

Preliminary Examination: Linear Models

Answer questions by showing all of your work.

1. Consider the linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, i.e.,

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \quad i = 1, \dots, n.$$

Now consider the centered model

$$y_i = \alpha + \beta_1(x_{i1} - \bar{x}_1) + \cdots + \beta_k(x_{ik} - \bar{x}_k) + \epsilon_i,$$

where

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad j = 1, \dots, k.$$

In matrix form, this model can be written as

$$\mathbf{y} = \begin{bmatrix} \mathbf{1} & \mathbf{X}_c \end{bmatrix} \begin{pmatrix} \alpha \\ \boldsymbol{\beta}_c \end{pmatrix} + \boldsymbol{\varepsilon}$$

with the centered form of the design matrix $\mathbf{X}_c$ and the corresponding coefficient $\boldsymbol{\beta}_c$.

Assume that, for each $j = 1, \dots, k$, the values $x_{1j}, \dots, x_{nj}$ are not all identical, and that $\mathbf{X}_c$ has full column rank. Then, the ordinary least squares estimators of α and $\boldsymbol{\beta}_c$ are

$$\hat{\alpha} = \bar{y} \quad \text{and} \quad \hat{\boldsymbol{\beta}}_c = (\mathbf{X}_c^\top \mathbf{X}_c)^{-1} \mathbf{X}_c^\top \mathbf{y}.$$

- (a) Prove the following theorem. (Note: Your proof should include all details.)

Let $\mathbf{y}$ be an n -dimensional random vector with

$$E(\mathbf{y}) = \boldsymbol{\mu}, \quad \text{Var}(\mathbf{y}) = \boldsymbol{\Sigma}.$$

If $\mathbf{A}$ is an $n \times n$ symmetric matrix, prove that

$$E(\mathbf{y}^\top \mathbf{A} \mathbf{y}) = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} + \text{tr}(\mathbf{A} \boldsymbol{\Sigma}).$$

- (b) Using the result in part (a), find an unbiased estimator of σ^2 . (Note: If the result in part (a) is not used, no partial point will be given.)
- (c) Define $\text{SSR}_c = \mathbf{y}^\top \mathbf{P}_c \mathbf{y}$ where $\mathbf{P}_c = \mathbf{X}_c (\mathbf{X}_c^\top \mathbf{X}_c)^{-1} \mathbf{X}_c^\top$. Find the distribution of SSR_c .

- Note: If you use any theorem, clearly state the theorem and **verify that all conditions required for the theorem hold**.
 - Hint: $\mathbf{X}\boldsymbol{\beta} = (\alpha\mathbf{1} + \mathbf{X}_c\boldsymbol{\beta}_c)$.
- (d) Let $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$, and define $\text{SSE} = \hat{\boldsymbol{\varepsilon}}^\top\hat{\boldsymbol{\varepsilon}}$, where $\hat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \mathbf{P})\mathbf{y}$ are the residuals (equivalently $\text{SSE} = \mathbf{y}^\top(\mathbf{I} - \mathbf{P})\mathbf{y}$).
- i. Show that $\mathbf{P}_c = \mathbf{P} - \frac{1}{n}\mathbf{J}$, where $\mathbf{J} = \mathbf{1}\mathbf{1}^\top$ is the matrix of all ones.
 - ii. Prove that SSR_c and SSE are independent.
2. This problem concerns the situation where doubts are cast on the stability assumption of the regression coefficient and the independence assumption of the errors. Consider the following change-point model:

$$Y_i = \mu_i + \epsilon_i, \quad i = 1, \dots, n,$$

where

$$\mu_i = \begin{cases} \beta_1, & \text{if } i \leq n/2, \\ \beta_2, & \text{otherwise.} \end{cases}$$

Suppose that we only observe $Y_1, \dots, Y_n$. For simplicity, assume that the sample size is even, namely $n = 2m$ for some integer $m > 0$.

- (a) Find the ordinary least squares estimator $(\hat{\beta}_1, \hat{\beta}_2)$ for (β_1, β_2) .

Assume now that the errors satisfy

$$\epsilon_i = e_i - a e_{i-1}, \quad i = 1, \dots, n,$$

where $a \in \mathbb{R}$ is a parameter controlling the dependence strength, and $e_0, e_1, \dots, e_n$ are independent normal random variables with mean zero and variance $\sigma^2 > 0$.

- (b) Assume that $a = 0$. Find the joint distribution of $(\hat{\beta}_1, \hat{\beta}_2)$. Are $\hat{\beta}_1$ and $\hat{\beta}_2$ independent in this case? How does the variance of $\hat{\beta}_1$ change when $n \rightarrow \infty$ (for example, whether it decreases to zero linearly in n , quadratically, or at some other rate)?
- (c) Assume that $a = 0$. Find an unbiased estimator $\hat{\sigma}^2$ of σ^2 , and derive a test for

$$H_0 : \beta_1 - \beta_2 = 0 \quad \text{vs.} \quad H_a : \beta_1 - \beta_2 \neq 0.$$

You need to specify the test statistic and its distribution under the null hypothesis.

- (d) Now assume that $a = 1$. Find the joint distribution of $(\hat{\beta}_1, \hat{\beta}_2)$. Are $\hat{\beta}_1$ and $\hat{\beta}_2$ independent in this case? How does the variance of $\hat{\beta}_1$ change when $n \rightarrow \infty$ in this case (for example, whether it decreases to zero linearly in n , quadratically, or at some other rate)? Compare your result with the one in part (b) and comment on the effect of dependence among the errors. Is dependence always a “bad” thing?
- (e) Now suppose that $0 < a < 1$. Find the distribution of $\hat{\beta}_1$. How does the variance of $\hat{\beta}_1$ change when $n \rightarrow \infty$ (for example, whether it decreases to zero linearly in n , quadratically, or at some other rate)? Compare your result with the ones in parts (b) and (d) and comment.