Linear Statistical Models

Week-3: Graded Assignment

Subjective Assignment: (Manual-grading) **Max. Marks: 35**

NOTE: All the plots should be properly labelled.

1. Consider the Linear Model:

$$y_{ij} = \mu + \mu_i + \epsilon_{ij} \quad ; \quad 1 \leq j \leq n_i \, , \, 1 \leq i \leq 3$$

Feel free to use the draft code available in the 'Supplementary Contents' → 'Week-3' on the portal.

(a) If $\epsilon_{ij} \sim N(0, 1)$, then find the distribution of y_{ij}. Elaborate on your answer. [3 Marks]

(b) Generate possible random values using R for μ, μ_i and n_i as generated in code draft provided. [1 Mark]

Note: *The value of n_i's should be greater than or equal to 100.*

(c) Using R, generate the dataset, i.e. y_{ij}, and store it in a data frame. [3 Marks]

(d) Using R, find the mean of y_i's denoted by y_{io} using the generated dataset.

Hint: $y_{io} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ [1 Mark]

(e) Using R, compute the following: [3 Marks]

i. Sum of squares within group, i.e. SSW

ii. Sum of squares between group, i.e. SSB

iii. Total sum of squares, i.e. TSS

(f) Using R, verify that $TSS = SSB + SSW$. [1 Mark]

(g) Iterate the steps in part $(c), (d)$ and (e) more than 100 times using R. [5 Marks]
Hint: *Randomly generate a number greater than 100, and perform that many iterations.*

(h) Using R, store the values of SSW, SSB and TSS computed in part (g) in a data frame. [2 Marks]

(i) Using R, plot the histogram of values obtained for each of SSW, SSB and TSS. [3 Marks]

(j) In each of the above plotted histogram, fit the density curve of 'Chi-square Distribution' to verify the following: [8 Marks]

i. $SSW \sim \chi^2$ with degree of freedom $= 2$.

ii. $SSB \sim \chi^2$ with degree of freedom $= n - 3$.
where, $n = n_1 + n_2 + n_3$

iii. $TSS \sim \chi^2$ with degree of freedom $= n - 1$.

Also, comment on each of the obtained histograms.

2. Suppose $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ be a linear model, where

$$\underset{\sim}{Y} = (y_1, y_2, y_3, \ldots, y_n)^T,$$

$$\underset{\sim}{\beta} = (\beta_1, \beta_2, \beta_3, \ldots, \beta_m)^T,$$

$$X = ((x_{ij})) \;\; ; 1 \leq i \leq n, 1 \leq j \leq m$$

and

$$\underset{\sim}{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3, \ldots, \epsilon_n)^T$$

where, $\underset{\sim}{\epsilon}$ has mean $\underset{\sim}{0}$ and variance covariance matrix $\sigma^2 I_{n \times n}$.

Let $\underset{\sim}{p}^T \underset{\sim}{\beta} = p_1\beta_1 + p_2\beta_2 + \ldots + p_m\beta_m \;\forall\; \underset{\sim}{p} \in \mathbb{R}^m$.

Suppose there is a $\underset{\sim}{l} \in \mathbb{R}^n$ such that:

$$E[\underset{\sim}{l}^T \underset{\sim}{Y}] = p^T \underset{\sim}{\beta}, \;\; \forall \;\underset{\sim}{\beta} \in \mathbb{R}^m$$

then show that there is a $\underset{\sim}{l} \in \mathbb{R}^n$ such that:

$$\underset{\sim}{l}^T X \beta = p^T \beta, \;\; \forall \;\underset{\sim}{\beta} \in \mathbb{R}^m$$

[5 Marks]