STAT5030 Assignment 4

Due: March 28, 2023

1. Consider the following full rank linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e} \quad \text{where e} \sim \mathrm{N}(0, \sigma^2 \mathrm{I}).$$

Let X be $n \times p$, and let y_i denote the elements of y. Also assume $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{b}}$ where $\hat{\boldsymbol{b}}$ is the OLS estimate of $\boldsymbol{b}$.

   (a) Prove that $\sum_{i=1}^{n} \hat{y}_i(y_i - \hat{y}_i) = 0$.

   (b) Prove that $\sum_{i=1}^{n} Var(\hat{y}_i) = p\sigma^2$.

2. Definition:

   (a) Mean Dispersion Error (MDE): The MDE of an estimator $\hat{\boldsymbol{\beta}}$ is defined as the matrix

$$M(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top.$$

   (b) MDE I criterion: Let $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ be two estimators of $\boldsymbol{\beta}$. Then $\hat{\boldsymbol{\beta}}_2$ is called MDE I superior to $\hat{\boldsymbol{\beta}}_1$ if the difference of their MDE matrices is nonnegative definite, that is , if

$$\Delta(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = M(\hat{\boldsymbol{\beta}}_1, \boldsymbol{\beta}) - M(\hat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}) \geq 0.$$

Theorem: Let $\boldsymbol{a}$ be a vector. Then

$$\boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^\top \geq 0 \quad \text{if and only if } \boldsymbol{a}^\top \boldsymbol{a} \leq 1.$$

Question: Consider the following linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \quad \text{where E(e)} = 0,\ \text{Var(y)} = \sigma^2 \mathrm{I}.$$

Assume $\boldsymbol{X}$ has full column rank and the first column is a column of ones. Let $\hat{\boldsymbol{\beta}}_1 = (1+\rho)^{-1}\hat{\boldsymbol{\beta}}$, $\rho > 0$ (ρ known). Furthermore, $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$. Prove that

$$\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} \leq \sigma^2$$

is a sufficient condition for MDE I superiority of $\hat{\boldsymbol{\beta}}_1$ compared to OLS estimator of β.

3. Let $\boldsymbol{\lambda}^\top \boldsymbol{\beta}$ be an estimable function of $\boldsymbol{\beta}$ in the model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{X}$ is $n \times p$ of rank $k < p \leq n$. Let $\hat{\beta}$ be any solution to the normal equations $\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{Y}$, and let $\boldsymbol{r}$ be any solution to $\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{r} = \boldsymbol{\lambda}$. Then, for the two estimators $\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}}$ and $\boldsymbol{r}^\top \boldsymbol{X}^\top \boldsymbol{Y}$, prove that

   (a) $E(\boldsymbol{\lambda}\hat{\boldsymbol{\beta}}) = E(\boldsymbol{r}^\top \boldsymbol{X}^\top \boldsymbol{Y}) = \boldsymbol{\lambda}^\top \boldsymbol{\beta}$.

(b) $\boldsymbol{r}^\top \boldsymbol{X}^\top \boldsymbol{Y}$ is invariant to the choice of $\boldsymbol{r}$.

4. Consider the model

$$\begin{aligned}
Y_1 &= \mu + \alpha_1 + \beta_1 + \varepsilon_1, \\
Y_2 &= \mu + \alpha_1 + \beta_2 + \varepsilon_2, \\
Y_3 &= \mu + \alpha_1 + \beta_3 + \varepsilon_3, \\
Y_4 &= \mu + \alpha_2 + \beta_1 + \varepsilon_4, \\
Y_5 &= \mu + \alpha_2 + \beta_2 + \varepsilon_5, \\
Y_6 &= \mu + \alpha_2 + \beta_3 + \varepsilon_6, \\
Y_7 &= \mu + \alpha_3 + \beta_1 + \varepsilon_7, \\
Y_8 &= \mu + \alpha_3 + \beta_2 + \varepsilon_8, \\
Y_9 &= \mu + \alpha_3 + \beta_3 + \varepsilon_9.
\end{aligned}$$

where ε_i, $i = 1, \cdots, 9$ are independently distributed as normal $(0, \sigma^2)$.

(a) State the conditions when $\lambda_0\mu + \lambda_1\alpha_1 + \lambda_2\alpha_2 + \lambda_3\alpha_3 + \lambda_4\beta_1 + \lambda_5\beta_2 + \lambda_6\beta_3$ is estimable.

(b) Is $\alpha_1 + \alpha_2$ estimable?

(c) Is $\beta_1 + \beta_2 + \beta_3$ estimable?

(d) Is $\mu + \alpha_2$ estimable?

(e) Is $6\mu + 2\alpha_1 + 2\alpha_2 + 2\alpha_3 + 3\beta_1 + 3\beta_3$ estimable?

(f) Is $\alpha_1 - 2\alpha_2 + \alpha_3$ estimable?

5. Consider the model

$$\begin{aligned}
Y_1 &= \beta_1 + \beta_2 + \varepsilon_1, \\
Y_2 &= \beta_1 + \beta_3 + \varepsilon_2, \\
Y_3 &= \beta_1 + \beta_2 + \varepsilon_3,
\end{aligned}$$

where ε_i, $i = 1, 2, 3$ are independently distributed as normal $(0, \sigma^2)$. Show that $\lambda_1\beta_1 + \lambda_2\beta_2 + \lambda_3\beta_3$ is estimable if and only if $\lambda_1 = \lambda_2 + \lambda_3$.

6. The period of oscillation t of a pendulum is $2\pi\sqrt{l/g}$, where l is the length and g is the gravitational constant. The periods observed are $t_{ij}(j = 1, 2, \cdots, n_i)$ and length $l_i(i = 1, \cdots, k)$ of the pendulum, in an experiment. Assuming the errors of observations to be uncorrelated with zero means and variance σ^2, obtain the best unbiased estimate of $2\pi/\sqrt{g}$ and an estimate of its variance. obtain the best unbiased estimate of $2\pi/\sqrt{g}$ and an estimate of its variance.

7. Let

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \cdots, k; , j = 1, \cdots, n_i;$$

and ε_{ij} are independently distributed as normal $(0, \sigma^2)$.

(a) Is the null hypothesis

$$H_0 : \frac{\mu + \alpha_1}{a_1} = \frac{\mu + \alpha_2}{a_2} = \cdots = \frac{\mu + \alpha_k}{a_k}$$

testable? Prove your claim. ($a_1, \cdots, a_k$ are constants.)

(b) Derive the testing procedure.

8. Suppose that there is a two-sample case:

Treatment data: $y_1, \cdots, y_n$, control data: $y_{n+1}, \cdots, y_{m+n}$. Suppose we model the response by an overall mean μ and group effects α_1 and α_2:

$$y_i = \mu + \alpha_1 + \varepsilon_i, \quad i = 1, \cdots, n;$$
$$y_i = \mu + \alpha_2 + \varepsilon_i, \quad i = n+1, \cdots, n+m.$$

(a) Is $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2)^\top$ identifiable?

(b) When an constraint $\alpha_1 + \alpha_2 = 0$ is imposed on $\boldsymbol{\beta}$, is $\boldsymbol{\beta}$ identifiable?

(c) When an constraint $\mu = 0$ is imposed on $\boldsymbol{\beta}$, is $\boldsymbol{\beta}$ identifiable?

9. Consider a quantile regression model

$$Y_i = X_i \beta_\tau + \varepsilon_{i,\tau},$$

where X_i and $\varepsilon_{i,\tau}$ are correlated.

Suppose $\hat{\beta}$ is the estimator for β by minimization check loss function

$$G(\beta) = \rho_\tau(y - x\beta)$$

and

$$\rho_\tau(u) = \begin{cases} \tau u & u > 0 \\ (\tau - 1)u & u \leq 0 \end{cases}$$

We can prove that $\sqrt{n}(\hat{\beta} - \beta) \sim N(0, A^{-1}B(A^{-1})^\top)$. Find the value of A and B.