

Modelos multivariados (marginais)

Prof. Caio Azevedo

Porque não usar os MNLM?

- Já vimos como analisar descritivamente alguns conjuntos de dados longitudinais.
- Vimos uma das primeiras tentativas de análise inferencial de dados longitudinais baseadas em metodologias desenvolvidas para análise de dados multivariados.
- Vimos também as limitações dessas metodologias.
- Estudaremos agora uma modelagem multivariada mas que trata cada unidade experimental individualmente.

Porque não usar os MNLM (análise multivariada)?

- Como analisar situações desbalanceadas e/ou incompletas? (os indivíduos são avaliados em diferentes condições de avaliação e/ou em quantidades diferentes delas).
- Como modelar a matriz de covariâncias apropriadamente?
- Como considerar a variabilidade intra/entre unidades experimentais de modo mais apropriado (p.e., diferentes curvas entre indivíduos)?
- Como reduzir o número de parâmetros de modo apropriado?
- Como modelar heterocedasticidade (variabilidade oriunda de outras fontes de informação)?

Formas de modelar dependência

- Modelos multivariados.
- Modelos marginais multivariados.
- Modelos de efeitos mistos.
- Modelagem de distribuições condicionais.
- Equações de estimação generalizadas.
- Cópulas.
- Modelagem livre de distribuição (“não-paramétrica”).

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)),

- 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.
- β_1 : é o incremento na concentração esperada de bilirrubina no intervalo de um dia.

Cont.

- Se considerarmos $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, 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). Por exemplo:
$$\text{Cov}(\xi_{ij}, \xi_{i'j}) = \rho, \rho \in \mathfrak{R}, \forall i, i', i \neq i' \text{ (correlação uniforme).}$$
- Vamos explorar diferentes estruturas de covariância (correlação) para os erros.

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 , $\boldsymbol{\xi}_j \stackrel{ind.}{\sim} N_{k_j}(\mathbf{0}, \boldsymbol{\Sigma}_j)$.

Voltando ao Exemplo 1

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

Cont.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{89} \end{bmatrix}; X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{89} \end{bmatrix}; \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{89} \end{bmatrix}$$

$$Y = X\beta + \xi$$

Algumas propriedades do modelo

- $\mathcal{E}(\mathbf{Y}_j) = \mathbf{X}_j\boldsymbol{\beta} = \boldsymbol{\mu}_j = (\mu_1, \dots, \mu_{k_j})^t$.
- $\text{Cov}(\mathbf{Y}_j) = \boldsymbol{\Sigma}_j$.
- Distribuição dos vetores aleatórios individuais:

$$\mathbf{Y}_j \stackrel{\text{ind.}}{\sim} N_{k_j}(\mathbf{X}_j\boldsymbol{\beta}, \boldsymbol{\Sigma}_j)$$

Estruturas para as matrizes de covariância

- Diferentes escolhas para Σ_j induzem diferentes estruturas de dependência para o vetor de respostas.
- Por exemplo, quando $\Sigma_j = \sigma^2 \mathbf{I}_{k_j}$, tem-se o modelo de regressão linear usual (homocedástico e com as observações independentes).
- Existem diversas técnicas para sugestão/escolha de matrizes de covariâncias (já vimos algumas dessas ferramentas, sob a ótica da análise descritiva).

Modelos para a estrutura de covariância

- Não estruturada (NE)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}$$

- Auto-regressiva de ordem 1 homocedástica (AR(1))

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Cont.

- Auto-regressiva com média móvel de ordem (1,1) homocedástica (ARMA(1,1))

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}$$

- Uniforme homocedástica (U)

$$\Sigma = \begin{bmatrix} \sigma^2 + \tau & \tau & \tau & \tau \\ \tau & \sigma^2 + \tau & \tau & \tau \\ \tau & \tau & \sigma^2 + \tau & \tau \\ \tau & \tau & \tau & \sigma^2 + \tau \end{bmatrix}$$

- Toeplitz homocedástica (T)

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

- Auto-regressiva de ordem 1 heterocedástica (ARH(1))

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_1\sigma_2\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_1\sigma_3\rho^2 & \sigma_2\sigma_3\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_1\sigma_4\rho^3 & \sigma_2\sigma_4\rho^2 & \sigma_3\sigma_4\rho & \sigma_4^2 \end{bmatrix}$$

- Auto-regressiva com média móvel de ordem (1,1) heterocedástica (ARMAH(1,1))

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\gamma & \sigma_1\sigma_3\gamma\rho & \sigma_1\sigma_4\gamma\rho^2 \\ \sigma_1\sigma_2\gamma & \sigma_2^2 & \sigma_2\sigma_3\gamma & \sigma_2\sigma_4\gamma\rho \\ \sigma_1\sigma_3\gamma\rho & \sigma_2\sigma_3\gamma & \sigma_3^2 & \sigma_3\sigma_4\gamma \\ \sigma_1\sigma_4\gamma\rho^2 & \sigma_2\sigma_4\gamma\rho & \sigma_3\sigma_4\gamma & \sigma_4^2 \end{bmatrix}$$

Cont.

■ Uniforme heterocedástica (UH)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_1\sigma_3\rho & \sigma_2\sigma_3\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_1\sigma_4\rho & \sigma_2\sigma_4\rho & \sigma_3\sigma_4\rho & \sigma_4^2 \end{bmatrix}$$

■ Toeplitz heterocedástica (TH)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_1\sigma_2\sigma_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ \sigma_1\sigma_3\rho_2 & \sigma_2\sigma_3\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_1\sigma_4\rho_3 & \sigma_2\sigma_4\rho_2 & \sigma_3\sigma_4\rho_1 & \sigma_4^2 \end{bmatrix}$$

- Ante-dependência de ordem 1 (AD(1))

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \sigma_1\sigma_4\rho_1\rho_2\rho_3 & \sigma_2\sigma_4\rho_2\rho_3 & \sigma_3\sigma_4\rho_3 & \sigma_4^2 \end{bmatrix}$$

Estimação

- Sob a ótica frequentista trabalha-se com o produtório das densidades de cada indivíduo, ou seja, com o produtório das distribuições multivariadas associadas à \mathbf{Y}_j .
- Também existem métodos Bayesianos.
- Suposição : $\Sigma_j = g(\theta)$, em que θ correspondem às variâncias ($\sigma_i^2, i = 1, 2, \dots, k$), em que k é quantidade de instante de avaliação ou a variância (σ^2) e aos parâmetros de correlação $(\rho, \gamma)'$.

Cont.

- Log-verossimilhança para n observações:

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

$$\boldsymbol{\Sigma}_j \equiv \boldsymbol{\Sigma}_j(\boldsymbol{\theta}).$$

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

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

Cont.

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

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

- 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:

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

Algoritmo (estimação por MV)

- A notação $\frac{\partial \Sigma_j}{\partial \theta_k}$ representa a derivada de Σ_j com relação à cada componente de θ o que resulta, para cada componente, numa matriz.
- Com as estimativas de θ , digamos $\tilde{\theta}$, obtem-se as estimativas de β , através de:

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

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

- 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, $\mathbf{Y}_j^* \sim N(\mathbf{0}_{k_j}, \mathbf{U}_j \boldsymbol{\Sigma}_j \mathbf{U}_j')$.
- 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.

Cont.

- A log-verossimilhança residual ou restrita é dada por

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

Cont.

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

$$\begin{aligned} l_R(\boldsymbol{\theta}) &= -\frac{1}{2} \ln(2\pi) \sum_{j=1}^n k_j - \frac{1}{2} \sum_{j=1}^n \ln |\boldsymbol{\Sigma}_j| \\ &\quad - \frac{1}{2} \sum_{j=1}^n \left(\mathbf{Y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}} \right)' \boldsymbol{\Sigma}_j^{-1} \left(\mathbf{Y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}} \right) \\ &\quad - \frac{1}{2} \ln \left| \sum_{j=1}^n \mathbf{X}_j' \boldsymbol{\Sigma}_j^{-1} \mathbf{X}_j \right| + \text{const.} \end{aligned} \quad (4)$$

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

- Uma vez que os estimadores de MVR de $\boldsymbol{\theta}$ forem obtidos, maximizando-se (4) (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 $\mathbf{Y}_j \stackrel{ind.}{\sim} N_{k_j}(\mathbf{X}_j\boldsymbol{\beta}, \boldsymbol{\Sigma}_j)$.
- 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:

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

Algoritmo (estimação por MVR)

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

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

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

- 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.

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.

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 “gls” 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 $\mu_{ij} = \mathcal{E}(Y_{ij}) = \mathbf{X}'_{ij}\boldsymbol{\beta}$, em que \mathbf{X}'_{ij} é i-ésima linha da matriz \mathbf{X}_j .
- A função gls trabalha com a estrutura $\text{Cov}(\mathbf{Y}_j) = \boldsymbol{\Sigma}_j = \sigma^2 \mathbf{R}_j = \sigma^2 \boldsymbol{\Lambda}_j \mathbf{C}_j \boldsymbol{\Lambda}_j$, 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

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

Cont.

- $\lambda(\cdot)$ é uma função de variância. Exemplos são dados abaixo (e_i denota um estrato i , o qual pode ser definido por covariáveis e/ou condições de avaliação e/ou grupos):

Função $\lambda(\cdot)$	$\lambda(\cdot)$	Comentário
$varPower(\mu_{ij}, \delta; \mathbf{v}_{ij})$	$ \mathbf{v}_{ijk} ^{\delta_{e_i}}; \mu_{ijk} ^{\delta_{e_i}}$	
$varExp(\mu_{ij}, \delta; \mathbf{v}_{ij})$	$\exp(\mathbf{v}_{ijk} \delta_{e_i}); \exp(\mu_{ijk} \delta_{e_i})$	
$varConstPower(\mu_{ij}, \delta; \mathbf{v}_{ij})$	$\delta_{1,e_i} + \mathbf{v}_{ijk} ^{\delta_{e_i}}; \delta_{1,e_i} + \mu_{ijk} ^{\delta_{e_i}}$	$\delta_{1,e_i} > 0, \forall i$
$varIdent(\delta; \mathbf{v}_{ij})$	δ_{e_i}	$\delta_{e_1} \equiv 1, \delta_{e_i} > 0, i \geq 2$

Cont.

- $Corr(Y_{ij}, Y_{i'j}) = h[d(\mathbf{t}_{ij}, \mathbf{t}_{i'j}), \boldsymbol{\varrho}]$, em que $d(\mathbf{t}_{ij}, \mathbf{t}_{i'j})$ é uma função de distância entre os vetores de posição \mathbf{t}_{ij} e $\mathbf{t}_{i'j}$ e $h(., .)$ é um função contínua com respeito à $\boldsymbol{\varrho}$ e $h(0, \boldsymbol{\varrho}) \equiv 1$.

Estrutura de correlação	$h(., .)$	Comentário
Serial		
corCompSymm	$h(k, \boldsymbol{\varrho}) \equiv \varrho$	$k = 1, 2, \dots; \varrho < 1$
corAR1	$h(k, \boldsymbol{\varrho}) \equiv \varrho^k$	$k = 0, 1, 2, \dots; \varrho < 1$
corCAR1	$h(s, \boldsymbol{\varrho}) \equiv \varrho^s$	$s \geq 0; \varrho > 0$
corSymm	$h(d(i, i'), \boldsymbol{\varrho}) \equiv \varrho_{ii'}$	$s \geq 0; \varrho > 0$

Mais a estrutura ARMA(p,q) que é definida como no contexto de séries temporais.

Cont.

Estrutura de correlação	$h(.,.)$	Comentário
Espacial		
corExp	$h(s, \varrho) \equiv e^{-s/\varrho}$	$s \geq 0; \varrho > 0$
corGaus	$h(s, \varrho) \equiv e^{(-s/\varrho)^2}$	$s \geq 0; \varrho > 0$
corLin	$h(s, \varrho) \equiv (1 - s/\varrho)\mathbb{1}_{(s < \varrho)}$	$s \geq 0; \varrho > 0$
corRatio	$h(s, \varrho) \equiv (1 - (s/\varrho)^2)/(1 + (s/\varrho)^2)$	$s \geq 0; \varrho > 0$
corSpher	$h(s, \varrho) \equiv (1 - 1,5(s/\varrho) + 0,5(s/\varrho)^3)\mathbb{1}_{(s < \varrho)}$	$s \geq 0; \varrho > 0$

Modelagem para os dados do Exemplo 1 (bilirrubina)

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

- (1) : $\mu_{ij} = \beta_0 + \beta_1(x_{ij} - 1)$; (2) : $\mu_{ij} = \beta_0 + \beta_1(x_{ij} - 1) + \beta_2(x_{ij} - 1)^2$;
(3) : $\mu_{ij} = \beta_0 + \beta_1(x_{ij} - 1)I_{(i \in \{1,2\})} + \beta_2(x_{ij} - 1)I_{(i \in \{3,4,5,6,7,8,9\})}$.
- (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/exponencial) (note que $x_{ij} = x_{ij'}$, $\forall i, j, j'$).
- *Corre*(Y_{ij}, Y_{i1j}) (1) AR(1), (2)(ARMA(1,1)), (3)Uniforme, (4) Não estruturada.

Modelos

Modelo	Preditor linear	Variância	Correlação
HRAR1	Reta	Homocedástico	AR(1)
HRARMA11	Reta	Homocedástico	ARMA(1,1)
HRU	Reta	Homocedástico	U
HRNE	Reta	Homocedástico	NE
HERAR1	Reta	Heterocedástico	AR(1)
HERARMA11	Reta	Heterocedástico	ARMA(1,1)
HERU	Reta	Heterocedástico	U
HENE	Reta	Heterocedástico	NE**

** Não foi possível ajustar

Modelos

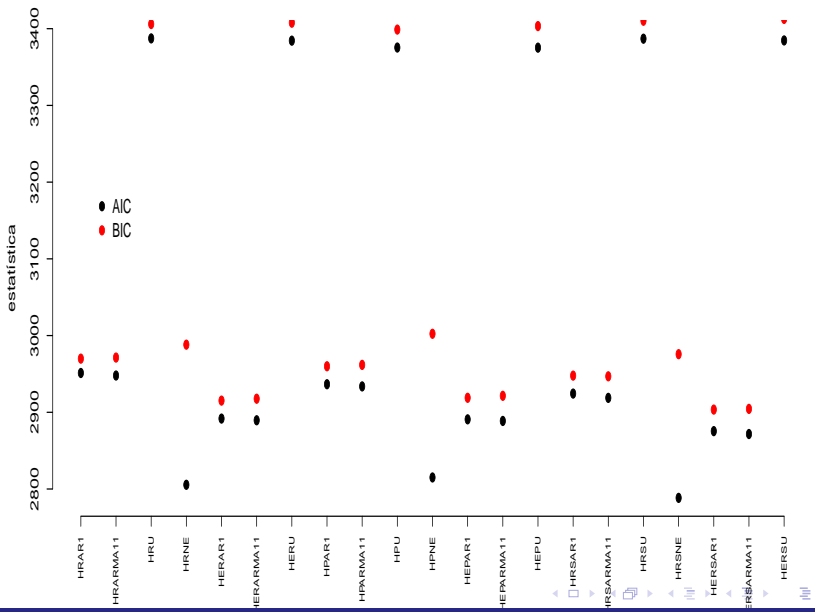
Modelo	Preditor linear	Variância	Correlação
HPAR1	Parábola	Homocedástico	AR(1)
HPARMA11	Parábola	Homocedástico	ARMA(1,1)
HPU	Parábola	Homocedástico	U
HPAR1	Parábola	Homocedástico	NE
HEPAR1	Parábola	Heterocedástico	AR(1)
HEPARMA11	Parábola	Heterocedástico	ARMA(1,1)
HEPU	Parábola	Heterocedástico	U
HEPNE	Parábola	Heterocedástico	NE**

** Não foi possível ajustar

Modelos

Modelo	Preditor linear	Variância	Correlação
HRSAR1	Regressão segmentada	Homocedástico	AR(1)
HRSARMA11	Regressão segmentada	Homocedástico	ARMA(1,1)
HRSU	Regressão segmentada	Homocedástico	U
HRSNE	Regressão segmentada	Homocedástico	NE
HERSAR1	Regressão segmentada	Heterocedástico	AR(1)
HERSARMA11	Regressão segmentada	Heterocedástico	ARMA(1,1)
HERSU	Regressão segmentada	Heterocedástico	U
HERSNE	Regressão segmentada	Heterocedástico	NE**

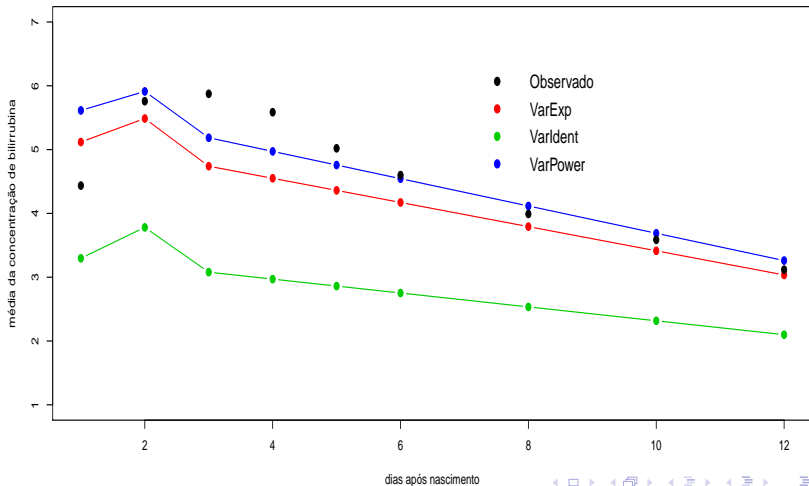
** Não foi possível ajustar



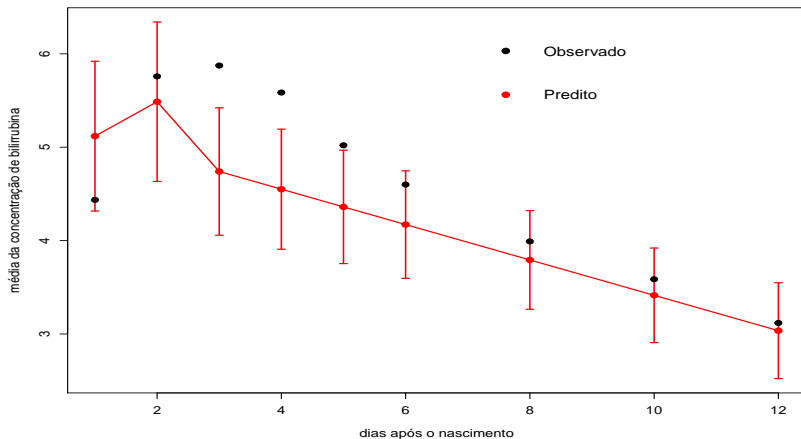
- Modelos ARMA(1,1): HER (AIC= 2889,62 ; BIC=2917,72), HEP(AIC= 2888,74, BIC=2921,51), HERS (AIC= 2871,71, BIC=2904,48).
- Modelos AR(1): HER (AIC= 2891,83; BIC=2915,24), HEP(AIC= 2890,82, BIC=2918,92), HERS (AIC= 2875,45, BIC=2903,54).
- Modelo heterocedástico com regressão segmentada (observações independentes): AIC= 4302,27; BIC = 4325, 68.

- Ajustou-se dois outros modelos heterocedásticos ARMA(1,1) com regressão segmentada: $\mathcal{V}(Y_{ij}) = \sigma^2 \delta_i^2 (\delta_1 = 1)$ (identidade) e $\mathcal{V}(Y_{ij}) = \sigma^2 |x_{ij}|^\delta$ (potência). Respectivamente AIC = 2767,91; BIC = 2833,46; AIC = 2905,23; BIC = 2948,00.
- Vamos comparar os modelos ARMA(1,1) com os três tipos de função de variância (identidade, exponencial e potência).

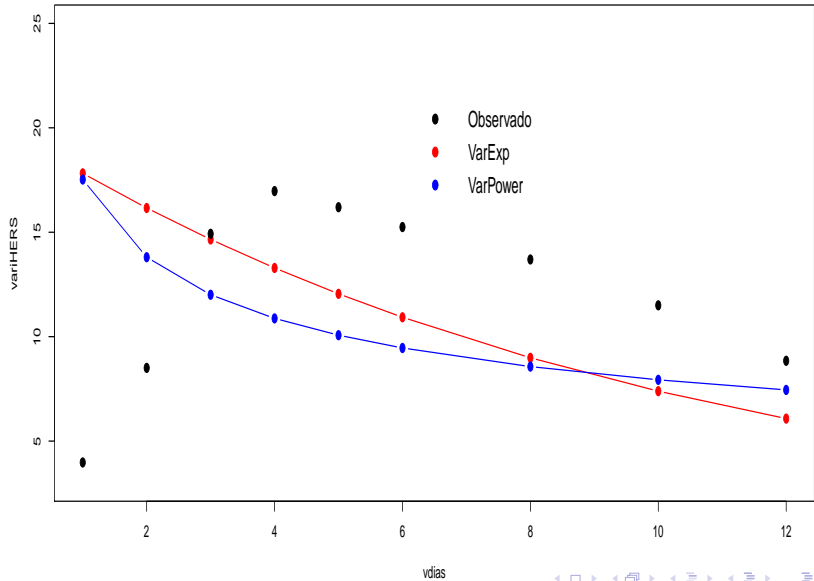
Modelo ARMA(1,1) com três tipos de função de variância

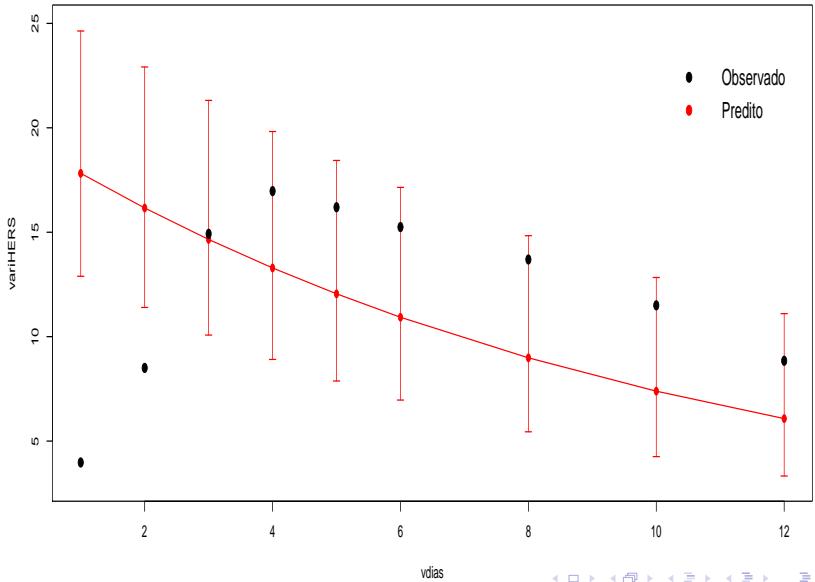


Modelo ARMA(1,1) com função de variância exponencial

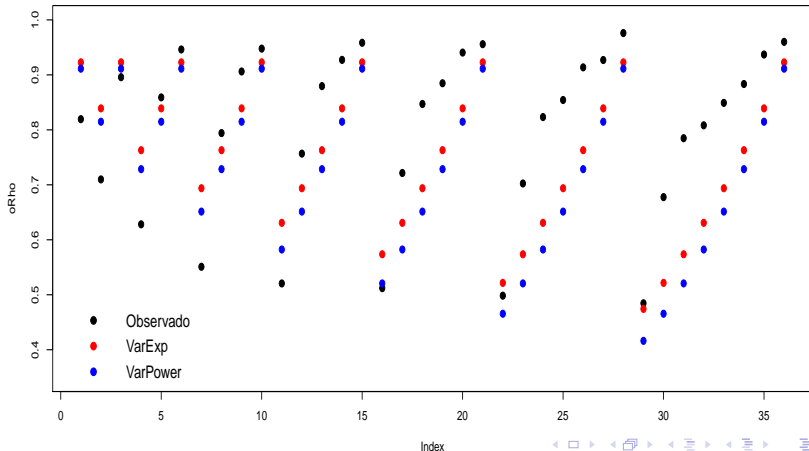


- Modelo final: ARMA(1,1), com regressão segmentada, função de variância exponencial.





Média dos desvios absolutos ($\text{VarExp} = 0,096$;
 $\text{VarPower} = 0,115$)



Estimativas dos parâmetros

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

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

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

Na parametrização $Y_{ij} = \phi_1 Y_{ij} + \theta_1 \xi_{(i-1)j} + \xi_{ij}$. A função de correlação pode ser encontrada em Pinheiro, J.; Bates, D. (2000). Mixed - Effects Models in S and S-plus (seção 5.3.1).

Testes de Hipótese

- Para implementar os testes $\mathbf{C}\beta = \mathbf{M}$, pode-se usar as estimativas dos parâmetros β (`vbeta < -fit.model$coeff`) e a estimativa da respectiva matriz de covariâncias (`fit.model$varBeta`).
- Em relação aos parâmetros (para testar $\mathbf{C}\theta = \mathbf{M}$) σ , δ (no caso de se usar a função `varIdent`) e ρ , o pacote trabalha com os chamados parâmetros irrestritos que correspondem à $\sigma^* = \ln(\sigma)$ e $\delta^* = \ln(\delta)$ e $\rho^* = \ln \frac{1+\rho}{1-\rho}$ (para as estruturas de correlação serial).
- A matriz de covariâncias obtida através de comando (`fit.model$apVar`) está relacionada aos parâmetros irrestritos.

Análise residual

- Resíduo ordinário: $R_o = Y_{ij} - \mathbf{X}'_{ij}\hat{\beta}$ ($\mathbf{R}_o = \mathbf{Y}_j - \mathbf{X}_j\hat{\beta}$)
(correlacionado e com variâncias, possivelmente, desiguais).
- Resíduo de Pearson: $R_p = \frac{R_o}{\sqrt{\widehat{\mathcal{V}}(Y_{ij})}}$ ($\mathbf{R}_p = \mathbf{R}_o ./ \sqrt{\widehat{\mathcal{V}}(\mathbf{Y}_j)}$)
(correlacionado e com variância unitária), em que
 $\widehat{\mathcal{V}}(\mathbf{Y}_j) = (\widehat{\mathcal{V}}(Y_{1j}), \dots, \widehat{\mathcal{V}}(Y_{kj}))'$ e “./” denota a divisão elemento por elemento.
- Resíduo normalizado: $\mathbf{R}_n = (\sigma \mathbf{U}_j)^{-1} \mathbf{R}_o$, em que \mathbf{U}_j é a matriz triangular superior da decomposição de Cholesky de $\mathbf{R}_j = \mathbf{U}'_j \mathbf{U}_j$
(não correlacionado e com variância unitária).

