Queen's University
Department of Economics

ECON 351* -- Introductory Econometrics

ASSIGNMENT 2: ANSWERS

Winter Term 2009 M.G. Abbott

TOPIC: **Statistical Inference 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 100. Marks are given for both content and presentation.

SOFT DUE DATE: **Friday February 27, 2009 by 4:00 pm**.

HARD DUE DATE: **Thursday March 5, 2009 by 4:00 pm**.

- Assignments submitted **on or before** the soft due date will receive a bonus of 3 points to a maximum total mark of 100.
- 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. NOTE: Assignment 2 uses that same dataset as Assignment 1.
- ***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:

(50 marks)

1. Compute and present OLS estimates of the following population regression equation for the full sample of 472 employees:

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

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

(10 marks)

(a) Report the OLS coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ computed by estimating population regression equation (1), the estimated standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$, the t-ratios for $\hat{\beta}_0$ and $\hat{\beta}_1$, the estimate $\hat{\sigma}^2$ of the constant error variance σ^2, the R^2 and F-statistic for the OLS sample regression equation, and the number of observations N used in estimation. .

<u>ANSWER Question 1(a)</u>

```
      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
------------------------------------------------------------------------------
```

Regressor	$\hat{\beta}_j$	$\hat{se}(\hat{\beta}_j)$	$t(\hat{\beta}_j)$	Lower 95% limit	Upper 95% limit
Constant	**-1.31250**	**0.761345**	**-1.72**	-2.80856	0.183562
ED_i	**0.575202**	**0.0589061**	**9.76**	0.459450	0.690954

N = **472**; $\hat{\sigma}^2$ = **11.84108**; R^2 = **0.1687**; F(1, 470) = **95.35**

(10 marks)

(b) Use the estimation results from OLS estimation of regression equation (1) to perform a t-test of the proposition that employees' hourly wage rates are unrelated to their years of formal education. State the null and alternative hypotheses. Show how the appropriate test statistic is calculated (give its formula). Report the sample value of the test statistic and its p-value. State the decision rule you use to decide between rejection and retention of the null hypothesis. State the inference you would draw from the test at the 5 percent and 1 percent significance levels. What would you conclude from the results of the test?

ANSWER Question 1(b): (total marks = 10)

- **Null hypothesis:** H_0: $\beta_1 = 0$
  Alternative hypothesis: H_1: $\beta_1 \neq 0$ a ***two-tail*** **test** **(2 marks)**

- **Calculation of t-test statistic:** The ***sample value*** $\mathbf{t_0}$ under the null hypothesis H_0 is

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - b_1}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)} = \frac{0.575202}{0.0589061} = \underline{\mathbf{9.76473}} = \underline{\mathbf{9.76}}$$ **(3 marks)**

- ***Two-tail*** **p-value for $\mathbf{t_0}$ = <u>0.0000</u>** **(1 mark)**

- **<u>Decision rule and inference</u>:** ***either formulation 1 or 2*** **(3 marks)**

(1) <u>Decision Rule -- Formulation 1</u>: This a ***two-tail*** **test**. Compare the sample value t_0 with the **α/2-level critical value of the t[N–2] = t[470] distribution**.

1. If $|t_0| \leq t_{\alpha/2}[N-2]$, ***retain*** **(do not reject) $\mathbf{H_0}$** at the 100α percent significance level.
2. If $|t_0| > t_{\alpha/2}[N-2]$, ***reject*** **$\mathbf{H_0}$** at the 100α percent significance level.

<u>Inference</u>:

Since $|t_0(\hat{\beta}_1)|$ = **9.76 > 1.965 = $\mathbf{t_{0.025}}$[470]**, ***reject*** **$\mathbf{H_0}$** at the ***5 %*** significance level.

Since $|t_0(\hat{\beta}_1)|$ = **9.76 > 2.586 = $\mathbf{t_{0.005}}$[470]**, ***retain*** **$\mathbf{H_0}$** at the ***1 %*** significance level.

(2) <u>Decision Rule -- Formulation 2</u>: This is a ***two-tail*** **test**. Compare the ***two-tail*** **p-value for $\mathbf{t_0}$** with the chosen **significance level** α.

1. If ***two-tail*** **p-value** for $t_0 \geq \alpha$, ***retain*** **(do not reject) $\mathbf{H_0}$** at significance level α.
2. If ***two-tail*** **p-value** for $t_0 < \alpha$, ***reject*** **$\mathbf{H_0}$** at significance level α.

<u>Inference</u>:

Since ***two-tail*** **p-value for $\mathbf{t_0}$ = 0.0000 < 0.05**, ***reject*** **$\mathbf{H_0}$** at the ***5 %*** significance level.
Since ***two-tail*** **p-value for $\mathbf{t_0}$ = 0. 0000 < 0.01**, ***reject*** **$\mathbf{H_0}$** at the ***1 %*** significance level.

- **<u>Result</u>:** **(1 mark)**
  The sample evidence favours the ***alternative*** **hypothesis** that **$\mathbf{\beta_1 \neq 0}$ at *both* the 5% and 1%**, significance levels.

Alternative ANSWER to Question 1(b): two-sided confidence interval for β_1, total marks = 10

- **Null hypothesis:** H_0: $\beta_1 = 0$
  Alternative hypothesis: H_1: $\beta_1 \neq 0$ a ***two-tail*** **test** **(2 marks)**

- **Two-sided 100(1−α) percent confidence interval for β_1** is:

  Lower 100(1−α) percent confidence limit for β_1 = $\hat{\beta}_1 - t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_1)$
  Upper 100(1−α) percent confidence limit for β_1 = $\hat{\beta}_1 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_1)$

- Compute **two-sided 95 percent confidence interval for β_1**

  Lower 95 percent confidence limit for β_1 = $\hat{\beta}_1 - t_{0.025}[470]\,\hat{se}(\hat{\beta}_1)$
  = 0.57520 − 1.965(0.058906) = 0.57520 − 0.11575 = **0.45945** = **0.4595** **(1 mark)**

  Upper 95 percent confidence limit for β_1 = $\hat{\beta}_1 + t_{0.025}[470]\,\hat{se}(\hat{\beta}_1)$
  = 0.57520 + 1.965(0.058906) = 0.57520 + 0.11575 = **0.69095** = **0.6910** **(1 mark)**

- Compute **two-sided 99 percent confidence interval for β_1**

  Lower 99 percent confidence limit for β_1 = $\hat{\beta}_1 - t_{0.005}[470]\,\hat{se}(\hat{\beta}_1)$
  = 0.57520 − 2.586(0.058906) = 0.57520 − 0.15233 = **0.42285** = **0.4229** **(1 mark)**

  Upper 99 percent confidence limit for β_1 = $\hat{\beta}_1 + t_{0.005}[470]\,\hat{se}(\hat{\beta}_1)$
  = 0.57520 + 2. 586(0.058906) = 0.57520 + 0.15233 = **0.72755** = **0.7276** **(1 mark)**

- **Decision rule and inference:** **(1 mark)**

  1. If **hypothesized value of β_1** lies ***inside*** the **two-sided 100(1−α) percent confidence interval for β_1**, ***retain*** **(do not reject) H_0** at significance level α.
  2. If **hypothesized value of β_1** lies ***outside*** the **two-sided 100(1−α) percent confidence interval for β_1**, ***reject*** **H_0** at significance level α.

  Inference: **(2 marks)**

  Since 0 lies **outside** the **two-sided 95 percent confidence interval for β_1**, ***reject*** **H_0** at the **5% significance level**.

  Since 0 lies **inside** the **two-sided 99 percent confidence interval for β_1**, ***retain*** **H_0** at the **1% significance level**.

  (1 mark)

  The sample evidence favours the ***alternative*** **hypothesis** that $\beta_1 \neq 0$ **at the 5%** significance level, **but *not* at the 1%** significance level.

(6 marks)

(c) How is the sample value of the test statistic computed in part (b) above related to the value of the F-statistic that *Stata* reports for the OLS sample regression equation corresponding to regression equation (1)?

ANSWER Question 1(c)

- The F-statistic that *Stata* reports for each simple OLS sample regression equation is the ANOVA F-statistic for testing the null hypothesis H_0: $\beta_1 = 0$ against the alternative hypothesis H_1: $\beta_1 \neq 0$. **(2 marks)**

- The **sample value of the ANOVA F-statistic *equals*** the ***square*** of the **sample value of the t-statistic** $t_0(\hat{\beta}_1)$ for testing the null hypothesis H_0: $\beta_1 = 0$ against the alternative hypothesis H_1: $\beta_1 \neq 0$: that is, $\text{ANOVA} - F_0 = \left(t_0(\hat{\beta}_1)\right)^2$.

  For the OLS sample regression equation corresponding to equation (1), ANOVA-F_0 = **95.35** and $t_0(\hat{\beta}_1)$ = **9.7647**, where $\sqrt{95.35} = 9.7647$ or $(9.7647)^2 = 95.35$. **(3 marks)**

- (1) The **p-value of ANOVA-F_0 = the two-tail p-value of** $t_0(\hat{\beta}_1)$ = **0.0000** **(1 mark)**

  or

  (2) The two-tail critical value of the t[N-2] = t[470] distribution is related to the critical value of the F[1, N-2] = F[1, 470] distribution as follows: $\left(t_{\alpha/2}[N-2]\right)^2 = F_\alpha[1, N-2]$.
  For example, at the **5% significance level**: $t_{0.025}[470] = \mathbf{1.965}$; $F_{0.05}[1, 470] = \mathbf{3.861}$; and $1.965^2 = 3.861$

(12 marks)

(d) Use the estimation results from OLS estimation of regression equation (1) to perform a test of the empirical proposition that workers' hourly earnings are positively related to their years of formal education. State the null and alternative hypotheses. Show how the appropriate test statistic is calculated (give its formula). Report the sample value of the test statistic and its p-value. Report the appropriate critical values of the test statistic at the 5 percent and 1 percent significance levels. State the decision rule you use to decide between rejection and retention of the null hypothesis. State the inference you would draw from the test at both the 5 percent and 1 percent significance levels. What would you conclude from the results of the test?

<u>ANSWER Question 1(d)</u> **(total marks = 12)**

Null hypothesis: H_0: $\beta_1 = 0$ (or $\beta_1 \le 0$) **(1 mark)**
Alternative hypothesis: H_1: $\beta_1 > 0$ a ***<u>right-tail</u>*** **t-test** **(2 marks)**

- **Calculation of t-test statistic:** The ***sample value*** $\mathbf{t_0}$ under the null hypothesis H_0 is

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - b_1}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)} = \frac{0.575202}{0.0589061} = \mathbf{\underline{9.76473}} = \mathbf{\underline{9.76}}$$ **(2 marks)**

- ***One-tail right-tail*** **p-value of** $\mathbf{t_0}$ = Pr(t > t_0) = Pr(t > 9.7647) = **<u>0.00000</u>** **(1 mark)**

- ***One-tail right-tail*** **critical values** of the t[N–2] = t[470] distribution are:

$t_\alpha[N-2] = t_{0.05}[38]$ = **<u>1.686</u>** at the **5%** significance level (α = 0.05) **(1 mark)**
$t_\alpha[N-2] = t_{0.01}[38]$ = **<u>2.429</u>** at the **1%** significance level (α = 0.01) **(1 mark)**

- **<u>Decision rule and inference</u>:** ***either formulation 1 or 2*** **(3 marks)**

<u>Decision Rule -- Formulation 1</u>: This a ***right-tail*** **test**. Compare the sample value t_0 with the ***upper*** **α-level critical value of the t[N–2] = t[470] distribution**.

1. If $t_0 \le t_\alpha[N-2]$, ***retain*** **(do not reject) H_0** at the 100α percent significance level.
2. If $t_0 > t_\alpha[N-2]$, ***reject*** **H_0** at the 100α percent significance level.

<u>Inference</u>: Since $t_0(\hat{\beta}_1)$ = **9.76 > 1.648 = $t_{0.05}$[470]**, ***reject*** **H_0** at the ***5 %*** significance level.
Since $t_0(\hat{\beta}_1)$ = **9.76 > 2.334 = $t_{0.01}$[470]**, ***reject*** **H_0** at the ***1 %*** significance level.

<u>Decision Rule -- Formulation 2</u>: This a ***right-tail*** **test**. Compare the ***right-tail*** **p-value for t_0** with the chosen **significance level α**.

1. If ***right-tail*** **p-value** for $t_0 \ge \alpha$, ***retain*** **(do not reject) H_0** at significance level α.
2. If ***right-tail*** **p-value** for $t_0 < \alpha$, ***reject*** **H_0** at significance level α.

<u>Inference</u>:

Since ***right-tail*** **p-value for t_0 = <u>0.00000</u> < 0.05**, ***reject*** **H_0** at the ***5 %*** significance level.
Since ***right-tail*** **p-value for t_0 = <u>0.00000</u> < 0.01**, ***reject*** **H_0** at the ***1 %*** significance level.

<u>Result</u>: **(1 mark)**
The sample evidence favours the ***alternative*** **hypothesis** that **β_1 > 0 at *both* the 5% and the 1%** significance levels; it indicates that individuals' hourly earnings are ***positively related*** to their years of formal education.

(12 marks)

(e) Use the results from OLS estimation of regression equation (1) to perform a test of the hypothesis that the mean hourly wage rate of workers with 16 years of schooling was equal to $8.00 per hour (in 1976 US dollars). State the null and alternative hypotheses. Show how the appropriate test statistic is calculated (give its formula). Report the sample value of the test statistic and its p-value. State the decision rule you use to decide between rejection and retention of the null hypothesis. State the inference you would draw from the test at the 5 percent and the 10 percent significance levels. What would you conclude from the results of the test?

<u>ANSWER Question 1(e)</u>: (total marks = 12)

- **Null hypothesis:** H_0: $\beta_0 + 16\beta_1 = 8$
  Alternative hypothesis: H_1: $\beta_0 + 16\beta_1 \neq 8$ a ***two-tail* test** **(3 marks)**

- **Calculation of t-test statistic:** The ***sample value*** $\mathbf{t_0}$ under the null hypothesis H_0 is

$$t_0(\hat{\beta}_0 + 50\hat{\beta}_1 - 225) = \frac{\hat{\beta}_0 + 16\hat{\beta}_1 - 8}{\hat{se}\left(\hat{\beta}_0 + 16\hat{\beta}_1 - 8\right)} = \frac{-0.109269}{0.253408} = \underline{\mathbf{-0.43120}} = \underline{\mathbf{-0.43}}$$ **(4 marks)**

- ***Two-tail* p-value for $\mathbf{t_0}$ = <u>0.6665</u> = <u>0.667</u>** **(1 mark)**

- **<u>Decision rule and inference</u>: *either formulation 1 or 2*** **(3 marks)**

(1) <u>Decision Rule -- Formulation 1</u>: This a ***two-tail* test**. Compare the sample value t_0 with the **α/2-level critical value of the t[N–2] = t[470] distribution**.

1. If $|t_0| \le t_{\alpha/2}[N-2]$, ***retain* (do not reject) $\mathbf{H_0}$** at the 100α percent significance level.
2. If $|t_0| > t_{\alpha/2}[N-2]$, ***reject* $\mathbf{H_0}$** at the 100α percent significance level.

<u>Inference</u>:
Since $\left|t_0(\hat{\beta}_1)\right|$ = **0.4312 < 1.965 = $\mathbf{t_{0.025}}$[470]**, ***retain* $\mathbf{H_0}$** at the ***5 %*** significance level.
Since $\left|t_0(\hat{\beta}_1)\right|$ = **0.4312 < 1.648 = $\mathbf{t_{0.05}}$[470]**, ***retain* $\mathbf{H_0}$** at the ***10 %*** significance level.

(2) <u>Decision Rule -- Formulation 2</u>: This a ***two-tail* test**. Compare the ***two-tail* p-value for $\mathbf{t_0}$** with the chosen **significance level α**.

1. If ***two-tail* p-value** for $t_0 \ge \alpha$, ***retain* (do not reject) $\mathbf{H_0}$** at significance level α.
2. If ***two-tail* p-value** for $t_0 < \alpha$, ***reject* $\mathbf{H_0}$** at significance level α.

<u>Inference</u>:

Since ***two-tail* p-value for $\mathbf{t_0}$ = 0.6665 > 0.05**, ***retain* $\mathbf{H_0}$** at the ***5 %*** significance level.
Since ***two-tail* p-value for $\mathbf{t_0}$ = 0.6665 > 0.10**, ***retain* $\mathbf{H_0}$** at the ***10 %*** significance level.

- **<u>Result</u>:** **(1 mark)**
  The sample evidence favours the ***null* hypothesis** that $\beta_0 + 16\beta_1 = 8$ **at *both* the 5% and 10%** significance levels.

<u>Alternative ANSWER 1 to Question 1(e): Two-sided confidence interval</u> (total marks = 12)

- **Null hypothesis:** H_0: $\beta_0 + 16\beta_1 = 8$
  Alternative hypothesis: H_1: $\beta_0 + 16\beta_1 \neq 8$ a ***two-tail* test** **(3 marks)**

- **Two-sided 100(1−α) percent confidence interval for β_0 + 16β_1** is:

  Lower 100(1−α) percent confidence limit for β_0 + 16β_1 = $\hat{\beta}_0 + 16\hat{\beta}_1 - t_{\alpha/2}[N-2]\,\hat{se}\left(\hat{\beta}_0 + 16\hat{\beta}_1\right)$

  Upper 100(1−α) percent confidence limit for β_0 + 16β_1 = $\hat{\beta}_0 + 16\hat{\beta}_1 + t_{\alpha/2}[N-2]\,\hat{se}\left(\hat{\beta}_0 + 16\hat{\beta}_1\right)$

- Compute **two-sided 95 percent confidence interval for β_0 + 16β_1**

  Lower 95 percent confidence limit for β_0 + 16β_1 = $\hat{\beta}_0 + 16\hat{\beta}_1 - t_{0.025}[470]\,\hat{se}\left(\hat{\beta}_0 + 16\hat{\beta}_1\right)$
  = 7.89073 − 1.965(0.253408) = 7.89073 − 47.7717 = **<u>7.39278</u>** = **<u>7.393</u>** **(1 mark)**

  Upper 95 percent confidence limit for β_0 + 16β_1 = $\hat{\beta}_0 + 16\hat{\beta}_1 + t_{0.025}[470]\,\hat{se}\left(\hat{\beta}_0 + 16\hat{\beta}_1\right)$
  = 7.89073 + 1.965(0.253408) = 7.89073 + 0.497947 = **<u>8.38868</u>** = **<u>8.389</u>** **(1 mark)**

- Compute **two-sided 90 percent confidence interval for β_0 + 16β_1**

  Lower 90 percent confidence limit for β_0 + 16β_1 = $\hat{\beta}_0 + 16\hat{\beta}_1 - t_{0.05}[470]\,\hat{se}\left(\hat{\beta}_0 + 16\hat{\beta}_1\right)$
  = 7.89073 − 1.648(0.253408) = 7.89073 − 0.417616 = **<u>7.4731</u>** = **<u>7.473</u>** **(1 mark)**

  Upper 90 percent confidence limit for β_0 + 16β_1 = $\hat{\beta}_0 + 16\hat{\beta}_1 + t_{0.05}[470]\,\hat{se}\left(\hat{\beta}_0 + 16\hat{\beta}_1\right)$
  = 7.89073 + 1.648(0.253408) = 7.89073 + 0.417616 = **<u>8.3084</u>** = **<u>8.308</u>** **(1 mark)**

- **<u>Decision rule and inference</u>:** **(2 marks)**

  1. If **hypothesized value of β_0 + 16β_1** lies ***inside*** the **two-sided 100(1−α) percent confidence interval for β_0 + 16β_1**, ***retain* (do not reject) H_0** at significance level α.
  2. If **hypothesized value of β_0 + 16β_1** lies ***outside*** the **two-sided 100(1−α) percent confidence interval for β_0 + 16β_1**, ***reject* H_0** at significance level α.

  <u>Inference</u>: **(2 marks)**

  Since 8 lies **<u>inside</u>** the **two-sided 95 percent confidence interval for β_0 + 16β_1**, ***retain* H_0** at the **5% significance level**.

  Since 8 lies **<u>inside</u>** the **two-sided 90 percent confidence interval for β_0 + 16β_1**, ***retain* H_0** at the **10% significance level**.

  (1 mark)

  The sample evidence favours the ***null* hypothesis** that $\beta_0 + 16\beta_1 = 8$ **at *both* the 5% and 10%** significance levels.

Alternative ANSWER 2 to Question 1(e): An F-test (total marks = 12)

- **Null hypothesis:** H$_0$: $\beta_0 + 16\beta_1 = 8$
  Alternative hypothesis: H$_1$: $\beta_0 + 16\beta_1 \neq 8$ a ***two-tail*** **test** **(3 marks)**

- **Calculation of F-test statistic:** The ***sample value*** $\mathbf{F_0}$ under the null hypothesis H$_0$ is

$$F_0(\hat{\beta}_0 + 16\hat{\beta}_1 - 8) = \frac{(\hat{\beta}_0 + 16\hat{\beta}_1 - 8)^2}{\hat{Var}(\hat{\beta}_0 + 16\hat{\beta}_1 - 8)} = \frac{0.0119397}{0.0642157} = \underline{\mathbf{0.18593}} = \underline{\mathbf{0.19}}$$ **(4 marks)**

- **(right-tail) p-value for $\mathbf{F_0}$ = <u>0.6665</u>** **(1 mark)**

- **<u>Decision rule and inference</u>:** ***either formulation 1 or 2*** **(3 marks)**

(1) <u>Decision Rule -- Formulation 1</u>: This a ***two-tail*** **test**. Compare the sample value F$_0$ with the **α-level critical value of the F[1, N–2] = F[1, 470] distribution**.

1. If $F_0 \le F_\alpha[1, N-2]$, ***retain*** **(do not reject) H$_0$** at the 100α percent significance level.
2. If $F_0 > F_\alpha[1, N-2]$, ***reject*** **H$_0$** at the 100α percent significance level.

<u>Inference</u>:
Since F$_0$ = **0.186 < <u>3.861</u> = $\mathbf{F_{0.05}}$[1, 470]**, ***retain*** **H$_0$** at the ***5 %*** significance level.
Since F$_0$ = **0.186 < <u>2.716</u> = $\mathbf{F_{0.10}}$[1, 470]**, ***retain*** **H$_0$** at the ***10 %*** significance level.

(2) <u>Decision Rule -- Formulation 2</u>: This a ***two-tail*** **test**. Compare the ***two-tail*** **p-value for $\mathbf{t_0}$** with the chosen **significance level α**.

1. If ***two-tail*** **p-value** for $F_0 \ge \alpha$, ***retain*** **(do not reject) H$_0$** at significance level α.
2. If ***two-tail*** **p-value** for $F_0 < \alpha$, ***reject*** **H$_0$** at significance level α.

<u>Inference</u>:

Since ***two-tail*** **p-value for $\mathbf{F_0}$ = 0.6665 > 0.05**, ***retain*** **H$_0$** at the ***5 %*** significance level.
Since ***two-tail*** **p-value for $\mathbf{F_0}$ = 0.6665 > 0.10**, ***retain*** **H$_0$** at the ***10 %*** significance level.

- **<u>Result</u>:** **(1 mark)**
  The sample evidence favours the ***null*** **hypothesis** that $\beta_0 + 16\beta_1 = 8$ **at *both* the 5% and 10%** significance levels.

(50 marks)

2. Compute OLS coefficient estimates of the following population regression equation for the full sample of 472 paid 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.

(10 marks)

(a) Report the OLS coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ computed by estimating population regression equation (1), the estimated standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$, the t-ratios for $\hat{\beta}_0$ and $\hat{\beta}_1$, the estimate $\hat{\sigma}^2$ of the constant error variance σ^2, the R^2 and F-statistic for the OLS sample regression equation, and the number of observations N used in estimation.

ANSWER Question 2(a)

```
. 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
------------------------------------------------------------------------------
```

Regressor	$\hat{\beta}_j$	$\hat{se}(\hat{\beta}_j)$	$t(\hat{\beta}_j)$	Lower 95% limit	Upper 95% limit
Constant	**7.18531**	**0.226903**	**31.67**	**6.73944**	**7.63118**
$FEMALE_i$	**-2.54949**	**0.327188**	**-7.79**	**-3.19242**	**-1.90656**

N = **472**; $\hat{\sigma}^2$ = **12.61379**; R^2 = **0.1144**; F(1, 470) = **60.72**

(6 marks)

(b) Report the value of the coefficient of determination R^2 for the sample regression equation that corresponds to regression equation (2). Explain in words what the value of the R^2 means.

<u>ANSWER Question 2(b)</u>

- $\mathbf{R^2}$ = **<u>0.1144</u>** for the OLS sample regression equation. **(2 marks)**

- The R^2 measures the ***proportion*** **or** ***fraction*** **of the total sample variation of the observed values of the dependent variable $\mathbf{WAGE_i}$** (the proportion of the total sum-of-squares TSS for $WAGE_i$) that is **explained by (1) the binary gender variable $\mathbf{FEMALE_i}$ *or* (2) the estimated OLS sample regression function**, i.e., by

$$\hat{WAGE}_i = \hat{\beta}_0 + \hat{\beta}_1 FEMALE_i = 7.1853 - 2.5495 FEMALE_i$$ **(1 mark)**

- The value $\mathbf{R^2 = 0.1144}$ thus means that (1) **the binary gender variable FEMALE** or (2) **the OLS sample regression function** explains **11.44 percent of the total sample variation of hourly wage rates, or 11.44 percent of the total sum-of-squares of the observed $\mathbf{WAGE_i}$ values.** **(3 marks)**

(10 marks)

(c) Use the estimation results from OLS estimation of regression equation (2) to perform a t-test of the empirical proposition that gender was unrelated to individuals' hourly wage rates in 1976, i.e., that there was no difference in hourly earnings between female and male workers in 1976. State the null and alternative hypotheses. Show how the appropriate test statistic is calculated (give its formula). Report the sample value of the test statistic and its p-value. State the decision rule you use to decide between rejection and retention of the null hypothesis. State the inference you would draw from the test at the 5 percent and 1 percent significance levels. What conclusion would you draw from the test?

ANSWER Question 2(c) **(total marks = 10)**

- **Null hypothesis:** H_0: $\beta_1 = 0$
  Alternative hypothesis: H_1: $\beta_1 \neq 0$ a ***two-tail*** **test** **(2 marks)**

- **Calculation of t-test statistic:** The ***sample value*** $\mathbf{t_0}$ under the null hypothesis H_0 is

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - b_1}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)} = \frac{-2.54949}{0.327188} = \underline{-7.79212} = \underline{-7.79}$$ **(3 marks)**

- ***Two-tail*** **p-value for $\mathbf{t_0}$ = 0.00000 = 0.000** **(1 mark)**

- **Decision rule and inference**: ***either formulation 1 or 2*** **(3 marks)**

(1) Decision Rule -- Formulation 1: This a ***two-tail*** **test**. Compare the sample value t_0 with the **α/2-level critical value of the t[N–2] = t[470] distribution**.

1. If $|t_0| \le t_{\alpha/2}[N-2]$, ***retain*** **(do not reject) $\mathbf{H_0}$** at the 100α percent significance level.
2. If $|t_0| > t_{\alpha/2}[N-2]$, ***reject*** **$\mathbf{H_0}$** at the 100$\alpha$ percent significance level.

Inference:

Since $|t_0(\hat{\beta}_1)|$ = **7.79 > 1.965 = $\mathbf{t_{0.025}}$[470]**, ***reject*** **$\mathbf{H_0}$** at the ***5 %*** significance level.

Since $|t_0(\hat{\beta}_1)|$ = **7.79 > 2.586 = $\mathbf{t_{0.005}}$[470]**, ***reject*** **$\mathbf{H_0}$** at the ***1 %*** significance level.

(2) Decision Rule -- Formulation 2: This a ***right-tail*** **test**. Compare the ***right-tail*** **p-value for $\mathbf{t_0}$** with the chosen **significance level** α.

1. If ***two-tail*** **p-value** for $t_0 \ge \alpha$, ***retain*** **(do not reject) $\mathbf{H_0}$** at significance level α.
2. If ***two-tail*** **p-value** for $t_0 < \alpha$, ***reject*** **$\mathbf{H_0}$** at significance level α.

Inference:

Since ***two-tail*** **p-value for $\mathbf{t_0}$ = 0.000 < 0.05**, ***reject*** **$\mathbf{H_0}$** at the ***5 %*** significance level.
Since ***two-tail*** **p-value for $\mathbf{t_0}$ = 0.000 < 0.01**, ***reject*** **$\mathbf{H_0}$** at the ***1 %*** significance level.

- **Result:** **(1 mark)**
  The sample evidence favours the alternative hypothesis that $\beta_1 \neq 0$ **at *both* the 5% and 1%** significance levels; it indicates that workers' average hourly earnings are related to gender.

<u>Alternative ANSWER Question 2(c)</u>: a two-sided confidence interval for β_1 (total marks = 12)

- **Null hypothesis:** H_0: $\beta_1 = 0$
  Alternative hypothesis: H_1: $\beta_1 \neq 0$ a ***two-tail*** **test** **(2 marks)**

- **Two-sided 100(1−α) percent confidence interval for β_1** is:

  Lower 100(1−α) percent confidence limit for β_1 = $\hat{\beta}_1 - t_{\alpha/2}[N-2]\,\hat{se}\left(\hat{\beta}_1\right)$
  Upper 100(1−α) percent confidence limit for β_1 = $\hat{\beta}_1 + t_{\alpha/2}[N-2]\,\hat{se}\left(\hat{\beta}_1\right)$

- Compute **two-sided 95 percent confidence interval for β_1**

  Lower 95 percent confidence limit for β_1 = $\hat{\beta}_1 - t_{0.025}[470]\,\hat{se}\left(\hat{\beta}_1\right)$
  = – 2.54949 – 1.965(0.327188) = – 2.54949 – 0.642924 = **<u>– 3.19242</u>** = **<u>– 3.192</u>** **(1 mark)**

  Upper 95 percent confidence limit for β_1 = $\hat{\beta}_1 + t_{0.025}[470]\,\hat{se}\left(\hat{\beta}_1\right)$
  = – 2.54949 + 1.965(0.327188) = – 2.54949 + 0.642924 = **<u>– 1.90656</u>** = **<u>– 1.907</u>** **(1 mark)**

- Compute **two-sided 99 percent confidence interval for β_1**

  Lower 99 percent confidence limit for β_1 = $\hat{\beta}_1 - t_{0.005}[470]\,\hat{se}\left(\hat{\beta}_1\right)$
  = – 2.54949 – 2.586(0.327188) = – 2.54949 – 0.846108 = **<u>– 3.39560</u>** = **<u>– 3.396</u>** **(1 mark)**

  Upper 99 percent confidence limit for β_1 = $\hat{\beta}_1 + t_{0.005}[470]\,\hat{se}\left(\hat{\beta}_1\right)$
  = – 2.54949 + 2.586(0.327188) = – 2.54949 + 0.846108 = **<u>– 1.70338</u>** = **<u>– 1.703</u>** **(1 mark)**

- **<u>Decision rule and inference</u>:** **(1 mark)**

  1. If **hypothesized value of β_1** lies ***inside*** the **two-sided 100(1−α) percent confidence interval for β_1**, ***retain*** **(do not reject) H_0** at significance level α.
  2. If **hypothesized value of β_1** lies ***outside*** the **two-sided 100(1−α) percent confidence interval for β_1**, ***reject*** **H_0** at significance level α.

  <u>Inference</u>: **(2 marks)**

  Since 0 lies **<u>outside</u>** the **two-sided 95 percent confidence interval for β_1**, ***reject*** **H_0** at the **5% significance level**.

  Since 0 lies **<u>outside</u>** the **two-sided 99 percent confidence interval for β_1**, ***reject*** **H_0** at the **1% significance level**.

  (1 mark)

  The sample evidence favours the ***alternative*** **hypothesis** that $\beta_1 \neq 0$ **at *both* the 5% and 1%** significance levels; it indicates that workers' average hourly earnings are related to gender.

(12 marks)

(d) Use the OLS estimation results for regression equation (2) to test the proposition that females' mean hourly wage rate was less than males' mean hourly wage rate in 1976. State the null and alternative hypotheses. Show how the appropriate test statistic is calculated (give its formula). Report the sample value of the test statistic and its p-value. Report the appropriate critical values of the test statistic at the 5 percent and 1 percent significance levels. State the decision rule you use to decide between rejection and retention of the null hypothesis. State the inference you would draw from the test at both the 5 percent and 1 percent significance levels. What would you conclude from the results of the test?

ANSWER Question 2(d) **(total marks = 12)**

Null hypothesis: H_0: $\beta_1 = 0$ (or $\beta_1 \geq 0$) **(1 mark)**
Alternative hypothesis: H_1: $\beta_1 < 0$ a ***left-tail*** **t-test** **(2 marks)**

- **Calculation of t-test statistic:** The ***sample value*** $\mathbf{t_0}$ under the null hypothesis H_0 is

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - b_1}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)} = \frac{-2.54949}{0.327188} = \underline{\mathbf{-7.79212}} = \underline{\mathbf{-7.79}}$$ **(2 marks)**

- ***One-tail left-tail*** **p-value of** $\mathbf{t_0}$ = Pr(t < $-t_0$) = Pr(t < –7.792) = **0.00000** = **0.0000** **(1 mark)**

- ***One-tail left-tail*** **critical values** of the t[N–2] = t[470] distribution are:

$-t_{\alpha}[N-2] = -t_{0.05}[470] = \underline{\mathbf{-1.648}}$ at the **5%** significance level ($\alpha = 0.05$) **(1 mark)**
$-t_{\alpha}[N-2] = -t_{0.01}[470] = \underline{\mathbf{-2.334}}$ at the **1%** significance level ($\alpha = 0.01$) **(1 mark)**

- **Decision rule and inference**: ***either formulation 1 or 2*** **(3 marks)**

(1) Decision Rule -- Formulation 1: This a ***left-tail*** **test**. Compare the sample value t_0 with the ***lower*** **α-level critical value of the t[N–2] = t[470] distribution**.

1. If $t_0 < -t_{\alpha}[N-2]$, ***reject*** $\mathbf{H_0}$ at the 100α percent significance level.
2. If $t_0 \geq -t_{\alpha}[N-2]$, ***retain*** **(do not reject)** $\mathbf{H_0}$ at the 100α percent significance level.

Inference: Since $t_0(\hat{\beta}_1)$ = **–7.792 < –1.648 = $-t_{0.05}$[470]**, ***reject*** $\mathbf{H_0}$ at the ***5 %*** significance level.

Since $t_0(\hat{\beta}_1)$ = **–7.792 < –2.334 = $-t_{0.01}$[470]**, ***reject*** $\mathbf{H_0}$ at the ***1 %*** significance level.

(2) Decision Rule -- Formulation 2: This a ***left-tail*** **test**. Compare the ***left-tail*** **p-value for t_0** with the chosen **significance level α**.

1. If ***left-tail*** **p-value** for $t_0 \geq \alpha$, ***retain*** **(do not reject)** $\mathbf{H_0}$ at significance level α.
2. If ***left-tail*** **p-value** for $t_0 < \alpha$, ***reject*** $\mathbf{H_0}$ at significance level α.

Inference:

Since ***left-tail*** **p-value for t_0 = 0.0000 < 0.05**, ***reject*** $\mathbf{H_0}$ at the ***5 %*** significance level.
Since ***left-tail*** **p-value for t_0 = 0.0000 < 0.01**, ***reject*** $\mathbf{H_0}$ at the ***1 %*** significance level.

Result: The sample evidence favours the ***alternative*** **hypothesis** that **β_1 < 0 at *both* the 5% and 1%**, significance levels; it indicates that females' mean hourly earnings are ***less than*** males' mean hourly earnings. **(1 mark)**

(12 marks)

(e) Use the results from OLS estimation of regression equation (2) to test the hypothesis that the mean hourly wage rate of female workers in 1976 was equal to $4.00 per hour (in 1976 US dollars). State the null and alternative hypotheses. Show how the appropriate test statistic is calculated (give its formula). Report the sample value of the test statistic and its p-value. State the decision rule you use to decide between rejection and retention of the null hypothesis. State the inference you would draw from the test at the 5 percent and the 1 percent significance levels. What conclusion would you draw from the results of the test?

<u>ANSWER Question 2(e):</u> (total marks = 12)

- **Null hypothesis:** H_0: $\beta_0 + \beta_1 = 4$
  Alternative hypothesis: H_1: $\beta_0 + \beta_1 \neq 4$ a ***two-tail*** **test** **(3 marks)**

- **Calculation of t-test statistic:** The ***sample value*** $\mathbf{t_0}$ under the null hypothesis H_0 is

$$t_0(\hat\beta_0 + \hat\beta_1 - 4) = \frac{\hat\beta_0 + \hat\beta_1 - 4}{\hat{se}(\hat\beta_0 + \hat\beta_1)} = \frac{\hat\beta_0 + \hat\beta_1 - 4}{\hat{se}(\hat\beta_0 + \hat\beta_1)} = \frac{0.635815}{0.235727} = \underline{\mathbf{2.69725}} = \underline{\mathbf{2.70}}$$ **(4 marks)**

- ***Two-tail*** **p-value for $\mathbf{t_0}$ = <u>0.0072428</u> = <u>0.0072</u>** **(1 mark)**

- **<u>Decision rule and inference</u>:** ***either formulation 1 or 2*** **(3 marks)**

(1) <u>Decision Rule -- Formulation 1</u>: This a ***two-tail*** **test**. Compare the sample value t_0 with the **α/2-level critical value of the t[N–2] = t[470] distribution**.

1. If $|t_0| \le t_{\alpha/2}[N-2]$, ***retain*** **(do not reject) $\mathbf{H_0}$** at the 100α percent significance level.
2. If $|t_0| > t_{\alpha/2}[N-2]$, ***reject*** **$\mathbf{H_0}$** at the 100α percent significance level.

<u>Inference</u>:
Since $\left|t_0(\hat\beta_0 + \hat\beta_1 - 200)\right|$ = **2.70 > 1.965 = $\mathbf{t_{0.025}}$[470]**, ***reject*** **$\mathbf{H_0}$** at the ***5 %*** significance level.
Since $\left|t_0(\hat\beta_0 + \hat\beta_1 - 200)\right|$ = **2.70 > 2.586 = $\mathbf{t_{0.005}}$[470]**, ***reject*** **$\mathbf{H_0}$** at the ***1 %*** significance level.

(2) <u>Decision Rule -- Formulation 2</u>: This a ***two-tail*** **test**. Compare the ***two-tail*** **p-value for $\mathbf{t_0}$** with the chosen **significance level α**.

1. If ***two-tail*** **p-value** for $t_0 \ge \alpha$, ***retain*** **(do not reject) $\mathbf{H_0}$** at significance level α.
2. If ***two-tail*** **p-value** for $t_0 < \alpha$, ***reject*** **$\mathbf{H_0}$** at significance level α.

<u>Inference</u>:

Since ***two-tail*** **p-value for $\mathbf{t_0}$ = 0.0072 < 0.05**, ***reject*** **$\mathbf{H_0}$** at the ***5 %*** significance level.
Since ***two-tail*** **p-value for $\mathbf{t_0}$ = 0.0072 < 0.01**, ***reject*** **$\mathbf{H_0}$** at the ***1 %*** significance level.

- **<u>Result</u>:** **(1 mark)**
  The sample evidence favours the ***alternative*** **hypothesis** that **$\boldsymbol{\beta_1} \neq$ 4 at *both* the 5% and the 1%** significance levels; it indicates that females' mean hourly wage rate was ***not equal*** **to $4.00 per hour** in 1976.

Alternative ANSWER 1 to Question 2(e): **two-sided confidence intervals (total marks = 12)**

- **Null hypothesis:** H$_0$: $\beta_0 + \beta_1 = 4$
  Alternative hypothesis: H$_1$: $\beta_0 + \beta_1 \neq 4$ a ***two-tail*** **test** **(3 marks)**

- **Two-sided 100(1–α) percent confidence interval for β_1** is:

  Lower 100(1–α) percent confidence limit for $\beta_0 + \beta_1$ = $\hat{\beta}_0 + \hat{\beta}_1 - t_{\alpha/2}[N-2]\,\hat{se}\left(\hat{\beta}_0 + \hat{\beta}_1\right)$

  Upper 100(1–α) percent confidence limit for $\beta_0 + \beta_1$ = $\hat{\beta}_0 + \hat{\beta}_1 + t_{\alpha/2}[N-2]\,\hat{se}\left(\hat{\beta}_0 + \hat{\beta}_1\right)$

- Compute **two-sided 95 percent confidence interval for $\beta_0 + \beta_1$**

  Lower 95 percent confidence limit for $\beta_0 + \beta_1$ = $\hat{\beta}_0 + \hat{\beta}_1 - t_{0.025}[470]\,\hat{se}\left(\hat{\beta}_0 + \hat{\beta}_1\right)$
  = 4.63582 – 1.965(0.235727) = 4.63582 – 0.463204 = **4.17261** = **4.173** **(1 mark)**

  Upper 95 percent confidence limit for $\beta_0 + \beta_1$ = $\hat{\beta}_0 + \hat{\beta}_1 + t_{0.025}[470]\,\hat{se}\left(\hat{\beta}_0 + \hat{\beta}_1\right)$
  = 4.63582 + 1.965(0.235727) = 4.63582 + 0.463204 = **5.09903** = **5.099** **(1 mark)**

- Compute **two-sided 99 percent confidence interval for $\beta_0 + \beta_1$**

  Lower 99 percent confidence limit for $\beta_0 + \beta_1$ = $\hat{\beta}_0 + \hat{\beta}_1 - t_{0.005}[470]\,\hat{se}\left(\hat{\beta}_0 + \hat{\beta}_1\right)$
  = 4.63582 – 2.586(0.235727) = 4.63582 – 0.609590 = **4.0262** = **4.026** **(1 mark)**

  Upper 99 percent confidence limit for $\beta_0 + \beta_1$ = $\hat{\beta}_0 + \hat{\beta}_1 + t_{0.005}[470]\,\hat{se}\left(\hat{\beta}_0 + \hat{\beta}_1\right)$
  = 4.63582 + 2.586(0.235727) = 4.63582 + 0.609590 = **5.24548** = **5.245** **(1 mark)**

- **Decision rule and inference:** **(2 marks)**

  1. If **hypothesized value of $\beta_0 + \beta_1$** lies ***inside*** the **two-sided 100(1–α) percent confidence interval for $\beta_0 + \beta_1$**, ***retain*** **(do not reject) H$_0$** at significance level α.
  2. If **hypothesized value of $\beta_0 + \beta_1$** lies ***outside*** the **two-sided 100(1–α) percent confidence interval for $\beta_0 + \beta_1$**, ***reject*** **H$_0$** at significance level α.

  Inference: **(2 marks)**

  Since 4 lies **outside** the **two-sided 95 percent confidence interval for $\beta_0 + \beta_1$**, ***reject*** **H$_0$** at the **5% significance level**.

  Since 4 lies **outside** the **two-sided 99 percent confidence interval for $\beta_0 + \beta_1$**, ***reject*** **H$_0$** at the **1% significance level**.

  The sample evidence favours the ***alternative*** **hypothesis** that $\beta_0 + \beta_1 \neq 200$ **at *both* the 5% and 1%** significance levels; it indicates that females' mean hourly wage rate was ***not equal*** **to $4.00 per hour** in 1976. **(1 mark)**

Alternative ANSWER 2 to Question 2(e): An F-test (total marks = 12)

- **Null hypothesis:** H$_0$: $\beta_0 + \beta_1 = 4$
  Alternative hypothesis: H$_1$: $\beta_0 + \beta_1 \neq 4$ a ***two-tail*** **F-test** **(3 marks)**

- **Calculation of F-test statistic:** The ***sample value*** $\mathbf{F_0}$ under the null hypothesis H$_0$ is

$$F_0(\hat{\beta}_0 + \hat{\beta}_1 - 4) = \frac{(\hat{\beta}_0 + \hat{\beta}_1 - 4)^2}{\hat{Var}(\hat{\beta}_0 + \hat{\beta}_1 - 4)} = \frac{(\hat{\beta}_0 + \hat{\beta}_1 - 4)^2}{\hat{Var}(\hat{\beta}_0 + \hat{\beta}_1)} = \frac{0.4042607}{0.05556735} = \underline{\mathbf{7.27515}} = \underline{\mathbf{7.28}}$$

(4 marks)

- **(right-tail) p-value for $\mathbf{F_0}$ = 0.007243 = 0.0072** **(1 mark)**

- **Decision rule and inference**: ***either formulation 1 or 2*** **(3 marks)**

(1) Decision Rule -- Formulation 1: This a ***two-tail*** **test**. Compare the sample value F$_0$ with the **α-level critical value of the F[1, N–2] = F[1, 470] distribution**.

1. If $F_0 \le F_\alpha[1, N-2]$, ***retain*** **(do not reject) H$_0$** at the 100α percent significance level.
2. If $F_0 > F_\alpha[1, N-2]$, ***reject*** **H$_0$** at the 100α percent significance level.

Inference:
Since F$_0$ = **7.28 > 3.861 = $\mathbf{F_{0.05}}$[1, 470]**, ***reject*** **H$_0$** at the ***5 %*** significance level.
Since F$_0$ = **7.28 > 6.689 = $\mathbf{F_{0.01}}$[1, 470]**, ***reject*** **H$_0$** at the ***1 %*** significance level.

(2) Decision Rule -- Formulation 2: This a ***two-tail*** **test**. Compare the **p-value for $\mathbf{F_0}$** with the chosen **significance level α**.

1. If **p-value** for $F_0 \ge \alpha$, ***retain*** **(do not reject) H$_0$** at significance level α.
2. If **p-value** for $F_0 < \alpha$, ***reject*** **H$_0$** at significance level α.

Inference:

Since **p-value for $\mathbf{F_0}$ = 0.0072 < 0.05**, ***reject*** **H$_0$** at the ***5 %*** significance level.
Since **p-value for $\mathbf{F_0}$ = 0.0072 < 0.01**, ***reject*** **H$_0$** at the ***1 %*** significance level.

- **Result:** **(1 mark)**
  The sample evidence favours the ***alternative*** **hypothesis** that $\beta_0 + \beta_1 \neq 4$ **at *both* the 5% and 1%** significance levels; it indicates that females' mean hourly wage rate was ***not equal*** **to $200 per hour** in 1976.

Output of Stata 'regress' command for Questions 1 and 2:

```
. *
. * Question 1: be = b0 + b1*income + u
. *
. * Question 1(a):
. *
. regress be income

      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  1,    38) =    5.95
       Model |  128366.057     1  128366.057           Prob > F      =  0.0195
    Residual |  819285.843    38  21560.1538           R-squared     =  0.1355
-------------+------------------------------           Adj R-squared =  0.1127
       Total |    947651.9    39  24298.7667           Root MSE      =  146.83

------------------------------------------------------------------------------
          be |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   1.457651    .597385     2.44   0.019     .2483081    2.666994
       _cons |   128.9803   34.59125     3.73   0.001     58.95401    199.0067
------------------------------------------------------------------------------


.
. *
. * Question 2: be = b0 + b1*female + u
. *
. * Question 2(a):
. *
. regress be female

      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  1,    38) =   21.01
       Model |  337400.872     1  337400.872           Prob > F      =  0.0000
    Residual |  610251.028    38  16059.2376           R-squared     =  0.3560
-------------+------------------------------           Adj R-squared =  0.3391
       Total |    947651.9    39  24298.7667           Root MSE      =  126.73

------------------------------------------------------------------------------
          be |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -183.9148   40.12417    -4.58   0.000    -265.1419   -102.6877
       _cons |   288.1053   29.07272     9.91   0.000     229.2506    346.9599
```