **(0.5 mark)**

Number of *males* in the sample = **245** **(0.5 mark)**

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       472    5.959174    3.770022        .53         25
          ed |       472    12.64195    2.691691          0         18
      female |       472    .4809322    .5001664          0          1

. tab1 female, missing

-> tabulation of female

     = 1 if |
female, = 0 |
    if male |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        245       51.91       51.91
          1 |        227       48.09      100.00
------------+-----------------------------------
      Total |        472      100.00
```

(20 marks)

2. Compute and present OLS estimates of the following population regression equation for the full sample of 472 paid workers:

$$WAGE_i = \beta_0 + \beta_1 ED_i + u_i \qquad (1)$$

where u_i is a random error term that is assumed to satisfy all the assumptions of the classical linear regression model.

(5 marks)

(a) Report the OLS coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ computed by estimating population regression equation (1).

ANSWER Question 2(a)

$\hat{\beta}_0 = -1.31250$ **(2.5 marks)**

$\hat{\beta}_1 = 0.575202$ **(2.5 marks)**

(5 marks)

(b) Interpret the value of the slope coefficient estimate $\hat{\beta}_1$; i.e., explain in words what the numerical value of $\hat{\beta}_1$ means.

ANSWER Question 2(b)

Answer must not be just a generic description of the slope coefficient estimate; it must explicitly account for the units in which WAGE and ED are measured.

WAGE is measured in ***dollars* per hour**; **ED** is measured in ***years***.

Therefore, the estimate $\hat{\beta}_1 = 0.5752$ means that a ***1-year increase* in education** is associated with an **increase in average hourly wages equal to 0.5752 *dollars* per hour**, or equivalently **57.52 *cents* per hour**. **(5 marks)**

(5 marks)

(c) Interpret the value of the intercept coefficient estimate $\hat{\beta}_0$; i.e., explain in words what the numerical value of $\hat{\beta}_0$ means.

ANSWER Question 2(c)

The estimate $\hat{\beta}_0 = -1.3125$ means that **the *average* (*mean*) hourly wage rate of workers with *zero* years of education (ED = 0) equals –1.3125 *dollars per hour*, or –131.25 *cents per hour***. **(4 marks)**

Note: There are almost no workers in the sample for whom ED = 0. In fact, **only *two* of 472 workers** in the sample have zero years of formal education: they are both female workers whose observed hourly wage rates are $2.90 per hour and $4.20 per hour. **Because there are so few sample observations for which ED = 0, it is difficult to obtain from this sample a good estimate of the mean hourly wage rate of all such workers in the population.** **(1 mark)**

(5 marks)

(d) On a set of appropriately labeled coordinate axes, draw the estimated sample regression function implied by OLS estimation of regression equation (1). That is, draw the graph of the equation $\widehat{WAGE}_i = \hat{\beta}_0 + \hat{\beta}_1 ED_i$, compute the coordinates of the two points on it that correspond to the values 12 and 16 of ED_i, and label these two points on your graph as A and B respectively. (Note: you do not need to use *Stata*, or any software program, to draw and label this graph.)

ANSWER Question 2(d)

The two points have the following coordinates:

Point A: For **ED_i = 12 years**, the estimated mean of average hourly earnings equals

$$\widehat{WAGE}_i = \hat{\beta}_0 + \hat{\beta}_1 ED_i = -1.31250 + 0.575202(12) = \textbf{5.589924 dollars per hour}$$
$$= \textbf{\$5.59 per hour} \quad \textbf{(1 mark)}$$

Point B: For **ED_i = 16 years**, estimated annual beer expenditure equals

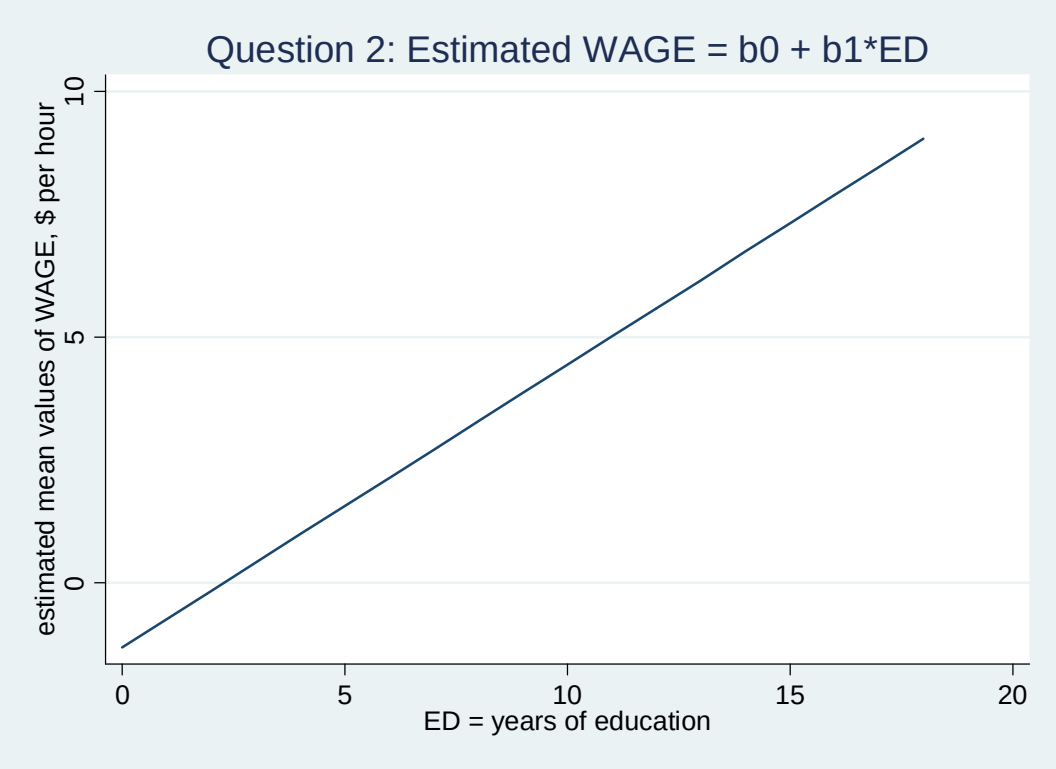
$$\widehat{WAGE}_i = \hat{\beta}_0 + \hat{\beta}_1 ED_i = -1.31250 + 0.575202(16) = \textbf{7.890732 dollars per hour}$$
$$= \textbf{\$7.89 per hour} \quad \textbf{(1 mark)}$$

Note: $\Delta\widehat{WAGE}_i / \Delta ED_i = (7.890732 - 5.589924)/(16 - 12) = 2.300808/4 = 0.575202 = \hat{\beta}_1$.

ANSWER Question 2(d), continued

Figure 1: Line Graph of $\widehat{\text{WAGE}}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{ED}_i = -1.31250 + 0.575202\,\text{ED}_i$

(3 marks) total: 2 marks for correct line graph; 1 mark for labeling points A and B



Stata output for Question 2:

```
. regress wage ed

      Source |       SS       df       MS              Number of obs =     472
-------------+------------------------------           F(  1,   470) =   95.35
       Model |  1129.04675     1  1129.04675           Prob > F      =  0.0000
    Residual |  5565.30746   470  11.8410797           R-squared     =  0.1687
-------------+------------------------------           Adj R-squared =  0.1669
       Total |   6694.3542   471  14.2130662           Root MSE      =  3.4411

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .5752019   .0589061     9.76   0.000     .4594501    .6909537
       _cons |  -1.312499   .7613451    -1.72   0.085    -2.808561    .1835623
------------------------------------------------------------------------------
```

(25 marks)

3. Compute OLS coefficient estimates of the following population regression equation for the full sample of 472 workers:

$$WAGE_i = \beta_0 + \beta_1 FEMALE_i + u_i \qquad (2)$$

where u_i is a random error term that is assumed to satisfy all the assumptions of the classical normal linear regression model.

(5 marks)

(a) Report the OLS coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ computed by estimating population regression equation (2).

ANSWER Question 3(a)

$\hat{\beta}_0$ = **7.185306** **(2.5 marks)**

$\hat{\beta}_1$ = **–2.549491** **(2.5 marks)**

(5 marks)

(b) Before attempting to interpret the OLS coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ in regression equation (2), compute and report the sample mean of $WAGE_i$ for females and the sample mean of $WAGE_i$ for males.

ANSWER Question 3(b)

sample mean of $WAGE_i$ for *females* = 4.635815 dollars ($4.64) per hour **(2.5 marks)**

sample mean of $WAGE_i$ for *males* = 7.185306 dollars ($7.19) per hour **(2.5 marks)**

```
. summarize wage if female == 0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       245    7.185306    4.236225        1.5         25

. summarize wage if female == 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       227    4.635815    2.618648        .53         22
```

(5 marks)

(c) Interpret the value of the slope coefficient estimate $\hat{\beta}_1$ in equation (2); i.e., explain in words what it means. Remember that $FEMALE_i$ is an indicator (or dummy) variable that takes only the two values 0 and 1.

ANSWER Question 3(c)

Since $\hat{E}(WAGE_i \mid FEMALE_i = 1) = \hat{\beta}_0 + \hat{\beta}_1$ and $\hat{E}(WAGE_i \mid FEMALE_i = 0) = \hat{\beta}_0$,

$\hat{E}(WAGE_i \mid FEMALE_i = 1) - \hat{E}(WAGE_i \mid FEMALE_i = 0) = \hat{\beta}_0 + \hat{\beta}_1 - \hat{\beta}_0 = \hat{\beta}_1$

$\hat{\beta}_1$ **= –2.549491 = the sample mean hourly wage rate of female workers *minus* the sample mean hourly wage rate of male workers**

The value **–2.549491** means that **the average hourly wage rate of *female workers* is *less than* the average hourly wage rate of *male workers* in the sample by <u>2.549491 dollars per hour</u>**, or **<u>$2.55 per hour</u>**. **(5 marks)**

(5 marks)

(d) Interpret the value of the intercept coefficient estimate $\hat{\beta}_0$ in equation (2); i.e., explain in words what it means.

ANSWER Question 3(d)

$\hat{E}(PRICE_i \mid FEMALE_i = 0) = \hat{\beta}_0$ **= 7.185306**

= the sample mean hourly wage rate of *male workers*.

The value **7.185306** means that **the average hourly wage rate of *male workers* in the sample** is **<u>7.185306 dollars per hour</u>**, or **<u>$7.19 per year</u>**. **(5 marks)**

(5 marks)

(e) Compute and interpret the value of $\hat{\beta}_0 + \hat{\beta}_1$, the sum of the OLS coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for equation (2); i.e., explain in words what this sum means.

ANSWER Question 3(e)

$\hat{\beta}_0 + \hat{\beta}_1$ is the predicted **sample *mean* hourly wage rate of *female workers***, since $\hat{E}(WAGE_i \mid FEMALE_i = 1) = \hat{\beta}_0 + \hat{\beta}_1$.

The sample mean value of the hourly wage rate of *female workers* is therefore equal to $\hat{\beta}_0 + \hat{\beta}_1 = 7.185306 - 2.549491 =$ **<u>4.635815 dollars per hour</u> = <u>$4.64 per hour</u>**.

(5 marks)

Stata output for Question 3:

```
. regress wage female

      Source |       SS       df       MS              Number of obs =     472
-------------+------------------------------           F(  1,   470) =   60.72
       Model |   765.87336     1   765.87336           Prob > F      =  0.0000
    Residual |  5928.48084   470   12.613789           R-squared     =  0.1144
-------------+------------------------------           Adj R-squared =  0.1125
       Total |   6694.3542   471  14.2130662           Root MSE      =  3.5516

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -2.549491   .3271883    -7.79   0.000    -3.192424   -1.906558
       _cons |   7.185306   .2269027    31.67   0.000     6.739437    7.631176
------------------------------------------------------------------------------

. summarize wage if female == 0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       245    7.185306    4.236225        1.5         25

. summarize wage if female == 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       227    4.635815    2.618648        .53         22

. display _b[female]
-2.5494912

. display _b[_cons]
7.1853061

. display _b[_cons] + _b[female]
4.635815
```

(20 marks)

4. Estimate separate OLS regressions of average hourly earnings on years of formal education for the subsample of females and the subsample of males. Write the female regression equation for $WAGE_i$ as

$$WAGE_i = \alpha_0 + \alpha_1 ED_i + u_i \qquad \text{if } FEMALE_i = 1 \qquad (3.1)$$

Write the male regression equation for $WAGE_i$ as

$$WAGE_i = \beta_0 + \beta_1 ED_i + u_i \qquad \text{if } FEMALE_i = 0 \qquad (3.2)$$

(8 marks)

(a) Compute and report the OLS coefficient estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$ for females and the OLS coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for males.

ANSWER Question 4(a)

For **females** ($FEMALE_i = 1$), the OLS coefficient estimates are:

$\hat{\alpha}_0$ = **–1.342252 = –1.342** **(2 marks)**

$\hat{\alpha}_1$ = **0.4851703 = 0.4852** **(2 marks)**

For **males** ($FEMALE_i = 0$), the OLS coefficient estimates are:

$\hat{\beta}_0$ = **–0.030571 = –0.0306** **(2 marks)**

$\hat{\beta}_1$ = **0.557694 = 0.5577** **(2 marks)**

(12 marks)

(b) Interpret the value of the female slope coefficient estimate $\hat{\alpha}_1$ for equation (3.1) and the value of the male slope coefficient estimate $\hat{\beta}_1$ for equation (3.2). Which slope coefficient estimate is larger? Can you legitimately conclude from this comparison of $\hat{\alpha}_1$ and $\hat{\beta}_1$ that the marginal effect of years of education on hourly earnings is different between males and females? Explain briefly.

ANSWER Question 4(b)

Interpretation of $\hat{\alpha}_1$ = 0.4852 for *females*:

For ***female workers***, a ***one-year increase*** **in years of formal education** is associated on average with an increase in average hourly earnings of **0.4852 *dollars* per hour**, or **48.52 *cents* per hour**. **(3 marks)**

Interpretation of $\hat{\beta}_1$ = 0.5577 for *males*:

For ***male workers***, a ***one-year increase*** **in years of formal education** is associated on average with an increase in average hourly earnings of **0.5577 *dollars* per hour**, or **55.77 *cents* per hour**. **(3 marks)**

The **male** slope coefficient estimate $\hat{\beta}_1$ is **larger** than the **female** slope coefficient estimate $\hat{\alpha}_1$, which suggests that the increase in mean hourly wage rates associated with a one-year increase in years of formal education is greater for male workers than for female workers: i.e., $\hat{\beta}_1 > \hat{\alpha}_1$. **(1 mark)**

Can you legitimately conclude from this comparison of $\hat{\alpha}_1$ and $\hat{\beta}_1$ that the marginal effect of years of education on hourly wage rates is different between male and female workers?
NO

(2 marks)

Explanation: The specific values of $\hat{\alpha}_1$ and $\hat{\beta}_1$ for the sample of 472 workers are point estimates that have some statistical error associated with them. Before we can legitimately conclude that β_1 for males is greater than α_1 for females, we would have to take account of these statistical errors. **(3 marks)**

Stata output for Question 4:

```
. *
. * Q 4(a): Eq (3.1) Regression of WAGE on ED for female == 1
. *
. regress wage ed if female == 1

      Source |       SS       df       MS              Number of obs =     227
-------------+------------------------------           F(  1,   225) =   60.08
       Model |  326.609596     1  326.609596           Prob > F      =  0.0000
    Residual |  1223.14414   225  5.43619616           R-squared     =  0.2107
-------------+------------------------------           Adj R-squared =  0.2072
       Total |  1549.75373   226   6.8573174           Root MSE      =  2.3316

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .4851703   .0625932     7.75   0.000     .3618264    .6085142
       _cons |  -1.342252     .78662    -1.71   0.089    -2.892337    .2078324
------------------------------------------------------------------------------
.
. * Q 4(a): Eq (3.2) Regression of WAGE on ED for female == 0
. regress wage ed if female == 0

      Source |       SS       df       MS              Number of obs =     245
-------------+------------------------------           F(  1,   243) =   39.77
       Model |  615.850074     1  615.850074           Prob > F      =  0.0000
    Residual |  3762.87704   243  15.4850907           R-squared     =  0.1406
-------------+------------------------------           Adj R-squared =  0.1371
       Total |  4378.72711   244  17.9456029           Root MSE      =  3.9351

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |    .557694   .0884332     6.31   0.000     .3835005    .7318874
       _cons |   -.030571   1.171511    -0.03   0.979    -2.338183    2.277041
------------------------------------------------------------------------------
```