CONFIDENTIAL

UTHM
Universiti Tun Hussein Onn Malaysia

# UNIVERSITI TUN HUSSEIN ONN MALAYSIA

## FINAL EXAMINATION
## SEMESTER II
## SESSION 2016/2017

| | | |
|---|---|---|
| COURSE NAME | : | APPLIED REGRESSION ANALYSIS |
| COURSE CODE | : | BWB 20803 |
| PROGRAMME CODE | : | BWQ |
| EXAMINATION DATE | : | JUNE 2017 |
| DURATION | : | 3 HOURS |
| INSTRUCTION | : | ANSWER **ALL** QUESTIONS |

TERBUKA

THIS QUESTION PAPER CONSISTS OF **EIGHT(8)** PAGES

CONFIDENTIAL

**Q1** (a) Write down the general model for first order multiple linear regression with two predictor variables. State **TWO (2)** assumptions regarding on the model.

(4 marks)

(b) 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.

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

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

(6 marks)

(c) State **THREE (3)** conditions in regression analysis where transformations can be employed.

(3 marks)

(d) Given that

$$\left(X'X\right)^{-1}=\begin{bmatrix}0.6 & 0 & 0\\ 0 & 0.4 & 0\\ 0 & 0 & 0.25\end{bmatrix},\ \left(X'Y\right)=\begin{bmatrix}3\\5\\8\end{bmatrix},\ Y=\begin{bmatrix}1.5\\4\\8\end{bmatrix}\ \text{ and } n=16$$

(i) Calculate the least squares coefficient and obtain the fitted regression function.

(5 marks)

(ii) Obtain the error sum of squares, SSE.

(5 marks)



**Q2** (a) Construct an appropriate diagrams to explain the difference between a leverage point and an influence point.

(6 marks)

(b) A factory manager would like to develop a regression model to estimate and predict the cost of producing their main product of an automobile braking system. A random sample of 20 observations were collected on the cost of making a unit of the braking system, $Y$ (in RM) and the weight of the metal, $X$ (in kg) used in the manfacturing of each unit of the braking system. A simple linear regression model relating the cost of one unit of the braking system to the weight of metal used is estimated as given in **Appendix 1**.

(i) What can you conclude based on the scatter plot?

(2 marks)

(ii) Test whether the model is significant. Use $\alpha = 0.05$. Comment on the proportion of total variability in cost explained by the model.

(6 marks)

(iii) Based on the values of DFBETAS and DFFITS for the last eight observations in **Table Q2(b)** below, identify which observations are potential outliers. Justify your answer.

**Table Q2(b)** Values of DFBETAS and DFFITS

| Observation | $Y$ (RM) | DFBETAS | DFFITS |
|---|---|---|---|
| 13 | 462 | 0.4281 | -0.4980 |
| 14 | 5928 | 0.5656 | 0.6215 |
| 15 | 2766 | -0.0033 | -0.0140 |
| 16 | 2597 | -2.7095 | -2.8525 |
| 17 | 5552 | 0.1899 | 0.5091 |
| 18 | 5623 | 0.2512 | 0.4363 |
| 19 | 3225 | 0.0105 | 0.0476 |
| 20 | 3616 | -0.1426 | 0.2389 |

(5 marks)

**Q3** A house developer company wants to investigate the influencing factors when the customers buy a house. The developer went to a local real estate agency and obtained the data consisting of the following variables:

$Y$ = selling price of house (RM'000)
$X_1$ = age of house (years)
$X_2$ = lot size (100 square feet)
$X_3$ = number of bedrooms

$$X_4 = \begin{cases} 1 & \text{if intermediate lot} \\ 0 & \text{if corner lot} \end{cases}$$



Refer to **Appendix 2** for the computer output.

(a) Write down the fitted regression function.

(2 marks)

(b) If the customer plan to buy a 5 year old intermediate house that has a lot size of 1400 square feet and 4 bedrooms, what price would you expect?

(3 marks)

(c) Does an increase in lot size by 100 square feet change the sales price by RM5000? Test by using the appropriate hypothesis. Use $\alpha = 0.05$.

(5 marks)

(d) Do the number of bedrooms and location of a lot cause changes in the mean sales price? Test $H_0: \beta_3 = \beta_4 = 0$ vs $H_1$: not both $\beta_3$ and $\beta_4$ equal zero. Use $\alpha = 0.05$.

(10 marks)

**Q4** A researcher studied the effects of the charge rate and temperature on the life of a new type of power cell in a preliminary small-scale experiment. The charge rate ($X_1$) was controlled at three levels (0.6, 1.0 and 1.4 amperes) and the ambient temperature ($X_2$) was controlled at three levels (10, 20, 30 $^0C$). Factors pertaining to the discharge of the power cell were held at fixed levels. The life of power cell, $Y$ was measured in terms of the number of discharge-charge cycles that a power cell underwent before it failed. The researcher was not sure about the nature of the response function in the range of the factors studied. Hence, the researcher decided to fit the second order polynomial regression model. Refer to the **Appendix 3** for the computer output.

(a) Write down the estimated regression function.

(2 marks)

(b) Synthesize the information of model adequacy based on residual plots.

(4 marks)

(c) Since there are replications in predictor variables, conduct an appropriate test to check whether the second order polynomial regression model is a good fit. Use $\alpha = 0.05$

(10 marks)

(d) Test whether a first order model would be sufficient or not. Use $\alpha = 0.05$.

(10 marks)

**Q5** Creatinine clearance ($Y$) is an important measure of kidney function, but it is difficult to obtain in a clinical office setting because it requires 24-hours urine collection. To determine whether this measure can be predicted from some data that are easily available, a kidney specialist obtained the data from 33 male subjects. The predictor variables are serum creatinine concentration ($X_1$), age ($X_2$) and weight ($X_3$).

Describe how to conduct the following model selection procedure using the above data as a reference (no calculation is necessary). Include graphs/plots and hypothetical results if necessary. Also discuss the advantage and disadvantage of each method.

(a) The all possible regression method using the adjusted $R^2$ criterion.

(6 marks)

(b) Forward Stepwise regression methods



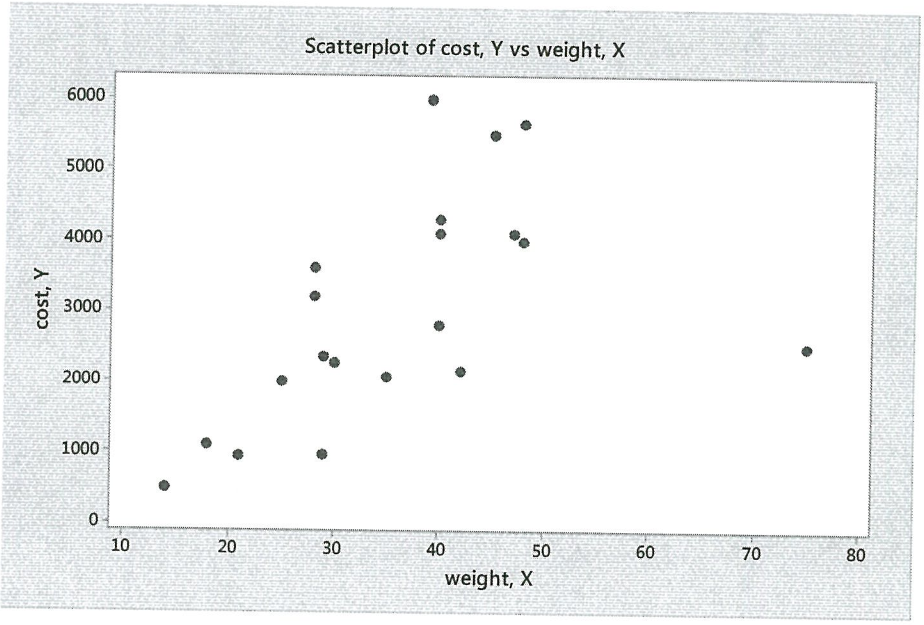
(6 marks)

**- END OF QUESTIONS –**

## FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2016/2017 PROGRAMME : BWQ
COURSE : APPLIED REGRESSION ANALYSIS COURSE CODE : BWB 20803

## APPENDIX 1




**Model Summary**[b]

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | 0.485[(a)] | 0.235 | 0.195 | 1315.9668 |

[a] Predictors : (Constant), Weight of Metal (kg)
[b] Dependent Variable : Cost per Unit (RM)

**ANOVA**[(b)]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 10119820 | 1 | 10119820 | 5.536 | 0.023[a] |
| | Residual | 32903607 | 18 | 1827978.17 | | |
| | Total | 43023427 | 19 | | | |

[a] Predictors : (Constant), Weight of Metal (kg)
[b] Dependent Variable : Cost per Unit (RM)

TERBUKA

## APPENDIX 2

| Model | Unstandardized Coefficients | |
|---|---|---|
| | $\beta$ | Std. Error |
| 1 Constant | 36.74 | 62.14 |
| $X_1$ | 8.917 | 3.044 |
| $X_2$ | 0.527 | 1.312 |
| $X_3$ | 61.46 | 13.70 |
| $X_4$ | -147.70 | 19.05 |

**Analysis of Variance**

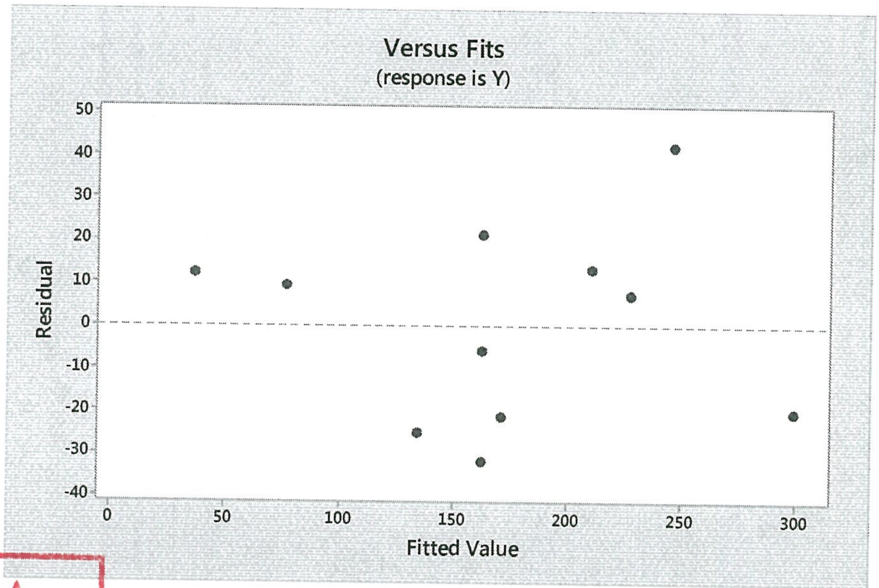
| Model | Sum of Squares | df |
|---|---|---|
| 1 Regression | 145778 | 4 |
| Residual | 1487 | 7 |
| Total | 145778 | 11 |

| SSR | Extra Sum of Squares | df |
|---|---|---|
| $X_1$ | 128143 | 1 |
| $X_2 \mid X_1$ | 1223 | 1 |
| $X_3 \mid X_1, X_2$ | 3646 | 1 |
| $X_4 \mid X_1, X_2, X_3$ | 12766 | 1 |

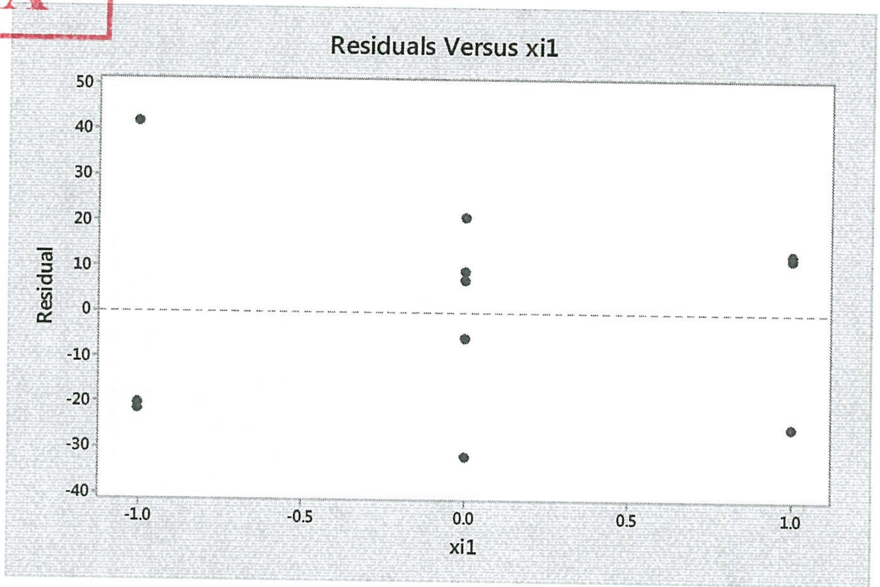
## APPENDIX 3

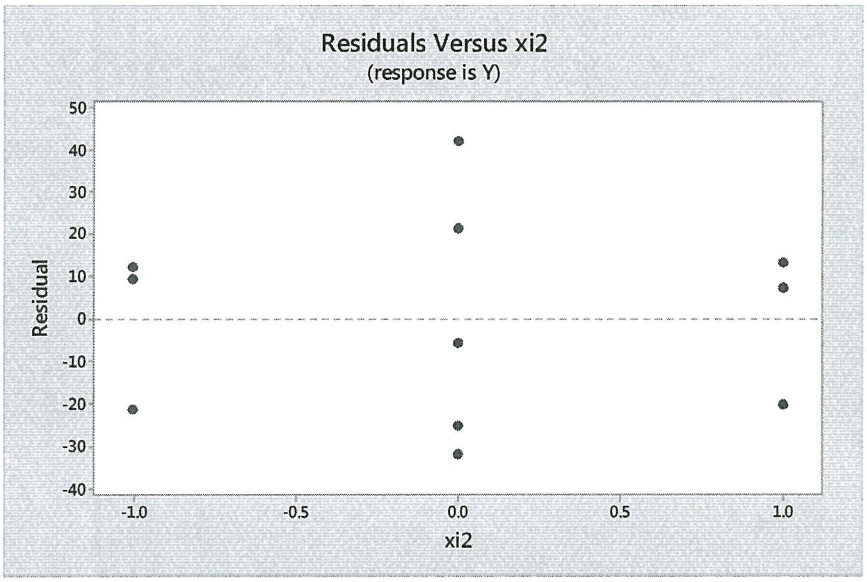
### Data Power Cells

| Sample, $i$ | $Y_i$ | $X_{i1}$ | $X_{i2}$ | Coded Values $x_{i1}$ | $x_{i2}$ | $x_{i1}^2$ | $x_{i2}^2$ | $x_{i1}x_{i2}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 150 | 0.6 | 10 | -1 | -1 | 1 | 1 | 1 |
| 2 | 86 | 1.0 | 10 | 0 | -1 | 0 | 1 | 0 |
| 3 | 49 | 1.4 | 10 | 1 | -1 | 1 | 1 | -1 |
| 4 | 288 | 0.6 | 20 | -1 | 0 | 1 | 0 | 0 |
| 5 | 157 | 1.0 | 20 | 0 | 0 | 0 | 0 | 0 |
| 6 | 131 | 1.0 | 20 | 0 | 0 | 0 | 0 | 0 |
| 7 | 184 | 1.0 | 20 | 0 | 0 | 0 | 0 | 0 |
| 8 | 109 | 1.4 | 20 | 1 | 0 | 1 | 0 | 0 |
| 9 | 279 | 0.6 | 30 | -1 | 1 | 1 | 1 | -1 |
| 10 | 235 | 1.0 | 30 | 0 | 1 | 0 | 1 | 0 |
| 11 | 224 | 1.4 | 30 | 1 | 1 | 1 | 1 | 1 |








**Regression Analysis: *Y* versus $x_{i1}$, $x_{i2}$, $x_{i1}^2$, $x_{i2}^2$, $x_1x_2$**

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 5 | 55365.6 | 11073.1 | 10.57 | 0.011 |
| Error | 5 | 5240.4 | 1048.1 | | |
| Total | 10 | 60606.0 | | | |
| | | | | | |
| Lack-of-Fit | 3 | 3835.8 | 1278.6 | 1.82 | 0.374 |
| Pure Error | 2 | 1404.7 | 702.3 | | |

Model Summary

| Root MSE | R-sq | R-sq(adj) |
|---|---|---|
| 32.3742 | 91.35% | 82.71% |

Coefficients

| Variable | DF | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|---|
| Constant | 1 | 162.8 | 16.6 | 9.81 | 0.000 | |
| xi1 | 1 | -55.8 | 13.2 | -4.22 | 0.008 | 1.00 |
| xi2 | 1 | 75.5 | 13.2 | 5.71 | 0.002 | 1.00 |
| xi1^2 | 1 | 27.4 | 20.3 | 1.35 | 0.236 | 1.08 |
| xi2^2 | 1 | -10.6 | 20.3 | -0.52 | 0.624 | 1.08 |
| x1x2 | 1 | 11.5 | 16.2 | 0.71 | 0.509 | 1.00 |

| Variable | DF | Extra Sum of Squares |
|---|---|---|
| Constant | 1 | 325424 |
| xi1 | 1 | 18704 |
| xi2 | 1 | 34202 |
| xi1^2 | 1 | 1645.9667 |
| xi2^2 | 1 | 284.9280 |
| x1x2 | 1 | 529.0 |