Preliminary Examination: LINEAR MODELS

Answer all questions and show all work. Students won't get full credits if there is not sufficient justification or explanations in the answers.
Q1 is 30 points; Q2 is 35 points, and Q3 is 35 points.

1. Let $Y \in \mathbb{R}$ be a random response, and let $\mathbf{x} \in \mathbb{R}^p$ be a *random* predictor. We would like to find a linear predictor $\alpha + \mathbf{x}^T\boldsymbol{\beta}$ that minimizes $E[(Y - \alpha - \mathbf{x}^T\boldsymbol{\beta})^2]$ over all choice of $\alpha \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$. Such a predictor is called a *best linear predictor* of Y.

   Define $\mu_y = E(Y)$, $\boldsymbol{\mu}_x = E(\mathbf{x})$, $\sigma_y^2 = \text{Var}(Y)$, $\boldsymbol{\Sigma}_{xx} = \text{Cov}(\mathbf{x})$, and $\boldsymbol{\sigma}_{xy} = \text{Cov}(\mathbf{x}, Y)$. Without loss of generality, we will write an arbitrary linear predictor in the form $\alpha + (\mathbf{x} - \boldsymbol{\mu}_x)^T\boldsymbol{\beta}$.

   a. Show that the optimal choice of α is simply $\hat{\alpha} = \mu_y$.

   b. Show that if $\boldsymbol{\beta}^*$ is a solution to the linear system $\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} = \boldsymbol{\sigma}_{xy}$, then $\mu_y + (\mathbf{x} - \boldsymbol{\mu}_x)^T\boldsymbol{\beta}^*$ is a best linear predictor of Y. *Without loss of generality, you may assume $\mu_y = 0$ and $\boldsymbol{\mu}_x = \mathbf{0}$ for this part.*

   c. Now add the assumption that the joint distribution of Y and $\mathbf{x}$ is $(p+1)$-dimensional multivariate normal distribution with a $(p+1) \times (p+1)$ positive definite covariance matrix. Give the conditional mean of Y given $\mathbf{x}$. Comment on how this conditional mean is related to the best linear predictor of Y.

2. Consider a model given by

$$y_{ij} = \mu_i + \epsilon_{ij}; \quad i = 1, ..., n, \ j = 1, 2, 3,$$

   where $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, and let $\mathbf{Y} = (y_{11}, y_{12}, y_{13}, ..., y_{n1}, y_{n2}, y_{n3})^T$. We call this our Model 1 and denote its sum of squares of errors by SSE_1.

   Now, consider an alternative model, referred to as Model 2:

$$y_{ij} = \beta_0 + \beta_1 x_i + \epsilon_{ij}; \quad i = 1, ..., n, \ j = 1, 2, 3,$$

   where x_i's are known and distinct real numbers, and ϵ_{ij}'s have the same distribution as in Model 1. We denote the sum of squares of errors based on Model 2 by SSE_2.

   a. Between Model 1 and Model 2, one model is a special case of the other. State which one is more general and explain why the other is its special case.

   b. Write SSE_1 and SSE_2 as quadratic forms, $\mathbf{Y}^T\mathbf{A}_1\mathbf{Y}$ and $\mathbf{Y}^T\mathbf{A}_2\mathbf{Y}$, respectively. Explicitly write down $\mathbf{A}_1$ and $\mathbf{A}_2$.

c. Construct an F test to compare Model 1 and Model 2. To get full credits, give the hypotheses, derive the test statistic and its distribution under the null hypothesis, and state the decision rule.

3. Suppose that in an experiment study, you suspect that some observations were contaminated. You want to test them jointly for being outliers. Therefore, you organize the suspected observations as the last q observations from a total of n and adopt the *mean shift outlier model* (MSOM) on these observations given below:

$$\begin{aligned} Y_1 &= \mathbf{x}_1^T\boldsymbol{\beta} + \epsilon_1 \\ &\vdots \\ Y_{n-q} &= \mathbf{x}_{n-q}^T\boldsymbol{\beta} + \epsilon_{n-q} \\ Y_{n-q+1} &= \mathbf{x}_{n-q+1}^T\boldsymbol{\beta} + \delta_1 + \epsilon_{n-q+1} \\ &\vdots \\ Y_n &= \mathbf{x}_n^T\boldsymbol{\beta} + \delta_q + \epsilon_n \end{aligned}$$

This model can be specified in matrix form by

$$\underbrace{\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{I}_q \end{bmatrix}}_{\mathbf{X}} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} + \boldsymbol{\varepsilon}$$

where $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$, $\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\mathbf{I}_n$, $\boldsymbol{\varepsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)^T$, and $\mathbf{I}_a$ represents the $a \times a$ identity matrix.

You can assume that $\mathbf{X}$ is full rank and use the fact below:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} (\mathbf{X}_1^T\mathbf{X}_1)^{-1} & -(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_2^T \\ -\mathbf{X}_2(\mathbf{X}_1^T\mathbf{X}_1)^{-1} & \mathbf{I}_q + \mathbf{X}_2(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_2^T \end{bmatrix}.$$

Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ denote the least squares estimator (LSE) for $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ under this MSOM model. We use $\hat{\boldsymbol{\beta}}_1$ to denote the LSE for $\boldsymbol{\beta}$ when we regress only $\mathbf{Y}_1$ on $\mathbf{X}_1$, that is, ignoring the last q suspected observations.

a. Show that (i) $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_1$ and (ii) $\hat{\boldsymbol{\delta}} = \mathbf{Y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_1$. That is, the LSE for $\boldsymbol{\delta}$ is the difference between the observed values $\mathbf{Y}_2$ and the fitted values for $\mathbf{X}_2$ in the model without the last q suspected observations.

b. Define $\hat{\mathbf{Y}}_2 = \mathbf{X}_2\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\delta}}$. Show that the last q observations are *perfectly* fit by the MSOM, that is, $\hat{\mathbf{Y}}_2 = \mathbf{Y}_2$. What can you say about the relation between the LSE for variance, $\hat{\sigma}^2$ for σ^2 under the MSOM and the LSE $\hat{\sigma}_1^2$ for σ^2 under the model without the last q observations?

c. Find the hat matrix for the MSOM and comment on the leverage for the suspected data points.

d. Assume normal distribution for $\epsilon_1, \ldots, \epsilon_n$. Conduct a joint outlier test by testing $\delta_1 = \cdots = \delta_q = 0$. To get full credits, derive the test statistic and its distribution under the null.