UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2015/2016

COURSE NAME	:	APPLIED REGRESSION ANALYSIS
COURSE CODE	:	BWB 20803
PROGRAMME CODE	:	BWQ
EXAMINATION DATE	:	JUNE / JULY 2016
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER **ALL** QUESTIONS

THIS QUESTION PAPER CONSISTS OF **SEVEN (7)** PAGES

Q1 (a) Write down the general model for first order multiple linear regression with two predictor variables. State two assumptions regarding on the model.

(4 marks)

(b) In the extensive application of regression analysis, one would come across three familiar measures namely Cook's Distance, DFBETAS and DFFITS. Explain the usage of these measures in regression analysis.

(6 marks)

(c) Verify that the total sum of squares (SST) is a decomposition of error sum of squares (SSE) and regression sum of squares (SSR) by showing that the left hand side of the following equation is the same as the right hand side.

$$\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2=\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2+\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2$$

$$\left[\text{Hint: } \sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2=\sum_{i=1}^{n}\left[\left(\hat{Y}_i-\bar{Y}\right)+\left(Y_i-\hat{Y}_i\right)\right]^2\right]$$

(5 marks)

(d) How does the problem of multicollinearity exist in multiple regression? State two effects of multicollinearity and two methods to detect the presence of multicollinearity.

(5 marks)

Q2 A soft drink bottler is analyzing the vending machine routes in his distribution system. He is interested in predicting the amount of time required by the route driver to service the vending machines in an outlet. The study has suggested that the two most important variables affecting the delivery time (Y) are the number of cased of product stocked (X_1) and the distance walked by the route driver (X_2). 25 data were randomly selected and recorded. The following computations were made.

$$(X'X)=\begin{bmatrix}25 & 219 & 10232\\ 219 & 3055 & 133889\\ 10232 & 133889 & 6725688\end{bmatrix},\quad (X'Y)=\begin{bmatrix}559.60\\ 7375.44\\ 337072.00\end{bmatrix},\quad \sum Y^2=18310.6290$$

(a) Calculate the least-square coefficient vector and obtain the fitted regression function for first order regression model with two predictor variables.

(6 marks)

(b) Obtain the estimated variance covariance matrix for **b**.

(8 marks)

(c) Interpret the meaning of coefficient b_2.

(2 marks)

(d) Calculate coefficient of determination, R^2 for the above problem. Interpret the result.

(4 marks)

Q3 A personnel officer in a govermental agency administered four newly developed aptitude tests (X_1, X_2, X_3, X_4) to each 50 applicants for entry-level clerical positions in the agency. For purpose of the study, all 50 applicants were accepted for positions irrespective of their test scores. After a probationary period, each applicant was rated for proficiency on the job (Y).

(a) The stepwise regression method is employed to select the best model for the job profeciency score of the employees. Describe the first two steps of the procedure.

(6 marks)

(b) Based on the MINITAB output of coefficient of correlation matrix below

Correlations: Y, X1, X2, X3, X4

	Y	X1	X2	X3
X1	0.529			
	0.000			
X2	0.412	0.054		
	0.003	0.710		
X3	0.895	0.187	0.424	
	0.000	0.193	0.002	
X4	0.881	0.331	0.302	0.826
	0.000	0.019	0.033	0.000

Cell Contents: Pearson correlation
P-Value

(i) identify which variables are potential predictors of job proficiency

(2 marks)

(ii) describe the nature and strength of the relationship

(2 marks)

(iii) identify which variable will be selected first to be in the model.

(2 marks)

(c) Based on the MINITAB output of the stepwise regression provided, write the equation of

(i) the model with only one variable estimated during the first step.

(2 marks)

(ii) the best two variable model.

(2 marks)

(iii) the best three variable model.

(2 marks)

Stepwise Regression: Y versus X1, X2, X3, X4

Alpha-to-Enter: 0.1 Alpha-to-Remove: 0.15

Response is Y on 4 predictors, with N = 50

Step	1	2	3	4
Constant	-109.7	-129.7	-123.4	-124.2
X3	1.996	1.840	1.369	1.288
T-Value	13.91	22.02	11.15	10.25
P-Value	0.000	0.000	0.000	0.000
X1		0.349	0.304	0.304
T-Value		9.99	9.89	10.20
P-Value		0.000	0.000	0.000
X4			0.49	0.51
T-Value			4.65	4.96
P-Value			0.000	0.000
X2				0.067
T-Value				2.00
P-Value				0.051
S	8.40	4.81	4.01	3.88
R-Sq	80.13	93.63	95.67	96.03
R-Sq(adj)	79.71	93.36	95.39	95.67
Mallows Cp	179.0	28.1	7.0	5.0

(d) Among the four variables, which is the most important predictor for the job profeciency? Explain.

(2 marks)

Q4 The manager of a company wishes to determine the important factors in predicting current salary of the company's employees. A statistical analysis was carried out on information obtained from 474 employees. The variables of interest are listed below.

Y = current salary (RM'000)
X_1 = beginning salary (RM'000)
X_2 = previous work experience (in months)
$X_3 = \begin{cases} 1 & \text{if female} \\ 0 & \text{if male} \end{cases}$

The manager wishes to determine the best predictor variables for predicting current salary. The results of the statistical analysis are shown in **Appendix 1**. Based on the results, answer the following questions.

(a) Is it true that on the average, male employees earn more than female employees? If yes, by how much? Test by using the 5% level of significance.

(6 marks)

(b) Interpret the meaning of each of the estimated regression coefficients.

(6 marks)

(c) Estimate the mean current salary of a female employee with the following details:
Beginning Salary = RM 1040.00
Previous Work Experience = 12 months.

(3 marks)

(d) Discuss how you would include the following variable into the existing estimated regression model.

$$\text{Employment category} = \begin{cases} \text{Manager} \\ \text{Supervisor} \\ \text{Technician} \\ \text{Clerk} \end{cases}$$

(5 marks)

Q5 A research was conducted to examine the effect of several factors on managerial performance. A sample of 100 management personnel from several departments within a government agency took part in the study. Each manager completed a questionnaire designed to measure the following variables:

Y = Performance index

$X_1 = \begin{cases} 1 & \text{if male} \\ 0 & \text{if female} \end{cases}$

X_2 = Job tenure (years)

X_3 = Manager and subordinate work relationship rating

X_4 = Effort level (average number hours per week invested in job

$X_5 = \begin{cases} 1 & \text{if upper-level manager} \\ 0 & \text{if lower-level manager} \end{cases}$

The data collected on the 100 managers were then used to fit several regression models of managerial performance.

(a) Initially, the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

was considered. For this model, $SSE = 352$ and $R^2 = 0.11$. Calculate the F statistic for testing model adequacy. Is this model useful for predicting performance rating Y? Use $\alpha = 0.05$.

(8 marks)

(b) Terms effort level and managerial level, $\beta_4 X_4 + \beta_5 X_5$ were added to the model in part (a), resulting in $SSE = 341$ and $R^2 = 0.14$. Do these terms contribute additional information for the prediction of performance rating Y? Test using $\alpha = 0.05$.

(6 marks)

(c) A third model was also considered

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_4 X_5 + \varepsilon .$$

This model resulted in $SSE = 321$ and $R^2 = 0.19$. Test the hypothesis that the interaction between effort level, X_4 and managerial level, X_5 is not important. Use $\alpha = 0.05$.

(6 marks)

- END OF QUESTION -

FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2015/2016 PROGRAMME : BWQ
COURSE : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

APPENDIX 1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.894[(a)]	0.799	0.798	7671.590

[a] Predictors : (Constant), Gender, Previous Experience (Months), Beginning Salary

ANOVA[(b)]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.10+E11	3	3.675+E10	624.465	0.000[a]
	Residual	2.77+E10	470	58853291.39		
	Total	1.38+E11	473			

[a] Predictors: (Constant), Gender, Previous Experience (Months), Beginning Salary

[b] Dependent Variable : Current Salary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		β	Std. Error	Beta		
1	(Constant)	6886.189	1203.343		5.723	0.000
	Beginning Salary	1.837	0.050	0.847	36.442	0.000
	Previous Experience	-24.622	3.422	0.150	-7.167	0.000
	Gender	-3014.517	805.979	-0.088	-3.740	0.000

[a] Dependent Variable : Current Salary