(102)輔仁大學碩士班招生考試試題	考試日期：102 年 3 月 8 日第 3 節
	本試題共 3 頁 (本頁為第 1 頁)

科目：迴歸分析　　系所組：應統碩士班

Regression Analysis

1. (20 points) MULTIPLE CHOICE QUESTIONS

   (a) (　　) If multicollinearity exists among the independent variables included in a multiple regression model, then (a) the regression coefficients will be difficult to interpret (b) the standard errors of the regression coefficients for the correlated independent variables will increase (c) one or more of the coefficients may have the wrong sign (d) All of these choices are true

   (b) (　　) In a multiple regression model, the value of Variance Inflation Factor (VIF) has to fall between (a) -1 and +1 (b) 0 and ∞ (c) 1 and ∞ (d) None of these choices.

   (c) (　　) The range of the Durbin-Watson statistic is between (a) -1 to 1 (b) 0 to 1 (c) $-\infty$ to $+\infty$ (d) 0 to 4

   (d) (　　) What value of Durbin-Watson statistic indicates no autocorrelation is present? (a) 1 (b) 2 (c) -2 (d) 0

   (e) (　　) A measure of the effect of an unusual x value on the regression results is called (a) Cook's D (b) Leverage (c) odd ratio (d) unusual regression

   (f) (　　) The following results were obtained from a multiple regression analysis. $SSR = 384$, $SST = 704$, $MSR = 48$, $MSE = 20$. How many independent variables were involved in this model? (a) 6 (b) 7 (c) 8 (d) 9

   (g) (　　) The adjusted multiple coefficient of determination is adjusted for (a) the number of dependent variables (b) the number of independent variables (c) the number of equations (d) detrimental situations

   (h) (　　) The numerical value of the coefficient of determination (a) is always larger than the coefficient of correlation (b) is always smaller than the coefficient of correlation (c) is negative if the coefficient of determination is negative (d) can be larger or smaller than the coefficient of correlation

   (i) (　　) The coefficients of the logistic regression model $\ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$ are estimated using a statistical technique called: (a) least squares method (b) maximum likelihood estimation (c) logistic likelihood estimation (d) All of these choices are true

   (j) (　　) If the Durbin-Watson statistic d has values smaller than 2, this indicates (a) a positive first-order autocorrelation (b) a negative first-order autocorrelation (c) no first-order autocorrelation at all (d) None of these choices

2. (15 points) Monthly total production costs and the number of units produced at a local company over a period of 10 months are shown below. x = units produced (in millions) and y = production costs (in millions \$). Data (x, y) are (2, 1), (3, 1), (4, 1), (5, 2), (6, 2), (7, 4), (8, 5), (9, 7), (10, 9), (10, 12).

(a) Assume that a model in the form of $y = \beta_0 + \beta_1 x + \varepsilon$ and using least squares to find $\hat{\beta}_0$ and $\hat{\beta}_1$.

(b) Construct and interpret a plot of residuals versus predicted response.

3. (15 points) In order to determine whether or not the number of automobiles sold per day (Y) is related to price (X_1 in \$1,000), and the number of advertising spots (X_2), data were gathered for 7 days. Part of the regression results is shown below.

	coefficient	standard error
constant	0.8051	
x_1	0.4977	0.4617
x_2	0.4733	0.0387

The Analysis of Variance Table

Source	*df*	*SS*	*MS*	*F*
Regression		40.70		
Residual		1.016		
Total				

(a). If the company charges \$20,000 for each car and uses 10 advertising spots, how many cars would you expect them to sell in a day?
(b). At $\alpha = 0.05$, test to determine if the fitted equation developed represents a significant relationship between the independent variables and the dependent variable.
(c). At 95% confidence, test to see if the number of advertising spots is a significant variable.

4. (15 points) The variables $(y_i, x_{1i}, x_{2i}), i = 1, 2, \ldots, 5$ and consider three regression modles and statistics as follows (SSE = residual sum of squares, AIC = Akaike's Information Criterion)

Model	*SSE*	*AIC*
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$	0.107	-13.222
$y = \beta_0 + \beta_1 x_1 + \varepsilon$	0.669	-6.0533
$y = \beta_0 + \beta_2 x_2 + \varepsilon$	8.927	6.8982

Please note that $P[F(1,2) > 18.51] = 0.05$, $P[F(1,3) > 10.13] = 0.05$, $P[F(1,4) > 7.71] = 0.05$ and $P[F(1,5) > 6.61] = 0.05$.
(a). Using AIC to select the best regression model and explain why.
(b). Let $\alpha = 0.05$, and $\sum_{i=1}^{5}(y_i - \overline{y})^2 = 25.172$. Using forward selection method to select a regressiom model.

5. (20 points) Given a dependent variable (y) and four independent variables x_1, x_2, x_3 and x_4. Base on the statistics in the table, the purpose is to find the best regression model. Note: p = number of parameters, MSE = mean square errors.

variables in model	p	df	SSE_p	R_p^2	MSE_p	C_P
None	1	53	4	0	0.075	1721
x_1	2	52	3.5	0.12	0.067	1511
x_2	2	52	2.58	0.35	0.050	1100
x_3	2	52	2.22	0.44	0.043	939
x_4	2	52	1.88	0.53	0.036	788
x_1, x_2	3	51	2.23	0.44	0.040	949
x_1, x_3	3	51	1.41	0.65	0.030	580
x_1, x_4	3	51	1.88	0.53	0.036	789
x_2, x_3	3	51	0.74	0.81	0.015	284
x_2, x_4	3	51	1.39	0.65	0.027	574
x_3, x_4	3	51	1.25	0.69	0.024	508
x_1, x_2, x_3	4	50	0.11	0.97	0.002	3.1
x_1, x_2, x_4	4	50	1.39	0.65	0.028	575
x_1, x_3, x_4	4	50	1.12	0.72	0.022	452
x_2, x_3, x_4	4	50	0.47	0.88	0.009	162
x_1, x_2, x_3, x_4	5	49	0.11	0.97	0.002	5

(a). Base on R_p^2, find the best regression model.
(b). Base on MSE_p, find the best regression model.
(c). Base on C_P, find the best regression model.

6. (15 points) Given data $(x_i, y_i), i = 1, 2, \ldots, 20$ and fit the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i are independent with $N(0, \sigma^2)$. We have $\overline{x} = 5, \overline{y} = 3$, $\sum_{i=1}^{20}(x_i - \overline{x})^2 = 160$, $\sum_{i=1}^{20}(y_i - \overline{y})^2 = 83.2$, and $\sum_{i=1}^{20}(x_i - \overline{x})(y_i - \overline{y}) = 80$.
(a). Write out the ANOVA Table.
(b). Suppose $x_i = 2, i = 1, 2, 3, 4$; $x_i = 4, i = 5, 6, 7, 8$; $x_i = 6, i = 9, 10, 11, 12$; $x_i = 8, i = 13, 14, 15, 16$; $x_i = 10, i = 17, 18, 19, 20$. The *pure error sum of squares* is 23.2. Is the model in part (a) a proper one? Write the hypothesis, test statistic and conclusion.