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

STAT 310: Regression Analysis

Term 211, Final Exam
Sunday December 02, 2022, 07:00 PM

Name: _____ ID #: _____

Question No	Full Marks	Marks Obtained
1	**06**	
2	**06**	
3	**05**	
4	**12**	
5	**10**	
6	**06**	
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: (6 points) Suppose we have the following sample of size $n = 3$ on response variable (y) and predictor (X):

X	y
x_1	y_1
x_2	y_2
x_3	y_3

We fit a second-degree polynomial regression model to these data i.e. $y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$. Assuming that $x_2 = x_3$, derive the variance inflation factor.

Q2: (3+3 = 6 points) Consider a multiple linear regression model $\boldsymbol{y}_{n\times 1} = \boldsymbol{X}_{n\times(k+1)}\boldsymbol{\beta}_{(K+1)\times 1} + \boldsymbol{\epsilon}_{n\times 1}$

a) Show that $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$ where $\boldsymbol{e}$ is the residuals vector obtained from OLS method and $\widehat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$.

b) Using result in part a), prove that the sum of residuals obtained through OLS estimation is always zero i.e. $\sum_{i=1}^{n} e_i = 0$.

Q3: (5 points) Mathematically show that maximizing R^2_{adj} is equivalent to minimizing MSE.

Good Luck

STAT 310: Regression Analysis

Term 211, Final Exam - Computational

Name: _____ ID #: _____

Q4: Code_____ **Report at least 4 decimal points.**

(12 points) Based on the performance of students in 1st Major exam, an instructor wants to predict the students' scores in 2nd Major exam. The data on the scores of students from two sections (i.e. Section 1 and Section 2) are given in the Excel file. Fit a single regression equation to predict the students' score in 2nd Major exam based on the performance in 1st Major exam. The regression equation should have the capacity to accommodate change in intercept and change in slope of lines for both sections.
Define your variables below:

$y \longrightarrow$

$X_1 \longrightarrow$

$X_2 \longrightarrow$

Write your regression equation here $\hat{y} =$

The predicted Major 2 score of a student from Section 1 who scored 62 in Major 1 is _____

Test the hypothesis that the regression lines for both sections are parallel. Use $\alpha = 0.05$

H_0: _____ H_1: _____

Test Statistic: $F/T =$ _____

Degrees of freedom:

P-value: _____

Decision: Circle the correct option $\longrightarrow$ $a)$ Reject H_0 / $b)$ Fail to reject H_0

Conclusion:

Q5: Code_____ **Report at least 4 decimal points.**

(10 points) A researcher wants to predict the quality of coffee (y). The candidate variables in the model are Aroma (X_1), Body (X_2), Flavor (X_3) and Oakiness (X_4).

a) Using the stepwise regression, reach to a best model using $\alpha_{IN} = \alpha_{OUT} = 0.1$.

Step 1: Variable Entered and its p-value: _____

Step 2: Variable Entered and its p-value: _____

Variable deleted and its p-value: _____

Step 3: Variable Entered and its p-value: _____

Variable deleted and its p-value: _____

Step 4: Variable Entered and its p-value: _____

Variable deleted and its p-value: _____

Final Model: $\hat{y} =$ _____

If $\alpha_{IN} = \alpha_{OUT} = 0.05$, final model is $\hat{y} =$ _____

b) Explain the reason why we cannot fix $\alpha_{OUT} < \alpha_{IN}$ in stepwise regression?

Q5: Code_____ **Report at least 4 decimal points.**

(6 points) Download the Excel file for this question containing the data on two variables y and X. Fit a linear spline to these data to predict y using two knots i.e. $k_1 = 15$ and $k_2 = 29$. The line should be continuous at the knots. Also, predict y when $x = 35$.

Hint: Fit the model: $y = \beta_0 + \beta_1 X + \gamma_1 S_1 + \gamma_2 S_2 + \epsilon$ where

$S_1 = \begin{cases} x - 15, & x > 15 \\ 0, & x \leq 15 \end{cases}$ and $S_2 = \begin{cases} x - 29, & x > 29 \\ 0, & x \leq 29 \end{cases}$

Final fitted model:

$\hat{y} =$ _____ + _____(X) + _____(S_1) + _____(S_2)

Predicted y when $x = 35$: