

4.5 - Análise de Variância com dois fatores, uma observação por célula sem interação

$$Y_{ij} = \mu + \alpha_i + \tau_j + \epsilon_{ij}$$

$$i = 1, 2, \dots, a > 1 \quad j = 1, 2, \dots, b > 1$$

$$n = ab$$

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

$$X = \begin{bmatrix} 1 & \overbrace{1 \ 0 \ \dots \ 0}^a & \overbrace{1 \ 0 \ \dots \ 0}^b \\ 1 & 1 \ 0 \ \dots \ 0 & 0 \ 1 \ \dots \ 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 \ 0 \ \dots \ 0 & 0 \ 0 \ \dots \ 1 \\ 1 & 0 \ 1 \ \dots \ 0 & 1 \ 0 \ \dots \ 0 \\ 1 & 0 \ 1 \ \dots \ 0 & 0 \ 1 \ \dots \ 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 \ 1 \ \dots \ 0 & 0 \ 0 \ \dots \ 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 \ 0 \ \dots \ 1 & 1 \ 0 \ \dots \ 0 \\ 1 & 0 \ 0 \ \dots \ 1 & 0 \ 1 \ \dots \ 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 \ 0 \ \dots \ 1 & 0 \ 0 \ \dots \ 1 \end{bmatrix} \begin{matrix} \left. \begin{matrix} 1 \\ b \text{ linhas} \end{matrix} \right\} \\ \left. \begin{matrix} 2 \\ b \text{ linhas} \end{matrix} \right\} \\ a \end{matrix}$$

$ab \times (a+b+1)$

$$\underline{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1b} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2b} \\ \vdots \\ Y_{a1} \\ Y_{a2} \\ \vdots \\ Y_{ab} \end{bmatrix}_{ab \times 1}$$

A \ B	B ₁	B ₂	...	B _b	
A ₁	Y ₁₁	Y ₁₂		Y _{1b}	Y _{1.}
A ₂	Y ₂₁	Y ₂₂		Y _{2b}	Y _{2.}
...					
A _a	Y _{a1}	Y _{a2}		Y _{ab}	Y _{a.}
	Y _{.1}	Y _{.2}		Y _{.b}	Y _{..}

$$\underline{\beta} = \begin{bmatrix} \mu \\ \alpha \\ \tau \end{bmatrix}$$

$$\underline{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_a \end{bmatrix}$$

$$\underline{\tau} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_b \end{bmatrix}$$

$$X'X = \begin{bmatrix} ab & b & b \dots b & a & a \dots a \\ & b & 0 & 0 & 1 & 1 \dots 1 \\ & & b & \dots 0 & 1 & 1 \dots 1 \\ & & & b & 1 & 1 \dots 1 \\ & & & & a & 0 \dots 0 \\ & & & & & a \dots 0 \\ & & & & & & a \end{bmatrix}$$

$$r(X) = r(X'X) = a+b-1$$

$$x_1 + x_2 + \dots + x_a = y_1 + y_2 + \dots + y_b$$

$$\Rightarrow y_b = x_1 + \dots + x_a - y_1 - \dots - y_{b-1}$$

$$(a+b+1) \times (a+b+1)$$

Equações Normais

$$\begin{cases} ab\hat{\mu} + b \sum \hat{\alpha}_i + a \sum \hat{\tau}_j = Y_{..} \\ b\hat{\mu} + b\hat{\alpha}_i + \sum \hat{\tau}_j = Y_{i.} \quad i=1, 2, \dots, a \\ a\hat{\mu} + \sum \hat{\alpha}_i + a\hat{\tau}_j = Y_{.j} \quad j=1, 2, \dots, b \end{cases}$$

$a+b+1$ parâmetros, $a+b+1$ equações

Uma solução das equações normais é

$$\hat{\mu}^* = \bar{Y}_{..}$$

$$\hat{\alpha}_i^* = \bar{Y}_{i.} - \bar{Y}_{..} \quad i=1, 2, \dots, a$$

$$\hat{\tau}_j^* = \bar{Y}_{.j} - \bar{Y}_{..} \quad j=1, 2, \dots, b$$

$$\bar{Y}_{.j} = \frac{\sum_{i=1}^a Y_{ij}}{a} = \frac{Y_{.j}}{a}$$

$$\bar{Y}_{..} = \frac{\sum_{i=1}^a \sum_{j=1}^b Y_{ij}}{ab} = \frac{Y_{..}}{ab}$$

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^b Y_{ij}}{b} = \frac{Y_{i.}}{b}$$

$$\bar{Y}_{.j} = \sum_{i=1}^a Y_{ij} - Y_{.j}$$

$$r(X) = r(X'X) \leq a + b - 1$$

Podem existir mais dependências lineares.

Para verificar isso Graybill pag 535 sugere analisar o sistema de equações normais da seguinte maneira;

- 1 - Dividir a 1ª linha por a
- 2 - Subtrair a linha α_i da linha α_1 e dividir por b , $i = 2, 3, \dots, a$
- 3 - Subtrair a linha τ_j da linha τ_1 e dividir por a , $j = 2, 3, \dots, b$

Ignorando as linhas α_1 e τ_1 obtém-se

$$\hat{\mu} + \frac{\sum \hat{\alpha}_i}{a} + \frac{\sum \hat{\tau}_j}{b} = \bar{Y}_{..}$$

$$\hat{\alpha}_1 - \hat{\alpha}_i = \bar{Y}_{1.} - \bar{Y}_{i.} \quad i = 2, 3, \dots, a$$

$$\hat{\tau}_1 - \hat{\tau}_j = \bar{Y}_{.1} - \bar{Y}_{.j} \quad j = 2, 3, \dots, b$$

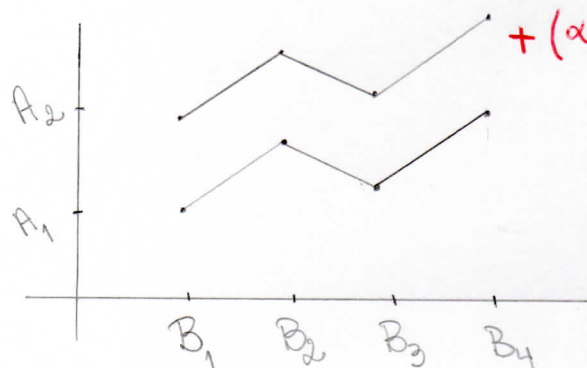
Temos $1 + a - 1 + b - 1 = a + b - 1$ eq. LI porque somente a 1ª equação contém μ e as restantes contêm parâmetros distintos.

Ex: $a=2$ 2 tipos de feijão $b=4$ 4 tipos de fertilizante

$$E(Y_{ij}) = \mu + \alpha_i + \tau_j \quad i=1,2 \quad j=1,2,3,4$$

$$E(Y) = \begin{bmatrix} E(Y_{11}) \\ E(Y_{12}) \\ E(Y_{13}) \\ E(Y_{14}) \\ E(Y_{21}) \\ E(Y_{22}) \\ E(Y_{23}) \\ E(Y_{24}) \end{bmatrix} = \begin{bmatrix} \mu + \alpha_1 + \tau_1 \\ \mu + \alpha_1 + \tau_2 \\ \mu + \alpha_1 + \tau_3 \\ \mu + \alpha_1 + \tau_4 \\ \mu + \alpha_2 + \tau_1 \\ \mu + \alpha_2 + \tau_2 \\ \mu + \alpha_2 + \tau_3 \\ \mu + \alpha_2 + \tau_4 \end{bmatrix}$$

Diagram illustrating the relationship between the two levels of factor A (1 and 2) and the four levels of factor B (1, 2, 3, 4). The difference between the means for level 2 and level 1 of factor A is constant across all levels of factor B, represented by the shaded regions labeled $+\alpha_2 - \alpha_1$.



$+(\alpha_2 - \alpha_1) \forall$ nível do fator B

Seriam 8 médias (8 parâmetros)
mas como é sem interação precisa-se de $4 + 1 = b + a - 1$ parâmetros
O modelo é "superparametrizado" porque usa $a + b + 1 = 7$ parâmetros.

$$\begin{aligned} E(Y_{21}) - E(Y_{11}) &= \alpha_2 - \alpha_1 = E(Y_{22}) - E(Y_{12}) = E(Y_{23}) - E(Y_{13}) = \\ &= E(Y_{24}) - E(Y_{14}) \end{aligned}$$

acrescimento quando se passa do nível 1 para o nível 2 do fator A fixado o nível de B é o mesmo para os quatro níveis de B.

Se o modelo fosse com interações, isso não aconteceria:

$$E(\gamma_{ij}) = \mu + \alpha_i + \tau_j + \gamma_{ij}$$

$$\begin{aligned} E(\gamma_{21}) - E(\gamma_{11}) &= \mu + \alpha_2 + \tau_1 + \gamma_{21} - (\mu + \alpha_1 + \tau_1 + \gamma_{11}) \\ &= \alpha_2 - \alpha_1 + (\gamma_{21} - \gamma_{11}) \end{aligned}$$

$$\begin{aligned} E(\gamma_{22}) - E(\gamma_{12}) &= \mu + \alpha_2 + \tau_2 + \gamma_{22} - (\mu + \alpha_1 + \tau_2 + \gamma_{12}) \\ &= \alpha_2 - \alpha_1 + (\gamma_{22} - \gamma_{12}) \end{aligned}$$

Verifica-se que (Graybill pag 536)

$$\mu + \bar{\alpha} + \bar{\tau}$$

$$\alpha_1 - \alpha_i \quad i = 2, 3, \dots, a$$

$$\tau_1 - \tau_j \quad j = 2, 3, \dots, b$$

$$\text{com } \bar{\alpha} = \frac{\sum_{i=1}^a \alpha_i}{a} \quad \text{e} \quad \bar{\tau} = \frac{\sum_{j=1}^b \tau_j}{b}$$

formam um conjunto base de funções estimáveis no modelo.

Os correspondentes estimadores ótimos destas quantidades são

$$\hat{\mu}^* + \frac{\sum_{i=1}^a \hat{\alpha}_i^*}{a} + \frac{\sum_{j=1}^b \hat{\tau}_j^*}{b} = \bar{Y}_{..}$$

$$\hat{\alpha}_j^* - \hat{\alpha}_i^* = \bar{Y}_{.j} - \bar{Y}_{.i} \quad i = 2, 3, \dots, a$$

$$\hat{\tau}_j^* - \hat{\tau}_i^* = \bar{Y}_{.j} - \bar{Y}_{.i} \quad j = 2, 3, \dots, b$$

$$\hat{\mu}^* = \bar{Y}_{..} \quad \hat{\alpha}_i^* = \bar{Y}_{.i} - \bar{Y}_{..} \quad \hat{\tau}_j^* = \bar{Y}_{.j} - \bar{Y}_{..} \quad \text{solução das}$$

equações normais

$$\bar{Y}_{..} = \frac{\sum_{i=1}^a \sum_{j=1}^b Y_{ij}}{ab} = \frac{Y_{..}}{ab}$$

$$\bar{Y}_{.i} = \frac{\sum_{j=1}^b Y_{ij}}{b} = \frac{Y_{.i}}{b}$$

$$\bar{Y}_{.j} = \frac{\sum_{i=1}^a Y_{ij}}{a} = \frac{Y_{.j}}{a}$$

Teorema

No modelo de análise de variância com dois fatores, sem interação, uma observação por célula, $\sum C_i \alpha_i$ é estimável se e somente se $\sum C_i \alpha_i$ é um contraste nos α_i 's. O estimador NVUVM de $\sum C_i \alpha_i$ é $\sum_{i=1}^a C_i \hat{\alpha}_i^* = \sum_{i=1}^a C_i \bar{Y}_{i.}$. Analogamente,

com relação aos τ_j , o estimador NVUVM do contraste $\sum_{j=1}^b d_j \tau_j$

$$\text{é } \sum_{j=1}^b d_j \hat{\tau}_j^* = \sum_{j=1}^b d_j \bar{Y}_{.j}.$$

Prova

$$\sum C_i \alpha_i \text{ estimável} \Rightarrow \sum_{i=1}^a C_i \alpha_i = d_0 (\mu + \bar{\alpha} + \bar{\tau}) + \sum_{i=1}^a d_i (\alpha_i - \bar{\alpha}) + \sum_{j=1}^b h_j (\tau_j - \bar{\tau})$$

$$d_0 = 0 \quad h_j = 0 \quad j = 1, 2, \dots, b$$

$$= \sum_{i=1}^a d_i \alpha_i - \sum_{i=1}^a d_i \bar{\alpha} = \sum_{i=1}^a d_i \alpha_i - \bar{\alpha} \sum_{i=1}^a d_i =$$

$$= d_1 \left(\sum_{i=1}^a d_i - d_1 \right) + \sum_{i=2}^a d_i \alpha_i = d_1 \sum_{i=2}^a d_i - \sum_{i=2}^a d_i \alpha_i$$

Estimável p/ α

$$C_1 = \sum_{i=2}^a d_i \quad C_2 = -d_2 \quad C_3 = -d_3 \quad \dots \quad C_a = -d_a$$

$$\therefore \sum_{i=1}^a C_i = \sum_{i=2}^a d_i - d_2 - d_3 - \dots - d_a = 0$$

Estimador NVUVM de $\sum C_i \alpha_i$, $\hat{\alpha}_i^* = \bar{Y}_{i.} - \bar{Y}_{..}$ solução dos eq. normais

$$\hat{\alpha}_i^* = \sum C_i (\bar{Y}_{i.} - \bar{Y}_{..}) = \sum C_i \bar{Y}_{i.} - \bar{Y}_{..} \sum C_i = \sum C_i \bar{Y}_{i.}$$

$$\hat{\sigma}^2 = \frac{1}{n-k} \left(\mathbf{y}'\mathbf{y} - \hat{\beta}^{*'}\mathbf{X}'\mathbf{y} \right) \rightarrow \underline{\text{Estimador NVUVM de } \sigma^2}$$

$$n=ab \quad k = a+b-1$$

$$n-k = ab - a - b + 1 = a(b-1) - (b-1) = (a-1)(b-1)$$

$$\mathbf{y}'\mathbf{y} = \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{a.} \\ y_{.1} \\ \vdots \\ y_{.b} \end{bmatrix} \quad \hat{\beta}^* = \begin{bmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{a.} - \bar{y}_{..} \\ \bar{y}_{.1} - \bar{y}_{..} \\ \vdots \\ \bar{y}_{.b} - \bar{y}_{..} \end{bmatrix}$$

$$\hat{\beta}^{*'}\mathbf{X}'\mathbf{y} = \hat{\mu}^* y_{..} + \sum_{i=1}^a \hat{\alpha}_i^* y_{i.} + \sum_{j=1}^b \hat{\tau}_j^* y_{.j}$$

$$= \frac{y_{..}^2}{ab} + \sum_{i=1}^a y_{i.} (\bar{y}_{i.} - \bar{y}_{..}) + \sum_{j=1}^b y_{.j} (\bar{y}_{.j} - \bar{y}_{..})$$

$$= \frac{y_{..}^2}{ab} + \left(\sum_{i=1}^a \frac{y_{i.}^2}{b} - \frac{y_{..}}{ab} \sum_{i=1}^a y_{i.} \right) + \left(\sum_{j=1}^b \frac{y_{.j}^2}{a} - \frac{y_{..}}{ab} \sum_{j=1}^b y_{.j} \right)$$

$$= \frac{y_{..}^2}{ab} + \left(\sum_{i=1}^a \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab} \right) + \left(\sum_{j=1}^b \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab} \right)$$

$$\underline{Y}'\underline{Y} - \hat{\beta}^* X' \underline{Y} = \text{cálculos}$$

$$= \sum_{i=1}^a \sum_{j=1}^b (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$$

$$\therefore \hat{\sigma}^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2}{(a-1)(b-1)}$$

Teorema

No modelo $Y_{ij} = \mu + d_i + \tau_j + \epsilon_{ij}$, $i=1,2,\dots,a$ $j=1,2,\dots,b$

$\epsilon_{ij} \sim N(0, \sigma^2)$ Y_{ij} v.a. independentes

$$a) Z_1 = \hat{\sum c_i d_i} = \sum c_i \bar{Y}_{i.} \sim N\left(\sum c_i \mu, \sigma^2 \frac{\sum c_i^2}{b}\right) \text{ se } \sum c_i = 0$$

$$b) Z_2 = \hat{\sum d_j \tau_j} = \sum d_j \bar{Y}_{.j} \sim N\left(\sum d_j \mu, \sigma^2 \frac{\sum d_j^2}{a}\right) \text{ se } \sum d_j = 0$$

$$c) U = \frac{(a-1)(b-1) \hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(a-1)(b-1)}$$

d) U, Z_1 e Z_2 são v.a. independentes

$$e) \text{Cov}\left[\hat{\sum c_i d_i}, \hat{\sum c_i^* d_i}\right] = \sigma^2 \frac{\sum c_i c_i^*}{b} \text{ se } \sum c_i = \sum c_i^* = 0$$

$$f) \text{Cov}\left[\hat{\sum d_j \tau_j}, \hat{\sum d_j^* \tau_j}\right] = \sigma^2 \frac{\sum d_j d_j^*}{a} \text{ se } \sum d_j = \sum d_j^* = 0$$

Consequência

Os intervalos de confiança com coeficiente $1-\alpha$ para $\sum c_i x_i$ com $\sum c_i = 0$ e $\sum d_j \bar{y}_j$ com $\sum \bar{y}_j = 0$ são respectivamente

$$\sum c_i \bar{y}_{i.} \pm t_{\alpha/2, (a-1)(b-1)} \hat{\sigma} \sqrt{\frac{\sum c_i^2}{b}}$$

$$\sum d_j \bar{y}_{.j} \pm t_{\alpha/2, (a-1)(b-1)} \hat{\sigma} \sqrt{\frac{\sum d_j^2}{a}}$$

Teste de Hipóteses

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a$$

H_a : Pelo menos uma das igualdades acima não se verifica

A estatística do teste da razão de Verossimilhança generalizada é

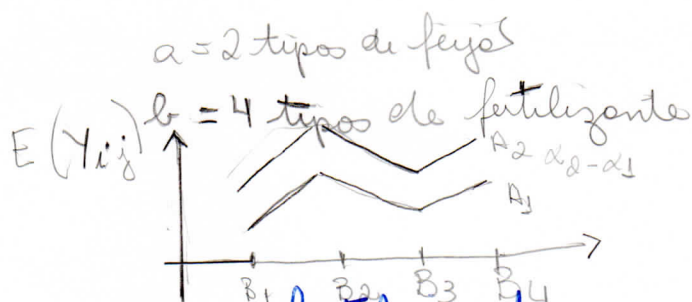
$$W_A = \frac{\frac{\sum_{i=1}^b \sum_{j=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2}{a-1}}{\frac{\sum_{i=1}^b \sum_{j=1}^a (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2}{(a-1)(b-1)}}$$

Rejeitamos H_0 se $W_A \geq w$, $P(F_{a-1, (a-1)(b-1)} \geq w) = \alpha$

No exemplo, se $H_0: \alpha_1 = \alpha_2$ e aceite temos

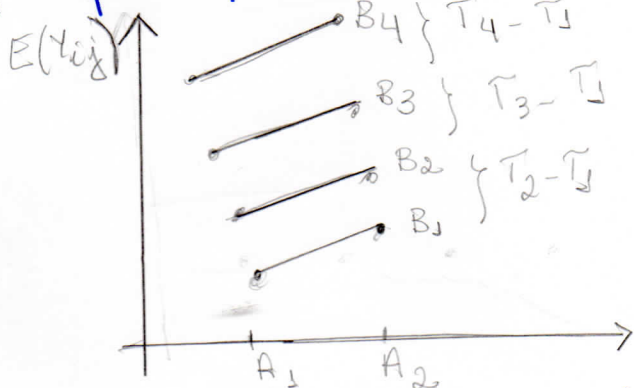
$$Y_{ij} = \mu + \alpha + \tau_j + \epsilon_{ij}$$

$$E(Y_{ij}) = \mu + \alpha + \tau_j$$



(a produção média depende apenas do fertilizante.

Fixado o fertilizante, a produção média é idêntica para os dois tipos de feijão.)



$$H_0: \tau_1 = \tau_2 = \dots = \tau_b$$

H_a : Pelo menos uma das igualdades não se verifica

$$W_B = \frac{\frac{\sum \sum (\bar{Y}_{.j} - \bar{Y}_{..})^2}{b-1}}{\frac{\sum \sum (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2}{(a-1)(b-1)}}$$

Rejeitamos H_0 se $W_B > w$, $P(F_{b-1, (a-1)(b-1)} > w) = \alpha$.

Sob H_0 , a produção média depende somente do tipo de feijão

Fixado o tipo de feijão, a produção média é idêntica para os quatro fertilizantes