QUEEN'S UNIVERSITY AT KINGSTON
Department of Economics

ECONOMICS 351* - Fall Term 2003

Introductory Econometrics

Fall Term 2003 **MID-TERM EXAM: ANSWERS** M.G. Abbott

DATE: **Tuesday October 28, 2003.**

TIME: **80 minutes; 1:00 p.m. - 2:20 p.m.**

INSTRUCTIONS: The exam consists of **FIVE (5)** questions. Students are required to answer **ALL FIVE (5)** questions.
Answer all questions in the exam booklets provided. Be sure your ***name*** and ***student number*** are printed clearly on the front of all exam booklets used.
Do not write answers to questions on the front page of the first exam booklet.
Please label clearly each of your answers in the exam booklets with the appropriate number and letter.
Please write legibly.

MARKING: The marks for each question are indicated in parentheses immediately above each question. **Total marks** for the exam **equal 100**.

GOOD LUCK!

All questions pertain to the simple (two-variable) linear regression model for which the population regression equation can be written in conventional notation as:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(1)}$$

where Y_i and X_i are observable variables, β_1 and β_2 are unknown (constant) regression coefficients, and u_i is an unobservable random error term. The Ordinary Least Squares (OLS) sample regression equation corresponding to regression equation (1) is

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(2)}$$

where $\hat{\beta}_1$ is the OLS estimator of the intercept coefficient β_1, $\hat{\beta}_2$ is the OLS estimator of the slope coefficient β_2, $\hat{u}_i$ is the OLS residual for the i-th sample observation, $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ is the OLS estimated value of Y for the i-th sample observation, and N is sample size (the number of observations in the sample).

QUESTIONS: **Answer ALL FIVE questions.**

(15 marks)

1. State the Ordinary Least Squares (OLS) estimation criterion. State the OLS normal equations. Derive the OLS normal equations from the OLS estimation criterion.

ANSWER:

(3 marks)

- State the Ordinary Least Squares (OLS) estimation criterion. **(3 marks)**

The OLS coefficient estimators are **those formulas or expressions for $\hat{\beta}_1$ and $\hat{\beta}_2$ that** **minimize the sum of squared residuals RSS** for any given sample of size N.

The **OLS estimation criterion** is therefore:

$$\underset{\{\hat{\beta}_j\}}{\text{Minimize}}\ \text{RSS}\left(\hat{\beta}_1, \hat{\beta}_2\right) = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2$$

(4 marks)

- State the OLS normal equations. **(4 marks)**

The ***first*** **OLS normal equation** can be written in *any one* of the following forms:

$$\Sigma_i\, Y_i - N\hat{\beta}_1 - \hat{\beta}_2 \Sigma_i\, X_i = 0$$

$$-N\hat{\beta}_1 - \hat{\beta}_2 \Sigma_i\, X_i = -\Sigma_i\, Y_i \qquad \textbf{(N1)}$$

$$N\hat{\beta}_1 + \hat{\beta}_2 \Sigma_i\, X_i = \Sigma_i\, Y_i$$

The ***second*** **OLS normal equation** can be written in *any one* of the following forms:

$$\Sigma_i\, X_i Y_i - \hat{\beta}_1 \Sigma_i\, X_i - \hat{\beta}_2 \Sigma_i\, X_i^2 = 0$$

$$-\hat{\beta}_1 \Sigma_i\, X_i - \hat{\beta}_2 \Sigma_i\, X_i^2 = -\Sigma_i\, X_i Y_i \qquad \textbf{(N2)}$$

$$\hat{\beta}_1 \Sigma_i\, X_i + \hat{\beta}_2 \Sigma_i\, X_i^2 = \Sigma_i\, X_i Y_i$$

Question 1 (continued)

(8 marks)

- Show how the OLS normal equations are derived from the OLS estimation criterion.

(4 marks)

<u>**Step 1**</u>: **Partially differentiate** the RSS $\left(\hat\beta_1, \hat\beta_2\right)$ function with respect to $\hat\beta_1$ and $\hat\beta_2$, using

$$\hat u_i = Y_i - \hat\beta_1 - \hat\beta_2 X_i \quad \Rightarrow \quad \frac{\partial \hat u_i}{\partial \hat\beta_1} = -1 \quad \text{and} \quad \frac{\partial \hat u_i}{\partial \hat\beta_2} = -X_i\,.$$

$$\frac{\partial RSS}{\partial \hat\beta_1} = \sum_{i=1}^{N} 2\hat u_i \left(\frac{\partial \hat u_i}{\partial \hat\beta_1}\right) = \sum_{i=1}^{N} 2\,\hat u_i(-1) = -2\sum_{i=1}^{N}\hat u_i = -2\sum_{i=1}^{N}\left(Y_i - \hat\beta_1 - \hat\beta_2 X_i\right) \qquad (1)$$

$$\frac{\partial RSS}{\partial \hat\beta_2} = \sum_{i=1}^{N} 2\hat u_i \left(\frac{\partial \hat u_i}{\partial \hat\beta_2}\right) = \sum_{i=1}^{N} 2\hat u_i(-X_i) = -2\sum_{i=1}^{N} X_i \hat u_i$$

$$= -2\sum_{i=1}^{N} X_i\left(Y_i - \hat\beta_1 - \hat\beta_2 X_i\right) \quad \text{since } \hat u_i = Y_i - \hat\beta_1 - \hat\beta_2 X_i \qquad (2)$$

$$= -2\sum_{i=1}^{N}\left(X_i Y_i - \hat\beta_1 X_i - \hat\beta_2 X_i^2\right).$$

(4 marks)

<u>**Step 2**</u>: **Obtain the first-order conditions (FOCs)** for a minimum of the RSS function by **setting the partial derivatives (1) and (2) *equal to zero*** and then dividing each equation by –2 and re-arranging:

$$\frac{\partial RSS}{\partial \hat\beta_1} = 0 \quad \Rightarrow \quad -2\,\Sigma_i\,\hat u_i = 0 \quad \Rightarrow \quad -2\,\Sigma_i\left(Y_i - \hat\beta_1 - \hat\beta_2 X_i\right) = 0$$

$$\Rightarrow \quad \Sigma_i\left(Y_i - \hat\beta_1 - \hat\beta_2 X_i\right) = 0$$

$$\Rightarrow \quad \Sigma_i\,Y_i - N\hat\beta_1 - \hat\beta_2\,\Sigma_i\,X_i = 0$$

$$\Rightarrow \quad \Sigma_i\,Y_i = N\hat\beta_1 + \hat\beta_2\,\Sigma_i\,X_i \qquad \textbf{(N1)}$$

$$\frac{\partial RSS}{\partial \hat\beta_2} = 0 \quad \Rightarrow \quad -2\,\Sigma_i\,X_i\hat u_i = 0 \quad \Rightarrow \quad -2\,\Sigma_i\,X_i\left(Y_i - \hat\beta_1 - \hat\beta_2 X_i\right) = 0$$

$$\Rightarrow \quad \Sigma_i\,X_i\left(Y_i - \hat\beta_1 - \hat\beta_2 X_i\right) = 0$$

$$\Rightarrow \quad \Sigma_i\left(X_i Y_i - \hat\beta_1 X_i - \hat\beta_2 X_i^2\right) = 0$$

$$\Rightarrow \quad \Sigma_i\,X_i Y_i - \hat\beta_1\,\Sigma_i\,X_i - \hat\beta_2\,\Sigma_i\,X_i^2 = 0$$

$$\Rightarrow \quad \Sigma_i\,X_i Y_i = \hat\beta_1\,\Sigma_i\,X_i + \hat\beta_2\,\Sigma_i\,X_i^2 \qquad \textbf{(N2)}$$

(15 marks)

2. Give a general definition of the F-distribution. Starting from this definition, derive the F-statistic for the OLS slope coefficient estimator $\hat{\beta}_2$. State all assumptions required for the derivation.

ANSWER:

(3 marks)

- ***General definition of the F-distribution:*** Consider the two random variables V_1 and V_2 such that

(1) $V_1 \sim \chi^2[m_1]$
(2) $V_2 \sim \chi^2[m_2]$
(3) V_1 and V_2 are *independent*.

Then the random variable

$$F = \frac{V_1/m_1}{V_2/m_2} \sim F[m_1, m_2]$$

where $F[m_1, m_2]$ denotes the **F-distribution with m_1 *numerator* degrees of freedom and m_2 *denominator* degrees of freedom**.

(2 marks)

- ***Error Normality Assumption:*** The random error term u_i is normally distributed with mean 0 and variance σ^2; that is,

$$u_i \mid X_i \sim N\left[0, \sigma^2\right] \text{ for all i} \quad \text{OR} \quad u_i \text{ is iid as } N\left[0, \sigma^2\right]$$

(4 marks)

- Use **three implications** of the error normality assumption to derive $F(\hat{\beta}_2)$:

1. $\hat{\beta}_2 \sim N\left[\beta_2, \text{Var}(\hat{\beta}_2)\right]$ where $\text{Var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum_{i=1}^{N} x_i^2} = \dfrac{\sigma^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$ **(1 mark)**

Implications of normality of $\hat{\beta}_2$:

$$Z(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \sim N[0, 1]$$

$$[Z(\hat{\beta}_2)]^2 = \frac{(\hat{\beta}_2 - \beta_2)^2}{\text{Var}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)^2}{\sigma^2 / \sum_i x_i^2} = \frac{(\hat{\beta}_2 - \beta_2)^2 \sum_i x_i^2}{\sigma^2} \sim \chi^2[1] \qquad (1)$$ **(1 mark)**

2. $\dfrac{(N-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2[N-2] \;\Rightarrow\; \dfrac{\hat{\sigma}^2}{\sigma^2} \sim \dfrac{\chi^2[N-2]}{(N-2)}$ (2) **(1 mark)**

3. The estimators $\hat{\beta}_2$ and $\hat{\sigma}^2$ are statistically independent
OR **(1 mark)**
The statistics $Z(\hat{\beta}_2)$ and $\hat{\sigma}^2/\sigma^2$ are statistically independent.

(6 marks)

- **The *F-statistic* for** $\hat{\beta}_2$. The F-statistic for $\hat{\beta}_2$ is therefore the ratio of (1) to (2):

$$F(\hat{\beta}_2) = \frac{\left(Z(\hat{\beta}_2)\right)^2}{\hat{\sigma}^2/\sigma^2}$$

$$= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2\left(\sum_i x_i^2\right)/\sigma^2}{\hat{\sigma}^2/\sigma^2}$$

$$= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2\left(\sum_i x_i^2\right)}{\hat{\sigma}^2}$$

$$= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\sigma}^2/\sum_i x_i^2}$$

$$= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{Var}(\hat{\beta}_2)} \qquad \text{since } \hat{\sigma}^2/\textstyle\sum_i x_i^2 = \hat{Var}(\hat{\beta}_2).$$

- ***Result:*** The **F-statistic for** $\hat{\beta}_2$ takes the form

$$F(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\sigma}^2/\left(\sum_i x_i^2\right)} = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{Var}(\hat{\beta}_2)} \sim F[1, N-2].$$

(10 marks)

3. Answer both parts (a) and (b) below. H_0 stands for the null hypothesis of a statistical test. For each of parts (a) and (b), select which of statements (1) to (4) best defines the concept in question.

<u>ANSWER</u>: Correct answers are highlighted in bold.

(5 marks)

(a) The ***significance level*** of a hypothesis test is best defined as:

(1) the probability of retaining H_0 when H_0 is true
(2) the probability of rejecting H_0 when H_0 is true ←
(3) the probability of retaining H_0 when H_0 is false
(4) the probability of rejecting H_0 when H_0 is false

(5 marks)

(b) The ***power*** of a hypothesis test is best defined as:

(1) the probability of retaining H_0 when H_0 is true
(2) the probability of rejecting H_0 when H_0 is true
(3) the probability of retaining H_0 when H_0 is false
(4) the probability of rejecting H_0 when H_0 is false ←

(36 marks)

4. A researcher is using data for a sample of 274 male employees to investigate the relationship between hourly wage rates Y_i (measured in *dollars per hour*) and firm tenure X_i (measured in *years*). Preliminary analysis of the sample data produces the following sample information:

$$N = 274 \qquad \sum_{i=1}^{N} Y_i = 1945.26 \qquad \sum_{i=1}^{N} X_i = 1774.00 \qquad \sum_{i=1}^{N} Y_i^2 = 18536.73$$

$$\sum_{i=1}^{N} X_i^2 = 30608.00 \qquad \sum_{i=1}^{N} X_i Y_i = 16040.72 \qquad \sum_{i=1}^{N} x_i y_i = 3446.226$$

$$\sum_{i=1}^{N} y_i^2 = 4726.377 \qquad \sum_{i=1}^{N} x_i^2 = 19122.32 \qquad \sum_{i=1}^{N} \hat{u}_i^2 = 4105.297$$

where $x_i \equiv X_i - \bar{X}$ and $y_i \equiv Y_i - \bar{Y}$ for i = 1, ..., N. Use the above sample information to answer all the following questions. **Show explicitly all formulas and calculations**.

(10 marks)

(a) Use the above information to compute OLS estimates of the intercept coefficient β_1 and the slope coefficient β_2.

- $\hat{\beta}_2 = \dfrac{\sum_i x_i y_i}{\sum_i x_i^2} = \dfrac{3446.226}{19{,}122.32} = \mathbf{0.1802201} = \underline{\mathbf{0.18022}}$ **(5 marks)**

- $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

$$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{1945.26}{274} = 7.09949 \quad \text{and} \quad \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{1774.00}{274} = 6.47445$$

Therefore

$$\hat{\beta}_1 = \bar{Y} - \beta_2 \bar{X} = 7.09949 - (0.18022)(6.47445) = 7.09949 - 1.166825 = \underline{\mathbf{5.93266}}$$

(5 marks)

(5 marks)

(b) Interpret the slope coefficient estimate you calculated in part (a) -- i.e., explain in words what the numeric value you calculated for $\hat{\beta}_2$ means.

<u>*Note*</u>: $\hat{\beta}_2$ = **0.18022**. Y_i is measured in ***<u>dollars</u>* <u>per hour</u>**, and X_i is measured in ***<u>years</u>***.

The estimate **0.18022** of β_2 means that an ***increase* (decrease) in firm tenure** X_i of ***1 year*** is associated on average with **an *increase* (decrease) in male employees' hourly wage rate equal to *0.18 <u>dollars</u> per hour***, or ***18 <u>cents</u> per hour***.

(5 marks)
(c) Calculate an estimate of σ^2, the error variance.

$$RSS = \sum_{i=1}^{N} \hat{u}_i^2 = 4105.297; \qquad N - 2 = 274 - 2 = 272$$

$$\hat{\sigma}^2 = \frac{RSS}{N-2} = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2} = \frac{4{,}105.297}{274-2} = \frac{4{,}105.297}{272} = \underline{\mathbf{15.0930}}$$ **(5 marks)**

(5 marks)
(d) Calculate an estimate of $Var(\hat{\beta}_2)$, the variance of $\hat{\beta}_2$.

$$V\hat{a}r(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} x_i^2} = \frac{15.0930}{19{,}122.32} = \underline{\mathbf{0.00078929}}$$ **(5 marks)**

(6 marks)
(e) Compute the value of R^2, the coefficient of determination for the estimated OLS sample regression equation. Briefly explain what the calculated value of R^2 means.

(4 marks)

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} y_i^2} = \frac{4726.377 - 4105.297}{4726.377} = \frac{621.08}{4726.377} = \underline{\mathbf{0.1314}}$$

OR

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} y_i^2} = 1 - \frac{4105.297}{4726.377} = 1 - 0.8686 = \underline{\mathbf{0.1314}}$$

(2 marks)
Interpretation of R^2 = 0.1314: The value of 0.1314 indicates that **13.14 percent of the total sample (or observed) variation in Y_i** (employees' hourly wage rates) **is *attributable to*, or *explained by*, the sample regression function or the regressor X_i** (firm tenure).

(5 marks)

(f) Calculate the sample value of the t-statistic for testing the null hypothesis H_0: $\beta_2 = 0$ against the alternative hypothesis H_1: $\beta_2 \neq 0$. (Note: You are not required to obtain or state the inference of this test.)

- t-statistic for $\hat{\beta}_2$ is $t(\hat{\beta}_2) = \dfrac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)}$ (1) **(1 mark)**

- From part (a), $\hat{\beta}_2 = 0.180220$; from part (d), $\hat{Var}(\hat{\beta}_2) = 0.00078929$.

- $\hat{se}(\hat{\beta}_2) = \sqrt{\hat{Var}(\hat{\beta}_2)} = \sqrt{0.00078929} = \mathbf{0.0280943}$ **(1 mark)**

- Calculate the ***sample value*** **of the t-statistic** (1) under H_0: set $\beta_2 = 0$, $\hat{\beta}_2 = 0.180220$ and $\hat{se}(\hat{\beta}_2) = 0.0280943$ in (1).

$$t_0(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)} = \frac{0.18022 - 0.0}{0.0280943} = \frac{0.18022}{0.0280943} = 6.4148 = \underline{\mathbf{6.415}}$$ **(3 marks)**

(24 marks)

5. You have been commissioned to investigate the relationship between the median selling prices of houses and the average number of rooms per house in 506 census districts of a large metropolitan area. The dependent variable is $\textbf{\textit{price}}_i$, the median selling price of a house in the i-th census district, measured in *thousands of dollars*. The explanatory variable is $\textbf{\textit{rooms}}_i$, the average number of rooms per house in the i-th census district. The model you propose to estimate is given by the population regression equation

$$\text{price}_i = \beta_1 + \beta_2 \,\text{rooms}_i + u_i$$

Your research assistant has used the 506 sample observations on price_i and rooms_i to estimate the following OLS sample regression equation, where the figures in parentheses below the coefficient estimates are the *estimated standard errors* of the coefficient estimates:

$$\underset{}{\text{price}_i} = \underset{(26.52)}{-347.96} + \underset{(4.193)}{91.1955}\,\text{rooms}_i + \hat{u}_i \qquad (i = 1, ..., N) \qquad N = 506 \qquad \textbf{(3)}$$

← (standard errors)

(8 marks)

(a) Compute the two-sided 95% confidence interval for the slope coefficient β_2.

- The **two-sided (1 – α)-level,** or **100(1 – α) percent, confidence interval for β_2** is computed as

$$\hat{\beta}_2 - t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) \qquad \textbf{(2 marks)}$$

where

- $\hat{\beta}_{2L} = \hat{\beta}_2 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_2)$ = the ***lower* 100(1 – α)% confidence limit for β_2**
- $\hat{\beta}_{2U} = \hat{\beta}_2 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_2)$ = the ***upper* 100(1 – α)% confidence limit for β_2**
- $t_{\alpha/2}[N-2]$ = the **α/2 critical value of the t-distribution with N–2 degrees of freedom**.

- Required results and intermediate calculations:

$N - k = 506 - 2 = 504$; $\hat{\beta}_2 = \mathbf{91.1955}$; $\hat{se}(\hat{\beta}_2) = \mathbf{4.193}$

$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \boldsymbol{\alpha/2 = 0.025}$: $t_{\alpha/2}[N-2] = \mathbf{t_{0.025}[504] = 1.96}$

$t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_2) = t_{0.025}[504]\hat{se}(\hat{\beta}_2) = 1.96(4.193) = \mathbf{8.21828}$

Question 5(a) -- continued

- **Lower 95% confidence limit for β_2** is: **(3 marks)**

$$\hat{\beta}_{2L} = \hat{\beta}_2 - t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) = \hat{\beta}_2 - t_{0.025}[504]\,\hat{se}(\hat{\beta}_2)$$

$$= 91.1955 - 1.96(4.193) = 91.1955 - 8.21828 = 82.97722 = \mathbf{\underline{82.98}}$$

- **Upper 95% confidence limit for β_2** is: **(3 marks)**

$$\hat{\beta}_{2U} = \hat{\beta}_2 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) = \hat{\beta}_2 + t_{0.025}[504]\,\hat{se}(\hat{\beta}_2)$$

$$= 91.1955 + 1.96(4.193) = 91.1955 + 8.21828 = 99.41378 = \mathbf{\underline{99.41}}$$

- **Result:** The **two-sided 95% confidence interval for β_2** is:

  [82.98, 99.41]

(8 marks)
(b) Perform a test of the null hypothesis H_0: $\beta_2 = 0$ against the alternative hypothesis H_1: $\beta_2 \neq 0$ at the 1% significance level (i.e., for significance level $\alpha = 0.01$). Show how you calculated the test statistic. State the decision rule you use, and the inference you would draw from the test. Briefly indicate the conclusion you would draw from the test.

H_0: $\beta_2 = 0$
H_1: $\beta_2 \neq 0$ a ***two-sided*** **alternative hypothesis** $\Rightarrow$ a ***two-tailed*** **test**

- Test statistic is $t(\hat{\beta}_2) = \dfrac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)} \sim t[N-2]$. (1)

- $\hat{\beta}_2 = \mathbf{91.1955}$ and $\hat{se}(\hat{\beta}_2) = \mathbf{4.193}$

Question 5(b) -- continued

- Calculate the ***sample value* of the t-statistic** (1) under H_0: set $\beta_2 = 0$, $\hat{\beta}_2 = 91.1955$ and $\hat{se}(\hat{\beta}_2) = 4.193$ in (1).

$$t_0(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)} = \frac{91.1955 - 0.0}{4.193} = \frac{91.1955}{4.193} = 21.74946 = \underline{\mathbf{21.75}}$$ **(3 marks)**

- **Null distribution** of $t_0(\hat{\beta}_2)$ is **t[N – 2] = t[506 – 2] = t[504]**

<u>Decision Rule</u>: At significance level α, **(2 marks)**

- ***reject* H_0** if $|t_0(\hat{\beta}_2)| > t_{\alpha/2}[504]$,
  i.e., if either (1) $t_0(\hat{\beta}_2) > t_{\alpha/2}[504]$ or (2) $t_0(\hat{\beta}_2) < -t_{\alpha/2}[504]$;

- ***retain* H_0** if $|t_0(\hat{\beta}_2)| \le t_{\alpha/2}[504]$, i.e., if $-t_{\alpha/2}[504] \le t_0(\hat{\beta}_2) \le t_{\alpha/2}[504]$.

Critical value of t[504]-distribution: from t-table, use **df = ∞**.

- ***two-tailed* <u>1 percent</u> critical value** = $t_{\alpha/2}[504] = t_{0.005}[504] = \underline{\mathbf{2.58}}$ **(1 mark)**

<u>Inference</u>: **(1 mark)**

♦ At **1 percent significance level**, i.e., for α = 0.01,

$|t_0(\hat{\beta}_2)|$ = **21.75 > 2.58** = $t_{0.005}[504]$ ⇒ ***reject* H_0** vs. H_1 **at 1 percent level**.

♦ **<u>Inference</u>: At the 1% significance level**, the null hypothesis $\beta_2 = 0$ is ***rejected*** in favour of the alternative hypothesis $\beta_2 \neq 0$.

Meaning of test outcome: **(1 mark)**
Rejection of the null hypothesis $\beta_2 = 0$ against the alternative hypothesis $\beta_2 \neq 0$ means that **the sample evidence favours the existence of a relationship between** the ***median selling price of a houses*** **and** the ***average number of rooms per house***.

(8 marks)
Question 5(b) – Alternative Answer (uses confidence interval approach)

- The **two-sided (1 – α)-level,** or **100(1 – α) percent, confidence interval for β_2** is:

$$\hat{\beta}_2 - t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2)$$
$$\hat{\beta}_{2L} \le \beta_2 \le \hat{\beta}_{2U}$$

- Required results and intermediate calculations:

$N - k = 506 - 2 = 504$; $\hat{\beta}_2 = \mathbf{91.1955}$; $\hat{se}(\hat{\beta}_2) = \mathbf{4.193}$

$1 - \alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \boldsymbol{\alpha/2 = 0.005}$: $t_{\alpha/2}[N-2] = \mathbf{t_{0.005}[504]} = \underline{\mathbf{2.58}}$ **(1 mark)**

$t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) = t_{0.005}[504]\,\hat{se}(\hat{\beta}_2) = 2.58(4.193) = \mathbf{10.81794}$

- **Lower 99% confidence limit for β_2** is: **(2 marks)**

$$\hat{\beta}_{2L} = \hat{\beta}_2 - t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) = \hat{\beta}_2 - t_{0.005}[504]\,\hat{se}(\hat{\beta}_2)$$
$$= 91.1955 - 2.58(4.193) = 91.1955 - 10.81794 = 80.37756 = \underline{\mathbf{80.38}}$$

- **Upper 99% confidence limit for β_2** is: **(2 marks)**

$$\hat{\beta}_{2U} = \hat{\beta}_2 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_2) = \hat{\beta}_2 + t_{0.005}[504]\,\hat{se}(\hat{\beta}_2)$$
$$= 91.1955 + 2.58(4.193) = 91.1955 + 10.\,81794 = 102.01344 = \underline{\mathbf{102.01}}$$

- **<u>Decision Rule</u>:** At significance level α, **(1 mark)**

  - ***reject* H_0** if the ***hypothesized* value b_2 of β_2** specified by H_0 **lies *outside*** the two-sided (1−α)-level confidence interval for β_2, i.e., if either
    (1) $b_2 < \hat{\beta}_2 - t_{\alpha/2}[504]\hat{se}(\hat{\beta}_2)$ or (2) $b_2 > \hat{\beta}_2 + t_{\alpha/2}[504]\hat{se}(\hat{\beta}_2)$.

  - ***retain* H_0** if the ***hypothesized* value b_2 of β_2** specified by H_0 **lies *inside*** the two-sided (1−α)-level confidence interval for β_2, i.e., if
    $\hat{\beta}_2 - t_{\alpha/2}[504]\hat{se}(\hat{\beta}_2) \le b_2 \le \hat{\beta}_2 + t_{\alpha/2}[504]\hat{se}(\hat{\beta}_2)$.

Question 5(b) – Alternative Answer (continued)

Inference: **(1 mark)**

- At **1 percent significance level**, i.e., for $\alpha = 0.01$,

  $b_2 = \mathbf{0} < \mathbf{80.38} = \hat{\beta}_{2L} = \hat{\beta}_2 - t_{\alpha/2}[504]\hat{se}(\hat{\beta}_2) \Rightarrow$ ***reject*** $\mathbf{H_0}$ vs. H_1 **at 1 percent level**.

- **Inference: At the 1% significance level**, the null hypothesis $\beta_2 = 0$ is ***rejected*** in favour of the alternative hypothesis $\beta_2 \neq 0$.

Meaning of test outcome: **(1 mark)**

Rejection of the null hypothesis $\beta_2 = 0$ against the alternative hypothesis $\beta_2 \neq 0$ means that **the sample evidence favours the existence of a relationship between** the ***median selling price of a houses*** **and** the ***average number of rooms per house***.

(8 marks)

(c) Perform a test of the proposition that a one-room increase in average house size is associated on average with an increase in median house price of *more than* $80,000. Use the 5 percent significance level (i.e., $\alpha = 0.05$). State the null hypothesis H_0 and the alternative hypothesis H_1. Show how you calculated the test statistic. State the decision rule you use, and the inference you would draw from the test.

H_0: $\beta_2 = 80.0$
H_1: $\beta_2 > 80.0$ ⇒ a ***right-tailed*** **test** **(1 mark)**

- Test statistic is $t(\hat{\beta}_2) = \dfrac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)} \sim t[N-2]$. (1)

- Calculate the ***sample value*** **of the t-statistic** (1) under H_0: set $\beta_2 = 80.0$, $\hat{\beta}_2 = 91.1955$ and $\hat{se}(\hat{\beta}_2) = 4.193$ in (1).

$$t_0(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)} = \frac{91.1955 - 80.0}{4.193} = \frac{11.1955}{4.193} = 2.6700 = \underline{\mathbf{2.67}}$$ **(3 marks)**

- **Null distribution** of $t_0(\hat{\beta}_2)$ is **t[N – 2] = t[506 – 2] = t[504]**

<u>Decision Rule</u>: At significance level α, **(1 mark)**

- ***reject*** $\mathbf{H_0}$ if $t_0(\hat{\beta}_2) > t_\alpha[504]$,
- ***retain*** $\mathbf{H_0}$ if $t_0(\hat{\beta}_2) \le t_\alpha[504]$.

Critical value of t[504]-distribution: from t-table, use **df =** ∞.

- ***right-tailed*** **<u>5 percent</u> critical value** = $t_{0.05}[504]$ = **<u>1.645</u>** **=** **<u>1.65</u>** **(1 mark)**

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

- At **5 percent significance level**, i.e., for $\alpha = 0.05$,

  $t_0(\hat{\beta}_2) = \mathbf{2.67 > 1.65} = t_{0.05}[504]$ ⇒ ***reject*** $\mathbf{H_0}$ vs. H_1 **at 1 percent level**.

- **<u>Inference</u>: At the 5% significance level**, the null hypothesis $\boldsymbol{\beta_2 = 80}$ **is *rejected*** in favour of the alternative hypothesis $\boldsymbol{\beta_2 > 80}$.

Percentage Points of the t-Distribution

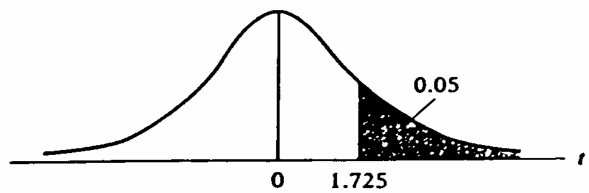
TABLE D.2
Percentage points of the *t* distribution

Example

Pr(t > 2.086) = 0.025

Pr(t > 1.725) = 0.05 for df = 20

Pr($|t|$ > 1.725) = 0.10



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika*.