

# CE-092: Extensões de modelos de regressão

## Modelos Não Lineares

Paulo Justiniano Ribeiro Jr

Curso de Estatística  
Departamento de Estatística  
Setor de Ciências Exatas  
Universidade Federal do Paraná



# Regressão: especificação geral

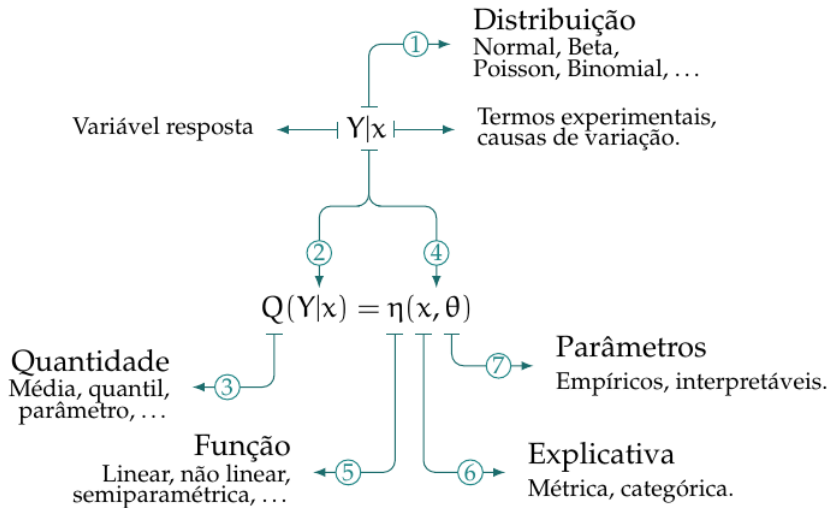


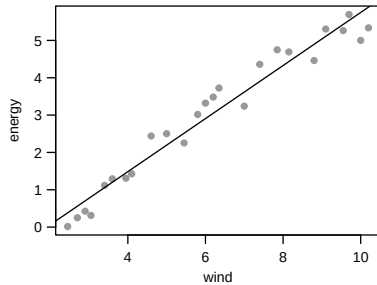
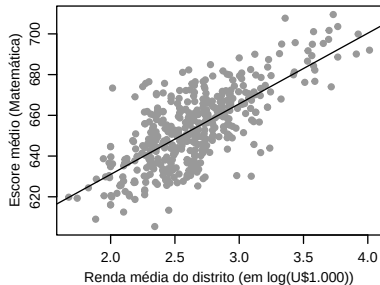
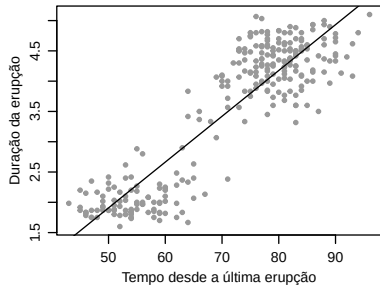
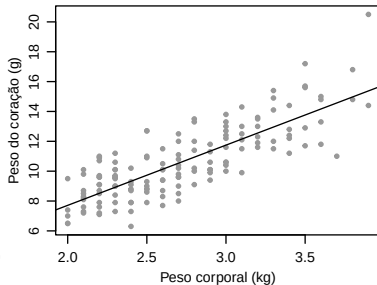
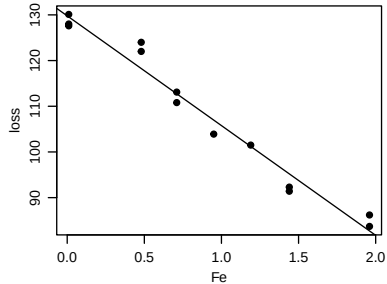
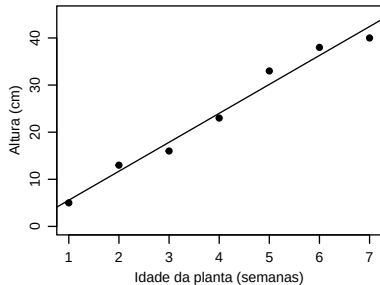
Figura 1. Modelo de regressão de forma geral.

# Mas afinal, de onde vem os modelos de regressão?

## Diferentes possibilidades:

- ▶ Não há qualquer modelo indicado a priori. A escolha de um modelo pode resultar da exploração dos dados e comparação de diferentes especificações.
- ▶ A relação linear do preditor pode ser substituída por equações não lineares (reg. não linear), funções suavizadas (reg. splines), dentre outras generalizações.
- ▶ Experiência empírica. Se um particular modelo proporcionou bom ajuste aos dados em estudos similares, possivelmente ele vai se ajustar bem aos dados do presente estudo;
- ▶ A teoria (física) pode sugerir o modelo. Por exemplo, segundo a lei de Ohm a voltagem aplicada nos terminais de um condutor é proporcional à corrente elétrica que o percorre. Logo, a relação entre as variáveis é linear, induzindo o modelo;

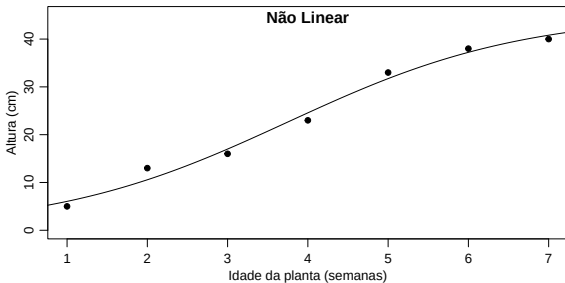
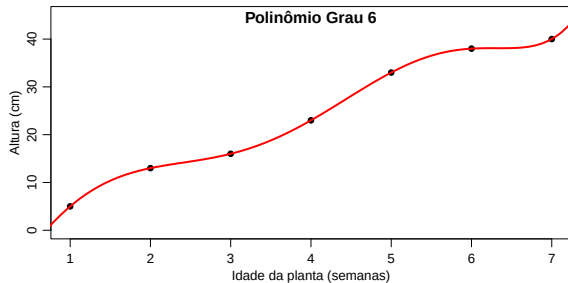
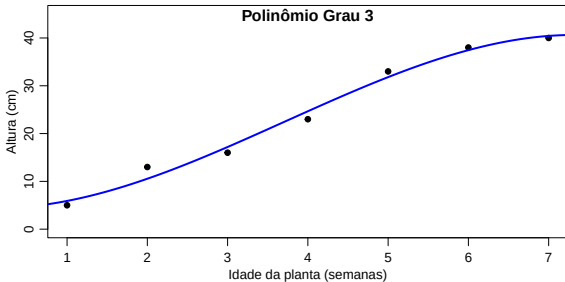
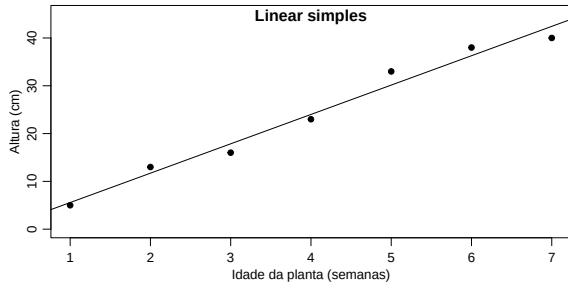
# Exemplos - motivação para regressão



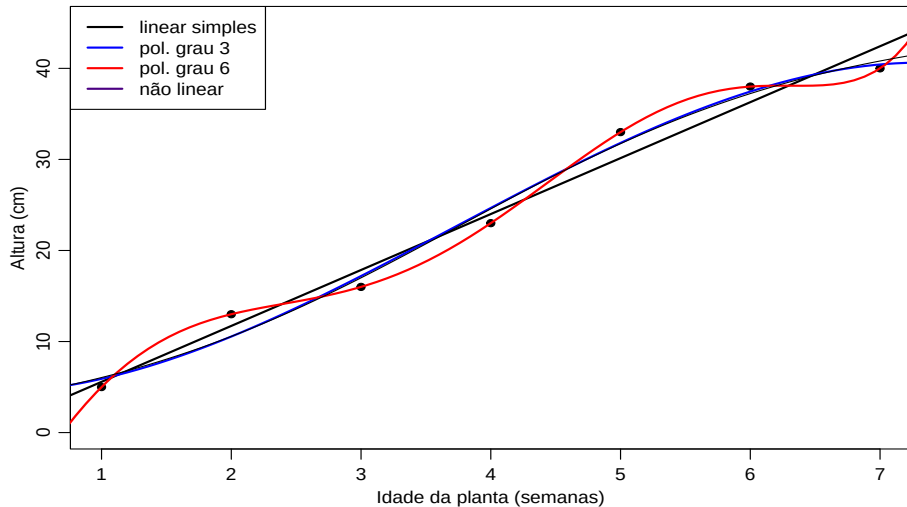
# Regressão linear simples

- ▶ Semelhanças e diferenças nos exemplos
- ▶ Retas de "regressão" nos exemplos:
  - ▶ Como são obtidas?
  - ▶ Como são interpretadas?
  - ▶ Como avaliar se descrevem bem os dados?
  - ▶ Como avaliar a incerteza?
  - ▶ Como avaliar se "valem a pena"?
  - ▶ Como utilizar para fazer inferência e predição?
- ▶ Os modelos e ajustes podem ser "melhorados"?

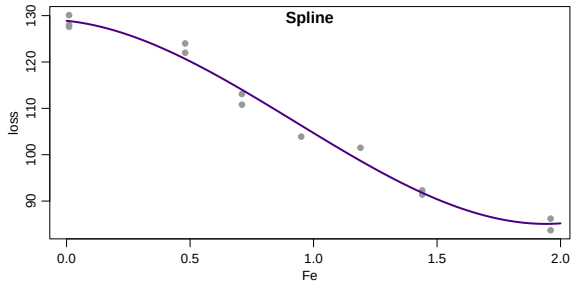
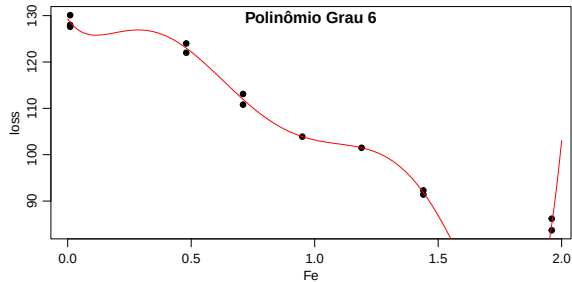
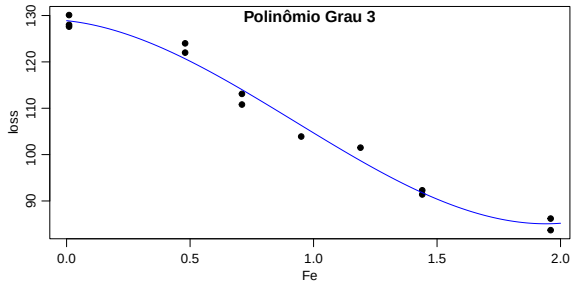
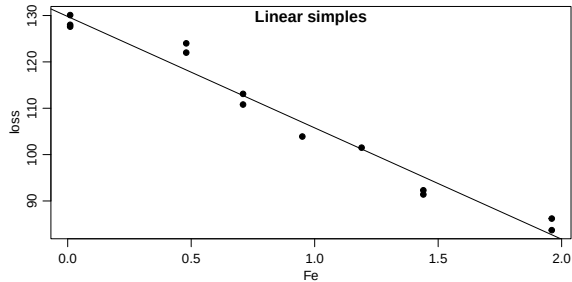
# Motivação para regressão: exemplo crescimento de planta



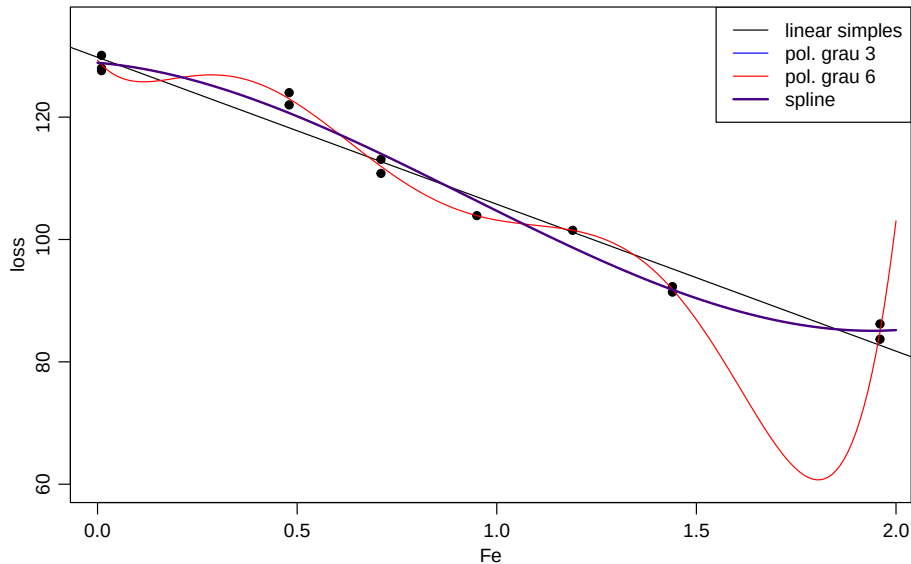
# Motivação para regressão: exemplo crescimento de planta



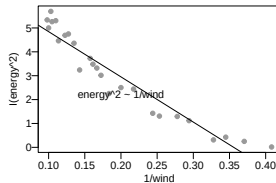
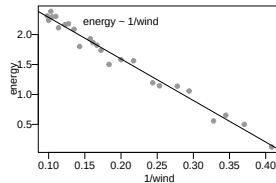
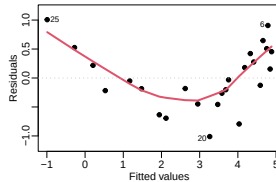
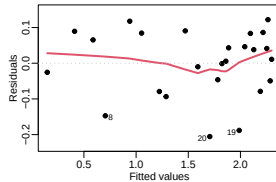
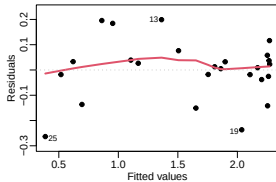
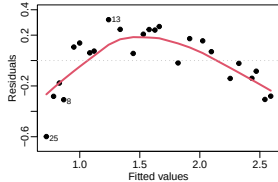
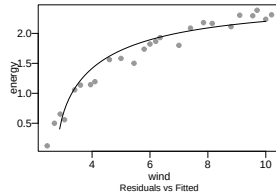
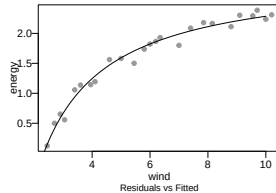
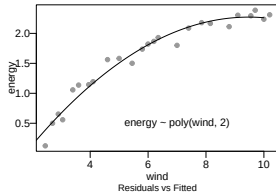
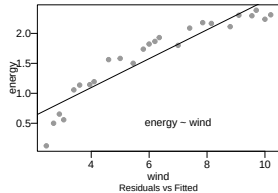
# Motivação para regressão: exemplo perda de material



## Motivação para regressão: exemplo perda de material



# Motivação para regressão: exemplo energia



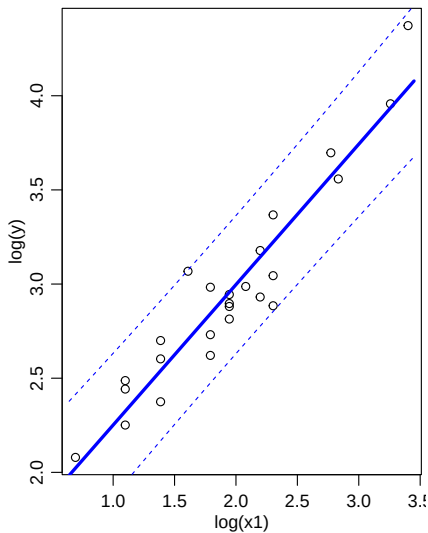
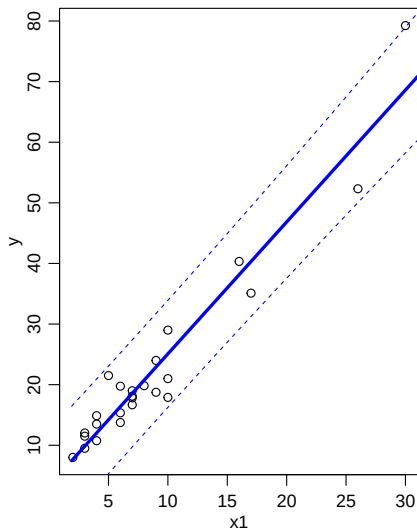
# Modelos de regressão

A variável resposta  $Y$  é explicada pelo preditor  $x$ :

$$h(Y_j) \sim \text{Distr.}(\mu(x_j), \phi)$$
$$E[Y|x_j] = \eta_j = f(x_j)$$

- ▶  $Y_i$  é a **variável resposta** (ou uma transformação desta)
- ▶ para a qual se assume uma **distribuição de probabilidades**
- ▶ da qual o comportamento resume (como, por exemplo, a **média**)
- ▶ é "explicado" por uma função  **$f(x_i)$**  da preditora  $x$ .

# Regressões: transformação dos dados



$$\log(Y_j) \sim N(\mu_j, \sigma^2)$$
$$\mu_j = \beta_0 + \beta_1 \log(x_j)$$

# Modelos diferentes:

$$Y = \beta_0 + \beta_1 x + \epsilon \quad \log(Y) = \beta_0 + \beta_1 \log(x) + \epsilon$$

$$Y = e^{\beta_0 + \beta_1 \log(x)} \cdot \epsilon^*$$

ou

$$Y \sim N(\mu, \sigma^2)$$

$$\mu = \beta_0 + \beta_1 x$$

ou

$$\log(Y) \sim N(\mu, \sigma^2)$$

$$\mu = \beta_0 + \beta_1 \log(x)$$

$$Y = e^{\beta_0 + \beta_1 x} + \epsilon$$

ou

$$Y \sim N(\mu, \sigma^2)$$

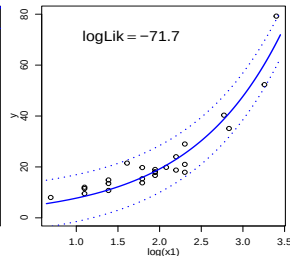
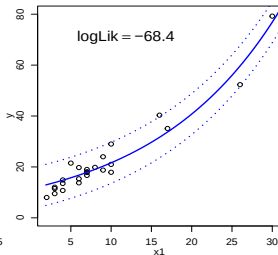
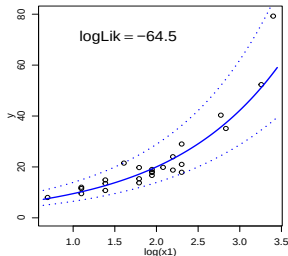
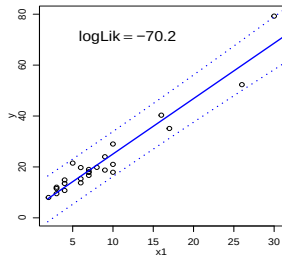
$$\log(\mu) = \beta_0 + \beta_1 x$$

$$Y = e^{\beta_0 + \beta_1 \log(x)} + \epsilon$$

ou

$$Y \sim N(\mu, \sigma^2)$$

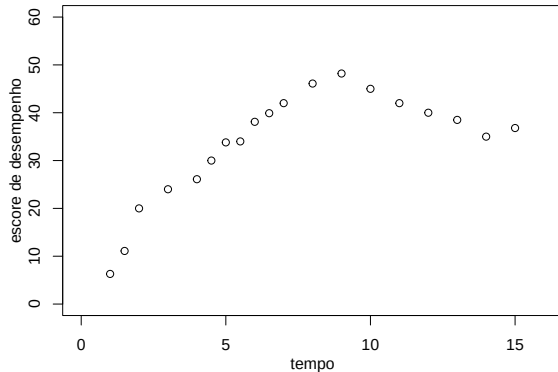
$$\log(\mu) = \beta_0 + \beta_1 \log(x)$$



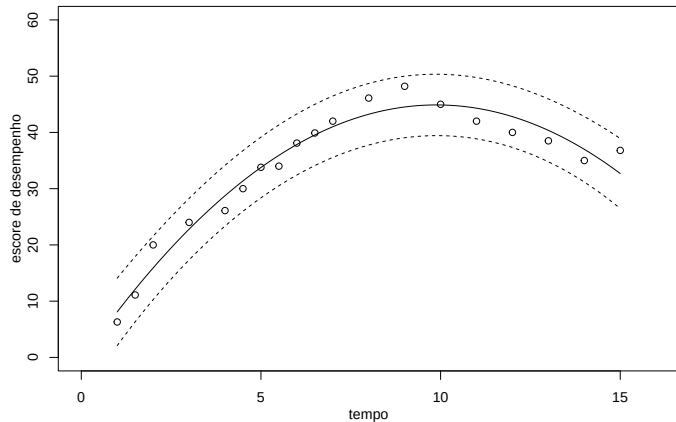
# Exemplo - Relação entre tempo de aquecimento e desempenho de ginastas

Tabela 1. Tempo de aquecimento e desempenho

| Atleta | $t$ | $y$  | Atleta | $t$ | $y$  |
|--------|-----|------|--------|-----|------|
| 1      | 1   | 6.3  | 11     | 7   | 42.0 |
| 2      | 1.5 | 11.1 | 12     | 8   | 46.1 |
| 3      | 2   | 20.0 | 13     | 9   | 48.2 |
| 4      | 3   | 24.0 | 14     | 10  | 45.0 |
| 5      | 4   | 26.1 | 15     | 11  | 42.0 |
| 6      | 4.5 | 30.0 | 16     | 12  | 40.0 |
| 7      | 5   | 33.8 | 17     | 13  | 38.5 |
| 8      | 5.5 | 34.0 | 18     | 14  | 35   |
| 9      | 6   | 38.1 | 19     | 15  | 36.8 |
| 10     | 6.5 | 39.9 |        |     |      |



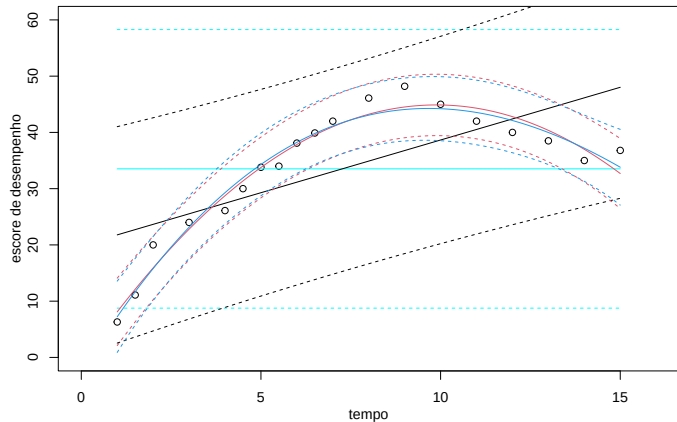
# Regressão Polinomial



$$Y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$$

## Exemplo - Tempo de aquecimento e desempenho de ginastas



$$Y_j \sim N(\mu_j, \sigma^2)$$

$$\mu_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \beta_3 x_j^3$$

## Exemplo – Tempo de aquecimento e desempenho de ginastas

### Summary do modelo

|                 | Estimate | Std. Error | t value | Pr(> t ) |
|-----------------|----------|------------|---------|----------|
| (Intercept)     | 33.5211  | 0.5616     | 59.69   | 0.0000   |
| poly(tempo, 3)1 | 34.1835  | 2.4478     | 13.96   | 0.0000   |
| poly(tempo, 3)2 | -33.3402 | 2.4478     | -13.62  | 0.0000   |
| poly(tempo, 3)3 | 2.3648   | 2.4478     | 0.97    | 0.3493   |

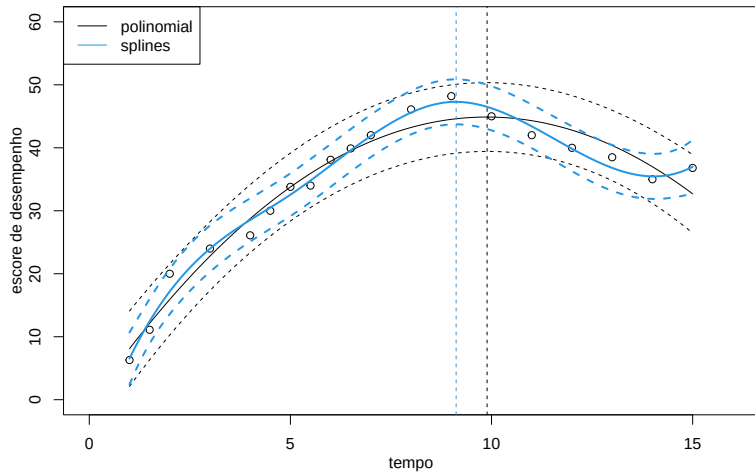
### ANOVA para modelos aninhados

|         | Res.Df | RSS     | Df | Sum of Sq | F      | Pr(>F) |
|---------|--------|---------|----|-----------|--------|--------|
| Modelo0 | 18     | 2375.55 |    |           |        |        |
| Modelo1 | 17     | 1207.04 | 1  | 1168.51   | 195.01 | 0.0000 |
| Modelo2 | 16     | 95.47   | 1  | 1111.57   | 185.51 | 0.0000 |
| Modelo3 | 15     | 89.88   | 1  | 5.59      | 0.93   | 0.3493 |

# Revisitando tempo de aquecimento de ginastas - *splines*

Ajuste versus interpretação

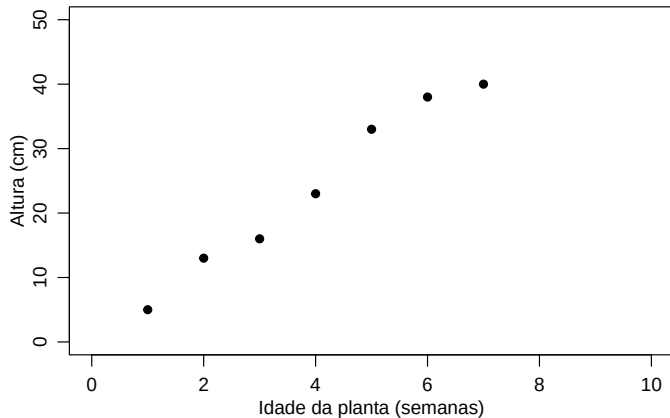
Estimação versus predição



$$Y_j \sim N(\mu_j, \sigma^2)$$

$$\mu_j = s(x_j)$$

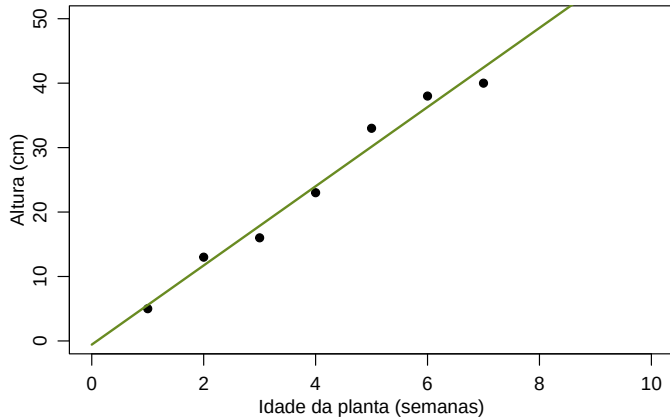
# Regressão Linear ou não linear?



$$Y_j \sim N(\mu_j, \sigma^2)$$

$$\mu_j = ???$$

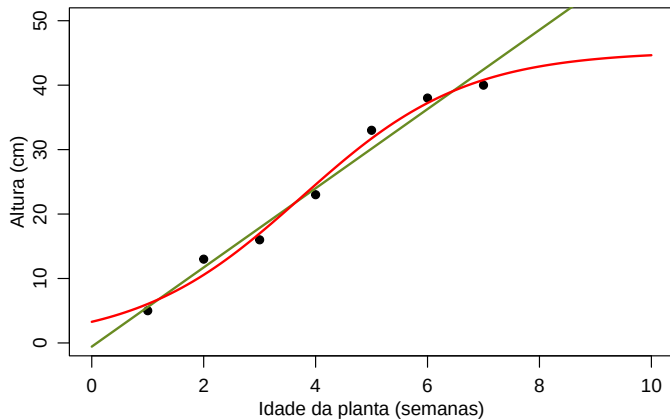
# Regressão Linear ou não linear?



$$Y_j \sim N(\mu_j, \sigma^2)$$

$$\mu_j = \beta_0 + \beta_1 x_j$$

# Regressão Linear ou não linear?



$$Y_j \sim N(\mu_j, \sigma^2)$$

$$\mu_j = \beta_0 + \beta_1 x_j$$

$$\mu_j = f(x_j)$$

# Um modelo não linear

Modelo de crescimento logístico:

$$f(x) = \frac{A}{1 + \exp(-(x - C)/S)}$$

Reescrevendo com letras **romanas para observadas** e **gregas para estimadas**:

$$Y_j \sim N(\mu_j, \sigma^2)$$

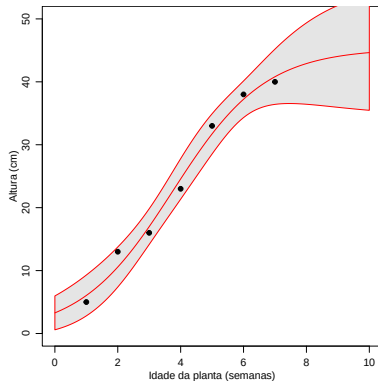
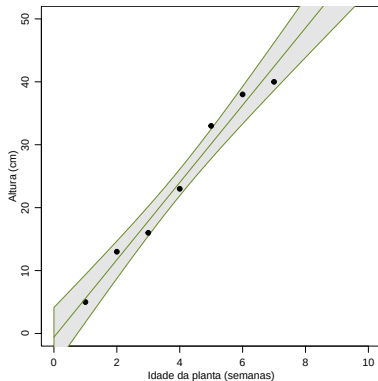
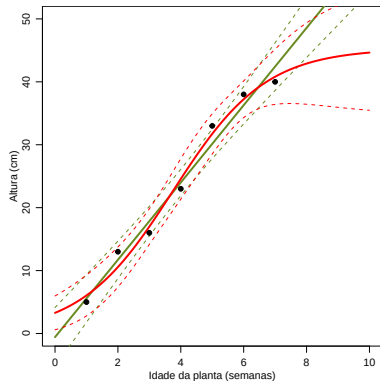
$$\mu_j = \frac{\beta_0}{1 + \exp(-(x_j - \beta_1)/\beta_2)}$$

Interpretações dos parâmetros:

- ▶  $A(\beta_0)$ : assíntota: *até onde a planta cresce*
- ▶  $C(\beta_1)$ : tempo de meia vida: *tempo para atingir metade do tamanho máximo*
- ▶  $S(\beta_2)$ : taxa: *velocidade que a planta cresce*

Ajuste de modelos não lineares por métodos numéricos que exigem "chutes" iniciais.  
As interpretações dos parâmetros podem ajudar na definição dos valores iniciais.

# Linear e não linear



# Linearidade

- Um modelo de regressão é linear se

$$\frac{\partial f(\mathbf{x}; \boldsymbol{\beta})}{\partial \beta_j} \quad (1)$$

não depender de  $\beta_j$ , para todo  $j = 0, 1, \dots, p$ .

- Ou seja, num modelo de regressão linear, ao se derivar parcialmente em relação a um parâmetro  $\beta_j$ , o resultado não depende de qualquer dos  $\beta_{j'}$ .

Modelos de regressão linear podem ser escritos na forma matricial:

$$\underset{n \times 1}{Y} = \underset{n \times p}{X} \cdot \underset{p \times 1}{\beta} + \underset{n \times 1}{\epsilon}$$

# Lineares e Não Lineares

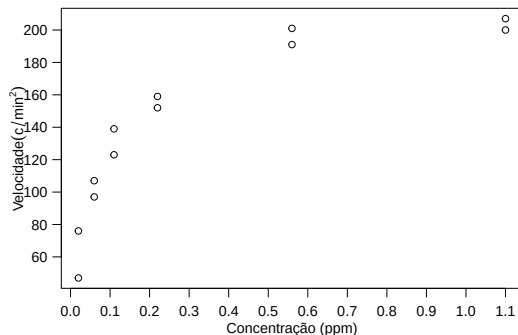
1.  $y = \beta_0 + \beta_1 x + \epsilon$
2.  $y = \beta_0 + \beta_1 \ln(x) + \epsilon$
3.  $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$
4.  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$
5.  $y = \beta_0 + \beta_1 \ln(x_1) + \frac{\beta_2}{x_2} + \beta_3 x_1 x_2 + \epsilon$
6.  $y = \beta_0 + \beta_1 e^{\beta_2 x_2} + \epsilon$
7.  $y = \frac{\beta_1}{1 + e^{\beta_2 x}} + \epsilon$
8.  $y = \beta_0 \sin(\beta_1 + \beta_2 x) + \epsilon$

## Exemplo - Reação química

Velocidade de uma reação enzimática ( $y$ , expressa em *contagem/min*<sup>2</sup>) e a concentração de certo substrato ( $x$ , em *partes por milhão*) em 12 repetições de um experimento.

Tabela 2. Dados - reação química

| Repetição | $x$  | $y$ | Repetição | $x$  | $y$ |
|-----------|------|-----|-----------|------|-----|
| 1         | 0.02 | 47  | 7         | 0.02 | 76  |
| 2         | 0.06 | 97  | 8         | 0.06 | 107 |
| 3         | 0.11 | 123 | 9         | 0.11 | 139 |
| 4         | 0.22 | 152 | 10        | 0.22 | 159 |
| 5         | 0.56 | 191 | 11        | 0.56 | 201 |
| 6         | 1.10 | 200 | 12        | 1.10 | 207 |



## Exemplo - Reação química

- ▶ Observamos uma relação não linear entre as variáveis.
- ▶ Ao invés de considerar um modelo polinomial, podemos recorrer à teoria da química e utilizar o [modelo de Michaelis-Menten](#), que descreve adequadamente problemas de cinética enzimática.
- ▶ Vamos considerar  $f(x, \beta) = \frac{\beta_1 x}{x + \beta_2}$ , ajustando o seguinte modelo de regressão não linear:

$$y = f(x, \beta) + \epsilon$$

ou

$$y = \frac{\beta_1 x}{x + \beta_2} + \epsilon$$

$$Y \sim N(\mu, \sigma^2)$$
$$\mu = \frac{\beta_1}{x + \beta_2}$$

## Ajuste do modelo

```
rquim.fit <- nls(y ~ b1*x/(x+b2), data = rquim, start = list(b1=200, b2=0.2))
summary(rquim.fit)

##
## Formula: y ~ b1 * x/(x + b2)
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## b1 2.127e+02  6.947e+00  30.615 3.24e-11 ***
## b2 6.412e-02  8.281e-03   7.743 1.57e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.93 on 10 degrees of freedom
##
## Number of iterations to convergence: 5
## Achieved convergence tolerance: 5.544e-06
```

|    | Estimate | Std. Error | t value | Pr(> t ) |
|----|----------|------------|---------|----------|
| b1 | 212.68   | 6.95       | 30.61   | 0.00     |
| b2 | 0.06     | 0.01       | 7.74    | 0.00     |

## Exemplo - Reação química

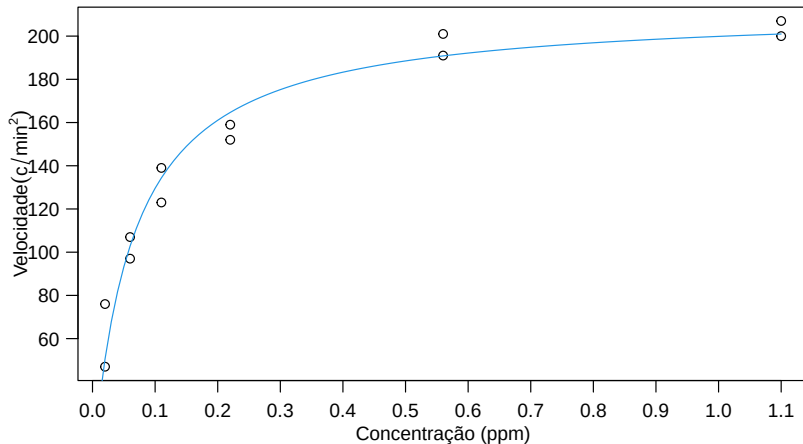


Figura 2. Dados de reação química e modelo não linear ajustado

# Reexpressando o modelo

Seria o mesmo modelo?

Modelo não linear:

$$y = \frac{\beta_1 x}{x + \beta_2} + \epsilon.$$

```
nls(y ~ b1*x/(x+b2), data=rquim, start=list(b1=200, b2=0.2))
```

Reexpressando o preditor:

$$\frac{1}{y} = \frac{x + \beta_2}{\beta_1 x} + \epsilon^*$$

$$\frac{1}{y} = \frac{x}{\beta_1 x} + \frac{\beta_2}{\beta_1 x} + \epsilon^*$$

```
lm(I(1/y) ~ I(1/x), data=rquim)
```

$$y^* = \beta_0^* + \beta_1^* x^* + \epsilon^*$$

em que:

$$y^* = \frac{1}{y} \quad ; \quad x^* = \frac{1}{x} \quad ; \quad \beta_0^* = \frac{1}{\beta_1} \quad ; \quad \beta_1^* = \frac{\beta_2}{\beta_1}$$

Modelo de regressão linear simples com transformação e reparametrização.

# Ajustes dos modelos não linear e linearizado

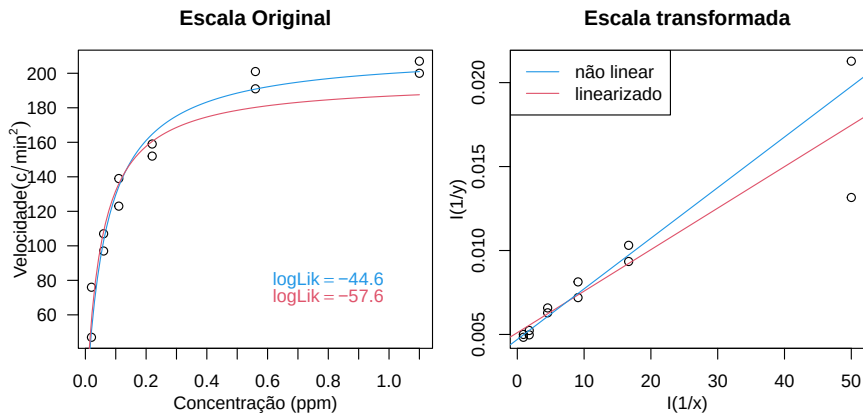


Figura 3. Dados de reação química e ajustes dos modelos

Modelo não-linear (base teórica) ajustado na **escala original** e linear (empírico) ajustado na **escala transformada**.

# Reexpressando de ainda outra forma

Seria o mesmo modelo?

Modelo não linear:

$$Y \sim N(\mu, \sigma^2)$$

$$\mu = \frac{\beta_1 x}{x + \beta_2}$$

```
nls(y ~ b1*x/(x+b2), data=rquim, start=list(b1=200, b2=0.2))
```

Modelo linear generalizado (GLM):

$$Y \sim N(\mu, \sigma^2)$$

$$h(\mu) = \frac{1}{\mu} = \beta_0^* + \beta_1^* x^*$$

```
glm(y ~ I(1/x), data=rquim, family = gaussian(link = "inverse"))
```

$$x^* = \frac{1}{x} \quad ; \quad \beta_0^* = \frac{1}{\beta_1} \quad ; \quad \beta_1^* = \frac{\beta_2}{\beta_1}$$

Modelo linear generalizado com distribuição normal e função de ligação inversa.

# Ajustes como modelo não linear e GLM

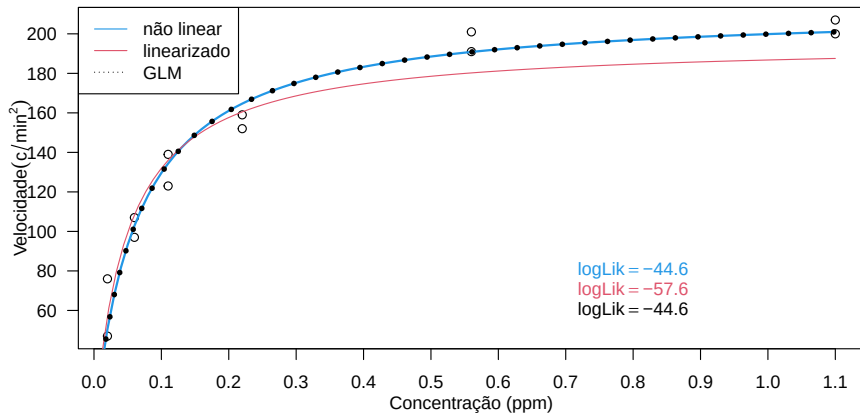


Figura 4. Dados de reação química e ajustes dos modelos