

# Modelos de regressão normais não lineares

Prof. Caio Azevedo

# Comentários

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

# Limitações dos MRNLH e vantagens dos MNL

- As interpretações dos parâmetros de modelos não lineares podem ser mais adequadas do que aquelas associadas aos parâmetros do MRNLH (particularmente, polinômios de grau  $\geq 3$  apresentam parâmetros de difícil interpretação), principalmente em termos do problema.
- Há modelos teóricos que explicam satisfatoriamente certos fenômenos, os quais se baseiam em soluções de equações diferenciais. Estas, por sua vez, podem induzir estruturas não-lineares.

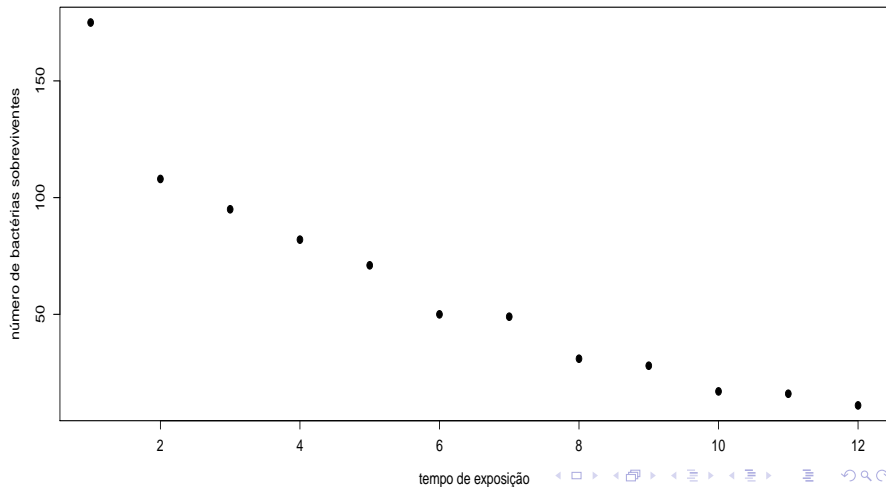
# Limitações dos MRNLH e vantagens dos MNL

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

## Exemplo 12: 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.
- 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.

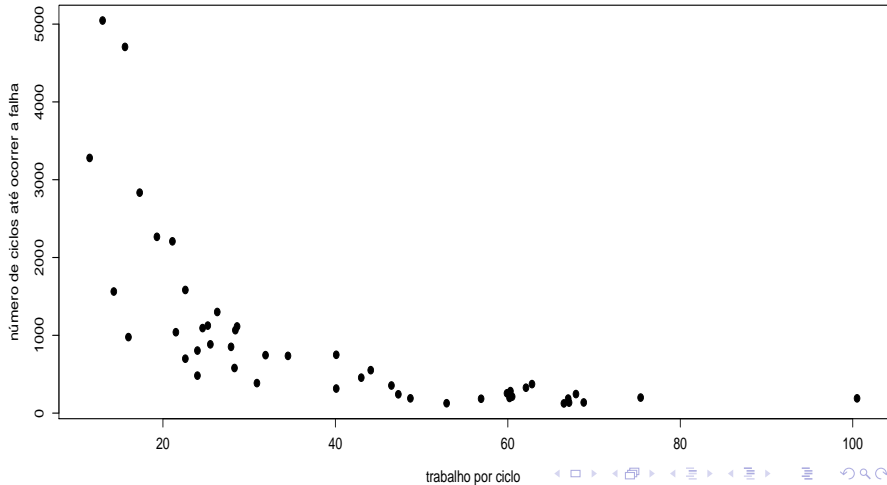
# Gráfico de dispersão



## Exemplo 13: 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 ( $\text{mJ} / \text{m}^3$ ) (intensidade do stress).

# Gráfico de dispersão

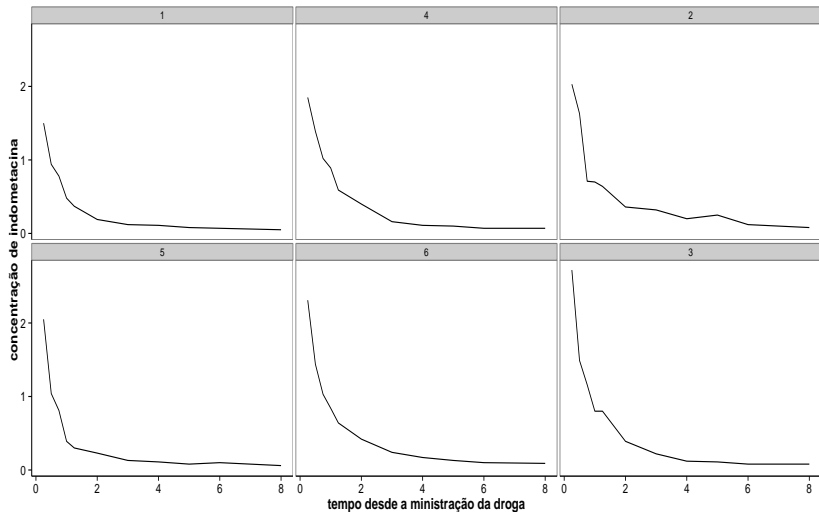




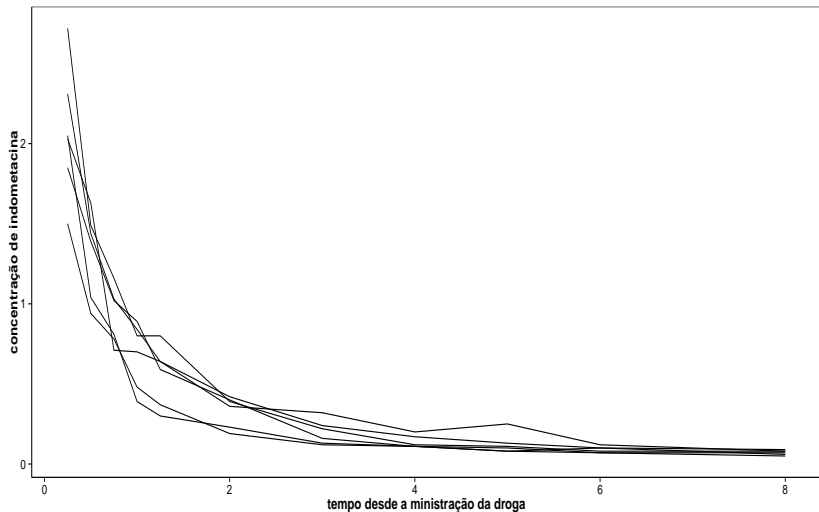
## Exemplo 14: cinética de indometacina, Kwan et al (1976)

- 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.

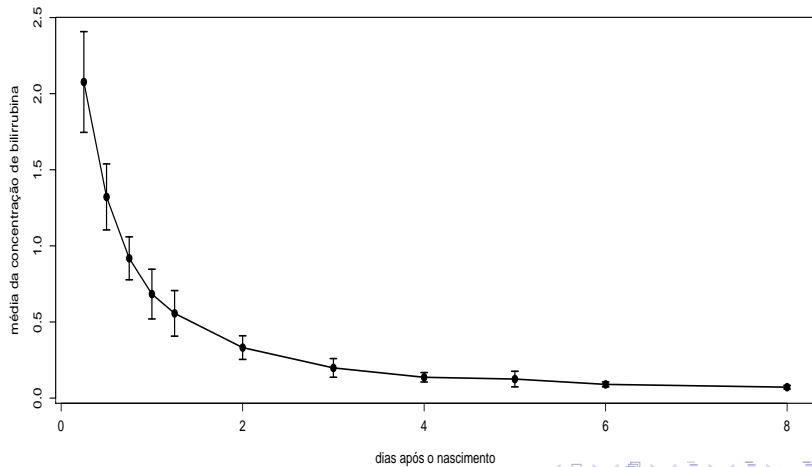
## Perfis individuais (separados)



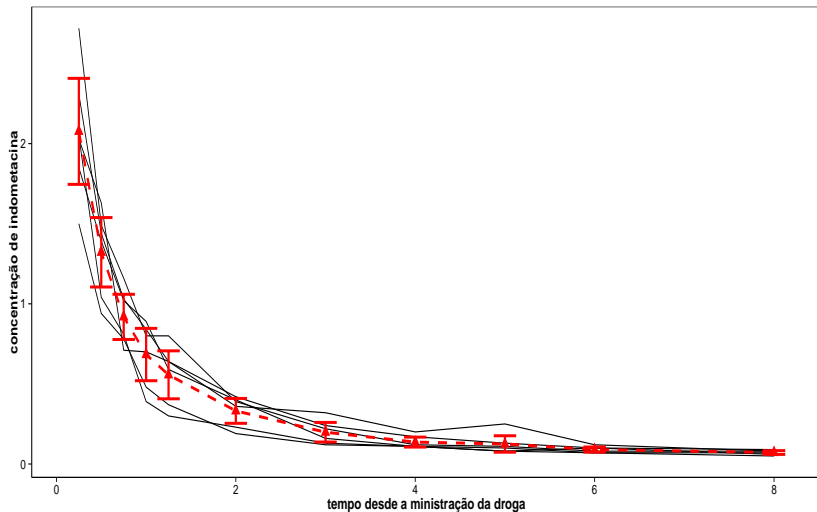
## Perfis individuais (juntos)



# Perfis médios



# Perfis médios e individuais



# 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  $\phi_i$ .
- $\xi_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ .

## Modelo não lineares: exemplos

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

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.

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

$$(\phi_i = \mathbf{X}_i\boldsymbol{\beta}), (M_1) : Y_i = \beta_1 + \beta_2 \exp(\beta_3/w_i) + \xi_i; (M_5) : Y_i = \beta_1 + \beta_2 w_i - e^{\beta_3 + \beta_4 w_i} + \xi_i.$$

$M_1$

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

$M_5$

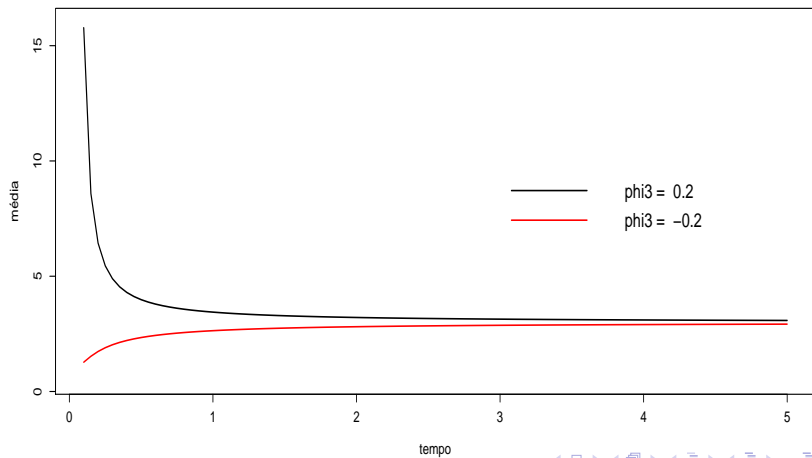
$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}; \mathbf{X}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & w_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & w_i \end{bmatrix}$$



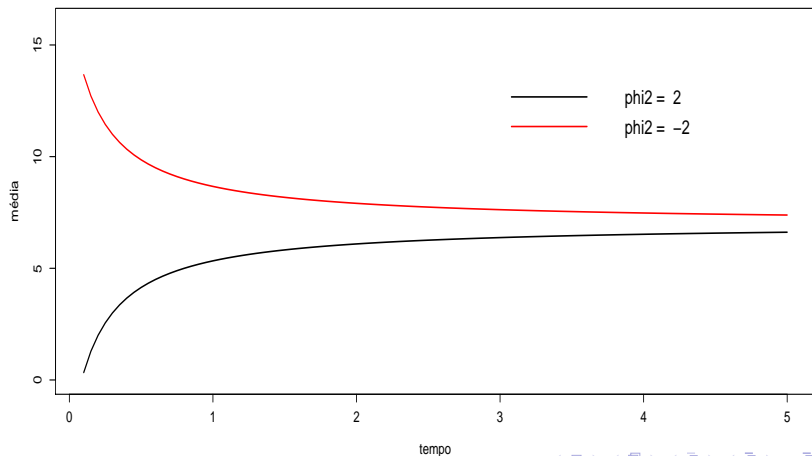
## Interpretação modelo 3

- $\phi_1 = \mathcal{E}(Y_i | w_i = 1)$
- Se  $w_i > 0$ ,  $\phi_2$  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.

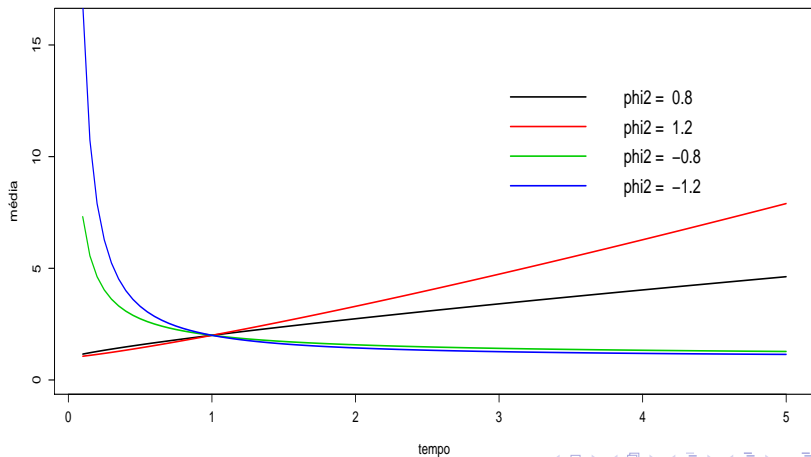
# Modelo 1: $\phi_1 = 1, \phi_2 = 2$



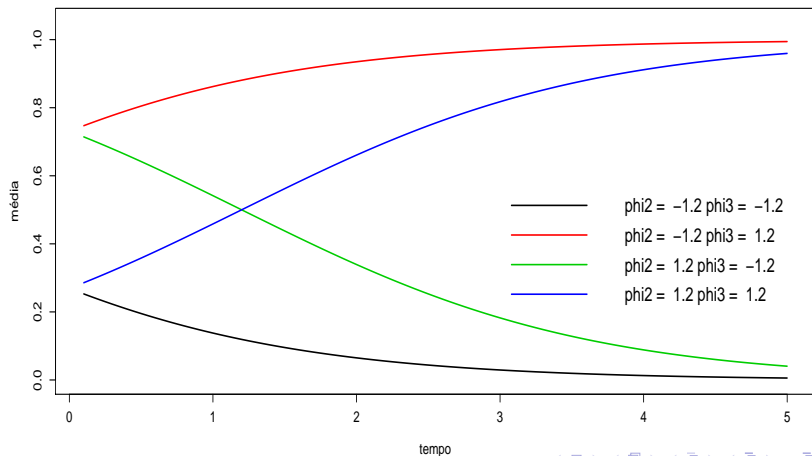
## Modelo 2: $\phi_1 = 7, \phi_3 = 0,2$



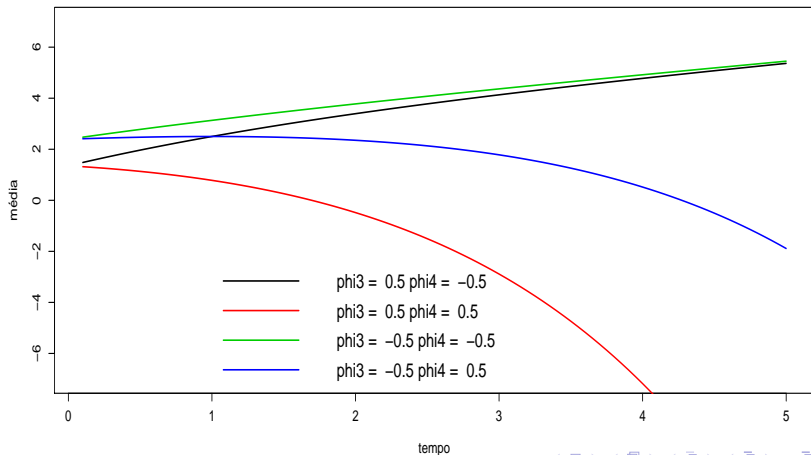
# Modelo 3: $\phi_1 = 1$



## 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)

$$Q(\boldsymbol{\beta}) = \sum_{j=1}^n (Y_i - \mu_i)^2$$

- Devemos derivar 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

- Os 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.
- Uma vez estimado  $\beta$ , podemos estimar  $\sigma^2$  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  $\beta$  é tal que

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

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

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

## 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:  $I(\beta) = \begin{bmatrix} I(\beta_1, \beta_1) & I(\beta_1, \beta_2) \\ I(\beta_2, \beta_1) & I(\beta_2, \beta_2) \end{bmatrix}$

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

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

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

# Análise residual

- Erro padronizado:  $\xi_i^* = \frac{Y_i - \mu_i}{\sigma^2}$ . Sob o bom ajuste do modelo  $\xi^* \sim N(0, 1)$ .
- Resíduo padronizado:  $R_i = \frac{Y_i - \hat{\mu}_i}{\hat{\sigma}^2}$ . Sob o bom ajuste do modelo e para  $n$  suficiente grande,  $R_i \approx N(0, 1)$ .
- Podemos utilizar os quatro gráficos usuais: resíduo  $\times$  índice, resíduo  $\times$  valores preditos, histograma e gráfico de envelopes.

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

## 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, também, a possibilidade de se criar modelos 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	$Dose * \exp(IKe + IKa - ICl) * (\exp(-\exp(IKe) * input) - \exp(-\exp(IKa) * input)) * (\exp(IKa) - \exp(IKe))$



## Cont.

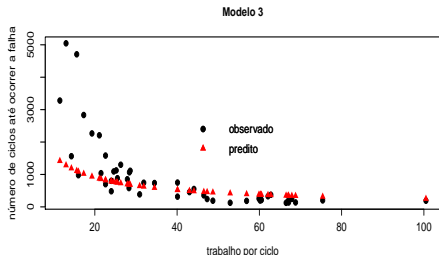
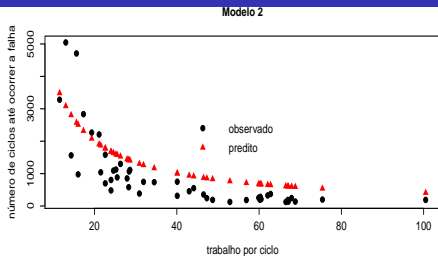
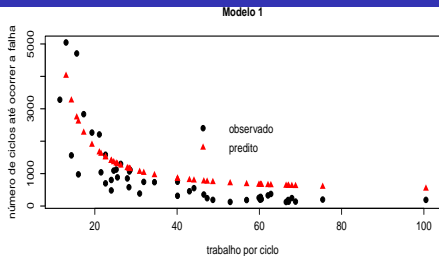
- ( $x$  : covariável)

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

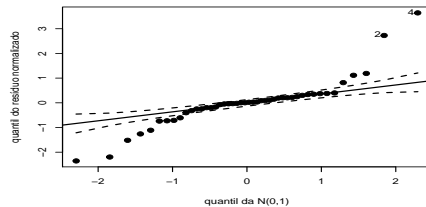
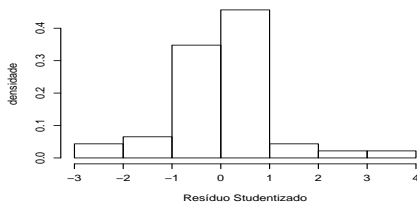
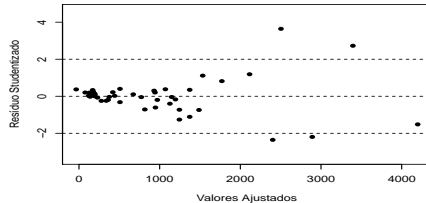
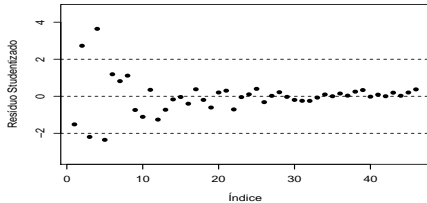
## Voltando ao Exemplo 13

- Ajustou-se os modelos  $M_1$ ,  $M_2$  e  $M_3$  (não houve convergência para os modelo  $M_4$  e  $M_5$ )
- $(M_1) : Y_i = \beta_1 + \beta_2 \exp(\beta_3/x_i) + \xi_i.$
- $(M_2) : Y_i = \beta_1 - \phi_2 (x_i + \phi_3)^{-1} + \xi_i.$
- $(M_3) : Y_i = \beta_1 x_i^{\phi_2} + \xi_i.$
- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2).$

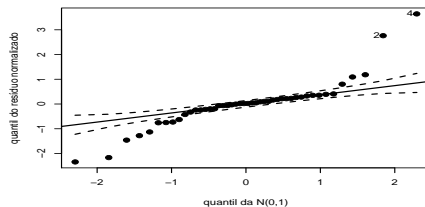
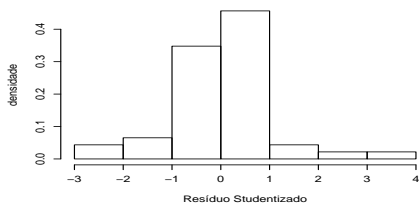
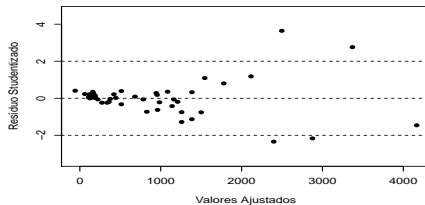
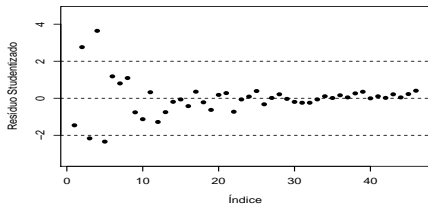
# Valores observados e preditos (para curvas empíricas)



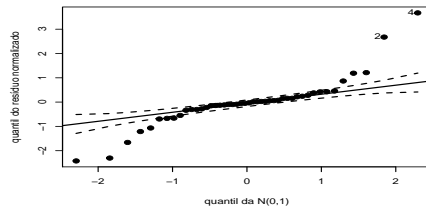
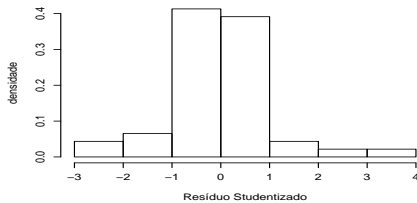
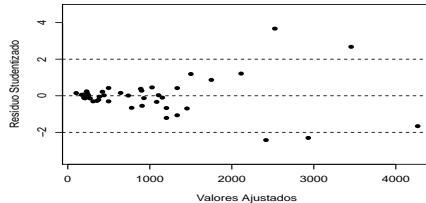
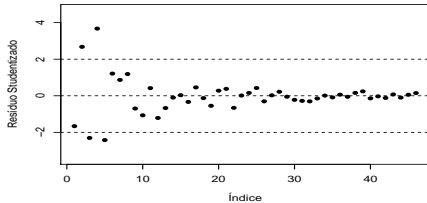
# Gráficos de diagnóstico: modelo 1



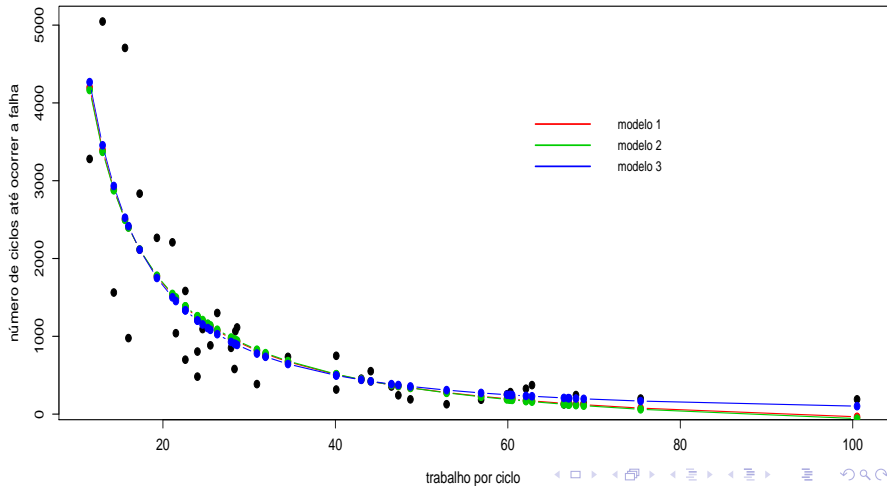
# Gráficos de diagnóstico: modelo 2



# Gráficos de diagnóstico: modelo 3



# Valores observados e preditos



# 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 caudades mais pesadas, heterocedásticos (assimétricos).