

**Instructor:** Dr. Jennifer Murdock

**Duration:** 50 minutes. You must stay in the test room for the entire time.

**Format:** 20 multiple choice questions with answers recorded on SCANTRON form

**Point values:** Each multiple question worth 5 points. There are 100 total possible points.

**Allowed aids:** A non-programmable calculator (and attached aid sheets, which you may detach)

### INSTRUCTIONS:

Do NOT write your answers to the multiple choice questions on these test papers  
ONLY those answers correctly marked on the SCANTRON form can earn marks  
You MAY do scratch work on these pages

---

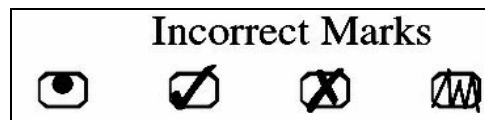
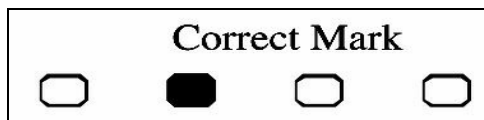
### SCANTRON INSTRUCTIONS

---

- Use only a blue or black ball point pen or pencil



- Pencil strongly recommended**, it can be erased if a mistake is made
- Make dark solid marks that fill the bubble completely



- Select the one best alternative
  - Questions with more than one answer selected will be scored incorrect
- Erase completely any marks you want to change
  - Do not use correction fluid
- Do not make stray marks on the form
- Answer every question: there is no penalty for guessing

---

**1<sup>st</sup>:** Print your **LAST NAME** and **INITIALS** in boxes provided

- Use exact name you are officially registered under
- Darken each letter in the corresponding bracket below each box

**2<sup>nd</sup>:** Print your 9 digit **STUDENT NUMBER** in the boxes provided

- Fill in zeros in front of the number if less than 9 digits
- Darken each number in the corresponding bracket below each box

**3<sup>rd</sup>:** Print 2 digit **FORM** number in the boxes provided

- Your FORM number is **03**
- Darken each number in the corresponding bracket below each box

**4<sup>th</sup>:** Sign your name in the **SIGNATURE** box

---

For the 20 questions, choose the most correct answer and mark it on the SCANTRON form.

(1) Which of the following is minimized by the least squares (OLS) coefficient estimates?

(a)  $\sum_{j=1}^n (y_j - \hat{y}_j)^2$

(b)  $\sum_{j=1}^n (y_j - \bar{y}_j)^2$

(c)  $\sum_{j=1}^n (\hat{y}_j - \bar{y}_j)^2$

(d)  $\sum_{j=1}^n (\hat{y}_j^2 - \bar{y}_j^2)$

(e)  $\sum_{j=1}^n (\hat{y}_j - b_0 - b_1 x_j)^2$

(2) Which of the following would cause the least squares (OLS) coefficient estimates to be biased?

- (a) Heteroscedasticity
- (b) An error with a large variance
- (c) A positive covariance between two explanatory (independent) variables
- (d) All of the above
- (e) None of the above

(3) In a simple regression, the standard error of estimate squared is an estimate of the variance of what?

- (a) The independent variable  $x$
- (b) The dependent variable  $y$
- (c) The unobservables  $\varepsilon$
- (d) The intercept estimate  $b_0$
- (e) The slope estimate  $b_1$

(4) In a multiple regression, which of the following will tend to INCREASE when the standard error of estimate INCREASES?

- (a)  $R^2$
- (b) Adjusted  $R^2$
- (c) Slope coefficient estimates
- (d) F statistic
- (e) p-value of the F test

► For Questions (5) - (7): Consider the following STATA output showing estimation results for a multiple regression model.

Source	SS	df	MS			
Model	6.75737639	3	2.2524588	Number of obs =	524	
Residual	502.288613	520	.965939641	F( 3, 520) =	2.33	
Total	509.04599	523	.973319292	Prob > F =	0.0733	
				R-squared =	0.0133	
				Adj R-squared =	0.0076	
				Root MSE =	.98282	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.1088136	.0435738			
x2	.0366306	.0416156			
x3	.0056433	.042454			
_cons	-.0910953	.0429565			

(5) For which variables is the slope coefficient statistically different from zero with  $\alpha = 0.05$ ?

- (a) x1
- (b) x2
- (c) x3
- (d) x2 and x3
- (e) x1, x2, and x3

(6) What is the rejection region associated with the test of overall statistical significance of the model with  $\alpha = 0.05$ ?

- (a)  $2.33 < F < \infty$
- (b)  $2.60 < F < \infty$
- (c)  $3.00 < F < \infty$
- (d)  $t < -1.645$  and  $t > 1.645$
- (e)  $t < -1.96$  and  $t > 1.96$

(7) Is this multiple regression statistically significant overall?

- (a) Yes, with a significance level of 0.05
- (b) Yes, with significance level between 0.05 – 0.10
- (c) Yes, with significance level between 0.01 – 0.05
- (d) Yes, with significance level between 0.005 – 0.01
- (e) No, not with any conventional significance level

(8) Suppose you read a report about a simple regression of the number of bicycles parked in front of a building ( $y$ ) on the outside air temperature ( $x$ ). The following table containing prediction intervals is given. What is the estimated least squares line?

Temperature	Prediction Interval
5	(17, 23)
15	(38, 42)
25	(57, 63)

- (a)  $\hat{y} = 5 + 10x$
- (b)  $\hat{y} = 10 + 2x$
- (c)  $\hat{y} = 10 + 5x$
- (d)  $\hat{y} = 17 + 2.1x$
- (e)  $\hat{y} = 23 + 1.9x$

(9) In a multiple regression what is the expected effect of adding a totally irrelevant variable to the right hand side of the equation?

- (a) SSE goes up
- (b) SST goes up
- (c) Parameter estimates become biased
- (d)  $R^2$  goes up
- (e) Heteroscedasticity becomes an issue

► For Questions (10) - (12): Consider the following STATA output showing estimation results for a multiple regression model.

Source	SS	df	MS				
Model	211.905119	3	70.6350397	Number of obs =	88		
Residual	737.622005	84	8.78121435	F( 3, 84) =	8.04		
Total	949.527125	87	10.9141049	Prob > F =	0.0001		
				R-squared =	0.2232		
				Adj R-squared =	0.1954		
				Root MSE =	2.9633		

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	.8521985	.3506446	2.43	0.017	.1549033	1.549494
w	.2530686	.347256	0.73	0.468	-.437488	.9436251
z	1.26054	.3218851	3.92	0.000	.6204358	1.900643
_cons	.4425827	.3172163	1.40	0.167	-.1882367	1.073402

(10) What is the variance of y?

- (a) 1.79
- (b) 2.96
- (c) 8.78
- (d) 10.91
- (e) 70.64

(11) What is the interpretation of “R-squared = 0.2232”?

- (a) 0.2232% of the variation in y is explained by variation in x, w, and z
- (b) 22.32% of the variation in y is explained by variation in x, w, and z
- (c) 22.32% of the variation in y is explained by variation in x and z
- (d) 22.32% of the variation in y is explained by variation in factors other than x, w, and z
- (e) The coefficient of correlation between y, x, w, and z is 0.2232, which indicates a weak positive linear relationship

(12) Given these results, which of the following actions should be taken?

- (a) The heteroscedasticity problem should be corrected by using an alternative functional form
- (b) The variable w should be dropped from the model and parameters re-estimated
- (c) Regression diagnostics should be done to check for violations of underlying assumptions
- (d) Observations that are outliers should be dropped to improve model fit
- (e) All of the above

► For Questions (13) - (14): Suppose you collect data on the quantity of hot dogs sold and prices for a random sample of street vendors in the city of Toronto. You obtain the following simple regression estimates with standard errors in parentheses:

$$\ln(\hat{Q}) - \hat{hat} = 4.77 - 1.65\ln(P)$$

(5.52) (0.57)

(13) Which of the following can you conclude?

- (a) The price elasticity of demand is -1.65
- (b) Demand is elastic
- (c) Demand is inelastic
- (d) A one dollar increase in price results in a 1.65 unit reduction in sales
- (e) None of the above

(14) Which of the following is the most serious flaw in the estimation results?

- (a) The logarithm of price is endogenous
- (b) Logarithms cannot be using in a simple linear regression
- (c) Heteroscedasticity
- (d) Statistical insignificance
- (e) Autocorrelation (serial correlation)

► For Question (15): The following multiple regression results are obtained:

$$\hat{Y} = 10.10 + 2.40X + 1.10Z - 0.55XZ + 22.10Q$$

(15) What is the point estimate of the effect that a one unit change in X will have on Y when Z = 10 and Q = 2?

- (a) Y decreases by 3.10 units
- (b) Y increases by 1.85 units
- (c) Y increases by 2.40 units
- (d) Y increases by 2.95 units
- (e) Y increases by 62.20 units

(16) For which of the following estimated multiple regression models does the point estimate of the underlined coefficient indicate that a 1 unit increase in X is associated with a 4 percent increase in Y?

- (a)  $\hat{Y} = 1.00 + \underline{4.00}X + 1.00Z$
- (b)  $\hat{Y} = 4.00 + \underline{0.04}X + 4.00Z$
- (c)  $\hat{Y} = 4.00 + \underline{400.00}\ln(X) + 2.00Z$
- (d)  $\ln(\hat{Y}) = 4.00 + \underline{0.04}X + 1.00Z$
- (e)  $\ln(\hat{Y}) = 4.00 + \underline{4.00}\ln(X) + 2.00\ln(Z)$

(17) For which of the following multiple regression results, with standard errors in parentheses, could you conclude that there is a quadratic relationship between X and Y? (Suppose the sample size in all cases is very large.)

Regression #1:  $\hat{Y} = 1.341 + 0.552X + 0.033X^2$   
 (0.440) (0.127) (0.011)

Regression #2:  $\hat{Y} = 0.552 + 1.224X + 0.018X^2$   
 (0.658) (2.044) (0.007)

Regression #3:  $\hat{Y} = 2.899 + 0.066X + 0.041X^2$   
 (1.112) (0.047) (0.287)

- (a) Regression #1
- (b) Regression #2
- (c) Regression #3
- (d) Regressions #1 and #2
- (e) Regressions #1, #2, and #3

(18) If you included 40 independent variables in a multiple regression model to explain y, how many would you expect to find are statistically significant at the 5% level if in fact there is no relationship between any of the 40 variables and y?

- (a) 0
- (b) 1
- (c) 2
- (d) 4
- (e) 5

► For Questions (19) - (20): Consider the following STATA output showing estimation results for a multiple regression model using panel data that follows 8 firms (Firms A, B, C, D, E, F, G, and H) over 10 years.

Source	SS	df	MS			
Model	995133.62	9	110570.402	Number of obs =	80	
Residual	83475.512	70	1192.50731	F( 9, 70) =	92.72	
Total	1078609.13	79	13653.2801	Prob > F =	0.0000	
				R-squared =	0.9226	
				Adj R-squared =	0.9127	
				Root MSE =	34.533	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
var1	1.458586	.7582898	1.92	0.058	-.053775	2.970947
var2	10.25862	.3677995	27.89	0.000	9.525072	10.99218
firm_B	80.6616	16.09547	5.01	0.000	48.5602	112.763
firm_C	-160.5668	17.41796	-9.22	0.000	-195.3058	-125.8278
firm_D	-25.03298	15.45627	-1.62	0.110	-55.85955	5.793584
firm_E	145.8896	18.13192	8.05	0.000	109.7267	182.0526
firm_F	97.04291	16.01392	6.06	0.000	65.10416	128.9817
firm_G	29.58329	15.69643	1.88	0.064	-1.72225	60.88882
firm_H	63.28781	15.90963	3.98	0.000	31.55706	95.01856
_cons	178.4531	97.79112	1.82	0.072	-16.58509	373.4914

(19) For Firm G, what is the relationship between y and var1 and var2?

- (a)  $\hat{y} = 148.87 + 1.46 \cdot \text{var1} + 10.26 \cdot \text{var2}$
- (b)  $\hat{y} = 178.45 + 1.46 \cdot \text{var1} + 10.26 \cdot \text{var2}$
- (c)  $\hat{y} = 208.04 + 1.46 \cdot \text{var1} + 10.26 \cdot \text{var2}$
- (d)  $\hat{y} = 230.87 + 1.46 \cdot \text{var1} + 10.26 \cdot \text{var2}$
- (e)  $\hat{y} = 409.32 + 1.46 \cdot \text{var1} + 10.26 \cdot \text{var2}$

(20) Suppose that  $\beta_2$  is the parameter that measures the marginal effect of var2 on y. Consider the following hypothesis test:

$$H_0: \beta_2 = 10$$

$$H_1: \beta_2 > 10$$

What is the test statistic and rejection region with  $\alpha = 0.01$ ?

- (a)  $t = 0.70$ ; reject if  $t > 2.374$
- (b)  $t = 0.70$ ; reject if  $t > 2.381$
- (c)  $t = 17.89$ ; reject if  $t > 2.381$
- (d)  $t = 27.89$ ; reject if  $t > 2.648$
- (e) None of the above