Preliminary Examination: Linear Models

Q1 counts for 30 points; Q2, 35 points; and Q3, 35 points.

Answer questions with showing all of your work.

1. (30 points) Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be $n \times p_1$ and $n \times p_2$ be matrices of predictors whose columns are linearly independent to each other. We consider the linear regression model below:

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon},$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p_1- and p_2- dimensional vectors, respectively, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

   (a) Express the ordinary least square estimator for $\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{pmatrix}$ using $\mathbf{X}_1$, $\mathbf{X}_2$, and $\boldsymbol{y}$.

   (b) Let

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix}\mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{X}_1\boldsymbol{y} \\ \mathbf{X}_2\boldsymbol{y}\end{pmatrix}$$

   be the ordinary least square estimator found above in (a). Find the explicit forms of $\mathbf{G}_{11}$, $\mathbf{G}_{12}$, $\mathbf{G}_{21}$, and $\mathbf{G}_{22}$.

   (c) Based on the results in part (b), show that $\boldsymbol{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1\boldsymbol{y}$ when $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$.

2. (35 points) Consider the following linear model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \ \boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I}_n)$$

where $\mathbf{X}$ is a $n \times p$ full rank matrix ($n > p$), $\boldsymbol{\beta}$ is a $(p+1)$-dimensional vector, and $\sigma^2 > 0$ is the variance of the error term. For a $k \times (p+1)$ matrix $\mathbf{C}$, consider the following hypotheses:

$$\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0} \text{ v.s. } \mathrm{H}_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$$

   (a) Find the F statistics for testing the hypotheses and its distribution under the null hypothesis.

   (b) Show that the t-test for individual coefficient is a special case of the hypothesis test in the previous part.
   Hint: Define the suitable $\mathbf{C}$ that leads to the t-test for an individual coefficient and find the sampling distribution of the test statistic.

(c) An estimator under the null hypothesis is given by

$$\hat{\boldsymbol{\beta}}_C = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\left[\mathbf{C}\mathbf{X}'\mathbf{X}\mathbf{C}'\right]\mathbf{C}\hat{\boldsymbol{\beta}},$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator. Find the expected value and variance of $\hat{\boldsymbol{\beta}}_C$ under the alternative hypothesis.

(d) Under the null hypothesis, between $\hat{\boldsymbol{\beta}}_C$ and $\hat{\boldsymbol{\beta}}$, which estimator should be preferred? Provide justification.

3. (35 points) Consider the one-way ANOVA model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ for $i = 1, 2, 3$ and $n_1 = 2$, $n_2 = 2$, $n_3 = 1$ where $E(\varepsilon_{ij}) = 0$, $V(\varepsilon_{ij}) = \sigma^2$, and ε_{ij} are independent each other for all i and j.

(a) Show that $\mu + \tau_1$, $\mu + \tau_2$ and $\mu + \tau_3$ are estimable functions of $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2, \tau_3)'$.

(b) Show that $\sum_{i=1}^{3} c_i\tau_i$ is estimable if and only if $\sum_{i=1}^{3} c_i = 0$.

(c) Show that $\bar{y}_{i\cdot}$ is the BLUE (best linear unbiased estimator) of $\mu + \tau_i$ for $i = 1, 2, 3$ by using the Gauss-Markov theorem with finding **an** LSE (least square estimator) $\boldsymbol{\beta}^0$ and **a** g-inverse $(\boldsymbol{X}'\boldsymbol{X})^-$.

(d) Find the $100(1-\alpha)\%$ **marginal** confidence interval for $\tau_1 - \tau_2$ by using the equation with a g-inverse $(\boldsymbol{X}'\boldsymbol{X})^-$. Write the degrees of freedom using the number.

(e) We now conduct the multiple hypothesis testing using the Scheffe intervals. Find the simultaneous confidence intervals for contrasts $\tau_1 - \tau_2$ and $\tau_1 - \tau_3$ under the significance level α.