Modelos não lineares hierárquicos de dois níveis

Prof. Caio Azevedo

Relembrando

- Modelos hierárquicos de dois níveis
  - MLH: $\mathcal{E}(\boldsymbol{Y}_j|\boldsymbol{u}_j) = \boldsymbol{X}_j'\boldsymbol{\beta}_j$ (nível 1) - $\boldsymbol{\beta}_j = \boldsymbol{W}_j\boldsymbol{\gamma} + \boldsymbol{u}_j$ (nível 2).
  - MLGH: $\mathcal{E}(\boldsymbol{Y}_{ji}|\boldsymbol{u}_j) = g^{-1}(\boldsymbol{X}_{ji}'\boldsymbol{\beta}_j)$ (nível 1) - $\boldsymbol{\beta}_j = \boldsymbol{W}_j\boldsymbol{\gamma} + \boldsymbol{u}_j$ (nível 2).
- Versão modelos mistos
  - MLH : $\mu_{ji} = \boldsymbol{Z}_j'\boldsymbol{\gamma} + \boldsymbol{X}_j'\boldsymbol{u}_j$.
  - MLGH: $\mu_{ji} = g^{-1}(\boldsymbol{Z}_j'\boldsymbol{\gamma} + \boldsymbol{X}_j'\boldsymbol{u}_j)$.

Comentários

- Nem sempre a relação entre as covariáveis ($\boldsymbol{X}_j$) e os efeitos aleatórios é adequamente descrita por um modelo linear ou por um modelo linear generalizado no nível 1 (isso também vale para o nível 2).
- Modelos não lineares (hierárquicos) (MNLH) podem ser mais apropriados em situações em que o comportamento das médias (em função de covariáveis como o tempo, por exemplo) não é satisfatoriamente modelado pelos MLH ou pelos MLGH.

Limitações dos MLH e MLGH e vantagens dos MNLH

- O comportamento linear e não linear induzidos, respectivamente, pelo MLH e MLGH podem não ser adequados para representar o comportamento da média.
- Existência de assíntotas (inferior e/ou superior).
- As interpretações dos parâmetros de modelos não lineares podem ser mais adequadas do que aquelas associadas aos parâmetros do MLN e do MLGH (particularmente, polinômios de grau ≥ 3 possuem parâmetros de difícil interpretação), principalmente em termos do problema. Isso também pode ocorrer para polinômios fracionários.

Limitações dos MLH e MLGH e vantagens dos MNLH

- Os MNLH, em geral, apresentam poucos parâmetros com interpretações úteis em termos do problema.
- Via de regra, eles costumam estar relacionados com (os) processos geradores de fenômenos de interesse (soluções de equações diferenciais).
- Garantia de que as médias preditas respeitarão o espaço paramétrico associado.
- Uma alternativas aos MNL(H) são os modelos que consideram polinômios fracionários.

Exemplo 6: cinética de indometacina, Kwan et al (1976)

- Os dados correspondem à um experimento de farmacocinética (etapas pelas quais a droga passa desde a ministração, introdução do fármaco no organismo, como tomar um comprimido, até sua eliminação, processo pela qual o fármaco deixa o organismo definitivamente) da droga indometacina (um tipo de anti-inflamatório).

Exemplo 6: cinética de indometacina, Kwan et al (1976)

- Seis indivíduos receberam de modo intravenoso a mesma dose de indometacina, e tiveram sua concentração de droga no plasma (em mcg/ml) medidas 11 vezes entre 15 minutos e 8 horas após o medicamento ser ministrado.
- Objetivo: estudar o comportamento da concentração ao longo do tempo.

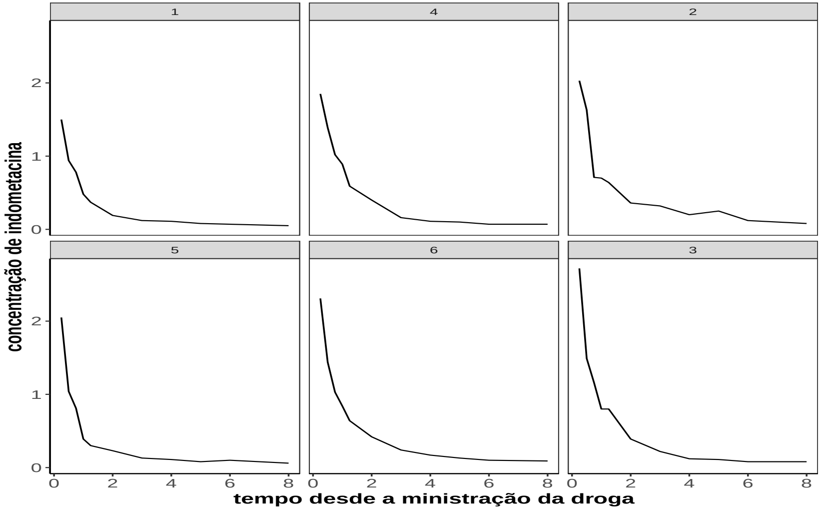
Dados: indivíduo 1

Indivíduo	Tempo	Concentração
1	0,25	1,50
1	0,50	0,94
1	0,75	0,78
1	1,00	0,48
1	1,25	0,37
1	2,00	0,19
1	3,00	0,12
1	4,00	0,11
1	5,00	0,08
1	6,00	0,07
1	8,00	0,05

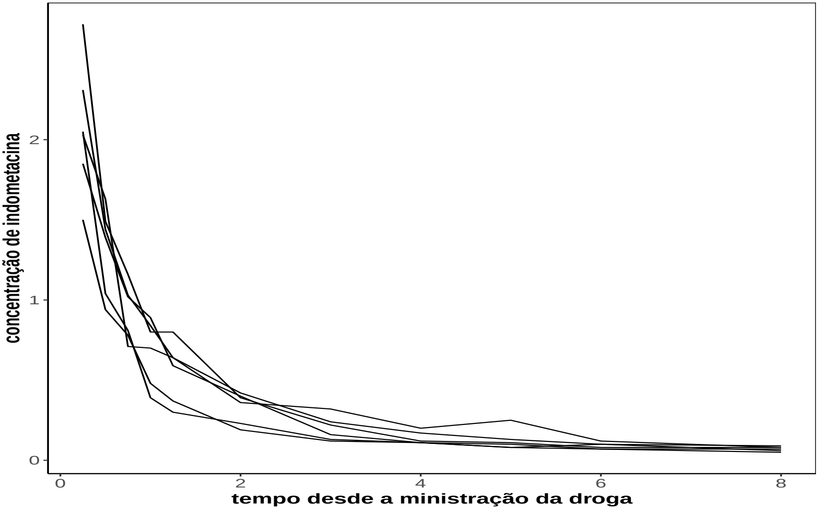
Descrição dos dados

- Temos que n $= 6$ e $n_j = 11, j = 1, 2, ..., 6$, com um total de 66 observaçoes.
- Este estudo (longitudinal) é completo (nenhuma observação foi perdida), balanceado (fo planejado para que todas as unidades experimentais fossem observadas em cada instante) e irregular (as distâncias entre os instantes não são as mesmas)

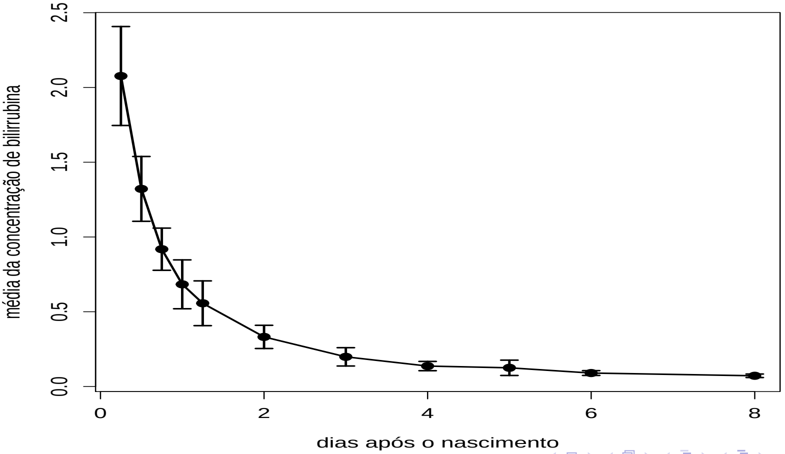
Perfis individuais (separados)



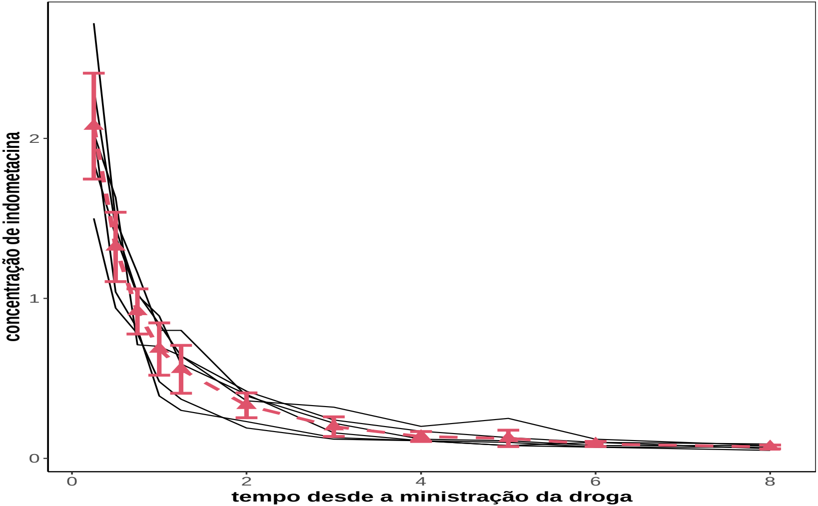
Perfis individuais (juntos)



Perfis médios



Perfis médios e individuais



Exemplo 7: crescimento de plantas de soja, Davidian and Giltinan (1995) (1976)

- Os dados correspondem à um experimento sobre crescimentos de dois tipos de soja: "Plant Introduction" # 416937 (P) um tipo experimental de cepa (parte da planta a que se cortou o caule e permanece viva no solo) e "Forrest" (F) (uma variedade comercial).
- O peso médio das folhas (em gramas) de seis plantas escolhidas aleatoriamente de cada parcela (área de cultivo) foi medida (aproximadamente) semanalmente entre duas e onze semanas depois de plantadas (embora os dias posssam ser diferentes e o espaçamento não ser o mesmo).

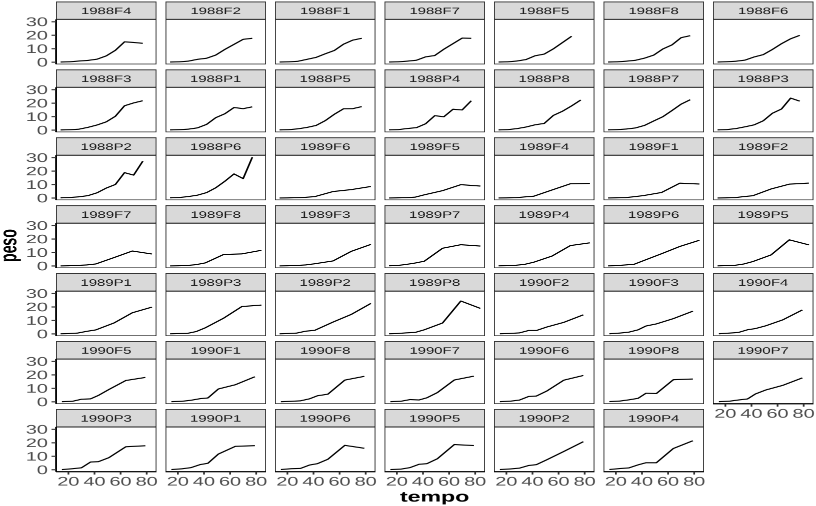
Cont.

- O estudo é (em princípio) completo, desbalanceado, em relação ao tempo, e irregular.
- $n = 48$ parcelas (16 em cada ano) com diferentes n_j.
- O experimento foi conduzido em três anos diferentes: (1988, 1989 e 1990).
- Objetivo: comparar os tipos de plantas, quanto ao crescimento, ao longo do tempo, bem como avaliar seus comportamentos ao longo do tempo.

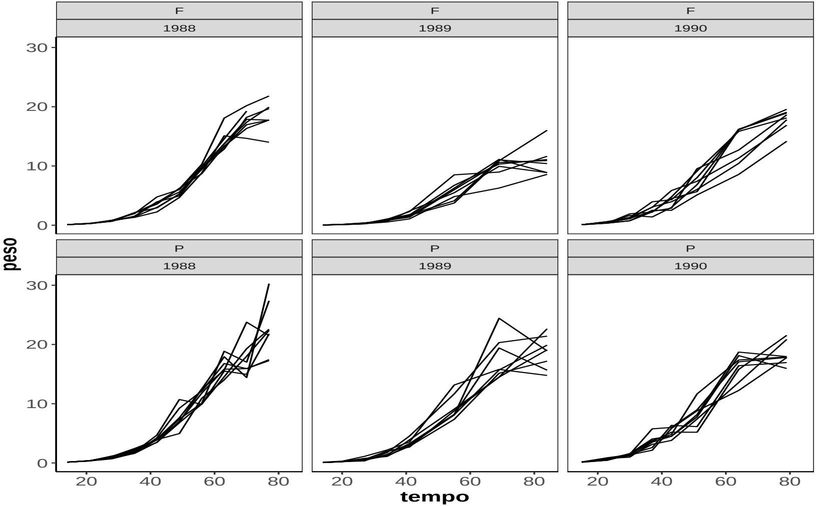
Dados da parcela 1988F1

parcela	variedade	ano	tempo (semanas)	peso
1988F1	F	1988	14	0,11
1988F1	F	1988	21	0,26
1988F1	F	1988	28	0,67
1988F1	F	1988	35	2,11
1988F1	F	1988	42	3,56
1988F1	F	1988	49	6,23
1988F1	F	1988	56	8,71
1988F1	F	1988	63	13,35
1988F1	F	1988	70	16,34
1988F1	F	1988	77	17,75

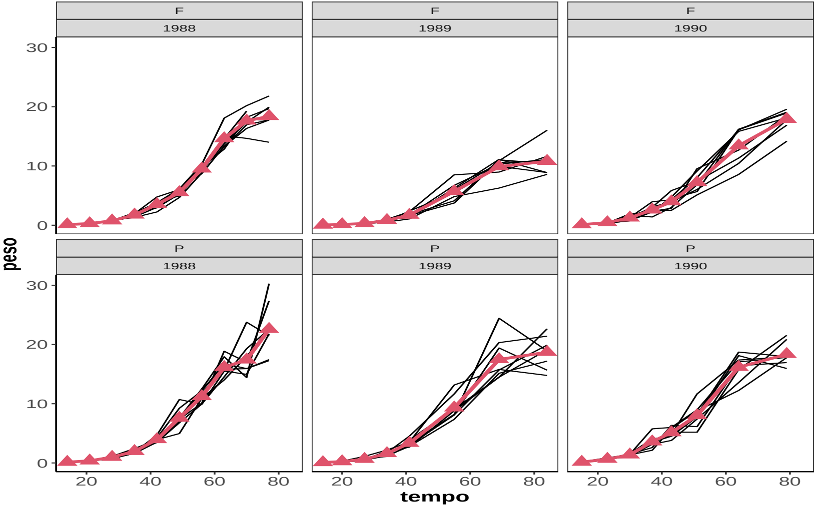
Perfis individuais por variedade x anos (separados)



Perfis individuais por variedade x anos (juntos)



Perfis individuais e médios por variedade x anos



Modelo não linear de um nível (sem efeito aleatório)

$$Y_j = f(\phi_j, \boldsymbol{X}_j, \boldsymbol{W}_j) + \xi_j \, , j = 1, ..., n \, (\text{UAE})$$

- Em que $\boldsymbol{X}_j$ representa covariáveis (matriz de planejamento) de interesse associadas aos efeitos fixos.
- $\phi_j = \boldsymbol{X}_j \boldsymbol{\beta}$.
- $\boldsymbol{W}_j$: outras covariáveis como o tempo, por exemplo.
- $f(.,.,.)$ é uma função geral, real e diferenciável e não linear em pelo menos uma componente do vetor ϕ_j.
- $\xi_j \overset{i.i.d}{\sim} N(0, \sigma^2)$.

Modelo não lineares: exemplos

- $(M_1): Y_j = \phi_1 + \phi_2 \exp\left(\phi_3/w_j\right) + \xi_j$.
- $(M_2): Y_j = \phi_1 - \phi_2 \left(w_j + \phi_3\right)^{-1} + \xi_j$.
- $(M_3): Y_j = \phi_1 w_j^{\phi_2} + \xi_j$.
- $(M_4): Y_j = \phi_1/(1 + exp(-(w_j - \phi_2)/\phi_3)) + \xi_j$.
- $(M_5): Y_j = \phi_1 + \phi_2 w_j - e^{\phi_3 + \phi_4 w_j} + \xi_j$.

Para os modelos M_1, M_2, M_4, $\boldsymbol{\phi}_j = (\phi_1, \phi_2, \phi_3)' = (\beta_1, \beta_2, \beta_3)'$, ou seja $\boldsymbol{X}_j = \boldsymbol{I}_3$. Para o modelo M_3 $\boldsymbol{\phi}_j = (\phi_1, \phi_2)' = (\beta_1, \beta_2)'$, ou seja $\boldsymbol{X}_j = \boldsymbol{I}_2$, enquanto que para o modelo M_5 $\boldsymbol{\phi}_j = (\phi_1, \phi_2, \phi_3, \phi_4)'$ $= (\beta_1, \beta_2, \beta_3, \beta_4)'$, ou seja $\boldsymbol{X}_j = \boldsymbol{I}_4$. Em todos os casos, a matriz $\boldsymbol{W}_j$ corresponde à variável w_j, que pode ser o tempo.

Cont.

- Os modelos M_1 e M_5 podem ser apresentados de uma outra forma:

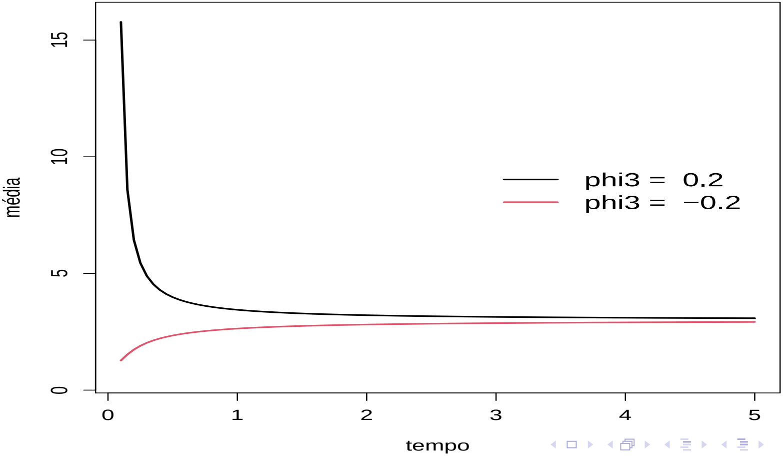
M_1

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} ; \boldsymbol{X}_j = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & w_j^{-1} \end{bmatrix}$$

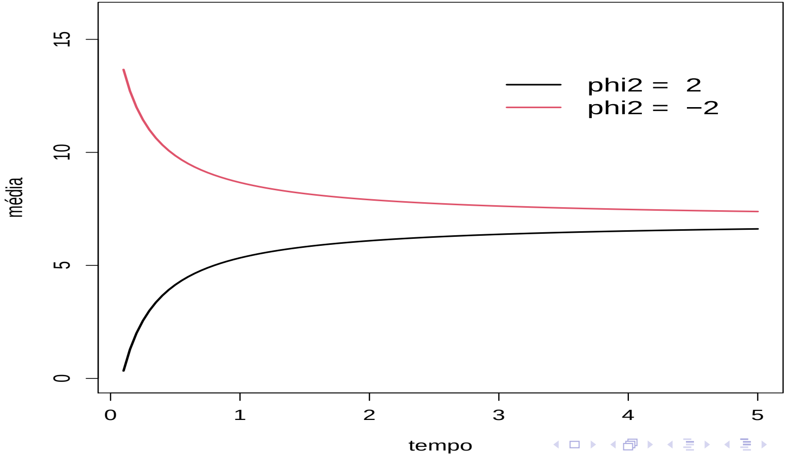
M_5

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} ; \boldsymbol{X}_j = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & w_j \end{bmatrix}$$

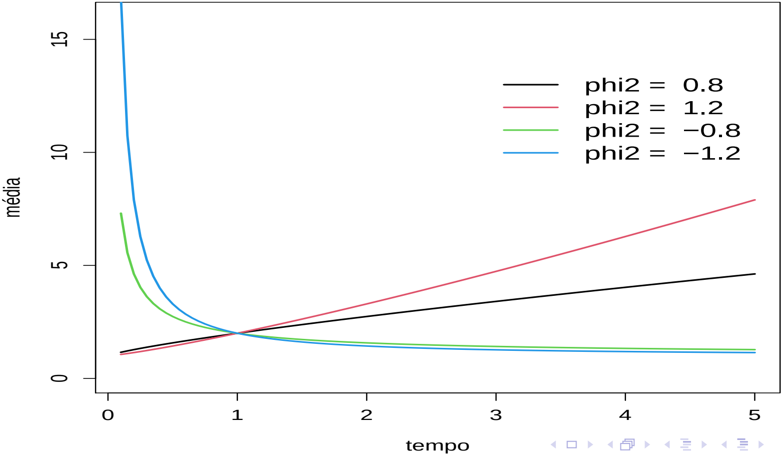
Modelo 1: $\phi_1 = 1, \phi_2 = 2$



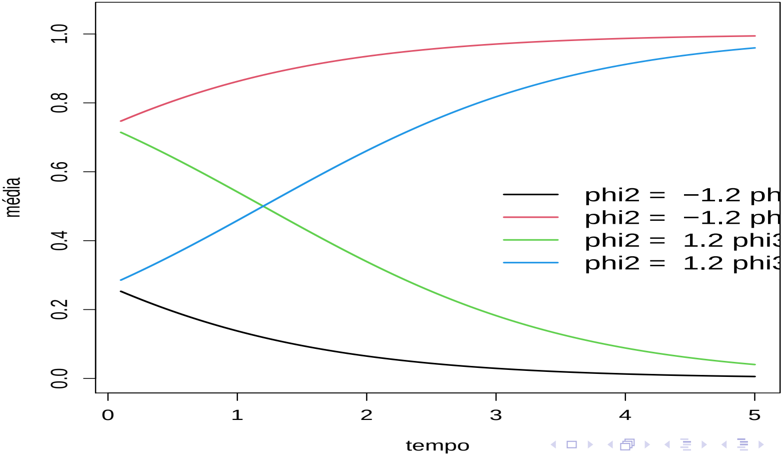
Modelo 2: $\phi_1 = 7, \phi_3 = 0,2$



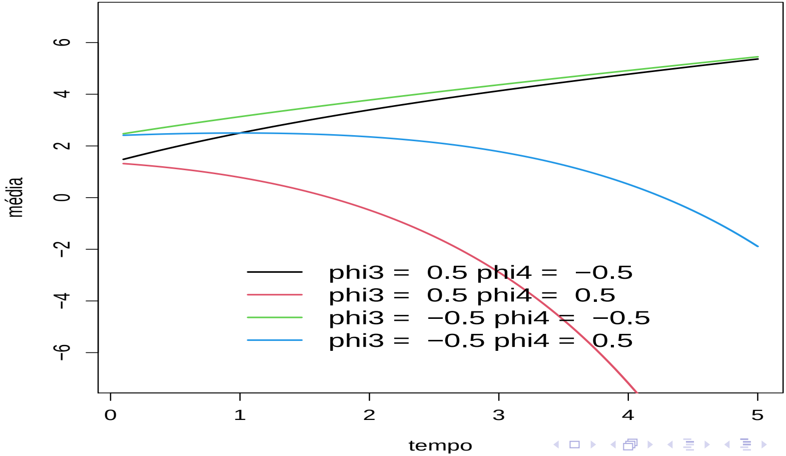
Modelo 3: $\phi_1 = 1$



Modelo 4: $\phi_1 = 1$



Modelo 5: $\phi_1 = 3, \phi_2 = 0,5$



Modelo não linear hierárquicos de dois níveis

$$
\begin{aligned}
Y_{ji} &= f(\phi_{ji}, \boldsymbol{X}_{ji}, \boldsymbol{V}_{ji}) + \xi_{ji} \, , \phi_{ji} = \boldsymbol{X}_{ji}\boldsymbol{\beta}_j \, (\text{nível 1}) \\
\boldsymbol{\beta}_j &= \boldsymbol{W}_j\gamma + \boldsymbol{u}_j \, (\text{nível 2}) \, , j = 1, ..., J, i = 1, ..., n_j
\end{aligned}
$$

- $\boldsymbol{V}_{ji}$: outras covariáveis como o tempo, por exemplo.
- $f(.,.,.)$ é uma função geral, real e diferenciável e não linear em pelo menos uma componente do vetor ϕ_{ji}.
- $\xi_{ji} \overset{i.i.d}{\sim} N(0, \sigma^2), \boldsymbol{u}_j \overset{i.i.d}{\sim} N_p(0, \boldsymbol{\Psi}), \xi_{ji} \perp \boldsymbol{u}_j \forall i, j$.

Modelo não linear hierárquicos de dois níveis (forma vetorial)

$$\boldsymbol{Y}_j = \boldsymbol{f}(\phi_j, \boldsymbol{X}_j, \boldsymbol{V}_j) + \xi_j; j = 1, ..., J$$

$$\boldsymbol{Y}_j = \begin{bmatrix} Y_{j1} \\ Y_{j2} \\ \vdots \\ Y_{jn_j} \end{bmatrix}; \boldsymbol{X}_j = \begin{bmatrix} \boldsymbol{X}_{j1} \\ \boldsymbol{X}_{j2} \\ \vdots \\ \boldsymbol{X}_{jn_j} \end{bmatrix}; \boldsymbol{V}_j = \begin{bmatrix} \boldsymbol{V}_{j1} \\ \boldsymbol{V}_{j2} \\ \vdots \\ \boldsymbol{V}_{jn_j} \end{bmatrix}$$

Modelo não linear hierárquicos de dois níveis (forma vetorial) cont.

$$\boldsymbol{\phi}_j = \begin{bmatrix} \phi_{j1} \\ \phi_{j2} \\ \vdots \\ \phi_{jn_j} \end{bmatrix}; \boldsymbol{\xi}_j = \begin{bmatrix} \xi_{j1} \\ \xi_{j2} \\ \vdots \\ \xi_{jn_j} \end{bmatrix}; \boldsymbol{f}(\boldsymbol{\phi}_j, \boldsymbol{X}_j, \boldsymbol{V}_j) = \begin{bmatrix} f(\boldsymbol{\phi}_{j1}, \boldsymbol{X}_{j2}, \boldsymbol{V}_{j1}) \\ f(\boldsymbol{\phi}_{j2}, \boldsymbol{X}_{j2}, \boldsymbol{V}_{j2}) \\ \vdots \\ f(\boldsymbol{\phi}_{jn_j}, \boldsymbol{X}_{jn_j}, \boldsymbol{V}_{jn_j}) \end{bmatrix};$$

$$\boldsymbol{V}_j = \begin{bmatrix} \boldsymbol{V}_{j1} \\ \boldsymbol{V}_{j2} \\ \vdots \\ \boldsymbol{V}_{jn_j} \end{bmatrix}$$

Modelos não lineares hieárquicos (de dois níveis): exemplos

- $(M_1) : Y_{ji} = \phi_{j1} + \phi_{j2} \exp\left(\phi_{j3}/v_{ji}\right) + \xi_{ji}$.
- $(M_2) : Y_{ji} = \phi_{j1} - \phi_{j2}\left(v_{ji} + \phi_{j3}\right)^{-1} + \xi_{ji}$.
- $(M_3) : Y_{ji} = \phi_{j1} v_{ji}^{\phi_{j2}} + \xi_{ji}$.
- $(M_4) : Y_{ji} = \phi_{j1}/(1 + exp(-(v_{ji} - \phi_{j2})/\phi_{j3})) + \xi_{ji}$.
- $(M_5) : Y_{ji} = \phi_{j1} + \phi_{j2} v_{ji} - e^{\phi_{j3} + \phi_{j4} v_{ji}} + \xi_{ji}$.

Modelo 1.1 hierárquico

- $Y_{ji} = \phi_{j1} + \phi_{j2} \exp\left(\phi_{j3}/v_{ji}\right) + \xi_{ji}$.

  $\phi_{j1} = \beta_{1j}, \beta_{1j} = \gamma_{10} + u_{1j}$, $\phi_{2j} = \beta_{2j} = \gamma_{20}$ e $\phi_{j3} = \beta_{3j} = \gamma_{30}$, nesse caso, $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j}, \beta_{3j})'$, $\boldsymbol{u}_j = u_{1j} \overset{i.i.d.}{\sim} N(0, \psi)$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \end{bmatrix} + \begin{bmatrix} u_{1j} \\ 0 \\ 0 \end{bmatrix}$$

Modelo 1.2 hierárquico

- $Y_{ji} = \phi_{j1} + \phi_{j2} \exp\left(\phi_{j3}/v_{ji}\right) + \xi_{ji}$.

  $\phi_{j1} = \beta_{1j}, \beta_{1j} = \gamma_{10} + u_{1j}$, $\phi_{j2} = \beta_{2j}, \beta_{2j} = \gamma_{20} + u_{2j}$ e

  $\phi_{j3} = \beta_{3j} = \gamma_{30}$, nesse caso, $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j}, \beta_{3j})'$,

  $\boldsymbol{u}_j = (u_{1j}, u_{2j})' \overset{i.i.d.}{\sim} N_2(0, \boldsymbol{\Psi})$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \end{bmatrix} + \begin{bmatrix} u_{1j} \\ u_{2j} \\ 0 \end{bmatrix}$$

Modelo 1.3 hierárquico

- $Y_{ji} = \phi_{j1} + \phi_{j2} \exp\left(\phi_{j3}/v_{ji}\right) + \xi_{ji}$.

  $\phi_{j1} = \beta_{1j}, \beta_{1j} = \gamma_{10} + u_{1j}$, $\phi_{j2} = \beta_{2j}, \beta_{2j} = \gamma_{20} + u_{2j}$ e

  $\phi_{3j} = \beta_{3j}, \beta_{3} = \gamma_{30} + u_{3j}$, nesse caso, $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j}, \beta_{3j})'$,

  $\boldsymbol{u}_j = (u_{1j}, u_{2j}, u_{3j})' \overset{i.i.d}{\sim} N_3(0, \boldsymbol{\Psi})$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_{1j} \\ u_{2j} \\ u_{3j} \end{bmatrix}$$

  Em todos os casos, a matriz $\boldsymbol{V}_{ji}$ corresponde à variável V_{ji}, que pode ser o tempo.

Modelo 3.1 hierárquico

- $Y_{ji} = \phi_{j1} v_{ji}^{\phi_{j2}} + \xi_{ji}$.

  $\phi_{j1} = \beta_{1j}, \beta_{1j} = \gamma_{10} + u_{1j}$ e $\phi_{2j} = \beta_{2j} = \gamma_{20}$, nesse caso,

  $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j})'$, $\boldsymbol{u}_j = u_{1j} \overset{i.i.d}{\sim} N(0, \psi)$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \end{bmatrix} + \begin{bmatrix} u_{1j} \\ 0 \end{bmatrix}$$

Modelo 3.2 hierárquico

- $Y_{ji} = \phi_{j1} v_{ji}^{\phi_{j2}} + \xi_{ji}$.

  $\phi_{j1} = \beta_{1j}, \beta_{1j} = \gamma_{10} + u_{1j}$ e $\phi_{j2} = \beta_{2j}, \beta_{2j} = \gamma_{20} + u_{2j}$, nesse caso,

  $\boldsymbol{\beta}_j = (\beta_{1j}, \beta_{2j})'$, $\boldsymbol{u}_j = (u_{1j}, u_{2j}) \overset{i.i.d}{\sim} N_2(0, \boldsymbol{\Psi})$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \end{bmatrix} + \begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix}$$

Estimação

- Aqui também utilizarmos a versão de modelos mistos de tais modelos (exercício).
- Parâmetros para estimar $(\gamma^t, \boldsymbol{u}^t, \boldsymbol{\theta}^t, \sigma^2)^t$, em que $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta})$.
- Verossimilhança completa:

$$
\begin{aligned}
L(\gamma, \boldsymbol{u}, \boldsymbol{\theta}, \sigma^2) & \propto & \exp\left\{-\sum_{j=1}^{J}\sum_{i=1}^{n_j}\frac{(y_{ji} - f(\phi_{ji}, \boldsymbol{X}_{ji}, \boldsymbol{V}_{ji}))^2}{2\sigma^2}\right\}\sigma^{-\sum_{j=1}^{n} n_j/2} \\
& \times & exp\left\{-0,5\sum_{j=1}^{J}\boldsymbol{u}_j'\boldsymbol{\Psi}^{-1}\boldsymbol{u}_j\right\}|\boldsymbol{\Psi}|^{-J/2}
\end{aligned}
$$

Verossimilhança marginal

$$
\begin{aligned}
L(\gamma, \boldsymbol{\theta}, \sigma^2) & \propto \sigma^{-\sum_{j=1}^{J} n_j/2} \prod_{j=1}^{J} \int_{\Re^p} \exp\left\{ -\sum_{i=1}^{n_j} \frac{(y_{ji} - f(\phi_{ji}, \boldsymbol{X}_{ji}, \boldsymbol{V}_{ji}))^2}{2\sigma^2} \right\} \\
& \times \; exp\left\{ -0,5\boldsymbol{u}_j' \boldsymbol{\Psi}^{-1} \boldsymbol{u}_j \right\} |\boldsymbol{\Psi}|^{-1/2} d\boldsymbol{u}
\end{aligned}
$$

A integral acima não tem solução explícita (lembrando que ϕ_{ji} é função de $\boldsymbol{u}_j$).

Resolução da integral

- Quadratura (gaussiana).
- Quadratura adaptativa.
- Aproximação da verossimilhança por um modelo linear misto.
- Aproximação de Laplace.
- Veja aqui, aqui, aqui, aqui e aqui.

Resolução da integral

- Aproximação do integrando usando o primeiro termo da expansão em séries de Taylor em torno do **valor esperado condicional** dos efeitos aleatórios (aqui).
- Aproximação do integrando usando o primeiro termo da expansão em séries de Taylor em torno da **moda condicional** dos efeitos aleatórios (aqui).
- Monte carlo.
- Algoritmos: EM, SEM, SAEM, MCEM.

Discutiremos brevemente

- Quadratura adaptativa.
- Aproximação da verossimilhança por um modelo linear misto (método de Lindstrom and Bates).
- Aproximação de Laplace.

Verossimilhança

- Vamos utilizar a seguinte decomposição da matriz $\boldsymbol{\Psi}^{-1} = \sigma^{-2}\boldsymbol{\Delta}'\boldsymbol{\Delta}$, em que $\boldsymbol{\Delta}$ (associado à decomposição de Cholesky) é uma matriz de precisão ($N = \sum_{j=1}^{n} n_j$). Assim:

$$
\begin{aligned}
L(\gamma, \boldsymbol{\theta}, \sigma^2) &\propto (\sigma^2)^{-(N+np)/2} \\
&\times \prod_{j=1}^{J} \int_{\Re^p} \exp\left\{ -\sum_{i=1}^{n_j} \frac{(y_{ji} - f(\phi_{ji}, \boldsymbol{X}_{ji}, \boldsymbol{V}_{ji}))^2 + ||\boldsymbol{\Delta}\boldsymbol{u}_j||^2}{2\sigma^2} \right\} \\
&\times |\boldsymbol{\Delta}| d\boldsymbol{u}_j \qquad (1)
\end{aligned}
$$

em que $||.||^2$ é a norma Euclidiana elevada ao quadrado.

Método de Lindstrom and Bates

- Tal método consiste em iterar entre dois passos: um passo de mínimos quadrados não lineares penalizado (MQNLP) e um passo de modelos lineares hierárquicos (MLH).
- No passo MQNLP, consideramos uma estimativa provisória de $\boldsymbol{\Delta}$ e obtemos estimativas provisórias de $\boldsymbol{u}_j$ e γ minimizando

$$\sum_{j=1}^{J}\left[\sum_{i=1}^{n_j}(y_{ji} - f(\phi_{ji}, \boldsymbol{X}_{ij}, \boldsymbol{V}_{ji}))^2 + ||\boldsymbol{\Delta}\boldsymbol{u}_j||^2\right]$$

Cont.

- No passo MLM a matriz $\boldsymbol{\Delta}$ é atualizada baseada na expansão de primeira ordem em séries de Taylor da função $f(.,.,.)$ em torno de estimativas provisórias de γ e $\boldsymbol{u}_j$, as quais serão denotadas por $\widetilde{\gamma}^{(w)}$ e $\widetilde{\boldsymbol{u}}_j^{(w)}$. Defina ainda

$$\widetilde{\boldsymbol{w}}_j^{(w)} = \boldsymbol{y}_j - \boldsymbol{f}_j(\widetilde{\gamma}^{(w)}, \widetilde{\boldsymbol{u}}_j^{(w)}) + \widetilde{\boldsymbol{Z}}_j^{(w)}\widetilde{\gamma}^{(w)} + \widetilde{\boldsymbol{X}}_j^{(w)}\widetilde{\boldsymbol{u}}_j^{(w)}$$

em que $\widetilde{\boldsymbol{Z}}_j^{(w)} = \left.\frac{\partial \boldsymbol{f}_j}{\partial \gamma'}\right|_{\widetilde{\gamma}^{(w)}, \widetilde{\boldsymbol{u}}_j^{(w)}}$ e $\widetilde{\boldsymbol{X}}_j^{(w)} = \left.\frac{\partial \boldsymbol{f}_j}{\partial \boldsymbol{u}_j'}\right|_{\widetilde{\gamma}^{(w)}, \widetilde{\boldsymbol{u}}_j^{(w)}}$

Cont.

- Assim a log-verossimilhança aproximada (para estimar $\boldsymbol{\Delta}$) é dada por:

$$
\begin{aligned}
l_{MLH}(\gamma, \boldsymbol{\Delta}, \sigma^2) &= -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{j=1}^{J}\Bigg\{\ln|\boldsymbol{\Sigma}_j(\boldsymbol{\Delta})| \\
&+ \sigma^{-2}\left[\boldsymbol{w}_j^{(w)} - \widetilde{\boldsymbol{Z}}_j^{(w)}\widetilde{\gamma}^{(w)}\right]'\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\Delta}) \\
&\times \left[\boldsymbol{w}_j^{(w)} - \widetilde{\boldsymbol{Z}}_j^{(w)}\widetilde{\gamma}^{(w)}\right]\Bigg\} \qquad (2)
\end{aligned}
$$

em que $\boldsymbol{\Sigma}_j(\boldsymbol{\Delta}) = \boldsymbol{I} + \widetilde{\boldsymbol{X}}_j^{(w)}\boldsymbol{\Delta}^{-1}(\boldsymbol{\Delta}^{-1})'\widetilde{\boldsymbol{X}}_j^{(w)'}$.

Cont.

- De modo semelhante ao que é feito nos MLH, podemos obter os valores que maximizam a função acima, em relação à γ e σ^2, explicitamente em função de $\boldsymbol{\Delta}$ e trabalhar com a verossimilhança perfilada de $\boldsymbol{\Delta}$ para estimá-la.

Cont.

- Pode-se também trabalhar com a logverossimilhança restrita (veja slides sobre MLH), ou seja:

$$l_R(\sigma^2, \boldsymbol{\Delta}) = l(\widetilde{\gamma}(\boldsymbol{\Delta}), \boldsymbol{\Delta}, \sigma^2) - \frac{1}{2}\sum_{j=1}^{J} \ln|\sigma^{-2}\widetilde{\boldsymbol{Z}}_j^{(w)'}\boldsymbol{\Sigma}_j(\boldsymbol{\Delta})\widetilde{\boldsymbol{Z}}_j^{(w)}| \quad (3)$$

em que $l(\widetilde{\gamma}(\boldsymbol{\Delta}), \boldsymbol{\Delta}, \sigma^2)$ é dada por (2), com γ substituido por $\widetilde{\gamma}$.

- O algoritmo alterna entre os passos MQNLP e MLM, até que algum critério de convergência seja alcançado.
- Para teoria assintótica, IC e Testes de hipótese, veja os respectivos slides de MLH e MLGH.

Aproximação de Laplace

- A integral que queremos é dada por:

$$
\begin{aligned}
L(\gamma, \boldsymbol{\theta}, \sigma^2) \quad &\propto \quad \prod_{j=1}^{J} \int_{\Re^p} \exp\left\{ -\sum_{i=1}^{n_j} \frac{(y_{ji} - f(\phi_{ji}, \boldsymbol{X}_{ji}, \boldsymbol{V}_{ji}))^2 + ||\boldsymbol{\Delta} \boldsymbol{u}_j||^2}{2\sigma^2} \right\} \\
&\times \quad (2\pi\sigma^2)^{-(n_j+1)/2} |\boldsymbol{\Delta}| d\boldsymbol{u}
\end{aligned}
$$

- Seja

$$
g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j) = \sum_{i=1}^{n_j} (y_{ji} - f(\phi_{ji}, \boldsymbol{X}_{ji}, \boldsymbol{V}_{ji}))^2 + ||\boldsymbol{\Delta} \boldsymbol{u}_j||^2. \tag{4}
$$

Aproximação de Laplace

- Defina:

$$
\begin{aligned}
\widetilde{\boldsymbol{u}}_j &= argmin_{\boldsymbol{u}_j} g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j) \\
g'(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j) &= \frac{\partial g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j)}{\partial \boldsymbol{u}_j} \\
g''(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j) &= \frac{\partial^2 g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j)}{\partial \boldsymbol{u}_j \partial \boldsymbol{u}_j'}
\end{aligned}
$$

- Considere a expansão em séries de Taylor de segunda ordem de g em torno de $\widetilde{\boldsymbol{u}}_j$:

$$
g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j) \approx g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j) + \frac{1}{2} (\boldsymbol{u}_j - \widetilde{\boldsymbol{u}}_j)' \, g''(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j) \, (\boldsymbol{u}_j - \widetilde{\boldsymbol{u}}_j)
$$

(note que $g'(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j) = 0$)

Cont.

- A aproximação de Laplace da verossimilhança é dada por:

$$
\begin{aligned}
L(\gamma, \boldsymbol{\theta}, \sigma^2) &\approx (2\pi\sigma^2)^{-N/2} |\boldsymbol{\Delta}|^J \\
&\times \; exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^{J} g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j) \right] \prod_{i=1}^{J} |g''(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j)|^{-1/2}
\end{aligned}
$$

- Temos que (veja Pinheiro and Bates (2013), pag. 316-317)

$$
g''(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j) \approx \boldsymbol{G}(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j) = \left. \frac{\partial \boldsymbol{f}_j(\gamma, \boldsymbol{u}_j)}{\partial \boldsymbol{u}_j} \right|_{\widetilde{\boldsymbol{u}}_j} \left. \frac{\partial \boldsymbol{f}_j(\gamma, \boldsymbol{u}_j)}{\partial \boldsymbol{b}_j'} \right|_{\widetilde{\boldsymbol{u}}_j} + \boldsymbol{\Delta}'\boldsymbol{\Delta}
$$

- A aproximação de Laplace da logverossimilhança modificada é dada por:

$$
\begin{aligned}
l_{LA}(\gamma, \sigma^2, \boldsymbol{\Delta}) &= -\frac{N}{2} + J \ln |\boldsymbol{\Delta}| \\
&- \frac{1}{2}\left\{\sum_{j=1}^{J} \ln |G(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j)| + \sigma^{-2} \sum_{j=1}^{J} g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j)\right\}
\end{aligned} \tag{5}
$$

- O procedimento itera entre a maximização de l_{LA} (Equação 5) em relação à $\gamma, \sigma^2, \boldsymbol{\Delta}$ e da minimização de $g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j)$ (Equação 4) com relação à $\boldsymbol{u}_j$.

Quadratura adaptativa

- A idéia é utilizar parte dos resultados da AL e substituir a integral de uma normal multivariada por sucessivas integrais de normais padrão independentes.
- Sejam $z_j, w_j, j = 1, 2, ..., N_Q$, respectivamente, as abscissas e os pesos para a integração por quadratura Gaussiana (unidimensional) baseado na N(0,1).
- Do resultado da AL temos que o integrando $exp(-g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j)/(2\sigma^2))$ pode ser aproximado por uma distribuição $N_q(\widetilde{\boldsymbol{u}}_j, \sigma^2 G^{-1}(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j))$

Quadratura adaptativa

- Assim

$$
\begin{aligned}
& \int exp(-g(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \boldsymbol{u}_j)/(2\sigma^2))d\boldsymbol{u}_j = \int \sigma^p |\boldsymbol{G}(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j)|^{-1/2} \\
\times \quad & \exp\left\{-g\left[\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j + \sigma \boldsymbol{G}^{-1/2}(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j)\boldsymbol{z}\right]/(2\sigma^2) + ||\boldsymbol{z}||^2/2\right\} \\
\times \quad & \exp(-||\boldsymbol{z}||^2/2)d\boldsymbol{z} \\
\approx \quad & \sigma^q |\boldsymbol{G}(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j)|^{-1/2} \\
\times \quad & \sum_{j_1=1}^{N_Q} \dots \sum_{j_q=1}^{N_Q} \exp\left\{-g\left[\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j, \widetilde{\boldsymbol{u}}_j + \sigma \boldsymbol{G}^{-1/2}(\gamma, \boldsymbol{\Delta}, \boldsymbol{y}_j)\boldsymbol{z}_j\right]/(2\sigma^2)\right. \\
+ \quad & \left. ||\boldsymbol{z}_j||^2/2\right\} \prod_{k=1}^{p} w_{jk}
\end{aligned}
$$

em que $\boldsymbol{z}_j = (z_{j1}, ..., z_{jq})'$.

Cont.

- Aproximação da logverossimilhança por quadratura adaptativa:

$$
\begin{aligned}
l_{AQA}(\gamma,\sigma^2,\boldsymbol{\Delta}) &= -\frac{N}{2}\ln(2\pi\sigma^2) + n\ln|\boldsymbol{\Delta}| - \frac{1}{2}\sum_{j=1}^{J}\ln|\boldsymbol{G}(\gamma,\boldsymbol{\Delta},\boldsymbol{y}_j)| \\
&= \sum_{j=1}^{J}\left(\sum_{j_1=1}^{N_Q}\ldots\sum_{j_q=1}^{N_Q}\exp\left\{-g\left[\gamma,\boldsymbol{\Delta},\boldsymbol{y}_j,\widetilde{\boldsymbol{u}}_j\right.\right.\right. \\
&+ \left.\sigma\boldsymbol{G}^{-1/2}(\gamma,\boldsymbol{\Delta},\boldsymbol{y}_j)\boldsymbol{z}_j\right]/(2\sigma^2) \\
&+ \left.\left.||\boldsymbol{z}_j||^2/2\right\}\prod_{k=1}^{q}w_{jk}\right)
\end{aligned}
$$

- A logverossimilhança acima pode ser maximizada através (em relação a $\gamma,\sigma^2,\boldsymbol{\Delta}$) de algoritmos de otimização adequados.

Resíduos normalizados

- São dados por:

$$\boldsymbol{R}_j = \boldsymbol{D}_j \left(\frac{\boldsymbol{Y}_j - \boldsymbol{f}(\widehat{\phi}_j, \boldsymbol{X}_j, \boldsymbol{V}_j)}{\sqrt{\widehat{\sigma}^2}} \right), j = 1, 2, ..., J,$$

- em que $\boldsymbol{D}_j$ corresponde ao inverso da matriz da decomposição de Cholesky da matriz de variâncias e covariâncias induzida pelo modelo (depende da estrutura dos efeitos aleatórios).

Recursos computacionais

- As funções "nls" (default no R) e "nlme" (pacote nlme) ajustam, respectivamente, modelos não lineares e modelos não lineares hierárquicos.
- A primeira estima os parâmetros via mínimos quadrados (ponderados) não lineares, ou seja, através da minimização de $\sum_{j=1}^{J}(Y_j - f(\phi_j, \boldsymbol{X}_j, \boldsymbol{W}_j))^2$, em relação à γ.
- Estimador para o σ^2 (função nls) : $\widehat{\sigma}^2 = \frac{1}{J-p}\sum_{j=1}^{J}\left(Y_j - f(\widehat{\phi}_j, \boldsymbol{X}_j, \boldsymbol{W}_j)\right)^2$.

Recursos computacionais

- A segunda o faz usando máxima verossimilhança (Equação 2) ou máxima verossimilhança restrita (Equação 3) (método de Lindstrom and Bates, veja aqui).
- Um ponto interessante é que ambas solicitam que o usuário forneçam estimativas iniciais para os parâmetros do preditor não linear.

Cont.

- Há uma certa flexibilidade na escolha do preditor não linear.
- Ele pode ser escrito diretamente na função ou ser inserido como uma função.
- Em geral é melhor entrar com o preditor não linear como uma função na qual conste a respectiva derivada (auxilia na convergência dos algoritmos).
- Há a possibilidade de criar uma outra função através da função "selfStart" na qual é possível inserir o cálculo automático das estimativas iniciais.

Cont.

- Há vários preditores não lineares implementados nesse pacote (derivada e estimativas iniciais já implementadas). Abaixo: $input, x$: covariável(is):

Nome	Modelo
SSasymp	$Asym + (R0 - Asym) * exp(-exp(lrc) * input)$
SSasympOff	$Asym * (1 - exp(-exp(lrc) * (input - c0)))$
SSasympOrig	$Asym * (1 - exp(-exp(lrc) * input))$
SSbiexp	$A1 * exp(-exp(lrc1) * input) + A2 * exp(-exp(lrc2) * input)$
SSfol	$Dose * exp(lKe + lKa - lCl) * (exp(-exp(lKe) * input)$
	$-exp(-exp(lKa) * input))(exp(lKa) - exp(lKe))$

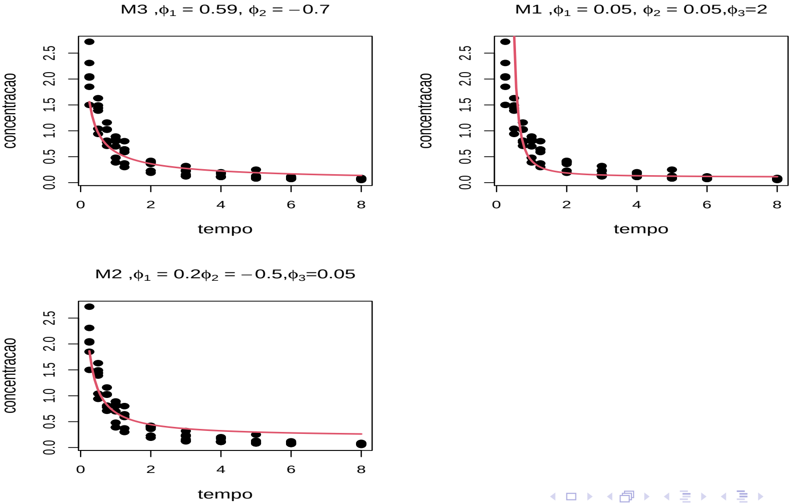
Cont.

Nome	Modelo
SSfpl	$A + (B - A)/(1 + exp((xmid - input)/scal))$
SSgompertz	$Asym * exp(-b2 * b3^{x})$
SSlogis	$Asym/(1 + exp((xmid - input)/scal))$
SSmicmen	$Vm * input/(K + input)$
SSweibull	$Asym - Drop * exp(-exp(lrc) * x^{pwr})$

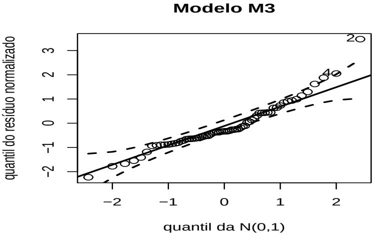
Voltando ao Exemplo 6

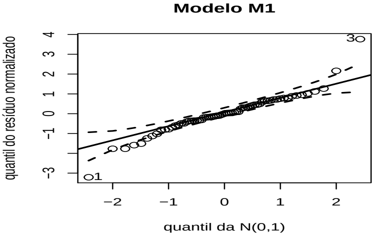
- Ajustou-se os modelos M_1, M_2 e M_3 (não houve convergência para os modelo M_4 e M_5)
- $(M_1) : Y_{ji} = \phi_{j1} + \phi_2 \exp\left(\phi_3 / v_{ji}\right) + \xi_{ji}$.
- $(M_2) : Y_{ji} = \phi_{j1} - \phi_2 \left(v_{ji} + \phi_3\right)^{-1} + \xi_{ji}$.
- $(M_3) : Y_{ji} = \phi_{j1} v_{ji}^{\phi_2} + \xi_{ji}$.
- Em todos os três modelos: $\phi_{j1} = \gamma_{10} + u_{1j}$, $u_{1j} \overset{i.i.d.}{\sim} N(0, \psi)$, $\phi_2 = \gamma_{20}$, $\phi_3 = \gamma_{30}$.
- $\xi_{ji} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $\xi_{ji} \perp u_{1j}$, $\forall i, j$.
- Em todos os modelos, $i = 1, 2, .., 11$ (condição de avaliação-nível 1), $j = 1, 2, ..., 6$ (indivíduo - nível 2).

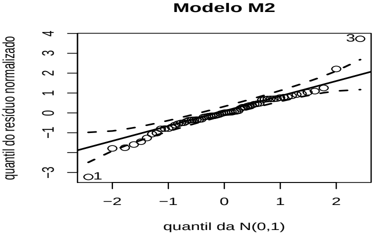
Sobre os valores iniciais do processo iterativo



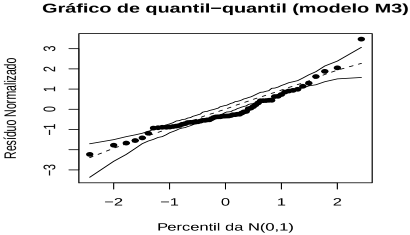
Resíduos normalizados com envelopes baseados na N(0,1)

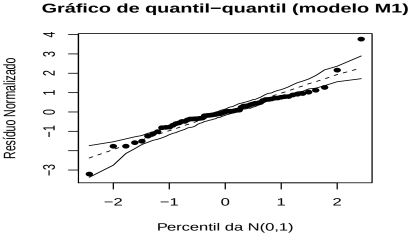


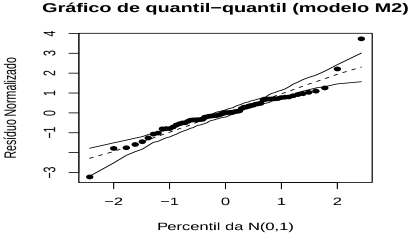




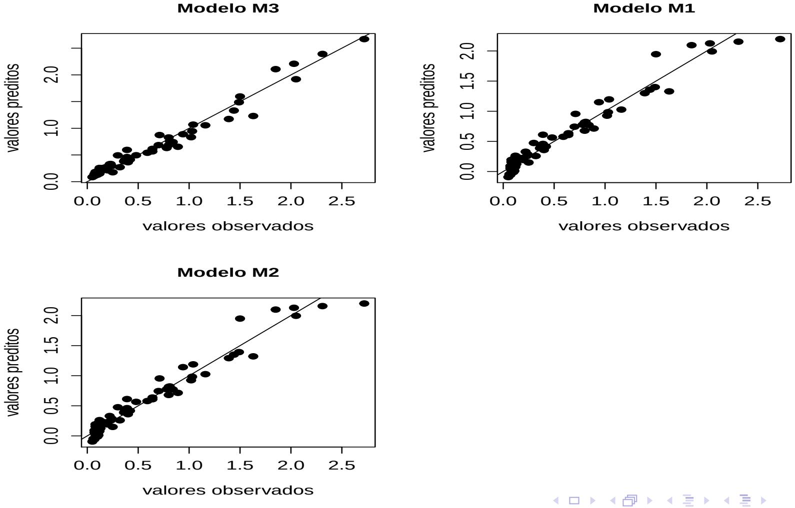
Resíduos normalizados com envelopes baseados no modelo



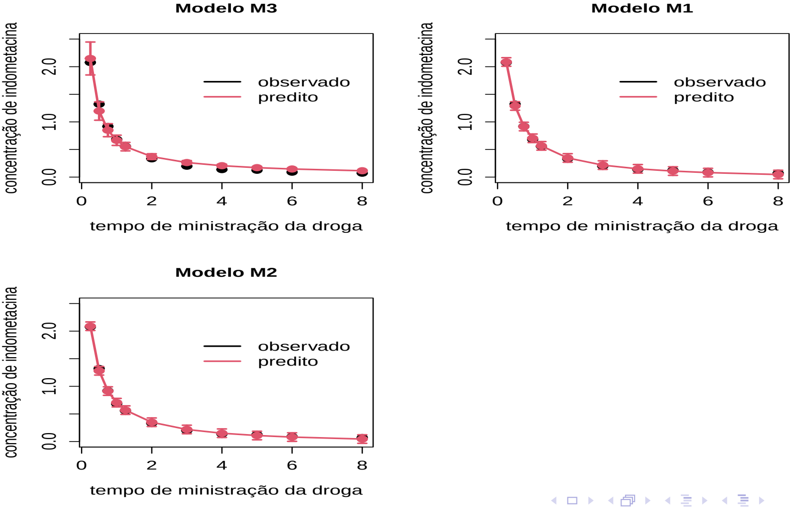




Valores individuais observados e preditos



Médias observadas e preditas (via valores preditos indiv.)

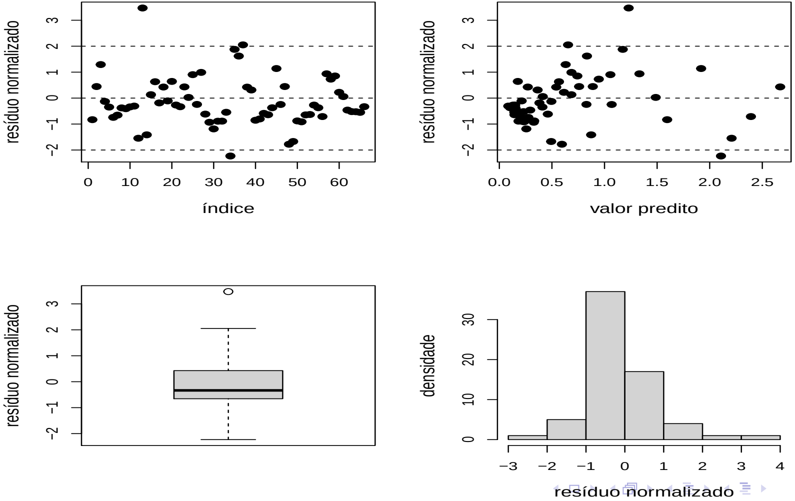


Estatísticas de comparação de modelos e somas de quadrados de resíduos

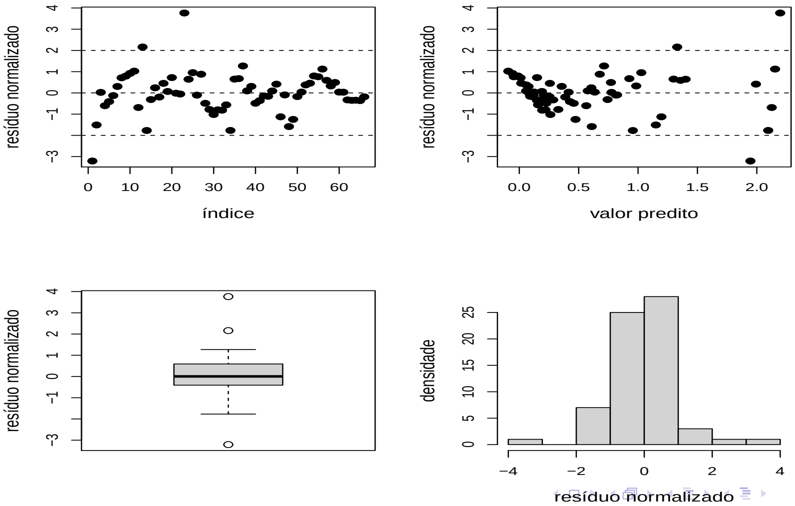
	AIC	BIC
M_3	-72,95	-64,20
M_2	-51,12	-41,18
M_1	-51,80	-40,85

M_3	M_2	M_1
0,915	0,924	0,924

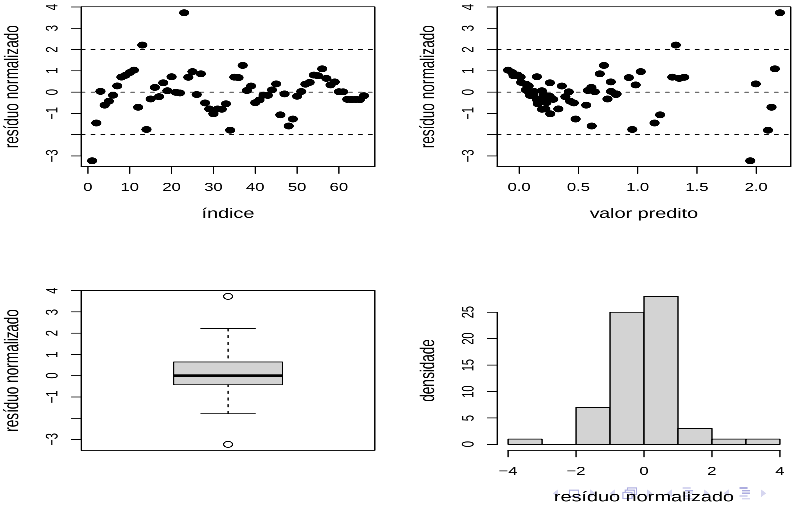
Gráficos de resíduos normalizados M3



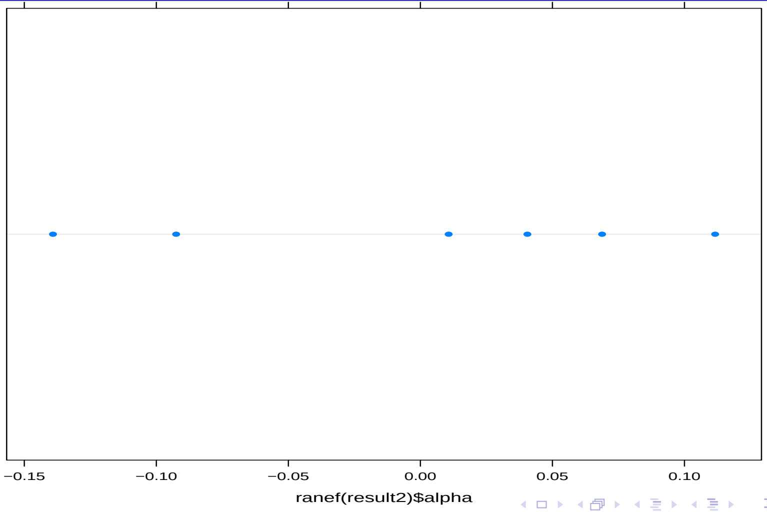
Gráficos de resíduos normalizados M1



Gráficos de resíduos normalizados M2



Efeitos aleatórios (M1)



Comentários

- Nenhum dos modelos apresentou bom ajuste.
- Dentre todos, o modelo 1 (M1), entretanto, apresentou o ajuste menos pior (incluindo a análise preditiva).
- Alternativa: utilizar uma distribuição com caudas pesadas para a distribuição do erro.
- Devido ao pequeno número de observações (seis) é difícil asseverar sobre o comportamento (distribuição) dos resíduos.

Estrutura de variância e covariância (M1)

- $\mathcal{E}(Y_{ji}) = \mathcal{E}(\gamma_{10} + u_{1j} + \gamma_{20}\exp(\gamma_{30}/v_{ji}) + \xi_{ij}) =$
  $\gamma_{10} + \gamma_{20}\exp(\gamma_{30}/v_{ji}) + \mathcal{E}(u_{1j}) + \mathcal{E}(\xi_{ji}) = \gamma_{10} + \gamma_{20}\exp(\gamma_{30}/w_{ji})$.
- $(u_{1j} \perp \xi_{ji}, \forall i,j) \mathcal{V}(Y_{ji}) = \mathcal{V}(\gamma_{10} + u_{1j} + \gamma_{20}\exp(\gamma_{30}/v_{ji}) + \xi_{ji}) =$
  $\mathcal{V}(u_{1j}) + \mathcal{V}(\xi_{ji}) = \psi + \sigma^2$.
- $Cov(Y_{ji}, Y_{ji'}) =$
  $Cov(\gamma_{10} + u_{1j} + \gamma_{20}\exp(\gamma_{30}/v_{ji}) + \xi_{ji}, \gamma_{10} + u_{1j} + \gamma_{20}\exp(\gamma_{30}/v_{ji'}) + \xi_{ji'})$
  $= \mathcal{V}(u_{1j}) = \psi$.
- $Corre(Y_{ji}, Y_{ji'}) = CCI = \frac{\psi}{\psi + \sigma^2}$.
- Exercício: fazer o mesmo para os modelos M2 e M3.

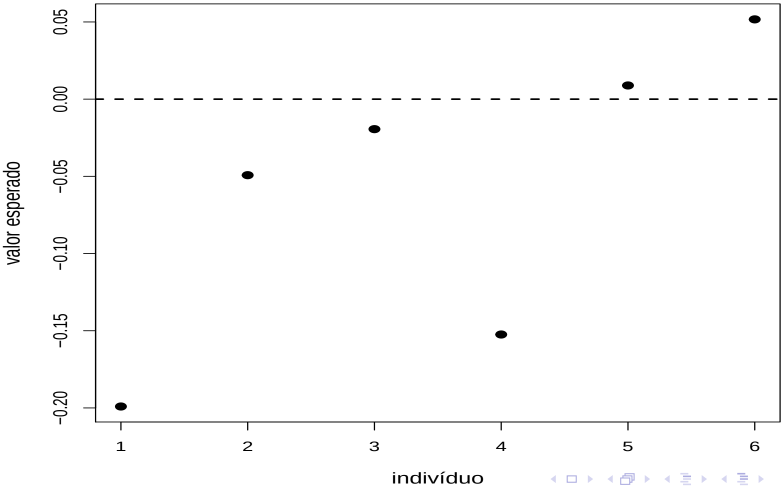
Estimativa dos parâmetros do modelo M1

Parâm.	Est.	EP	IC(95%)	Estat. t	p-valor (aprox. p/ normal)
γ_{10}	3,23	0,41	[2,42 ; 4,04]	7,84	$<$ 0,0001
γ_{20}	-3,29	0,39	[-4,06 ; -2,53]	-8,42	$<$ 0,0001
γ_{30}	-0,26	0,05	[-0,37 ;-0,16]	-5,09	$<$ 0,0001

Interpretação: valores esperados quando $v \to 0$, $\mathcal{E}(Y) = \widetilde{\gamma_{10}} = 3,23$ e $v \to \infty$, $\mathcal{E}(Y) = \widetilde{\gamma_{10}} + \widetilde{\gamma_{20}} = -0,06$.

ψ	σ^2	CCI
0,009	0,019	0,325

Valores esperados condicionais ($v \to \infty$)



Comentários

- Os parâmetros de regressão se mostraram significativos.
- As correlações entre as medidas feitas no mesmo indivíduo aparentam ser significativas.
- Há problemas na estimativa da média marginal (e em algumas médias condicionais) quando o tempo tende à infinito.
- Apesar do ajuste ruim, em termos da análise de resíduos, os valores preditos (individuais e médios) mostraram-se razoáveis.