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

STAT 310: Linear Regression
Term 171, Final Exam (Mathematical)
Thursday January 04, 2018 (7:00 pm)

Name: ID #:

Question No	Full Marks	Marks Obtained
1	**08**	
2	**06**	
3	**06**	
4	**06**	
Total	**26**	

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. Make sure you have 12 unique pages of exam paper (including this title page).
4. Show all the calculation steps. There are points for the steps so if your miss them, you would lose points.

Blank page

Q.No.1: - (4+4 = 8 points) Consider a simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where β_0, β_1 are unknown parameters and x_i's are fixed.

a) Derive the point estimates of β_0 and β_1 using the method of ordinary least squares.

b) The i[th] *residual* is defined to be the difference between the observed and fitted value of the response for point i.

$$e_i = y_i - \hat{y}_i$$

Mathematically, Show that

i. $\sum \hat{y}_i \, e_i = 0$

ii. The regression line passes through $(\bar{y}, \bar{x})$

Q.No.2: - (2+4 = 6 points) Consider the following data and assume that steam usage and temperature are jointly normally distributed.

Temp (X)	21	24	32	47	50	59	68	74	62	50	41	30
Usage (Y)	185.79	214.47	288.03	424.84	454.68	539.03	621.55	675.06	562.03	452.93	369.95	273.98

$\sum x = 558$ $\sum y = 5062.34$ $\sum x^2 = 29256$ $\sum y^2 = 2416234.61$ $\sum xy = 265869.63$

(a) Find and interpret the correlation between steam usage and monthly average ambient temperature.

(b) Test the hypothesis that there is a strong positive linear association between the stem usage and the temperature. Assume $\alpha = 0.05$ and write down all the testing steps.

$\tanh\left(\tan^{-1} r - \frac{Z_{\alpha/2}}{\sqrt{n-2}}\right) \leq \rho \leq \tanh\left(\tan^{-1} r + \frac{Z_{\alpha/2}}{\sqrt{n-2}}\right)$	$T = \frac{r\sqrt{n-2}}{(1-r^2)}$

Q.No.3:- (6 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 $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y}$. Now, the estimated residual vector is defined as $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$. Mathematically, derive the variance-covariance matrix of $\boldsymbol{e}$.

Q.No.4: - (6 points) Consider a multiple linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ where r_{12} represents the simple correlation coefficient between x_1 and x_2. We transform all the variables using the unit length scaling and the resulting model is $y^0 = \gamma_1 w_1 + \gamma_2 w_2 + \varepsilon$. This model can also be written in matrix notation as $\boldsymbol{y^0} = \boldsymbol{\gamma w} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\gamma} = \begin{bmatrix}\gamma_1 \\ \gamma_2\end{bmatrix}$, $\boldsymbol{\hat{\gamma}} = (\boldsymbol{w'w})^{-1}\boldsymbol{w'y^0}$ and $\textbf{Var} - \textbf{Cov}(\boldsymbol{\hat{\gamma}}) = \sigma^2(\boldsymbol{w'w})^{-1}$.

It is to be noted that $\boldsymbol{w'w}$ is a correlation matrix in terms of original variables i.e. $\boldsymbol{w'w} = \begin{bmatrix}1 & r_{12} \\ r_{12} & 1\end{bmatrix}$

If there is a perfect linear relationship between x_1 and x_2 (i.e. $r_{12} = \pm 1$) then mathematically show that:

a) $\mathrm{Var}(\hat{\gamma}_1) = \mathrm{Var}(\hat{\gamma}_2) = \infty$

b) $\mathrm{Cov}(\hat{\gamma}_1, \hat{\gamma}_2) = \pm\infty$

Blank page