

INTRODUÇÃO AO MODELO DE REGRESSÃO NÃO LINEAR





INTRODUÇÃO

- Os modelos de regressão apresentam, de modo geral, a seguinte forma:

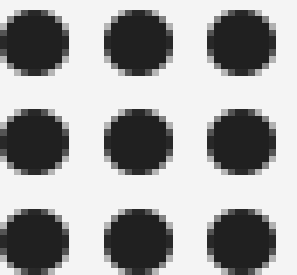
$$Y = f(X, \theta) + \varepsilon$$

Sendo:

- $f(X, \theta)$ a função das variáveis preditoras X e parâmetros θ a serem estimados, e ε são os erros aleatórios.




- Para variáveis biológicas, modelos lineares raramente explicam a relação entre as variáveis. Essas são, portanto, melhores descritas por **modelos não lineares**.





Modelos de Regressão Não Linear

- Modelos não lineares são caracterizados por funções $f(\mathbf{X}, \boldsymbol{\theta})$ não lineares em pelo menos um dos parâmetros θ_i .

- 
- Existem diversos modelos de regressão não lineares. Eles possuem finalidades específicas. Alguns dos mais conhecidos são:
 - Modelo de Mitscherlich;
 - Modelo Logístico;
 - Modelo de Gompertz;
 - Modelo de Michaelis-Menten.

Modelo de Mitscherlich

- Esse modelo tem como uso principal o estudo da variação do crescimento de vegetais ou produção de determinada cultura (Y) em função da quantidade de nutrientes fornecida (X). Ele é dado pela seguinte função:

$$y = \alpha(1 - 10^{-\gamma(x+\theta)}) + \varepsilon$$

Modelo Logístico

- É um modelo utilizado para crescimento de populações humanas, animais e vegetais. De modo usual $\theta > 0$ e $\beta > 0$, em que a função logística aparenta uma curva sigmoide.

$$y = \frac{\theta}{1 + e^{-(\alpha + \beta x)}} + \varepsilon$$

Modelo de Gompertz

- Modelo utilizado para descrever crescimentos, onde este é proporcional não ao tamanho da população, mas ao logaritmo deste tamanho. É dado pela função:

$$y = e^{\alpha + \beta \rho x} + \varepsilon$$

Em que $\beta < 0$ e $0 < \rho < 1$, e a função de Gompertz $\exp(\alpha + \beta \rho x)$ é monotonicamente crescente.

Modelo de Michaelis-Menten

- Modelo utilizado para estudos da taxa de reações químicas catalizadas por enzimas. É dado pela função:

$$y = \frac{\theta_1 x}{x + \theta_2} + \varepsilon$$



Dados de Secagem

- Além dos citados anteriormente, existem diversos outros modelos para diversos fins.
- No que diz respeito a dados de secagem de plantas, por exemplo, uma série de modelos são utilizados na literatura (Tabela 1).



Dados de Secagem

Tabela 1: Modelos de Regressão não-linear utilizados na elaboração da curva de perda de massa (y) de flores em relação ao tempo de desidratação (x).

Designação do modelo	Modelo*	Equação
Page	$y = \exp(-k \cdot x^n)$	(1)
Page Modificado	$y = \exp[-(k \cdot x)^n]$	(2)
Henderson e Pabis	$y = a \cdot \exp(-k \cdot x)$	(3)
Midilli et al.	$y = a \cdot \exp(-k \cdot x^n) + b \cdot x$	(4)
Wang e Sing	$y = 1 + a \cdot x + bx^2$	(5)
Newton	$y = \exp(-k \cdot x)$	(6)
Logarítmico	$y = a \cdot \exp(-k \cdot x) + c$	(7)
Exponencial de dois termos	$y = a \cdot \exp(-k \cdot x) + (1 - a) \exp(-k \cdot a \cdot x)$	(8)
Dois termos	$y = a \cdot \exp(-k_0 \cdot x) + b \cdot \exp(-k_1 \cdot x)$	(9)
Thompson	$x = a \cdot \ln(y) + b \cdot [\ln(y)]^2$	(10)

*As constantes k , n , a , b , c , k_0 e k_1 são parâmetros dos modelos.

Estimação de parâmetros

- Como visto, os modelos de regressão possuem parâmetros, os quais devem ser estimados. Para isso, pode-se usar o Método de Estimação por Mínimos Quadrados Ordinários.

Estimação por Mínimos Quadrados Ordinários

- As estimativas do vetor de parâmetros θ para um modelo não linear são aquelas que minimizam a soma de quadrados dos erros, dada por:

$$S(\theta) = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2$$

Estimação por Mínimos Quadrados Ordinários

- Para encontrar os estimadores que geram tais estimativas, é preciso derivar $S(\boldsymbol{\theta})$ com relação a cada elemento do vetor $\boldsymbol{\theta}$ e igualar a 0, obtendo-se o seguinte vetor de equações normais:

$$\sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2 \left[\frac{\delta f(x_i, \boldsymbol{\theta})}{\delta \theta_j} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = 0$$



Estimação por Mínimos Quadrados Ordinários

- Para modelos lineares esse sistema normalmente possui um vetor único de soluções para os parâmetros θ , em que cada parâmetro depende unicamente dos dados observados.
- Já para modelos não lineares, os estimadores $\hat{\theta}$ resultantes do sistema, em geral não apresentam forma analítica e as soluções precisam ser obtidas por algum método de aproximações iterativo.

Métodos Iterativos

- Os principais métodos iterativos são o método de Gauss-Newton, método de Davidon-Fletcher-Powell, método da descida mais íngreme (steepest descent), método da procura pelo gradiente conjugado, método da procura de Nelder-Mead, algoritmo de Golub-Pereyra, entre outros.





Linearização

- Todos os processos iterativos dependem da escolha de boas estimativas iniciais e os métodos de ajuste de uma regressão não linear não são exceção.
- Uma técnica utilizada para encontrar estimativas iniciais para os parâmetros é a técnica da linearização.



Linearização

- Essa técnica consiste em transformar a função a fim de torná-la linear nos parâmetros.
- Quando isso for possível, pode-se realizar uma regressão linear da variável transformada sobre os novos parâmetros e usar as estimativas dessa regressão (após aplicar a transformação inversa) como estimativas iniciais para o modelo não linear, ignorando o erro aleatório.



Linearização

- O conjunto de dados que será analisado se refere ao peso final após secagem de flores de *Viola × wittrockiana* Gams em relação ao tempo de secagem.
- Vamos ajustar ao conjunto de dados dois dos modelos não lineares citados anteriormente. Para isso, vamos utilizar a técnica de linearização para encontrar chutes iniciais para os parâmetros dos modelos.

Modelo de Newton

- Vamos começar linearizando o modelo para encontrar o chute inicial do parâmetro k :

$$y = e^{(-kx)}$$

$$\log(y) = \log(e^{(-kx)})$$

$$\log(y) = -kx$$

Portanto:

$$z = -kx \quad \longrightarrow \quad \begin{array}{c} \text{MODELO} \\ \text{LINEARIZADO} \end{array}$$



Modelo de Newton

Modelo linear: $z = \beta_0 + \beta_1 x$

No nosso modelo linearizado: $k = -\beta_1$

Em que: $z = \log(y)$



Modelo de Newton

Assim, para encontrar um chute inicial para o parâmetro k , devemos estimar β_1 no modelo linear:

$$z = \beta_0 + \beta_1 x$$

Em que:

$$z = \log(y)$$



Modelo de Newton

- No R, após a inserção dos dados, utilizando os comandos já usados no módulo 1, vamos utilizar os seguintes comandos:

```
z <- log(y)
```

```
lm1 <- lm(z ~ x)
```

```
summary(lm1)
```



Modelo de Newton

Call:

lm(formula = z ~ x)

Residuals:

<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
<i>-0.30216</i>	<i>-0.14723</i>	<i>0.01023</i>	<i>0.12889</i>	<i>0.35055</i>

Coefficients:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(> t)</i>
<i>(Intercept)</i>	<i>-0.1687580</i>	<i>0.0484649</i>	<i>-3.482</i>	<i>0.0011 **</i>
<i>x</i>	<i>-0.0077612</i>	<i>0.0003295</i>	<i>-23.553</i>	<i><2e-16 ***</i>

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 0.1623 on 46 degrees of freedom

Multiple R-squared: 0.9234, Adjusted R-squared: 0.9218

F-statistic: 554.7 on 1 and 46 DF, p-value: < 2.2e-16



Modelo de Newton

$$k = -\beta_1$$

$$k = 0.0077612$$

- Encontramos o chute inicial para o parâmetro k !



Modelo de Newton

Agora, ajustaremos o modelo da regressão não linear por meio da função *nls()*. Essa função tem como estrutura e principais argumentos os seguintes:

nls(formula, data, start, control(list), algorithm, trace)



- *formula* é a fórmula do modelo que será utilizado.
- *data* é o conjunto de dados a ser utilizado.
- *start* é uma lista com valores iniciais para cada um dos parâmetros.



- *control* é uma função que controla características do algoritmo, como o número máximo de iterações e o critério de tolerância. Por exemplo, poderia-se definir `control=nls.control(maxiter=40, tol=0.00001)`.
- *algorithm* define o tipo de algoritmo utilizado. O default é o algoritmo de Gauss-Newton.
- *trace* define se são apresentados a `SQRes` e as estimativas dos parâmetros a cada passo ou não (`trace=TRUE` ou `trace=FALSE`).



Modelo de Newton

No nosso exemplo usaremos:

```
m1 <- nls(y~exp(-k*x), data=dados, start=c(k=0.0077612 ),  
trace=T)  
summary(m1)
```



Modelo de Newton

```
> summary(m1)
```

*Formula: $y \sim \exp(-k * x)$*

Parameters:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(> t)</i>
<i>k</i>	0.0095043	0.0001934	49.15	<2e-16 ***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 0.04281 on 47 degrees of freedom

Number of iterations to convergence: 8

Achieved convergence tolerance: 6.159e-07



Modelo de Newton

- Assim, chegamos que a estimativa obtida para o parâmetro do modelo é $\hat{k} = 0,0095044$ e, portanto, o modelo de regressão não linear estimado é da forma:

$$\hat{y} = e^{-0,0095043x}$$



Modelo de Newton

- Vamos plotar a curva estimada no gráfico de dispersão utilizando os comandos:

`plot(y~x)`

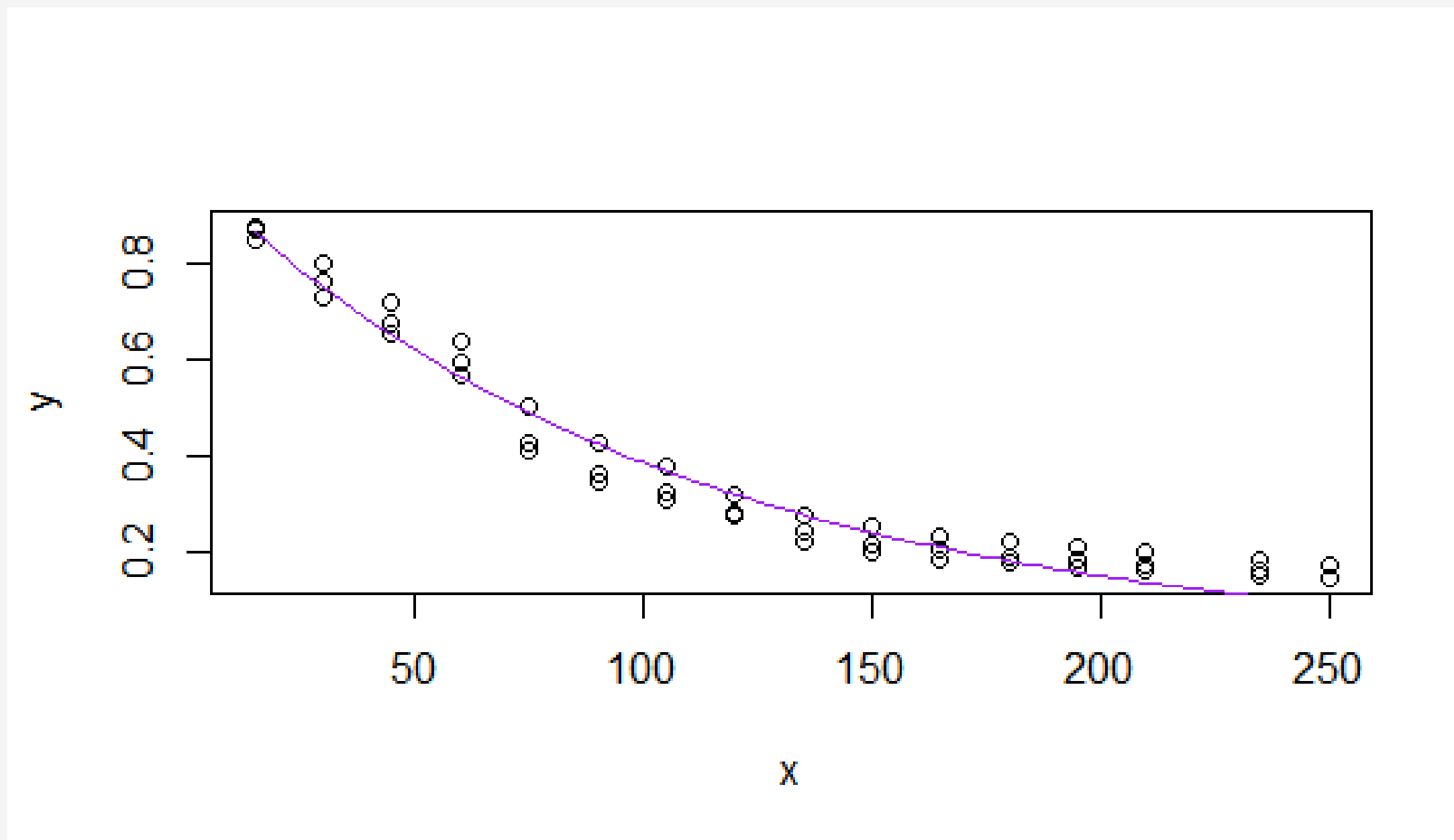
*`curve(exp(-0.0095043*x), add=T, col="purple")`*



Modelo de Newton

- Dessa forma, obtemos:

Figura 1: Modelo de regressão de Newton ajustado aos dados de desidratação das flores de *Viola x wittrockiana*.



Modelo de Michaelis-Menten

- Ajustaremos agora o modelo de Michaelis-Menten aos dados. Iniciaremos o processo linearizando o modelo para encontrar os chutes iniciais dos parâmetros:

$$y = \frac{\theta_1 x}{x + \theta_2}$$

$$\frac{1}{y} = \frac{1}{\frac{\theta_1 x}{x + \theta_2}}$$

$$\frac{1}{y} = \frac{x + \theta_2}{\theta_1 x}$$



Modelo de Michaelis-Menten

$$\frac{1}{y} = \frac{x + \theta_2}{\theta_1 x}$$

$$\frac{1}{y} = \frac{x}{\theta_1 x} + \frac{\theta_2}{\theta_1 x}$$

$$\frac{1}{y} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \frac{1}{x}$$

$$y^* = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} x^*$$



Modelo de Michaelis-Menten

Modelo linear: $z = \beta_0 + \beta_1 x$

No nosso modelo linearizado: $\beta_0 = \frac{1}{\theta_1}$

Em que: $z = \frac{1}{y}$

$$\beta_1 = \frac{\theta_2}{\theta_1}$$



Modelo de Michaelis-Menten

Assim, para encontrar um chute inicial para os parâmetros θ_1 e θ_2 , devemos estimar β_0 e β_1 no modelo linear:

$$z = \beta_0 + \beta_1 x$$

Em que:

$$z = \frac{1}{y}$$



Modelo de Michaelis-Menten

- No R, vamos utilizar os seguintes comandos:

```
z <- 1/y
```

```
lm2 <- lm(z ~ x)
```

```
summary(lm2)
```



Modelo de Michaelis-Menten

Call:

lm(formula = z ~ x)

Residuals:

<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
-0.94042	-0.27321	0.03342	0.24913	0.84077

Coefficients:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(> t)</i>
<i>(Intercept)</i>	0.5173052	0.1255433	4.121	0.000156 ***
<i>x</i>	0.0249078	0.0008536	29.180	< 2e-16 ***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 0.4205 on 46 degrees of freedom

Multiple R-squared: 0.9487, Adjusted R-squared: 0.9476

F-statistic: 851.5 on 1 and 46 DF, p-value: < 2.2e-16



Modelo de Michaelis-Menten

$$\beta_0 = 0,5173052$$

$$\frac{1}{\theta_1} = 0,5173052$$

$$\therefore \theta_1 = 1,933$$

$$\beta_1 = 0,0249078$$

$$\frac{\theta_2}{\theta_1} = 0,0249078$$

$$\theta_2 = 0,0249078 \times 1,933$$

$$\therefore \theta_2 = 0,04814$$

- Encontramos o chute inicial para os parâmetros!



Modelo de Michaelis-Menten

Ajustaremos o modelo da regressão por meio da função *nls()*:

```
m2 <- nls(y~(theta1*x)/(x+theta2), data=dados,  
start=c(theta1=1.933,theta2=0.04814), trace=T)  
summary(m2)
```



Modelo de Michaelis-Menten

> summary(m2)

*Formula: $y \sim (\text{theta1} * x) / (x + \text{theta2})$*

Parameters:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(> t)</i>
<i>theta1</i>	0.30214	0.02291	13.19	<2e-16 ***
<i>theta2</i>	-10.09489	0.64456	-15.66	<2e-16 ***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 0.1608 on 46 degrees of freedom

Number of iterations to convergence: 9

Achieved convergence tolerance: 2.449e-06



Modelo de Michaelis-Menten

- Assim, as estimativas obtidas para os parâmetros do modelo são $\hat{\theta}_1 = 0,30214$ e $\hat{\theta}_2 = -10,09489$ e, portanto, o modelo de regressão não linear estimado é da forma:

$$\hat{y} = \frac{0,30214 x}{x - 10,09489}$$



Modelo de Michaelis-Menten

- Vamos plotar a curva estimada no gráfico de dispersão utilizando os comandos:

plot(y~x)

curve(0.30214*x/(x-10.09489), add=T, col="orange")



Modelo de Michaelis-Menten

- Dessa forma, obtemos:

Figura 2: Modelo de regressão de Michaelis-Menten ajustado aos dados de desidratação das flores de *Viola x wittrockiana*.

