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. Consider the following summary statistics as obtained from a dataset with three variables x_1, x_2, x_3.
   $\overline{x_1} = 68$, $\overline{x_2} = 70$, $\overline{x_3} = 74$, $s_1^2 = 100$, $s_2^2 = 25$, $s_3^2 = 81$, $r_{12} = 0.6$, $r_{13} = 0.7$, $r_{23} = 0.65$.
   i) Construct the regression equation for x_1 on x_2 and x_3.
   ii) Compute the multiple correlation coefficient of x_1 on x_2 and x_3
   iii) Compute the partial correlation coefficient between x_1 and x_3.
   iv) Give your comments. 3+3+3+3= 12

2. Show that if in a p-variate distribution, all the pair-wise correlations are equal to ρ, then $\geq -\frac{1}{p-1}$. (8)

3. If **X** ~ N_q (**0**, I_p) and P_1 is a m X_p matrix such that, $P_1P_1' = I_m$, then find the distributions of
   i) **Z** = P_1**X**
   ii) **U** = $\frac{1}{\sigma^2}$ (**X'X** – **Z'Z**).
   iii) Show that **Z** and **U** are mutually independent.
   (3+3+2)

4. Let $X_1, X_2, \ldots, X_k$ be independent Poisson variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ respectively. Show that the conditional distribution of the sum of $X_1, X_2, \ldots, X_{k-1}$ given $X_1 + X_2 + \ldots + X_k = n$ follows multinomial distribution. (6)

5. Prove that under a multiple linear regression set-up, the regressed values of the response variable are independent of the error in regression. (6)