QUEEN'S UNIVERSITY AT KINGSTON
Department of Economics

CONFIDENTIAL
turn in exam
question paper

ECONOMICS 351* - Fall Term 2008

Introductory Econometrics

Fall Term 2008 **MID-TERM EXAM: ANSWERS** M.G. Abbott

DATE: **Monday October 27, 2008.**

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

INSTRUCTIONS: The exam consists of **FOUR (4)** questions. Students are required to answer **ALL FOUR (4)** questions.

Answer all questions in the exam booklets provided. Be sure your ***student number*** is printed clearly on the front of all exam booklets used. Your ***name*** is optional.

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.

Tables of percentage points of the t-distribution and F-distribution are given on the last two pages of the exam.

This exam is **CONFIDENTIAL**. This question paper must be submitted in its entirety with your answer booklet(s); otherwise your exam will not be marked.

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

GOOD LUCK!

QUESTIONS: **Answer ALL FOUR questions.**

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

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \textbf{(1)}$$

where Y_i and X_i are observable variables, β_0 and β_1 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

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(2)}$$

where $\hat{\beta}_0$ is the OLS estimator of the intercept coefficient β_0, $\hat{\beta}_1$ is the OLS estimator of the slope coefficient β_1, $\hat{u}_i$ is the OLS residual for the i-th sample observation, and N is sample size (the number of observations in the sample).

(15 marks)

1. Show that the OLS slope coefficient estimator $\hat{\beta}_1$ is a linear function of the Y_i sample values. Stating explicitly all required assumptions, prove that the OLS slope coefficient estimator $\hat{\beta}_1$ is an unbiased estimator of the slope coefficient β_1.

ANSWER to Question 1:

(5 marks)

- Show that the OLS slope coefficient estimator $\hat{\beta}_1$ is a *linear* function of the Y_i sample values.

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i x_i (Y_i - \overline{Y})}{\sum_i x_i^2} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2} - \frac{\overline{Y} \sum_i x_i}{\sum_i x_i^2}$$

$$= \frac{\sum_i x_i Y_i}{\sum_i x_i^2} \qquad \text{because } \sum_i x_i = 0$$

$$= \Sigma_i k_i Y_i \qquad \text{where } k_i \equiv \frac{x_i}{\Sigma_i x_i^2}.$$

(5 marks)

ANSWER to Question 1 -- continued:

(10 marks)

- Stating explicitly all required assumptions, prove that the OLS slope coefficient estimator $\hat{\beta}_1$ is an unbiased estimator of the slope coefficient β_1.

(1) **Substitute for Y_i** the expression $Y_i = \beta_0 + \beta_1 X_i + u_i$ from the population regression equation (or PRE). **(5 marks)**

$$\begin{aligned}
\hat{\beta}_1 &= \sum_i k_i Y_i \\
&= \sum_i k_i(\beta_0 + \beta_1 X_i + u_i) && \text{since } Y_i = \beta_0 + \beta_1 X_i + u_i \text{ by assumption (A1)} \\
&= \sum_i (\beta_0 k_i + \beta_1 k_i X_i + k_i u_i) \\
&= \beta_0 \sum_i k_i + \beta_1 \sum_i k_i X_i + \sum_i k_i u_i \\
&= \beta_1 + \sum_i k_i u_i, && \text{since } \sum_i k_i = 0 \text{ and } \sum_i k_i X_i = 1
\end{aligned}$$

(2) Now **take expectations** of the above expression for $\hat{\beta}_1$ conditional on the value X_i of X: **(5 marks)**

$$\begin{aligned}
E(\hat{\beta}_1) &= E(\beta_1) + E[\sum_i k_i u_i] \\
&= \beta_1 + \sum_i k_i E(u_i | X_i) && \text{since } \beta_1 \text{ is a constant and the } k_i \text{ are nonstochastic} \\
&= \beta_1 + \sum_i k_i 0 && \text{since } E(u_i | X_i) = E(u_i) = 0 \text{ by assumption (A2)} \\
&= \beta_1
\end{aligned}$$

(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 $\hat{\beta}_1$. State all assumptions required for the derivation.

ANSWER to Question 2:

(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,

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

- Use **three implications** of the error normality assumption to derive $F(\hat{\beta}_1)$:

1. $\hat{\beta}_1 \sim N\left[\beta_1, \text{Var}(\hat{\beta}_1)\right]$ where $\text{Var}(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_{i=1}^{N} x_i^2} = \dfrac{\sigma^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$ **(1 mark)**

Implications of normality of $\hat{\beta}_1$:

$$Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \sim N[0, 1]$$

<u>ANSWER to Question 2 -- continued</u>:

$$[Z(\hat{\beta}_1)]^2 = \frac{(\hat{\beta}_1-\beta_1)^2}{Var(\hat{\beta}_1)} = \frac{(\hat{\beta}_1-\beta_1)^2}{\sigma^2/\sum_i x_i^2} = \frac{(\hat{\beta}_1-\beta_1)^2\sum_i x_i^2}{\sigma^2} \sim \chi^2[1] \qquad (1)$$ **(1 mark)**

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

3. The estimators $\hat{\beta}_1$ and $\hat{\sigma}^2$ are statistically independent
OR **(1 mark)**
The statistics $Z(\hat{\beta}_1)$ and $\hat{\sigma}^2/\sigma^2$ are statistically independent.

♦ **The *F-statistic* for** $\hat{\beta}_1$. The F-statistic for $\hat{\beta}_1$ is therefore the ratio of (1) to (2): **(6 marks)**

$$\begin{aligned} F(\hat{\beta}_1) &= \frac{\left(Z(\hat{\beta}_1)\right)^2}{\hat{\sigma}^2/\sigma^2} \\ &= \frac{\left(\hat{\beta}_1-\beta_1\right)^2\left(\sum_i x_i^2\right)/\sigma^2}{\hat{\sigma}^2/\sigma^2} \\ &= \frac{\left(\hat{\beta}_1-\beta_1\right)^2\left(\sum_i x_i^2\right)}{\hat{\sigma}^2} \\ &= \frac{\left(\hat{\beta}_1-\beta_1\right)^2}{\hat{\sigma}^2/\sum_i x_i^2} \\ &= \frac{\left(\hat{\beta}_1-\beta_1\right)^2}{V\hat{a}r(\hat{\beta}_1)} \qquad \text{since } \hat{\sigma}^2/\sum_i x_i^2 = V\hat{a}r(\hat{\beta}_1). \end{aligned}$$

❑ ***<u>Result</u>:*** The **F-statistic for** $\hat{\beta}_1$ takes the form

$$F(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1-\beta_1\right)^2}{\hat{\sigma}^2/\left(\sum_i x_i^2\right)} = \frac{\left(\hat{\beta}_1-\beta_1\right)^2}{V\hat{a}r(\hat{\beta}_1)} \sim F[1, N-2].$$

(32 marks)

3. A researcher is using data for a sample of 3240 female employees 25 years of age and over to investigate the relationship between employees' hourly wage rates Y_i (measured in *dollars per hour*) and their age X_i (measured in *years*). The population regression equation takes the form of equation (1): $Y_i = \beta_0 + \beta_1 X_i + u_i$. Preliminary analysis of the sample data produces the following sample information:

$N = 3240$ $\quad \sum_{i=1}^{N} y_i^2 = 78434.97$ $\quad \sum_{i=1}^{N} x_i^2 = 25526.17$ $\quad \sum_{i=1}^{N} x_i y_i = 3666.426$

$\sum_{i=1}^{N} Y_i = 34379.16$ $\quad \sum_{i=1}^{N} X_i = 96143.00$ $\quad \sum_{i=1}^{N} Y_i^2 = 443227.1$

$\sum_{i=1}^{N} X_i^2 = 2878451.0$ $\quad \sum_{i=1}^{N} X_i Y_i = 1023825.0$ $\quad \sum_{i=1}^{N} \hat{u}_i^2 = 77908.35$

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_0 and the slope coefficient β_1.

- $\hat{\beta}_1 = \dfrac{\sum_i x_i y_i}{\sum_i x_i^2} = \dfrac{3666.426}{25526.17} =$ **0.143634 = <u>0.1436</u>** **(5 marks)**

- $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

  $\bar{Y} = \dfrac{\sum_{i=1}^{N} Y_i}{N} = \dfrac{34379.16}{3240} = 10.61085$ and $\bar{X} = \dfrac{\sum_{i=1}^{N} X_i}{N} = \dfrac{96143.00}{3240} = 29.67377$

  Therefore

  $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 10.610852 - (0.143634)(29.67377) = 10.610852 - 4.262162 =$ **<u>6.34869</u>**

  = **<u>6.349</u>**

  (5 marks)

(4 marks)

(b) Interpret the slope coefficient estimate you calculated in part (a) – i.e., explain in words what the numeric value you calculated for $\hat{\beta}_1$ means.

Note: $\hat{\beta}_1$ = **0.1436**. Y_i is measured in ***dollars per hour***, and X_i is measured in ***years***.

The estimate **0.143634** of β_1 means that a **1-year *increase* (decrease) in age** X_i is associated on average with **an *increase* (decrease) in hourly wage rate equal to *0.1436 dollars per hour***, or ***14.36 cents per hour***.

(4 marks)

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

$$\text{RSS} = \sum_{i=1}^{N} \hat{u}_i^2 = 77908.35; \qquad N - 2 = 3240 - 2 = 3238$$

$$\hat{\sigma}^2 = \frac{\text{RSS}}{N-2} = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2} = \frac{77908.35}{3240-2} = \frac{77908.35}{3238} = \underline{24.060639} = \underline{24.0606}$$

(4 marks)

(6 marks)

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

(4 marks)

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} y_i^2} = \frac{78434.97 - 77908.35}{78434.97} = \frac{526.62}{78434.97} = \underline{0.006714}$$

OR

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} y_i^2} = 1 - \frac{77908.35}{78434.97} = 1 - 0.993286 = \underline{0.006714}$$

(2 marks)

Interpretation of R^2 = 0.006714: The value of 0.006714 indicates that **0.67 percent of the total sample (or observed) variation in Y_i** (hourly wage rates) **is *attributable to*, or *explained by*, the regressor X_i** (age) **or the sample regression function**.

(8 marks)

(e) Calculate the sample value of the t-statistic for testing the null hypothesis H_0: $\beta_1 = 0$ against the alternative hypothesis H_1: $\beta_1 \neq 0$. (Note: You are not required to obtain or state the inference of this test.)

- t-statistic for $\hat{\beta}_1$ is $t(\hat{\beta}_1) = \dfrac{(\hat{\beta}_1 - \beta_1)}{\hat{se}(\hat{\beta}_1)}$ (1) **(2 marks)**

- From part (a), $\hat{\beta}_1 = \mathbf{0.143634}$; $\hat{Var}(\hat{\beta}_1) = \dfrac{\hat{\sigma}^2}{\sum_{i=1}^{N} x_i^2} = \dfrac{24.060639}{25526.17} = 0.000942587.$

- The estimated standard error of $\hat{\beta}_1$ is:

$$\hat{se}(\hat{\beta}_1) = \sqrt{\hat{Var}(\hat{\beta}_1)} = \sqrt{0.000942587} = 0.03070158 = 0.0307016$$

- Calculate the ***sample value*** **of the t-statistic** (1) under H_0: set $\beta_1 = 0$, $\hat{\beta}_1 = \mathbf{0.143634}$ and $\hat{se}(\hat{\beta}_1) = 0.0307016$ in (1) above.

$$t_0(\hat{\beta}_1) = \frac{(\hat{\beta}_1 - \beta_1)}{\hat{se}(\hat{\beta}_1)} = \frac{0.143634 - 0}{0.0307016} = \frac{0.143634}{0.0307016} = \underline{\mathbf{4.678388}} = \underline{\mathbf{4.68}}$$ **(6 marks)**

(38 marks)

4. You have been commissioned to investigate the relationship between the birth weights of newborn females and the number of prenatal visits to a physician or midwife that their mothers made during pregnancy. The dependent variable is ***bwght$_i$***, the birth weight of the i-th newborn female, measured in *grams*. The explanatory variable is ***pnvisits$_i$***, the number of prenatal visits of the i-th newborn's mother during pregnancy, measured in *number of visits*. The model you propose to estimate is given by the population regression equation

$$\text{bwght}_i = \beta_0 + \beta_1 \text{pnvisits}_i + u_i .$$

Your research assistant has used 857 sample observations on ***bwght$_i$*** and ***pnvisits$_i$*** 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:

$$\underset{}{\text{bwght}_i} = \underset{(65.6909)}{3199.02} + \underset{(5.36347)}{14.1219}\,\text{pnvisits}_i + \hat{u}_i \qquad (i = 1, ..., N) \qquad N = 857 \qquad \textbf{(3)}$$

$\leftarrow$ (standard errors)

(8 marks)

(a) Perform a test of the null hypothesis H$_0$: $\beta_1 = 0$ against the alternative hypothesis H$_1$: $\beta_1 \neq 0$ 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. What would you conclude from the results of the test?

H$_0$: $\beta_1 = 0$
H$_1$: $\beta_1 \neq 0$ a ***two-sided*** **alternative hypothesis** $\Rightarrow$ a ***two-tailed*** **test**

- Test statistic is either $t(\hat{\beta}_1) = \dfrac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} \sim t[N-2]$ *or* $F(\hat{\beta}_1) = \dfrac{(\hat{\beta}_1 - \beta_1)^2}{\hat{Var}(\hat{\beta}_1)} \sim F[1, N-2]$.

- $\hat{\beta}_1$ = **14.1219**; $\hat{se}(\hat{\beta}_1)$ = **5.36347**; $\hat{Var}(\hat{\beta}_1) = \left(\hat{se}(\hat{\beta}_1)\right)^2$ = **28.76681**

- Calculate the ***sample value*** **of *either* the t-statistic *or* the F-statistic** under H$_0$: set $\beta_1 = 0$, $\hat{\beta}_1$ = **14.1219**, $\hat{se}(\hat{\beta}_1)$ = **5.36347,** and $\hat{Var}(\hat{\beta}_1)$ = **28.76681**.

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} = \frac{14.1219 - 0.0}{5.36347} = \frac{14.1219}{5.36347} = \mathbf{2.63298} = \underline{\mathbf{2.63}}$$

or **(3 marks)**

ANSWER to Question 4(a) -- continued:

$$F_0(\hat{\beta}_1) = \frac{(\hat{\beta}_1 - \beta_1)^2}{\hat{Var}(\hat{\beta}_1)} = \frac{(14.1219 - 0.0)^2}{28.76681} = \frac{199.42806}{28.76681} = \mathbf{6.9326} = \underline{\mathbf{6.93}}$$

- **Null distribution** of $t_0(\hat{\beta}_1)$ is **t[N – 2] = t[857 – 2] = t[855]**
- **Null distribution** of $F_0(\hat{\beta}_1)$ is **F[1, N – 2] = F[1, 857 – 2] = F[1, 855]**

Decision Rule: At significance level α, **(2 marks)**

- ***reject* H₀** if $F_0(\hat{\beta}_1) > F_\alpha[1, 855]$ *or* $|t_0(\hat{\beta}_1)| > t_{\alpha/2}[855]$,

  i.e., if either (1) $t_0(\hat{\beta}_1) > t_{\alpha/2}[855]$ or (2) $t_0(\hat{\beta}_1) < -t_{\alpha/2}[855]$;

- ***retain* H₀** if $F_0(\hat{\beta}_1) \le F_\alpha[1, 855]$ *or* $|t_0(\hat{\beta}_1)| \le t_{\alpha/2}[855]$,

  i.e., if $-t_{\alpha/2}[855] \le t_0(\hat{\beta}_1) \le t_{\alpha/2}[855]$.

Critical values of t[855]-distribution or F[1, 855]-distribution: in t-table, use **df = ∞**.

- ***two-tailed*** **1 percent critical value** = $t_{\alpha/2}[855] = t_{0.005}[855] = 2.5816 =$ **2.58** **(1 mark)**

  $= t_{0.005}[\infty] = 2.576 =$ **2.58**

Critical values of F[1, 855]-distribution: in F-table, use **denominator df = ∞**.

- **1 percent critical value** = $F_\alpha[1, 855] = F_{0.01}[1, 855] = 6.6646 =$ **6.66**

  $= F_{0.01}[1, \infty] =$ **6.63**

Inference: **(1 mark)**

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

$|t_0(\hat{\beta}_1)|$ = **2.63 > 2.58** = $t_{0.005}[855]$ ⇒ ***reject* H₀** vs. H_1 **at 1 percent level**.

$F_0(\hat{\beta}_1)$ = **6.93 > 6.63** = $F_{0.01}[1, 855]$ ⇒ ***reject* H₀** vs. H_1 **at 1 percent level**.

<u>ANSWER to Question 4(a) -- continued</u>:

- **<u>Inference</u>: At the 1% significance level**, the null hypothesis H_0: $\beta_1 = 0$ is ***rejected*** in favour of the alternative hypothesis H_1: $\beta_1 \neq 0$.

<u>Conclusion implied by test outcome</u>: **(1 mark)**

Rejection of the null hypothesis $\beta_1 = 0$ against the alternative hypothesis $\beta_1 \neq 0$ means that **the sample evidence favours the existence of a relationship between the *number of prenatal visits*** of mothers and the ***birth weights*** of their female newborns.

<u>Question 4(a) – Alternative Answer</u> -- uses ***confidence interval*** **approach**

- The **two-sided (1 – α)-level,** or **100(1 – α) percent, confidence interval for β_1** is:

$$\hat{\beta}_1 - t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_1)$$

$$\hat{\beta}_{1L} \le \beta_1 \le \hat{\beta}_{1U}$$

- Required results and intermediate calculations:

$N - K = 857 - 2 = 855$; $\hat{\beta}_1 = \mathbf{14.1219}$; $\hat{se}(\hat{\beta}_1) = \mathbf{5.36347}$

$1 - \alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \boldsymbol{\alpha/2 = 0.005}$: $t_{\alpha/2}[N-2] = \mathbf{t_{0.005}[855]} = \underline{\mathbf{2.576}}$ **(1 mark)**

$t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_1) = t_{0.005}[855]\,\hat{se}(\hat{\beta}_1) = 2.576(5.36347) = \mathbf{13.816299}$

- **Lower 99% confidence limit for β_1** is: **(2 marks)**

$$\hat{\beta}_{1L} = \hat{\beta}_1 - t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_1) = \hat{\beta}_1 - t_{0.005}[855]\,\hat{se}(\hat{\beta}_1)$$

$= 14.1219 - \mathbf{2.576}(5.36347) = 14.1219 - 13.816299 = 0.305601 = \underline{\mathbf{0.3056}}$

$= 14.1219 - \mathbf{2.617}(5.36347) = 14.1219 - 14.036201 = 0.0696086 = \underline{\mathbf{0.08570}}$

$= 14.1219 - \mathbf{2.58}(5.36347) = 14.1219 - 13.837753 = 0.284147 = \underline{\mathbf{0.2841}}$

- **Upper 99% confidence limit for β_1** is: **(2 marks)**

$$\hat{\beta}_{1U} = \hat{\beta}_1 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_1) = \hat{\beta}_1 + t_{0.005}[855]\,\hat{se}(\hat{\beta}_1)$$

$= 14.1219 + \mathbf{2.576}(5.36347) = 14.1219 + 13.816299 = 27.938199 = \underline{\mathbf{27.9382}}$

$= 14.1219 + \mathbf{2.617}(5.36347) = 14.1219 + 14.036201 = 28.158101 = \underline{\mathbf{28.1581}}$

$= 14.1219 + \mathbf{2.58}(5.36347) = 14.1219 + 13.837753 = 27.959653 = \underline{\mathbf{27.9597}}$

Question 4(a) – Alternative Answer (continued)

- **Two-sided 99% confidence interval for β_1** is therefore:

  $\mathbf{0.3056 \le \beta_1 \le 27.9382}$ *or* $\mathbf{0.08570 \le \beta_1 \le 28.1581}$ *or* $\mathbf{0.2841 \le \beta_1 \le 27.9597}$

- **Decision Rule:** At significance level α, **(1 mark)**

- ***reject* H$_0$** if the ***hypothesized* value b$_1$ of β_1** specified by H_0 **lies *outside*** the two-sided $(1-\alpha)$-level confidence interval for β_1, i.e., if either

  (1) $b_1 < \hat{\beta}_1 - t_{\alpha/2}[855]\hat{se}(\hat{\beta}_1)$ or (2) $b_1 > \hat{\beta}_1 + t_{\alpha/2}[855]\hat{se}(\hat{\beta}_1)$.

- ***retain* H$_0$** if the ***hypothesized* value b$_1$ of β_1** specified by H_0 **lies *inside*** the two-sided $(1-\alpha)$-level confidence interval for β_1, i.e., if

  $\hat{\beta}_1 - t_{\alpha/2}[855]\hat{se}(\hat{\beta}_1) \le b_1 \le \hat{\beta}_1 + t_{\alpha/2}[855]\hat{se}(\hat{\beta}_1)$.

Inference: **(1 mark)**

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

  $$\mathbf{b_1 = 0 < 0.3056} = \hat{\beta}_{1L} = \hat{\beta}_1 - t_{0.005}[855]\hat{se}(\hat{\beta}_1)$$

  $\Rightarrow$ ***reject* H$_0$** vs. H_1 **at 1 percent level**.

- **Inference: At the 1% significance level**, the null hypothesis H_0: $\beta_1 = 0$ is ***rejected*** in favour of the alternative hypothesis H_1: $\beta_1 \neq 0$.

Conclusion implied by test outcome: **(1 mark)**

Rejection of the null hypothesis $\beta_1 = 0$ against the alternative hypothesis $\beta_1 \neq 0$ means that **the sample evidence favours the existence of a relationship between the *number of prenatal visits*** of mothers and the ***birth weights*** of their female newborns.

(8 marks)

(b) Perform a test of the proposition that newborn females' birth weights are positively related to their mother's number of prenatal visits during pregnancy, i.e., that an increase in the number of prenatal visits *increases* the birth weights of newborn females. Use the 1 percent significance level (i.e., $\alpha = 0.01$). 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.

<u>ANSWER to Question 4(b)</u>:

<u>Null and Alternative Hypotheses</u>:

H_0: $\beta_1 = 0$ **(1 mark)**

H_1: $\beta_1 > 0$ $\Rightarrow$ a ***right-tailed*** **t-test** **(1 mark)**

- Test statistic is $t(\hat{\beta}_1) = \dfrac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} \sim t[N-2]$; $\hat{\beta}_1 = \mathbf{14.1219}$; $\hat{se}(\hat{\beta}_1) = \mathbf{5.36347}$.

- Calculate the ***sample value*** **of the t-statistic** under H_0:
  set $\beta_1 = 0$, $\hat{\beta}_1 = \mathbf{14.1219}$, and $\hat{se}(\hat{\beta}_1) = \mathbf{5.36347}$.

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} = \frac{14.1219 - 0.0}{5.36347} = \frac{14.1219}{5.36347} = \mathbf{2.63298} = \underline{\mathbf{2.63}}$$ **(2 marks)**

- **Null distribution** of $t_0(\hat{\beta}_1)$ is **t[N – 2] = t[857– 2] = t[855]**

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

- ***reject*** $\mathbf{H_0}$ if $t_0(\hat{\beta}_1) > t_\alpha[855]$,
- ***retain*** $\mathbf{H_0}$ if $t_0(\hat{\beta}_1) \le t_\alpha[855]$.

Critical value of t[855]-distribution: from t-table, use **df = 120** or **df = ∞** or any number in between.

- ***right-tail*** **<u>1 percent</u> critical value** = $t_{0.01}[855]$ = **<u>2.331</u>** = **<u>2.33</u>** **(1 mark)**
  = $t_{0.01}[855]$ = **<u>2.358</u>** = **<u>2.36</u>**
  = $t_{0.01}[\infty]$ = **<u>2.326</u>** = **<u>2.33</u>**

ANSWER to Question 4(b) -- continued:

Inference: **(1 mark)**

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

  $t_0(\hat{\beta}_1) = \mathbf{2.63} > \mathbf{2.33} = t_{0.01}[855] \Rightarrow$ ***reject*** $\mathbf{H_0}$ vs. H_1 **at 1 percent level**.

- **Inference: At the 1% significance level**, the null hypothesis $\mathbf{H_0: \beta_1 = 0}$ **is *rejected*** in favour of the alternative hypothesis $\mathbf{H_1: \beta_1 > 0}$.

(14 marks)

(c) Compute the two-sided 95% confidence interval for the intercept coefficient β_0. Use this two-sided 95% confidence interval for β_0 to test the hypothesis that the mean birth weight of newborn females whose mothers made no prenatal visits to a physician or midwife equals 3,000 grams. State the null hypothesis H_0 and the alternative hypothesis H_1. State the decision rule you use, and the inference you would draw from the test.

<u>ANSWER to Question 4(c)</u>:

- The **two-sided (1 – α)-level,** or **100(1 – α) percent, confidence interval for β_1** is computed as

$$\hat{\beta}_0 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0)$$ **(2 marks)**

where

- $\hat{\beta}_{0L} = \hat{\beta}_0 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0)$ = the ***lower*** **100(1 – α)% confidence limit for β_0**
- $\hat{\beta}_{0U} = \hat{\beta}_0 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0)$ = the ***upper*** **100(1 – α)% confidence limit for β_0**
- $t_{\alpha/2}[N-2]$ = the **α/2 critical value of the t-distribution with N–2 degrees of freedom**.

- Required results and intermediate calculations:

$N - K = 857 - 2 = 855$; $\hat{\beta}_0 = 3199.02$; $\hat{se}(\hat{\beta}_0) = 65.6909$

$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow$ **$\alpha/2 = 0.025$**: $t_{\alpha/2}[N-2] = \mathbf{t_{0.025}[855] = 1.963}$
$= \mathbf{t_{0.025}[855] = 1.980}$
$= \mathbf{t_{0.025}[\infty] = 1.960}$

$t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) = t_{0.025}[855]\hat{se}(\hat{\beta}_0) = 1.963(65.6909) = \mathbf{128.95124}$

- **Lower 95% confidence limit for β_0** is: **(3 marks)**

$$\hat{\beta}_{0L} = \hat{\beta}_0 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) = \hat{\beta}_0 - t_{0.025}[855]\hat{se}(\hat{\beta}_0)$$

$= 3199.02 - \mathbf{1.963}(65.6909) = 3199.02 - 128.95124 = 3070.069 =$ **<u>3070.07</u>**

$= 3199.02 - \mathbf{1.980}(65.6909) = 3199.02 - 130.06798 = 3068.952 =$ **<u>3068.95</u>**

$= 3199.02 - \mathbf{1.960}(65.6909) = 3199.02 - 128.75416 = 3070.266 =$ **<u>3070.27</u>**

ANSWER to Question 4(c) -- continued:

- **Upper 95% confidence limit for β_0** is: **(3 marks)**

$$\hat{\beta}_{0U} = \hat{\beta}_0 + t_{\alpha/2}[N-2]\,\hat{se}(\hat{\beta}_0) = \hat{\beta}_0 + t_{0.025}[855]\,\hat{se}(\hat{\beta}_0)$$

$$= 3199.02 + \mathbf{1.963}(65.6909) = 3199.02 + 128.95124 = 3327.971 = \underline{\mathbf{3327.97}}$$

$$= 3199.02 + \mathbf{1.980}(65.6909) = 3199.02 + 130.06798 = 3329.088 = \underline{\mathbf{3329.09}}$$

$$= 3199.02 + \mathbf{1.960}(65.6909) = 3199.02 + 128.75416 = 3327.774 = \underline{\mathbf{3327.77}}$$

- **Result:** The **two-sided 95% confidence interval for β_0** is: **[3070.07, 3327.97]**

  3070.07 ≤ β_0 ≤ 3327.97 *or*

  3068.95 ≤ β_0 ≤ 3329.09 *or*

  3070.27 ≤ β_0 ≤ 3327.77

- **Null and Alternative Hypotheses:** **(2 marks)**

  H_0: $\beta_0 = 3000$
  H_1: $\beta_0 \neq 3000$ a ***two-sided* alternative hypothesis** ⇒ a ***two-tailed* test**

- **Decision Rule:** At significance level α, **(2 marks)**

- ***reject* H_0** if the ***hypothesized* value b_0 of β_0** specified by H_0 **lies *outside*** the two-sided $(1-\alpha)$-level confidence interval for β_0, i.e., if either
  (1) $b_0 < \hat{\beta}_0 - t_{\alpha/2}[855]\hat{se}(\hat{\beta}_0)$ or (2) $b_0 > \hat{\beta}_0 + t_{\alpha/2}[855]\hat{se}(\hat{\beta}_0)$.

- ***retain* H_0** if the ***hypothesized* value b_0 of β_0** specified by H_0 **lies *inside*** the two-sided $(1-\alpha)$-level confidence interval for β_0, i.e., if
  $\hat{\beta}_0 - t_{\alpha/2}[855]\hat{se}(\hat{\beta}_0) \le b_0 \le \hat{\beta}_0 + t_{\alpha/2}[855]\hat{se}(\hat{\beta}_0)$.

Inference: **(2 marks)**

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

$$b_0 = \mathbf{3000} < \mathbf{3070.07} = \hat{\beta}_{0L} = \hat{\beta}_0 - t_{0.025}[855]\hat{se}(\hat{\beta}_0)$$

⇒ ***reject* H_0** vs. H_1 **at 5 percent level**.

- **Inference: At the 5% significance level**, the **null hypothesis H_0: $\beta_0 = 3000$** is ***rejected*** in favour of the alternative hypothesis H_1: $\beta_0 \neq 3000$.

(8 marks)

(d) Perform a test of the proposition that each additional prenatal visit made by the mother is associated on average with an increase in their newborn females' birth weight of less than 25 grams. 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.

<u>ANSWER to Question 4(d)</u>:

<u>Null and Alternative Hypotheses</u>:

H_0: $\beta_1 = 25$ **(1 mark)**

H_1: $\beta_1 < 25$ $\Rightarrow$ a ***left-tailed*** **t-test** **(1 mark)**

- Test statistic is $t(\hat{\beta}_1) = \dfrac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} \sim t[N-2]$; $\hat{\beta}_1 = \mathbf{14.1219}$; $\hat{se}(\hat{\beta}_1) = \mathbf{5.36347}$.

- Calculate the ***sample value* of the t-statistic** under H_0:
  set $\beta_1 = 25$, $\hat{\beta}_1 = \mathbf{14.1219}$, and $\hat{se}(\hat{\beta}_1) = \mathbf{5.36347}$.

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} = \frac{14.1219 - 25.0}{5.36347} = \frac{-10.8781}{5.36347} = \mathbf{-2.02818} = \underline{\mathbf{-2.03}}$$ **(2 marks)**

- **Null distribution** of $t_0(\hat{\beta}_1)$ is **t[N – 2] = t[857– 2] = t[855]**

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

- ***reject*** $\mathbf{H_0}$ if $t_0(\hat{\beta}_1) < -t_{\alpha}[855]$,
- ***retain*** $\mathbf{H_0}$ if $t_0(\hat{\beta}_1) \ge -t_{\alpha}[855]$.

Critical value of t[855]-distribution: from t-table, use **df = 120** or **df = ∞** or any number in between.

- ***left-tail*** **<u>1 percent</u> critical value** $= -t_{0.05}[855] = \underline{\mathbf{-1.6466}} = \underline{\mathbf{-1.65}}$ **(1 mark)**
  $= -t_{0.05}[855] = \underline{\mathbf{-1.658}} = \underline{\mathbf{-1.66}}$
  $= -t_{0.05}[\infty] = \underline{\mathbf{-1.645}} = \underline{\mathbf{-1.65}}$

ANSWER to Question 4(d) -- continued:

Inference: **(1 mark)**

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

  $t_0(\hat{\beta}_1) = \mathbf{-2.03} < \mathbf{-1.65} = -\,t_{0.05}[855] \;\Rightarrow\;$ ***reject* $\mathbf{H_0}$** vs. H_1 **at 5 percent level**.

- **Inference: At the 5% significance level**, the null hypothesis **H_0: $\beta_1 = 25$ is *rejected*** in favour of the alternative hypothesis **H_1: $\beta_1 < 25$**.

Percentage Points of the t-Distribution

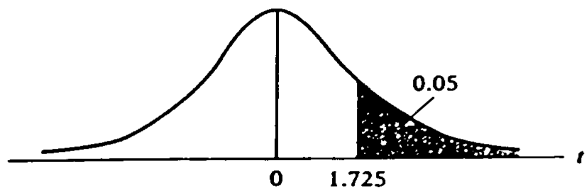
TABLE D.2
Percentage points of the *t* distribution

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*.

Source: Damodar N. Gujarati, *Basic Econometrics*, Third Edition. New York: McGraw-Hill, 1995, p. 809.

Selected Upper Percentage Points of the F-Distribution

TABLE D.3
Upper percentage points of the *F* distribution (*continued*)

df for denominator N_2	df for numerator N_1												
	Pr	1	2	3	4	5	6	7	8	9	10	11	12
22	.25	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.38	1.37
	.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	.25	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.37	1.36
	.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
26	.25	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.36	1.35
	.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	.25	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
	.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
30	.25	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.35	1.34
	.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
40	.25	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31
	.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	.25	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.29
	.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
120	.25	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.27	1.26
	.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	.25	1.33	1.39	1.38	1.36	1.34	1.32	1.31	1.29	1.28	1.27	1.26	1.25
	.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
∞	.25	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.24
	.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

Source: Damodar N. Gujarati, *Basic Econometrics*, Third Edition. New York: McGraw-Hill, 1995, p. 814.