Class Activity Problems

Probability and Statistics 2022
Indian Institute of Management, Udaipur

21st July, 2022

Name:

Name:

- There are are a total of 114 points in this Question Paper. Answer as much as you can. If your acquired score is greater than equal to 100, it will be counted as 100%.
- The Exam is scheduled for 3 hours. "Time Left" reminders will be posted in 1.5 hrs, 2:30 hrs, 2:45 hrs from the beginning of the Exam time.
- There are three $\star$ marked problems that are more involved than the rest. In case you are stuck in one of those, it might be a good idea to consider solving other problems first and then continue with the * marked problems.
- You may take help from the "Exam Assistance Note" containing a few required definitions, lemma and theorem statements.

1. (a) Let Y be Response variable, X_1, X_2 denote the Explanatory variables, ε be unknown random errors and $\beta_1, \ldots, \beta_3$ are unknown parameters of interest. Determine whether the following relationship equation is a linear model. Relationship Equation: $Y = \beta_0 + \beta_1 \sin(X_1) + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$.

Score:
Total Score: 5

Ans:

☐ Not Linear Model ☐ Linear Model

(b) Identify if the following matrix is a Orthogonal Projection matrix.

$$M = \frac{1}{9}\begin{bmatrix} 3 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 1 \end{bmatrix}$$

Score:
Total Score: 5

Ans:

☐ Orthogonal Projection ☐ Not Orthogonal Projection

(c) Consider the matrices $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Does the matrix A have a Generalized Inverse ? If yes, Construct a generalized inverse of A.

Score:
Total Score: 1+4

Ans:

(d) Consider the model $\underset{\sim}{\mathbf{y}} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, where $\underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & -1 \end{bmatrix}$ and $\underset{\sim}{\varepsilon}$ is a mean zero error vector with variance co-variance matrix $\sigma^2 I_{4\times 4}$. Prove that the parameter β_1 is not estimable

Score: ______

Total Score: 7

Ans:

2. A spring balance is used to weigh three objects with unknown weights β_1, β_2 , β_3, β_4. The objective is to estimate the $\underset{\sim}{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{bmatrix}^T$. Assume that each the measurements in the spring balance are subject to (independent) Normally distributed random errors with mean 0 and unknown variance $\sigma^2 > 0$. Consider the data, when each object are weighted twice to get observations

$$\underset{\sim}{\mathbf{Y}} = \begin{bmatrix} y_{1,1} & y_{1,2} & y_{1,3} & y_{2,1} & \dots & y_{4,2} & y_{4,3} \end{bmatrix}^T,$$

where $y_{i,j}$ denotes the j^{th} replicated measurement corresponding to the experiment when i^{th} object is placed on the spring balance. Note that $\underset{\sim}{\mathbf{Y}} \in \mathbb{R}^{12}$.

(a) Represent the estimation problem in terms of the matrix notation $\underset{\sim}{\mathbf{Y}} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ by identifying the response vector $\underset{\sim}{\mathbf{Y}}$, design matrix $\mathbf{X}$, regression coefficients $\underset{\sim}{\beta}$ and the vectors of the random errors $\underset{\sim}{\varepsilon}$.

Score: ______
Total Score: 2+ 5+2+1

Ans:

(b) Prove that all the linear parametric functions of the parameters $\underset{\sim}{\beta}$ are estimable for the associated design matrix.

Score: ______
Total Score: 5

Ans:

(c) Find the **Orthogonal Projection Matrix** for $\mathscr{C}(\mathbf{X})$, the column space of $\mathbf{X}$. (Justify your steps with the reference to the results/theorem/lemma you are using.)

Score: ____
Total Score: 8

Ans:

(d) Consider a linear parametric function $\theta = \beta_1 + \beta_2 + \beta_3$. Identify a vector $\underset{\sim}{\lambda}$ such that $\theta = \underset{\sim}{\lambda}^T \underset{\sim}{\beta}$.

Score: ____
Total Score: 2

Ans:

(e) Find the Best Linear Unbiased Estimator for θ (Show your steps and justify your steps with the reference to the results/theorem/lemma you are using.)

Score: ______
Total Score: 7

Ans:

(f) A practitioner is interested in estimating all the parameters $\beta_1, \beta_2, \beta_3, \beta_4$ simultaneously. write down the definition of $(1-\alpha)100\%$ simultaneous confidence intervals for the parameters.

Score: ____
Total Score: 5

Ans:

(g) $\star$ Which type of simultaneous Confidence interval would you prefer for the parameters $\{\beta_1, \beta_2, \beta_3, \beta_4\}$ and why?

Score: ____
Total Score: 5

Ans:

(h) Construct a set of 95% simultaneous confidence intervals for the parameters $\{\beta_i\}_{i=1}^{4}$.

Score: ____
Total Score: 8

Ans:

3. Consider the standard linear model $\underset{\sim}{\mathbf{Y}} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ where $\underset{\sim}{\varepsilon}$ is Normally distributed ($\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{\mathbf{0}}, \sigma^2 V)$), where V is a known positive definite matrix. Assume the design matrix $\mathbf{X}$ has full column rank. Also, assume that σ is a known positive constant.

(a) Show that the log likelihood function can be represented as

$$l_{\underset{\sim}{\mathbf{Y}}}(\underset{\sim}{\beta}) = K(V, \sigma^2) - \frac{(\underset{\sim}{\mathbf{Y}} - \mathbf{X}\underset{\sim}{\beta})^T V^{-1} (\underset{\sim}{\mathbf{Y}} - \mathbf{X}\underset{\sim}{\beta})}{2\sigma^2}$$

where $K(V, \sigma^2)$ is a constant that does not involve $\underset{\sim}{\beta}$.

Score: ____
Total Score: 5

Ans:

(b) Use part(a) to derive $\hat{\underset{\sim}{\beta}}_{MLE}$, the Maximum Likelihood Estimator for the parameter $\underset{\sim}{\beta}$.

Score: ___
Total Score:7

Ans:

4. Consider a linear model given by

$$\mathbf{Y} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

where $\mathbf{Y} \in \mathbb{R}^n, \underset{\sim}{\beta} \in \mathbb{R}^p$ and $\underset{\sim}{\varepsilon}$ is a Normally distributed random vector with mean $\underset{\sim}{\mathbf{0}}$ and variance $\sigma^2 I_{n\times n}$ and $\mathbf{X}$ is an $n \times p$ matrix with rank $r < p < n$ (i.e. The design matrix $\mathbf{X}$ does not have full column rank). The parameter σ^2 is an unknown positive number. Assume that $\underset{\sim}{\varepsilon} \sim N\left(\underset{\sim}{\mathbf{0}}, \sigma^2 I_{n\times n}\right)$. Let $\mathscr{C}(\mathbf{X})$ denotes the column space of $\mathbf{X}$, the vector space containing all possible linear combination of the column vectors of the matrix $\mathbf{X}$. It can be shown that $P = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ is the **Orthogonal Projection** matrix for the $\mathscr{C}(\mathbf{X})$, where $(\mathbf{X}^T\mathbf{X})^{-}$ is any Generalized inverse of $(\mathbf{X}^T\mathbf{X})$.

(a) Show that $E(\mathbf{Y}^T(I_{n\times n} - P)\mathbf{Y}) = (n-r)\sigma^2$.

Score: ______
Total Score: 5

Ans:

Prove that the statistics $\mathbf{Y}^T\mathbf{P}\mathbf{Y}$ and $\mathbf{Y}^T(\mathbf{I}_{n\times n} - \mathbf{P})\mathbf{Y}$ are independent (Mention if you are using any result).

(b)

Score: ____
Total Score:5

Ans:

Derive the distribution of $\frac{(\mathbf{Y}^T P\mathbf{Y})/r}{(\mathbf{Y}^T(I_{n\times n}-P)\mathbf{Y})/(n-r)}$? (Show your steps)

(c)

Score: ____
Total Score: 10

Ans:

(d) $\star$ Let $\underset{\sim}{\widehat{\beta}}_{LSE}$ be the **Least Square Estimator** for the parameter $\underset{\sim}{\beta}$. Show that $(\underset{\sim}{\mathbf{Y}} - \mathbf{X}\underset{\sim}{\beta})^T P(\underset{\sim}{\mathbf{Y}} - \mathbf{X}\underset{\sim}{\beta}) = (\underset{\sim}{\widehat{\beta}}_{LSE} - \underset{\sim}{\beta})^T \mathbf{X}^T \mathbf{X}(\underset{\sim}{\widehat{\beta}}_{LSE} - \underset{\sim}{\beta})$.

Score: ____
Total Score: 5

Ans:

(e) $\star$ Derive the distribution of $\frac{\|X(\underset{\sim}{\widehat{\beta}}_{LSE} - \underset{\sim}{\beta})\|^2}{\sigma^2}$. (You may use the relation in part(d) to get your answer.)

Score: ______
Total Score: 5

Ans: