

---

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

**STAT 310: Linear Regression**

Semester 161

Final Exam

Wednesday January 11, 2017

7:00 pm

Name: \_\_\_\_\_

ID #: \_\_\_\_\_

---

<b>Question No</b>	<b>Full Marks</b>	<b>Marks Obtained</b>
1	15	
2	10	
3	10	
4	10	
5	13	
6	12	
<b>Total</b>	<b>70</b>	

Q.No.1:- (3×5 = 15 points)

(a) The coefficient of determination for a multiple linear regression with  $n = 10$  observations and  $k = 5$  predictors is equal to  $R^2 = 0.86$  and the F-value is equal to  $F_0 = 20.5$ . What is the value of the adjusted coefficient of determination  $R_{adj}^2$ ?

(b) We consider the model  $y = \beta_0 + \beta_1 x + \epsilon$ . Let  $[-0.01, 1.5]$  be the 95% confidence interval for  $\beta_1$ . In this case, test the significance of  $\beta_1$  using t-test at 1% level of significance.

(c) Consider the linear regression model where  $\mathbf{Y}$  is the vector of observations and  $\mathbf{X}$  is the vector of predictor variables. Also,  $\hat{\boldsymbol{\beta}}$  be the vector of OLS estimates. Show mathematically that  $SSE = \mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$

Q.No.2:- ( $2 \times 5 = 10$  points) The economic structure of Major League Baseball allows some teams to make substantially more money than others, which in turn allows some teams to spend much more on player salaries. These teams might therefore be expected to have better players and win more games on the field as a result. Suppose that after collecting data on team payroll (in millions of dollars) and season win total for 2010, we find a regression equation of  $Wins = 71.87 + 0.101Payroll - 0.060League$  where  $League$  is an indicator variable that equals 0 if the team plays in the National League or 1 if the team plays in the American League.

(a) Suppose that Teams A and B both play in the same league, and Team A's payroll is \$1 million higher than Team B's. On average, we would expect Team A to win how many games more than Team B?

(b) Suppose that Teams A and B have the same payroll, but Team A plays in the National League while Team B plays in the American League. On average, we would expect Team A to win how many games more than Team B?

(c) Suppose we plotted the data and drew the regression lines for National League and American League teams. What would be the slope of the line for American League teams?

(d) Suppose we plotted the data and drew the regression lines for National League and American League teams. What would be the intercept of the line for American League teams?

(e) One American League team in the data set had a payroll of \$108 million and won 88 games. Calculate the residual for this observation.

Q.No.3:- (10 points) Suppose that we want to fit the model  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \epsilon$  to a data that has following summary:

$$n = 10, \quad \sum y = 1057, \quad \sum x_1 = 93, \quad \sum x_2 = 97, \quad \sum x_1y = 12023, \quad \sum x_2y = 11107, \\ \sum x_1^2 = 1421, \quad \sum x_2^2 = 1037, \quad \sum x_1x_2 = 972.$$

Compute the least square estimates for  $\beta_0$ ,  $\beta_1$  and  $\beta_2$ .



Q.No.4:- (3+2+5 = 10 points) Download the data for this question from Blackboard. Run the standardized regression using unit length scaling.

(a) Find the variance inflation factor for all the 8 regressors in the model.

$X_1$ : \_\_\_\_\_  $X_2$ : \_\_\_\_\_  $X_3$ : \_\_\_\_\_  $X_4$ : \_\_\_\_\_

$X_5$ : \_\_\_\_\_  $X_6$ : \_\_\_\_\_  $X_7$ : \_\_\_\_\_  $X_8$ : \_\_\_\_\_

(b) Write  $W'y$  matrix and interpret its elements.

(c) Run the ridge regression and find the optimal value of  $k$ . Predict the value of  $y$  from the finally selected model when  $x_1=0$ ,  $x_2=55.3$ ,  $x_3=814.2$ ,  $x_4=2.15$ ,  $x_5=200$ ,  $x_6=320$ ,  $x_7=70$  and  $x_8=15$ .

Q.No.5:- (5+5+3 = 13 points) The excel file contains the data on the thrust of a jet turbine engine and six candidate regressors.

(a) Use stepwise regression method to find the best model at the significance levels  $\alpha_{IN} = \alpha_{OUT} = 0.1$ . Write the details of every step separately i.e. which variable is added/deleted and why.

(b) Make the residual plot and normal probability plot (using externally studentized residuals) for the best model selected in part (a). Comment on these plots. Also find the  $R^2_{\text{prediction}}$  for the best model.

(c) For the best model, do we have any influential observation in the data? Clearly mention the method (and all its steps) you used for answering this question.



Q.No.6:- (3+2+2+5 = 12 points) Twenty observations on Tool life,  $y$  (hours), lathe speed,  $x_1$  (rpm) and tool type are given in the excel sheet. Fit a single regression model assuming that we need two different regression lines to adequately model these data, with both intercept and slope depending on the tool type.

(a) Write down your estimated model.

(b) What is the average change in tool life due to a unit change in lathe speed for Tool A?

(c) What is the average change in tool life due to a unit change in lathe speed for Tool B?

(d) Test whether or not the two regression lines are identical.

$H_0$ :

$H_1$ :

$F_0$ =

with  $v_1$ =

and  $v_2$ =

P-value=

Decision and conclusion: