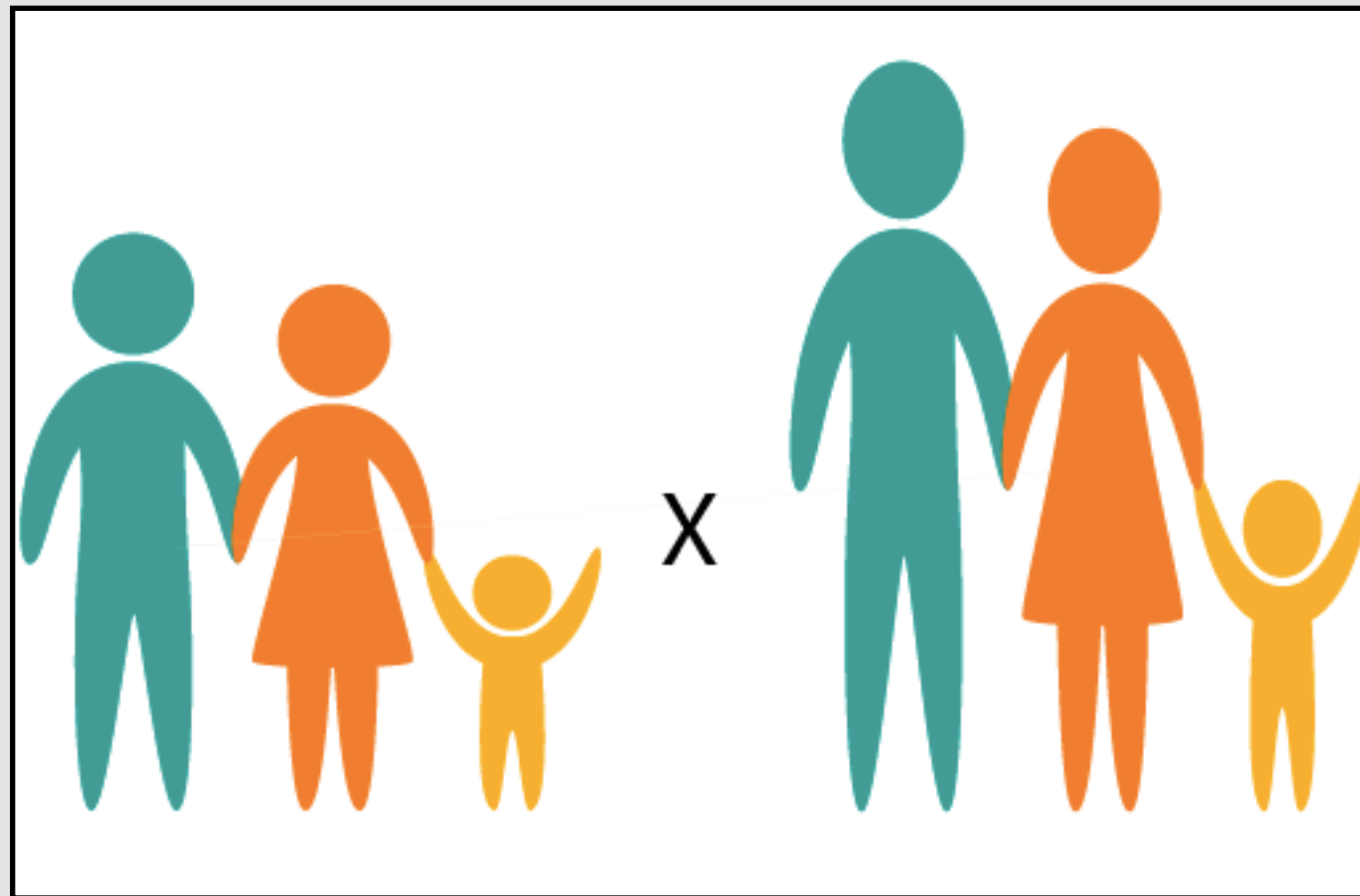


Regressão Linear Simples

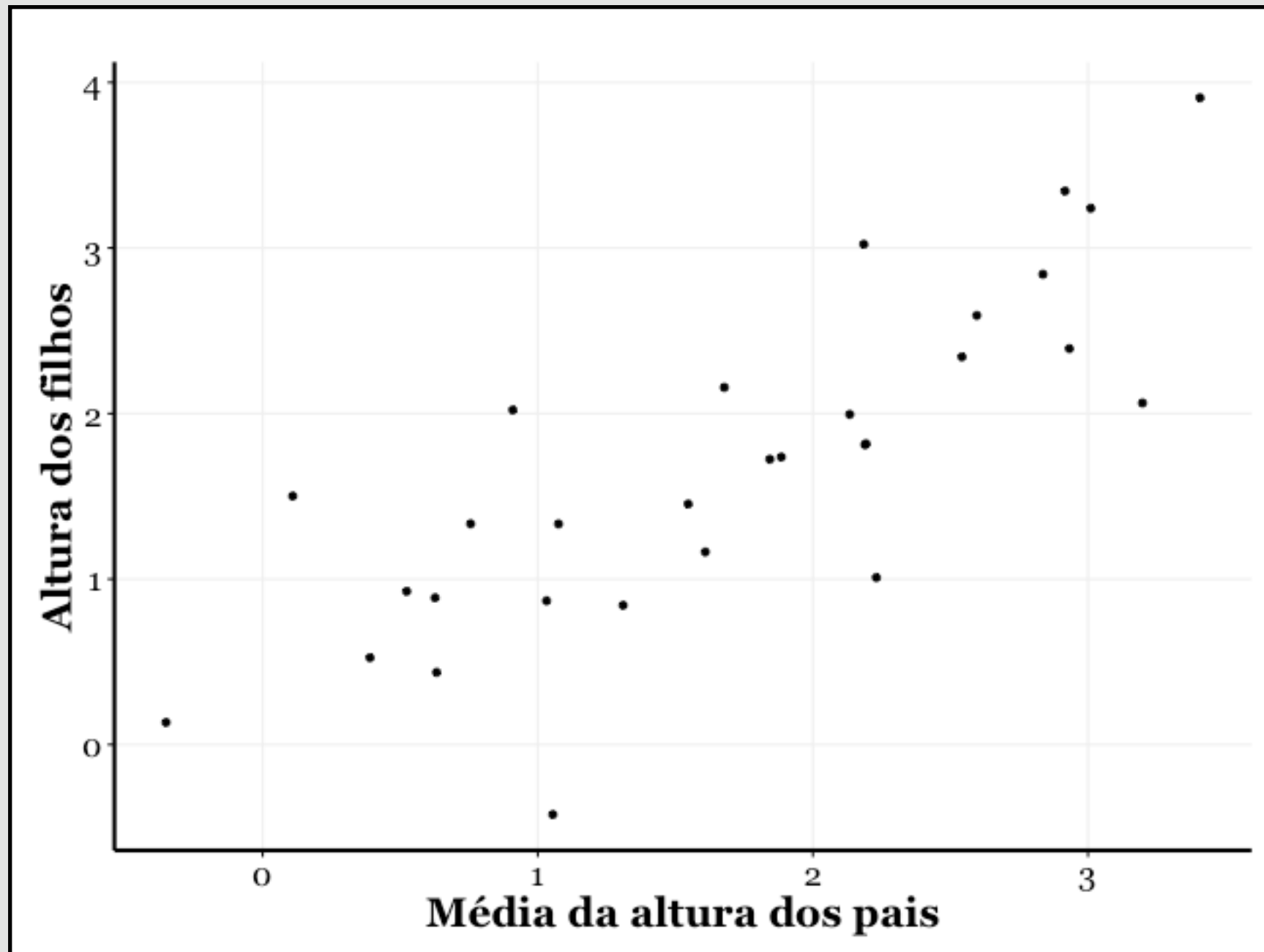
Semana Temática da Biologia (USP) - 2016

Problema

- Existe relação entre a altura dos pais e a altura dos filhos?



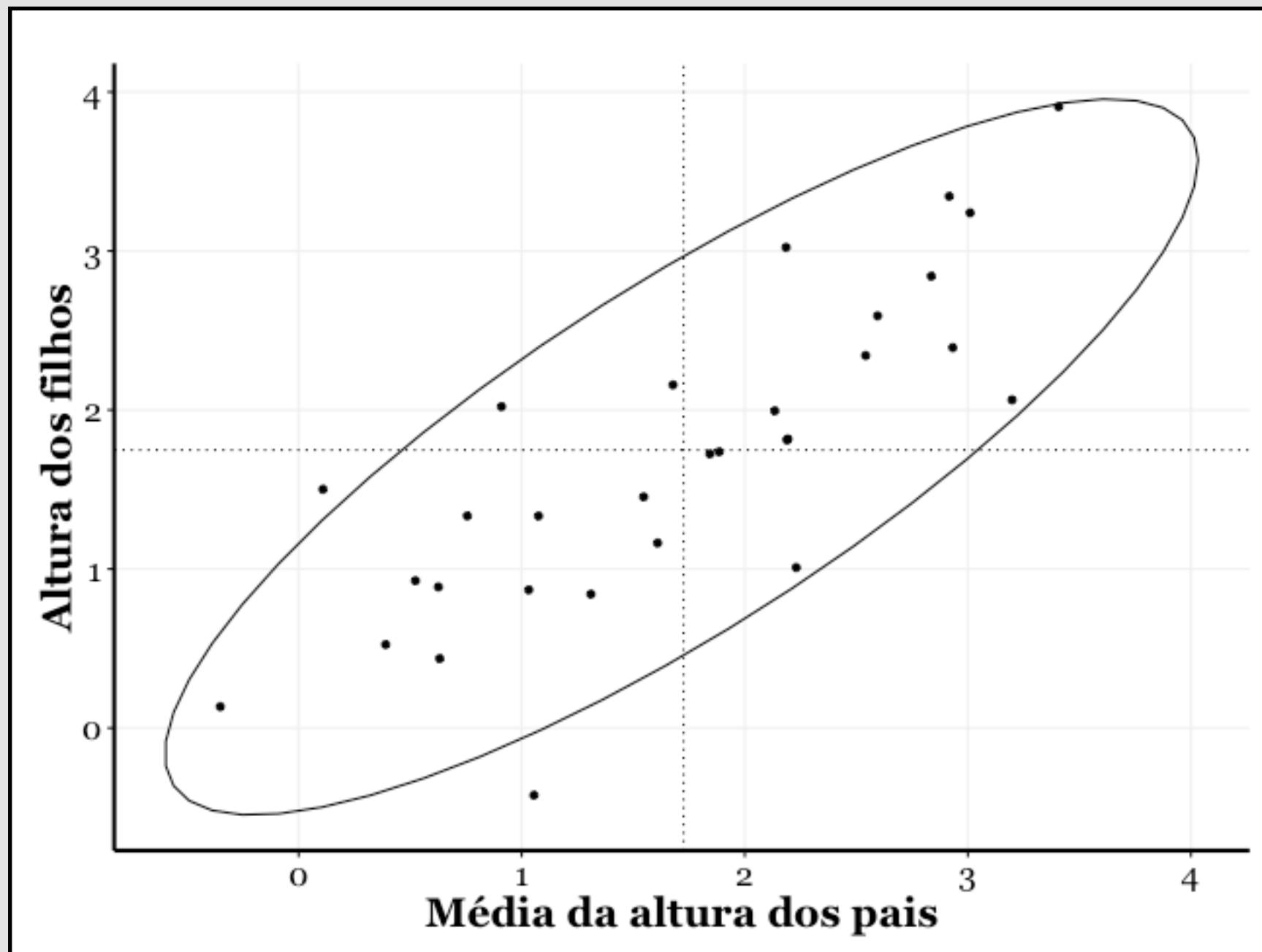
Como visualizar?



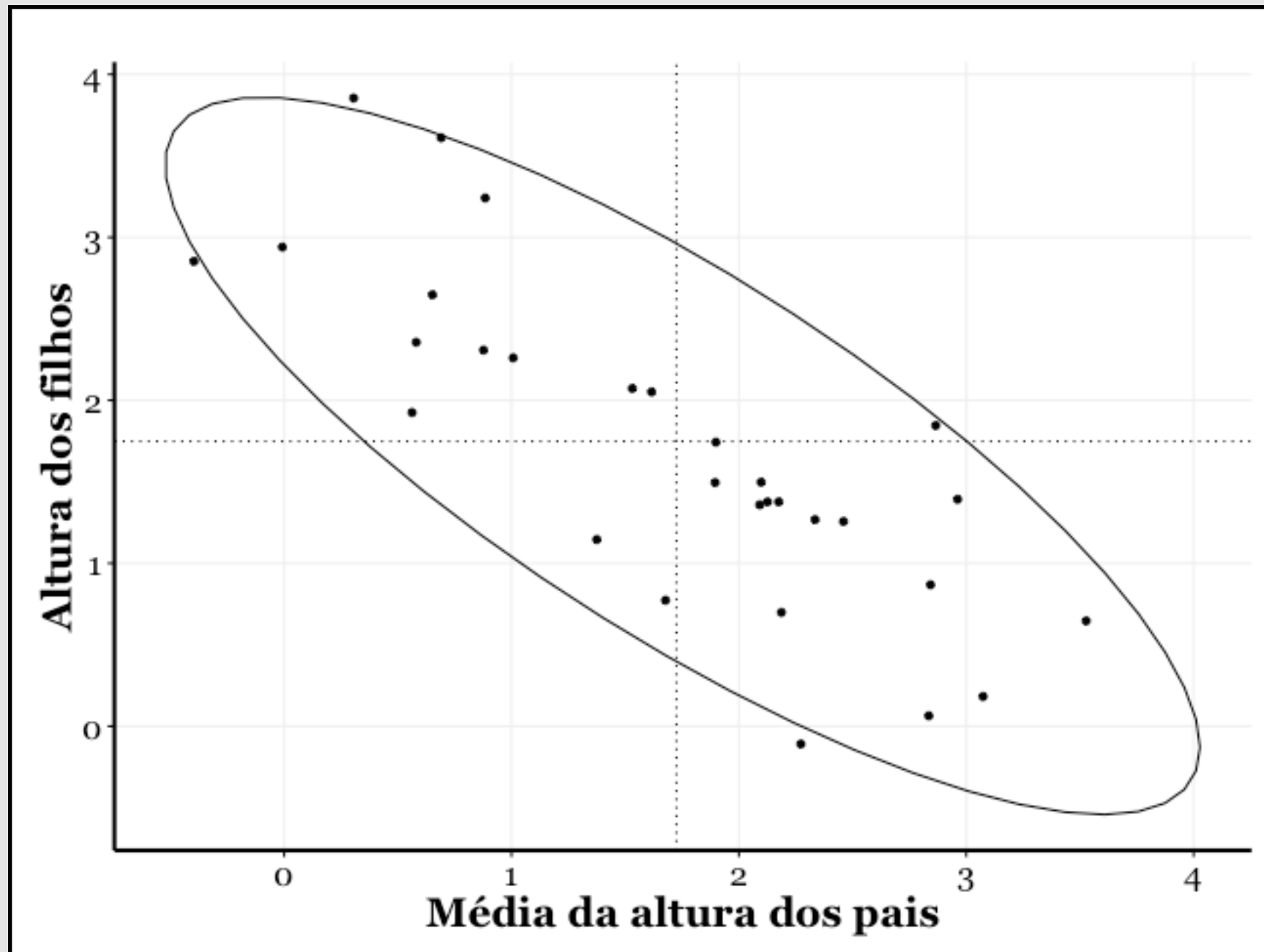
Covariância

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

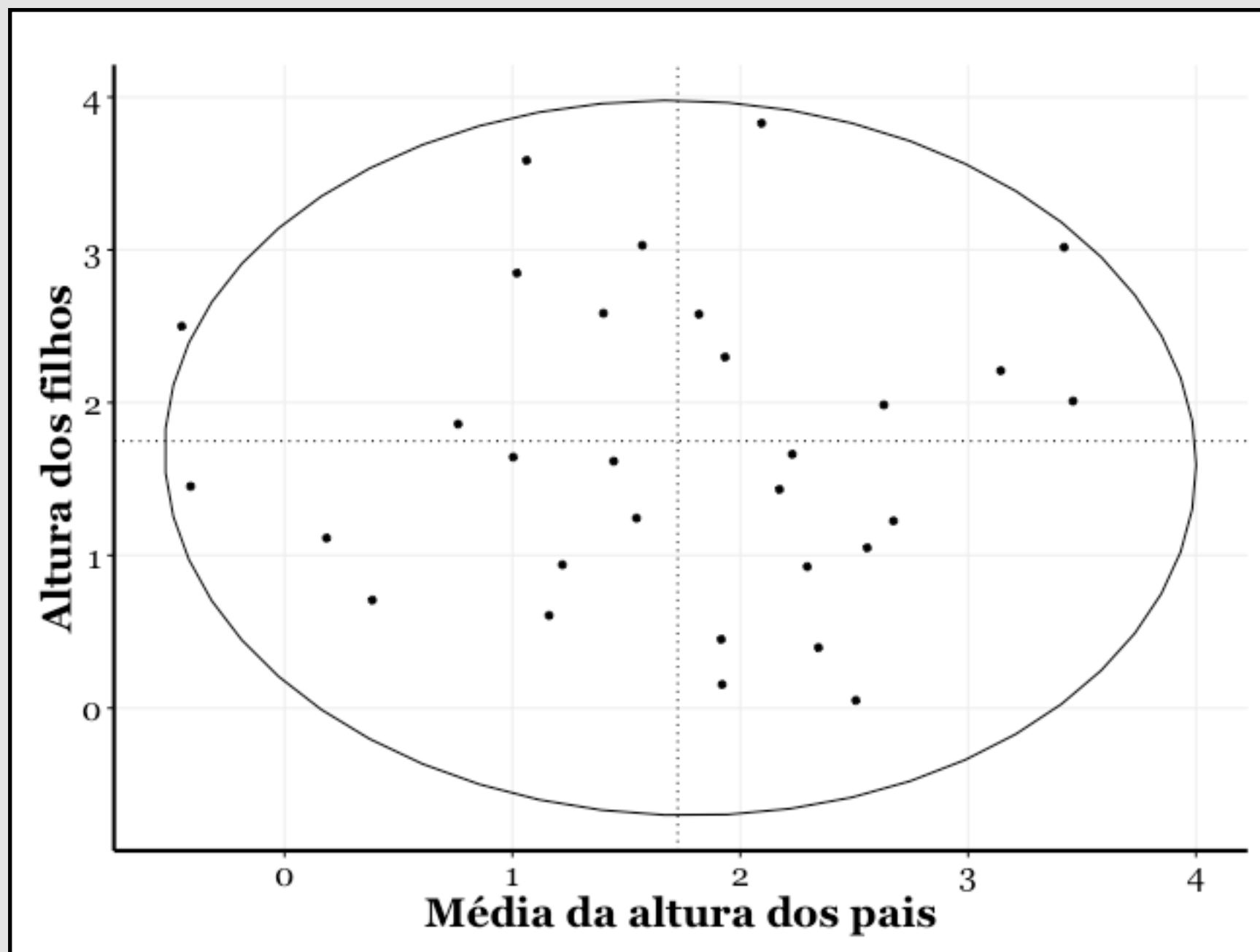
Covariância



Covariância



Covariância



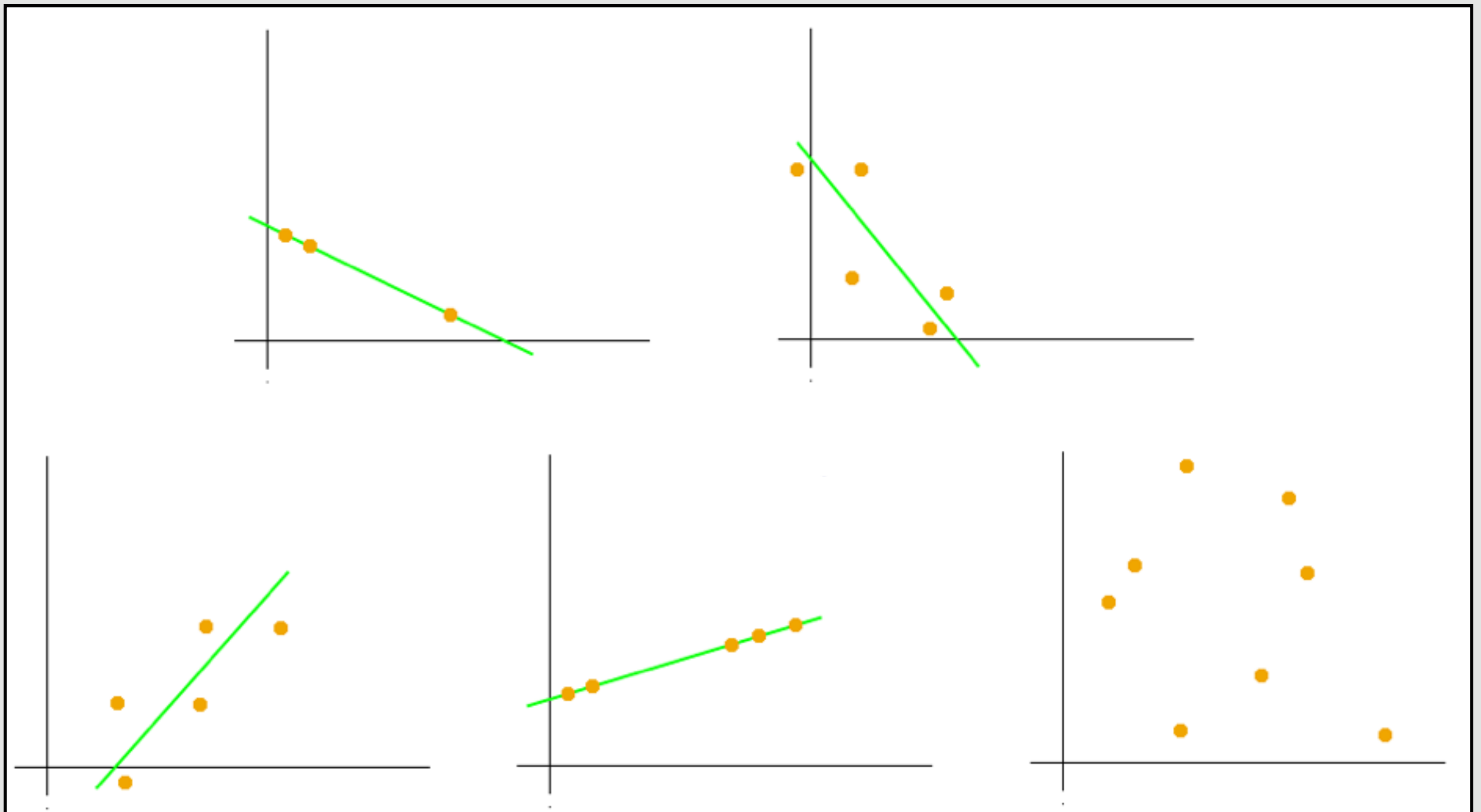
Correlação

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

$$\rho = \frac{\text{cov}(x, y)}{s_x s_y}$$

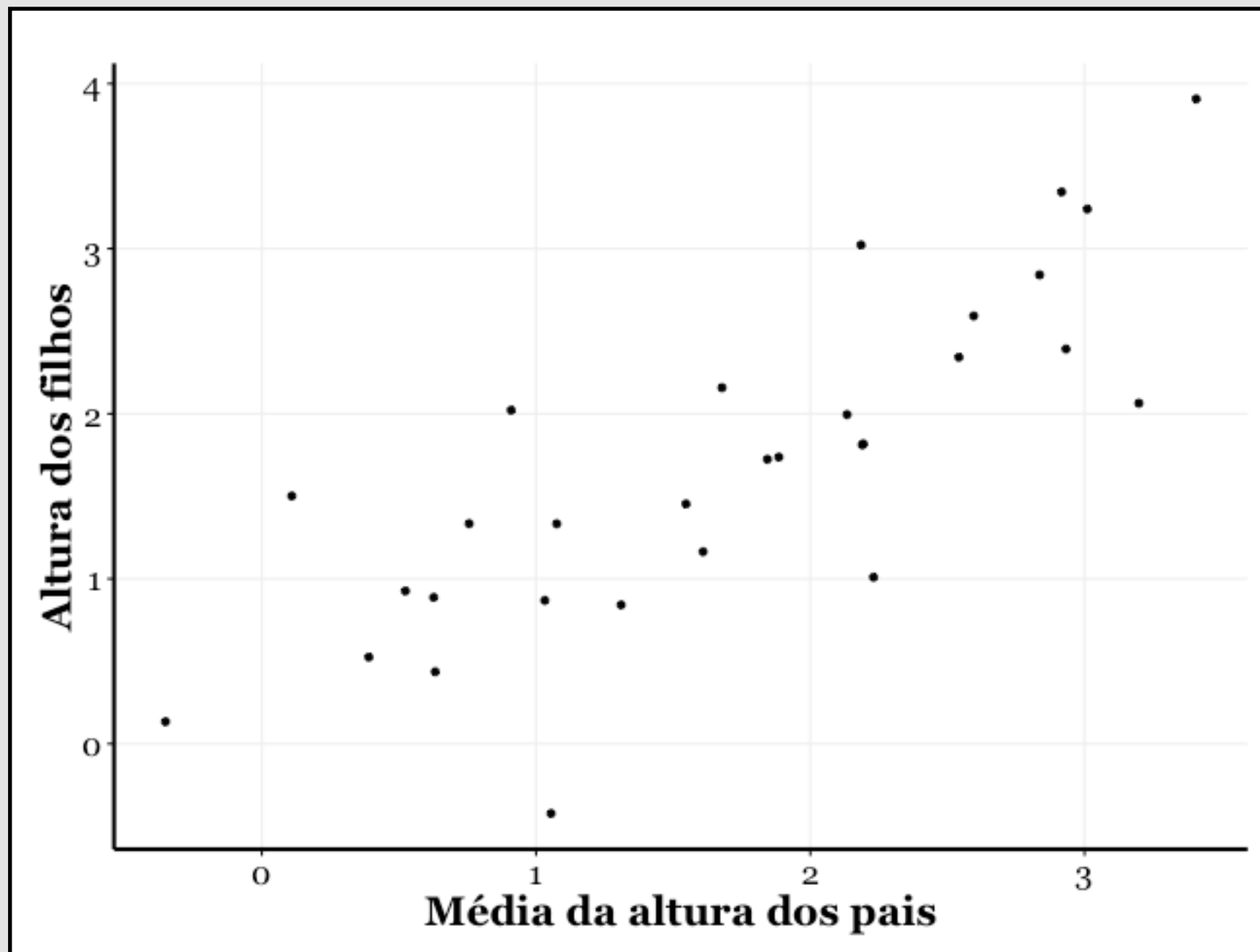
O que o coeficiente de correlação nos informa?

Correlação



Regressão Linear

Altura do filho = Altura média dos pais



Regressão Linear

- Explica como duas variáveis se relacionam por meio de uma função matemática

$$y = \alpha + \beta x$$

Regressão Linear

- Modelo estatístico:

Observação = parte previsível + erro aleatório

- Modelo da regressão linear:

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Regressão Linear

- Modelo estatístico:

Observação = parte previsível + erro aleatório

- Modelo da regressão linear:

$$y_i = \alpha + \beta x_i + \epsilon_i$$



Intercepto

Regressão Linear

- Modelo estatístico:

Observação = parte previsível + erro aleatório

- Modelo da regressão linear:

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Inclinação da reta

Regressão Linear

- Modelo estatístico:

Observação = parte previsível + erro aleatório

- Modelo da regressão linear:

$$y_i = \alpha + \beta x_i + \epsilon_i$$



Erro

Regressão Linear

- Significado dos parâmetros:

$$y_i = \alpha + \beta x_i + \epsilon_i$$

- Intercepto:
- Inclinação:
- Erro:

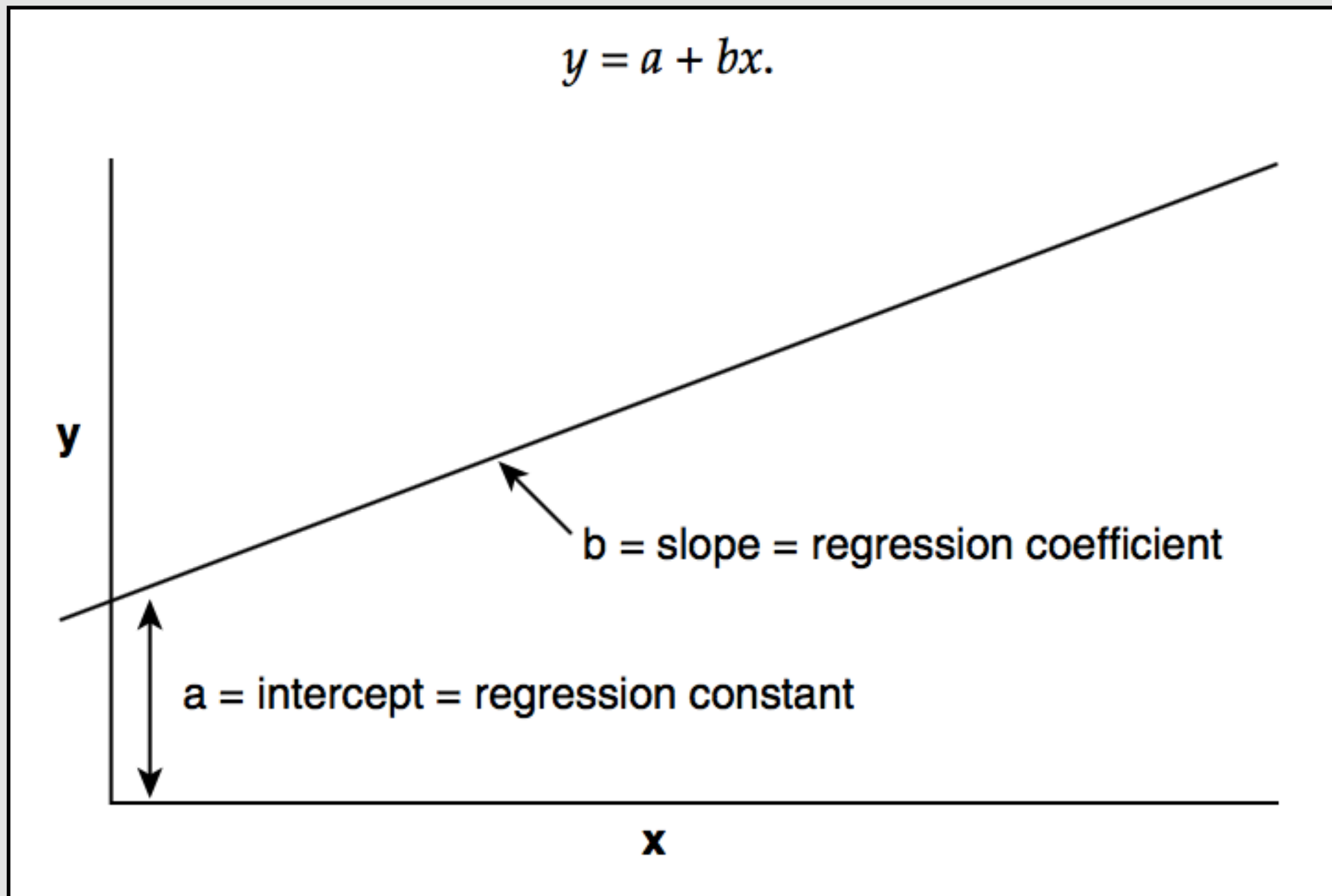
Regressão Linear

- Significado dos parâmetros:

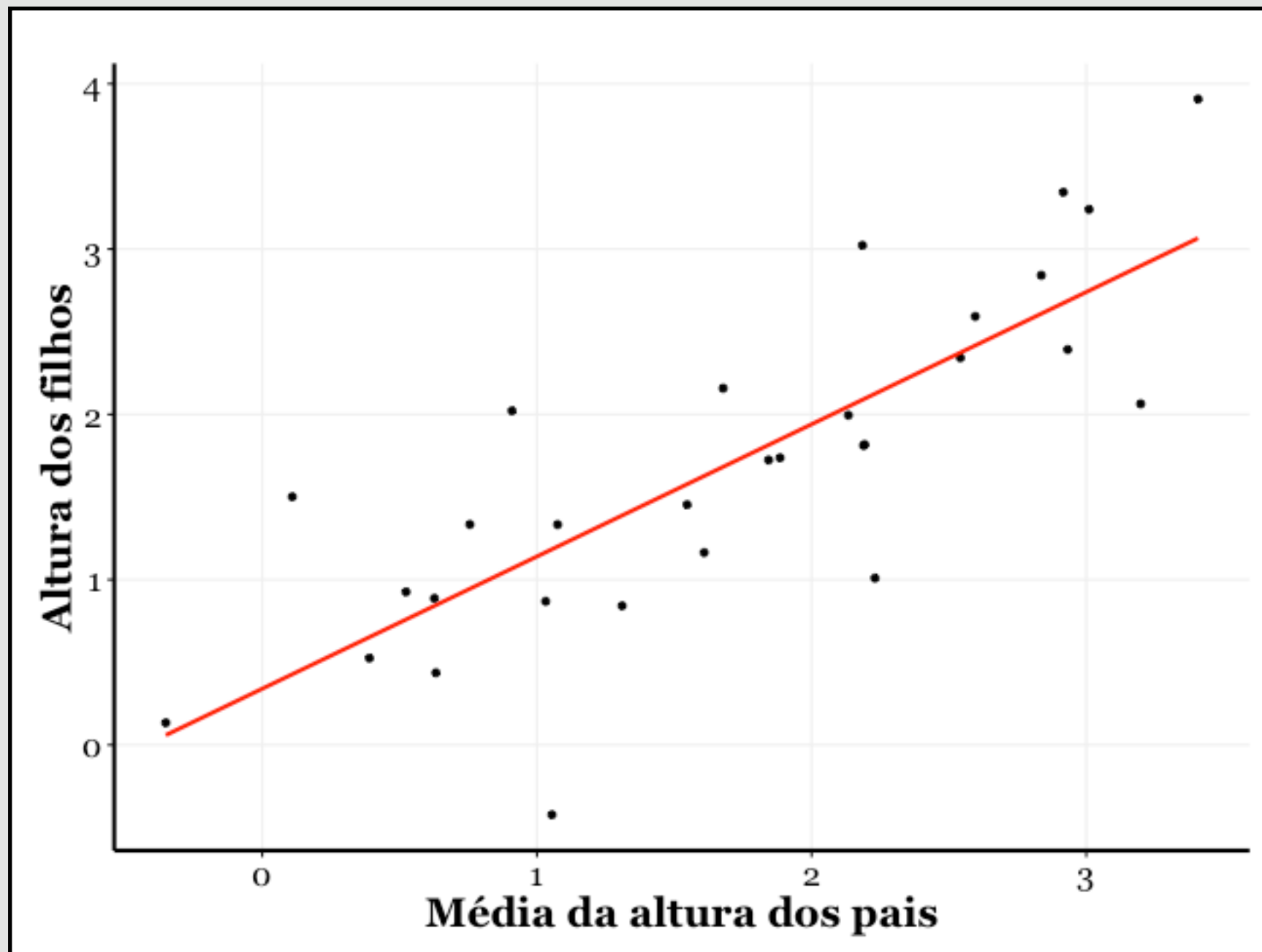
$$y_i = \alpha + \beta x_i + \epsilon_i$$

- Intercepto: valor médio de Y quando $X = 0$
- Inclinação: inclinação da reta. Para cada incremento de 1 unidade em X temos um incremento de B em Y.
- Erro: erro aleatório

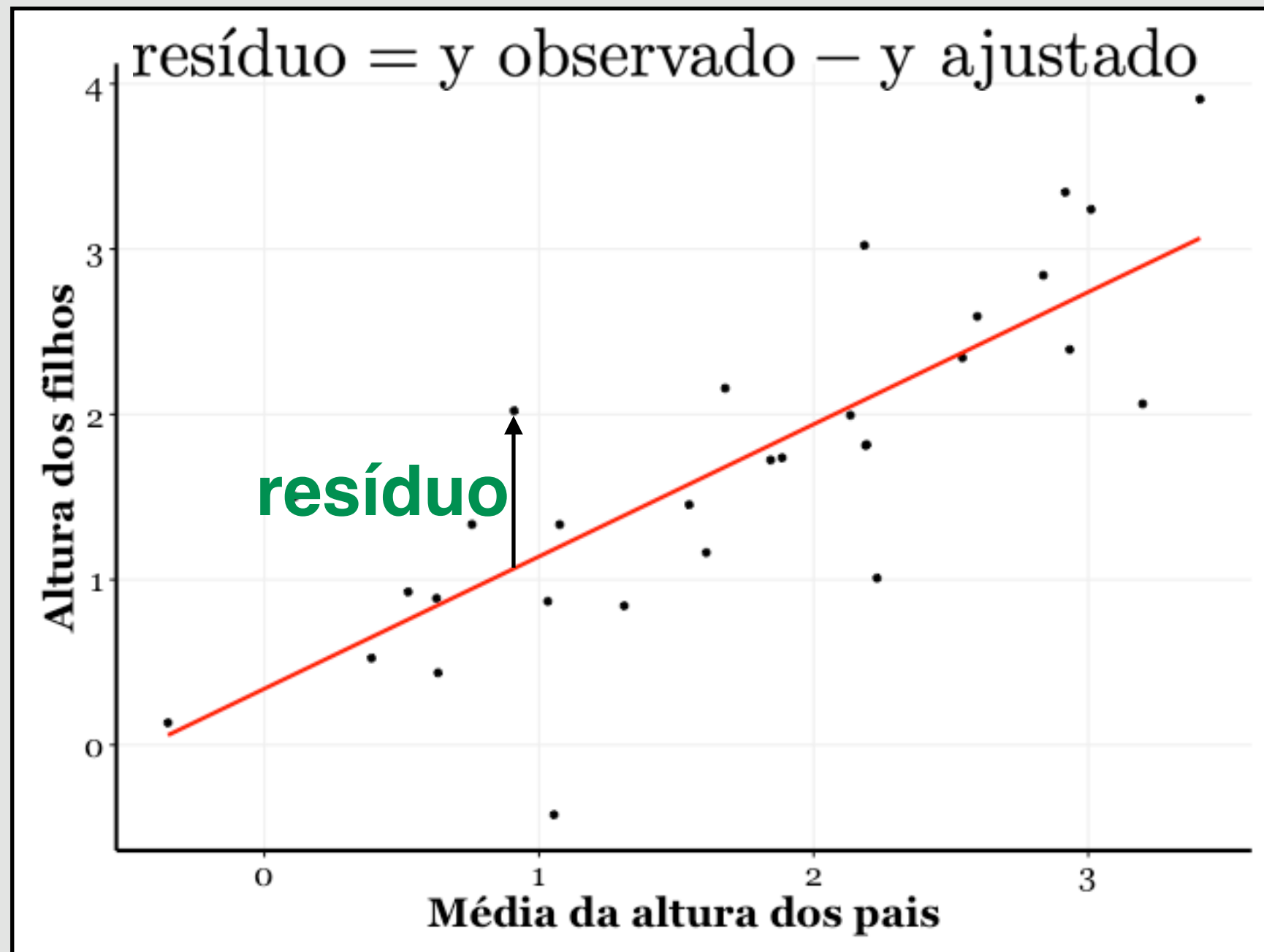
Regressão Linear



Regressão Linear



Regressão Linear



Regressão Linear

$$y_i = \alpha + \beta x_i + \epsilon_i$$

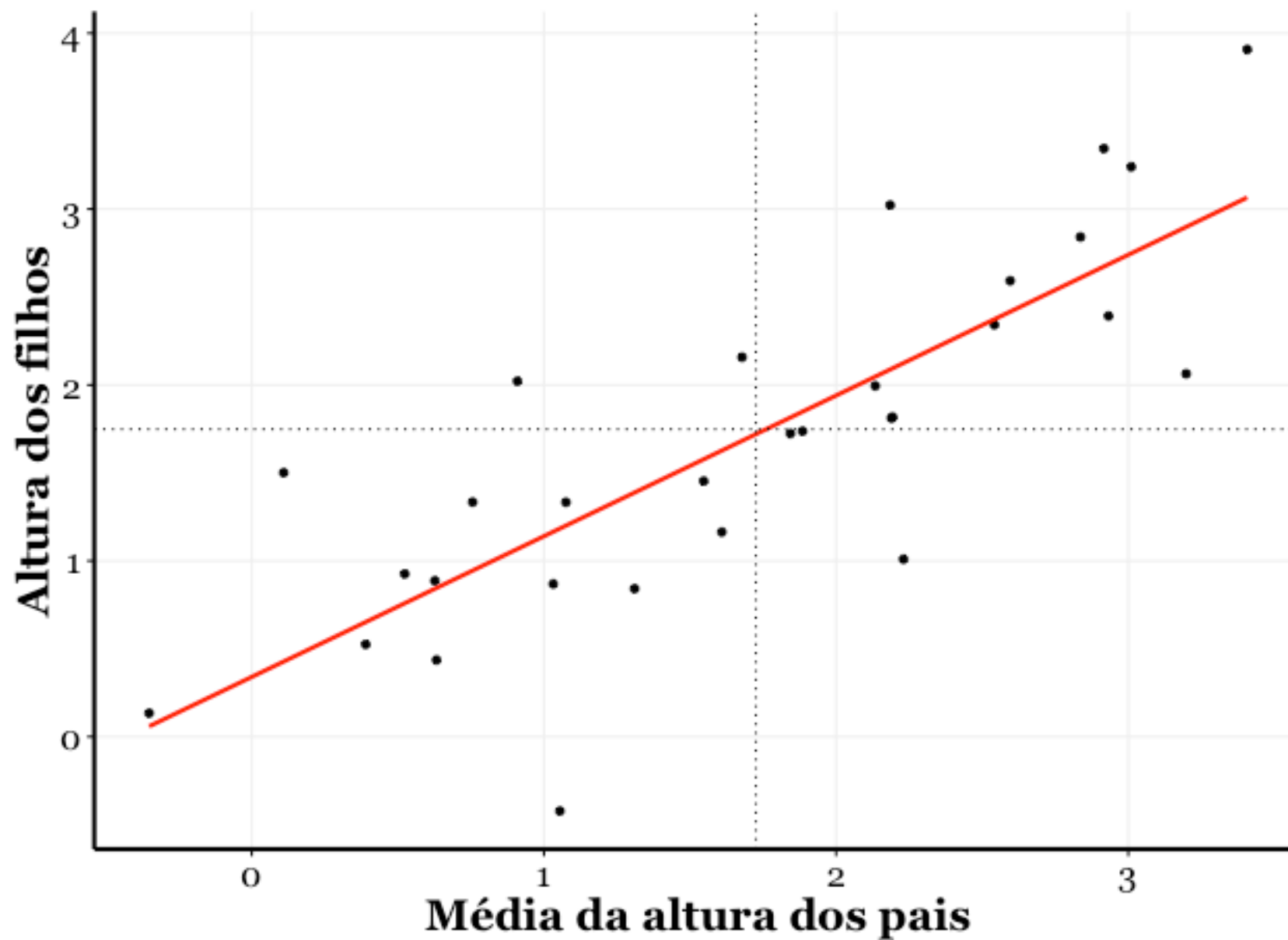
- O que precisamos estimar?

Regressão Linear

$$y_i = \alpha + \beta x_i + \epsilon_i$$

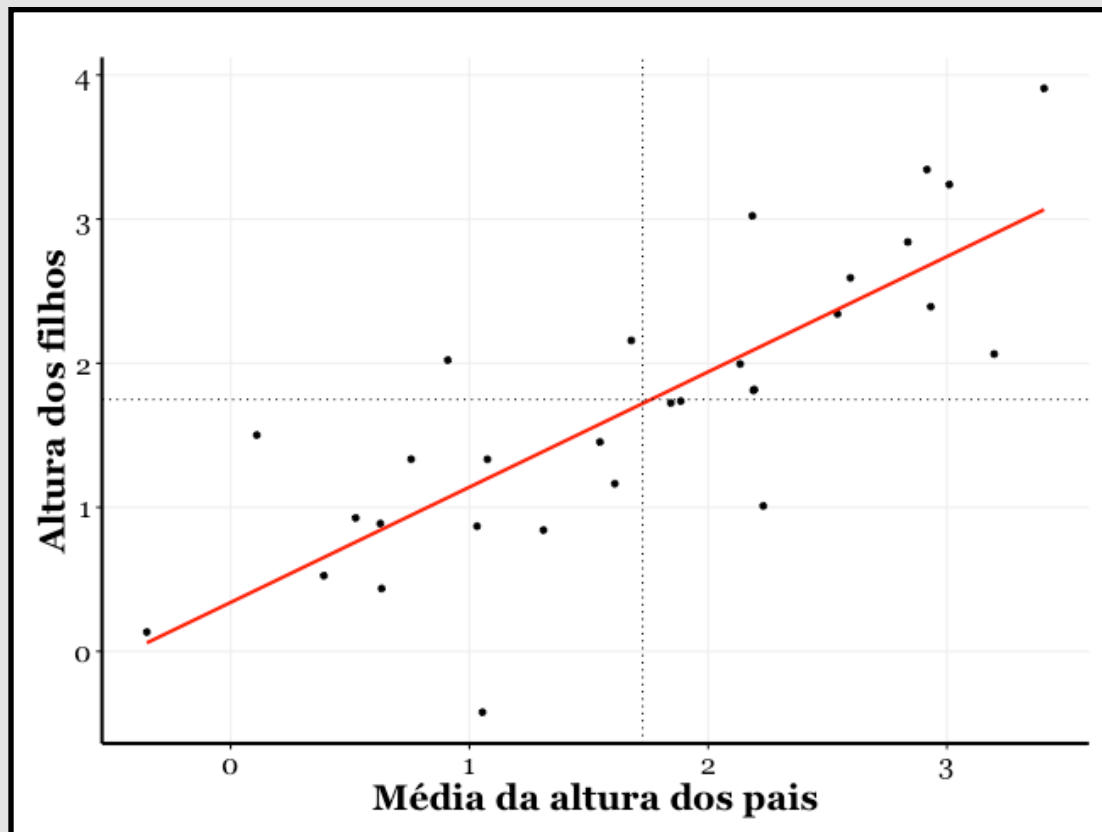
$$\beta = \frac{\text{cov}(x, y)}{s_x^2} \quad \bar{y} = \alpha + \beta \bar{x}$$

Regressão Linear



$$\bar{y} = \alpha + \beta \bar{x}$$

ANOVA da Regressão



Fonte de variação	g.l.	Soma dos Quadrados	F	P-valor
Regressão	1	18.56	49.78	<0,001
Resíduo	28	10.44		

$$R^2 = 0.64$$

- Conclusão?
- Na ANOVA, testa-se a hipótese nula de que a inclinação da reta é igual a zero

Avaliando o modelo

$$R^2 = \frac{SQReg}{SQTot}$$

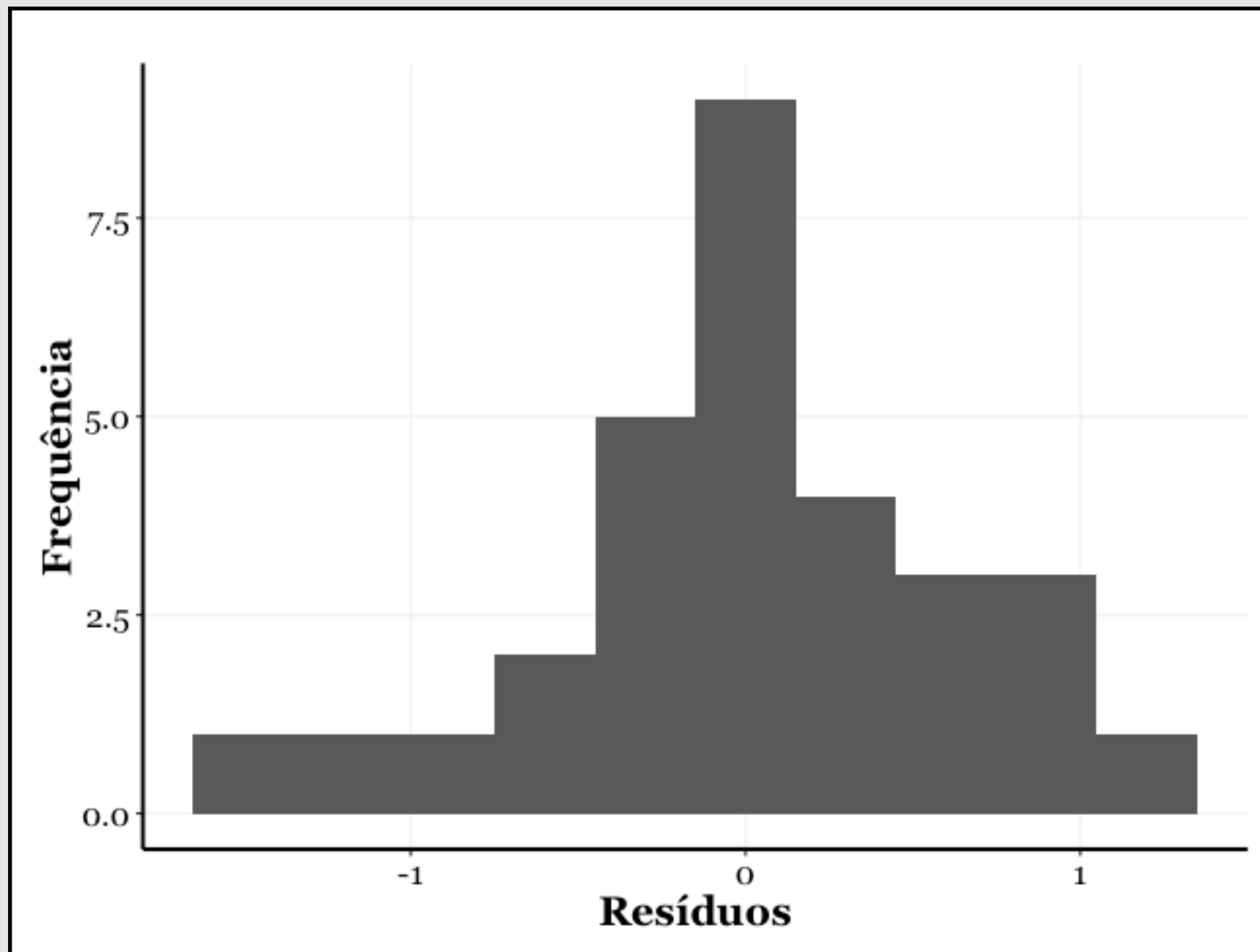
- O que essa medida nos informa?
- Quanto menor o coeficiente de explicação, menor é o grau de explicação do modelo (*i.e.*, o modelo não é adequado)

Premissas

- Independência
- Normalidade
- Homogeneidade
- X é fixo

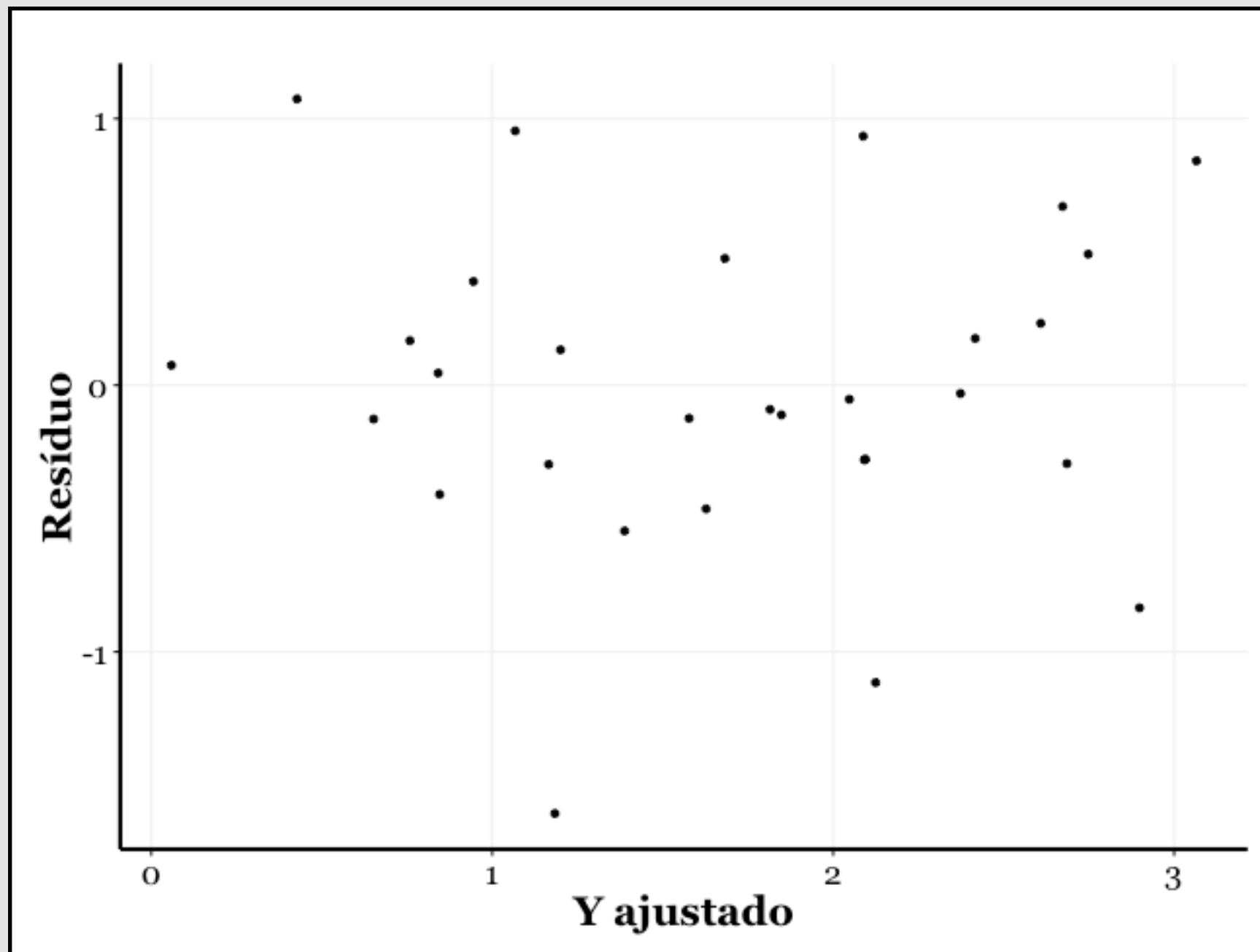
Premissas

- Normalidade



Premissas

- Homogeneidade



Premissas

- Homogeneidade

