KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS & STATISTICS
DHAHRAN, SAUDI ARABIA

STAT 310: Linear Regression
Semester 161
Second Major Exam (Mathematical)
Monday December 05, 2016
8:00 pm

Name: ID #:

Question No	Full Marks	Marks Obtained
1	**18**	
2	**08**	
3	**08**	
4	**12**	
5	**06**	
6	**08**	
Total	**60**	

Q.No.1:- (2+4+5+7 = 18 points) Suppose we have a multiple linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$.

(a) What is the order of these matrices/vectors?

$\boldsymbol{y}$:______________ $\boldsymbol{X}$:_______________ $\boldsymbol{\beta}$:_________________ $\boldsymbol{\epsilon}$:_______________

(b) Derive the least square estimators ($\widehat{\boldsymbol{\beta}}$) for the unknown parameters ($\boldsymbol{\beta}$).

(c) Show that $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$. Also derive the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$.

(d) Let the regression model having only one regressor.

(i) Write down the following matrices/vectors for only one regressor i.e. $k = 1$:

$\boldsymbol{y} =$ $\boldsymbol{X} =$ $\boldsymbol{X}' =$

(ii) Using the matrix multiplication, mathematically show that (for only one regressor) the $\widehat{\boldsymbol{\beta}}$ vector reduces to $\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \frac{n\sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \end{bmatrix}$.

Q.No.2:- (8 points) Suppose that we want to fit the model $y = \beta_0 + \beta_1 x + \epsilon$ using weighted least squares. Assume that the observations are uncorrelated but have unequal variance. Find a general formula for the weighted least-square estimator of β_0 and β_1. Simplify as much as possible.

Q.No.3:- (2+6 = 8 points) Suppose we have a multiple linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ which is estimated using least-square technique and the estimated model is $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$. Now, the estimated error vector is defined as $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$.

(a) Show that $\boldsymbol{e} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$ where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'$ is the hat matrix.

(b) Mathematically, derive the variance-covariance matrix of $\boldsymbol{e}$.

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS & STATISTICS
DHAHRAN, SAUDI ARABIA

STAT 310: Linear Regression
Semester 161
Second Major Exam (Computational)
Monday December 05, 2016
8:00 pm

Name: ID #:

Q.No.4:- (12 points)Consider the gasoline mileage data in file data-table-B3.
Fill in the following blanks:

$(\boldsymbol{X}'\boldsymbol{X}) = \begin{bmatrix} \quad \end{bmatrix}$ $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} \quad \end{bmatrix}$ $(\boldsymbol{X}'\boldsymbol{y}) = \begin{bmatrix} \quad \end{bmatrix}$

$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \quad \end{bmatrix}$ $\boldsymbol{Var} - \widehat{\boldsymbol{Cov}}(\widehat{\boldsymbol{\beta}}) = \begin{bmatrix} \quad \end{bmatrix}$

Analysis of Variance table

Source	d.f.	SS	MS	F
Regression	2	( )	( )	( )
Error	( )	( )	( )	
Total	31	( )		

Conclusions about the regression based on F calculated:

A 95% confidence interval for β_1 is ( ) $\pm$ 2.045( )

A 95% prediction interval for a new observation on gasoline mileage when $x_1 = 257$ and $x_2 = 2$ is
( ) $\pm$ 2.045( )

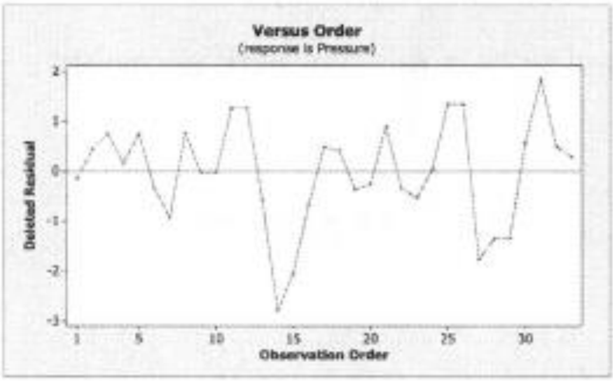
Q.No.5:- (6 points) Consider the data in data-prob-2-16. Fit a linear regression model to the data.

(a) Find the externally studentized residuals (denoted by t_i in the formula sheet).

(b) Using the residuals from part (a), construct and interpret a normal probability plot by plotting arranged t_is against the cumulative probability $P_i = \frac{(i-0.5)}{n}$, $i = 1,2,3, \dots \dots, n$.

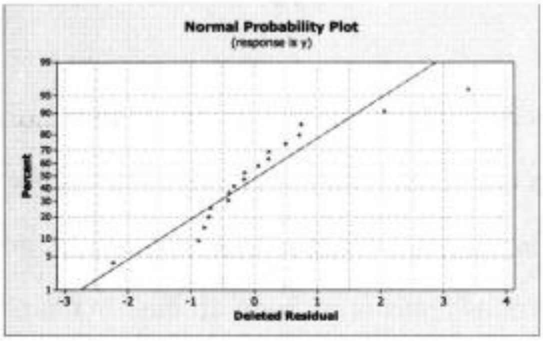
Interpretation:

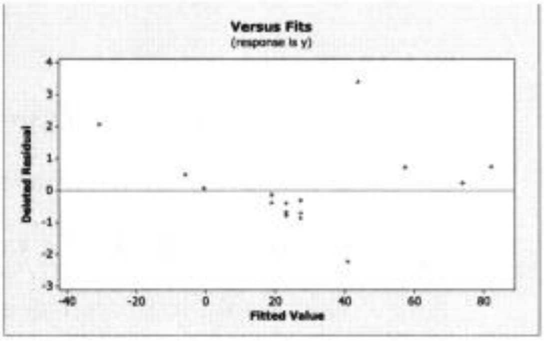
(c) Suppose that the data were collected in the order shown in the table. The plot of the residuals versus time order is given as:



Comment of this graph.

Q.No.6:- (8 points) Consider the methanol oxidation data in Table B20. Perform a thorough analysis of these data. The normal probability and the residual plots are given as follows:





(a) Comments on Normal probability plot:

(b) Comments on residual plot:

(c) Perform a log transformation on the response percent conversion only. Fit the linear regression model.

(i) Are all the variables in this transformed model significant? If yes, then construct the normal probability and the residual plots and comment.

(ii) If the coefficient of any variable is insignificant, then remove that variable and fit the regression model again. Now construct the normal probability and the residual plots and comment.