

CONFIDENTIAL



UNIVERSITI TUN HUSSEIN ONN MALAYSIA

**FINAL EXAMINATION
SEMESTER I
SESSION 2021/2022**

COURSE NAME : APPLIED REGRESSION ANALYSIS
COURSE CODE : BWB 20803
PROGRAMME CODE : BWQ
EXAMINATION DATE : JANUARY / FEBRUARY 2022
DURATION : 2 HOURS
INSTRUCTION : 1. ANSWER **ALL** QUESTIONS.
2. THIS FINAL EXAMINATION IS AN **ONLINE** ASSESSMENT AND CONDUCTED VIA **OPEN BOOK**.

THIS QUESTION PAPER CONSISTS OF TWELVE (12) PAGES

TERBUKA

CONFIDENTIAL

Q1 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)

(a) Write down the general linear equation for the above problem. (1 mark)

(b) A new variable will be included into the existing estimated regression model.

Employment category = $\left\{ \begin{array}{l} \text{Manager} \\ \text{Supervisor} \\ \text{Technician} \\ \text{Clerk} \end{array} \right.$

(i) Write down the new general linear equation when new variable added into the model. (1 mark)

(ii) Determine the response function for each type of employment category. (4 marks)

(iii) Specify the alternative statement of H_0 and H_a for the appropriate test with X_1 and X_2 fixed, does working as a manager increase the expected salary as compared with working as a supervisor? (2 marks)

(iv) Specify the alternative statement of H_0 and H_a for the appropriate test with X_1 and X_2 fixed, does the expected salary of working as a technician greater than working as a clerk? (2 marks)

Q2 Institut Penyelidikan Keselamatan Jalan Raya Malaysia (MIROS) is concerned about the increasing number of accidents along the MEX Highway. The number of accidents (Y) seems to be related to the number of vehicles that travel over it (X_1) and the speed (X_2) in miles per hour which they are travelling. Data regarding these variables were collected with the intention of examining it statistically so that the policy makers can introduced the new speed laws on reducing the number of accidents. The first order model with an interaction between both predictor variables was adopted in the study. The Minitab output obtained is presented in **Appendix 1**.

(a) Write down the estimated model.

TERBUKA
(1 mark)

(b) Interpret the intercept value in the model. Is it possible to interpret the intercept? Give the reasons of your answer. (5 marks)

(c) Test the appropriateness of the model. Use the 5% significance level. (5 marks)

(d) After completing the analysis, the research assistant decides to modify the model as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \varepsilon.$$

The sum of squares error obtained from the analysis was 1327.545. Show if the quadratic terms are significant in explaining the number of accidents? Provide evidence to support the finding. Use the 5% significance level.

(7 marks)

Q3 The Minitab output in **Appendix 2** shows the multiple regression analysis on the sales of cosmetics. The variables involved in the study are X_1 denotes expenditures for point-of-sale displays in beauty salons and department stores (in thousand dollars), X_2 and X_3 represent the corresponding expenditures for local media advertising and prorated share of national media advertising, respectively. Meanwhile the response variable Y is sales (in thousand cases).

(a) Determine the best model by applying all possible regression method. You can choose any criterion. (7 marks)

(b) Use the stepwise regression method to find the best subset of predictor variables for the response variable. Use α -to-enter = 0.10 and α -to-remove = 0.15. (6 marks)

(c) Write the equation of the best model to predict Y obtained in the stepwise regression method. Is it the same as your proposed model in **Q3(a)**? Elaborate your answer. (4 marks)

TERBUKA

Q4 In a small-scale regression study, the data in **Table Q4** were obtained.

Table Q4

Observation	Y	X_1	X_2	X_3
1	49	45	36	45
2	55	30	28	40
3	85	11	16	42
4	32	30	46	40
5	26	39	76	43
6	28	42	78	27
7	95	17	24	36
8	26	63	80	42
9	74	25	12	52
10	37	32	27	35
11	31	37	37	55
12	49	29	34	47
13	38	26	32	28
14	41	38	45	30
15	12	38	99	26
16	44	25	38	47
17	29	27	51	44
18	40	37	32	54
19	31	34	40	36

- (a) Identify and show any outlying points with respect to X or Y direction. Explain your findings. (7 marks)
- (b) Compute DFFITS and Cook's distance values to assess their influences for all cases in **Q4(a)**. Conclude your findings. (8 marks)

- END OF QUESTIONS -

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2021/2022 PROGRAMME CODE : BWQ
 COURSE NAME : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

Appendix 1

Coefficients^a

Model	Unstandardized Coefficients		t
	β	Std. Error	
1 (Constant)	8.7514	20.1345	0.435
X_1	0.0229	0.0115	1.980
X_2	0.0731	0.0240	3.051
X_1X_2	0.00071	0.00015	4.720

a. Dependent Variable : Y

ANOVA^b

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	3954.712	3	1318.238	287.2288	0.000 ^a
Residual	1588.141	346	4.590		
Total	5542.853	349			

a. Predictors: (Constant), X_1 , X_2

b. Dependent Variable : Y

TERBUKA
CONFIDENTIAL

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2021/2022 PROGRAMME CODE : BWQ
 COURSE NAME : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

APPENDIX 2

MODEL 1

Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate
1	0.805	0.648	0.620	12.9965

Predictors : (Constant), Expenditures

Coefficients

Model	Unstandardized Coefficients		t	Sig.
	β	Std. Error		
1 (Constant)	18.354	14.808	1.239	0.237
Expenditures	3.879	0.794	4.887	0.000

Dependent Variable: Sales

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	4034.414	1	4034.414	23.885	0.000
Residual	2195.822	13	168.909		
Total	6230.236	14			

Predictors: (Constant), Expenditures

Dependent Variable : Sales

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2021/2022 PROGRAMME CODE : BWQ
 COURSE NAME : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

MODEL 2

Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate
	0.521	0.271	0.215	18.6875

Predictors : (Constant), Advertising expenditures

Coefficients

Model	Unstandardized Coefficients		t	Sig.
	β	Std. Error		
(Constant)	112.463	11.776	9.551	0.000
Advertising expenditures	-2.727	1.239	-2.200	0.046

Dependent Variable: Sales

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1690.364	1	1690.364	4.840	0.046
Residual	4539.872	13	349.221		
Total	6230.236	14			

Predictors: (Constant), Advertising expenditures

Dependent Variable: Sales

TERBUKA

CONFIDENTIAL

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2021/2022 PROGRAMME CODE : BWQ
 COURSE NAME : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

MODEL 3

Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate
	0.372	0.139	0.072	20.3171

Predictors : (Constant), Prorated share

Coefficients

Model	Unstandardized Coefficients		t	Sig.
	β	Std. Error		
(Constant)	66.972	16.000	4.186	0.001
Prorated share	3.344	2.311	1.447	0.172

Dependent Variable: Sales

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	864.026	1	864.026	2.093	0.172
Residual	5366.210	13	412.785		
Total	6230.236	14			

Predictors: (Constant), Prorated share

Dependent Variable: Sales

TERBUKA

FINAL EXAMINATION			
SEMESTER / SESSION	: SEM I / 2021/2022	PROGRAMME CODE	: BWQ
COURSE NAME	: APPLIED REGRESSION ANALYSIS	COURSE CODE	: BWB 20803

MODEL 4

Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate
	0.835	0.697	0.646	12.5461

Predictors : (Constant), Advertising expenditures, Expenditures

Coefficients

Model	Unstandardized Coefficients		t	Sig.
	β	Std. Error		
(Constant)	-25.580	34.556	-0.74	0.473
Expenditures	5.353	1.304	4.104	0.001
Advertising expenditures	1.978	1.417	1.396	0.188

Dependent Variable: Sales

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4341.373	2	2170.686	13.790	0.001
Residual	1888.863	12	157.405		
Total	6230.236	14			

Predictors: (Constant), Advertising expenditures, Expenditures

Dependent Variable : Sales



FINAL EXAMINATION					
SEMESTER / SESSION	: SEM I / 2021/2022	PROGRAMME CODE	: BWQ		
COURSE NAME	: APPLIED REGRESSION ANALYSIS	COURSE CODE	: BWB 20803		

MODEL 5

Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate
	0.957	0.916	0.901	6.6220

Predictors : (Constant), Prorated share, Expenditures

Coefficients

Model	Unstandardized Coefficients		t	Sig.
	β	Std. Error		
(Constant)	-20.372	9.814	-2.076	0.060
Expenditures	4.312	0.410	10.506	0.000
Prorated share	4.718	0.765	6.170	0.000

Dependent Variable: Sales

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	5704.027	2	2852.014	65.039	0.000
Residual	526.209	12	43.851		
Total	6230.236	14			

Predictors: (Constant), Prorated share, Expenditures

Dependent Variable : Sales

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2021/2022 PROGRAMME CODE : BWQ
 COURSE NAME : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

MODEL 6

Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate
	0.827	0.684	0.631	12.8151

Predictors : (Constant), Prorated share, Advertising expenditures

Coefficients

Model	Unstandardized Coefficients		t	Sig.
	β	Std. Error		
(Constant)	84.219	10.781	7.812	0.000
Advertising expenditures	-4.237	0.932	-4.547	0.001
Prorated share	6.321	1.598	3.955	0.002

Dependent Variable: Sales

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4259.516	2	2129.758	12.968	0.001
Residual	1970.720	12	164.227		
Total	6230.236	14			

Predictors: (Constant), Prorated share, Advertising expenditures

Dependent Variable : Sales



FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2021/2022 PROGRAMME CODE : BWQ
 COURSE NAME : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

MODEL 7

Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate
	0.957	0.916	0.893	6.8940

Predictors : (Constant), Prorated share, Advertising expenditures, Expenditures

Coefficients

Model	Unstandardized Coefficients		t	Sig.
	β	Std. Error		
(Constant)	-16.058	19.071	-0.842	0.418
Expenditures	4.146	0.751	5.520	0.000
Advertising expenditures	-0.236	0.881	-0.268	0.794
Prorated share	4.831	0.901	5.361	0.000

Dependent Variable: Sales

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	5707.438	3	1902.479	40.029	0.000
Residual	522.798	11	47.527		
Total	6230.236	14			

Predictors : (Constant), Prorated share, Advertising expenditures, Expenditures

Dependent Variable : Sales

TERBUKA