QUEEN'S UNIVERSITY AT KINGSTON
Department of Economics

ECONOMICS 351* - Section A

Introductory Econometrics

Fall Term 1999 **FINAL EXAMINATION** M.G. Abbott

DATE: **Friday December 17, 1999.**

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 three questions; students are required to answer **ANY TWO** of the three questions 1, 2, and 3 in Part A.

PART B contains two questions; students are required to answer **BOTH** of the two questions 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, 2, and 3 (40 marks for each question) Answer ***any two of*** **Questions 1, 2, and 3**.	....	**80 marks**
PART B:	Question 4 (60 marks) and Question 5 (60 marks) Answer ***all parts of both*** **Questions 4 *and* 5**.	....	**120 marks**
TOTAL MARKS	....		**200 marks**

PART A (80 marks)

Instructions: Answer **ANY TWO (2)** of the three questions 1, 2, and 3 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).

(15 marks)

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

(15 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.

(10 marks)

(c) Starting with the t-statistic for $\hat{\beta}_2$, derive the two-sided $100(1-\alpha)$ percent confidence interval for the slope coefficient β_2. Define all terms required for the derivation. Explain how the confidence interval for β_2 is interpreted.

(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 sample).

(15 marks)

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

(15 marks)

(b) 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.

(10 marks)

(c) Explain how you would use the estimation results for the OLS sample regression equation (2) to test the *joint* significance of the slope coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_3$. State the null and alternative hypotheses, give the formula for the appropriate test statistic, and state its null distribution. Define the p-value for the calculated test statistic. Explain how you would use the p-value for the calculated test statistic to decide between rejection and nonrejection of the null hypothesis at significance level α.

(40 marks)

3. You are conducting an empirical investigation into the beer expenditures of male and female office employees of a company. You have assembled for this project a random sample of observations on 40 office employees, 21 of whom are females and 19 of whom are males. The sample data provide observations on the following observable variables:

BE_i = the annual beer expenditures of employee i, measured in dollars per year;

INC_i = the annual income of employee i, in thousands of dollars per year;

AGE_i = the age of employee i, in years;

F_i = a female indicator variable, defined such that $F_i = 1$ if employee i is female and $F_i = 0$ if employee i is male.

The beer expenditure equation for ***female*** **employees** can be written as:

$$BE_i = \alpha_1 + \alpha_2 INC_i + \alpha_3 AGE_i + u_{fi} \qquad \textbf{(1a)}$$

where the α_j (j = 1, 2, 3) are the female regression coefficients.

The beer expenditure equation for ***male*** **employees** can be written as:

$$BE_i = \beta_1 + \beta_2 INC_i + \beta_3 AGE_i + u_{mi} \qquad \textbf{(1b)}$$

where the β_j (j = 1, 2, 3) are the male regression coefficients.

Your research assistant provides you with OLS estimates of the following pooled beer expenditure equation, and some restricted variants thereof, on the combined sample of all 40 employees:

$$BE_i = \theta_1 + \theta_2 INC_i + \theta_3 AGE_i + \theta_4 F_i + \theta_5 F_i\, INC_i + \theta_6 F_i\, AGE_i + u_i \qquad \textbf{(2)}$$

where the θ_j (j = 1, …, 6) are the regression coefficients in pooled equation (2). OLS estimation results for the pooled beer expenditure equations on the full sample of N = 40 employees are given in the following table (with estimated standard errors given in parentheses below the coefficient estimates).

3. (continued)

<u>Question 3</u>: Pooled OLS Sample Regression Equations for BE_i

Regressors		(2)	(3)	(4)
Intercept	$\hat{\theta}_1$	451.36 (63.945)	461.86 (51.344)	342.88 (72.343)
INC_i	$\hat{\theta}_2$	3.5729 (0.68599)	2.4109 (0.40316)	2.3822 (0.60359)
AGE_i	$\hat{\theta}_3$	−9.3632 (1.9155)	−8.1828 (1.5501)	−7.5756 (2.3170)
F_i	$\hat{\theta}_4$	−208.34 (96.508)	−190.25 (27.768)	----
$F_i\, INC_i$	$\hat{\theta}_5$	−1.7731 (0.84823)	----	----
$F_i\, AGE_i$	$\hat{\theta}_6$	2.8337 (3.0871)	----	----
	RSS =	244466.5	275888.8	635636.7
	TSS =	947651.9	947651.9	947651.9
	N =	40	40	40

<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 OLS estimation results for pooled regression equation (2) in column (2) of the table to compute the values of the OLS coefficient estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$ for the female beer expenditure equation (1a), and the values of the OLS coefficient estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ for the male beer expenditure equation (1b).

(10 marks)

(b) Use the OLS estimation results for pooled regression equation (2) in column (2) of the table to test the proposition that $\alpha_2 < \beta_2$, i.e., to test the proposition that the marginal effect of income INC_i on beer expenditure BE_i is smaller for females than for males. 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.

3. (continued)

(10 marks)
(c) Use the estimation results in the table to compute a test of the following hypothesis:

$$H_0: \alpha_j = \beta_j \quad \forall\ j = 1, 2, 3$$
$$H_1: \alpha_j \neq \beta_j \quad j = 1, 2, 3.$$

Perform the test at the 1 percent significance level. 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. What would you conclude from the results of the test about the similarity of the male and female beer expenditure functions?

(10 marks)
(d) Use the estimation results in the table to compute a *joint* test of the hypothesis that (1) the male and female slope coefficients for INC_i are equal and (2) the male and female slope coefficients for AGE_i are equal. 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.

PART B (120 marks)

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

(60 marks)

4. You are conducting an econometric investigation into the annual salaries of CEOs (Chief Executive Officers) of major private companies. The sample data consist of observations for 177 private firms on the following variables:

$SALARY_i$ = the annual salary of the CEO of firm i, measured in thousands of dollars;
TR_i = the annual total sales revenues of firm i, measured in millions of dollars;
$VALUE_i$ = the market value of firm i, measured in millions of dollars;
TEN_i = the number of years the CEO has been employed with firm i;
AGE_i = the age of the CEO of firm i, in years.

The regression model of CEO salaries you propose to use is:

$$\ln SALARY_i = \beta_1 + \beta_2 \ln TR_i + \beta_3 \ln VALUE_i + \beta_4 TEN_i + \beta_5 TEN_i^2 + \beta_6 AGE_i + \beta_7 AGE_i^2 + u_i \qquad \textbf{(1)}$$

where the $\beta_j\ (j = 1, 2, \ldots, 7)$ are regression coefficients, $\ln X_i$ denotes the natural logarithm of the variable X_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 OLS estimates of two restricted versions of equation (1). For each of the three sample regression equations estimated on the sample of N = 177 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 SALARY$_i$

Regressors		(1)	(2)	(3)
Intercept	$\hat{\beta}_1$	5.572 (1.144)	4.369 (0.2587)	6.060 (1.152)
$\ln TR_i$	$\hat{\beta}_2$	0.1820 (0.04120)	0.1646 (0.03864)	0.1800 (0.04229)
$\ln VALUE_i$	$\hat{\beta}_3$	0.1023 (0.04928)	0.1085 (0.04883)	0.09887 (0.05054)
TEN_i	$\hat{\beta}_4$	0.04593 (0.01422)	0.04512 (0.01412)	----
TEN_i^2	$\hat{\beta}_5$	–0.001219 (0.0004764)	–0.001210 (0.0004747)	----
AGE_i	$\hat{\beta}_6$	–0.04201 (0.04122)	----	–0.05334 (0.04152)
AGE_i^2	$\hat{\beta}_7$	0.0003324 (0.0003621)	----	0.0004607 (0.0003613)
	RSS =	42.060	42.474	44.879
	TSS =	64.646	64.646	64.646
	N =	177	177	177

<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) Interpret the slope coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ in sample regression equation (1); that is, explain what the numerical values of the slope coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ mean. Use the results for sample regression equation (1) to test the *individual* significance of each of the slope coefficient estimates $\hat{\beta}_2$ for $\ln TR_i$ and $\hat{\beta}_3$ for $\ln VALUE_i$. 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?

4. (continued)

(10 marks)

(b) Use the estimation results for regression equation (1) in column (1) of the table 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)

(c) Use the estimation results for regression equation (1) in column (1) of the table to test the hypothesis $\beta_2 = \beta_3$. Explain in words what this hypothesis means. Perform this test at the 5 percent significance level. In addition to the information given in column (1) of the table, your research assistant tells you that the estimated covariance of the coefficient estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ is $\hat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.001473$. 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 (1) in column (1) of the table to test the proposition that $\beta_3 > 0$ at the 1 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. Is your inference the same at the 5 percent significance level as it is at the 1 percent significance level?

(10 marks)

(e) State the coefficient restrictions that are imposed on regression equation (1) in estimating sample regression equation (2) in column (2) of the table. Perform a test of these coefficient restrictions at the 5 percent significance level. State the null and alternative hypotheses, and explain 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. Is your inference the same at the 10 percent significance level as it is at the 5 percent significance level? Based on the outcome of the test, would you choose equation (2) or equation (1)?

(10 marks)

(f) State the coefficient restrictions that are imposed on regression equation (1) in estimating sample regression equation (3) in column (3) of the table. Perform a test of these coefficient restrictions at the 1 percent significance level. State the null and alternative hypotheses, and explain 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. Based on the outcome of the test, would you choose equation (3) or equation (1)?

(60 marks)

5. You are conducting an econometric investigation into the list prices of cars sold in North America in 1978. The sample data consist of 74 observations on the following variables:

P_i = the list price of car i, measured in dollars;
WT_i = the weight of car i, measured in pounds;
FE_i = the fuel efficiency of car i, measured in miles per gallon.

Using the sample data described above, you compute OLS estimates of two log-log (double log) regression equations. The first is

$$\ln P_i = \beta_1 + \beta_2 \ln WT_i + \beta_3 \ln FE_i + \beta_4 (\ln WT_i)^2 + \beta_5 (\ln FE_i)^2 + \beta_6 (\ln WT_i)(\ln FE_i) + u_{1i} \quad \textbf{(1)}$$

where the $\beta_j\,(j = 1, 2, \ldots, 6)$ are regression coefficients, $\ln X_i$ denotes the natural logarithm of the variable X_i, and u_{1i} is a random error term. The second regression equation is

$$\ln P_i = \beta_1 + \beta_2 \ln WT_i + \beta_3 \ln FE_i + u_{2i} \quad \textbf{(2)}$$

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

OLS estimation of regression equation (1) on the sample of N = 74 observations yields the following OLS coefficient estimates (with estimated standard errors in parentheses below the coefficient estimates) and summary statistics:

OLS Estimates of Equation (1)

	Constant (β_1)	$\ln WT_i$ (β_2)	$\ln FE_i$ (β_3)	$(\ln WT_i)^2$ (β_4)	$(\ln FE_i)^2$ (β_5)	$(\ln WT_i)(\ln FE_i)$ (β_6)
$\hat{\beta}_j$ $\left(\widehat{se}(\hat{\beta}_j)\right)$	480.0 (118.6)	–86.77 (22.17)	–82.06 (22.41)	4.144 (1.071)	4.377 (1.153)	6.843 (1.980)

Summary Statistics -- OLS Estimates of Equation (1):

$RSS_{(1)} = \sum_{i=1}^{N} \hat{u}_{1i}^2 = 5.5503$; $TSS_{(1)} = \sum_{i=1}^{N} \left(\ln P_i - \overline{\ln P}\right)^2 = 11.224$; N = 74

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

5. (continued)

OLS estimation of regression equation (2) on the sample of N = 74 observations yields the following OLS coefficient estimates (with estimated standard errors in parentheses below the coefficient estimates) and summary statistics:

OLS Estimates of Equation (2)

	Constant (β_1)	$\ln WT_i$ (β_2)	$\ln FE_i$ (β_3)	$(\ln WT_i)^2$ (β_4)	$(\ln FE_i)^2$ (β_5)	$(\ln WT_i)(\ln FE_i)$ (β_6)
$\tilde{\beta}_j$ $\left(\hat{se}(\tilde{\beta}_j)\right)$	9.118 (2.917)	0.1910 (0.2720)	−0.6616 (0.2785)	----	----	----

Summary Statistics -- OLS Estimates of Equation (2):

$$RSS_{(2)} = \sum_{i=1}^{N} \hat{u}_{2i}^2 = 7.7912; \qquad TSS_{(2)} = \sum_{i=1}^{N} \left(\ln P_i - \overline{\ln P}\right)^2 = 11.224; \qquad N = 74$$

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

(10 marks)

(a) Use the results from OLS estimation of regression equation (2) to compute the two-sided 95 percent confidence intervals for (1) the slope coefficient β_2 of the regressor $\ln WT_i$, and (2) the slope coefficient β_3 of the regressor $\ln FE_i$.

(10 marks)

(b) Use the results from OLS estimation of regression equation (2) to test the proposition that $\beta_3 < 0$, i.e., that $\ln FE_i$ has a negative marginal effect on $\ln P_i$ in equation (2). Perform the test at the 5 percent significance level (i.e., for significance level α = 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. What would you conclude from the results of the test?

(10 marks)

(c) Use the results from OLS estimation of regression equation (2) to compute a two-tail test of the restriction $\beta_2 = -\beta_3$ in equation (2). Perform the test at the 5 percent significance level (i.e., for significance level α = 0.05). Explain in words what this restriction means. 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 $\tilde{\beta}_2$ and $\tilde{\beta}_3$ in equation (2) is $\hat{Cov}(\tilde{\beta}_2, \tilde{\beta}_3)$ = 0.06399. 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.

5. (continued)

(10 marks)

(d) State the coefficient restrictions that regression equation (2) imposes on regression equation (1). Use the results given above from OLS estimation of regression equations (1) and (2) to perform a test of these coefficient restrictions 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. Based on the outcome of the test, which of the two regression equations would you choose, equation (1) or equation (2)?

(10 marks)

(e) Write the expression for the marginal effect of $\ln WT_i$ on $\ln P_i$ implied by regression equation (1). State the coefficient restrictions on regression equation (1) that make the marginal effect of $\ln WT_i$ on $\ln P_i$ equal to zero. Compute a test of the null hypothesis that the marginal effect of $\ln WT_i$ on $\ln P_i$ equals zero 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 a Residual Sum-of-Squares **RSS = 6.8178** and an Explained Sum-of-Squares **ESS = 4.4058**. 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.

(10 marks)

(f) Write the expression for the marginal effect of $\ln FE_i$ on $\ln P_i$ implied by regression equation (1). State the coefficient restrictions on regression equation (1) that make the marginal effect of $\ln FE_i$ on $\ln P_i$ equal to a constant. Compute a test of the null hypothesis that the marginal effect of $\ln FE_i$ on $\ln P_i$ equals a constant in regression equation (1). Perform the test at the 1 percent significance level. 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 $\mathbf{R^2 = 0.4007}$. 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.

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:

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

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

- 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 $E(Y_0 | X_0) = \beta_1 + \beta_2 X_0$,

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

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

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

Selected Formulas (continued)

<u>For the Multiple (Three-Variable) Linear Regression Model</u>

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

- 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 $\hat{\beta}_3$:

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

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

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

Percentage Points of the t-Distribution

Selected Upper Percentage Points of the F-Distribution