

Modelos lineares mistos

Prof. Caio Azevedo

Modelo normal linear multivariado (marginal)

$$\mathbf{Y}_{j(k_j \times 1)} = \mathbf{X}_{j(k_j \times p)} \boldsymbol{\beta}_{(p \times 1)} + \boldsymbol{\xi}_{j(k_j \times 1)}, j = 1, \dots, n \quad (1)$$

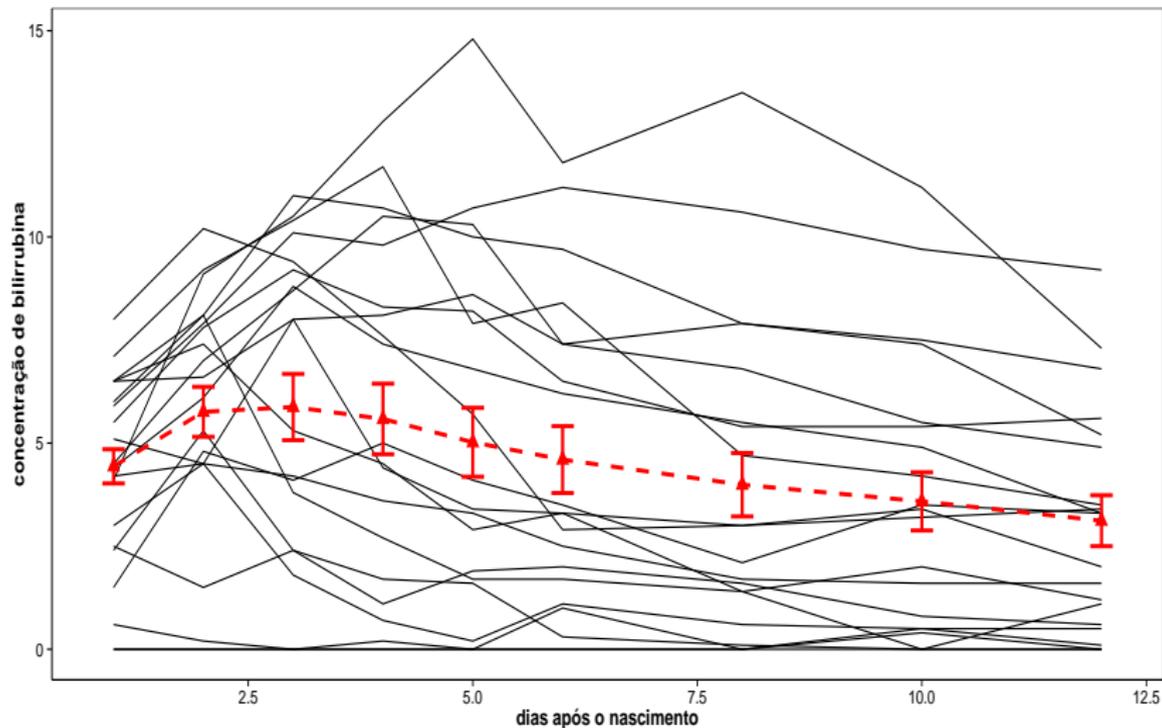
(indivíduo)

- $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{k_j j})'$, k_j : número de condições de avaliação em que o indivíduo j é avaliado.
- \mathbf{X}_j : matriz de planejamento associada aos efeitos fixos (parâmetros de regressão) para o indivíduo j (não-aleatória e conhecida).
- $\boldsymbol{\beta}$: vetor de efeitos fixos ou parâmetros de regressão (não-aleatório e desconhecido).
- $\boldsymbol{\xi}_j$: vetor de erros associado ao indivíduo j , $\boldsymbol{\xi}_j \stackrel{ind.}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_j)$.

Limitações dos modelos multivariados marginais (MMM)

- Como considerar a variabilidade dos perfis individuais?
- Como levar em consideração fontes de variabilidade não contempladas pelos erros?
- Modelagem da matriz de covariâncias recai totalmente na estrutura dos erros.
- Em muitas situações (por exemplo em Psicometria) observa-se independência condicional, entre as observações, à variáveis latentes (feitos aleatórios).
- Em muitas situações (por exemplo em Psicometria) certas variáveis latentes possuem interpretações de valioso interesse.

Cont.: 20 RN selecionados aleatoriamente e perfil médio



Modelagem para os dados do Exemplo 1 (bilirrubina)

$$Y_{ij} = \beta_0 + (x_{ij} - 1)\beta_1 + \xi_{ij},$$

$j = 1, 2, \dots, 89$, (indivíduo), $i = 1, \dots, 9$ (dia (condição de avaliação)),

- Anteriormente (MMM), consideramos que $\text{Corr}(\xi_{ij}, \xi_{i'j}) = h[d(\mathbf{t}_{ij}, \mathbf{t}_{i'j}), \boldsymbol{\rho}]$ e $\mathcal{V}(\xi_{ij}) = \sigma^2 \lambda^2(\boldsymbol{\delta}, \mathbf{v}_{ij})$.
- Com isso foi possível modelar a dependência entre as observações e a variabilidade delas (permitindo heterocedasticidade).
- Contudo, as características anteriormente mencionadas não foram contempladas.
- Uma forma de fazê-lo passa por incluir mais componentes aleatórias (não observáveis) no modelo.

Modelagem alternativa para os dados do Exemplo 1 (bilirrubina)

$$Y_{ij} = \beta_0 + (x_{ij} - 1)\beta_1 + b_j + \xi_{ij},$$

$j = 1, 2, \dots, 89$, (indivíduo), $i = 1, \dots, 9$ (dia (condição de avaliação)),

- Admita que: $b_j \perp \xi_{ij}, \forall i, j$, $b_j \stackrel{i.i.d.}{\sim} N(0, \psi)$ e $\text{Cov}(\xi_{ij}, \xi_{i'j}) = \rho$ e $\mathcal{V}(\xi_{ij}) = \sigma^2$, $\xi_j \stackrel{ind.}{\sim} N(\mathbf{0}, \Sigma_j)$.

- Assim,

- $\text{Cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} 0, & \text{se } j \neq j' \\ \psi + \sigma^2, & \text{se } i = i' \text{ e } j = j' \\ \psi + \rho, & \text{se } i \neq i', \text{ e } j = j' \end{cases}$

- Se $\rho = 0$, teremos o chamado modelo de independência condicional, pois, condicionado à b_j , as observações serão independentes.

Cont.

- x_{ij} : é o dia (1,2,3,4,5,6,8,10,12), em que a concentração de bilirrubina, correspondente ao instante i , foi medida no indivíduo j .
- Y_{ij} : é a concentração de bilirrubina no instante i do indivíduo j .
- $E(Y_{ij}|x_{ij} = 1) = \beta_0$ é a concentração esperada de bilirrubina no primeiro dia de vida (perfil marginal).
- β_1 : é o incremento na concentração esperada de bilirrubina no intervalo de um dia.
- No entanto, $E(Y_{ij}|x_{ij} = 1, b_j) = \beta_0 + b_j$ (perfil condicional).

Cont.

$$Y_{ij} = \beta_0 + (x_{ij} - 1)\beta_1 + b_j + \xi_{ij},$$

$j = 1, 2, \dots, 89$, (indivíduo), $i = 1, \dots, 9$ (dia (condição de avaliação)),

- Admita agora que: $b_j \perp \xi_{ij}, \forall i, j$, $b_j \stackrel{i.i.d.}{\sim} N(0, \psi)$ e

$$\text{Corr}(\xi_{ij}, \xi_{i'j'}) = h[d(\mathbf{t}_{ij}, \mathbf{t}_{i'j}), \boldsymbol{\varrho}] \text{ e } \mathcal{V}(\xi_{ij}) = \sigma^2 \lambda^2(\boldsymbol{\delta}, \mathbf{v}_{ij}), \boldsymbol{\xi}_j \stackrel{ind.}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_j).$$

- Assim,

$$\text{Cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} 0, & \text{se } j \neq j' \\ \psi + \sigma^2 \lambda^2(\boldsymbol{\delta}, \mathbf{v}_{ij}), & \text{se } i = i' \text{ e } j = j' \\ \psi + \sigma^2 h[d(\mathbf{t}_{ij}, \mathbf{t}_{i'j}), \boldsymbol{\varrho}], & \text{se } i \neq i', \text{ e } j = j' \end{cases}$$

Cont.

- Se considerarmos $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ e $b_j \equiv 0$, ou de modo equivalente ($\psi \equiv 0$), teremos o modelo de regressão linear tradicional (homocedástico e com as observações independentes).
- Entretanto, temos indícios de que é mais apropriado considerar alguma estrutura de dependência entre os erros (em relação às medidas feitas no mesmo indivíduo) bem como considerar perfis (individuais) diferentes.

Modelo normal linear misto

$$\mathbf{Y}_{j(k_j \times 1)} = \mathbf{X}_{j(k_j \times p)}\boldsymbol{\beta}_{(p \times 1)} + \mathbf{Z}_{j(k_j \times q)}\mathbf{b}_{j(q \times 1)} + \boldsymbol{\xi}_{j(k_j \times 1)}, j = 1, \dots, n$$

(indivíduo)

- $\mathbf{Y}_j = (Y_{j1}, \dots, Y_{jk_j})'$, k_j : número de condições de avaliação em que o indivíduo j é avaliado.
- \mathbf{X}_j : matriz de planejamento associada aos efeitos fixos para o indivíduo j (não-aleatória e conhecida).
- \mathbf{Z}_j : matriz de planejamento associada aos efeitos aleatórios para o indivíduo j (não-aleatória e conhecida).
- $\boldsymbol{\beta}$: vetor de efeitos fixos (não-aleatório e desconhecido).
- \mathbf{b}_j : vetor de efeitos aleatórios associado ao indivíduo j (aleatório e desconhecido), $\mathbf{b}_j \stackrel{ind.}{\sim} N(\mathbf{0}, \boldsymbol{\Psi})$.
- $\boldsymbol{\xi}_j$: vetor de erros associado ao indivíduo j , $\boldsymbol{\xi}_j \stackrel{ind.}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_j)$, $\mathbf{b}_j \perp \boldsymbol{\xi}_j, \forall i$.

Cont.

- Efeitos fixos: modelam características populacionais.
- Efeitos aleatórios: modelam características individuais.

Voltando ao Exemplo 1

$$\mathbf{Y}_j = \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{9j} \end{bmatrix}; \mathbf{X}_j = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 11 \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}; \mathbf{Z}_j = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}; \mathbf{b}_j = b_j;$$

$$\boldsymbol{\xi}_j = \begin{bmatrix} \xi_{1j} \\ \xi_{2j} \\ \vdots \\ \xi_{9j} \end{bmatrix}$$

Cont.

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_{89} \end{bmatrix}; \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_{89} \end{bmatrix}; \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \vdots \\ \boldsymbol{\xi}_{89} \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & 0 & \dots & 0 \\ 0 & \mathbf{Z}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{Z}_{89} \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\xi}, \mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_n)'$$

Algumas propriedades do modelo

- $\mathcal{E}(\mathbf{Y}_j|\mathbf{b}_j) = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j$.
- $\mathcal{E}(\mathbf{Y}_j) = \mathbf{X}_j\boldsymbol{\beta}$.
- $\text{Cov}(\mathbf{Y}_j|\mathbf{b}_j) = \boldsymbol{\Sigma}_j$.
- $\text{Cov}(\mathbf{Y}_j) = \mathbf{V}_j = \mathbf{Z}_j\boldsymbol{\Psi}\mathbf{Z}_j' + \boldsymbol{\Sigma}_j$.
- $\mathbf{Y}_j|\mathbf{b}_j \sim N_{k_j}(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j, \boldsymbol{\Sigma}_j)$. Além disso, como

$$\mathbf{Y}_j|\mathbf{b}_j \sim N_{k_j}(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j, \boldsymbol{\Sigma}_j)$$

$$\mathbf{b}_j \sim N_q(\mathbf{0}, \boldsymbol{\Psi})$$

portanto

$$\mathbf{Y}_j \sim N_{k_j}(\mathbf{X}_j\boldsymbol{\beta}, \mathbf{Z}_j\boldsymbol{\Psi}\mathbf{Z}_j' + \boldsymbol{\Sigma}_j)$$

Estruturas para as matrizes de covariância

- Diferentes escolhas para Ψ e Σ_j induzem diferentes estruturas de dependência para o vetor de respostas.
- Por exemplo, quando $\Sigma_j = \sigma^2 I_{k_j}$, tem-se o modelo de independência condicional homocedástico. Modelos de independência condicional são bastante considerados em psicometria (Teoria de Resposta ao item).
- Por outro lado, quando $\Sigma_j = \sigma^2 I_{k_j}$ e $\Psi \equiv \mathbf{0}$, tem-se o modelo de regressão linear usual (homocedástico e com as observações independentes).
- Dependendo da importância dos efeitos aleatórios para o estudo, podemos pensar em diferentes estruturas de covariância para eles.

Modelos para a estrutura de covariância

- Podemos, essencialmente, escolher para Ψ e Σ_j alguma das opções já vistas anteriormente.
- A covariâncias do vetor de resposta será, portanto, uma combinação das matrizes escolhidas para os erros e para os efeitos aleatórios.
- Naturalmente, as escolhas são limitadas pelo recurso computacional a ser utilizado.
- Em geral a matriz Ψ é assumida ser não estruturada ou diagonal.

Estimação

- Sob a ótica frequentista, em geral, trabalha-se com a distribuição marginal de \mathbf{Y}_j em relação à \mathbf{b}_j , ou seja $\mathbf{Y}_j \sim N(\mathbf{X}_j\boldsymbol{\beta}, \mathbf{Z}_j\boldsymbol{\Psi}\mathbf{Z}_j' + \boldsymbol{\Sigma}_j)$
- Alternativa: algoritmo EM utilizando a distribuição conjunta de (\mathbf{Y}, \mathbf{b}) .
- Também existem métodos Bayesianos.
- Suposição : $\boldsymbol{\Sigma}_j = g(\boldsymbol{\theta}_1)$ e $\boldsymbol{\Psi} = h(\boldsymbol{\theta}_2)$ de modo que $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ não possuem componentes comuns.

Cont.

- Log-verossimilhança (marginal) para n observações:

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\theta}) &= -\frac{1}{2} \ln(2\pi) \sum_{j=1}^n k_j - \frac{1}{2} \sum_{j=1}^n \ln |\mathbf{V}_j| \\ &\quad + -\frac{1}{2} \sum_{j=1}^n (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})' \mathbf{V}_j^{-1} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta}) \end{aligned} \quad (2)$$

$$\mathbf{V}_i \equiv \mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_j \boldsymbol{\Psi}(\boldsymbol{\theta}_2) \mathbf{Z}_j' + \boldsymbol{\Sigma}_j(\boldsymbol{\theta}_1), \quad \boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$$

- Se $\boldsymbol{\theta}$ for conhecido, o estimador de MV (que corresponde ao estimador de MQG) de $\boldsymbol{\beta}$ é dado por:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^n \mathbf{X}_j' \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \left(\sum_{j=1}^n \mathbf{X}_j' \mathbf{V}_j^{-1} \mathbf{Y}_j \right) \quad (3)$$

Cont.

- Para estimar (θ) , substituímos (3) em (2), obtendo uma log-verossimilhança perfilada:

$$\begin{aligned} l(\theta) &= -\frac{1}{2} \ln(2\pi) \sum_{j=1}^n k_j - \frac{1}{2} \sum_{j=1}^n \ln |\mathbf{V}_j| \\ &\quad - \frac{1}{2} \sum_{j=1}^n (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta})' \mathbf{V}_j^{-1} (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta}) \end{aligned} \quad (4)$$

- A maximização da log-verossimilhança (4) tem de ser feita através de métodos iterativos como os algoritmos de Newton-Raphson, Escore de Fisher, Gauss-Newton, BFGS.
- Uma vez que tais estimativas forem obtidas, as inserimos em (3).

Cont.

- As distribuições assintóticas dos estimadores podem ser obtidas através do TCL.
- Os erros-padrão assintóticos podem ser obtidos através das inversas das informações de Fisher para (para θ) e através de uma fórmula analítica (para β).
- Os estimadores de MV para β são não viesados, mas o mesmo não acontece com os estimadores de MV de θ .
- Alternativa: estimadores de MV restritos (MVR) (também chamados de estimadores MV residuais).

Algoritmo (estimação por MV)

- Estima-se θ através de algum algoritmo de maximização conveniente (NR, RF, Gauss-Newton, BFGS), resolvendo-se o sistema de equações dado por:

$$\begin{aligned} S(\theta_k) &= \frac{\partial l(\theta)}{\partial \theta_k} = -\frac{1}{2} \sum_{j=1}^n \frac{\partial \ln |\mathbf{V}_j|}{\partial \theta_k} - \frac{1}{2} \sum_{j=1}^n \text{tr} \left[\frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta}) (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta})' \right. \\ &+ \left. \mathbf{V}_j^{-1} \frac{\partial (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta}) (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta})'}{\partial \theta_k} \right] = -\frac{1}{2} \sum_{j=1}^n \text{tr} \left[\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_k} \right] \\ &- \frac{1}{2} \sum_{j=1}^n \text{tr} \left[\frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta}) (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta})' \right. \\ &+ \left. \mathbf{V}_j^{-1} \frac{\partial (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta}) (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta})'}{\partial \theta_k} \right] \end{aligned}$$

Algoritmo (estimação por MV)

- A notação $\frac{\partial \mathbf{V}_j}{\partial \boldsymbol{\theta}}$ representa a derivada de \mathbf{V}_j com relação à cada componente de $\boldsymbol{\theta}$ o que resulta, para cada componente, numa matriz.
- Com as estimativas de $\boldsymbol{\theta}$, digamos $\tilde{\boldsymbol{\theta}}$, obtem-se as estimativas de $\boldsymbol{\beta}$, através de:

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{x}_j \right)^{-1} \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j^{-1}(\tilde{\boldsymbol{\theta}}) \mathbf{y}_j \right)$$

- A matriz de covariâncias de $\hat{\boldsymbol{\beta}}$ é dada por $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j(\boldsymbol{\theta})^{-1} \mathbf{x}_j \right)^{-1}$ e uma estimativa é dada por:
$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{x}_j \right)^{-1}.$$

- Para o estimador $\hat{\theta}$ uma aproximação da matriz de covariâncias pode ser obtida através da inversa da matriz $-\mathbf{H}(\theta) = -\frac{\partial l(\theta)}{\partial \theta \partial \theta'}$ e uma estimativa é dada pela inversa de: $-\mathbf{H}(\tilde{\theta}) = -\frac{\partial l(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\tilde{\theta}}$, respectivamente $\boldsymbol{\Sigma}_{\theta} = -\mathbf{H}(\theta)^{-1}$ e $\tilde{\boldsymbol{\Sigma}}_{\theta} = -\mathbf{H}(\tilde{\theta})^{-1}$.
- Os erros-padrão dos estimadores $\hat{\beta}$ e $\hat{\theta}$ correspondem à raiz quadrada dos elementos da diagonal principal das respectivas matrizes de covariância.

- A distribuição dos estimadores (exata ou assintótica) pode ser obtida através de um dos seguintes métodos:
 - Convergência em distribuição dos estimadores de máxima verossimilhança ($\hat{\beta} \approx N_p(\beta, \Sigma_\beta)$ e $\hat{\theta} \approx N_r(\theta, \Sigma_\theta)$), para n suficientemente grande.
 - Métodos de reamostragem.
 - Método Delta (para funções, não lineares, dos parâmetros, que sejam de interesse).

Máxima verossimilhança restrita (ou residual)

- MVR: consiste em maximizar a verossimilhança de uma transformação ortogonal do vetor de respostas, ou seja, da verossimilhança induzida por $\mathbf{Y}_j^* = \mathbf{U}_j \mathbf{Y}_j$,
- Em geral, $\mathbf{U}_j = \mathbf{I}_{k_j} - \mathbf{X}_j(\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j'$.
- Assim, $\mathbf{Y}_j^* \sim N(\mathbf{0}_{k_j}, \mathbf{U}_j \mathbf{V}_j \mathbf{U}_j')$, em que $\mathbf{V}_j = \text{Cov}(\mathbf{Y}_j) = \mathbf{Z}_j \boldsymbol{\Psi} \mathbf{Z}_j' + \boldsymbol{\Sigma}_j$.
- Os estimadores de MVR de β são não viesados enquanto que o viés do estimadores de MVR de θ são menores em comparação com os estimadores de MV.
- O nome “residual” vem do fato de que a matriz \mathbf{U}_j gera os resíduos no ajuste por mínimos quadrados ordinários.

- A log-verossimilhança residual ou restrita é dada por

$$\begin{aligned} l_R(\boldsymbol{\theta}) &= -\frac{1}{2} \ln(2\pi) \sum_{j=1}^n k_j - \frac{1}{2} \sum_{j=1}^n \ln |\mathbf{U}_j \mathbf{V}_j \mathbf{U}_j| \\ &\quad - \frac{1}{2} \sum_{j=1}^n (\mathbf{Y}_j^*)' \mathbf{U}_j \mathbf{V}_j^{-1} \mathbf{U}_j (\mathbf{Y}_j^*). \end{aligned}$$

Cont.

- A log-verossimilhança residual ou restrita pode ser escrita como

$$\begin{aligned}l_R(\boldsymbol{\theta}) &= -\frac{1}{2} \ln(2\pi) \sum_{j=1}^n k_j - \frac{1}{2} \sum_{j=1}^n \ln |\mathbf{V}_j| \\ &\quad - \frac{1}{2} \sum_{j=1}^n \left(\boldsymbol{y}_j - \mathbf{x}_j \hat{\boldsymbol{\beta}} \right)' \mathbf{V}_j^{-1} \left(\boldsymbol{y}_j - \mathbf{x}_j \hat{\boldsymbol{\beta}} \right) \\ &\quad - \frac{1}{2} \ln \left| \sum_{j=1}^n \mathbf{x}_j' \mathbf{V}_j^{-1} \mathbf{x}_j \right| + \text{const.}\end{aligned}\tag{5}$$

em que $\hat{\boldsymbol{\beta}}$ é dado em (3).

- Uma vez que os estimadores de MVR de $\boldsymbol{\theta}$ forem obtidos, maximizando-se (5) (numericamente), os estimadores de MVR de $\boldsymbol{\beta}$ podem ser obtidos inserindo aqueles em (3).

Cont.

- As distribuições exatas ou assintóticas dos estimadores de MVR podem ser obtidas de modo semelhante aos dos estimadores de MV.
- Lembrem-se de que estamos lidando com um conjunto de vetores aleatórios independentes mas não identicamente distribuídos $\mathbf{Y}_j \stackrel{ind.}{\sim} N_{k_j}(\mathbf{X}_j\boldsymbol{\beta}, \mathbf{V}_j)$.
- TLC's que levem tal estrutura em consideração devem ser utilizados.

Algoritmo (estimação por MVR)

- Estima-se θ através de algum algoritmo de maximização conveniente (NR, RF, Gauss-Newton, BFGS), resolvendo-se o sistema de equações dado por:

$$\begin{aligned} S(\theta_k) &= \frac{\partial l_R(\theta)}{\partial \theta_k} = -\frac{1}{2} \sum_{j=1}^n \text{tr} \left[\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_k} \right] \\ &- \frac{1}{2} \sum_{j=1}^n \text{tr} \left[\frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta}) (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta})' \right. \\ &+ \left. \mathbf{V}_j^{-1} \frac{\partial (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta}) (\mathbf{Y}_j - \mathbf{X}_j \hat{\beta})'}{\partial \theta_k} \right] - \frac{1}{2} \sum_{i=1}^n \text{tr} \left[\mathbf{V}_j^{-1} \mathbf{X}_i' \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} \mathbf{V}_j^{-1} \mathbf{X}_i \right] \end{aligned}$$

Algoritmo (estimação por MVR)

- Com as estimativas de θ , digamos $\tilde{\theta}_R$, obtem-se as estimativas de β_R , ou seja:

$$\tilde{\beta}_R = \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j(\tilde{\theta})^{-1} \mathbf{x}_j \right)^{-1} \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j^{-1}(\tilde{\theta}) \mathbf{y}_j \right)$$

- A matriz de covariâncias de $\hat{\beta}_R$ é dada por $\Sigma_{\beta_R} = \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j(\theta)^{-1} \mathbf{x}_j \right)^{-1}$ e uma estimativa é dada por:
 $\tilde{\Sigma}_{\beta_R} = \left(\sum_{j=1}^n \mathbf{x}'_j \mathbf{V}_j(\tilde{\theta})^{-1} \mathbf{x}_j \right)^{-1}$.

- Para o estimador θ_R uma aproximação da matriz de covariâncias pode ser obtida através da inversa da matriz

$$\Sigma_{\theta_R} = -\mathbf{H}_R(\theta) = -\frac{\partial l_R(\theta)}{\partial \theta \partial \theta'}$$

e uma estimativa é dada pela inversa de: $-\mathbf{H}_R(\tilde{\theta}) = -\frac{\partial l_R(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\tilde{\theta}}$, respectivamente $\Sigma_{\theta_R} = -\mathbf{H}_R(\theta)^{-1}$ e $\tilde{\Sigma}_{\theta_R} = -\mathbf{H}_R(\tilde{\theta})^{-1}$.

- Os erros-padrão dos estimadores $\hat{\beta}_R$ e $\hat{\theta}_R$ correspondem à raiz quadrada dos elementos da diagonal principal das respectivas matrizes de covariância.

- A distribuição dos estimadores pode ser obtida através de um dos seguintes métodos:
 - Convergência em distribuição dos estimadores de máxima verossimilhança.
 - Métodos de reamostragem
 - Método Delta.

Cont.

- Preditores para os efeitos aleatórios podem ser obtidos através da distribuição condicional (à posteriori), de $\mathbf{b}_j | \mathbf{y}_j$, ou seja

$$p(\mathbf{b}_j | \mathbf{y}_j) = \frac{p(\mathbf{y}_j | \mathbf{b}_j) p(\mathbf{b}_j)}{\int_{\mathbb{R}^q} p(\mathbf{y}_j | \mathbf{b}_j) p(\mathbf{b}_j) d\mathbf{b}}$$

a qual corresponde à

$$\mathbf{b}_j | \mathbf{y}_j \sim N_q \left(\boldsymbol{\Psi} \mathbf{Z}_j' \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}), (\mathbf{Z}_j \boldsymbol{\Sigma}_j^{-1} \mathbf{Z}_j' + \boldsymbol{\Psi}^{-1})^{-1} \right)$$

- Assim, um preditor para \mathbf{b}_j seria sua média condicional (à posteriori) ou seja, $\hat{\mathbf{b}}_j = \hat{\boldsymbol{\Psi}} \mathbf{Z}_j' \hat{\mathbf{V}}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}})$, em que $\hat{\cdot}$ denota um dos estimadores vistos anteriormente (MV ou MVR).

Cont.

- Medida de precisão de $\hat{\mathbf{b}}_j - \mathbf{b}_j$, $\text{Cov}(\hat{\mathbf{b}}_j - \mathbf{b}_j) = \Psi - \text{Cov}(\hat{\mathbf{b}}_j)$. Em que

$$\text{Cov}(\hat{\mathbf{b}}_j) = \Psi \mathbf{Z}'_j \left(\mathbf{V}_j^{-1} - \mathbf{V}_j \mathbf{X}_j \left(\sum_{j=1}^n \mathbf{X}'_j \mathbf{V}_j \mathbf{X}_j \right)^{-1} \mathbf{X}'_j \mathbf{W}_j \right) \mathbf{X}_j \Psi$$

Intervalos de Confiança

- Seja $\hat{\vartheta}$ o componente de interesse do vetor $\hat{\beta}$ ou do vetor $\hat{\theta}$ e $\widehat{EP}(\hat{\vartheta})$ um estimador consistente (como aqueles apresentados) do respectivo erro-padrão.
- IC assintótico com coeficiente de confiança de γ

$$\hat{\vartheta} \pm z_{(1+\gamma)/2} \widehat{EP}(\hat{\vartheta})$$

$$P(Z \leq z_{(1+\gamma)/2}) = \frac{1+\gamma}{2}$$

Testes de Hipótese

- Seja $\widehat{\Sigma}_\beta$ um estimador consistente da matriz de covariâncias de $\widehat{\beta}$ (como aqueles apresentados).
- Desejamos testar $H_0 : \mathbf{C}\beta = \mathbf{M}$ vs $H_1 : \mathbf{C}\beta \neq \mathbf{M}$
- Podemos usar a seguintes estatística (do tipo Wald)

$$Q = (\mathbf{C}\widehat{\beta} - \mathbf{M})' (\mathbf{C}\widehat{\Sigma}_\beta\mathbf{C}')^{-1} (\mathbf{C}\widehat{\beta} - \mathbf{M})$$

para n suficientemente grande, temos que $Q \sim \chi^2_{(r(\mathbf{C}),\delta)}$,

$$\delta = (\mathbf{C}\beta - \mathbf{M})' (\mathbf{C}\Sigma_\beta\mathbf{C}')^{-1} (\mathbf{C}\beta - \mathbf{M})$$

Comentários

- Em relação aos testes de hipótese para θ , podemos proceder de modo análogo ao que fizemos para β .
- Note, contudo, que existem três tipos de parâmetros em θ : parâmetros de variância (σ^2), de correlação (ρ) e de covariância (σ_1). Para os parâmetros de variância, faz-se necessário testes mais específicos quando $M = 0$.
- Para outros detalhes, veja as referências.

Seleção de modelos: Teste da razão de verossimilhanças

- Seja $\hat{\theta}_i$ o estimador de máxima verossimilhança obtido sob o modelo i e $\tilde{\theta}_i$ sua respectiva estimativa.
- Denote por $L_i(\hat{\theta}_i)$ e $l_i(\hat{\theta}_i)$ o máximo da verossimilhança e da log-verossimilhança do modelo i , respectivamente, avaliados nos respectivos estimadores de MV, enquanto que $L_i(\tilde{\theta}_i)$ e $l_i(\tilde{\theta}_i)$ são os respectivos máximos avaliados nas estimativas de MV.
- No caso dos modelos mistos usa-se a log-verossimilhança marginal.

Teste da razão de verossimilhanças (cont.)

- A estatística do TRV é dada por $\Delta = \frac{L_1(\hat{\theta}_1)}{L_2(\hat{\theta}_2)}$.
- Rejeita-se H_0 se $\Delta \leq \delta_c$, em que δ_c é um valor crítico adequado.
- Alternativamente, rejeitamos H_0 se

$$\Lambda = -2\ln(\Delta) = -2 \left(l_1(\hat{\theta}_1) - l_2(\hat{\theta}_2) \right) \geq \lambda_c,$$

em que $P(Q \geq \lambda_c) = \alpha$, $Q \approx \chi^2_{(\gamma)}$ e

$\gamma =$ número de parâmetros do modelo M_2 - número de parâmetros do modelo M_1 .

- Nesse caso, p -valor $\approx P(Q \geq \lambda | H_0)$, em que λ é o valor observado da estatística Λ e $Q \sim \chi^2_{(\gamma)}$. Assim, rejeita-se H_0 se p -valor $\leq \alpha$.

Estatísticas de comparação de modelos

- O TRV é apropriado na comparação somente de modelos encaixados (o modelo com menor número de parâmetros é um caso particular do modelo com maior número de parâmetros).
- Além disso, ele não leva em consideração (diretamente) o número de parâmetros do modelo (somente na distribuição da estatística).
- Existem várias alternativas, em termos de estatísticas para comparar modelos, que “penalizam” a verossimilhança em relação ao número de parâmetros, tamanho da amostra entre outros fatores.
- Veremos o AIC e o BIC.

Estatísticas de comparação de modelos (cont.)

- O AIC e BIC, para o i -ésimo modelo, são dados, respectivamente, por:

$$AIC_i = -2l_i(\tilde{\theta}_i) + 2k$$

$$BIC_i = -2l_i(\tilde{\theta}_i) + k \ln(n)$$

que $l_i(\tilde{\theta}_i)$ denota a log-verossimilhança do i -ésimo modelo avaliada em alguma estimativa (p.e. máxima verossimilhança), k é o número de parâmetros e n é o número de observações.

- Portanto, o modelo que apresentar os menores valores, será o modelo “melhor ajustado” aos dados.

Função “lme” pacote “nlme”

- Ajusta a classe de modelos (1) por MV ou MRV, permitindo heterocedasticidade (entre as condições de avaliação e/ou grupos) sob diversas estruturas de correlação.
- Seja $\mu_{(c)ij} = \mathcal{E}(Y_{ij}) = \mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_j$ (média condicional), $\mu_{(m)ij} = \mathcal{E}(Y_{ij}) = \mathbf{X}'_{ij}\boldsymbol{\beta}$ (média marginal), em que \mathbf{X}'_{ij} é i -ésima linha da matriz \mathbf{X}_j e o equivalente em relação à matriz \mathbf{Z}'_{ij} .
- A função lme trabalha com a estrutura $\text{Cov}(\mathbf{Y}_j) = \mathbf{V}_j = \sigma^2 \mathbf{Z}_j \mathbf{D}(\boldsymbol{\theta}_D) \mathbf{Z}'_j + \sigma^2 \mathbf{R}_j = \sigma^2 \mathbf{Z}_j \mathbf{D}(\boldsymbol{\theta}_D) \mathbf{Z}'_j + \sigma^2 \boldsymbol{\Lambda}_j \mathbf{C}_j \boldsymbol{\Lambda}_j$, em que $\sigma^2 > 0$ é um parâmetro de escala, $\boldsymbol{\Lambda}_j$ é uma matriz diagonal com elementos positivos (para permitir heterocedasticidade) e \mathbf{C}_j é uma matriz de correlações (como aquelas vistas anteriormente).

Cont.

- Temos ainda que

$$\mathbf{R}_j(\mu_{ij}, \boldsymbol{\theta}; \mathbf{v}_{ij}) = \begin{cases} \mathbf{R}_j(\boldsymbol{\theta}; \mathbf{v}_{ij}) = \Lambda_j(\boldsymbol{\delta}; \mathbf{v}_{ij}) \mathbf{C}_j(\boldsymbol{\varrho}) \Lambda_j(\boldsymbol{\delta}; \mathbf{v}_{ij}) \\ \Lambda_j(\mu_{ij}, \boldsymbol{\delta}; \mathbf{v}_{ij}) \mathbf{C}_j(\boldsymbol{\varrho}) \Lambda_j(\mu_{ij}, \boldsymbol{\delta}; \mathbf{v}_{ij}) \end{cases}$$

em que $\boldsymbol{\theta} = (\boldsymbol{\delta}', \boldsymbol{\varrho}')$, $\boldsymbol{\delta}$ é um vetor de parâmetros de variância e \mathbf{v}_{ij} é um vetor de covariáveis (conhecidas) das variâncias.

- Portanto, a variância de cada observação é dada por

$$\mathcal{V}(Y_{ij}) = \sigma^2 D_j + \begin{cases} \sigma^2 \lambda^2(\boldsymbol{\delta}; \mathbf{v}_{ij}) \\ \sigma^2 \lambda^2(\mu_{ij}, \boldsymbol{\delta}; \mathbf{v}_{ij}) \end{cases}$$

em que D_j é o j -ésimo elemento da diagonal principal de

$$\mathbf{Z}_j \mathbf{D}(\boldsymbol{\theta}_D) \mathbf{Z}_j'$$

Cont.

- Especificação da matriz de covariâncias dos efeitos aleatórios (Ψ) na notação do pacote lme (D).

| Nome | Descrição |
|------------|--|
| pdIdent | ψI_q |
| pdDiag | $\text{diag}(\psi_1, \dots, \psi_q)$ |
| pdCompSymm | uniforme |
| pdLogChol | não estruturada com parametrização log-Cholesky (default) |
| pdSymm | não estruturada com parametrização SVD (decomposição) |
| pdNatural | não estruturada com parametrização "natural" |
| pdBlocked | matriz bloco diagonal com os blocos definidos como uma das classes acima |

Modelagem para os dados do Exemplo 1 (bilirrubina)

$$Y_{ij} = \mu_{ij} + \xi_{ij},$$

- (1) $\mu_{ij} = \beta_0 + \beta_1(x_{ij} - 1)I_{(i \in \{1,2\})} + \beta_2(x_{ij} - 1)I_{(i \in \{3,4,5,6,7,8,9\})} + b_{1j}$
; (2) $\mu_{ij} = \beta_0 + \beta_1(x_{ij} - 1)I_{(i \in \{1,2\})} + \beta_2(x_{ij} - 1)I_{(i \in \{3,4,5,6,7,8,9\})} + b_{1j}$
+ $b_{2j}x_{ij}$.
- (1) : $\mathcal{V}(Y_{ij}) = \sigma^2$ (homocedástico);
(2) $\mathcal{V}(Y_{ij}) = \sigma_i^2 = \sigma^2 \exp(x_{ij}\delta)$, $i = 1, 2, \dots, 9$ (heterocedástico) (note que $x_{ij} = x_{ij'}$, $\forall i, j, j'$).
- $\text{Corre}(Y_{ij}, Y_{i'j})$ (1) AR(1), (2)(ARMA(1,1)).
- $(b_{1j}, b_{2j}) \stackrel{i.i.d}{\sim} N_2(\mathbf{0}, \Psi)$, $\Psi = \begin{bmatrix} \psi_1 & \psi_0 \\ \psi_0 & \psi_2 \end{bmatrix}$

Modelos

| Modelo | Variância | Correlação | Efeitos aleatórios |
|-------------|-----------------|------------|---------------------------------|
| HAR1I | Homocedástico | AR(1) | intercepto |
| HAR1ICA | Homocedástico | AR(1) | intercepto, coeficiente angular |
| HARMA11I | Homocedástico | ARMA(1,1) | intercepto |
| HARMA11ICA | Homocedástico | ARMA(1,1) | intercepto, coeficiente angular |
| HEAR1I | Heterocedástico | AR(1) | intercepto |
| HEAR1ICA | Heterocedástico | AR(1) | intercepto, coeficiente angular |
| HEARMA11I | Heterocedástico | ARMA(1,1) | intercepto |
| HEARMA11ICA | Heterocedástico | ARMA(1,1) | intercepto, coeficiente angular |

Modelos

| Modelo | AIC | BIC |
|---------------|---------|---------|
| HAR1I | 2926,41 | 2954,51 |
| HAR1ICA | 2930,41 | 2967,87 |
| HARMA11I | 2920,82 | 2953,59 |
| HARMA11ICA | 2924,82 | 2966,95 |
| HEAR1I | 2854,91 | 2887,69 |
| HEAR1ICA | 2858,91 | 2901,05 |
| HEARMA11I | 2852,96 | 2890,41 |
| HEARMA11ICA | 2856,96 | 2903,78 |

O MMM selecionado foi HEARMA11 (mesma função de variância e com regressão segmentada): AIC= 2871,71; BIC=2904,48

Estimativas dos parâmetros

MLM

| Parâmetro | Estimativa | EP | IC(95%) | Estatística | p-valor |
|-----------|------------|------|------------------|-------------|---------|
| β_0 | 5,37 | 0,40 | [4,59 ; 6,15] | 13,48 | <0,0001 |
| β_1 | 0,32 | 0,14 | [0,05 ; 0,59] | 2,33 | 0,0198 |
| β_2 | -0,21 | 0,02 | [-0,26 ; -0,16] | -8,36 | <0,0001 |

MMM

| Parâmetro | Estimativa | EP | IC(95%) | Estatística | p-valor |
|-----------|------------|------|-----------------|-------------|---------|
| β_0 | 5,12 | 0,42 | [4,30 ; 5,94] | 12,28 | <0,0001 |
| β_1 | 0,37 | 0,12 | [0,13 ; 0,60] | 3,07 | 0,0021 |
| β_2 | -0,19 | 0,03 | [-0,25 ; -0,13] | -6,08 | <0,0001 |

Estimativas dos parâmetros

MLM

| Parâmetro | Estimativa | IC(95%) |
|------------|------------|------------------|
| σ^2 | 13,20 | [9,45 ; 18,44] |
| δ | -0,08 | [-0,10 ; -0,07] |
| ψ_1 | 6,45 | [4,29 ; 9,68] |

OBS: No modelo com intercepto e coeficiente aleatório, semelhante ao modelo escolhidos, $\tilde{\psi}_2 = < 0,0001$ (indicando uma não significância do efeito aleatório).

MMM

| Parâmetro | Estimativa | IC(95%) |
|------------|------------|------------------|
| σ^2 | 19,65 | [14,58 ; 26,49] |
| δ | -0,05 | [-0,06 ; -0,04] |

Estimativas dos parâmetros

MLM

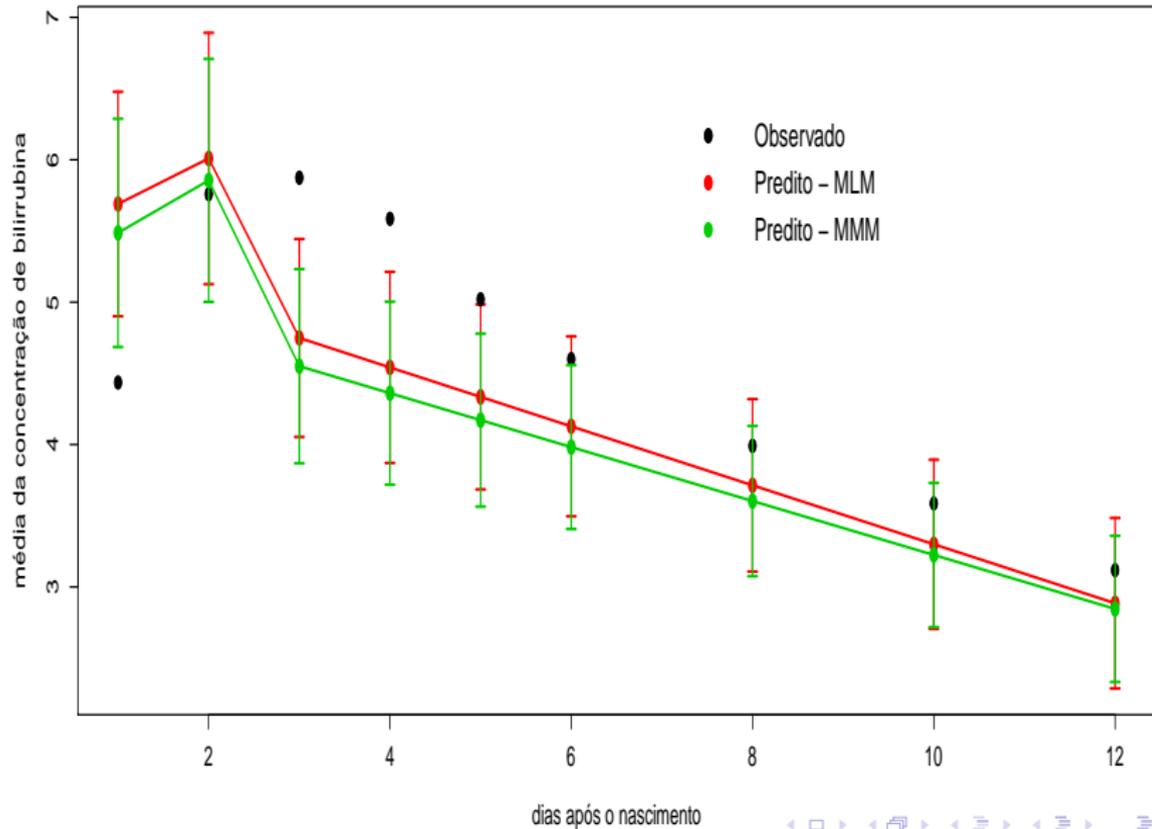
| Parâmetro | Estimativa | IC(95%) |
|-----------|------------|----------------|
| ϕ | 0,84 | [0,78 ; 0,89] |

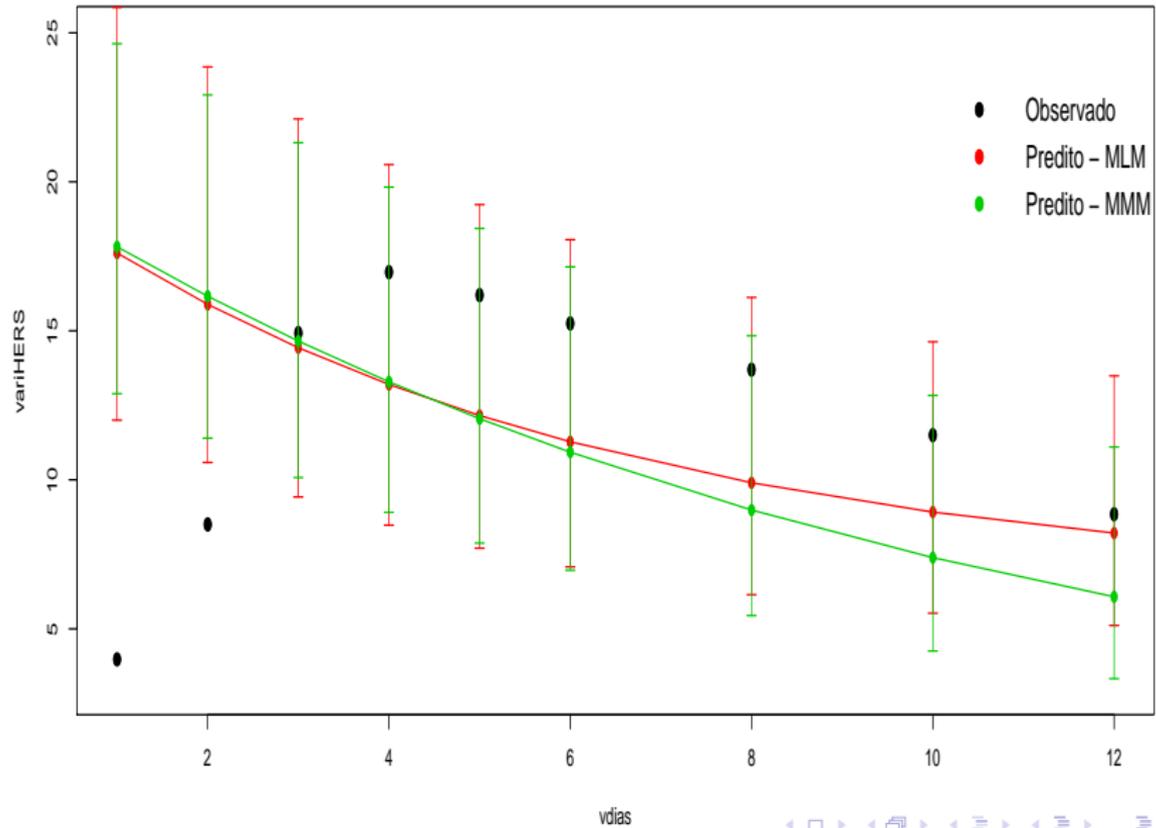
OBS: No modelo ARMA(1,1), semelhante ao modelo escolhido,

$\tilde{\theta}_1 = 0,086[-0,002; 0,173]$ (indicando uma contribuição menos significativa desse coeficiente).

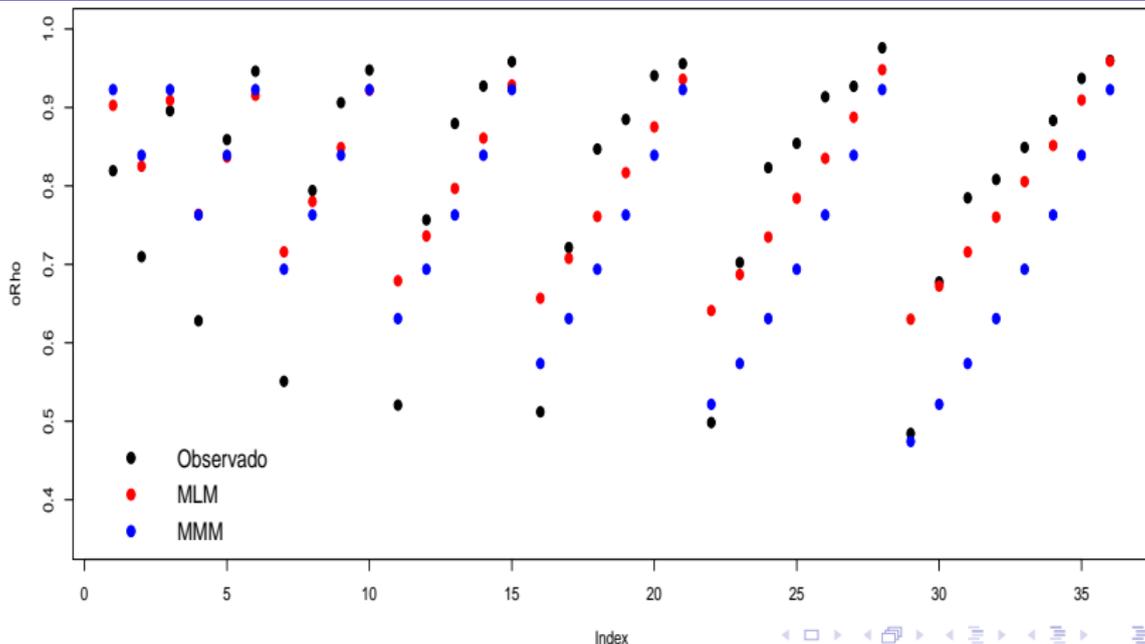
MMM

| Parâmetro | Estimativa | IC(95%) |
|------------|------------|--------------|
| ϕ_1 | 0,91 | [0,88 ;0,93] |
| θ_1 | 0,09 | [0,02 ;0,17] |

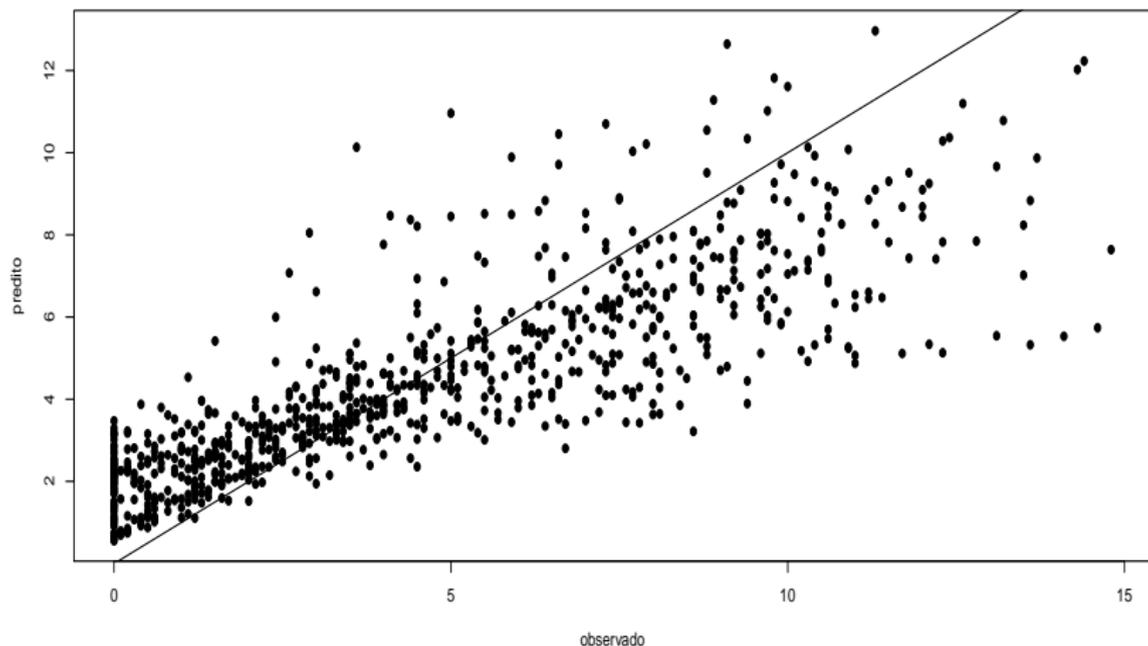




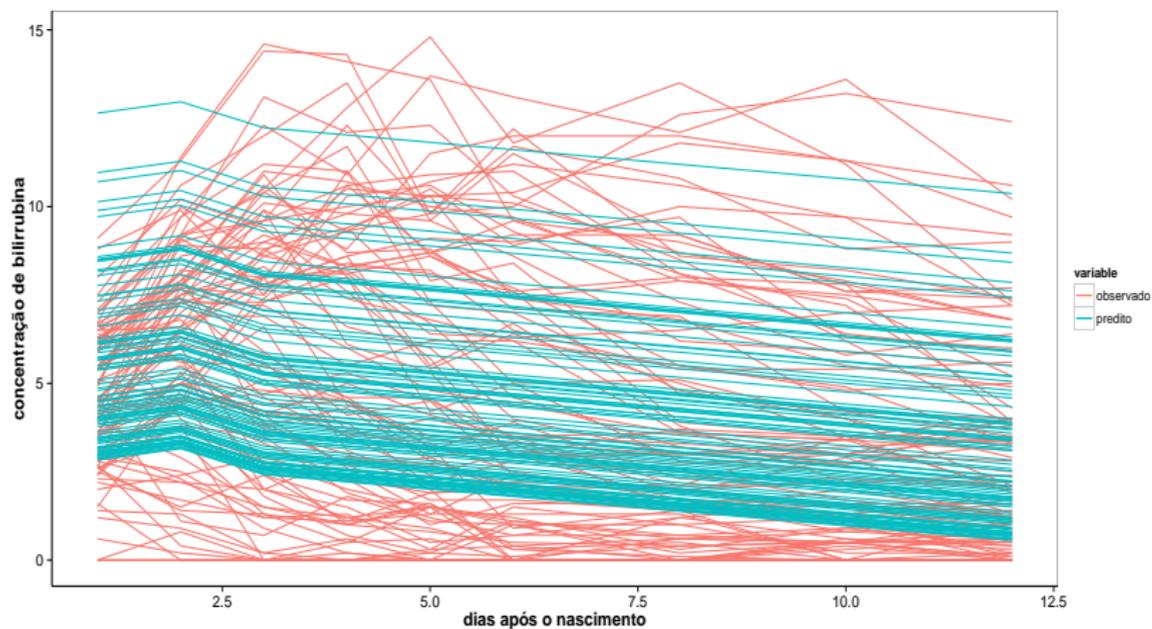
Média dos desvios absolutos (MLM = 0,063;
MMM=0,096)



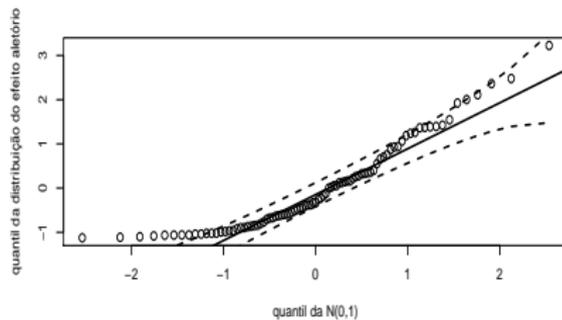
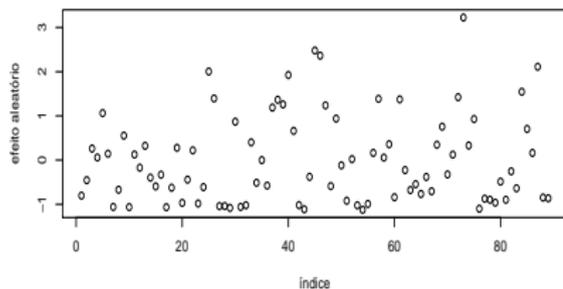
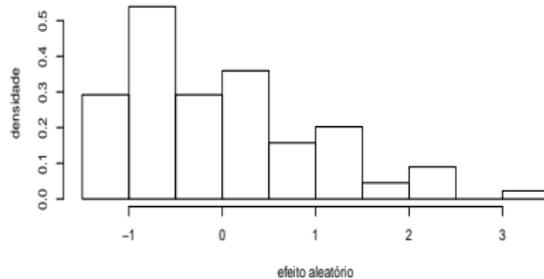
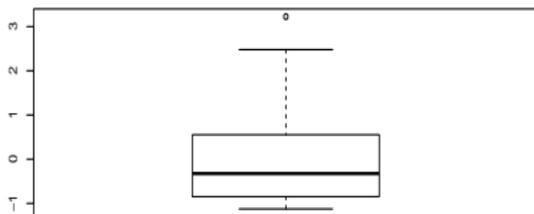
Valores individuais preditos: $\widehat{Y}b_j = \mathbf{X}_j\widehat{\beta} + \mathbf{Z}_j\widehat{b}_j$



Perfis individuais preditos e observados



Efeitos aleatórios



Análise de resíduos para modelos mistos

- Existem duas fontes de variação: os efeitos aleatórios \mathbf{b} e os erros (marginais) ξ .
- Tipos de erros:
 - Erros condicionais: $\xi_j = \mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j\mathbf{b}_j$
 - Erros marginais: $\epsilon = \mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta} = \mathbf{Z}_j\mathbf{b}_j + \xi_j$
 - Efeitos aleatórios: $\mathbf{Z}_j\mathbf{b}_j = \mathcal{E}(\mathbf{Y}_j|\mathbf{b}_j) - \mathcal{E}(\mathbf{Y}_j)$.
- Respectivos resíduos (valores preditos):
 - Resíduos condicionais: $\hat{\xi}_j = \mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}} - \mathbf{Z}_j\hat{\mathbf{b}}_j$
 - Resíduos marginais: $\hat{\epsilon}_j = \mathbf{Y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}} = \mathbf{Z}_j\hat{\mathbf{b}}_j + \hat{\xi}_j$

Tipos de resíduos

- Segundo Hilden-Minton (1995)
 - Resíduo puro (para um específico tipo de erro): se ele depende apenas das componentes fixas e do erro que ele pretende predizer.
 - Resíduo confundido: depende de outros tipos de erros.

Tipos de resíduos

- Na função “lme”

- Resíduo condicional normalizado: $\hat{\xi}_j^* = (\hat{\sigma} \hat{\mathbf{U}}_{(c)j})^{-1} \hat{\xi}_j$

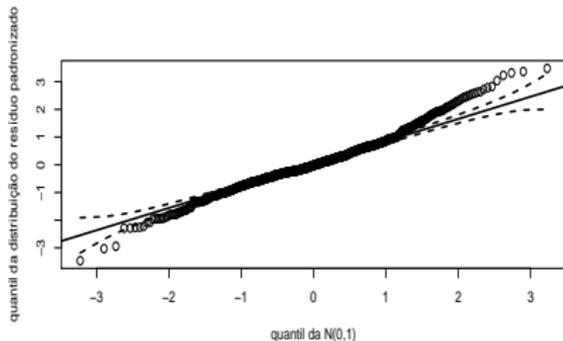
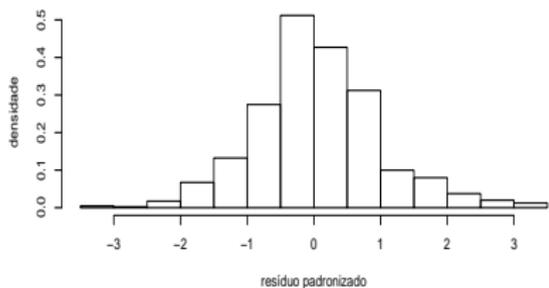
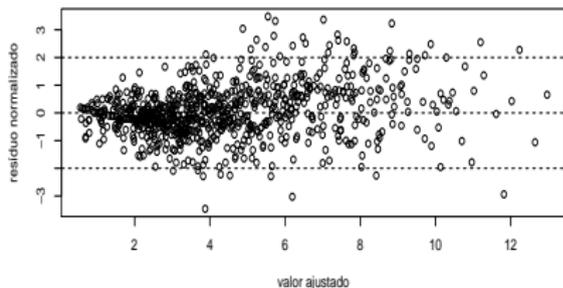
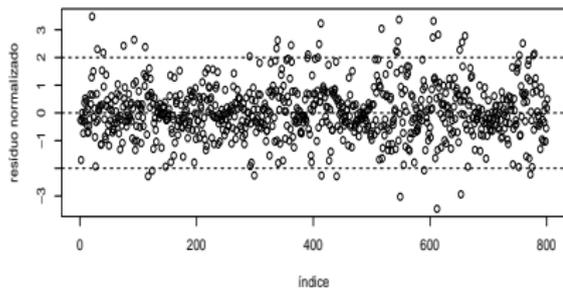
- Resíduo marginal normalizado: $\hat{\epsilon}_j^* = (\hat{\sigma} \hat{\mathbf{U}}_{(m)j})^{-1} \hat{\epsilon}_j$

em que $\hat{\mathbf{U}}_{(c)j}$ é a matriz triangular superior da decomposição de Cholesky de $\hat{\mathbf{R}}_j = \hat{\mathbf{U}}'_{(c)j} \hat{\mathbf{U}}_{(c)j}$ e $\hat{\mathbf{U}}_{(m)j}$ é a matriz triangular superior da decomposição de Cholesky de $\hat{\mathbf{V}}_j = \hat{\mathbf{U}}'_{(m)j} \hat{\mathbf{U}}_{(m)j}$.

Tipos de resíduos

- Segundo Pinheiro and Bates (2000), página 239, e Schabenberger (2004), respectivamente, $\hat{\xi}_j^*$ e $\hat{\epsilon}_j^*$ devem seguir, aproximadamente uma distribuição $N(0,1)$, no caso do modelo estar bem ajustado.
- No entanto, Nobre and Singer (2007) sugerem a utilização do resíduo de confundimento mínimo proposto por Hilden-Milton (1995), veja também Nobre (2004).

Usando os resíduos $\hat{\xi}_j^*$



Usando os resíduos $\hat{\xi}_j^*$

