

Modelos de regressão normais não lineares

Prof. Caio Azevedo

Introdução

- MRNLH : $\mu_i = \mathbf{X}'_i\beta$.
- MLG : $\mu_i = g^{-1}(\mathbf{X}'_i\beta)$.
- Nem sempre a relação entre as covariáveis (\mathbf{X}_i) e a resposta pode ser descrita adequadamente por um modelo linear.

Introdução

- Modelos não lineares (MNL) 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 MRNLH. Além disso, podemos ter assíntotas (inferior e/ou superior).
- Vamos utilizar a sigla MRNNLH: modelos de regressão normais não lineares homocedásticos.

Limitações dos MRNLH e MLG e vantagens dos MRNNLH

- As interpretações dos parâmetros de modelos não lineares podem ser mais adequadas do que aquelas associadas aos parâmetros do MRNLH (particularmente, polinômios de grau ≥ 3 apresentam parâmetros de difícil interpretação), principalmente em termos do problema.
- À rigor podemos ter modelos para a classe dos MLG semelhantes à modelos não lineares (e.g., $\mu_i = e^{\eta_i}$), mas estes não são modelos intrinsecamente não-lineares ($\mu_i = g^{-1}(\eta_i)$), em que η_i é um preditor linear.
- Existência de assíntotas (inferior e/ou superior)

Limitações dos MRNLH e MLG e vantagens dos MRNNLH

- O comportamento linear e não linear induzidos, respectivamente, pelo MRNLH e MLGM podem não ser adequados para representar o comportamento da média.
- Há modelos teóricos que explicam satisfatoriamente certos fenômenos, os quais se baseiam em soluções de equações diferenciais (e outras metodologias matemáticas). Estas, por sua vez, podem induzir estruturas não-lineares.
- Veja [aqui](#), [aqui](#), [aqui](#), [aqui](#), [aqui](#), [aqui](#).

Cont.

- Nem todos os MRNLH são linearizáveis ou, mesmo quando o são, tal procedimento pode levar a inferências ruins ([aqui](#)).
- Os MRNNLH, em geral, apresentam poucos parâmetros, com interpretações úteis em termos do problema.
- Modelam o problema de forma satisfatória, em geral, com um número pequeno de parâmetros.
- Garantia de que as médias previstas respeitarão o espaço paramétrico associado.

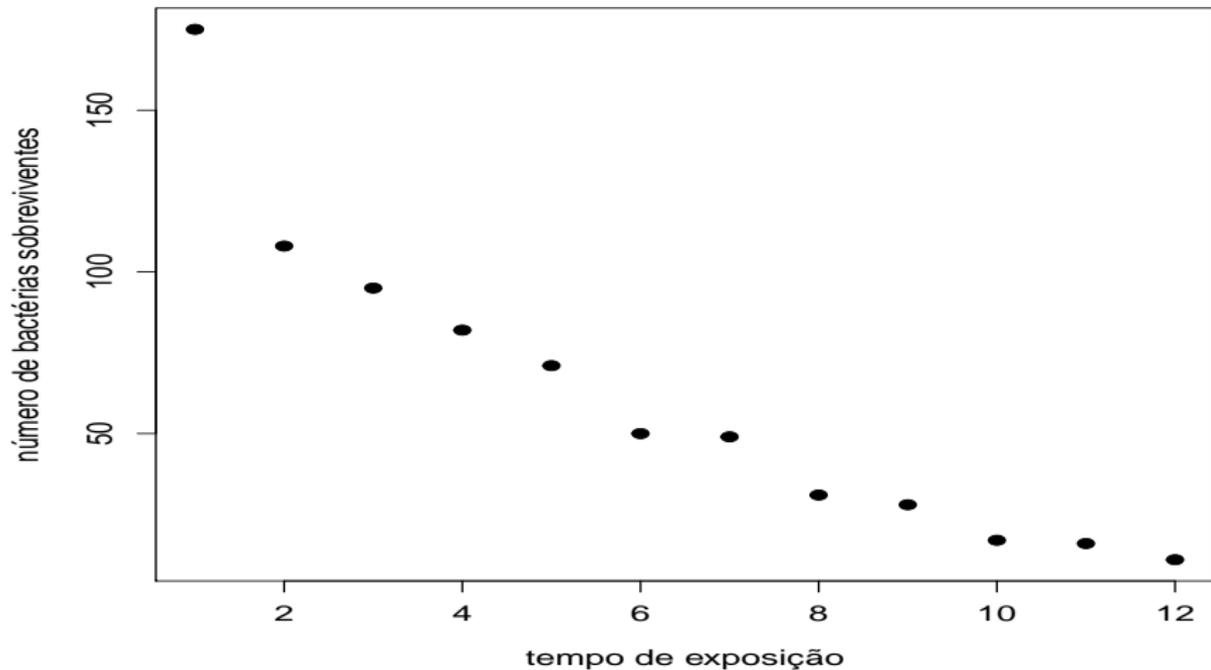
(Voltando ao) Exemplo 3: tempo de sobrevivências de bactérias

- Os dados correspondem ao número de bactérias sobreviventes em amostras de um produto alimentício segundo o tempo (em minutos) de exposição do produto à uma temperatura de $300^{\circ}F$.
- Resposta: número (contagem) de bactérias sobreviventes.
- Variável explicativa: tempo de exposição.

(Cont.)

- Nessas amostras de alimentos foram feitas 12 medições, a cada minuto, contabilizando a quantidade de bactérias vivas (do total original) sobreviventes.
- Novamente temos uma situação de medidas repetidas e, assim, as observações podem ter algum tipo de dependência.
- Objetivo: modelar o número de bactérias sobreviventes em função da temperatura.

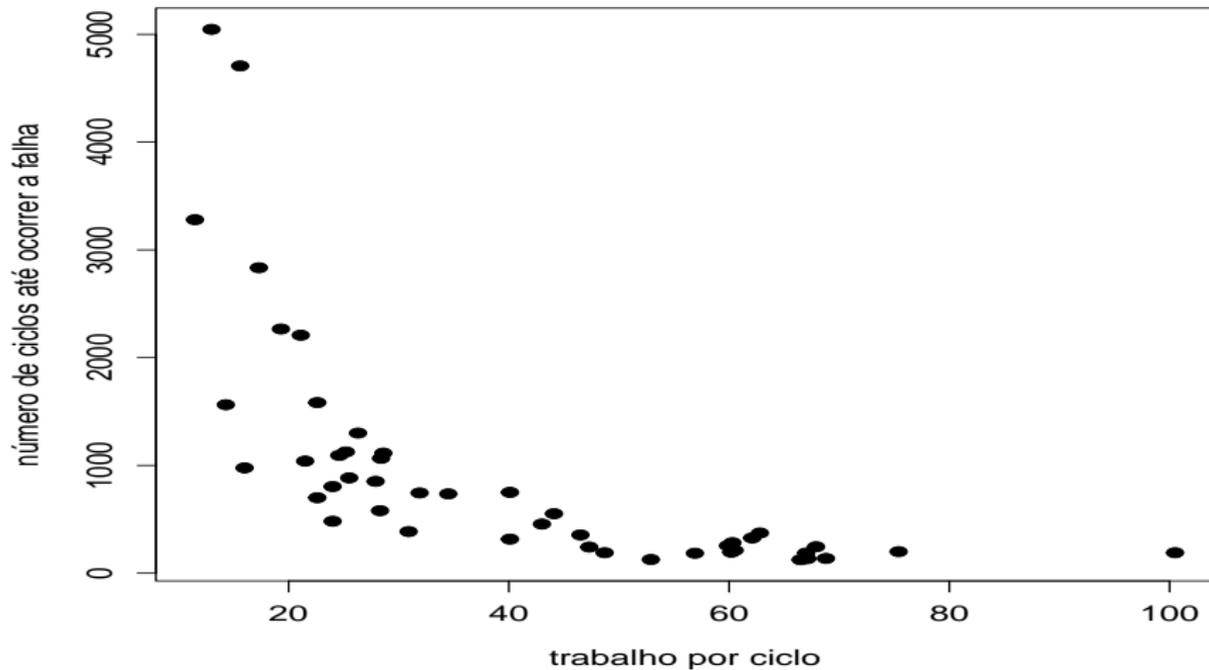
Gráfico de dispersão



Exemplo 16: Dados sobre fadiga biaxial

- Diz respeito sobre fadiga (rachadura/falha no funcionamento) de peças de metal (46 observações).
- Resposta: número de ciclos até ocorrer a falha.
- Variável explicativa: trabalho por ciclo (mJ / m^3) (intensidade do stress).
- Objetivo: modelar o número de ciclos em função do trabalho por ciclo.

Gráfico de dispersão



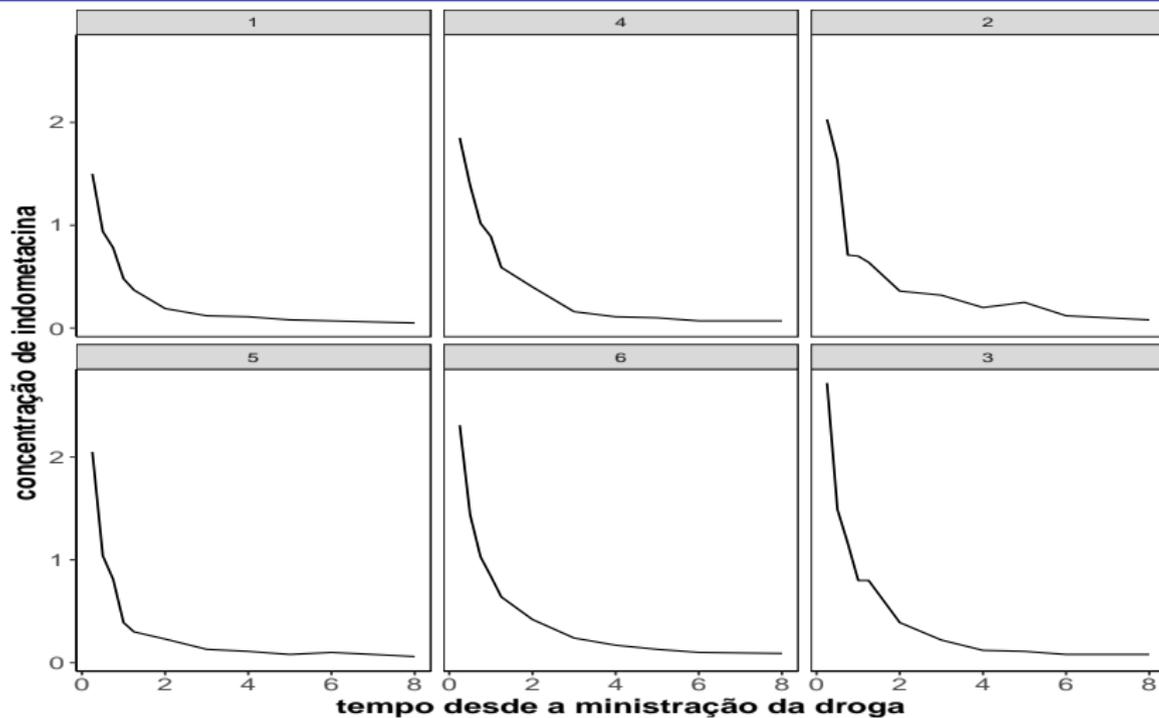
Exemplo 17: cinética de indometacina, Kwan et al (1976)

- 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).

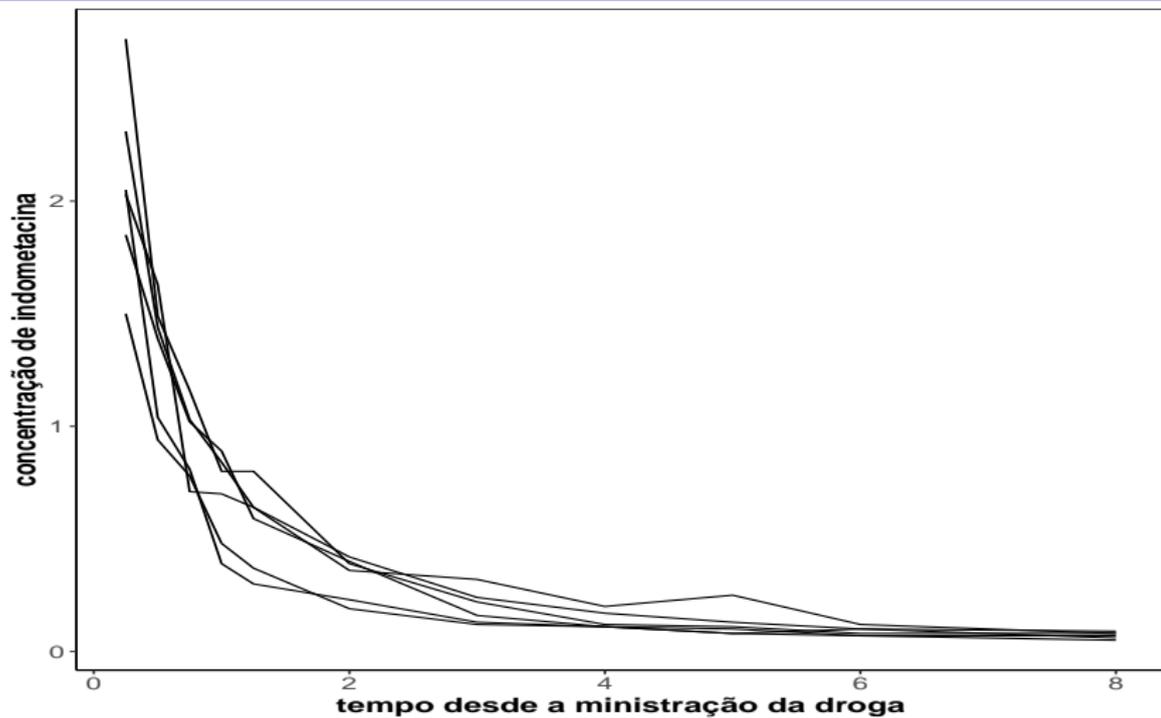
Exemplo 17: cinética de indometacina, Kwan et al (1976)

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

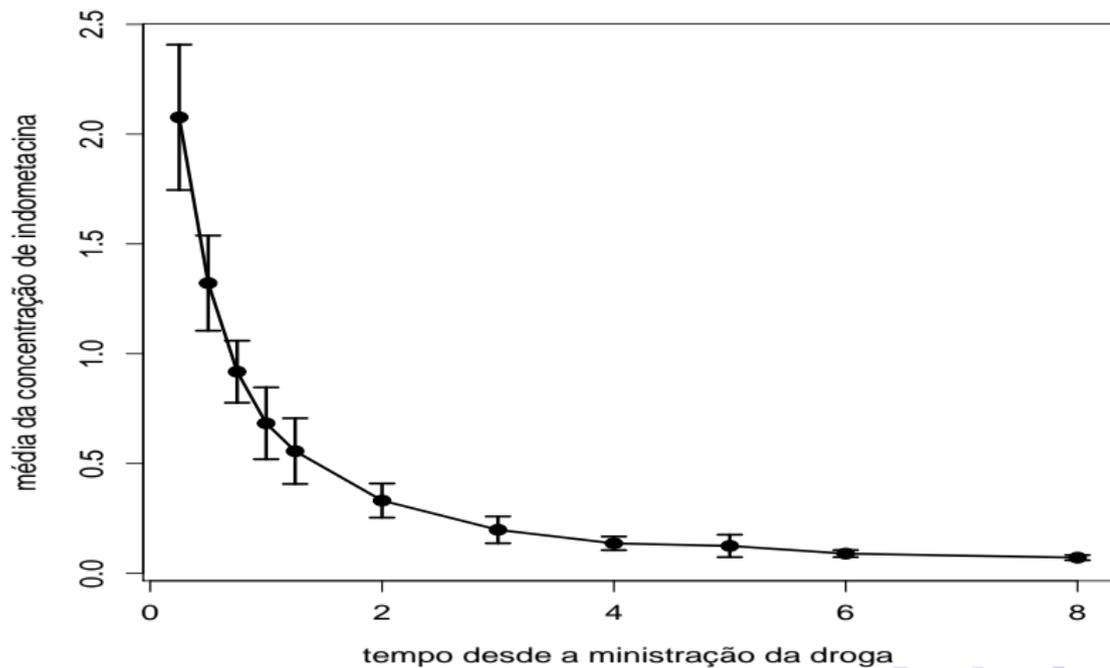
Perfis individuais (separados)



Perfis individuais (juntos)



Perfil médio



Modelo normal não linear

$$Y_i = f(\phi_i, \mathbf{X}_i, \mathbf{W}_i) + \xi_i; i = 1, \dots, n \text{ (indivíduo)}$$

- Em que \mathbf{X}_i representa covariáveis (matriz de planejamento) de interesse associadas aos efeitos fixos.
- $\phi_i = \mathbf{X}_i \beta$.
- \mathbf{W}_i : 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_i .
- $\xi_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$.

Modelo não lineares: exemplos

- $(M_1) : Y_i = \phi_1 + \phi_2 \exp(\phi_3/w_i) + \xi_i.$
- $(M_2) : Y_i = \phi_1 - \phi_2 (w_i + \phi_3)^{-1} + \xi_i.$
- $(M_3) : Y_i = \phi_1 w_i^{\phi_2} + \xi_i.$
- $(M_4) : Y_i = \phi_1 / (1 + \exp(-(w_i - \phi_2)/\phi_3)) + \xi_i.$
- $(M_5) : Y_i = \phi_1 + \phi_2 w_i - e^{\phi_3 + \phi_4 w_i} + \xi_i.$

Modelo não lineares: exemplos

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

Cont.

- Os modelos M_1 e M_5 podem ser apresentados de uma outra forma ($\phi_i = \mathbf{X}_i\boldsymbol{\beta}$),

$$(M_1) : Y_i = \beta_1 + \beta_2 \exp(\beta_3/w_i) + \xi_i;$$

M_1

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

Cont.

- Cont.

$$(M_5) : Y_i = \beta_1 + \beta_2 w_i - e^{\beta_3 + \beta_4 w_i} + \xi_i;$$

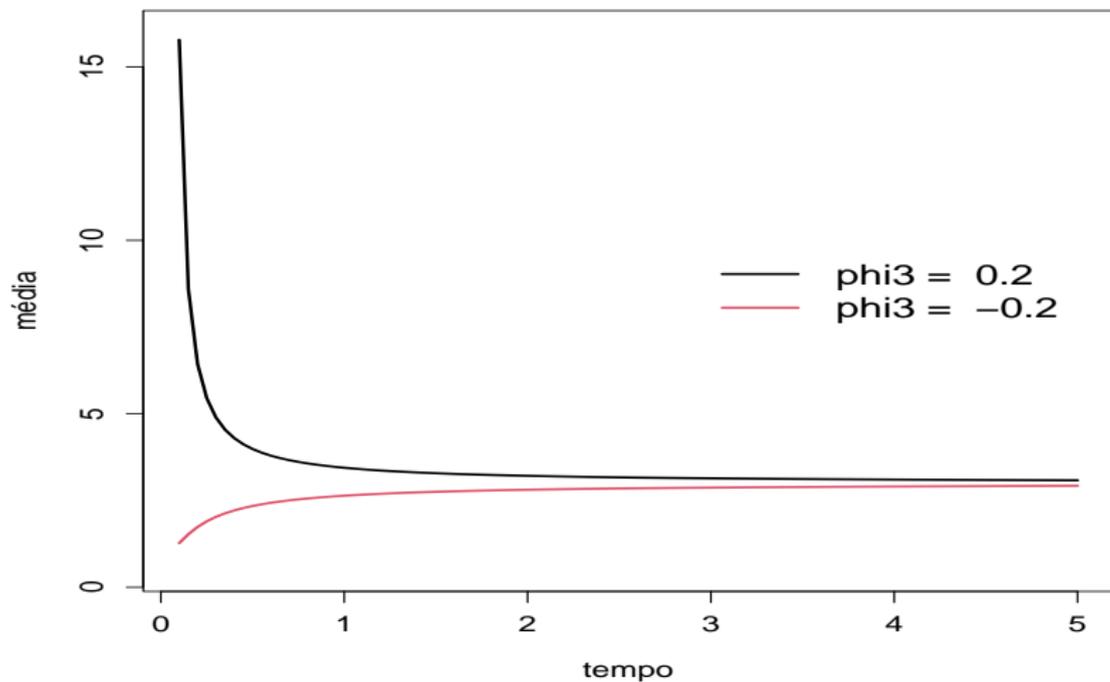
M_5

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}; \mathbf{X}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & w_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & w_i \end{bmatrix}$$

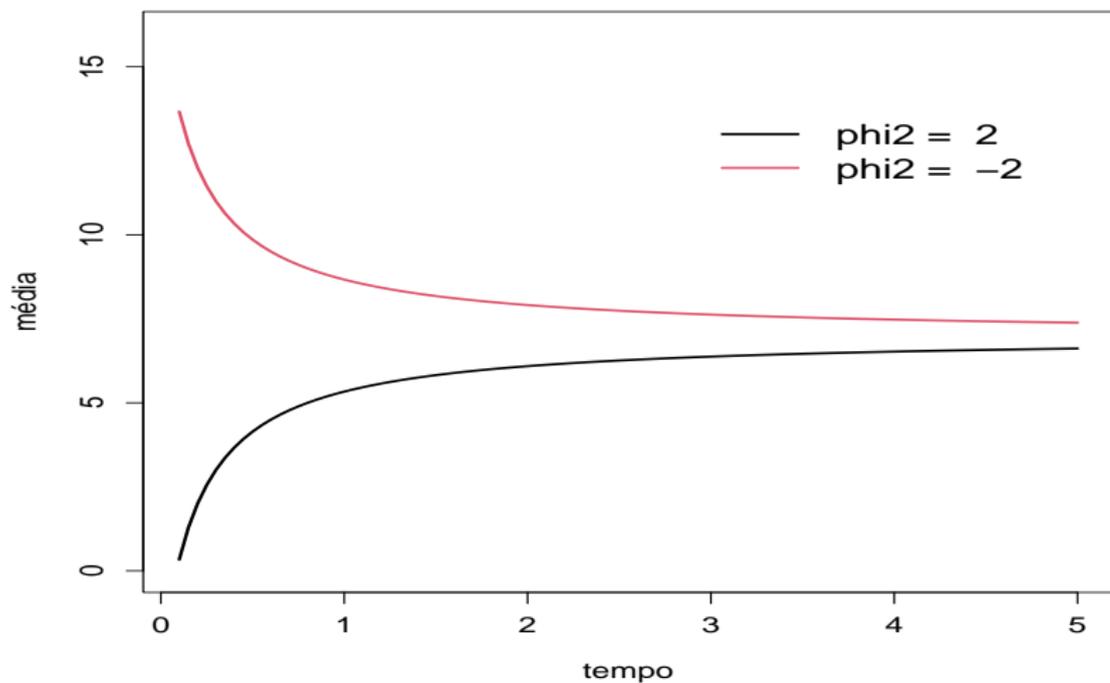
Interpretação modelo 3

- $\phi_1 = \mathcal{E}(Y_i | w_i = 1)$
- Se $w_i > 0$, ϕ_2 regula a taxa de crescimento (se positivo) ou de decaimento (se negativo) da resposta esperada em função de w_i .
- As interpretações dos parâmetros, muitas vezes, têm de ser estabelecidas conforme seu sinal e/ou os das variáveis explicativas.

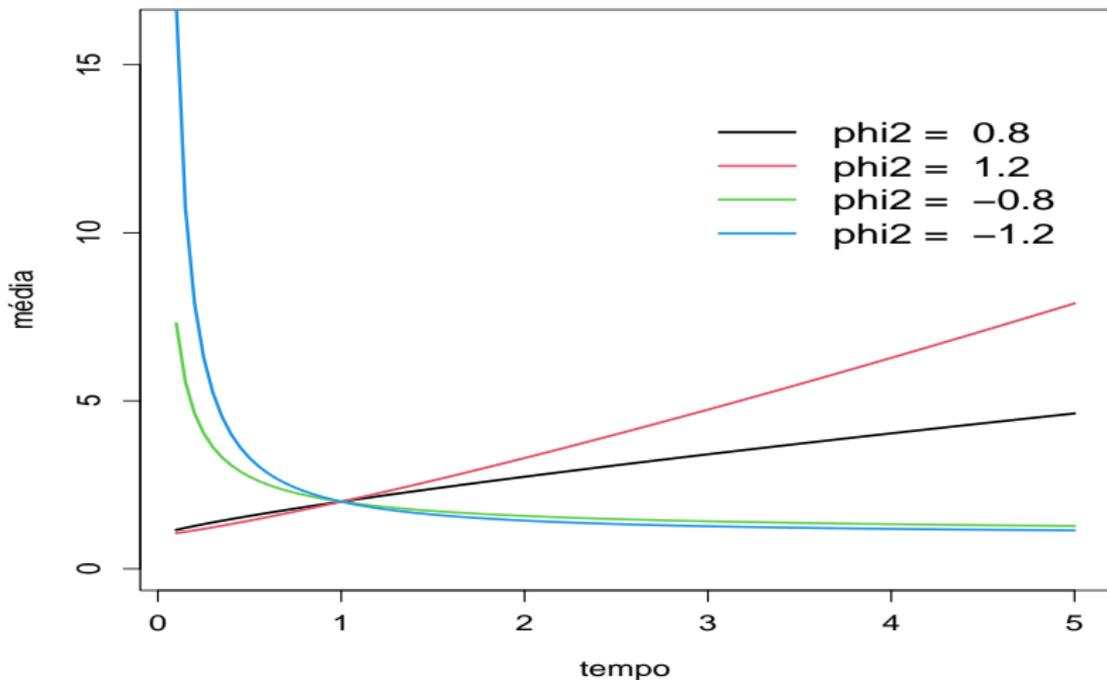
Modelo 1: $\phi_1 = 1, \phi_2 = 2$



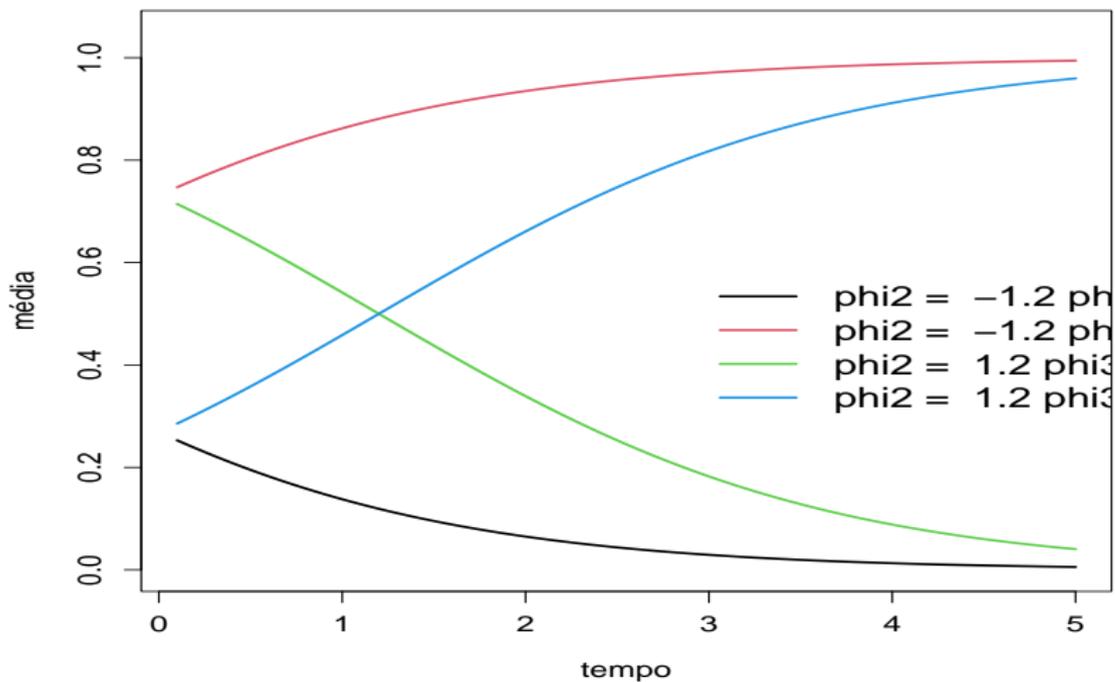
Modelo 2: $\phi_1 = 7, \phi_3 = 0,2$



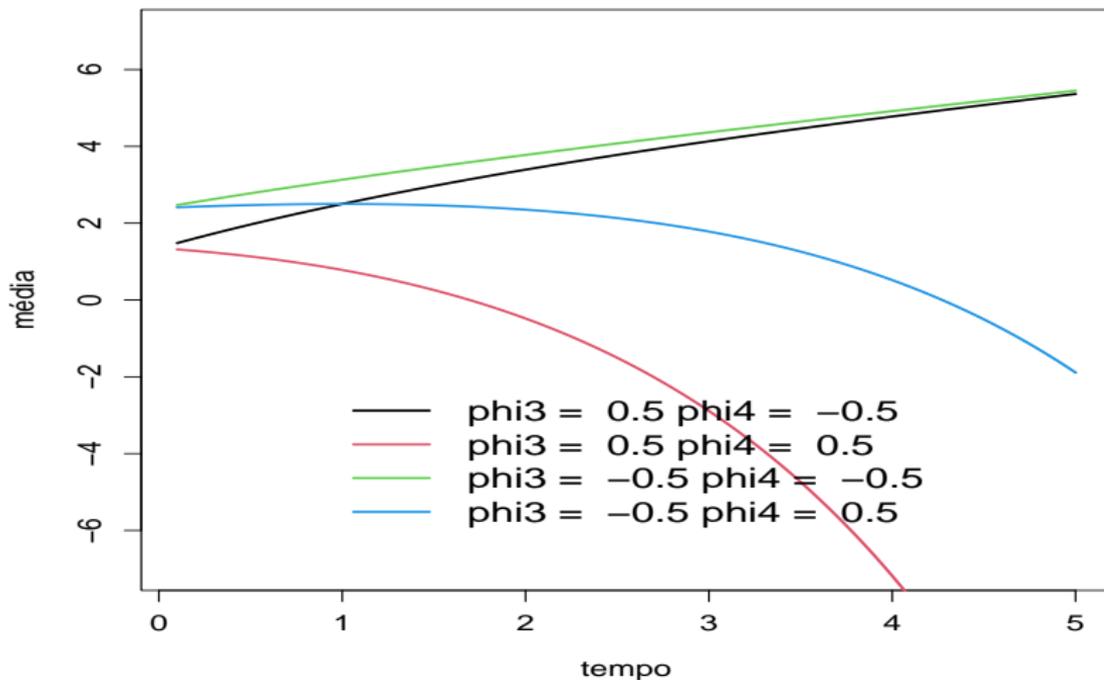
Modelo 3: $\phi_1 = 1$



Modelo 4: $\phi_1 = 1$



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



Estimação

- Parâmetros de interesse: $(\boldsymbol{\beta}, \sigma^2)'$.
- Mínimos quadrados não-lineares ($\mu_i = f(\phi_i, \mathbf{X}_i, \mathbf{W}_i)$). Minimizar a seguinte soma de quadrados (dos erros)

$$Q(\boldsymbol{\beta}) = \sum_{j=1}^n (Y_j - \mu_j)^2.$$

- Devemos derivar $Q(\boldsymbol{\beta})$ com relação à cada componente de $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, gerando $\mathbf{S}(\boldsymbol{\beta})$, e resolver o sistema oriundo de $\mathbf{S}(\tilde{\boldsymbol{\beta}}) = \mathbf{0}$.

Estimação

- O sistema de equações resultante será não linear e, portanto, algum algoritmo de otimização deverá ser utilizado.
- Neste caso, algoritmos como Gauss-Newton, BFGS e Nelder-Meder podem ser preferíveis aos usuais Newton-Raphson e Escore de Fisher.
- Particularmente, a escolha de valores iniciais para as utilização das rotinas de otimização, requer bastante cuidado.

Estimação

- Uma vez estimado β , podemos estimar σ^2 através de

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n (Y_i - \hat{\mu}_i)^2,$$

em que $\hat{\mu}_i = f(\hat{\phi}_i, \mathbf{X}_i, \mathbf{W}_i)$

Inferência

- Sob certas **condições de regularidade** e para n suficientemente grande, o estimador de mínimos quadrados não lineares de β é tal que

$$\hat{\beta} \approx N_p(\beta, Cov(\beta)),$$

em que $Cov(\beta)$ depende de algumas formas matriciais (e também de σ^2).

- Portanto, todos os resultados inferenciais (intervalos de confiança e testes de hipótese) como apresentados para os MRNLH, continuam válidos com as devidas adaptações (e, essencialmente, são todos assintóticos).

Exemplo: Modelo 3

- $Q(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \beta_1 x_i^{\beta_2})^2$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$.
- Componentes do veto escore $\mathbf{S}(\boldsymbol{\beta}) = \begin{bmatrix} S(\beta_1) \\ S(\beta_2) \end{bmatrix}$:
$$S(\beta_1) = -2 \sum_{i=1}^n (y_i - \beta_1 x_i^{\beta_2}) x_i^{\beta_2}.$$
$$S(\beta_2) = -2 \sum_{i=1}^n (y_i - \beta_1 x_i^{\beta_2}) \beta_1 x_i^{\beta_2} \ln x_i.$$

Exemplo: Modelo 3

■ Matriz Hessiana: $\mathbf{H}(\boldsymbol{\beta}) = \begin{bmatrix} H(\beta_1, \beta_1) & H(\beta_1, \beta_2) \\ H(\beta_2, \beta_1) & H(\beta_2, \beta_2) \end{bmatrix}$

$$H(\beta_1, \beta_1) = 2 \sum_{i=1}^n x_i^{2\beta_2},$$

$$H(\beta_2, \beta_2) = -2\beta_1 \sum_{i=1}^n (y_i - 2\beta_1 x_i^{\beta_2}) x_i^{\beta_2} (\ln x_i)^2,$$

$$H(\beta_1, \beta_2) = -2 \sum_{i=1}^n (y_i - 2\beta_1 x_i^{\beta_2}) x_i^{\beta_2} \ln x_i.$$

Exemplo: Modelo 3

■ Informação de Fisher: $I(\boldsymbol{\beta}) = \begin{bmatrix} I(\beta_1, \beta_1) & I(\beta_1, \beta_2) \\ I(\beta_2, \beta_1) & I(\beta_2, \beta_2) \end{bmatrix}$

$$I(\beta_1, \beta_1) = -2 \sum_{i=1}^n x_i^{2\beta_2},$$

$$I(\beta_2, \beta_2) = -2\beta_1 \sum_{i=1}^n x_i^{2\beta_2} (\ln x_i)^2,$$

$$I(\beta_1, \beta_2) = -2 \sum_{i=1}^n \beta_1 x_i^{2\beta_2} \ln x_i.$$

Análise residual

- Erro padronizado: $\xi_i^* = \frac{Y_i - \mu_i}{\sigma^2}$. Sob o bom ajuste do modelo $\xi_i^* \stackrel{iid}{\sim} N(0, 1)$.
- Resíduo padronizado: $R_i = \frac{Y_i - \hat{\mu}_i}{\hat{\sigma}^2}$. Sob o bom ajuste do modelo e para n suficiente grande, $R_i \stackrel{iid}{\approx} N(0, 1)$.
- Podemos utilizar os cinco gráficos usuais: resíduo \times índice, resíduo \times valores preditos, histograma, box-plot e gráfico de envelopes.
- É possível também construir gráficos de envelopes simulados sob o modelo, para R_i .
- Outros resíduos podem ser considerados.

Recursos computacionais

- A função “`nls`” (default no R) ajusta modelos de regressão não lineares via mínimos quadrados (ponderados) não lineares.
- Um ponto interessante é que ela solicita que o usuário forneça estimativas iniciais para os parâmetros do preditor não linear.
- Utiliza (como default) o algoritmo de Gauss-Newton (trabalha apenas com as primeiras derivadas).
- Permite a utilização de pesos para cada observação (mínimos quadrados não lineares ponderados, que equivale a minimizar

$$Q^*(\beta) = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{w_i}, \text{ em que } w_i \text{ são os pesos}.$$

Cont.

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

Cont.

- Há vários preditores não lineares implementados já implementados nesse pacote (derivada e estimativas iniciais) (abaixo - *input*, *x*, *xmid*, : covariável).

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	$Dose * \exp(IKe + IKa - ICl) * (\exp(-\exp(IKe) * input) - \exp(-\exp(IKa) * input)) * (\exp(IKa) - \exp(IKe))$

Cont.

- (x : covariável)

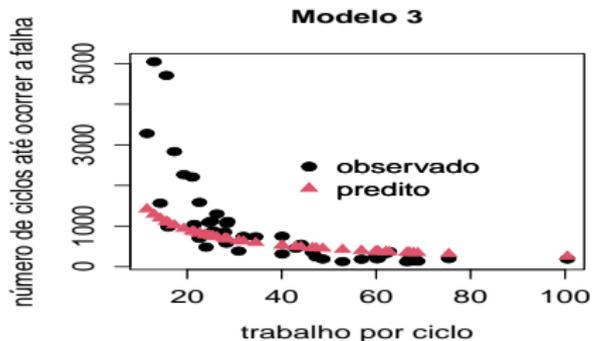
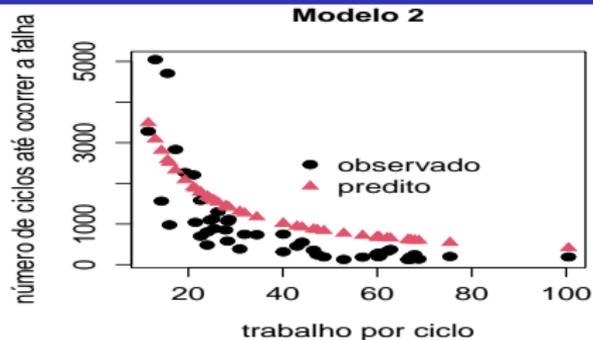
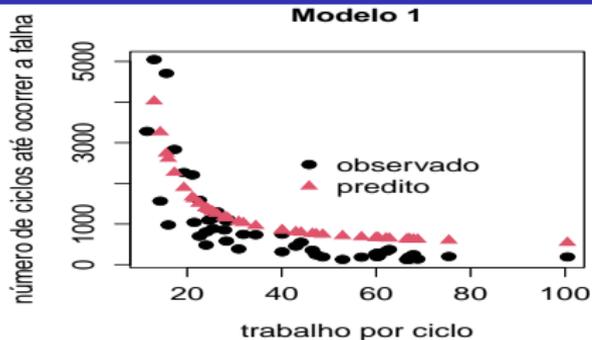
Nome	Modelo
SSfpl	$A + (B - A)/(1 + \exp((xmid - input)/scal))$
SSgompertz	$Asym * \exp(-b2 * b3^x)$
SSlogis	$Asym/(1 + \exp((xmid - input)/scal))$
SSmicmen	$Vm * input/(K + input)$
SSweibull	$Asym - Drop * \exp(-\exp(lrc) * x^{pwr})$

- As funções “**nlsr**” e “**nlmrt**” apresentam alternativas para o ajuste dos MRNNLH, no sentido de disponibilizarem algoritmos numéricos mais estáveis.

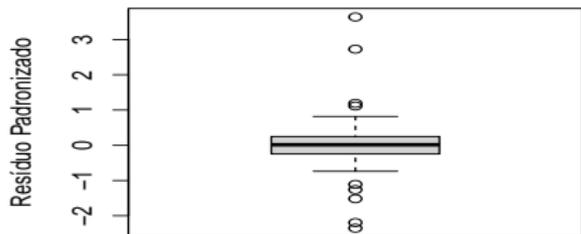
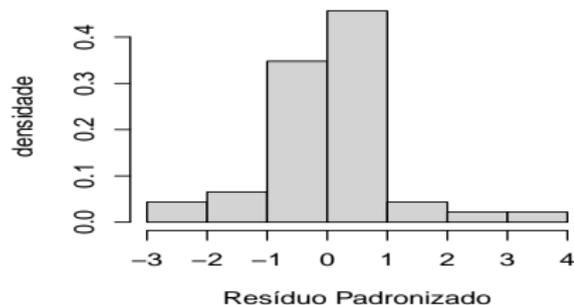
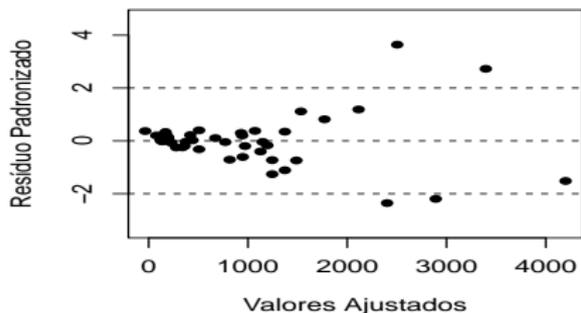
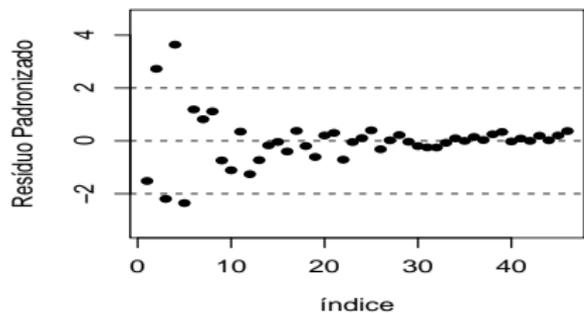
Voltando ao Exemplo 16

- Ajustou-se os modelos M_1 , M_2 e M_3 (não houve convergência para os modelo M_4 e M_5), ou seja:
- $(M_1) : Y_i = \beta_1 + \beta_2 \exp(\beta_3/x_i) + \xi_i$.
- $(M_2) : Y_i = \beta_1 - \beta_2 (x_i + \beta_3)^{-1} + \xi_i$.
- $(M_3) : Y_i = \beta_1 x_i^{\beta_2} + \xi_i$.
- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.

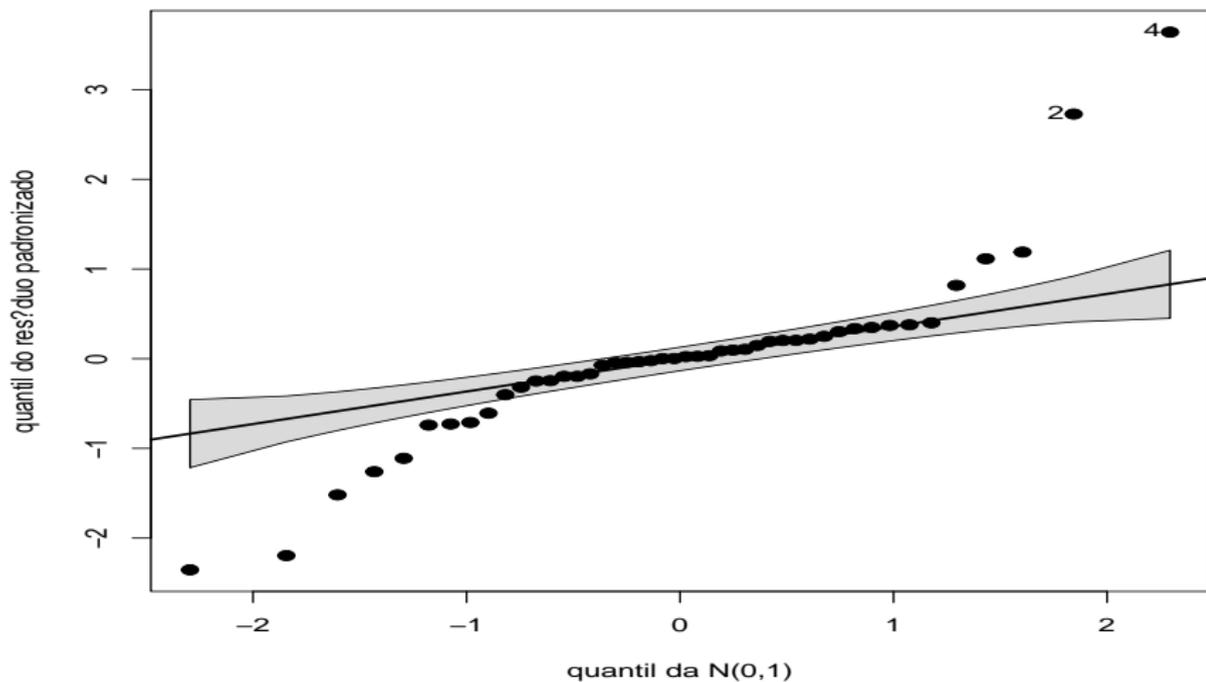
Valores observados e preditos (“chutes” iniciais)



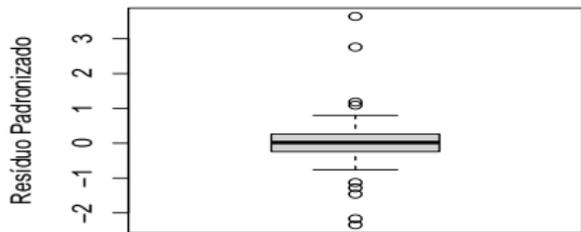
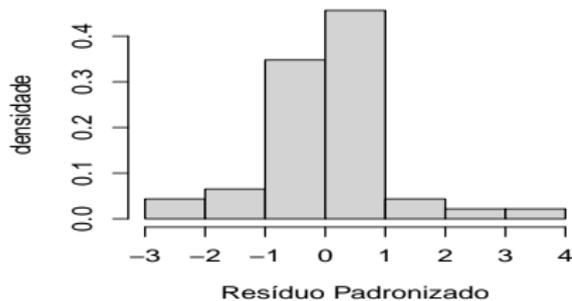
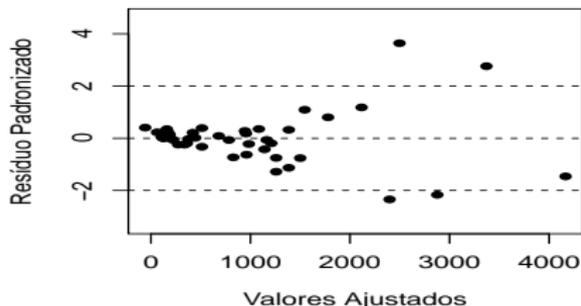
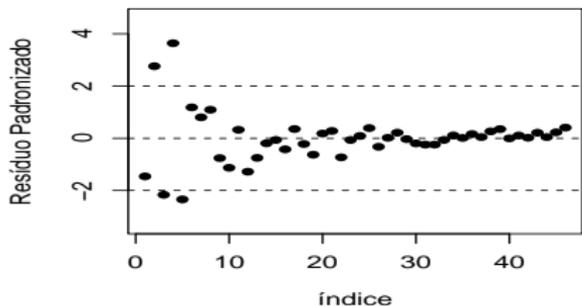
Gráficos de diagnóstico: modelo 1



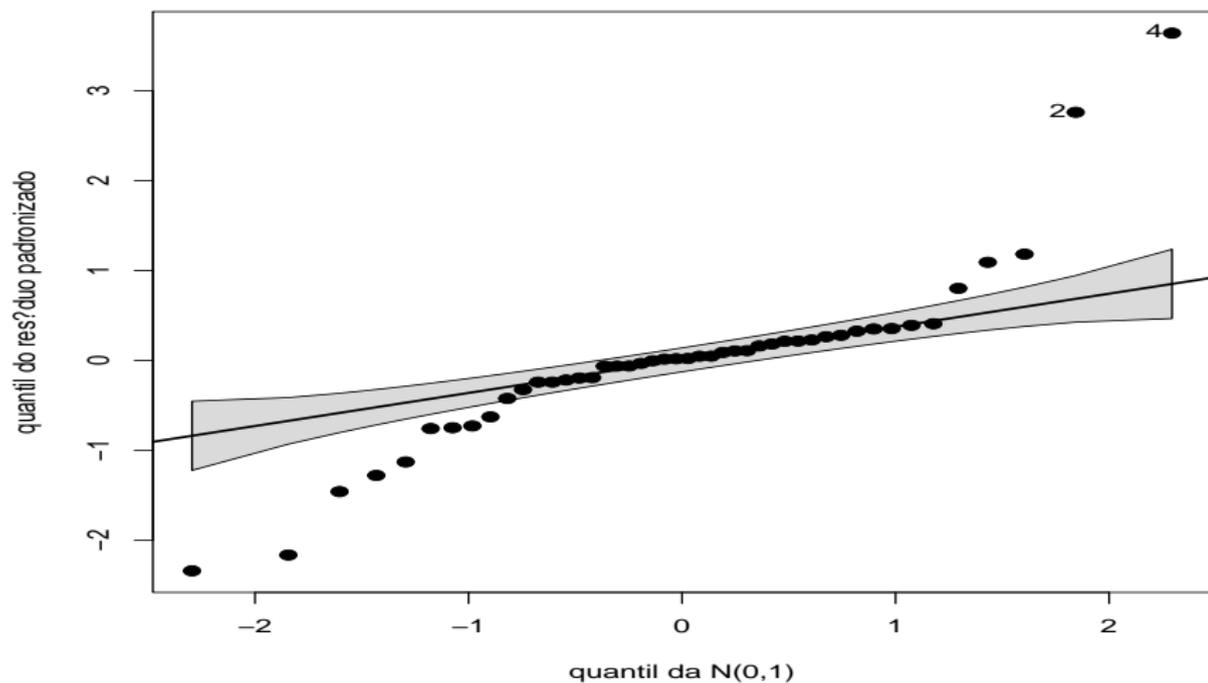
Gráficos de envelopes: modelo 1



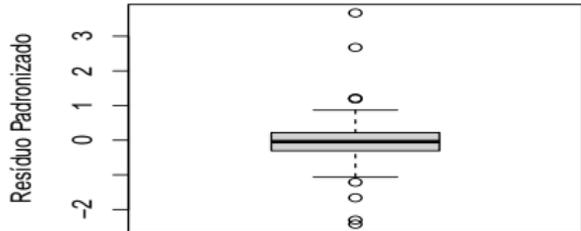
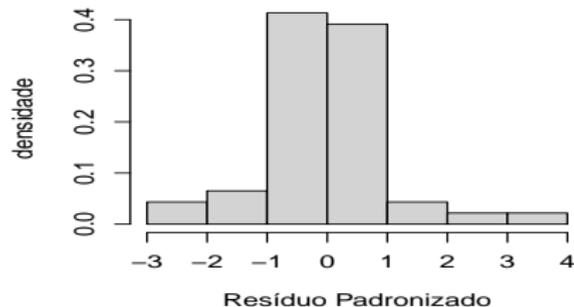
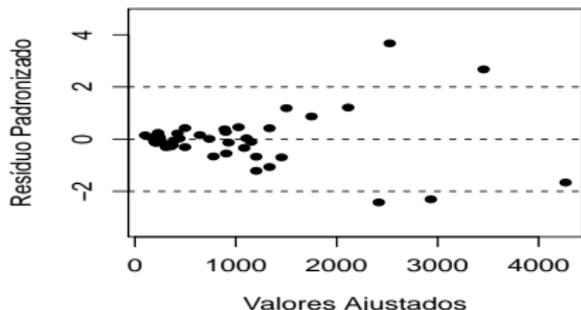
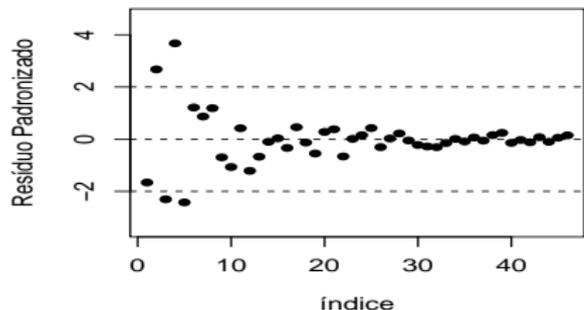
Gráficos de diagnóstico: modelo 2



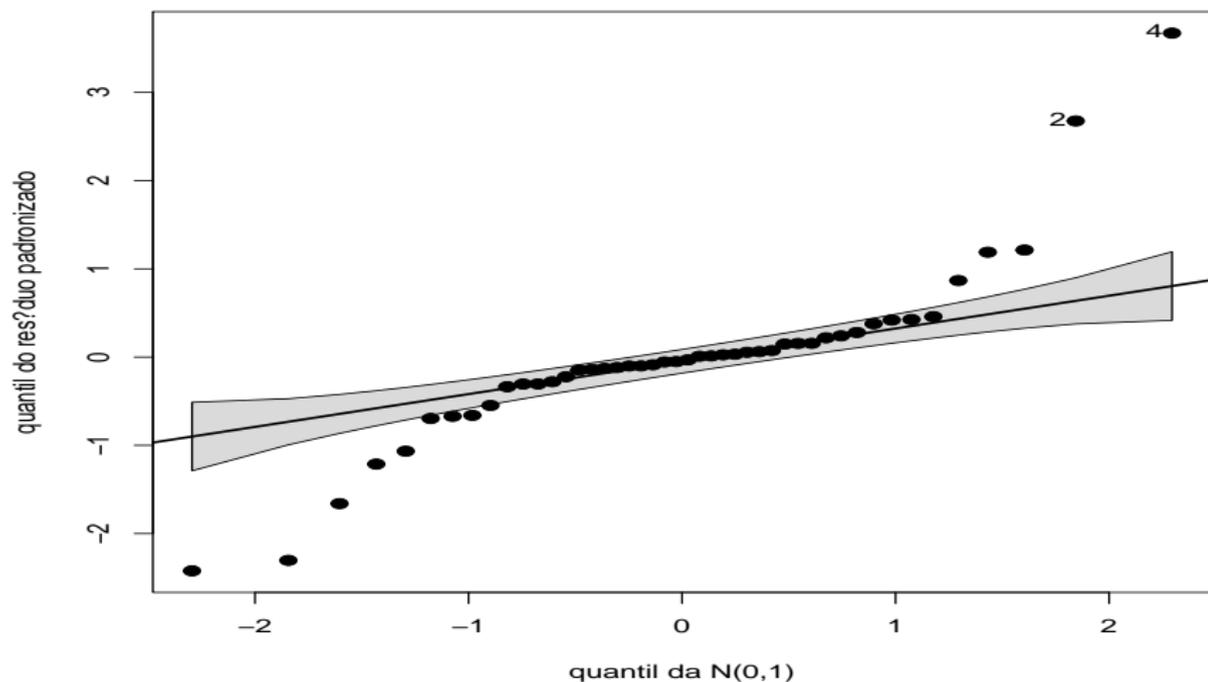
Gráficos de envelopes: modelo 2



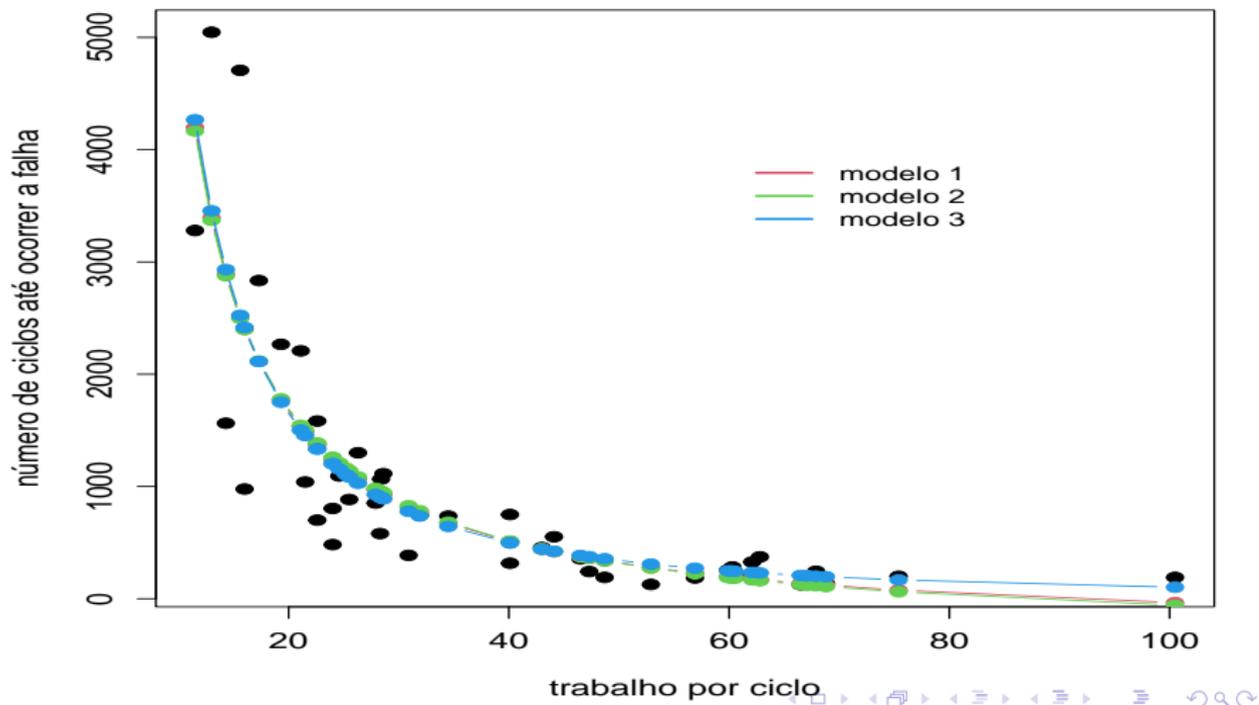
Gráficos de diagnóstico: modelo 3



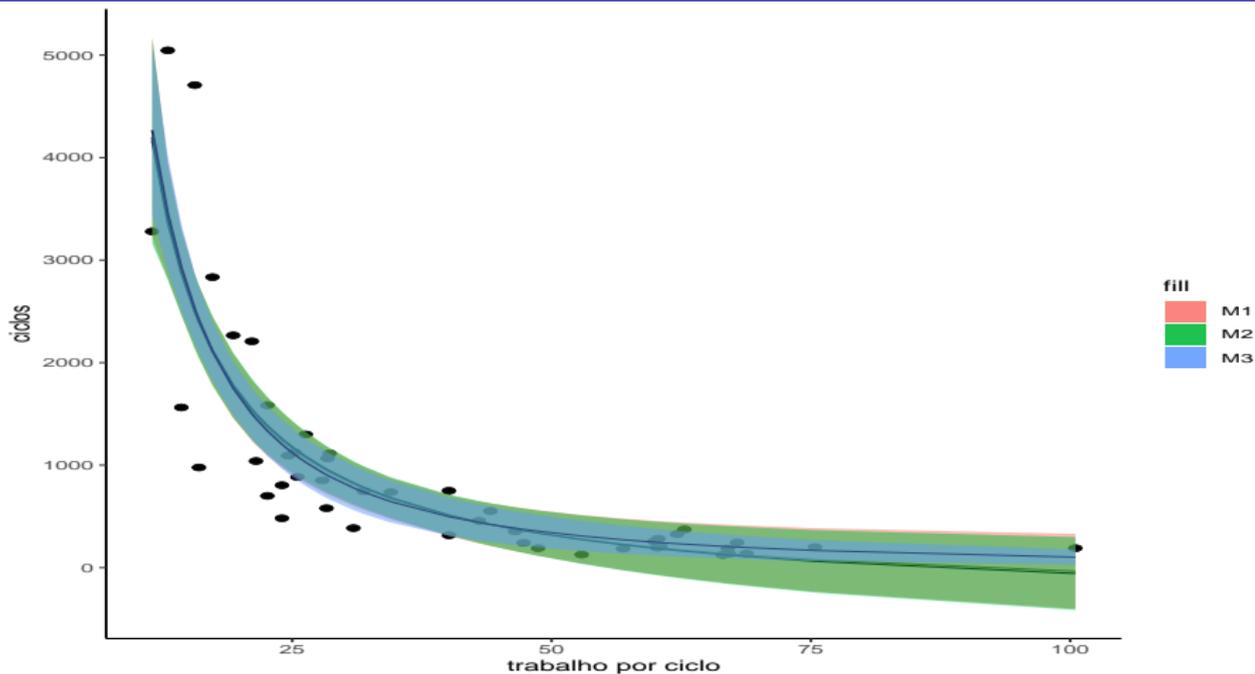
Gráficos de envelopes: modelo 3



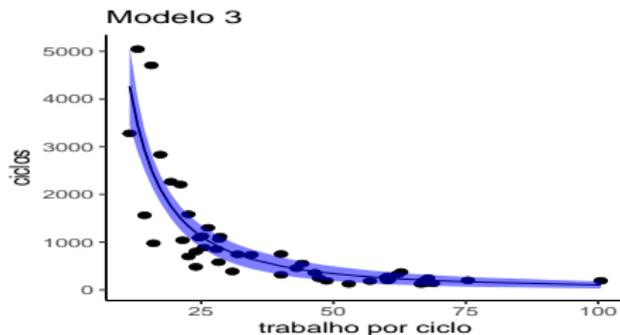
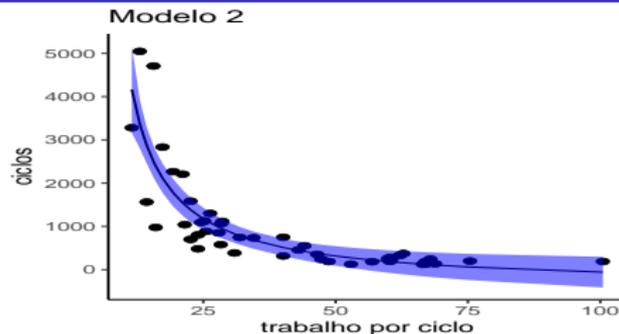
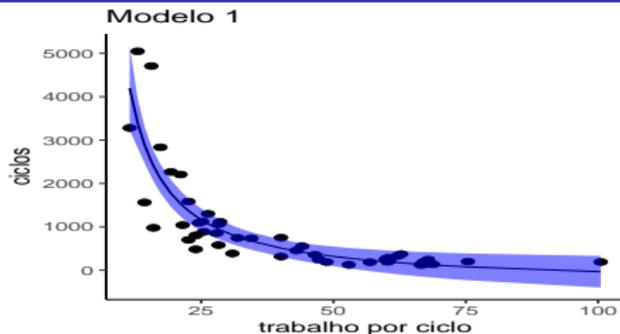
Valores observados e preditos



Valores observados e preditos com IC's



Valores observados e preditos com IC's



Critérios de Informação (CI)

CI	modelo 1	modelo 2	modelo 3
AIC	724,76	724,98	722,16
BIC	732,07	732,29	727,64
AICc	725,33	725,55	722,43
SABIC	718,84	719,06	717,54
HQCIC	724,81	725,03	721,53
CAICe	731,24	731,46	725,81
-2logver	716,76	716,98	716,16

Comentários

- Há indícios de heterocedasticidade e caudas pesadas nos resíduos.
- Os três modelos apresentaram razoáveis previsões mas, aparentemente, outro preditor deve ser considerado.
- Sugestão: modelos com caudas mais pesadas, heterocedásticos (assimétricos): modelos “**gamlss**” (no pacote **gamlss** pode-se usar a função “**nl**”) - erros heterocedásticos com distribuição t de Student.
- Veja uma abordagem Bayesiana nesse sentido [aqui](#).

Estimativas e Inferência

Par.	Est.	EP	IC(95%)	Est. Z	p-valor
α	285.627,86	155.748,31	[-19.633,21 ; 590.888,94]	1,834	0,0734
β	-1,721	0,193	[-2,100 ; -1,343]	-8,912	< 0,0001

Note que $\mathcal{E}(Y_i|x_i) = \beta_1 x_i^{\beta_2}$, assim, $\mathcal{E}(Y_i|x_i = 1) = \beta_1$ e $(\forall a > 0)$

$$\mathcal{E}(Y_i|x_i^* = ax_i) = \beta_1 x_i^{\beta_2} a^{\beta_2} = \mathcal{E}(Y_i|x_i) a^{\beta_2} \rightarrow \frac{\mathcal{E}(Y_i|x_i^* = ax_i)}{\mathcal{E}(Y_i|x_i^* = x_i)} = a^{\beta_2}$$

Estimativas e Inferência

- Assim, o valor esperado do número de ciclos para um trabalho por ciclo igual a 1 é $\approx 285.627,86 [0;590.888,94]$.
- Por outro lado, a taxa de decrescimento do número de ciclos, para um trabalho por ciclo igual a a é $\approx a^{-1,721} [a^{-2,100}; a^{-1,343}]$.
- Exercício, encontra o IC acima através do método Delta.