QUEEN'S UNIVERSITY AT KINGSTON
Department of Economics

CONFIDENTIAL
turn in exam
question paper

ECONOMICS 351* - Winter Term 2008

Introductory Econometrics

Winter Term 2008 **FINAL EXAMINATION** M.G. Abbott

DATE: **Tuesday April 15, 2008**

TIME: **Three (3) hours (180 minutes); 9:00 a.m. – 12 noon**

INSTRUCTIONS: The examination is divided into two parts.

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

PART B contains **3** questions; students are required to **answer ALL THREE** questions in Part B.

- Answer all questions in the exam booklets provided. Be sure **your *student number*** is printed clearly and legibly on the front page 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.
- 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.
- **Please write legibly.** GOOD LUCK!

The candidate is urged to submit with the answer paper a clear statement of any assumptions made if doubt exists as to the interpretation of any question that requires a written answer.

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 (25 marks for each question) .... **25 marks**
Answer ***either one of*** **Questions 1 and 2**.

PART B: Questions 3 (105 marks), 4 (40 marks) 5 (30 marks) .... **175 marks**
Answer ***all parts of*** **Questions 3, 4 *and* 5**.

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

PART A (25 marks)

Instructions: Answer **EITHER ONE (1) of questions 1 and 2** in this part. Total marks for each question equal 25; marks for each part are given in parentheses.

(25 marks)

1. Consider the simple (two-variable) linear regression model for which the population regression equation can be written in conventional notation as:

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

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).

(17 marks)

(a) Prove that the OLS slope coefficient estimator $\hat{\beta}_1$ is an unbiased estimator of the slope coefficient β_1. Include in your answer a definition of the unbiasedness property. State those assumptions of the classical linear regression model that are required to prove the unbiasedness of the OLS estimator $\hat{\beta}_1$.

(4 marks)

(b) The significance level of a statistical test of a null hypothesis H_0 against an alternative hypothesis H_1 is defined as:

(1) the probability of rejecting H_0 when H_0 is false;
(2) the probability of rejecting H_0 when H_0 is true;
(3) the probability of not rejecting H_0 when H_0 is false;
(4) the probability of not rejecting H_0 when H_0 is true.

(4 marks)

(c) The p-value of a calculated t-statistic or F-statistic is best defined as:

(1) the lowest significance level at which the null hypothesis H_0 can be rejected;
(2) the probability that the null hypothesis H_0 is true;
(3) the probability that the null hypothesis H_0 is false;
(4) the highest significance level at which the null hypothesis H_0 can be rejected.

(25 marks)

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

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad \textbf{(1)}$$

where Y_i, X_{1i} and X_{2i} are observable variables; β_0, 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

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{u}_i \qquad (i = 1, ..., N) \qquad \textbf{(2)}$$

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, $\hat{\beta}_2$ 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 estimation sample).

(17 marks)

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

(8 marks)

(b) Define the p-value of the calculated t-statistic for $\hat{\beta}_2$ when the null hypothesis H_0: $\beta_2 = 0$ is tested against each of the following three alternative hypotheses:

(1) H_1: $\beta_2 \neq 0$;

(2) H_1: $\beta_2 < 0$.

PART B (175 marks)

Instructions: Answer ***all*** **parts of questions 3, 4 and 5** in this part. Question 3 is worth a total of 105 marks. Question 4 is worth a total of 40 marks. Question 5 is worth a total of 30 marks. Marks for each part are given in parentheses. **Show explicitly all formulas and calculations**.

(105 marks)

3. You are investigating the relationship between the selling prices of houses in a single large metropolitan area and the level of nitrogen oxide air pollution. You have sample data for 506 census tracts (small geographical areas) in a large metropolitan area on the following variables:

$PRICE_i$ = the median selling price of houses in the i-th census tract, measured in *thousands of dollars*;
NOX_i = level of nitrogen oxide concentration in the air of the i-th census tract, measured in *parts per 100 million*;
$PTAX_i$ = property taxes per \$1,000 of assessed property value in the i-th census tract, measured in *dollars per \$1,000*;
$DIST_i$ = weighted distance of the i-th census tract from five employment centers in the metropolitan area, measured in *miles*;
$ROOMS_i$ = average number of rooms in houses located in the i-th census tract, measured in *rooms*.

Using the given sample data, your trusty research assistant estimates three different regression equations and reports the following estimation results (with estimated *standard errors* given in parentheses below the regression coefficient estimates):

$$PRICE_i = \beta_0 + \beta_1 NOX_i + \beta_2 PTAX_i + \beta_3 DIST_i + \beta_4 ROOMS_i + u_{1i} \qquad \textbf{(1)}$$

$\hat{\beta}_0 = -90.176$ (40.360) $\hat{\beta}_1 = -16.978$ (4.1655) $\hat{\beta}_2 = -1.3060$ (0.21545) $\hat{\beta}_3 = -7.2207$ (2.0000) $\hat{\beta}_4 = 78.013$ (4.0409)

$RSS_{(1)} = 1{,}817{,}577.05$ $TSS_{(1)} = 4{,}282{,}553.21$ $N = 506$

$$\ln PRICE_i = \alpha_0 + \alpha_1 \ln NOX_i + \alpha_2 \ln PTAX_i + \alpha_3 \ln DIST_i + \alpha_4 ROOMS_i + u_{2i} \qquad \textbf{(2)}$$

$\hat{\alpha}_0 = 5.8647$ (0.30057) $\hat{\alpha}_1 = -0.61070$ (0.12685) $\hat{\alpha}_2 = -0.31605$ (0.041126) $\hat{\alpha}_3 = -0.12463$ (0.043810) $\hat{\alpha}_4 = 0.28647$ (0.018107)

$RSS_{(2)} = 36.44308$ $TSS_{(2)} = 84.58221$ $N = 506$

3. (continued)

$$\ln PRICE_i = \gamma_0 + \gamma_1 NOX_i + \gamma_2 PTAX_i + \gamma_3 DIST_i + \gamma_4 ROOMS_i + u_{3i} \qquad \textbf{(3)}$$

$\hat{\gamma}_0 = 4.3748$	$\hat{\gamma}_1 = -0.080099$	$\hat{\gamma}_2 = -0.0077925$	$\hat{\gamma}_3 = -0.020440$	$\hat{\gamma}_4 = 0.28665$
(0.18045)	(0.018624)	(0.0009633)	(0.0089418)	(0.018067)

$RSS_{(3)} = 36.33319$ $TSS_{(3)} = 84.58221$ N = 506

$RSS_{(n)}$ is the Residual Sum-of-Squares and $TSS_{(n)}$ is the Total Sum-of-Squares for sample regression equation (n), where n = 1, 2, 3. $\ln X_i$ denotes the natural logarithm of the variable X_i. Sample size N = 506. Estimated *standard errors* are given in parentheses below the coefficient estimates.

(12 marks)

(a) Interpret the slope coefficient estimate $\hat{\beta}_1$ in sample regression equation (1), the slope coefficient estimate $\hat{\alpha}_1$ in sample regression equation (2), and the slope coefficient estimate $\hat{\gamma}_1$ in sample regression equation (3). That is, explain in words what the numerical values of the slope coefficient estimates $\hat{\beta}_1$, $\hat{\alpha}_1$ and $\hat{\gamma}_1$ mean.

(8 marks)

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

(10 marks)

(c) Use the estimation results given above for regression equations (2) and (3) to test the *individual* significance of the slope coefficient estimate $\hat{\alpha}_1$ for $\ln NOX_i$ in equation (2) and the *individual* significance of the slope coefficient estimate $\hat{\gamma}_1$ for NOX_i in equation (3). 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?

(8 marks)

(d) Use the estimation results given above for regression equation (1) to test the proposition that house prices are negatively related to the level of nitrogen oxide pollution in the air. 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 at the 5 percent significance level. Would your inference be the same at the 1 percent significance level?

3. (continued)

(12 marks)

(e) Use the above estimation results to compute the value of R^2 for OLS sample regression equations (1), (2) and (3). For which of the regression equations (1), (2) and (3) can the values of R^2 be used to compare the goodness-of-fit to the sample data? Explain.

(10 marks)

(f) Use the estimation results for regression equation (1) to test the proposition that the explanatory variable ROOMS_i is positively related to house prices PRICE_i, i.e., that median house prices are higher in census tracts with larger houses (more rooms). Perform the test 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)

(g) Use the estimation results for regression equation (3) to test the *joint* significance of the slope coefficient estimates $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{\gamma}_3$ and $\hat{\gamma}_4$ at the 1 percent significance level (i.e., for significance level $\alpha = 0.01$); i.e., test the proposition that all slope coefficients in regression equation (3) are jointly equal to zero. 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.

(15 marks)

(h) Use the estimation results for regression equation (2) to compute the two-sided 95 percent confidence interval for the slope coefficient α_1 of $\ln \mathrm{NOX}_i$. Use this confidence interval to perform a two-tail test of the hypothesis that $\alpha_1 = -0.75$ at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the null and alternative hypotheses, and explain in words what the null hypothesis means. State the decision rule you use, and the inference you would draw from the test.

(10 marks)

(i) Use the estimation results for regression equation (3) to perform a two-tail test of the proposition that $\gamma_1 = -0.10$ at the 5 percent significance level. Explain in words what this proposition means. 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 at the 5 percent significance level. Would your inference be the same at the 10 percent significance level?

3. (continued)

(10 marks)

(j) Use the estimation results for regression equation (2) to perform a two-tail test of the null hypothesis $\alpha_1 = \alpha_2$ at the 5 percent significance level. Explain in words what this null hypothesis means. The estimated covariance of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ is $\hat{Cov}(\hat{\alpha}_1, \hat{\alpha}_2) = -0.0018268$. 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 at the 5 percent significance level.

(40 marks)

4. You are conducting an econometric investigation into the hourly wage rates of female and male employees in the Toronto metropolitan area. The sample data consist of observations for 13,118 employees on the following variables:

W_i = the hourly wage rate of the i-th employee, in dollars per hour;

ED_i = years of formal education completed by the i-th employee, in years;

AGE_i = age of the i-th employee, in years;

F_i = an indicator variable defined such that $F_i = 1$ if the i-th employee is female, and $F_i = 0$ if the i-th employee is male.

The regression model you propose to use is

$$W_i = \beta_0 + \beta_1 ED_i + \beta_2 AGE_i + \beta_3 AGE_i^2 + \beta_4 ED_i AGE_i + \beta_5 F_i + \beta_6 F_i ED_i + \beta_7 F_i AGE_i + \beta_8 F_i AGE_i^2 + \beta_9 F_i ED_i AGE_i + u_i \qquad \textbf{... (1)}$$

where the β_j (j = 0, 1, 2, …, 9) 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 of two restricted versions of equation (1). For each of the sample regression equations estimated on the sample data, 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)

Question 4: OLS Sample Regression Equations for W_i
(*standard errors* in parentheses)

Regressors		(1)	(2)	(3)
Intercept	$\hat{\beta}_0$	–23.07 (3.918)	–27.23 (3.375)	–18.93 (3.013)
ED_i	$\hat{\beta}_1$	0.01365 (0.1906)	0.2341 (0.1464)	0.3359 (0.1472)
AGE_i	$\hat{\beta}_2$	1.436 (0.1436)	1.594 (0.1332)	1.263 (0.1088)
AGE_i^2	$\hat{\beta}_3$	–0.01755 (0.001387)	–0.01828 (0.001374)	–0.01552 (0.001029)
ED_iAGE_i	$\hat{\beta}_4$	0.03656 (0.004385)	0.02940 (0.003403)	0.02732 (0.003424)
F_i	$\hat{\beta}_5$	1.416 (6.077)	12.48 (3.439)	–6.551 (0.1962)
F_iED_i	$\hat{\beta}_6$	0.6049 (0.2980)	----	----
F_iAGE_i	$\hat{\beta}_7$	–0.2699 (0.2186)	–0.6792 (0.1689)	----
$F_iAGE_i^2$	$\hat{\beta}_8$	0.003030 (0.002054)	0.004784 (0.001992)	----
$F_iED_iAGE_i$	$\hat{\beta}_9$	–0.01935 (0.006962)	----	----
	RSS =	1,613,827.86	1,615,838.22	1,638,981.08
	TSS =	2,159,012.05	2,159,012.05	2,159,012.05
	N =	13,118	13,118	13,118

Note: 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.

(14 marks)

(a) 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?

4. (continued)

(10 marks)

(b) Use the estimation results for regression equation (3) in the above table to perform a test of the proposition that female employees of any given education and age have lower average hourly wages than male employees of the same education and age. 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. 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)

(c) State the coefficient restrictions that regression equation (2) in the table imposes on regression equation (1). Explain in words what the restrictions mean. Use the estimation results given in the table to perform a test of these coefficient restrictions 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 1 percent significance level (i.e., for significance level $\alpha = 0.01$)? Based on the outcome of the test, which of the two regression equations would you choose, equation (1) or equation (2)?

(6 marks)

(d) Use the estimation results for regression equation (1) in the above table to compute an estimate of the female-male difference in mean hourly wages for employees who have 16 years of formal education and who are 30 years of age. That is, use the estimates of regression equation (1) to compute an estimate of

$$E\left(W_i \middle| F_i = 1, ED_i = 16, AGE_i = 30\right) - E\left(W_i \middle| F_i = 0, ED_i = 16, AGE_i = 30\right)$$

<u>Question 5</u> begins on the next page.

(30 marks)

5. You are investigating the relationship between the final exam grades of university students in an introductory economics course and those students' class attendance, as measured by the percentage of classes each student attended during the term. You also have sample data on two additional explanatory variables: each student's cumulative GPA (Grade Point Average) prior to the term in which the introductory economics course was taken; and each student's score on a standardized college entrance exam, the ACT exam. You have sample data for 680 students on the following variables:

$finalpct_i$ = final exam grade of the i-th student, measured in *percentage points*;

$attrate_i$ = percentage of classes attended by the i-th student during the term, measured in *percentage points*;

GPA_i = cumulative Grade Point Average (GPA) of the i-th student prior to the term in which the introductory economics course was taken, measured *out of 4.0*;

ACT_i = ACT score of the i-th student on the ACT college entrance exam, measured in *points*.

Using the given sample data on 680 students, your trusty research assistant has estimated regression equation (1) and obtained the following estimation results (with estimated *standard errors* given in parentheses below the coefficient estimates):

$$finalpct_i = \beta_0 + \beta_1 attrate_i + \beta_2 GPA_i + \beta_3 GPA_i^2 + \beta_4 GPA_i attrate_i + \beta_5 ACT_i + u_i \qquad \ldots \textbf{(1)}$$

$\hat{\beta}_0 = 62.261$ (10.026) $\hat{\beta}_1 = -0.061858$ (0.12175) $\hat{\beta}_2 = -20.117$ (5.7271)

$\hat{\beta}_3 = 3.8356$ (1.1959) $\hat{\beta}_4 = 0.057852$ (0.051337) $\hat{\beta}_5 = 0.90182$ (0.13359)

$\hat{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 0.13006$ $\hat{Cov}(\hat{\beta}_1, \hat{\beta}_3) = 0.069642$ $\hat{Cov}(\hat{\beta}_1, \hat{\beta}_4) = -0.0060842$

$\hat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = -5.0408$ $\hat{Cov}(\hat{\beta}_2, \hat{\beta}_4) = -0.066730$ $\hat{Cov}(\hat{\beta}_2, \hat{\beta}_5) = 0.073538$

$RSS = 73080.413$ $TSS = 94137.169$ N = 680

RSS is the Residual Sum-of-Squares and TSS is the Total Sum-of-Squares for sample regression equation (1). Sample size N = 680. $\hat{Cov}(\hat{\beta}_j, \hat{\beta}_h)$ is the estimated covariance between coefficient estimates $\hat{\beta}_j$ and $\hat{\beta}_h$. Estimated *standard errors* are given in parentheses below the coefficient estimates $\hat{\beta}_j$ (j = 0, 1, …, 5).

5. (continued)

(10 marks)

(a) Compute a test of the proposition that class attendance as measured by attrate_i, the percentage of classes attended by the i-th student during the term, has no effect on students' final exam grade. Perform the test at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the coefficient restrictions on regression equation (1) implied by this proposition; that is, 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 a Residual Sum-of-Squares value of **RSS = 73934.237**. 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 at the 5 percent significance level.

(10 marks)

(b) Compute a test of the proposition that students' cumulative Grade Point Average as measured by GPA_i has no effect on students' final exam grade. Perform the test at the 5 percent significance level (i.e., for significance level $\alpha = 0.05$). State the coefficient restrictions on regression equation (1) implied by this proposition; that is, 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 $\mathbf{R^2 = 0.1701}$. 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 at the 5 percent significance level.

(10 marks)

(c) The average student in the course attends 80 percent of the classes and has a cumulative Grade Point Average of 2.6; that is, the sample mean value of attrate_i equals 80, and the sample mean value of GPA_i equals 2.6. Write the expression (or formula) for the marginal effect on finalpct_i of attrate_i implied by regression equation (1). Use the estimation results for regression equation (1) to test the proposition that the marginal effect of class attendance (attrate_i) on students' final exam grade (finalpct_i) equals zero for the average student whose class attendance rate is 80 percent and GPA is 2.6. 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 implied by this proposition. 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 at the 1 percent significance level.

Selected Formulas

For the Simple (Two-Variable) Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (i = 1, \ldots, N)$$

- 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 $\hat{\beta}_0$ and the covariance of the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ in the two-variable linear regression model:

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2 \Sigma_i X_i^2}{N \Sigma_i (X_i - \overline{X})^2} = \frac{\sigma^2 \Sigma_i X_i^2}{N \Sigma_i x_i^2};$$

$$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\overline{X}\left(\frac{\sigma^2}{\Sigma_i x_i^2}\right).$$

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

  - When $\hat{Y}_0$ is used as a mean predictor of $E(Y_0 | X_0) = \beta_0 + \beta_1 X_0$,

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

  - When $\hat{Y}_0$ is used as an individual predictor of $Y_0 | X_0 = \beta_0 + \beta_1 X_0 + u_i$,

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

Selected Formulas (continued)

For the Multiple (Three-Variable) Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (i = 1, \ldots, N)$$

- Deviations from sample means are defined as:

$$y_i \equiv Y_i - \overline{Y}; \qquad x_{1i} \equiv X_{1i} - \overline{X}_1; \qquad x_{2i} \equiv X_{2i} - \overline{X}_2;$$

where

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

$\overline{X}_{1i} = \Sigma_i X_{1i} / N = \dfrac{\Sigma_i X_{1i}}{N}$ is the sample mean of the X_{1i} values;

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

- The OLS slope coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ and in deviation-from-means form are:

$$\hat{\beta}_1 = \frac{\left(\Sigma_i x_{2i}^2\right)\left(\Sigma_i x_{1i} y_i\right) - \left(\Sigma_i x_{1i} x_{2i}\right)\left(\Sigma_i x_{2i} y_i\right)}{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i}^2\right) - \left(\Sigma_i x_{1i} x_{2i}\right)^2};$$

$$\hat{\beta}_2 = \frac{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i} y_i\right) - \left(\Sigma_i x_{1i} x_{2i}\right)\left(\Sigma_i x_{1i} y_i\right)}{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i}^2\right) - \left(\Sigma_i x_{1i} x_{2i}\right)^2}.$$

- Formulas for the variances and covariances of the slope coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$\mathrm{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2 \Sigma_i x_{2i}^2}{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i}^2\right) - \left(\Sigma_i x_{1i} x_{2i}\right)^2};$$

$$\mathrm{Var}\left(\hat{\beta}_2\right) = \frac{\sigma^2 \Sigma_i x_{1i}^2}{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i}^2\right) - \left(\Sigma_i x_{1i} x_{2i}\right)^2};$$

$$\mathrm{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) = \frac{\sigma^2 \Sigma_i x_{1i} x_{2i}}{\left(\Sigma_i x_{1i}^2\right)\left(\Sigma_i x_{2i}^2\right) - \left(\Sigma_i x_{1i} x_{2i}\right)^2}.$$

Percentage Points of the t-Distribution

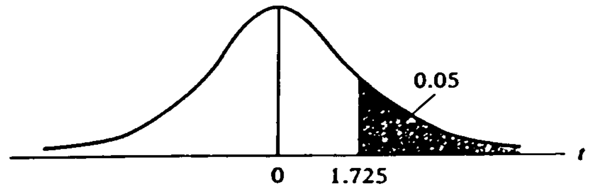
TABLE D.2
Percentage points of the *t* distribution

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
	.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
22	.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
	.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
24	.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
	.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
26	.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
	.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
28	.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
	.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
30	.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
	.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
40	.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
	.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
60	.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
	.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
120	.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
	.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
200	.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.