QUEEN'S UNIVERSITY AT KINGSTON
Department of Economics

ECONOMICS 351* - Section A

Introductory Econometrics

Fall Term 2000 **FINAL EXAMINATION** M.G. Abbott

DATE: **Friday December 15, 2000.**

TIME: **Three (3) hours (180 minutes); 2:00 p.m. - 5:00 p.m.**

INSTRUCTIONS: The examination is divided into two parts.

PART A contains two questions; students are required to answer **ONE** of the two questions 1 and 2 in Part A.

PART B contains three questions; students are required to answer **ALL THREE** of the questions 3, 4 and 5 in Part B.

- 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.** GOOD LUCK! Happy Holidays!

If the instructor is unavailable in the examination room and if doubt exists as to the interpretation of any question, the candidate is urged to submit with the answer paper a clear statement of any assumptions made.

MARKING: The marks for each question are indicated in parentheses immediately above each question. **Total marks** for the exam **equal 200**.

PART A: Questions 1 and 2 (40 marks for each question) .... **40 marks**
Answer ***either one of*** **Questions 1 and 2**.

PART B: Questions 3 (60 marks), 4 (50 marks) and 5 (50 marks) .... **160 marks**
Answer ***all parts of*** **Questions 3, 4 *and* 5**.

TOTAL MARKS .... **200 marks**

PART A (40 marks)

Instructions: Answer **EITHER ONE (1) of questions 1 and 2** in this part. Total marks for each question equal 40; marks for each part are given in parentheses.

(40 marks)

1. Consider 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 and 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, and N is sample size (the number of observations in the sample).

(20 marks)

(a) Stating explicitly all required assumptions, derive the expression (or formula) for $Var(\hat{\beta}_2)$, the variance of the OLS slope coefficient estimator $\hat{\beta}_2$. How do you compute an unbiased estimator of $Var(\hat{\beta}_2)$?

(20 marks)

(b) Give a general definition of a t-statistic. Starting from this definition, derive the t-statistic for $\hat{\beta}_2$ in OLS sample regression equation (2). State the assumptions required for the derivation.

(40 marks)

2. Consider the multiple linear regression model for which the population regression equation can be written in conventional notation as:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(1)}$$

where Y_i, X_{2i} and X_{3i} are observable variables; β_1, β_2 and β_3 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

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(2)}$$

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, $\hat{\beta}_3$ is the OLS estimator of the slope coefficient β_3, $\hat{u}_i$ is the OLS residual for the i-th sample observation, and N is sample size (the number of observations in the estimation sample).

(20 marks)

(a) Derive the OLS decomposition equation for $TSS \equiv \sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} (Y_i - \bar{Y})^2$, the total sum-of-squares of the observed Y_i values around their sample mean $\bar{Y}$ in sample regression equation (2). State the computational properties of the OLS sample regression equation on which the OLS decomposition equation depends.

(20 marks)

(b) Write an interpretive formula for $Var(\hat{\beta}_2)$, the variance of the OLS slope coefficient estimator $\hat{\beta}_2$ in sample regression equation (2). Use this formula to explain how each of the following factors affects $Var(\hat{\beta}_2)$:

(1) the error variance;

(2) the size of the estimation sample;

(3) the total sample variation of the X_{2i} values;

(4) the degree of linear dependence between the sample values X_{2i} and X_{3i}.

PART B (160 marks)

Instructions: Answer ***all*** **parts of questions 3, 4 and 5** in this part. Question 3 is worth a total of 60 marks. Questions 4 and 5 are each worth a total of 50 marks. Marks for each part are given in parentheses. **Show explicitly all formulas and calculations**.

(60 marks)

3. You are conducting an empirical investigation into the prices of houses in a single large metropolitan area. The sample data consist of 88 observations on the following observable variables:

P_i = the selling price of house i, in *thousands* of dollars;

HS_i = the house size of house i, in *hundreds* of square feet;

YS_i = the yard size of house i, in *hundreds* of square feet;

DC_i = an indicator variable defined such that $DC_i = 1$ if house i is a colonial-style house, and $DC_i = 0$ if house i is not a colonial-style house.

Your research assistant estimates two alternative regression models of house prices on the sample of N = 88 observations for recently sold houses in a single large metropolitan area. The estimation results for the two models are given below.

Model 1

$$P_i = \beta_1 + \beta_2 HS_i + \beta_3 YS_i + \beta_4 DC_i + u_{1i} \qquad \textbf{(1)}$$

where the β_j (j = 1, 2, 3, 4) are regression coefficients and u_{1i} is a random error term.

OLS Estimates of Equation (1):

$\hat{\beta}_1 = -5.2945$	$\hat{\beta}_2 = 13.236$	$\hat{\beta}_3 = 0.21117$	$\hat{\beta}_4 = 19.123$
(24.772)	(1.1361)	(0.064319)	(13.898)

$$RSS_{(1)} = \sum_{i=1}^{N} \hat{u}_{1i}^2 = 302371.03; \qquad TSS_{(1)} = \sum_{i=1}^{N} \left(P_i - \bar{P}\right)^2 = 917854.51; \qquad N = 88$$

where $RSS_{(1)}$ is the Residual Sum-of-Squares and $TSS_{(1)}$ is the Total Sum-of-Squares from OLS estimation of regression equation (1).

3. (continued)

The *Stata* listing of the estimated variance-covariance matrix for the coefficient estimates of regression equation (1) is:

```
symmetric VC1[4,4]
                 HS          YS          DC       _cons
   HS      1.290663
   YS    -.01339656    .00413694
   DC      -1.00957   -.00181495    193.1588
_cons    -24.081826   -.10212255  -113.40082   613.65797
```

Model 2

$$\ln P_i = \alpha_1 + \alpha_2 \ln HS_i + \alpha_3 \ln YS_i + \alpha_4 DC_i + u_{2i} \qquad \textbf{(2)}$$

where the α_j (j = 1, 2, 3, 4) are regression coefficients, $\ln X_i$ denotes the natural logarithm of the variable X_i, and u_{2i} is a random error term.

OLS Estimates of Equation (2):

$\hat{\alpha}_1 = 2.6389$	$\hat{\alpha}_2 = 0.75001$	$\hat{\alpha}_3 = 0.16811$	$\hat{\alpha}_4 = 0.066069$
(0.24431)	(0.080633)	(0.038150)	(0.042769)

$RSS_{(2)} = \sum_{i=1}^{N} \hat{u}_{2i}^2 = 2.843202$; $TSS_{(2)} = \sum_{i=1}^{N} \left(\ln P_i - \overline{\ln P}\right)^2 = 8.017604$; N = 88

where $RSS_{(2)}$ is the Residual Sum-of-Squares and $TSS_{(2)}$ is the Total Sum-of-Squares from OLS estimation of regression equation (2).

The *Stata* listing of the estimated variance-covariance matrix for the coefficient estimates of regression equation (2) is:

```
symmetric VC2[4,4]
               lnHS        lnYS          DC       _cons
 lnHS     .00650165
 lnYS    -.00095107    .00145545
   DC    -.00034231  -9.593e-06    .00182922
_cons    -.01496642   -.00342946  -.00021096   .05968928
```

(10 marks)

(a) Interpret each of the slope coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_4$ in regression equation (1); that is, explain in words what the numerical values of the slope coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_4$ mean. Interpret each of the slope coefficient estimates $\hat{\alpha}_2$ and $\hat{\alpha}_4$ in regression equation (2); that is, explain in words what the numerical values of the slope coefficient estimates $\hat{\alpha}_2$ and $\hat{\alpha}_4$ mean.

3. (continued)

(10 marks)

(b) Use the estimation results for regression equation (1) to test the *individual* significance of each of the slope coefficient estimates $\hat{\beta}_2$ for HS_i and $\hat{\beta}_4$ for DC_i. For each test, state the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. Which of these two slope coefficient estimates are individually significant at the 5 percent significance level? Which of these two slope coefficient estimates are individually significant at the 1 percent significance level?

(10 marks)

(c) Use the estimation results for regression equation (1) to test the *joint* significance of all the slope coefficient estimates at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(d) Use the estimation results for regression equation (2) to test the proposition that $\alpha_2 < 1$, i.e., to test the proposition that the marginal effect of $\ln HS_i$ on $\ln P_i$ is less than 1. Explain in words what this proposition means. Perform the test at the 1 percent significance level. State the null and alternative hypotheses. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(e) Use the estimation results for regression equation (2) to test the proposition that $\alpha_2 > \alpha_3$, i.e., to test the proposition that the marginal effect of $\ln HS_i$ on $\ln P_i$ is greater than the marginal effect of $\ln YS_i$ on $\ln P_i$. Explain in words what this proposition means. Perform the test at the 1 percent significance level. State the null and alternative hypotheses. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(f) Use the estimation results for regression equation (2) to test the proposition that $\alpha_2 + \alpha_3 = 1$, i.e., that the sum of the marginal effect of $\ln HS_i$ on $\ln P_i$ and the marginal effect of $\ln YS_i$ on $\ln P_i$ equals 1. Explain in words what this proposition means. Perform the test at the 5 percent significance level. State the null and alternative hypotheses. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(50 marks)

4. You are conducting an econometric investigation into the hourly wage rates of female employees. The sample data consist of 252 observations on the following variables:

W_i = the hourly wage rate of employee i, measured in dollars per hour;
ED_i = the number of years of formal education completed by employee i;
EXP_i = the number of years of work experience accumulated by employee i.

The regression model you propose to use is the log-lin (semi-log) regression equation

$$\ln W_i = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 ED_i^2 + \beta_5 EXP_i^2 + \beta_6 ED_i EXP_i + u_i \qquad \textbf{(1)}$$

where the β_j (j = 1, 2, …, 6) are regression coefficients, $\ln W_i$ denotes the natural logarithm of the variable W_i, and u_i is a random error term.

Using the sample data described above, your research assistant computes OLS estimates of regression equation (1) and of two restricted versions of equation (1). For each of the sample regression equations estimated on the sample of N = 252 observations, the following table contains the OLS coefficient estimates (with estimated standard errors in parentheses below the coefficient estimates) and the summary statistics RSS (residual sum-of-squares), TSS (total sum-of-squares), and N (number of sample observations).

4. (continued)

<u>Question 4</u>: OLS Sample Regression Equations for $\ln W_i$

Regressors		**(1)**	**(2)**	**(3)**
Intercept	$\hat{\beta}_1$	1.150 (0.4410)	1.162 (0.2399)	0.3348 (0.1415)
ED_i	$\hat{\beta}_2$	–0.08667 (0.05912)	–0.08817 (0.03822)	0.08231 (0.01046)
EXP_i	$\hat{\beta}_3$	0.02102 (0.01461)	0.02059 (0.006428)	0.004116 (0.001894)
ED_i^2	$\hat{\beta}_4$	0.007417 (0.002065)	0.007458 (0.001644)	---
EXP_i^2	$\hat{\beta}_5$	–0.0003852 (0.0001518)	–0.0003835 (0.000143)	---
ED_iEXP_i	$\hat{\beta}_6$	–0.000031 (0.0009307)	---	---
RSS =		35.4317	35.4319	39.6443
TSS =		49.5336	49.5336	49.5336
N =		252	252	252

<u>Note</u>: The symbol "----" means that the corresponding regressor was omitted from the estimated sample regression equation. The figures in parentheses below the coefficient estimates are the estimated standard errors. RSS is the Residual Sum-of-Squares, TSS is the Total Sum-of-Squares, and N is the number of sample observations.

(10 marks)

(a) Write the expression for the marginal effect of ED_i on $\ln W_i$ implied by regression equation (1). State the coefficient restrictions on regression equation (1) that make the marginal effect of ED_i on $\ln W_i$ equal to zero for all employees. Compute a test of the null hypothesis that the marginal effect of ED_i on $\ln W_i$ equals zero for all employees in regression equation (1). Perform the test at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$). State the null hypothesis H_0 and the alternative hypothesis H_1. Write the *restricted* regression equation implied by the null hypothesis H_0. OLS estimation of this *restricted* regression equation yields an **R-squared value of R^2 = 0.0428**. Use this information, together with the results from OLS estimation of equation (1), to calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

4. (continued)

(10 marks)

(b) State the coefficient restriction that regression equation (2) imposes on regression equation (1). Explain in words what the restriction means. Use the estimation results given above in the table to perform a test of this coefficient restriction at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. Would your inference be the same at the 10 percent significance level (i.e., for significance level $\alpha = 0.10$)? Based on the outcome of the test, which of the two regression equations would you choose, equation (1) or equation (2)?

(10 marks)

(c) Use the results from OLS estimation of regression equation (2) in column (2) to compute a two-tail test of the hypothesis that the marginal log-wage effect of ED_i equals zero for all women with 16 years of education (for whom $ED_i = 16$). Perform the test at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$). In addition to the information given above for the OLS estimates of regression equation (2), you are told that the estimated covariance of the coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_4$ in equation (2) is $\hat{Cov}(\hat{\beta}_2, \hat{\beta}_4) = -0.00006066$. State the null and alternative hypotheses. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(d) Use the results from OLS estimation of regression equation (2) in column (2) to compute an estimate of the conditional mean log-wage difference between female employees with 16 years of formal education and 3 years of work experience and female employees with 12 years of formal education and 3 years of work experience. That is, compute from the coefficient estimates for equation (2) in column (2) an estimate of

$$E(\ln W \mid ED_i = 16, EXP_i = 3) - E(\ln W \mid ED_i = 12, EXP_i = 3).$$

Explain in words the numerical value of the estimate you obtain. Write the formula you would use to compute the estimated variance of the conditional mean log-wage difference $E(\ln W \mid ED_i = 16, EXP_i = 3) - E(\ln W \mid ED_i = 12, EXP_i = 3)$ in regression equation (2); but note that you do not actually have to calculate this estimated variance.

4. (continued)

(10 marks)

(e) State the coefficient restriction(s) that regression equation (3) imposes on regression equation (1). Explain in words what the restriction(s) mean. Use the results given above from OLS estimation of regression equations (1) and (3) to perform a test of these coefficient restrictions at the 1 percent significance level (i.e., for significance level α = 0.01). State the null and alternative hypotheses, show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test. Based on the outcome of the test, which of the two regression equations would you choose, equation (1) or equation (3)?

(50 marks)

5. You are conducting an econometric investigation into the list prices of cars sold in North America in 1978. Your particular interest is in comparing the determinants of list prices for foreign and domestic cars. The sample data consist of 74 observations on the following variables:

P_i = the list price of car i, measured in *hundreds* of dollars;
WGT_i = the weight of car i, measured in *hundreds* of pounds;
F_i = 1 if car i is a foreign car, = 0 if car i is a domestic car.

The regression model of car prices you propose to use is the pooled regression model

$$P_i = \beta_1 + \beta_2 WGT_i + \beta_3 WGT_i^2 + \beta_4 F_i + \beta_5 F_i WGT_i + \beta_6 F_i WGT_i^2 + u_i \quad \textbf{(1)}$$

where the $\beta_j \ (j = 1, 2, \ldots, 6)$ are regression coefficients, and u_i is a random error term.

Using the sample data described above, your research assistant computes OLS estimates of regression equation (1) and OLS estimates of two restricted versions of equation (1). For each of the three sample regression equations estimated on the sample of N = 74 observations, the following table contains the OLS coefficient estimates (with estimated standard errors in parentheses below the coefficient estimates) and the summary statistics RSS (residual sum-of-squares), TSS (total sum-of-squares), and N (number of sample observations).

5. (continued)

<u>Question 5</u>: OLS Sample Regression Equations for Car Prices P_i

Regressors		**(1)**	**(2)**	**(3)**
Intercept	$\hat{\beta}_1$	159.34 (41.295)	55.267 (37.995)	134.19 (39.978)
WGT_i	$\hat{\beta}_2$	–9.8414 (2.5845)	–3.7146 (2.4334)	–7.2731 (2.6917)
WGT_i^2	$\hat{\beta}_3$	0.19851 (0.039583)	0.11211 (0.038311)	0.15142 (0.043373)
F_i	$\hat{\beta}_4$	–75.918 (126.55)	32.470 (6.4941)	----
$F_i WGT_i$	$\hat{\beta}_5$	3.4648 (9.9967)	---	----
$F_i WGT_i^2$	$\hat{\beta}_6$	0.032625 (0.19341)	----	---
	RSS =	21496.33	28352.19	38477.99
	TSS =	63506.54	63506.54	63506.54
	N =	74	74	74

<u>Note</u>: The symbol "----" means that the corresponding regressor was omitted from the estimated sample regression equation. The figures in parentheses below the coefficient estimates are the estimated standard errors. RSS is the Residual Sum-of-Squares, TSS is the Total Sum-of-Squares, and N is the number of sample observations.

(10 marks)
(a) Use the estimation results for regression equation (1) in column (1) to compute the OLS estimates of the intercept coefficient, the slope coefficient of WGT_i and the slope coefficient of WGT_i^2 for *domestic* cars. Use the estimation results for regression equation (1) in column (1) to compute the OLS estimates of the intercept coefficient, the slope coefficient of WGT_i and the slope coefficient of WGT_i^2 for *foreign* cars.

(10 marks)
(b) Compare the goodness-of-fit to the sample data of the three sample regression equations (1), (2) and (3) in the table. Calculate the value of an appropriate goodness-of-fit measure for each of the sample regression equations (1), (2) and (3) in the table. Which of the three sample regression equations provides the best fit to the sample data? Which of the three sample regression equations provides the worst fit to the sample data?

5. (continued)

(15 marks)

(c) Use the estimation results in the table to compute a *joint* test of the hypothesis that (1) the foreign-car intercept coefficient equals the domestic-car intercept coefficient, (2) the foreign-car slope coefficient of WGT_i equals the domestic-car slope coefficient of WGT_i, and (3) the foreign-car slope coefficient of WGT_i^2 equals the domestic-car slope coefficient of WGT_i^2. Perform the test at the 1 percent significance level. State the null and alternative hypotheses, and explain in words what the null hypothesis means. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

(15 marks)

(d) Write the expression (or formula) for the marginal effect of car weight (WGT_i) on car price (P_i) for *domestic* cars implied by pooled regression equation (1). Write the expression (or formula) for the marginal effect of car weight (WGT_i) on car price (P_i) for *foreign* cars implied by pooled regression equation (1). Use the estimation results in the table to compute a test of the hypothesis that the marginal effect of car weight on car price for *foreign* cars equals the marginal effect of car weight on car price for *domestic* cars. Perform the test at the 1 percent significance level (i.e., for significance level α = 0.01). State the null and alternative hypotheses. Show how you calculate the required test statistic, and state its null distribution. State the decision rule you use, and the inference you would draw from the test.

Selected Formulas

For the Simple (Two-Variable) Linear Regression Model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (i = 1, \ldots, N)$$

❑ Deviations from sample means are defined as:

$$y_i \equiv Y_i - \overline{Y}; \qquad x_i \equiv X_i - \overline{X};$$

where

$\overline{Y} = \Sigma_i Y_i / N = \dfrac{\Sigma_i Y_i}{N}$ is the sample mean of the Y_i values;

$\overline{X} = \Sigma_i X_i / N = \dfrac{\Sigma_i X_i}{N}$ is the sample mean of the X_i values.

❑ Formulas for the variance of the OLS intercept coefficient estimator $\hat{\beta}_1$ and the covariance of the OLS coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ in the two-variable linear regression model:

$$\mathrm{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2 \Sigma_i X_i^2}{N\,\Sigma_i \left(X_i - \overline{X}\right)^2} = \frac{\sigma^2 \Sigma_i X_i^2}{N\,\Sigma_i x_i^2};$$

$$\mathrm{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) = -\overline{X}\left(\frac{\sigma^2}{\Sigma_i x_i^2}\right).$$

❑ Formulas for the variance of the conditional predictor $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$:

- When $\hat{Y}_0$ is used as a mean predictor of $E\left(Y_0 \mid X_0\right) = \beta_1 + \beta_2 X_0$,

$$\mathrm{Var}\left(\hat{Y}_0^m\right) = \sigma^2\left[\frac{1}{N} + \frac{\left(X_0 - \overline{X}\right)^2}{\Sigma_i x_i^2}\right].$$

- When $\hat{Y}_0$ is used as an individual predictor of $Y_0 \mid X_0 = \beta_1 + \beta_2 X_0 + u_0$,

$$\mathrm{Var}\left(\hat{Y}_0\right) = \sigma^2 + \sigma^2\left[\frac{1}{N} + \frac{\left(X_0 - \overline{X}\right)^2}{\Sigma_i x_i^2}\right].$$

Selected Formulas (continued)

For the Multiple (Three-Variable) Linear Regression Model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (i = 1, \ldots, N)$$

- Deviations from sample means are defined as:

$$y_i \equiv Y_i - \overline{Y}; \qquad x_{2i} \equiv X_{2i} - \overline{X}_2; \qquad x_{3i} \equiv X_{3i} - \overline{X}_3;$$

where

$\overline{Y} = \Sigma_i Y_i / N = \dfrac{\Sigma_i Y_i}{N}$ is the sample mean of the Y_i values;

$\overline{X}_{2i} = \Sigma_i X_{2i} / N = \dfrac{\Sigma_i X_{2i}}{N}$ is the sample mean of the X_{2i} values;

$\overline{X}_{3i} = \Sigma_i X_{3i} / N = \dfrac{\Sigma_i X_{3i}}{N}$ is the sample mean of the X_{3i} values.

- The OLS slope coefficient estimators $\hat{\beta}_2$ and $\hat{\beta}_3$ in deviation-from-means form are:

$$\hat{\beta}_2 = \frac{\left(\Sigma_i x_{3i}^2\right)\left(\Sigma_i x_{2i} y_i\right) - \left(\Sigma_i x_{2i} x_{3i}\right)\left(\Sigma_i x_{3i} y_i\right)}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2};$$

$$\hat{\beta}_3 = \frac{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i} y_i\right) - \left(\Sigma_i x_{2i} x_{3i}\right)\left(\Sigma_i x_{2i} y_i\right)}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2}.$$

- Formulas for the variances and covariances of the slope coefficient estimators $\hat{\beta}_2$ and $\hat{\beta}_3$:

$$\mathrm{Var}\left(\hat{\beta}_2\right) = \frac{\sigma^2 \Sigma_i x_{3i}^2}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2};$$

$$\mathrm{Var}\left(\hat{\beta}_3\right) = \frac{\sigma^2 \Sigma_i x_{2i}^2}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2};$$

$$\mathrm{Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right) = \frac{\sigma^2 \Sigma_i x_{2i} x_{3i}}{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{3i}^2\right) - \left(\Sigma_i x_{2i} x_{3i}\right)^2}.$$

Percentage Points of the t-Distribution

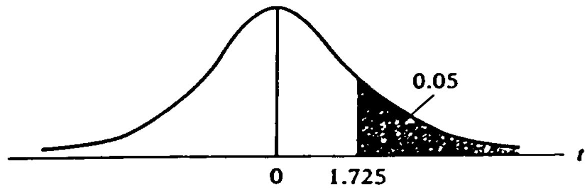
TABLE D.2
Percentage points of the *t* distribution

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	Pr	df for numerator N_1: 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.