Preliminary Examination: Linear Models

Q1 counts for 40 points and Q2, 60 points.

Answer questions by showing all of your work.

1. We consider the multiple linear regression model (referred to as Model 1)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y} = (y_1, \dots, y_n)' \in \mathbb{R}^n$, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ of rank p, and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$. The parameters of this model are $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\sigma^2 > 0$, withe the true values $\boldsymbol{\beta}_*$ and σ_*^2.

(a) Derive the least squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_*$ and an unbiased estimator $\hat{\sigma}^2$ of σ_*^2 as the function of all y_i's and $\boldsymbol{X}$ for $i = 1, \dots, n$.

(b) What is the distribution of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, respectively?

(c) Using the distributions of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ derived in part (b), explain how to construct a 95% confidence interval for $\beta_{*,k}$ that is the k-th element of $\boldsymbol{\beta}_*$.

(d) For the questions below, we partition the covariate as $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, where $\boldsymbol{X}_1 \in \mathbb{R}^{n\times p_1}$ of rank p_1, and $\boldsymbol{X}_2 \in \mathbb{R}^{n\times p_2}$ with some integers $1 \le p_1 < p$ and $p_2 = p - p_1$. Suppose now we fit the model below (referred to as Model 2) with the same error distribution assumption on $\boldsymbol{\varepsilon}$ of Model 1:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}.$$

We partition the true regression coefficient vector accordingly as $\boldsymbol{\beta}_* = \begin{pmatrix} \boldsymbol{\beta}_{1*} \\ \boldsymbol{\beta}_{2*} \end{pmatrix}$.

Show that regardless of $\boldsymbol{\beta}_{2*}$, the coefficient of determination (R^2) from fitting Model 1 is no smaller than that from fitting Model 2.

(e) Let Model 1 denote the full model. Consider a reduced model derived from Model 1 that includes the intercept term and a subset of $k-1$ predictors, where $1 \le k \le p$. That is, the reduced model contains k total regression coefficients (including the intercept). Mallows' C_p statistic for this reduced model is given by:

$$C_k = \frac{\|\boldsymbol{y} - \hat{\boldsymbol{y}}_k\|_2^2}{\hat{\sigma}^2} - (n - 2k),$$

where $\hat{\boldsymbol{y}}_k$ is the fitted response vector from the reduced model, and $\hat{\sigma}^2$ is the unbiased estimator of the error variance σ_*^2 obtained from the full model (Model 1). Show that if $\boldsymbol{\beta}_{2*} = \mathbf{0}$, then Mallows' C_p criterion will favor the reduced model (Model 2) over the full model (Model 1).

2. (60 points) Consider a one-way model

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

for $i = 1, 2$ and $j = 1, 2, 3$. We write the linear model with

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{\beta}^\top = (\mu,\ \alpha_1,\ \alpha_2)^\top$.

(a) Show that μ in Model (1) is NOT an estimable function.

(b) We now think about introducing a side condition such that the model is written by

$$\begin{pmatrix} \boldsymbol{y} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{T} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \mathbf{0} \end{pmatrix} \tag{2}$$

- Introduce your side condition(s) and express it/them as $\boldsymbol{T}$. Explain that the condition(s) satisfy three requirements that make $\hat{\boldsymbol{\beta}}$ for Model (2) uniquely determined.
- Derive the least squares estimate of $\boldsymbol{\beta}$ in (2).

(c) Now, we assume that one is interested in estimating $\mu + \alpha_1$ from Model (1).

- Show that $\mu + \alpha_1$ is an estimable function.
- Let $\hat{\boldsymbol{\beta}}_{\text{ols}}$ be <u>an</u> ordinary least sqaure estimate of $\boldsymbol{\beta}$. Express $\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}_{\text{ols}}$ as a function of $\boldsymbol{\lambda}$, $\boldsymbol{X}$, and $\boldsymbol{y}$.
- Derive <u>a</u> matrix $(\boldsymbol{X}^\top \boldsymbol{X})^-$. Then, express $\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}_{\text{ols}}$ as a function of y_{ij} for $i = 1, 2$ and $j = 1, 2, 3$.

(d) Now, we add the following random error assumptions to Model (1):

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2).$$

Prove that $\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}_{\text{ols}}$ in Part (c) is the best linear unbiased estimator of $\boldsymbol{\lambda}^\top \boldsymbol{\beta}$ of Model (1).

(e) Derive and express the 100(1-α)% confidence interval of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ as the function of $\boldsymbol{\lambda}$, $\hat{\boldsymbol{\beta}}_{\text{ols}}$, $\boldsymbol{X}$, and MSE (mean squared error). Clearly mention the degrees of freedom and the quantile.