Part I: Linear Models

Chapter I.4

Linear Models – BLUEs & MLEs

Topics

To be discussed...

- Theorem of Gauß-Markov & BLUEs
- Maximum Likelihood estimation under normal assumptions

I.4.26 Definition

Let $\mathbf{Y} \sim P_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \Theta \subseteq \mathbb{R}^p$, be a statistical model and let $g : \Theta \to \mathbb{R}^k$ be a known function. Then,

- An estimator $\widehat{\boldsymbol{\xi}} = \widehat{\boldsymbol{\xi}}(\mathbf{Y})$ is called an **unbiased estimator** of $g(\boldsymbol{\beta})$ if $E\widehat{\boldsymbol{\xi}}(\mathbf{Y}) = g(\boldsymbol{\beta})$ for any $\boldsymbol{\beta} \in \Theta$.
- An estimator $\widehat{\boldsymbol{\xi}}(\mathbf{Y}) = A\mathbf{Y}$ with a known matrix $A \in \mathbb{R}^{k \times n}$ is called a **linear estimator**.

I.4.27 Remark

- Given a LM with design matrix B, the LSE $\widehat{\boldsymbol{\beta}}^+ = (B'B)^+ B'\mathbf{Y}$ is a linear estimator.
- Given a LM with design matrix B, a linear unbiased estimator $\widehat{\boldsymbol{\xi}}(\mathbf{Y}) = A\mathbf{Y}$ of $\boldsymbol{\beta}$ with $A \in \mathbb{R}^{p \times n}$ has to satisfy the condition:
$$E\widehat{\boldsymbol{\xi}}(\mathbf{Y}) = E(A\mathbf{Y}) = E(AB\boldsymbol{\beta} + A\boldsymbol{\varepsilon}) = AB\boldsymbol{\beta} = \boldsymbol{\beta} \quad \forall\, \boldsymbol{\beta} \in \mathbb{R}^p.$$
Hence, $AB = I_p$ must hold.
- Further, it should be noticed that the assumption of an existing linear unbiased estimator ensures both identifiability and uniqueness of the LSE.

  Suppose that a linear unbiased estimator exists. Then, $AB = I_p$ so that $p = \text{rank}(AB) \leqslant \text{rank}(B)$. Therefore, B has full rank and $B'B$ is a regular matrix. According to Corollary I.4.12, the LSE is unique. Furthermore, it is unbiased due to Theorem I.4.25.

I.4.28 Theorem (Gauß-Markov)

Consider a LM with design matrix B. Let $\widehat{\boldsymbol{\xi}} = \widehat{\boldsymbol{\xi}}(\mathbf{Y}) = A\mathbf{Y}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ with $A \in \mathbb{R}^{p \times n}$. Then,

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\xi}}\right) - \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}\right) \geqslant 0\,.$$

In particular, this yields $\mathrm{Var}\left(\mathbf{c}'\widehat{\boldsymbol{\beta}}\right) \leqslant \mathrm{Var}\left(\mathbf{c}'\widehat{\boldsymbol{\xi}}\right)$ for any $\mathbf{c} \in \mathbb{R}^p$ and, thus,

$$\mathrm{Var}\left(\widehat{\beta}_j\right) \leqslant \mathrm{Var}\left(\widehat{\xi}_j\right), \quad j = 1, \ldots, p.$$

BLUEs and MLEs

I.4.29 Definition

In the situation of Theorem I.4.28, the LSE $\widehat{\boldsymbol{\beta}}$ is called **best linear unbiased estimator (BLUE)** of $\boldsymbol{\beta}$. For $\mathbf{c} \in \mathbb{R}^p$, $\mathbf{c}'\widehat{\boldsymbol{\beta}}$ is called BLUE of $\mathbf{c}'\beta$.

I.4.30 Remark

Notice that the derivation of LSEs and BLUEs as well as their means and variances do not depend on the particular distributional assumption. However, their distributions depend on the distribution of the error term $\boldsymbol{\varepsilon}$.

I.4.31 Theorem

Given a NoLM with a regular matrix $B'B$ and unknown variance parameter $\sigma^2 > 0$, $\widehat{\boldsymbol{\beta}}$ is also the **Maximum-Likelihood-Estimator (MLE)** of $\boldsymbol{\beta}$.

With $P = I_n - B(B'B)^{-1}B'$, the MLE of σ^2 is given by $\widetilde{\sigma^2} = \frac{1}{n}\mathbf{Y}'P\mathbf{Y}$.

I.4.32 Theorem

Let $n \geqslant p$. Given a NoLM as in Definition I.4.3 and $\operatorname{rank}(B) = r \leqslant p$, we have:

1. $\widehat{\boldsymbol{\beta}}^+$ and $\mathbf{Y} - B\widehat{\boldsymbol{\beta}}^+$ are stochastically independent.
2. $\|\mathbf{Y} - B\widehat{\boldsymbol{\beta}}^+\|^2/\sigma^2 \sim \chi^2(n-r)$
3. $\|B(\widehat{\boldsymbol{\beta}}^+ - \boldsymbol{\beta})\|^2/\sigma^2 \sim \chi^2(r)$
4. If $r = p$ then $\dfrac{1}{\sqrt{\mathbf{c}'(B'B)^{-1}\mathbf{c}}} \cdot \dfrac{\mathbf{c}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{\|\mathbf{Y} - B\widehat{\boldsymbol{\beta}}\|^2/(n-p)}} \sim t(n-p)$ for any $\mathbf{c} \neq \mathbf{0}$.

I.4.33 Remark

The condition $\operatorname{rank}(B) = p$ in Theorem I.4.32 4 can be replaced by the condition $\mathbf{c} \notin \operatorname{Ker}((B'B)^+)$.

Part I: Linear Models

Chapter I.4

Linear Models –Testing

Topics

To be discussed...

- Parametrization of LMs
- Full & reduced LM
- Variance decomposition
- Testing statistical hypothesis
- F-test

I.4.34 Remark

- Consider a LM $\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ as in Definition I.4.2 with $\text{rank}(B) = p \leqslant n$ and orthogonal projection $Q = B(B'B)^{-1}B'$. Then, $\text{Im}(Q) = \text{Im}(B)$ is called the *estimation space* or *model space*.
- Two LMs that have the same model space are called *equivalent* since they lead to the same $E\mathbf{Y}$. The LSE of $E\mathbf{Y}$ under both models are identical:

  Let $\mathbf{Y} = B_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$ and $\mathbf{Y} = B_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$ be two LMs with $\text{Im}(B_1) = \text{Im}(B_2)$. Then, Q is identical for both B_1 and B_2 since it depends only on $\text{Im}(B_1) = \text{Im}(B_2)$. Thus, the (unique) LSE of $E\mathbf{Y}$ is given by (see Remark I.4.11)
  $$B_1\hat{\boldsymbol{\beta}}_1 = B_2\hat{\boldsymbol{\beta}}_2 = Q\mathbf{Y}.$$
- In this sense, one may write for the LM
  $$E\mathbf{Y} \in \text{Im}(B), \quad \text{Cov}\,(\mathbf{Y}) = \sigma^2 I_n,$$
  so that it is independent of the particular chosen parametrization.

Testing in LM

I.4.35 Example

Consider a simple linear regression with $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ and

- $Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j$, $j = 1, \ldots, n$. Then, $\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and design matrix $B_1 = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = [\mathbb{1}_n \mid \boldsymbol{x}] \in \mathbb{R}^{n\times 2}$.
- $Y_j = \gamma_0 + \gamma_1(x_j - \overline{x}) + \varepsilon_j$, $j = 1, \ldots, n$. Then, $\mathbf{Y} = B_2\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ with parameter $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$ and design matrix $B_2 = \begin{pmatrix} 1 & x_1 - \overline{x} \\ 1 & x_2 - \overline{x} \\ \vdots & \vdots \\ 1 & x_n - \overline{x} \end{pmatrix} = [\mathbb{1}_n \mid \boldsymbol{x} - \overline{x}\mathbb{1}_n] \in \mathbb{R}^{n\times 2}$.
- Then, $\mathrm{Im}(B_1) = \mathrm{Im}(B_2)$. Furthermore, provided $\mathrm{rank}(B_1) = 2$, $\gamma_0 = \beta_0 + \beta_1\overline{x}$, $\gamma_1 = \beta_1$ and the respective LSEs are
$$\widehat{\gamma}_1 = \widehat{\beta}_1, \quad \widehat{\gamma}_0 = \widehat{\beta}_0 + \widehat{\beta}_1\overline{x}.$$

I.4.36 Remark

Testing in LMs means that a restriction is put on the model space, that is, we consider a matrix B_0 with

$$\mathrm{Im}(B_0) \subseteq \mathrm{Im}(B).$$ B0 contains subset of column spaces of B

Thus, we have

- the **full NoLM**:

$$Y = B\beta + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n). \tag{I.3}$$

- a **reduced NoLM**:

$$Y = B_0\gamma + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n), \quad \mathrm{Im}(B_0) \subseteq \mathrm{Im}(B). \tag{I.4}$$

- Then, Model I.4 implies Model I.3. Hence, one has to ask whether Model I.4 is true.

From Corollary I.4.10, we get immediately the following identity.

I.4.37 Corollary (variance decomposition formula)

Let $\widehat{\beta}$, and $\widehat{\gamma}$ be LSEs in a LM with design matrix B and B_0 such that $\mathrm{Im}(B_0) \subseteq \mathrm{Im}(B)$. Then,

$$\psi(\widehat{\gamma}) = \psi(\widehat{\beta}) + \|B(\widehat{\gamma} - \widehat{\beta})\|^2$$

I.4.38 Example

Consider a regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad 1 \leqslant i \leqslant n.$$

We would like to test whether variable x_2 contributes significantly to the regression model. The reduced model is

$$Y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i, \quad 1 \leqslant i \leqslant n.$$

Then,

- the design matrices are given by

$$B = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \\ 1 & x_{n1} & x_{n2} \end{pmatrix} = [\mathbb{1}_n \mid \boldsymbol{x}_1 \mid \boldsymbol{x}_2] = [B_0 \mid \boldsymbol{x}_2] \in \mathbb{R}^{n\times 3}, \qquad B_0 = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{pmatrix} = [\mathbb{1}_n \mid \boldsymbol{x}_1] \in \mathbb{R}^{n\times 2}$$

- Clearly $\mathrm{Im}(B_0) \subseteq \mathrm{Im}(B)$.
- Consider LSE under both models with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)', \boldsymbol{\gamma} = (\beta_0, \beta_1)'$:

$$\widehat{\boldsymbol{\beta}} = (B'B)^{+}B'\mathbf{Y}, \quad \widehat{\boldsymbol{\gamma}} = (B_0'B_0)^{+}B_0'\mathbf{Y}.$$

Remark

A sequence of LMs $\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_k, \mathcal{M}$ with corresponding model spaces

$$\mathrm{Im}(B_0) \subseteq \mathrm{Im}(B_1) \ldots \subseteq \mathrm{Im}(B_k) \subseteq \mathrm{Im}(B)$$

are called *nested models*.
In particular $\mathcal{M}_0$ is nested in $\mathcal{M}_1$, $\mathcal{M}_1$ in $\mathcal{M}_2$, ..., $\mathcal{M}_k$ in $\mathcal{M}$.

Nested models play an important role in model selection procedures.

I.4.39 Theorem

Consider a full NoLM as in (I.3), that is,

$$\mathbf{Y} = B\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n)$$

and a reduced NoLM as in (I.4), that is,

$$\mathbf{Y} = B_0\boldsymbol{\gamma} + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n), \quad \mathrm{Im}(B_0) \subsetneq \mathrm{Im}(B).$$

Furthermore, let $Q = B(B'B)^{-1}B'$ and $Q_0 = B_0(B_0'B_0)^{-1}B_0'$ be the orthogonal projections on the images of B and B_0, respectively. Furthermore, let $r = \mathrm{rank}(Q)$ and $\mathrm{rank}(Q - Q_0) = r - r_0$.

Then:

1. If the full Modell (I.3) is true then

F-distribution with (m1-1, m2-1) degrees of freedom if the samples of sizes m1 and m2 are drawn from normal populations with equal variances.

P = I_{n} - Q

$$\frac{\mathbf{Y}'(Q - Q_0)\mathbf{Y}/(r - r_0)}{\mathbf{Y}'(I_n - Q)\mathbf{Y}/(n - r)} \sim F(r - r_0, n - r, \boldsymbol{\beta}'B'(Q - Q_0)B\boldsymbol{\beta}/(2\sigma^2)).$$

non-central parameter

2. If the reduced Modell (I.4) is true then ß2 = 0

$$\frac{\mathbf{Y}'(Q - Q_0)\mathbf{Y}/(r - r_0)}{\mathbf{Y}'(I_n - Q)\mathbf{Y}/(n - r)} \sim F(r - r_0, n - r, 0) = F(r - r_0, n - r).$$

Testing

From Theorem I.4.39, we construct the following testing procedure.

I.4.40 Testing procedure (F-test)

In the situation of Theorem I.4.39, consider the hypothesis

$$H_0 : E\mathbf{Y} = B_0\boldsymbol{\gamma} \text{ for some } \boldsymbol{\gamma}.$$

Then, an α-level statistical test for H_0 is given by the decision rule

$$\text{Reject } H_0 \text{ if } \frac{\mathbf{Y}'(Q - Q_0)\mathbf{Y}/(r - r_0)}{\mathbf{Y}'(I_n - Q)\mathbf{Y}/(n - r)} > F_{1-\alpha}(r - r_0, n - r)$$

where $F_{1-\alpha}(r - r_0, n - r)$ denotes the $(1 - \alpha)$-quantile of the $F(r - r_0, n - r)$-distribution.