Introduction to Linear Models and Regression

Full Marks: 30 Time : 2.5 hrs

Answer question no. 1 and from the rest, as many as you want, but the maximum you can score is 30.

1. Following are the scores of 8 students in pre-test (x3), test (x2) and final examination (x1).
   i. Fit a suitable multiple linear regression equation.
   ii. Check how good the regression equation will perform?
   iii. Among the pre-test and test scores, which one has closer alignment with the final scores? (3+2+3)

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Pre-test	43	38	27	28	35	21	19	13
Test	22	29	23	33	20	8	17	19
Final	66	38	55	63	25	17	33	18

2. Justify the statement with suitable derivations: "Although error variables of a simple linear model are assumed to be homoscedastic and mutually independent, the residuals generated due to the use of least square approach to estimate the regression coefficients (of the same model), may violate one or more of the assumptions". (5)
3. Suppose $X_1, X_2, \ldots, X_k$ are k discrete random variables having multinomial distribution with parameters n, $p_1, p_2, \ldots, p_k$. Find the expression for the partial correlation between X_1 and X_2 keeping rest of the variables fixed. (5)
4. Prove – "The role of the inverse of dispersion matrix in Mahalanobis distance is to re-scale the original variables to be homoscedastic and to make the new variables uncorrelated". (5)
5. Suppose $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$. The find the distribution of $\mathbf{X}' \Sigma^{-1}\mathbf{X}$. (5)
6. Suppose $r_{1.23..p}$ denotes the multiple correlation coefficient between X_1 (response variable) and $X_2, X_3, \ldots, X_p$. Show that, $r_{1.23\ldots p}$ will be zero, iff the pair-wise correlation coefficients of X_1 with the rest of the variables is 0. Can you think of one or more cases triggering this situation? (5+2)