



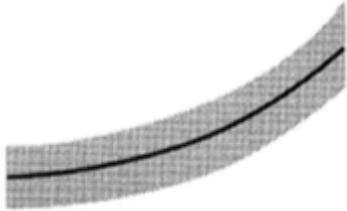
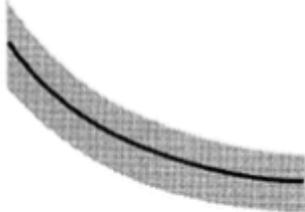
ME613 - Análise de Regressão

Parte 4

Samara F. Kiihl - IMECC - UNICAMP

Transformações

Transformações: relações não lineares

	Prototype Regression Pattern	Transformations of X
(a)		$X' = \log_{10} X$ $X' = \sqrt{X}$
(b)		$X' = X^2$ $X' = \exp(X)$
(c)		$X' = 1/X$ $X' = \exp(-X)$

Transformações: relações não lineares

- Linearizar uma relação linear, em casos que a suposição de normalidade dos erros com variância constante é adequada .
- Tentar encontrar transformação em X .
- Avaliar os gráficos de resíduo para decidir qual a melhor transformação.

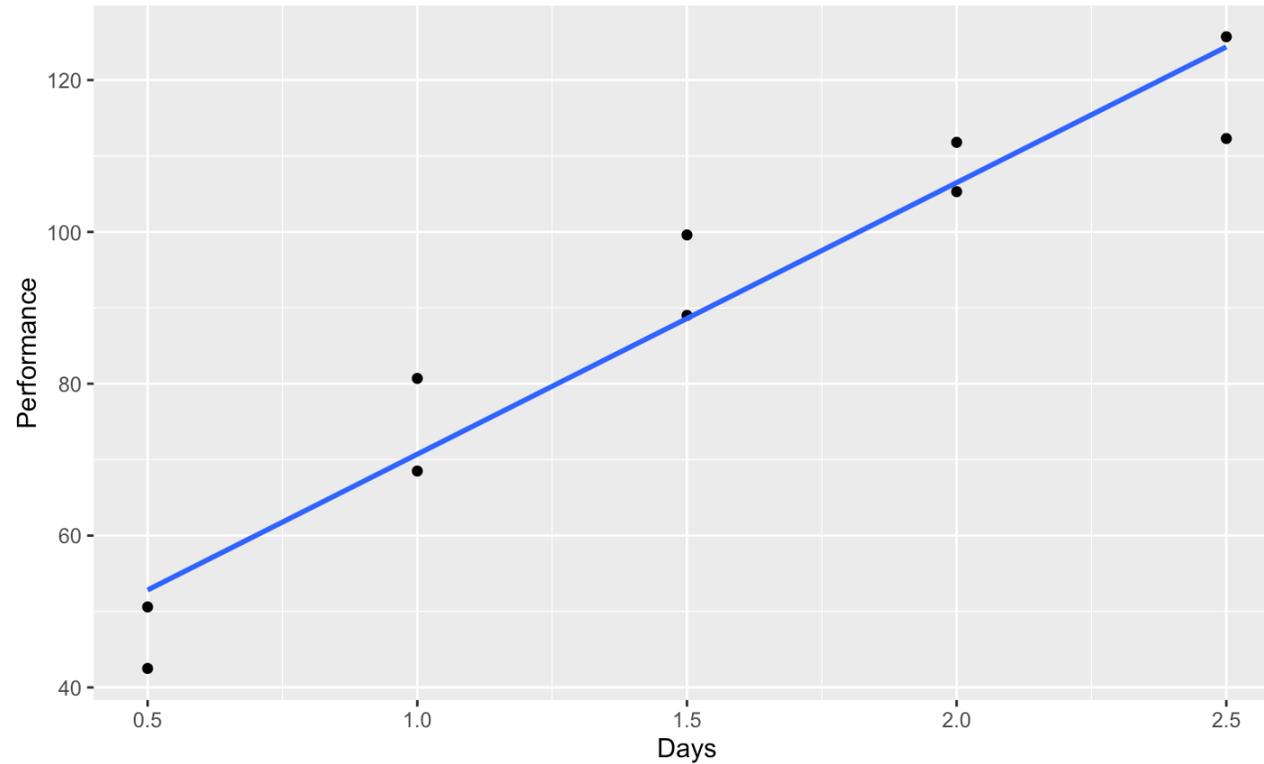
Exemplo: Treinamento e Performance

X: número de dias de treinamento recebido.

Y: performance nas vendas.

##	Days	Performance
## 1	0.5	42.5
## 2	0.5	50.6
## 3	1.0	68.5
## 4	1.0	80.7
## 5	1.5	89.0
## 6	1.5	99.6
## 7	2.0	105.3
## 8	2.0	111.8
## 9	2.5	112.3
## 10	2.5	125.7

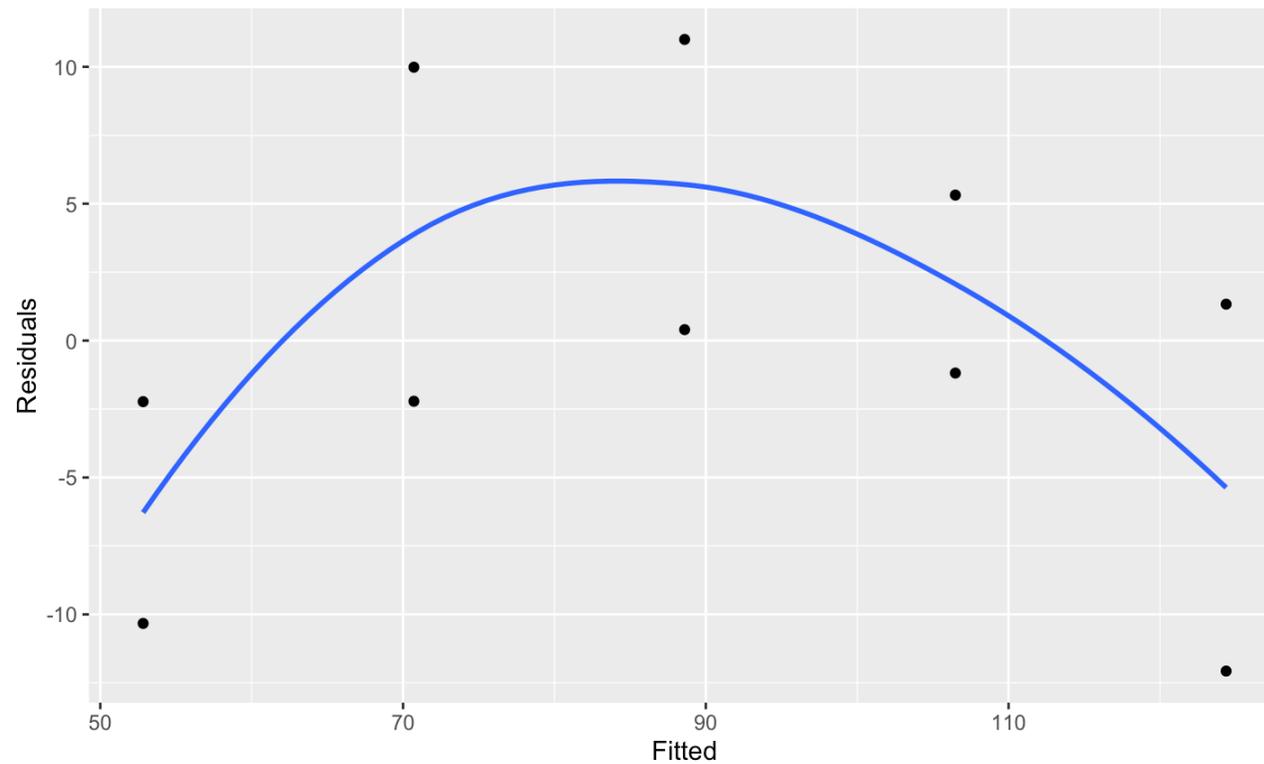
Exemplo: Treinamento e Performance



$$\hat{Y} = 34.945 + 35.77X$$

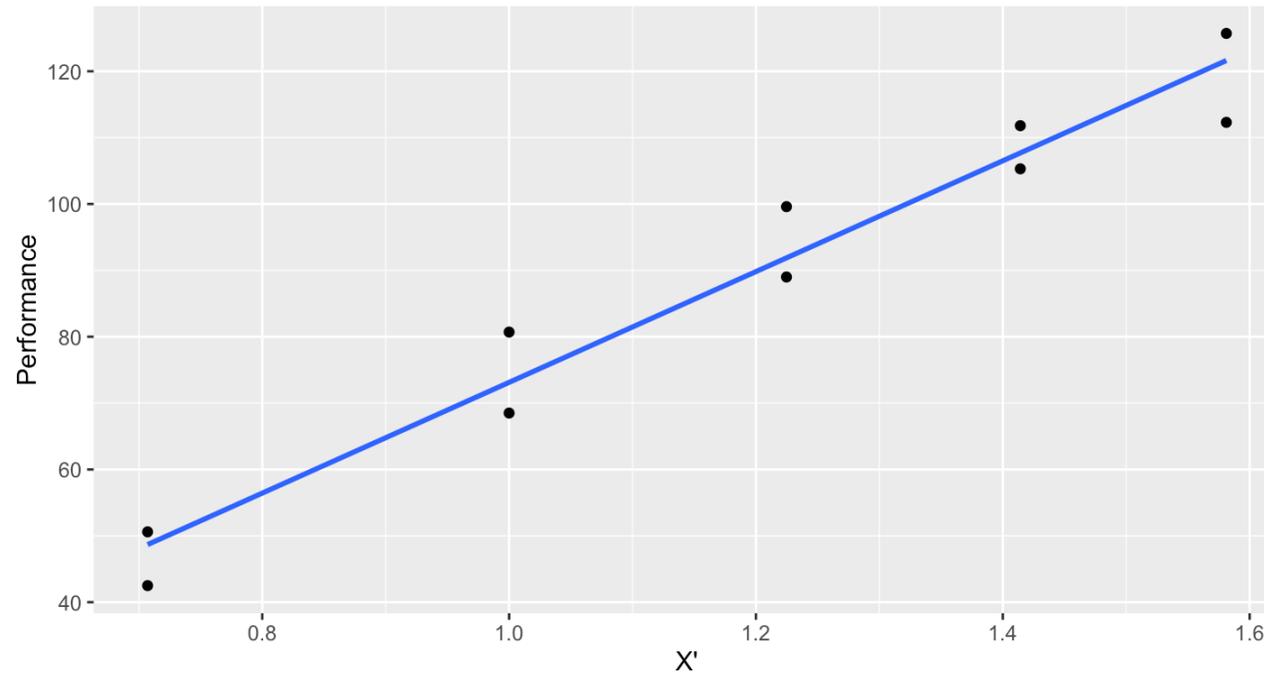
Exemplo: Treinamento e Performance

Gráfico de resíduos: e_i versus \hat{Y}_i .



Exemplo: Treinamento e Performance

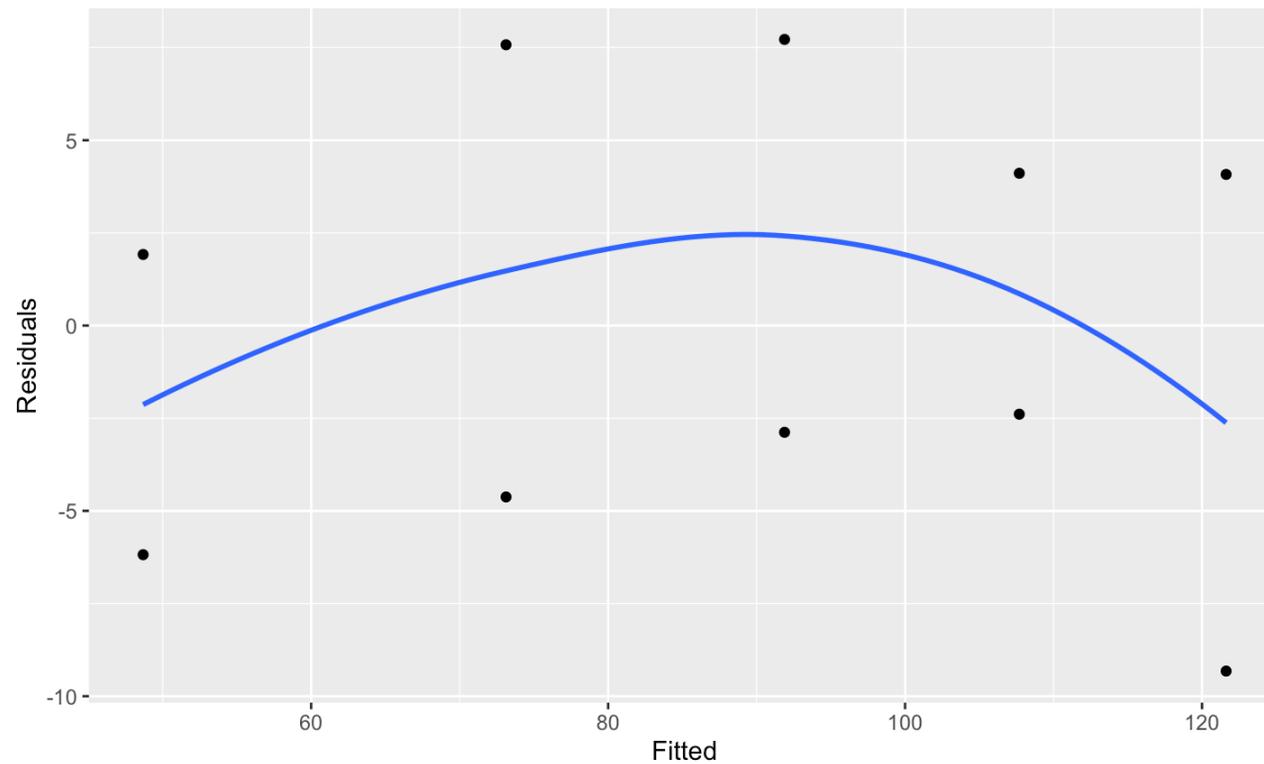
$$X' = \sqrt{X}$$



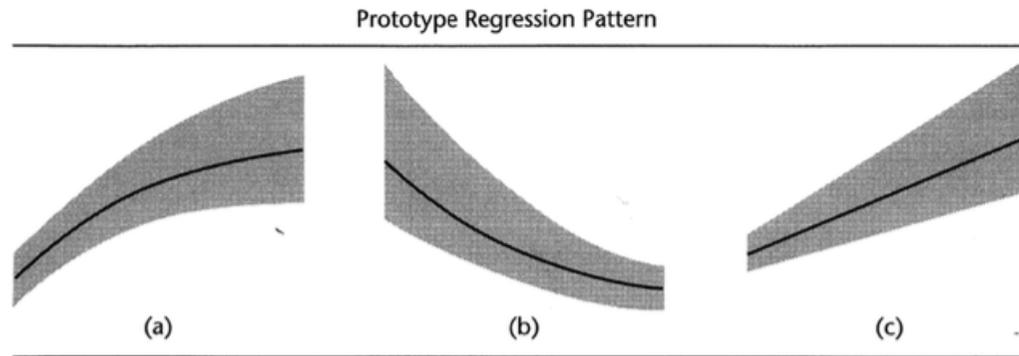
$$\hat{Y} = -10.33 + 83.45\sqrt{X}$$

Exemplo: Treinamento e Performance

Gráfico de resíduos: e_i versus \hat{Y}_i usando X' no modelo.



Transformações: não normalidade e variância não constante



Transformações: não normalidade e variância não constante

- Tentar encontrar transformação em Y .
- Pode ser combinada com uma transformação também em X .

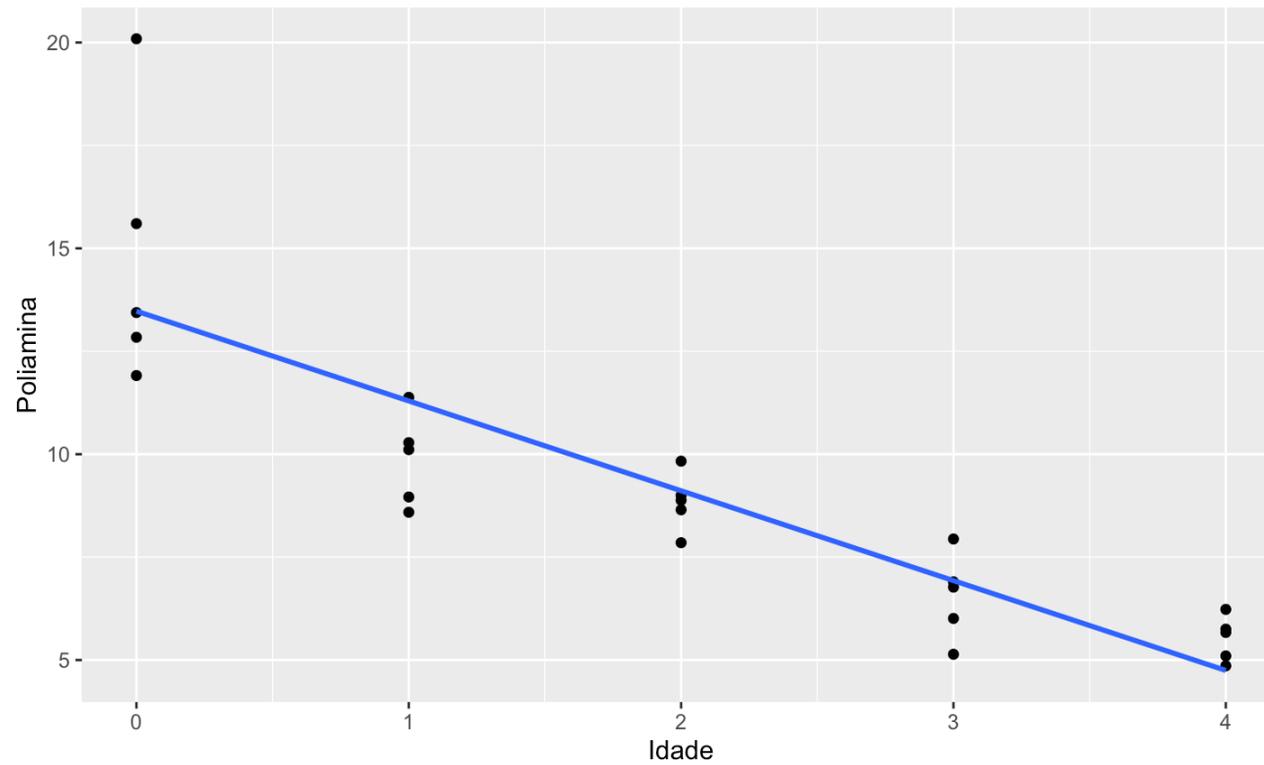
Exemplo: Idade e nível de poliamina

X: Idade

Y: nível de poliamina no plasma

##	Idade	Poliamina
## 1	0	13.44
## 2	0	12.84
## 3	0	11.91
## 4	0	20.09
## 5	0	15.60
## 6	1	10.11
## 7	1	11.38
## 8	1	10.28
## 9	1	8.96
## 10	1	8.59
## 11	2	9.83
## 12	2	9.00
## 13	2	8.65
## 14	2	7.85
## 15	2	8.88
## 16	3	7.94
## 17	3	6.01
## 18	3	5.11

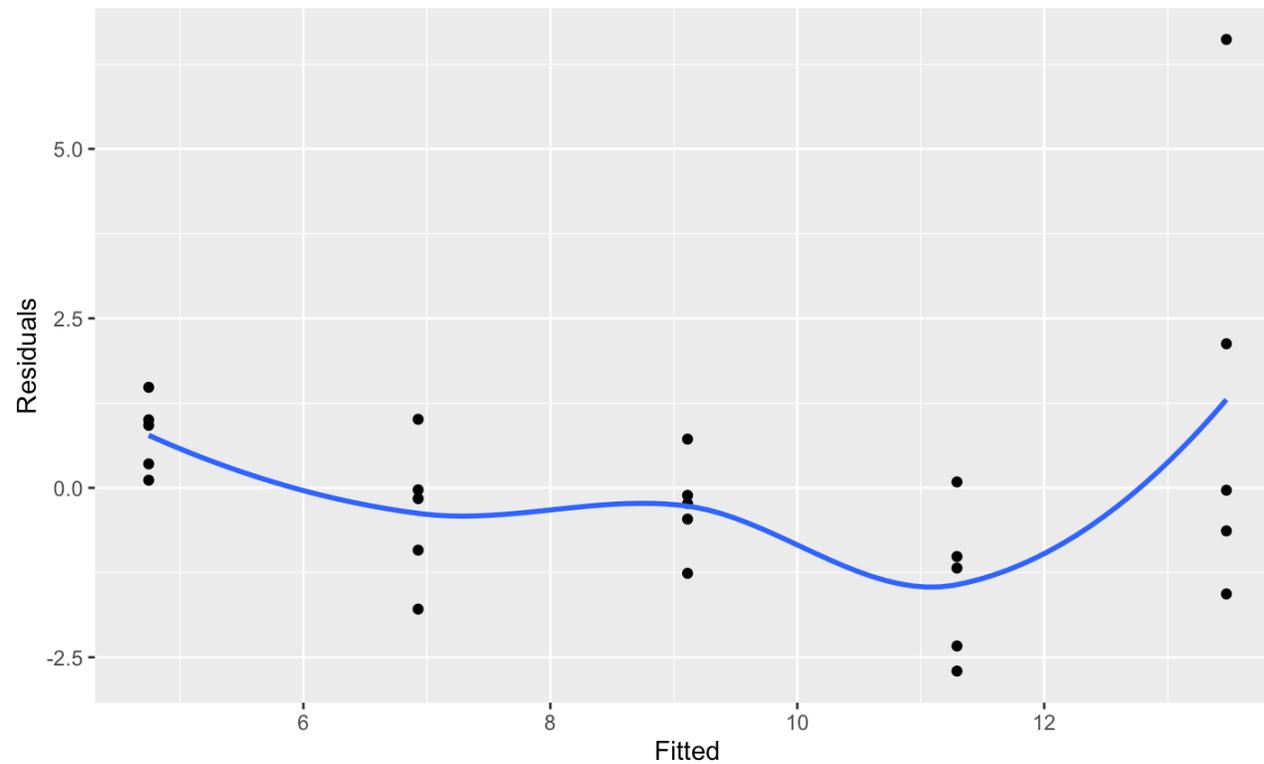
Exemplo: Idade e nível de poliamina



$$\hat{Y} = 13.4752 + -2.182X$$

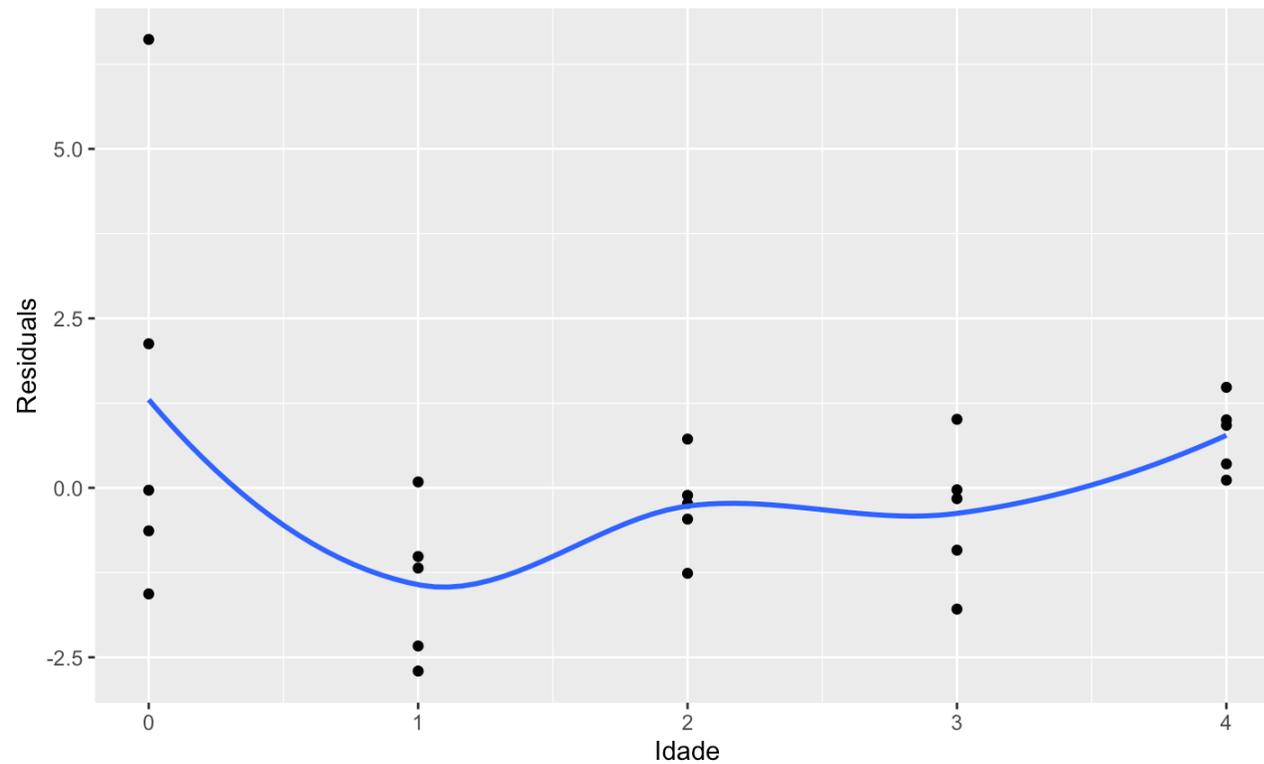
Exemplo: Idade e nível de poliamina

Gráfico de resíduos: e_i versus \hat{Y}_i .



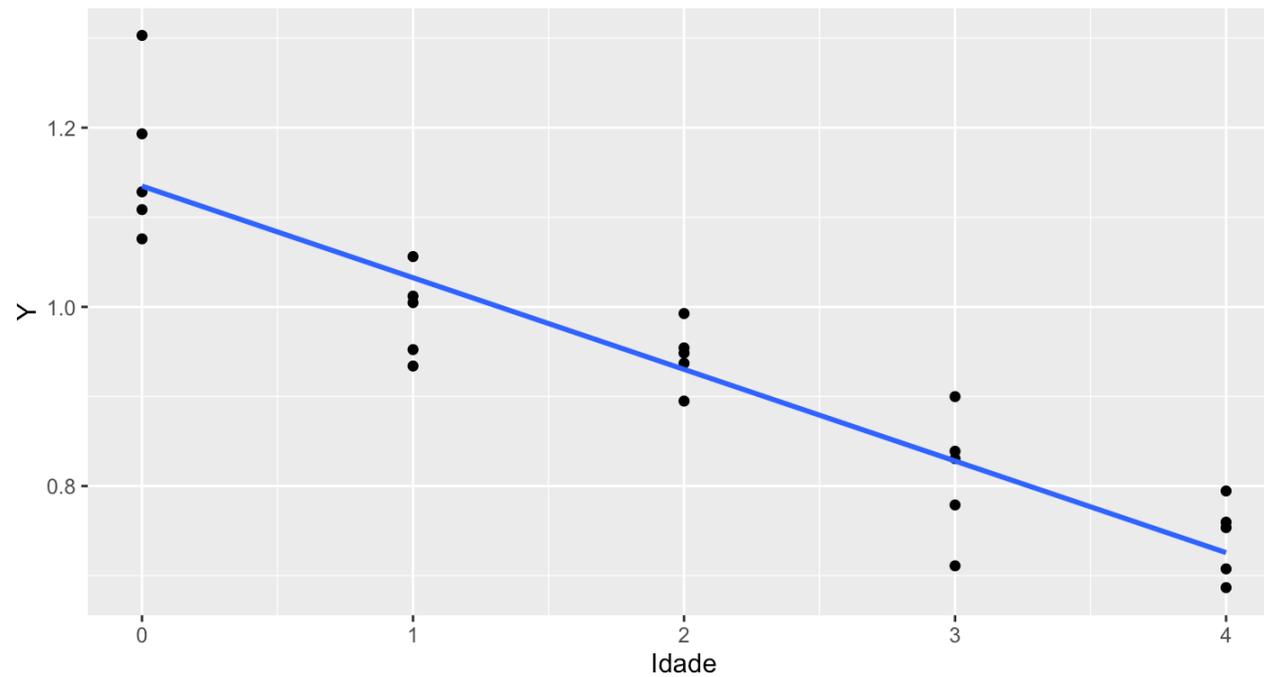
Exemplo: Idade e nível de poliamina

Gráfico de resíduos: e_i versus X_i .



Exemplo: Idade e nível de poliamina

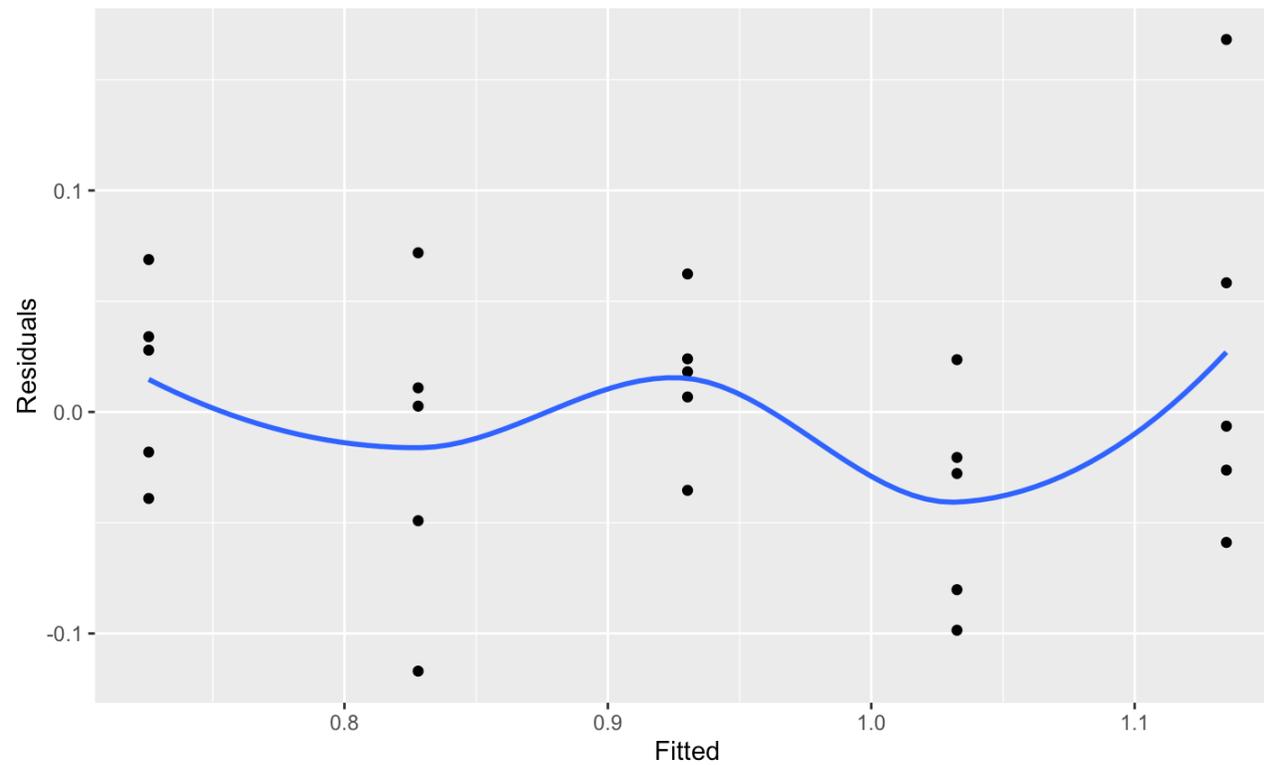
$$Y' = \log_{10} Y$$



$$\hat{Y}' = 1.13 + -0.1X$$

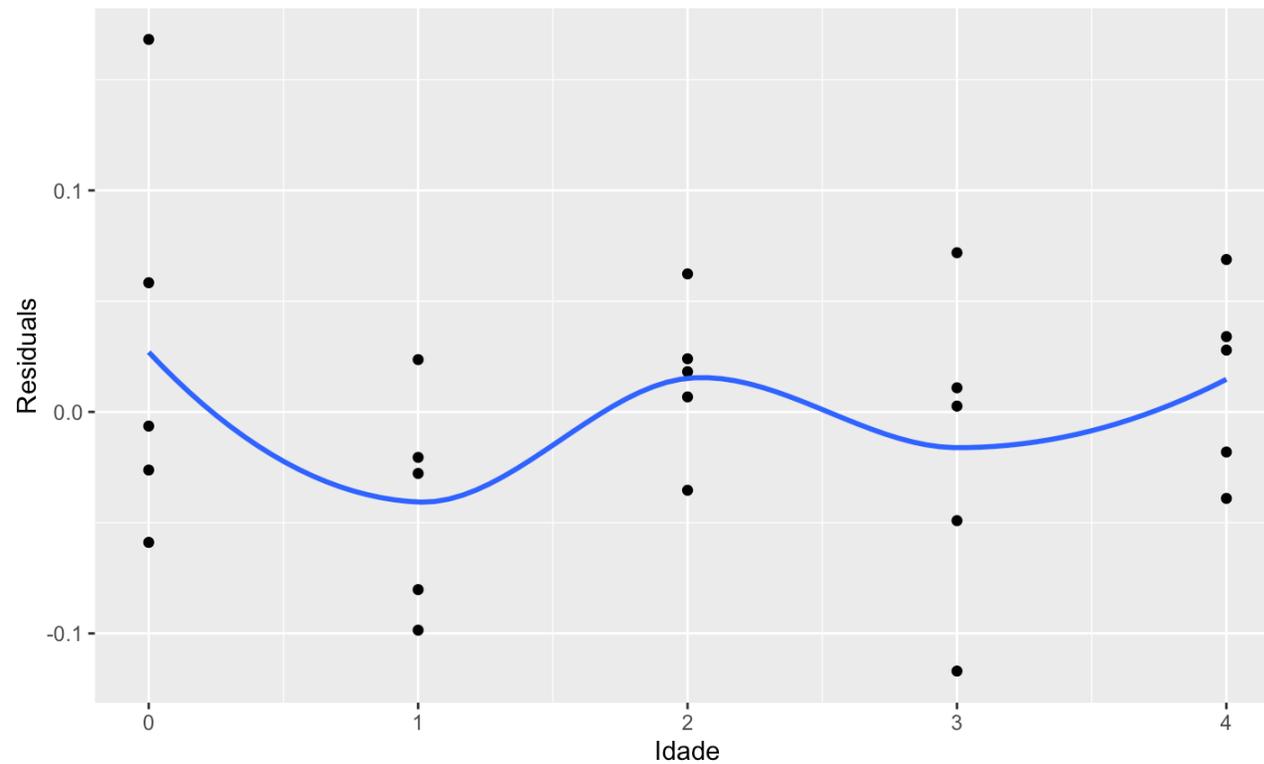
Exemplo: Idade e nível de poliamina

Gráfico de resíduos: e_i versus \hat{Y}'_i usando Y' no modelo.



Exemplo: Idade e nível de poliamina

Gráfico de resíduos: e_i versus X_i .



Algumas transformações em Y

- $\log_e(Y)$: para estabilizar a variância quando esta tende a crescer à medida que Y cresce.
- \sqrt{Y} : estabilizar a variância quando esta é proporcional à média dos Y 's.
- $\frac{1}{Y}$: estabilizar a variância, minimizando o efeito de valores muito altos de Y .
- Y^2 : estabilizar a variância quando esta tende a decrescer com a média de Y 's.
- $\arcsin \sqrt{Y}$: estabilizar a variância quando os dados são proporções.
- etc...

Transformações de Box-Cox

- Muitas vezes é difícil determinar, através de gráficos, qual a melhor transformação a ser feita.
- O procedimento de [Box-Cox](#) identifica automaticamente uma transformação:

$$Y' = Y^\lambda$$

em que λ é um parâmetro a ser determinado a partir dos dados.

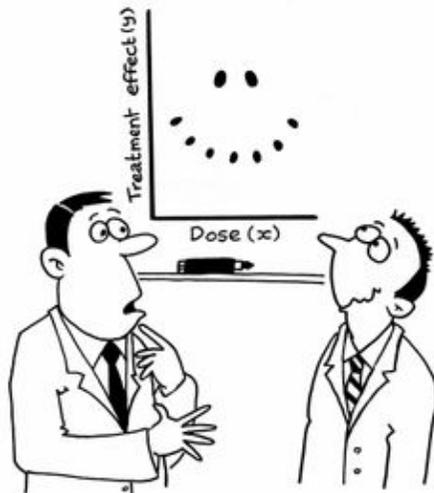
- Modelo com dados transformados:

$$Y_i^\lambda = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- O procedimento de Box-Cox utiliza o método de máxima verossimilhança para estimar λ .

Leitura

- Applied Linear Statistical Models: 3.8-3.11.
- Weisberg - [Applied Linear Regression](#): Capítulo 8.
- Faraway - [Linear Models with R](#): Capítulo 9.



"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."