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

STAT 310: Regression Analysis

Term 211, Second Major Exam
Monday November 22, 2021, 06:00 PM

Name: ________________________________ ID #: ____________________

Question No	Full Marks	Marks Obtained
1	**08**	
2	**08**	
3	**18**	
4	**11**	
5	**6 (Bonus)**	
Total	**45**	

Instructions:

1. Formula sheet will be provided to you in exam. You are not allowed to bring, with you, formula sheet or any other printed/written paper.
2. Mobiles are not allowed in exam. If you have your **mobile** with you, **turn it off** and put it **under your seat** so that it is visible to proctor.
3. Show all the calculation steps. There are points for the steps so if your miss them, you lose points.
4. Derive every result that you use in your solution, unless mentioned otherwise.
5. Anything bold in a question indicates that it is a vector or matrix.

[Blank Page]

Q1: (4+4 = 8 points) Consider a multiple linear regression model $\boldsymbol{y}_{n\times1} = \boldsymbol{X}_{n\times(k+1)}\boldsymbol{\beta}_{(K+1)\times1} + \boldsymbol{\epsilon}_{n\times1}$ where the OLS estimates are given as $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'$.

a) Show that the vector of residuals can be written as $\boldsymbol{e}_{n\times1} = (\boldsymbol{I}_{n\times n} - \boldsymbol{H}_{n\times n})\boldsymbol{\epsilon}_{n\times1}$.

b) Mathematically, prove that $\text{Var}(\boldsymbol{e}_{n\times1}) = \sigma^2(\boldsymbol{I}_{n\times n} - \boldsymbol{H}_{n\times n})$

In case you need more space, use page 2

Q2: (8 points) Consider a multiple linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$ where β_0, β_1 and β_2 are unknown parameters and x_i's are fixed. We transform all the variables using unit length scaling and fit the model $y_i^0 = b_1 w_{i1} + b_2 w_{i2} + \varepsilon_i$ which can be written in the matrix notation as:

$$\boldsymbol{y}^0 = \boldsymbol{W}\boldsymbol{b} + \boldsymbol{\varepsilon}.$$

Mathematically, show that the diagonal elements of $(\boldsymbol{W'W})^{-1}$ matrix are equal to $\frac{1}{1-r_{x_1x_2}^2}$ where $r_{x_1x_2} = \frac{\sum[(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)]}{\sqrt{\sum(x_{i1}-\bar{x}_1)^2 \times \sum(x_{i2}-\bar{x}_2)^2}}$ is the correlation coefficient between x_1 and x_2.

$w_{ij} = \frac{x_{ij}-\bar{x}_j}{\sqrt{s_{jj}}}$, $y_i^0 = \frac{y_i-\bar{y}}{\sqrt{s_{yy}}}$, $s_{jj} = \sum_{i=1}^{n}\left(x_{ij}-\bar{x}_j\right)^2$, $s_{yy} = \sum_{i=1}^{n}(y_i-\bar{y})^2$ $\forall\ i = 1,2,\dots,n$ and $j = 1,2$

Good Luck

Report at least 6 decimal points.

Q3: Code_____
Write code # before starting.
Attached file contains the data on gasoline mileage performance for 25 automobiles where the description of variables is as follows:
y→ Miles/gallon
x_1→ Compression ratio
x_2→ Rear axle ratio
x_3→ Overall length (in.)
x_4→ Weight (lb)

Fit the model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$ and test the following linear constraints:
H_0: $2\beta_1 - \beta_2 = 1$ and $\beta_3 + 10\beta_4 = 0.1$
H_1: At least one constraint in H_0 is not true.
Answer the following questions:

(4 pts.)

$\boldsymbol{T} = \left[\quad \right], \boldsymbol{c} = \left[\quad \right]$

(3 pts.) $F =$________________

(2 pts.) $p - value =$_______________

(1 pt.) Decision and conclusion:

Perform a thorough influential analysis on the given data and answer the following questions:

(2 pts.) Cook's D_9 = _________________ Comment:

(2 pts.) $DFBETAS_{2,13}$ = _________________ Comment:

(2 pts.) $DFFITS_{17}$ = ___________________ Comment:

(2 pts.) $COVRATIO_{21}$ = __________________ Comment:

Report at least 6 decimal points.

Q4: Code_____
Write code # before starting.
Attached file contains the data on response variable y and predictor x. Suppose that the relationship between y and x is intrinsically linear and is given as:

$$y = \sqrt{\cos^{-1}\left[\beta_0 + \beta_1 \sin\sqrt{X}\right]}.$$

Transform the variables such that the relationship becomes linear. The new variables are

(4 pts.) $y' =$____________________ and $x' =$____________________

Fit a simple linear regression model on the transformed variables. The fitted model is given as:

(2 pts.) $\hat{y}' = [$_________$] + [$_________$]x'$

Predict the original response y when $x = 0.9$. Also construct a 95% prediction interval for original y when $x = 0.9$.

(2 pts.) $\hat{y}_{x=0.9} =$_____________

(3 pts.) Lower Prediction Limit = __________________ Upper Prediction Limit = _______________

[Good Luck]