Modelos não lineares mistos

Prof. Caio Azevedo

Relembrando

- MLM : $\mu_{ij} = \mathbf{X}_j'\beta + \mathbf{Z}_j\mathbf{b}_j$.
- MLGM : $\mu_{ij} = g^{-1}(\mathbf{X}_j'\beta + \mathbf{Z}_j\mathbf{b}_j)$.
- Nem sempre a relação entre as covariáveis ($\mathbf{X}_j$) e os efeitos aleatórios é adequamente descrita por um modelo linear ou por um modelo linear generalizado.
- Modelos não lineares (mistos) (MNLM) 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 MLM ou pelos MLGM.

Limitações dos MLM e MLGM e vantagens dos MNLM

- O comportamento linear e não linear induzidos, respectivamente, pelo MLM e MLGM 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 MLM e do MLGM (particularmente, polinômios de grau ≥ 3 possuem parâmetros de difícil interpretação), principalmente em termos do problema.

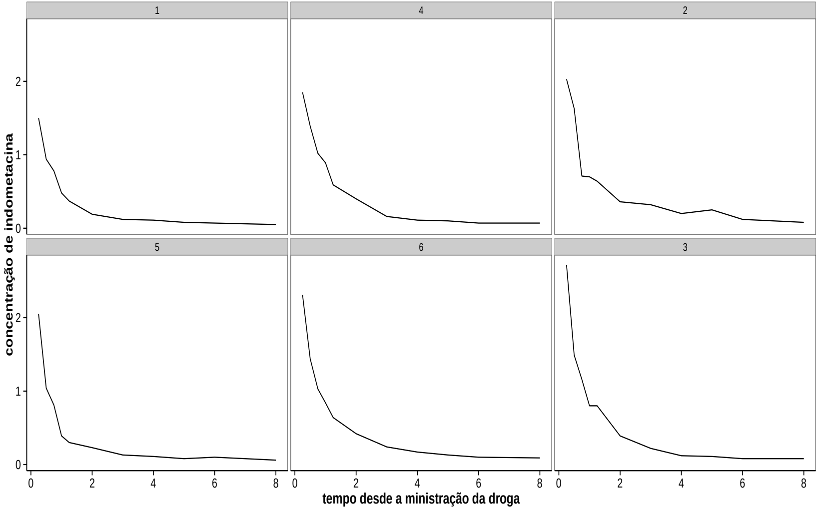
Limitações dos MLM e MLGM e vantagens dos MNLM

- Os MNLM, em geral, apresentam poucos parâmetros com interpretações úteis em termos do problema.
- Garantia de que as médias preditas respeitarão o espaço paramétrico associado.

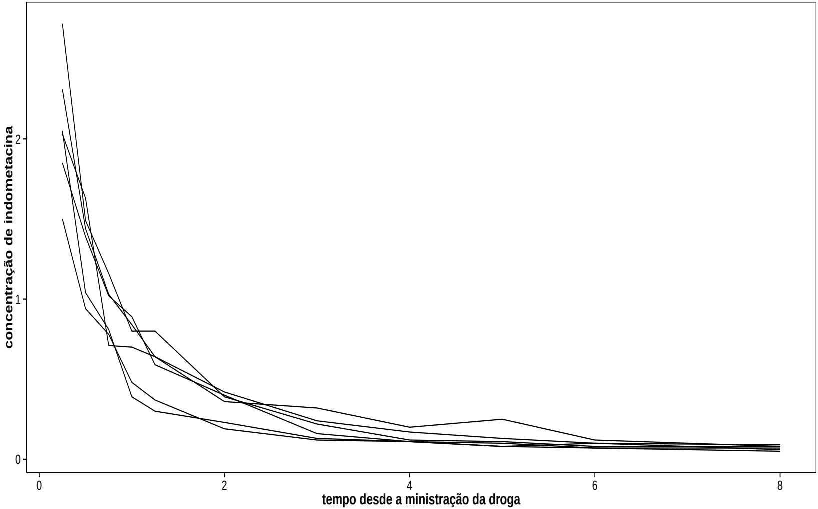
Exemplo 6: cinética de indometacina, Kwan et al (1976)

- 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).
- 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.
- Estudar o comportamento da concentração ao longo do tempo.

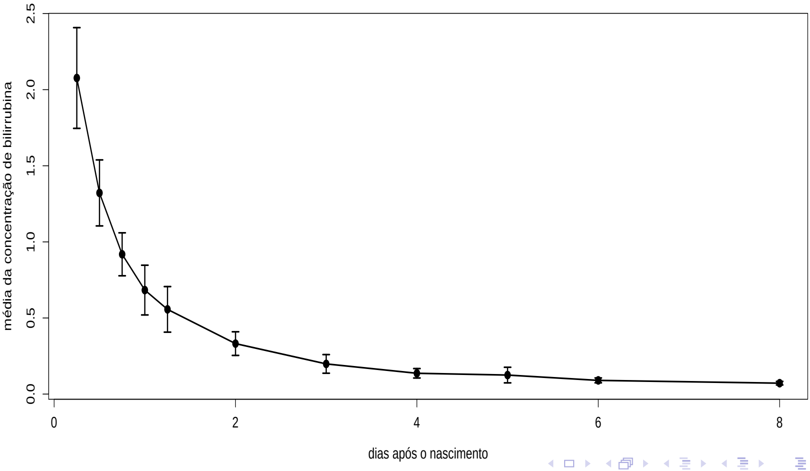
Perfis individuais (separados)



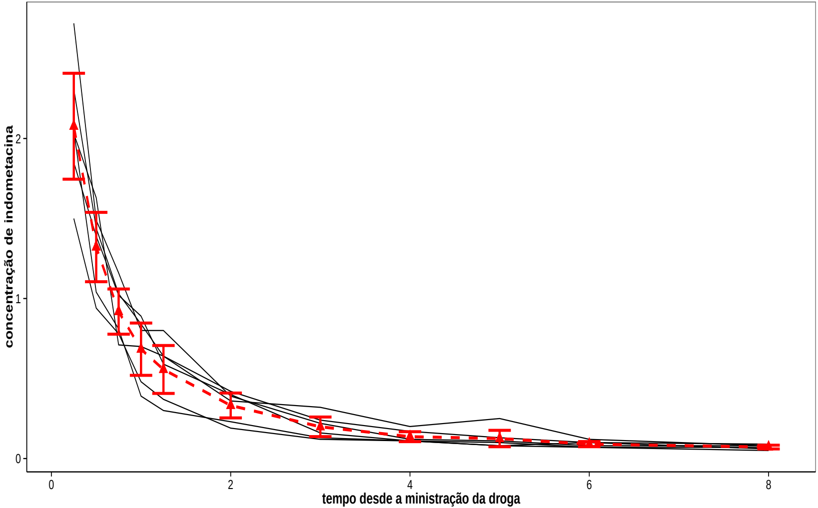
Perfis individuais (juntos)



Perfis médios



Perfis médios e individuais



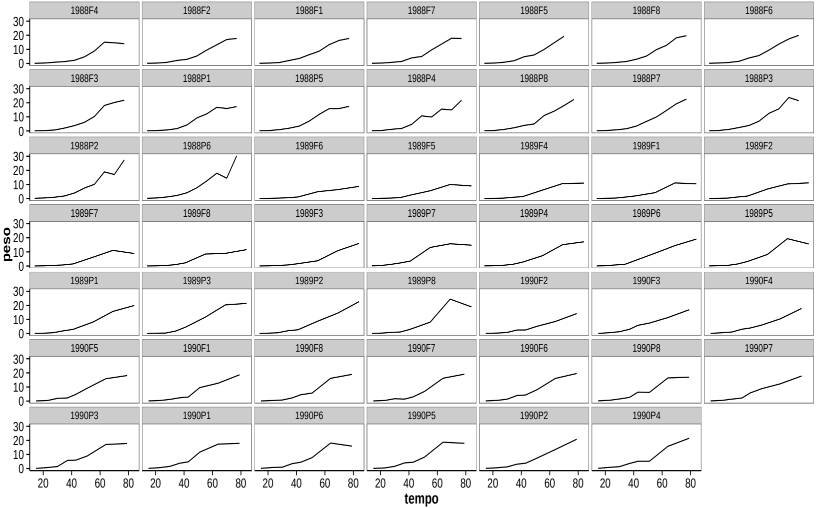
Exemplo 7: crescimento de plantas de soja, Davidian and Giltinan (1995) (1976)

- 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) numa frequência semanal entre duas à sete semanas depois de plantadas.

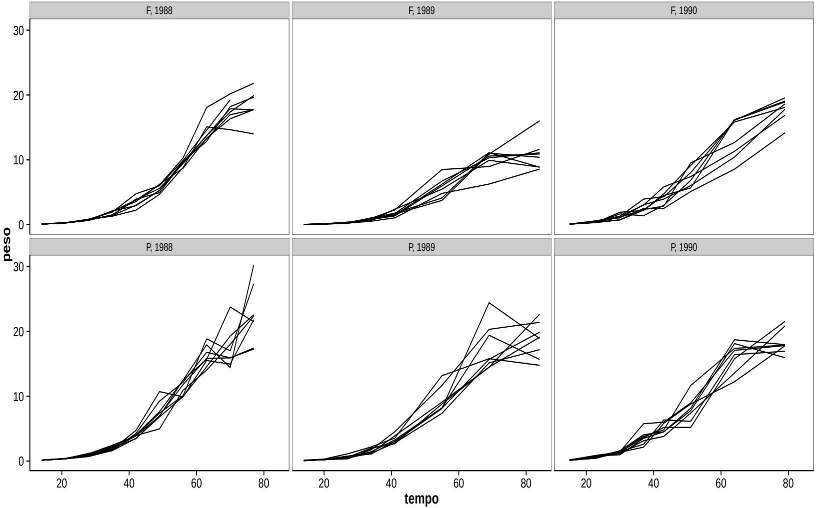
Cont.

- O experimento foi conduzido em três anos diferentes: (1988, 1989 e 1990).
- Objetivo: comparar os tipos de plantas, quanto crescimento, ao longo do tempo.

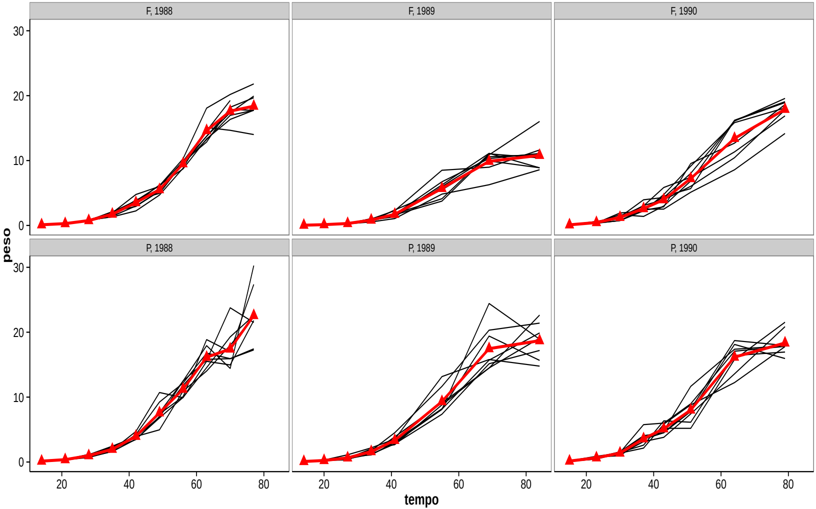
Perfis individuais por variedade x anos (separados)



Perfis individuais por variedade x anos (juntos)



Perfis individuais e médios por variedade x anos



Modelo não linear (sem efeito aleatório)

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

- Em que $\mathbf{X}_j$ representa covariáveis (matriz de planejamento) de interesse associadas aos efeitos fixos.
- $\phi_j = \mathbf{X}_j\boldsymbol{\beta}$.
- $\mathbf{W}_j$: 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 $\mathbf{X}_j = \mathbf{I}_3$. Para o modelo M_3 $\boldsymbol{\phi}_j = (\phi_1, \phi_2)' = (\beta_1, \beta_2)'$, ou seja $\mathbf{X}_j = \mathbf{I}_2$, 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 $\mathbf{X}_j = \mathbf{I}_4$. Em todos os casos, a matriz $\mathbf{W}_j$ 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:

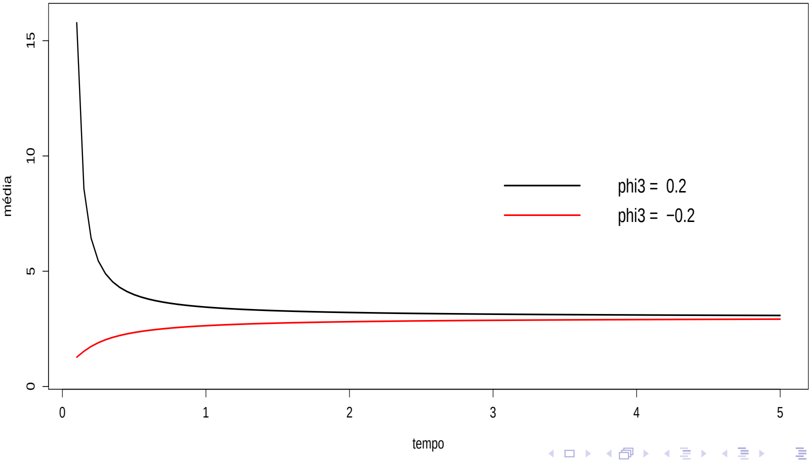
M_1

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

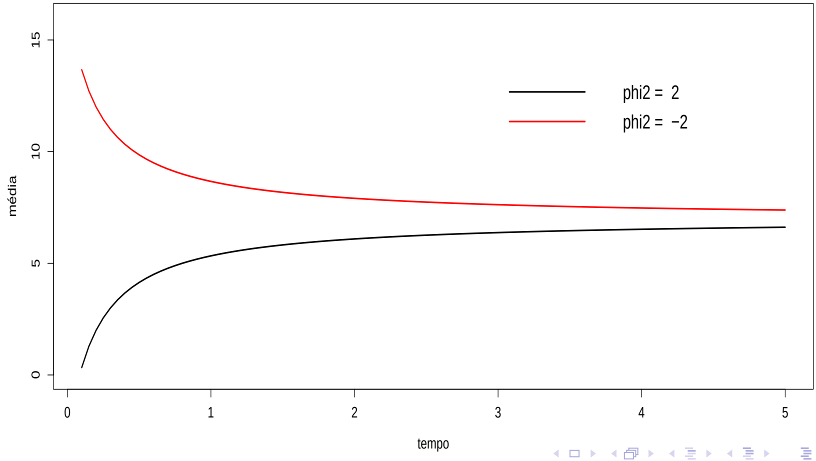
M_5

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

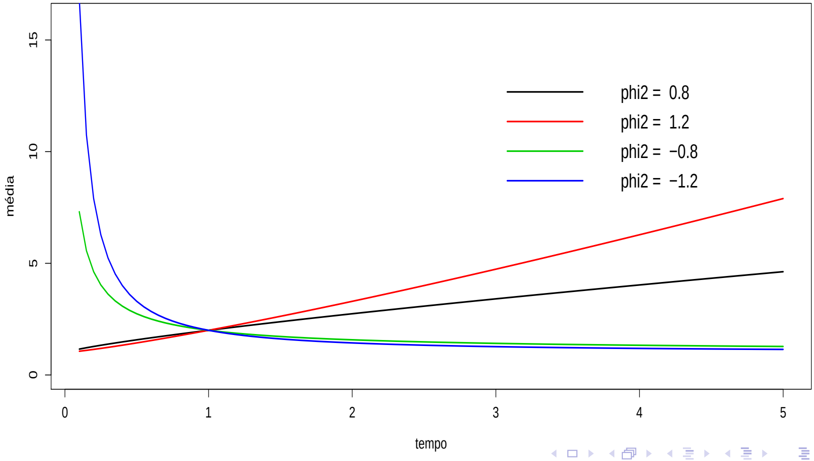
Modelo 1: $\phi_1 = 1, \phi_2 = 2$



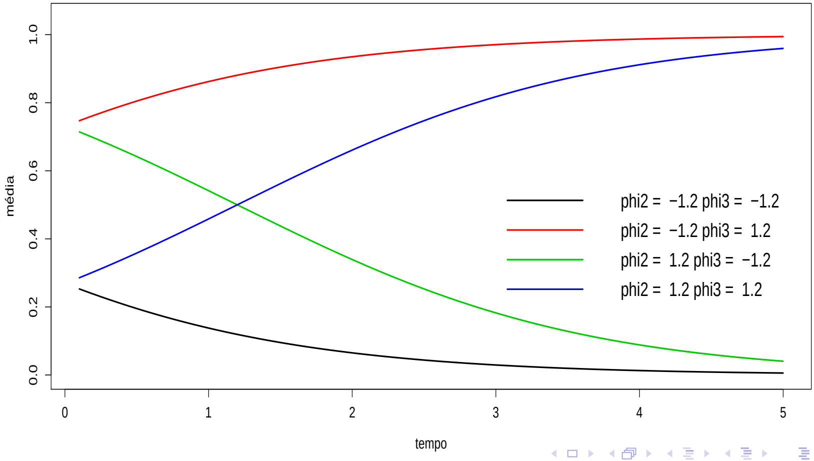
Modelo 2: $\phi_1 = 7, \phi_3 = 0,2$



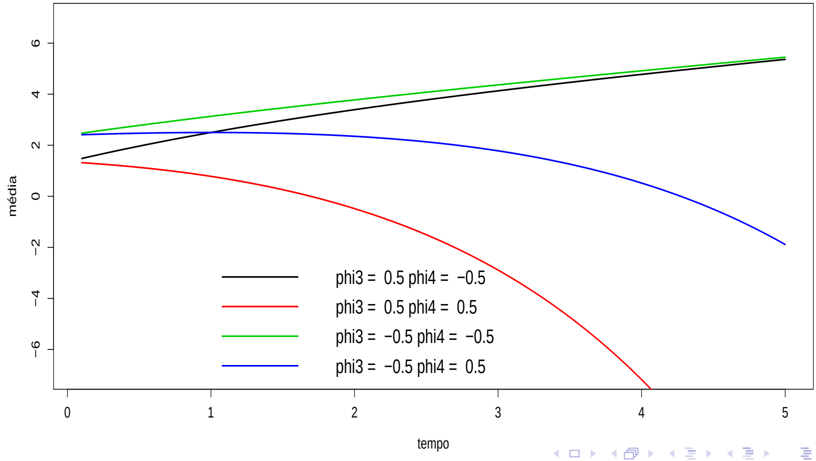
Modelo 3: $\phi_1 = 1$



Modelo 4: $\phi_1 = 1$



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



Modelo não linear misto

$$Y_{ij} = f(\phi_{ij}, \mathbf{X}_{ij}, \mathbf{W}_{ij}) + \xi_{ij}; j = 1, ..., n(\text{indivíduo}); i = 1, ..., k_j(\text{medida repetida})$$

- Em que $\mathbf{X}_{ij}$ representa covariáveis (matrizes de planejamento) de interesse associadas aos efeitos fixos ($\mathbf{A}_{ij}$) e aleatórios ($\mathbf{B}_{ij}$).
- $\phi_{ij} = \mathbf{A}_{ij}\boldsymbol{\beta} + \mathbf{B}_{ij}\mathbf{b}_j$.
- $\mathbf{W}_{ij}$: 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_{ij}.
- $\xi_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2), \mathbf{b}_j \overset{i.i.d}{\sim} N_q(\mathbf{0}, \boldsymbol{\Psi}), \xi_{ij} \perp \mathbf{b}_j \forall i, j$.

Modelo não linear misto (forma vetorial)

$$\mathbf{Y}_j = \mathbf{f}(\phi_j, \mathbf{X}_j, \mathbf{W}_j) + \xi_j; j = 1, ..., n$$

$$\mathbf{Y}_j = \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{k_j j} \end{bmatrix}; \mathbf{X}_j = \begin{bmatrix} \mathbf{X}_{1j} \\ \mathbf{X}_{2j} \\ \vdots \\ \mathbf{X}_{k_j j} \end{bmatrix}; \mathbf{A}_j = \begin{bmatrix} \mathbf{A}_{1j} \\ \mathbf{A}_{2j} \\ \vdots \\ \mathbf{A}_{k_j j} \end{bmatrix}; \mathbf{B}_j = \begin{bmatrix} \mathbf{B}_{1j} \\ \mathbf{B}_{2j} \\ \vdots \\ \mathbf{B}_{k_j j} \end{bmatrix};$$

$$\phi_j = \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \vdots \\ \phi_{k_j j} \end{bmatrix}; \boldsymbol{\xi}_j = \begin{bmatrix} \xi_{1j} \\ \xi_{2j} \\ \vdots \\ \xi_{k_j j} \end{bmatrix}; \mathbf{f}(\phi_j, \mathbf{X}_j) = \begin{bmatrix} f(\phi_{1j}, \mathbf{X}_{1j}, \mathbf{W}_{1j}) \\ f(\phi_{2j}, \mathbf{X}_{2j}, \mathbf{W}_{2j}) \\ \vdots \\ f(\phi_{k_j j}, \mathbf{X}_{k_j j}, \mathbf{W}_{k_j j}) \end{bmatrix}; \mathbf{W}_j = \begin{bmatrix} \mathbf{W}_{1j} \\ \mathbf{W}_{2j} \\ \vdots \\ \mathbf{W}_{k_j j} \end{bmatrix}$$

Modelo não lineares mistos: exemplos

- $(M_1): Y_{ij} = \phi_{1j} + \phi_{2j}\exp\left(\phi_{3j}/w_{ij}\right) + \xi_{ij}$.
- $(M_2): Yij = \phi_{1j} - \phi_{2j}\left(w_{ij} + \phi_{3j}\right)^{-1} + \xi_{ij}$.
- $(M_3): Yij = \phi_{1j} w_{ij}^{\phi_{2j}} + \xi_{ij}$.
- $(M_4): Yij = \phi_{1j}/(1 + exp(-(w_{ij} - \phi_{2j})/\phi_{3j})) + \xi_{ij}$.
- $(M_5): Yij = \phi_{1j} + \phi_{2j} w_{1j} - e^{\phi_{3j} + \phi_{4j} w_{ij}} + \xi_{ij}$.

Modelo 1 misto

- $Y_{ij} = \phi_{1j} + \phi_{2j}\exp\left(\phi_{3j}/w_{ij}\right) + \xi_{ij}$.

  $\phi_{1j} = \beta_1 + b_{1j}$, $\phi_{2j} = \beta_2$ e $\phi_{3j} = \beta_3$, nesse caso, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$, $\mathbf{b}_j = b_{1j} \overset{i.i.d.}{\sim} N(0, \psi)$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} b_{1j} \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{A}_{ij} = \mathbf{I}_3, \mathbf{B}_{ij} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Modelo 1 misto

- $Y_{ij} = \phi_{1j} + \phi_{2j} \exp\left(\phi_{3j}/w_{ij}\right) + \xi_{ij}$.

  $\phi_{1j} = \beta_1 + b_{1j}$, $\phi_{2j} = \beta_2 + b_{2j}$ e $\phi_{3j} = \beta_3$, nesse caso,

  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$, $\mathbf{b}_j = (b_{1j}, b_{2j})' \overset{i.i.d.}{\sim} N_2(\mathbf{0}, \boldsymbol{\Psi})$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} b_{1j} \\ b_{2j} \\ 0 \end{bmatrix}$$

$$\mathbf{A}_{ij} = \mathbf{I}_3, \mathbf{B}_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Modelo 1 misto

- $Y_{ij} = \phi_{1j} + \phi_{2j} \exp\left(\phi_{3j}/w_{ij}\right) + \xi_{ij}$.

  $\phi_{1j} = \beta_1 + b_{1j}$, $\phi_{2j} = \beta_2 + b_{2j}$ e $\phi_{3j} = \beta_3 + b_{3j}$, nesse caso,

  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$, $\mathbf{b}_j = (b_{1j}, b_{2j}, b_{3j})' \overset{i.i.d}{\sim} N_3(\mathbf{0}, \mathbf{\Psi})$ e

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

  $\mathbf{A}_{ij} = \mathbf{I}_3, \mathbf{B}_{ij} = \mathbf{I}_3$

  Em todos os casos, a matriz $\mathbf{W}_{ij}$ corresponde à variável w_{ij}, que pode ser o tempo.

Modelo 3 misto

- $Y_{ij} = \phi_{1j} w_{ij}^{\phi_{2j}} + \xi_{ij}$.

  $\phi_{1j} = \beta_1 + b_{1j}$ e $\phi_{2j} = \beta_2$, nesse caso, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$,

  $\mathbf{b}_j = b_{1j} \overset{i.i.d}{\sim} N(0, \psi)$ e

$$\boldsymbol{\phi}_j = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} b_{1j} \\ 0 \end{bmatrix}$$

$$\mathbf{A}_{ij} = \mathbf{I}_2, \mathbf{B}_{ij} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Modelo 3 misto

- $Y_{ij} = \phi_{1j} w_{ij}^{\phi_{2j}} + \xi_{ij}$.

  $\phi_{1j} = \beta_1 + b_{1j}$ e $\phi_{2j} = \beta_2 + b_{2j}$, nesse caso, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, $\mathbf{b}_j = (b_{1j}, b_{2j}) \overset{i.i.d}{\sim} N_2(\mathbf{0}, \boldsymbol{\Psi})$ e

$$\phi_j = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} b_{1j} \\ b_{2j} \end{bmatrix}$$

  $\mathbf{A}_{ij} = \mathbf{I}_2, \mathbf{B}_{ij} = \mathbf{I}_2$

Estimação

- Parâmetros para estimar $(\boldsymbol{\beta}, \mathbf{b}, \boldsymbol{\theta}, \sigma^2)$, em que $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta})$.
- Verossimilhança completa:

$$
\begin{aligned}
L(\boldsymbol{\beta}, \mathbf{b}, \boldsymbol{\theta}, \sigma^2) &= \exp\left\{-\sum_{j=1}^{n}\sum_{i=1}^{k_j}\frac{(y_{ij} - f(\phi_{ij}, \mathbf{X}_{ij}, \mathbf{W}_{ij}))^2}{2\sigma^2}\right\} \\
&\times \; exp\left\{-0,5\sum_{j=1}^{n}\mathbf{b}_j'\boldsymbol{\Psi}^{-1}\mathbf{b}_j\right\}|\boldsymbol{\Psi}|^{-n/2}
\end{aligned}
$$

Verossimilhança marginal

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &\propto \prod_{j=1}^{n} \int_{\Re^q} \exp\left\{ - \sum_{i=1}^{k_j} \frac{(y_{ij} - f(\phi_{ij}, \mathbf{X}_{ij}, \mathbf{W}_{ij}))^2}{2\sigma^2} \right\} \\
&\times \quad exp\left\{ -0,5\mathbf{b}_j' \boldsymbol{\Psi}^{-1} \mathbf{b}_j \right\} |\boldsymbol{\Psi}|^{-1/2} d\mathbf{b}
\end{aligned}
$$

A integral acima não tem solução explícita (lembrando que ϕ_{ij} é função de $\mathbf{b}_j$).

Resolução da integral

- Quadratura (gaussiana).
- Quadratura adaptativa.
- Aproximação da verossimilhança por um modelo linear misto.
- Aproximação de Laplace.
- Aproximação do integrando usando o primeiro termo da expansão em séries de Taylor em torno do valor esperado condicional dos efeitos aleatórios.
- Aproximação do integrando usando o primeiro termo da expansão em séries de Taylor em torno da moda condicional dos efeitos aleatórios.
- Monte carlo.

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 ($N = \sum_{j=1}^{n} k_j$). Assim:

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) &\propto (\sigma^2)^{-(N+nq)/2} \\
&\times \prod_{j=1}^{n} \int_{\Re^q} \exp\left\{ -\sum_{i=1}^{k_j} \frac{(y_{ij} - f(\phi_{ij}, \mathbf{X}_{ij}, \mathbf{W}_{ij}))^2 + ||\boldsymbol{\Delta}\mathbf{b}_j||^2}{2\sigma^2} \right\} \\
&\times |\boldsymbol{\Delta}| d\mathbf{b}_j \qquad (1)
\end{aligned}
$$

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 mistos (MLM).
- No passo MQNLP, consideramos uma estimativa provisória de $\boldsymbol{\Delta}$ e obtemos estimativas provisórias de $\mathbf{b}_j$ e $\boldsymbol{\beta}$ minimizando

$$\sum_{j=1}^{n}\left[\sum_{i=1}^{k_j}(y_{ij} - f(\phi_{ij}, \mathbf{X}_{ij}, \mathbf{W}_{ij}))^2 + ||\boldsymbol{\Delta}\mathbf{b}_j||^2\right]$$

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 $\boldsymbol{\beta}$ e $\mathbf{b}_j$, as quais serão denotadas por $\widetilde{\boldsymbol{\beta}}^{(w)}$ e $\widetilde{\mathbf{b}}_j^{(w)}$. Defina ainda

$$\widetilde{\mathbf{w}}_j^{(w)} = \mathbf{y}_j - \mathbf{f}_j(\widetilde{\boldsymbol{\beta}}^{(w)}, \widetilde{\mathbf{b}}_j^{(w)}) + \widetilde{\mathbf{X}}_j^{(w)} \widetilde{\boldsymbol{\beta}}^{(w)} + \widetilde{\mathbf{Z}}_j^{(w)} \widetilde{\boldsymbol{\beta}}_j^{(w)}$$

em que $\widetilde{\mathbf{X}}_j^{(w)} = \left.\dfrac{\partial \mathbf{f}_j}{\partial \boldsymbol{\beta}'}\right|_{\widetilde{\boldsymbol{\beta}}^{(w)}, \widetilde{\mathbf{b}}_j^{(w)}}$ e $\widetilde{\mathbf{Z}}_j^{(w)} = \left.\dfrac{\partial \mathbf{f}_j}{\partial \mathbf{b}_j'}\right|_{\widetilde{\boldsymbol{\beta}}^{(w)}, \widetilde{\mathbf{b}}_j^{(w)}}$

Cont.

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

$$
\begin{aligned}
l_{MLM}(\beta, \mathbf{\Delta}, \sigma^2) &= -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{j=1}^{n}\Big\{ \ln|\mathbf{\Sigma}_j(\mathbf{\Delta})| \\
&+ \sigma^{-2}\left[\mathbf{w}_j^{(w)} - \widetilde{\mathbf{X}}_j^{(w)}\beta\right]' \mathbf{\Sigma}_j^{-1}(\mathbf{\Delta}) \\
&\times \left[\mathbf{w}_j^{(w)} - \widetilde{\mathbf{X}}_j^{(w)}\beta\right]\Big\} \qquad (2)
\end{aligned}
$$

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

Cont.

- De modo semelhante ao que é feito nos MLM, podemos obter os valores que maximizam a função acima, em relação à $\boldsymbol{\beta}$ 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 MLM), ou seja:

$$l_R(\sigma^2, \mathbf{\Delta}) = l(\widetilde{\beta}(\mathbf{\Delta}), \mathbf{\Delta}, \sigma^2) - \frac{1}{2}\sum_{j=1}^{n} \ln|\sigma^{-2}\widetilde{\mathbf{X}}_j^{(w)'}\mathbf{\Sigma}_j(\mathbf{\Delta})\widetilde{\mathbf{X}}_j^{(w)}|$$

em que $l(\widetilde{\beta}(\mathbf{\Delta}), \mathbf{\Delta}, \sigma^2)$ é dada por (2), com β substituido por $\widetilde{\beta}$.

- O algoritmo alterna entre os passos MQNLP e MLM, até que algum critério de convergência seja alcançado.

Aproximação de Laplace

- A integral que queremos aproximar é a constante na expressão abaixo:

$$
\begin{aligned}
L(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2) &\propto \prod_{j=1}^{n}\int_{\Re^q} \exp\left\{-\sum_{i=1}^{k_j}\frac{(y_{ij}-f(\phi_{ij},\mathbf{X}_{ij},\mathbf{W}_{ij}))^2+||\boldsymbol{\Delta}\mathbf{b}_j||^2}{2\sigma^2}\right\} \\
&\times (2\pi\sigma^2)^{-(k_j+1)/2}|\boldsymbol{\Delta}|d\mathbf{b}
\end{aligned}
$$

- Seja $g(\boldsymbol{\beta},\boldsymbol{\Delta},\mathbf{y}_j,\mathbf{b}_j)=\sum_{i=1}^{k_j}(y_{ij}-f(\phi_{ij},\mathbf{X}_{ij},\mathbf{W}_{ij}))^2+||\boldsymbol{\Delta}\mathbf{b}_j||^2$.

Aproximação de Laplace

- Defina:

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

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

$$g(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j, \mathbf{b}_j) \approx g(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j, \widetilde{\mathbf{b}}_j) + \frac{1}{2}\left(\mathbf{b}_j - \widetilde{\mathbf{b}}_j\right)' g''(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j, \widetilde{\mathbf{b}}_j)\left(\mathbf{b}_j - \widetilde{\mathbf{b}}_j\right)$$

(note que $g'(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j, \widetilde{\mathbf{b}}_j) = \mathbf{0}$)

Cont.

- A aproximação de Laplace da verossimilhança é dada por:

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

- Temos que (veja Pinheiro and Bates (2009), pag. 316-317)

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

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

$$
\begin{aligned}
l_{LA}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}) &= -\frac{N}{2} + n \ln |\boldsymbol{\Delta}| \\
&- \frac{1}{2} \left\{ \sum_{j=1}^{n} \ln |G(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j)| + \sigma^{-2} \sum_{j=1}^{n} g(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j, \widetilde{\mathbf{b}}_j) \right\}
\end{aligned}
$$

- O procedimento itera entre a maximização de l_{LA} em relação à $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}$ e da minimiização de $g(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j)$ com relação à $\mathbf{b}_j$.

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 $exp(-g(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j, \mathbf{b}_j)/(2\sigma^2))$ pode ser aproximado por uma distribuição $N_q(\widetilde{\mathbf{b}}_j, \sigma^2 G^{-1}(\boldsymbol{\beta}, \boldsymbol{\Delta}, \mathbf{y}_j))$

Quadratura adaptativa

- Assim

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

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

Cont.

- A aproximação da logverossimilhança por quadratura adaptativa:

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

- A logverossimilhança acima pode ser maximizada através de algoritmos de otimização adequados.

Recursos computacionais

- As funções “nls” (default no R) e “nlme” (pacote nlme) ajustam, respectivamente, modelos não lineares e modelos não lineares mistos.
- A primeira estima os parâmetros via mínimos quadrados (ponderados) não lineares.
- A segunda o faz usando máxima verossimilhança ou máxima verossimilhança restrita.
- 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 com esse pacote (derivada e estimativa iniciais) já estão disponíveis na função ($input$: covariável, $dose$: valor inicial de interesse).

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.

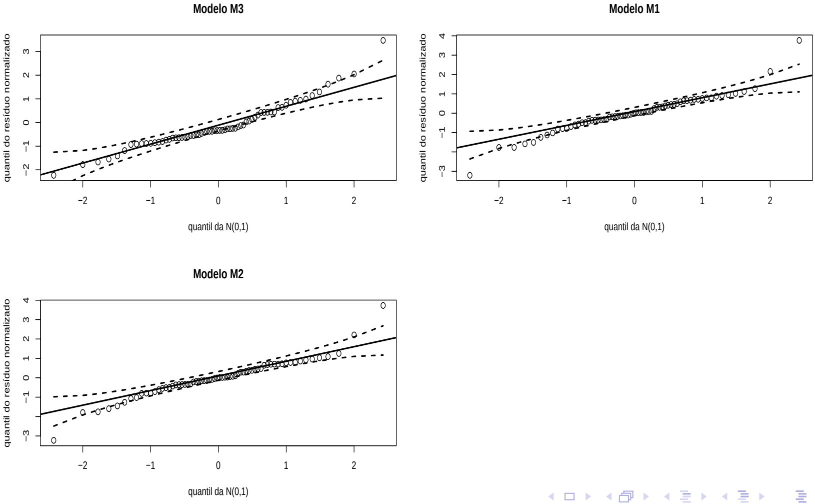
- (x : covariável)

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	$Asym - Drop * exp(-exp(lrc) * x^{pwr})$

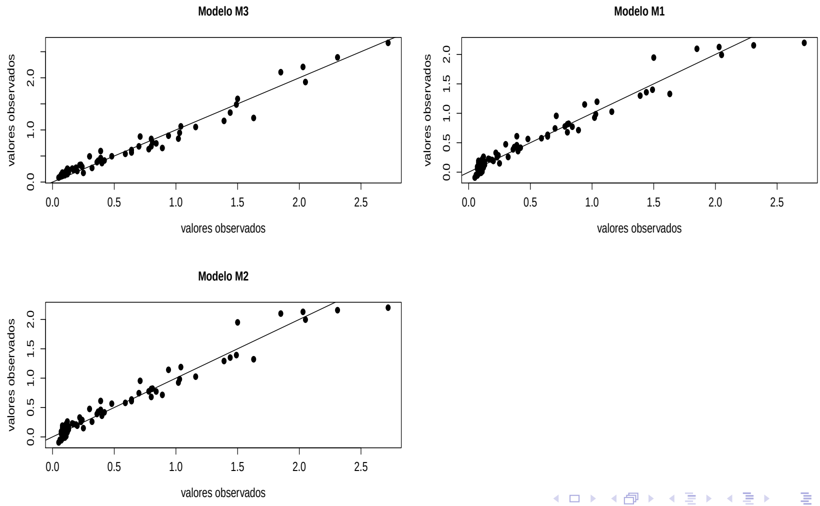
Voltando ao Exemplo 6

- Ajustou-se os modelos M_1, M_2 e M_3 (não houve convergência para os modelo M_4 e M_5)
- $(M_1): Y_{ij} = \phi_{1j} + \phi_2 \exp\left(\phi_3/w_{ij}\right) + \xi_{ij}$.
- $(M_2): Y_{ij} = \phi_{1j} - \phi_2 \left(w_{ij} + \phi_3\right)^{-1} + \xi_{ij}$.
- $(M_3): Y_{ij} = \phi_{1j} w_j^{\phi_2} + \xi_{ij}$.
- $\phi_{1j} = \phi_1 + b_j$, $b_j \overset{i.i.d.}{\sim} N(0, \psi)$.
- $\xi_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $\xi_{ij} \perp b_j$.

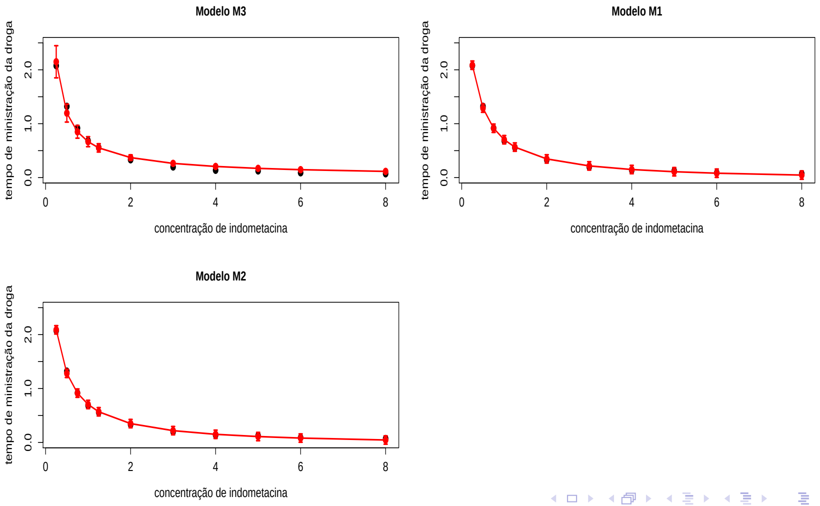
Resíduos normalizados com envelopes baseados na N(0,1)



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

Resíduos normalizados com envelopes baseados no modelo