

Modelo $\underline{Y} = X\underline{\beta} + \underline{\epsilon} \quad \underline{\epsilon} \sim N(\underline{0}, \sigma^2 I)$

X - matriz de planejamento $n \times p$ $r(X) = k < p < n$

Estimações Pontual

Máxima Verossimilhança e Mínimos Quadrados

$$m.v. \begin{cases} X'X \hat{\beta}_{m.v.} = X'Y \\ \hat{\sigma}_{m.v.}^2 = \frac{1}{n} (Y'Y - \hat{\beta}_{m.v.}' X'Y) \end{cases}$$

Mínimos Quadrados

Equações Normais: $X'X\hat{\beta} = X'Y \quad *$

$r(X) = k < p \Rightarrow r(X'X) = k < p$ $X'X$ é não inversível

$\hat{\beta}^* = GX'Y$ onde G é uma I G de $X'X$ é uma solução.

O sistema $*$ é consistente porque $X'Y \in$ espaço linear gerado pelas colunas de $X'X$.

Solução Geral: $\hat{\beta}^* = GX'Y + (GX'X - I)\underline{z} \quad \underline{z} \in \mathbb{R}^p$

$$X'X \hat{\beta}^* = X'XGX'\underline{y} + (X'XGX'X - X'X)\underline{z} =$$

$$G \text{ IG de } X'X \Rightarrow X'XGX'X = X'X \quad \text{def. IG}$$

$$\begin{cases} XGX'X = X \\ XGX' \text{ é invariante} \\ G \text{ é IG} \end{cases}$$

$$= \underbrace{X'XGX'}_{X'}\underline{y} + (X'X - X'X)\underline{z} = X'\underline{y}$$

$$\hat{\sigma}_{mv}^2 = \frac{1}{n} (\underline{y}'\underline{y} - \hat{\beta}_{mv}' X'\underline{y}) \quad \text{é invariante em } G$$

$$\text{pois } \hat{\beta}_{mv}' X'\underline{y} = \underline{y}' \underbrace{XGX'}_{XGX'} \underline{y} \quad \text{e } XGX' \text{ é invariante.}$$

Def: Funções Estimáveis

No modelo $\underline{y} = X\beta + \underline{\varepsilon}$, $\underline{\varepsilon} \sim N(0, \sigma^2 I)$, $r(X) = K < p < n$
 $(X_{n \times p})$ uma função linear de β , $\underline{l}'\beta$ é uma função
 estimável se existir um estimador não viciado de $\underline{l}'\beta$
 que é linear em $\underline{y}$.

$$\Leftrightarrow \exists \underline{t}' \mid E(\underline{t}'\underline{y}) = \underline{l}'\beta.$$

Teorema 1

No modelo $Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$, X de posto incompleto, seja $L'\beta$ uma função linear de β .

Nestas condições, $L'\beta$ é uma função estimável se e somente se existir um vetor $\underline{t}$, $n \times 1$, tal que $L' = \underline{t}'X$.

Prova:

$$L'\beta \text{ é estimável} \iff E(\underline{t}'Y) = L'\beta \quad \forall \beta$$

$$\iff \underline{t}'X\beta = L'\beta \quad \forall \beta \iff L' = \underline{t}'X.$$

Exemplo

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad i=1,2,3 \quad j=1,2,\dots,n_i$$

Análise de Variância com um fator

n_i - número de observações do i -ésimo nível do fator

$$n_1 = 3 \quad n_2 = 2 \quad n_3 = 1$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{31} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}_{6 \times 4} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \varepsilon$$

$$r(X) = 3 = k < p = 4$$

a) $E(Y_{11})$ é estimável

$E(Y_{11}) = \mu + \alpha_1$ é da forma $\underline{\ell}'\beta$ para

$$\underline{\ell}' = [1 \ 1 \ 0 \ 0]$$

$E(Y_{11})$ é estimável pois tomando $\underline{\tau}' = [1 \ 0 \ 0 \ 0 \ 0 \ 0]$

$$E(\underline{\tau}'\underline{Y}) = E\left\{ [1 \ 0 \ 0 \ 0 \ 0 \ 0] \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{31} \end{bmatrix} \right\} = E(Y_{11})$$

Verificamos aqui, pela definição, que $E(Y_{11})$ é estimável
($\exists \underline{\tau}'\underline{Y} \mid E(\underline{\tau}'\underline{Y}) = \text{parâmetro que desejamos estimar}$)

Realmente $\underline{\ell}' = \underline{\tau}'X$ para

$$\underline{\tau}' = [1 \ 0 \ 0 \ 0 \ 0 \ 0] \quad (\text{verificar})$$

b) $E(Y_{11}) - E(Y_{21}) = \alpha_1 - \alpha_2$ é estimável

Tomar $\underline{\tau}' = [1 \ 0 \ 0 \ -1 \ 0 \ 0]$, $\underline{\tau}'\underline{Y} = Y_{11} - Y_{21}$

$E(\underline{\tau}'\underline{Y}) = E(Y_{11} - Y_{21}) = \alpha_1 - \alpha_2 \therefore \alpha_1 - \alpha_2$ é estimável
pela definição

c) μ não é estimável

$$\underline{l}'\beta \text{ é estimável} \Leftrightarrow \underline{l}' = \underline{x}'X$$

$$\mu = \underline{l}'\beta \text{ para } \underline{l}' = [1 \ 0 \ 0 \ 0]$$

$$[1 \ 0 \ 0 \ 0] = [x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6] \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 1 \\ x_1 + x_2 + x_3 = 0 \\ x_4 + x_5 = 0 \\ x_6 = 0 \end{cases} \quad \text{impossível}$$

$$\nexists \underline{x}' \mid \underline{l}' = \underline{x}'X$$

Definição

$\underline{l}_1' \beta, \underline{l}_2' \beta, \dots, \underline{l}_m' \beta$ é um conjunto de m funções estimáveis linearmente independentes se cada uma delas é estimável e os vetores $p \times 1$, $\underline{l}_1, \underline{l}_2, \dots, \underline{l}_m$ são linearmente independentes. (Nestas condições, o posto de $L = [\underline{l}_1 \ \underline{l}_2 \ \dots \ \underline{l}_m]$ é m).

Teorema 2

No modelo $Y = X\beta + \epsilon$, $X_{n \times p}$, $r(X) = k < p$, o número de funções estimáveis linearmente independentes é $r(X) = k$.

Obs: No modelo com matriz de planejamento de posto completo, qualquer função $\underline{l}'\beta$ é estimável. ($\therefore$ este conceito não tem importância no modelo com matriz de planejamento de posto completo).

Teorema 3

No modelo $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ $X_{n \times p}$, $r(X) = K < p < n$

$$\underline{\varepsilon} \sim N(0, \sigma^2 I)$$

i) Se $\underline{l}'\underline{\beta}$ é estimável, então $\underline{l}'\underline{\beta}^*$ é invariante para qualquer solução $\underline{\beta}^*$ das equações normais

$$\underline{l}'\underline{\beta} \text{ estimável} \Rightarrow \underline{l}' = \underline{t}'X \Rightarrow \underline{l}'\underline{\beta}^* = \underline{t}'XGX'\underline{y}$$

invariante em G

ii) Se $\underline{l}'\underline{\beta}$ é estimável então $\underline{l}'\underline{\beta}^*$ é NVUVM para $\underline{l}'\underline{\beta}$, onde $\underline{\beta}^*$ é qualquer solução das equações normais.

$$\text{iii) } \hat{\sigma}^2 = \frac{\underline{y}'\underline{y} - \underline{\beta}^{*'}X'\underline{y}}{n-K} \text{ é o estimador NVUVM de}$$

σ^2 (é invariante).

iv) Cada linha de $X\underline{\beta} = E(y_{ij})$ é uma função estimável.

Resultados Adicionais

$$a) \hat{\beta}^* = G'X'Y \sim N_p(G'X'X\beta, G'X'XG'\sigma^2)$$

↳ pode ser singular

b) $X\beta$ (n dimensional) é estimável e

$\therefore X\hat{\beta}^*$ é invariante,

$XG'X' = XGX'$ invariante

$$X\hat{\beta}^* \sim N_n(\underbrace{XGX'X}_{\downarrow} \beta, \underbrace{XGX'XG'X'}_X \sigma^2) \sim N_n(X\beta, XGX'\sigma^2)$$

invariante em $G=X$

Exemplo

Consideremos o peso de 6 plantas de três diferentes tipos

Normal	atípica	aberrante
101	84	32
105	88	
94		

Y_{ij} - peso da j -ésima planta do i -ésimo tipo

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i=1,2,3 \quad j=1,2,\dots,n_i$$

$$\begin{bmatrix} 101 \\ 105 \\ 94 \\ 84 \\ 88 \\ 32 \end{bmatrix} = \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \end{bmatrix}$$

$$X'X = \begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$X'y = \begin{bmatrix} \gamma_{11} + \gamma_{12} + \gamma_{13} + \gamma_{21} + \gamma_{22} + \gamma_{31} \\ \gamma_{11} + \gamma_{12} + \gamma_{13} \\ \gamma_{21} + \gamma_{22} \\ \gamma_{31} \end{bmatrix} = \begin{bmatrix} \gamma_{..} \\ \gamma_{1.} \\ \gamma_{2.} \\ \gamma_{3.} \end{bmatrix}$$

$$= \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix}$$

Equações Normais

$$\begin{bmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_0 \\ \alpha_1^0 \\ \alpha_2^0 \\ \alpha_3^0 \end{bmatrix} = \begin{bmatrix} 504 \\ 300 \\ 172 \\ 32 \end{bmatrix}$$

Obter uma submatriz de dimensão $k=3$ inversível.
Inverter, transpor, substituir
(e transpor o resultado,
→ Preencher as demais posições com 0.

Verifica-se que $G = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ é uma IG de $X'X$.

$$\hat{\beta}^* = GX'Y = [0 \ 100 \ 86 \ 32]'$$

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\beta}^{*'}X'Y}{n-k} \quad k = r(X) = 3$$

$$Y'Y = 45886 \quad \hat{\beta}^{*'}X'Y = 45816 \quad \hat{\sigma}^2 = \frac{70}{3}$$

$\alpha_1 - \alpha_2$ é estimável

$$\alpha_1 - \alpha_2 = [0 \ 1 \ -1 \ 0] \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

$$\alpha_1 - \alpha_2 = [0 \ 1 \ -1 \ 0] \hat{\beta}^* \text{ é invariante com } \hat{\beta}^*$$

$$\therefore \alpha_1 - \alpha_2 = [0 \ 1 \ -1 \ 0] \begin{bmatrix} 0 \\ 100 \\ 86 \\ 32 \end{bmatrix} = 14$$

Verifica-se que α_1 e α_2 não são estimáveis.

$E(Y_{11}) = E(Y_{12}) = E(Y_{13}) = \mu + \alpha_1$ é estimável

(1ª, 2ª e 3ª linhas de $X(\beta)$)

$$\hat{\mu} + \alpha_1 = [1 \ 1 \ 0 \ 0] \hat{\beta}^* = [1 \ 1 \ 0 \ 0] \begin{bmatrix} 0 \\ 100 \\ 86 \\ 32 \end{bmatrix} = 100$$

$$(\mu + \alpha_1 = \underline{\ell}' \underline{\beta} \text{ p/ } \underline{\ell}' = [1 \ 1 \ 0 \ 0])$$

Observar que $\hat{\mu} + \alpha_1 = \bar{y}_1 = \frac{y_{1.}}{3} = \frac{300}{3}$

média amostral do grupo normal

Cálculo de $\text{Var}(\alpha_1 - \alpha_2)$

$$\text{Var}(\hat{\beta}^*) = \text{Var}(GX'Y) = GX'XG'\sigma^2$$

$$\text{Var}(\underline{q}' \hat{\beta}^*) = \underline{q}' \text{Var}(\hat{\beta}^*) \underline{q} = \underline{q}' GX'XG' \underline{q} \sigma^2$$

Para $\underline{q}' \underline{\beta}$ estimável, $\underline{q}' = \underline{t}' X \Rightarrow$

$$\text{Var}(\underline{q}' \hat{\beta}^*) = \underline{q}' \underbrace{GX'XG'}_{X'} \underline{t} \sigma^2 = \underline{q}' GX' \underline{t} \sigma^2 =$$

$$= \underline{q}' G \underline{q} \sigma^2$$

$$\text{Var}(\underline{q}' \hat{\beta}^*) = \underline{t}' XGX' \underline{t} \sigma^2 \text{ constante } \forall G \text{ pois } XGX' \text{ é invariante.}$$

$$\alpha_1 - \alpha_2 = q' \beta \quad \text{para} \quad q' = [0 \quad 1 \quad -1 \quad 0]$$

$$\therefore \text{Var}(\hat{\alpha}_1 - \hat{\alpha}_2) = [0 \quad 1 \quad -1 \quad 0] G \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \cdot \hat{\sigma}^2 = \frac{5}{6} \hat{\sigma}^2$$

$$\widehat{\text{Var}}(\hat{\alpha}_1 - \hat{\alpha}_2) = \frac{5}{6} \hat{\sigma}^2 = \frac{5}{6} \cdot \frac{70}{3}$$

Pelo Teorema 3, $\hat{\alpha}_1 - \hat{\alpha}_2$ e $\hat{\mu} + \hat{\alpha}_1$ são NVUM para $E(Y_{11}) - E(Y_{21})$ e $E(Y_{11})$.

$K=3 \Rightarrow$ obteríamos no máximo três funções estimáveis linearmente independentes.

$r(X)=3 \Rightarrow X$ tem três linhas que formam vetores LI, $\tilde{l}'_1, \tilde{l}'_2, \tilde{l}'_3$

Como cada linha de $X\beta$ é estimável, $\tilde{l}'_1\beta, \tilde{l}'_2\beta$ e $\tilde{l}'_3\beta$ são três funções estimáveis linearmente independentes.

No exemplo,

$$\underline{l}'_1 = [1 \ 1 \ 0 \ 0] \quad \underline{l}'_2 = [1 \ 0 \ 1 \ 0] \quad \text{e} \quad \underline{l}'_3 = [1 \ 0 \ 0 \ 1]$$

são LI pois

$$a_1 \underline{l}'_1 + a_2 \underline{l}'_2 + a_3 \underline{l}'_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow a_1 = a_2 = a_3 = 0.$$

Portanto $\underline{l}'_1 \beta$, $\underline{l}'_2 \beta$ e $\underline{l}'_3 \beta$ são funções estimáveis linearmente independentes.

$$\underline{l}'_1 \beta = \mu + \alpha_1$$

$$\underline{l}'_2 \beta = \mu + \alpha_2$$

$$\underline{l}'_3 \beta = \mu + \alpha_3$$

Dizemos que $\underline{l}'_1 \beta$, $\underline{l}'_2 \beta$ e $\underline{l}'_3 \beta$ é um conjunto base de funções estimáveis.

É fácil verificar que combinações lineares de funções estimáveis são estimáveis.

Todas as funções estimáveis nesse modelo serão combinações lineares de $\underline{l}'_1 \beta$, $\underline{l}'_2 \beta$ e $\underline{l}'_3 \beta$.