

5. Tópicos Especiais

Modelos com Medidores Repetidos

$\hookrightarrow$ medidor num mesmo indivíduo

Modelos com medidores repetidos podem ser vistos como caso particular do modelo linear Genl.

$$y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \Sigma)$$

Σ precisa ser definida, a partir de uma conveniente especificação da matriz Σ , que neste caso, não será diagonal.

De acordo com essa formulação, se Σ é conhecida, o estimador único de β é:

$$\hat{\beta}_G = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y$$

Hipóteses sobre β podem ser testadas através da Hipótese Linear Genl:

$$H_0: C\beta = m \quad n(C) = p \quad C: p \times (k+1) \quad p \leq k+1$$

com estatística:

$$W = \frac{(C\hat{\beta}_G - m)' [C(X' \Sigma^{-1} X)^{-1} C']^{-1} (C\hat{\beta}_G - m)}{p \hat{\sigma}_G^2}$$

$$\hat{\sigma}_G^2 = \frac{1}{n-k-1} y' \left[\Sigma^{-1} - \Sigma^{-1} X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} \right] y$$

que sob $H_0 \sim F_{p, n-k-1}$ Central

Na maior parte das situações de interesse, Σ é desconhecida.

Graybill (1976) prova que:

a) se $\Sigma = X F X'$, então $\hat{\beta}_G = (X' X)^{-1} X' y$

b) Teorema 6.8.2, Graybill

Seção 3.4 do Curso

Nestes casos, a inferência é possível mesmo que os parâmetros presentes em Σ sejam desconhecidos (a forma de Σ deve ser conhecida).

11 / 06 / 2020

No entanto, planejamentos com medidor repetido são encontrados com mais frequência no contexto de análise de Variância:

- Planejamento em blocos
- Modelos de Split-plot
- Experimentos fatoriais com Repetições

O último tópico se apresenta de diferentes formas, de acordo com o tipo de economia do fator de repetição:

- 1) Experimentos com um fator e medidor repetido em todos os tratamentos
- 2) Experimentos com 2 fatores com medidor repetido em um fator
- 3) Experimentos com 2 fatores com medidor repetido em ambos os fatores.

A análise dos modelos 2 e 3) bem como a de experimentos com 3 fatores e medidor repetido pode ser encontrada em Winov (1971) - *Statistical Principles in Experimental Design*.

Modelo 1) 1 fator e medidor repetido em todos os tratamentos.

$$y_{ij} = \mu + \alpha_i + \tau_j + \epsilon_{ij} \quad i=1,2,\dots,n \quad j=1,2,\dots,K$$

com $\sum_{j=1}^K \tau_j = 0$

α_i : efeito de indivíduos, não é parâmetro, não v.a.s
 τ_j : parâmetro associado aos tratamentos
 ϵ_{ij} : parâmetros

α_i v.a. independentes $\sim N(0, \sigma_\alpha^2)$

ϵ_{ij} v.a. independentes entre si e independentes dos α_i .

$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$

τ_j - parâmetro associado ao j -ésimo tratamento $i=1,2,\dots,n \quad j=1,2,\dots,K$.

Como consequência das suposições

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

$$\text{Var}(y_{ij}) = \text{Var}(\alpha_i + \epsilon_{ij}) = \sigma_\alpha^2 + \sigma_\epsilon^2 = \sigma_y^2$$

α_i e ϵ_{ij} independentes

$i \rightarrow$ índice de indivíduos

$\rightarrow$ no mesmo indivíduo

$$\begin{aligned} \text{Cor}(y_{ij}, y_{ij'}) &= \text{Cor}(\mu + \alpha_i + \tau_j + \epsilon_{ij}, \mu + \alpha_i + \tau_{j'} + \epsilon_{ij'}) = \\ &= \text{Cor}(\alpha_i, \alpha_i) + \text{Cor}(\alpha_i, \epsilon_{ij'}) + \text{Cor}(\epsilon_{ij}, \alpha_i) + \text{Cor}(\epsilon_{ij}, \epsilon_{ij'}) \\ &= 1 + 0 + 0 + 0 \end{aligned}$$

$$= \text{Var}(x_i) = \sigma_x^2, \quad i \neq j'$$

$$\text{Cov}(y_{ij}, y_{i'j'}) = 0, \quad i \neq i' \quad (\text{covariância entre indivíduos diferentes})$$

O coeficiente de correlação entre duas medidas no mesmo elemento amostral é $\rho = \frac{\sigma_{x^2}}{\sigma_x^2 + \sigma_e^2}$ coeficiente de correlação entre clones
 $\rho(y_{ij}, y_{ij'})$

A matriz de covariância dos r. a. medidos na mesma unidade experimental é:

$$\Sigma = \sigma_y^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix} = V(y_{i1}, y_{i2}, \dots, y_{ik}), \quad i=1, \dots, n$$

matriz de variância e covariância de um vetor como em

$$\text{Var} \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1k} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2k} \\ \vdots \\ y_{n1} \\ y_{n2} \\ \vdots \\ y_{nk} \end{bmatrix} = \begin{bmatrix} \sum & 0 & \dots & 0 \\ 0 & \sum & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sum \end{bmatrix}$$

matriz bloco diagonal

Ex: Nota dada por 6 juízes para 4 tipos de vinho

		VINHOS				
	juiz	1	2	3	4	$\bar{y}_{i\cdot}$
1	20	24	28	28	25	
2	15	18	23	24	20	
3	18	19	24	23	21	$n=6$
4	26	26	30	30	28	$K=4$
5	22	24	28	26	25	
6	19	21	27	25	23	$\bar{y}_{\cdot j} = 23,67$
$\bar{y}_{\cdot j}$	20	22	26,67	26		

11 / 06 / 2010

O padrão apresentado no matriz Σ é denominado padrão "compound symmetry" (estrutura uniforme, estrutura de simetria completa)

$$\mu + \tau_1 = \mu + \tau_2 = \dots = \mu + \tau_K$$

$$\tau_1 = \tau_2 = \dots = \tau_K$$

$$\left(\sum_{j=1}^K \tau_j = 0 \right)$$

Para testar a hipótese

$$H_0: \tau_j = 0 \quad j = 1, 2, \dots, K$$

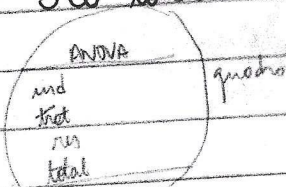
define-se

$$SQ \text{ tratamentos} = n \sum_{j=1}^K (\bar{y}_{\cdot j} - \bar{y}_{\cdot \cdot})^2$$

G L

K-1

$$SQ \text{ total} = \sum_i \sum_j (y_{ij} - \bar{y}_{\cdot \cdot})^2 \quad nK-1$$



$$* SQ \text{ Resíduos} = \sum_i \sum_j (y_{ij} - \bar{y}_{i \cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot \cdot})^2$$

não é SQ_{trat} SQ_{trat} tem um $SQ_{\text{Indivíduos}}$

com $(n-1)(K-1)$ graus de liberdade.

Estatística de teste:

$$F = \frac{QM_{\text{trat}}}{QM_{\text{res}}} = \frac{\frac{SQ_{\text{trat}}}{K-1}}{\frac{SQ_{\text{Resíduos}}}{(n-1)(K-1)}}$$

$$F \leq F_{\alpha, K-1, (n-1)(K-1)} \quad \text{aceitamos } H_0$$

$$F > F_{\alpha, K-1, (n-1)(K-1)} \quad \text{rejeitamos } H_0$$

Verifique - se que, sob H_0 , esta estatística tem distribuição F central e, de modo geral, tem distribuição F não central com

$$\lambda = n \frac{\sum \bar{\epsilon}_{ij}^2}{\sigma_{\epsilon}^2}$$

Obs: A suposição de matriz Σ compound symmetric, que permite o uso do teste F habitual no caso de dados correlacionados, é bastante restritiva. Existem outros padrões possíveis que podem ser encontrados em Huynh e Feldt [1970] - JASA, 65 (332) :1582-1589.

Em particular, vale sob o padrão de esfericidade:

$$\text{Var}(\bar{y}_{\cdot j} - \bar{y}_{\cdot j'}) \text{ constante } j \neq j' \quad * \text{ Referência Total}$$

* SPSS faz esse teste logo no início quando se declara que são modelos medidos repetidos.

Existem testes para verificar a validade do padrão "compound symmetric".

Caso não seja possível usar a estatística F habitual as soluções

são: - utilizar correções nos números de graus de liberdade; (*)

- uso de testes aproximados

- adotar a técnica de análise de variâncias multivariadas.

(*) p.1106 - Neter, Wasserman, Natchheim, Kitterer, 4ª edição.

- uso de modelos de componentes de variância. (modelos de efeitos aleatórios)

Aula: 14/6/2010 Continuação do Exemplo

y - nota dada por 6 juizes para 4 tipos de vinhos

$$y_{ij} = \mu + \alpha_i + \tau_j + \epsilon_{ij} \quad \begin{matrix} i = 1, 2, \dots, n \\ j = 1, 2, \dots, k \end{matrix}$$

$$\sum \tau_j = 0$$

$$\alpha_i \sim N(0, \sigma_{\alpha}^2)$$

$$\epsilon_{ij} \sim N(0, \sigma_{\epsilon}^2)$$

τ_j - parâmetros associados
ao j -ésimo tratamento

14 / 06 / 2010

S T O Q S S D
X ☐ ☐ ☐ ☐ ☐ ☐

$$\rho(y_{ij}, y_{ij'}) = \frac{\sigma_{\epsilon}^2}{\sigma_{\epsilon}^2 + \sigma_{\alpha}^2}$$

$$H_0: \tau_j = 0 \quad j=1, 2, \dots, K$$

$$SQ_{\text{Trat}} = n \sum_{j=1}^K (\bar{y}_{.j} - \bar{y}_{..})^2 \quad K-1 \text{ graus de liberdade}$$

$$SQ_{\text{Res}} = \sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \quad (K-1)(n-1) \text{ gl.}$$

$$F = \frac{QM_{\text{Trat}}}{QM_{\text{Res}}}$$

ANOVA

FV	SS	gl	F
indivíduos	$\sum Q_{\text{Indivíduos}}$	$n-1$	
tratamentos	$\sum Q_{\text{Trat}}$	$K-1$	$\frac{QM_{\text{Trat}}}{QM_{\text{Res}}}$
resíduos	$\sum Q_{\text{Res}}$	$(K-1)(n-1)$	

(*) $SQT = \sum Q_{\text{Trat}} + \sum Q_{\text{Res}}$

$$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4$$

No Exemplo:

FV	SQ	gl	QM	F
juizes	173,333	5	34,667	
vinhos	184,000	3	61,333	$\frac{61,333}{1,067} = 57,5$
erro	16,000	15	1,067	
	373,333	23		

$$\alpha = 0,01 \quad F_{0,3,15} = 5,42$$

Rejeitamos H_0 , concluímos que as
notas médias (para os 4 tipos de vinhos) são diferentes

Exemplo 2

O deseja-se estimar a quantidade média de nitrogênio nas folhas das árvores de um pomar.

Para isto, tomou-se uma amostra de I árvores e uma amostra de J folhas de cada árvore.

Y_{ij} - quantidade de nitrogênio na j -ésima folha da i -ésima árvore

$$i = 1, 2, \dots, I \quad j = 1, 2, \dots, J$$

Modelo :

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

α_i - efeito aleatório de árvore

ϵ_{ij} - efeito da j -ésima folha na i -ésima árvore

Suposições:

$$\alpha_i \sim N(0, \sigma_\alpha^2) \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

α_i e ϵ_{ij} v. a. independentes

$$\Rightarrow E(Y_{ij}) = E(E(Y_{ij} | \alpha_i)) = E(\mu + \alpha_i) = \mu + E(\alpha_i) = \mu$$

$$\text{Var}(Y_{ij}) = \sigma_a^2 + \sigma_e^2$$

6''

$\text{Cov}(Y_{ij}, Y_{ij'}) = \sigma_a^2 \rightarrow$ covariância entre medidas em duas folhas da mesma árvore

Além disso

$\text{Cov}(Y_{ij}, Y_{i'j'}) = 0 \rightarrow$ cov. entre medidas em árvores diferentes.

No modelo de efeitos fixos, $F = \frac{MQE}{MQD}$ teste

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

O teste similar no modelo de efeitos aleatórios é

$$H_0: \sigma_a^2 = 0 \quad [\cancel{\sigma_a^2} = 0 \Rightarrow \alpha_i = 0 \text{ com prob } 1, \\ E(Y_{ij} | \alpha_i) = \mu \quad \forall i)$$

Para fazer o teste

$$SQE = \sum_{i=1}^I J (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$SQD = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - \bar{Y}_{i.})^2$$

$$QME = \frac{SQE}{I-1}$$

$$QMD = \frac{SQD}{I(J-1)}$$

$$F = \frac{QME}{QMD}$$

Verifica-se que $\frac{\sigma_e^2 F}{J \sigma_d^2 + \sigma_e^2} \sim F_{I-1, I(J-1)}$

Sob H_0 ,

$$F = \frac{QME}{QMD} \sim F_{I-1, I(J-1)}$$

O modelo estudado é o de componentes de Variação com um fator (balanceado) ou análise de variância com um fator aleatório.

Ref:

Cap 15 - Graybill

Variance Components Models - Searle, Casella, McCulloch, 1992

10/06

Pasta 20

Nº cópias 5

8

Modelos de Efeitos Aleatórios com p variáveis explicativas
 Dados provenientes de medidas repetidas podem
 ser analisados

$$\tilde{y}_j = \begin{bmatrix} y_{j1} \\ y_{j2} \\ \vdots \\ y_{jk_j} \end{bmatrix} = X_j \beta + X_j \tilde{\alpha}_j + \tilde{\epsilon}_j \quad j=1, 2, \dots, n \quad (1)$$

$$X_j = \begin{bmatrix} x_{j11} & \dots & x_{jp1} \\ x_{j12} & & x_{jp2} \\ \vdots & & \vdots \\ x_{j1k_j} & & x_{jp k_j} \end{bmatrix} \quad k_j \text{ instantes ou situações}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{vetor de parâmetros}$$

Suposições:

i) $\tilde{\alpha}_j$ é um vetor aleatório (ou vetor de efeitos aleatórios)
 associado à j -ésima unidade

$$\tilde{\alpha}_j \sim N_p(0, \Sigma_\alpha) \quad j=1, 2, \dots, n$$

n° de variáveis explicativas no modelo

$$ii) \tilde{\epsilon}_j \sim N_{k_j}(0, \sigma^2 I_{k_j}) \quad j=1, 2, \dots, n$$

iii) $\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_n, \tilde{\epsilon}_1, \tilde{\epsilon}_2, \dots, \tilde{\epsilon}_n$ são vetores aleatórios
 independentes

$\Rightarrow$ independência entre $Y_j, Y_l \quad j \neq l$

(independência entre medidas em indivíduos distintos)

Consequências

i) $E(Y_j) = X_j \beta$

ii) $Var(Y_j) = Var(X_j \beta + X_j \tilde{b}_j + \tilde{\epsilon}_j) = Var(X_j \tilde{b}_j) + Var(\tilde{\epsilon}_j)$
 $= X_j \Sigma_b X_j' + \sigma^2 I_{K_j} = \Sigma_j$

iii) $Y_j \sim N_{K_j}(X_j \beta, \Sigma_j)$

$Y \sim N_{\sum K_j'}(X \beta, \Sigma)$

$$\Sigma = \begin{bmatrix} \Sigma_j & 0 & 0 \\ 0 & \Sigma_j & 0 \\ & & \Sigma_j \end{bmatrix}$$

Esta formulação é bastante geral pois permite diferentes quantidades de medidas para cada unidade amostral ($K_1, K_2, \dots, K_n$) e diferentes valores das variáveis explicativas.

β e Σ_j são estimados por máxima verossimilhança

~~O modelo (1) pode ser escrito na forma linear geral~~

~~$Y = X \beta + \epsilon^*$~~

~~com $\epsilon^* = X \tilde{b} + \epsilon \quad Var(\epsilon^*) = \Sigma_j$~~

$$L(\underline{y}_1, \underline{y}_2, \dots, \underline{y}_n, \underline{\beta}, \underline{\Theta}) = (2\pi)^{-\frac{n}{2}} \prod_{j=1}^n |\Sigma_j|^{-1/2} \exp \left[-\frac{1}{2} \sum_{j=1}^n (\underline{y}_j - X_j \underline{\beta})' \Sigma_j^{-1} (\underline{y}_j - X_j \underline{\beta}) \right] \quad 10$$

$\underline{\Theta}$ - parâmetros presentes em Σ_j $(\underline{y}_j - X_j \underline{\beta})$

Se $\underline{\Theta}$ é m -dimensional, verifica-se que

$\hat{\underline{\beta}}$ e $\hat{\underline{\Theta}}$ correspondem à solução de um sistema de $p+m$ equações: tel que $\Sigma_j = \sum_{g=1}^m \theta_g G_{gj}$

$$\sum_{j=1}^n X_j' \hat{\Sigma}_j^{-1} (\underline{y}_j - X_j \hat{\underline{\beta}}) = \underline{0}_p$$

$$\frac{1}{2} \sum_{j=1}^n \text{tr} \hat{\Sigma}_j^{-1} \left[(\underline{y}_j - X_j \hat{\underline{\beta}}) (\underline{y}_j - X_j \hat{\underline{\beta}})' - \hat{\Sigma}_j \right] \hat{\Sigma}_j^{-1} G_{jl} = 0$$

$$l = 1, 2, \dots, m$$

Modelos de Efeitos Aleatórios para Análise de Dados Longitudinais não Balanceados em Relação ao Tempo -

Adalberto Andreoni - Dissertação de Mestrado, 1989

pag 36

Searle, Casella, McCulloch 1992
Variance Components

Verifica-se que

11

$$\hat{\beta} = \left(\sum_{j=1}^n X_j' \hat{\Sigma}_j^{-1} X_j \right)^{-1} \left(\sum_{j=1}^n X_j' \hat{\Sigma}_j^{-1} y_j \right)$$

Se os Σ_j são conhecidos,

$$\hat{\beta} = \left(\sum_{j=1}^n X_j' \Sigma_j^{-1} X_j \right)^{-1} \left(\sum_{j=1}^n X_j' \Sigma_j^{-1} y_j \right)$$

As soluções explícitas das equações podem não ser obtidas de forma simples e em muitos casos, processos iterativos são necessários.

Caso Particular

~~$K_1 = K_2 = \dots = K_j = K$~~ e X_j é constante para todo j .
 $X_j = x^* \quad \forall j = 1, 2, \dots, n$

Ex: Distância (em mm) entre o centro da pituitária e a fissura pterio-maxilar em crianças em 4 idades

criança	Idade (anos)			
	8	10	12	14
1	21	20	21,5	23,0
2	21,0	21,5	24,0	25,5

Variáveis Explicativas

X_2 - Idade $k = 4$

$p = 2$ variáveis explicativas

$$X_j = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} = X^* \quad 4 \times 2$$

$$\underline{y}_1 = \begin{bmatrix} 21,0 \\ 20,0 \\ 21,5 \\ 23,0 \end{bmatrix}$$

$$\underline{y}_2 = \begin{bmatrix} 21,0 \\ 21,5 \\ 24,0 \\ 25,5 \end{bmatrix}$$

$$\underline{y}_{27} = \begin{bmatrix} 22,0 \\ 21,5 \\ 23,5 \\ 25,0 \end{bmatrix}$$

Lembrando que se

$$\underline{y} \sim N(X\beta, V) \quad V \text{ conhecida}$$

pag 6

Estimador de Max. Veross. de β

$$\hat{\beta}_G = (X' V^{-1} X)^{-1} X' V^{-1} \underline{y}$$

O modelo (1) pode ser escrito na forma linear geral

$$\underline{y} = X\beta + \underline{\epsilon} \quad \text{para}$$

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} x^* \\ x^* \\ \vdots \\ x^* \end{bmatrix} \quad m \times p$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad p \times 1$$

$$V = \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma & 0 \\ 0 & 0 & \Sigma \end{bmatrix}$$

$m \times m$

$$\Sigma = X^* \Sigma_\epsilon X^{*'} + \sigma^2 I_k \quad (\text{Problema: } \Sigma \text{ é desconhecida})$$

Neste caso, temos

$$VX = XF$$

$$VX = \begin{bmatrix} \sum X^* \\ \sum X^* \\ \vdots \\ \sum X^* \end{bmatrix}_{m \times n \times m \times p} \quad XF = \begin{bmatrix} X^* \\ X^* \\ \vdots \\ X^* \end{bmatrix}_{m \times p \times p \times p} \cdot F = \begin{bmatrix} X^* F \\ X^* F \\ \vdots \\ X^* F \end{bmatrix}_{m \times p}$$

$$\Sigma = X^* \sum_{2 \times 2} X^{*'} + \sigma^2 I_K$$

$$\sum_{4 \times 4} X^* = X^* \sum_{2 \times 2} X^{*'} X^* + \sigma^2 I_K X^* = X^* F$$

$$\text{para } F = \sum_{2 \times 2} X^{*'} X^* + \sigma^2 I_p$$

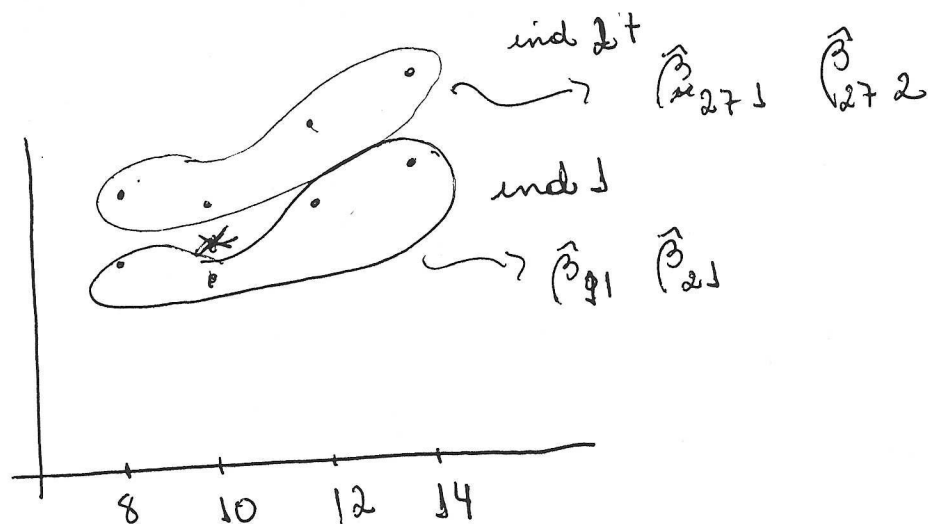
Portanto

$$\hat{\beta}_G = \hat{\beta} = (X'X)^{-1} X'Y$$

$$X'X = [n X^{*'} X^*]$$

$$X'Y = \sum_{j=1}^n X^{*'} y_j$$

$$\hat{\beta} = \frac{1}{n} [X^{*'} X^*]^{-1} \sum_{j=1}^n X^{*'} y_j = \frac{\sum_{j=1}^n [X^{*'} X^*]^{-1} X^{*'} y_j}{n}$$



$$[\hat{\beta}_1, \hat{\beta}_2] = \frac{1}{27} \sum_{j=1}^{27} (\hat{\beta}_{j1}, \hat{\beta}_{j2})$$

Inferência sobre β , Σ e σ^2

12.4.1 The Random Coefficient Growth Model
Graybill.

—11—11—

No geral, utiliza-se procedimentos iterativos para estimar β e Σ_j .

BMDP 5V - estima β e Σ_j por máxima verossimilhança, através de procedimentos iterativos (Newton Raphson, Fisher scoring, EM)

o procedimento permite a existência de dados incompletos que são imputados

Apresenta o valor das estatísticas de Wald para testar a significância dos parâmetros do modelo.

Como consequência de

$$\hat{\beta} = \frac{1}{n} \left[X^*{}' X^* \right]^{-1} \sum_{j=1}^n X^*{}' y_{\sim j}$$

$$E(\hat{\beta}) = \frac{1}{n} \left[X^*{}' X^* \right]^{-1} \sum_{j=1}^n X^*{}' X^* \beta = \frac{1}{n} \left[X^*{}' X^* \right]^{-1} n$$

$$\left[X^*{}' X^* \right] \hat{\beta} = \beta$$

$$\text{Var}(\hat{\beta}) = \frac{1}{n} \left[X^*{}' X^* \right]^{-1} X^*{}' \text{Var} \left(\sum_{j=1}^n y_{\sim j} \right) X^* \left[X^*{}' X^* \right]^{-1} \frac{1}{n}$$

$$= \frac{1}{n^2} \left[X^*{}' X^* \right]^{-1} X^*{}' n \Sigma X^* \left[X^*{}' X^* \right]^{-1}$$

$$\Sigma = X^* \Sigma_{\epsilon} X^*{}' + \sigma^2 I$$

$$= \frac{1}{n} \left[X^*{}' X^* \right]^{-1} X^*{}' \left[X^* \Sigma_{\epsilon} X^*{}' + \sigma^2 I \right] X^* \left[X^*{}' X^* \right]^{-1}$$

$$= \frac{1}{n} \left[\underbrace{\sum b + (X^*{}' X^*)^{-1} \sigma^2}_{\Sigma_0} \right] = \frac{\Sigma_0}{n}$$

$$\hat{\beta} \sim N \left(\beta, \frac{\Sigma_0}{n} \right)$$

$$\text{Var}(\hat{\beta}) = \frac{1}{n} [X^*{}' X^*]^{-1} X^*{}' \underbrace{\sum X^*}_{\Sigma = \text{Var}(\tilde{y}_j)} [X^*{}' X^*]^{-1} = \frac{\Sigma_0}{n}$$

é estimado por

$$\frac{\hat{\Sigma}_0}{n} \quad \text{onde} \quad \hat{\Sigma}_0 = [X^*{}' X^*]^{-1} X^*{}' \underbrace{\sum_{j=1}^n (\tilde{y}_j - \bar{\tilde{y}})(\tilde{y}_j - \bar{\tilde{y}})}_{\substack{\text{matriz de} \\ \text{produto cruzado}}}_{n-1}$$

$$X^* [X^*{}' X^*]^{-1}$$

$$\bar{\tilde{y}} = \frac{1}{n} \sum_{j=1}^n \tilde{y}_j$$

Intervalos de confiança para $\tilde{l}'\beta$ são da forma

$$\tilde{l}'\hat{\beta} \pm t_{\alpha/2, n-1} \sqrt{\frac{1}{n} \tilde{l}' \hat{\Sigma}_0 \tilde{l}}$$