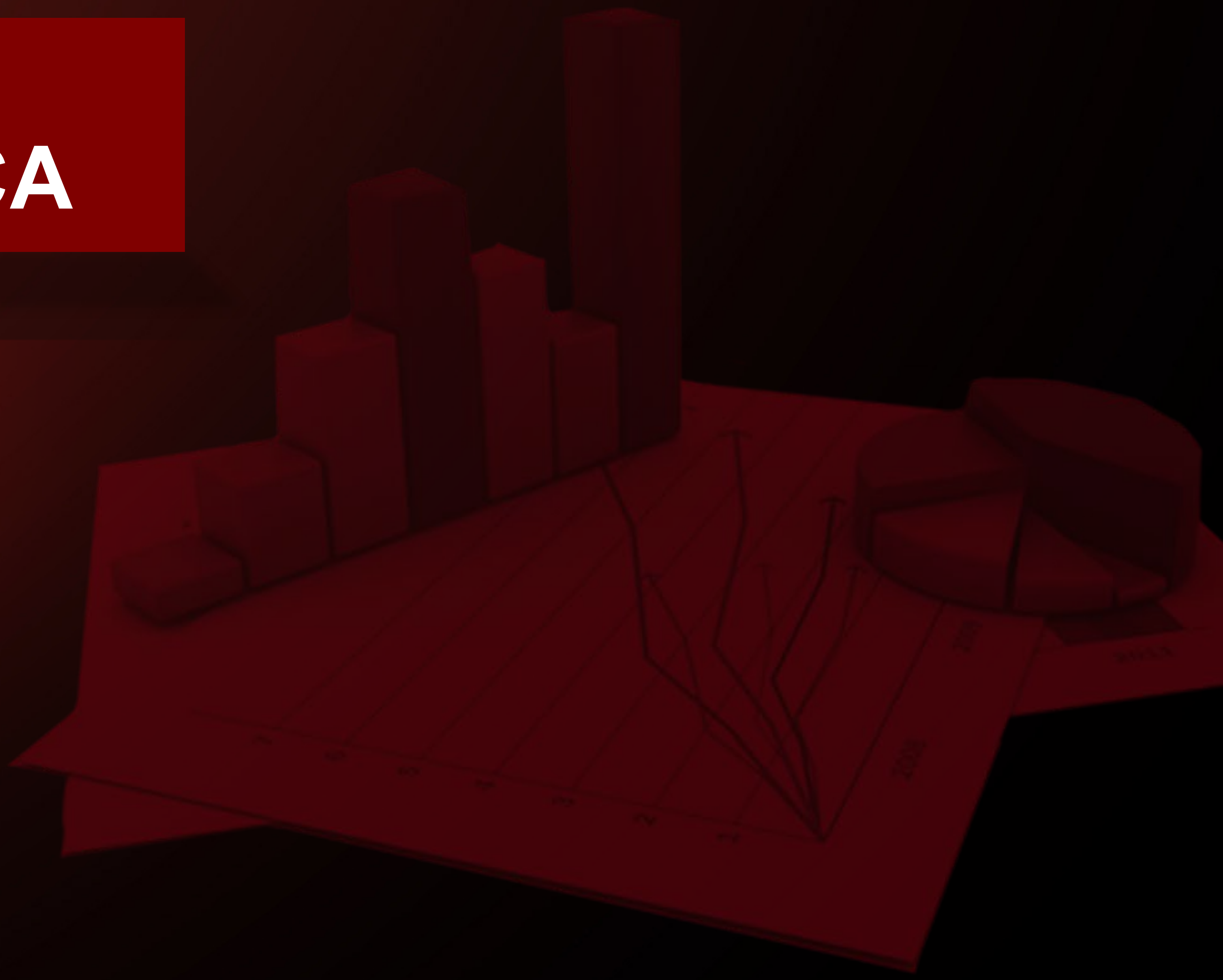
MODELAGEM E INFERÊNCIA ESTATÍSTICA

Análise de ajustes



Helios Neto
Doutor em ciência e tecnologia da computação
Gerente de inovação

O QUE VOU ESTUDAR HOJE?

Como realizar o cálculo dos parâmetros da reta:

- **Método de mínimos quadrados**

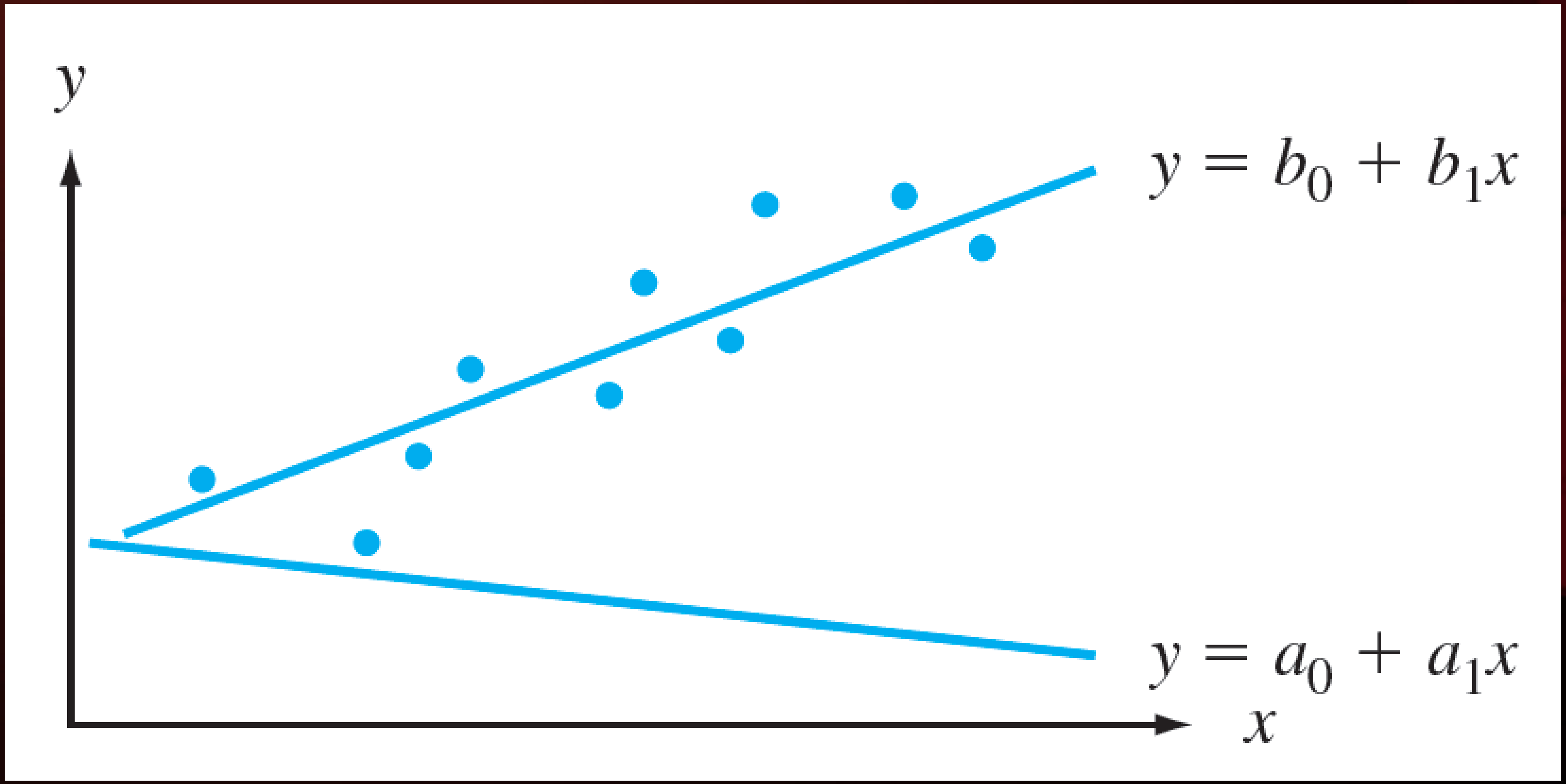
Estimar σ^2 e σ

Coeficiente de determinação r^2

Inferências sobre β_1

ESTIMAR OS PARÂMETROS

Duas estimativas diferentes da reta de regressão verdadeira



Fonte: (DEVORE, 2018, p. 464)

ESTIMAR OS PARÂMETROS

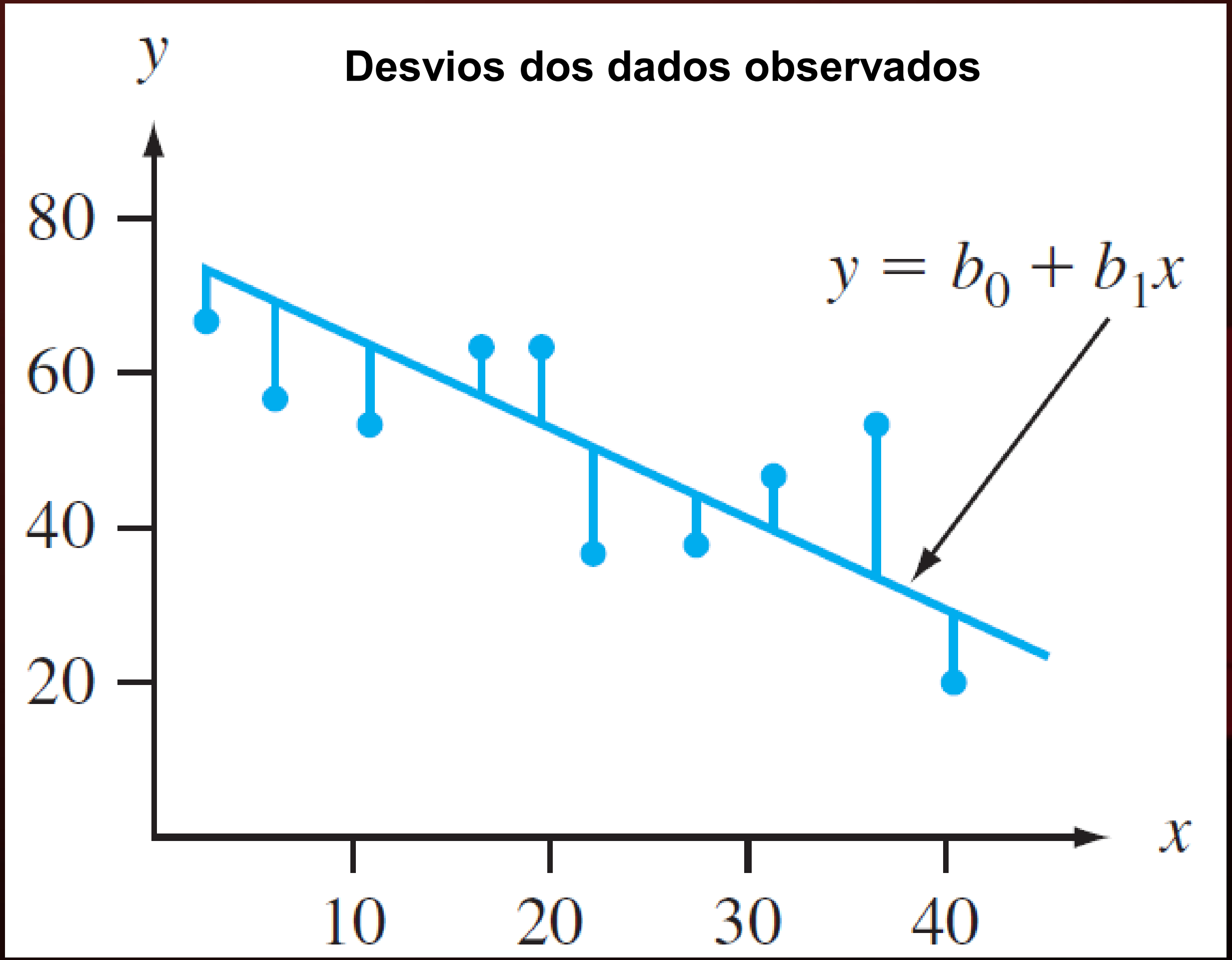
Se os parâmetros são desconhecidos, do que eu disponho?

Dados amostrais

n pares → (x_1,y_1), (x_2,y_2), (x_3,y_3), … (x_n,y_n).
y_i é o valor observado de Y_i

$$Y_i = \beta_o + \beta_1 x + \varepsilon_i$$

ESTIMAR OS PARÂMETROS



Fonte: (DEVORE, 2018, p. 464)

O desvio vertical do ponto (x_i, y_i) da reta $y = b_0 + b_1x$, é:
altura do ponto – altura da reta

SIMULAÇÃO DO MÉTODO DOS MÍNIMOS QUADRADOS

https://phet.colorado.edu/en/simulations/least-squares-regression

MÉTODO DOS MÍNIMOS QUADRADOS

Resolver o sistema

$$nb_0 + (\Sigma xi)b_1 = \Sigma y_i$$
$$(\Sigma x_i)b_0 + (\Sigma x_i^2)b_1 = \Sigma x_i y_i$$

A estimativa dos mínimos quadrados do coeficiente de inclinação β_1 da reta de regressão é

$$b_1 = \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

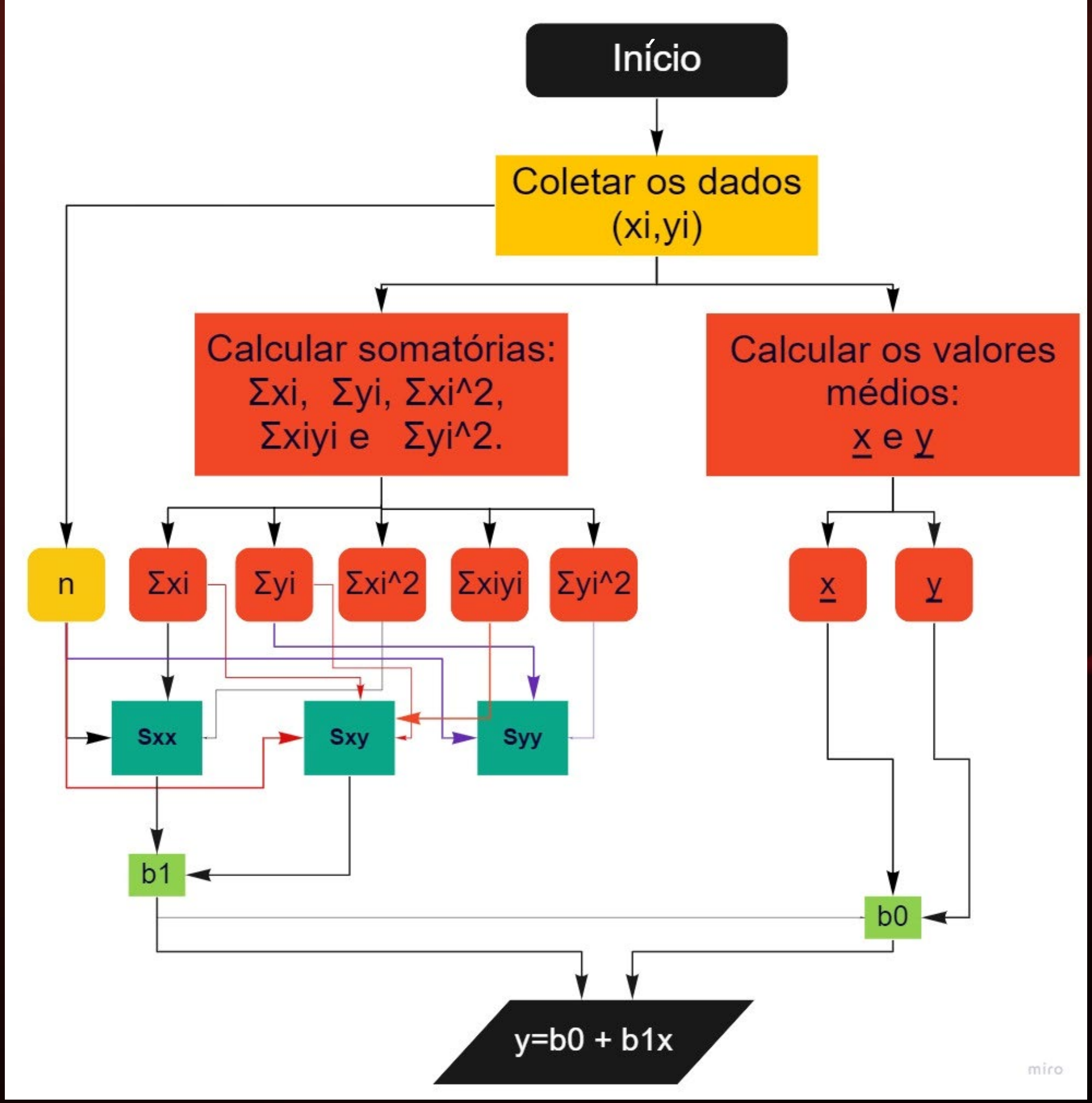
As fórmulas de cálculo do numerador e denominador de β_1 são

$$S_{xy} = \Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)/n \qquad S_{xx} = \Sigma x_i^2 - (\Sigma x_i)^2/n$$

A estimativa dos mínimos quadrados do β_0 da reta de regressão verdadeira é

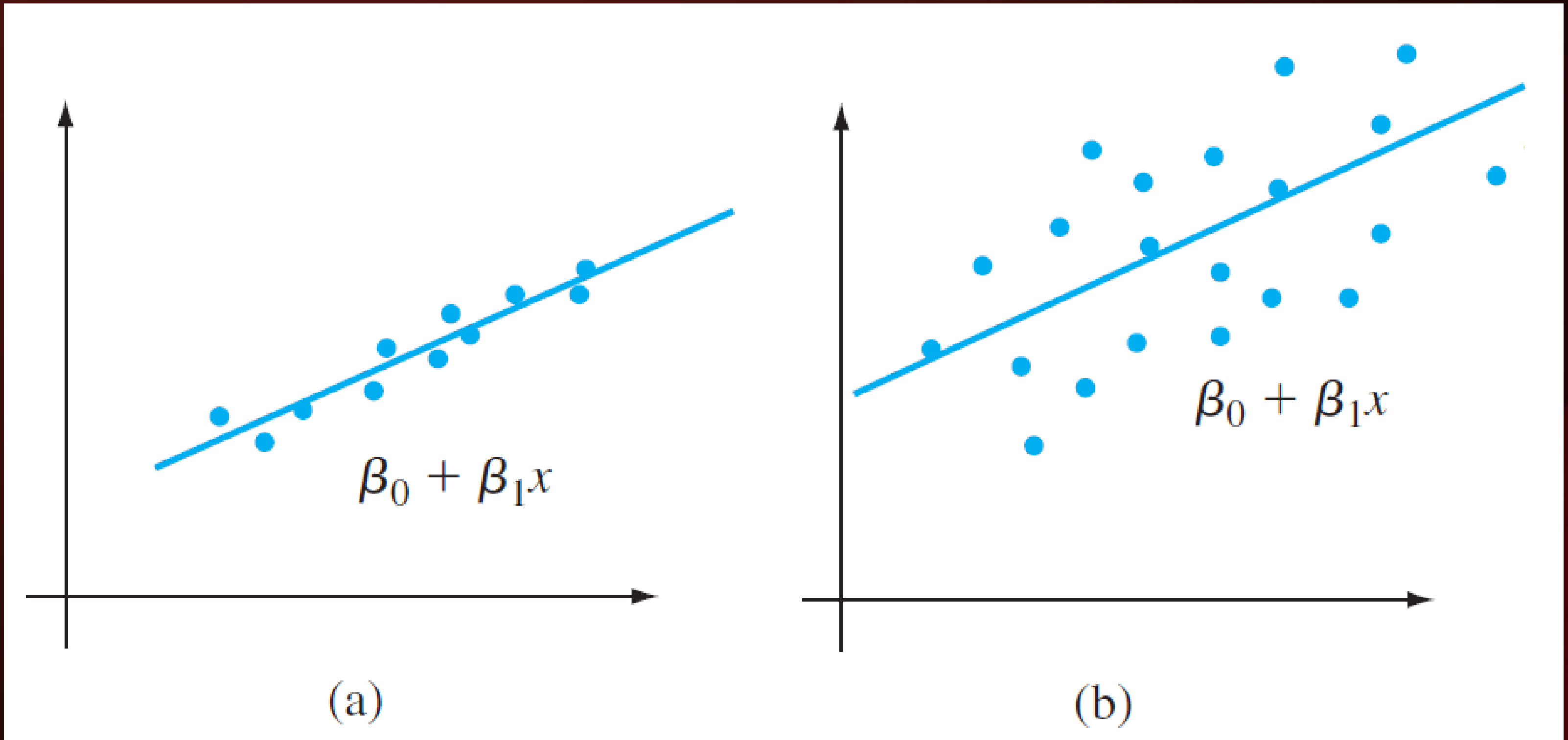
$$b_0 = \hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$$

ESTIMAR OS PARÂMETROS



Resumo do processo

ESTIMAR σ^2 e σ



Resíduos:
Diferenças +/-
$y_i - \hat{y}_i$
Estimativa de σ^2

Amostra típica para σ^2: (a) pequeno; (b) grande.
Fonte: (DEVORE, 2018, p. 467)

ESTIMAR σ^2 e σ

A variância é definida pela seguinte equação

$$\sum (y_i - \bar{y})^2$$

A soma de quadrados dos erros (que equivale à soma de quadrados dos resíduos), representada por SQE, é

$$SQE = \sum (y_i - \hat{y}_i)^2 = \sum [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

e a estimativa de σ^2 é

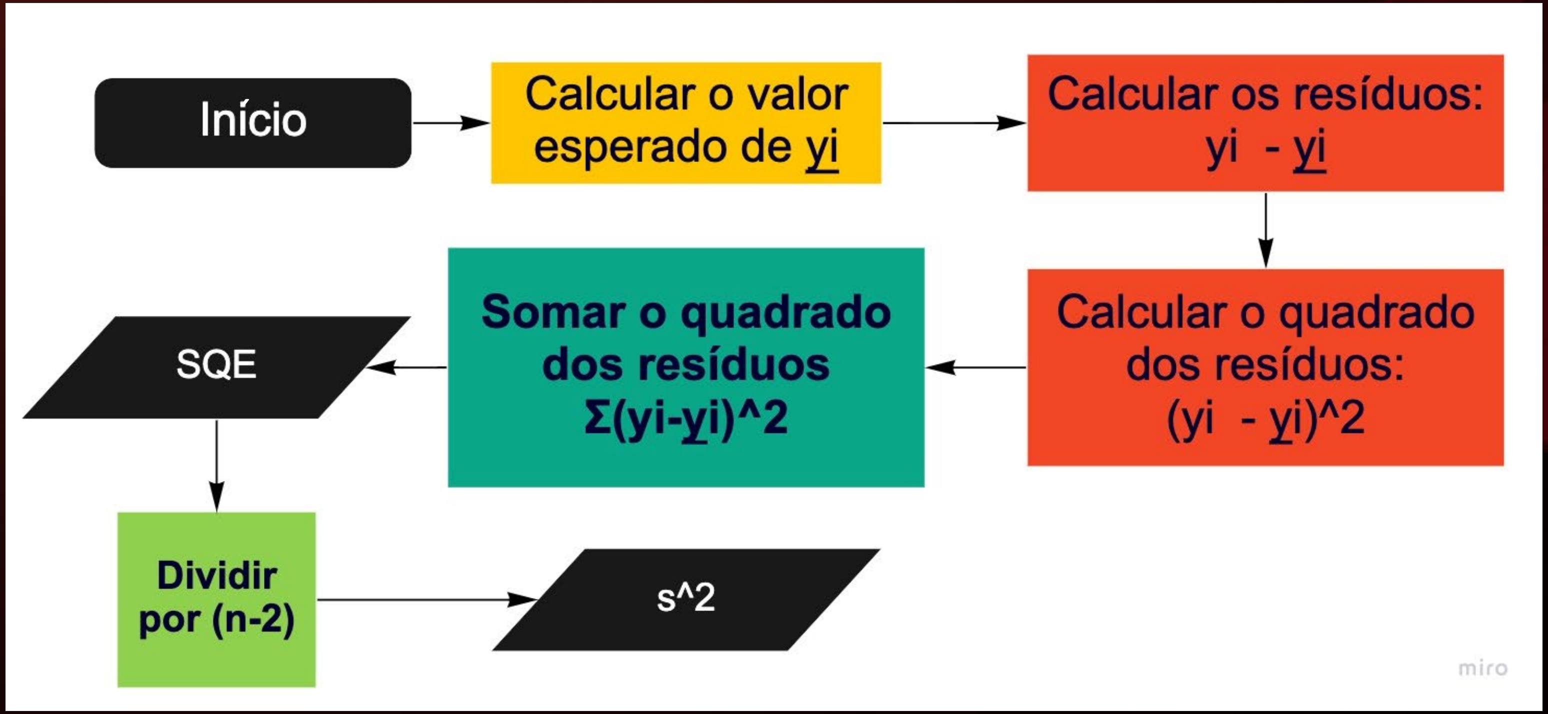
$$\hat{\sigma}^2 = s^2 = \frac{SQE}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

$$SQE = \Sigma y_i^2 - \hat{\beta}_0 \Sigma y_i - \hat{\beta}_1 \Sigma x_i y_i = S_{yy} - \hat{\beta}_1 S_{xy}$$

Fonte: (DEVORE, 2018, p. 469)

ESTIMAR σ^2 e σ

Procedimento



COEFICIENTE DE DETERMINAÇÃO r^2

SQE (Soma de Quadrados dos Erros) → medida da quantidade de variação em y não explicada.

$$\text{SQE} = S_{yy} - \hat{\beta}_1 S_{xy}$$

SQT (Soma de Quadrados Total)

$$\text{SQT} = S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

SQR (Soma de Quadrados da Regressão)

$$SQR = SQT - SQE$$

COEFICIENTE DE DETERMINAÇÃO r²

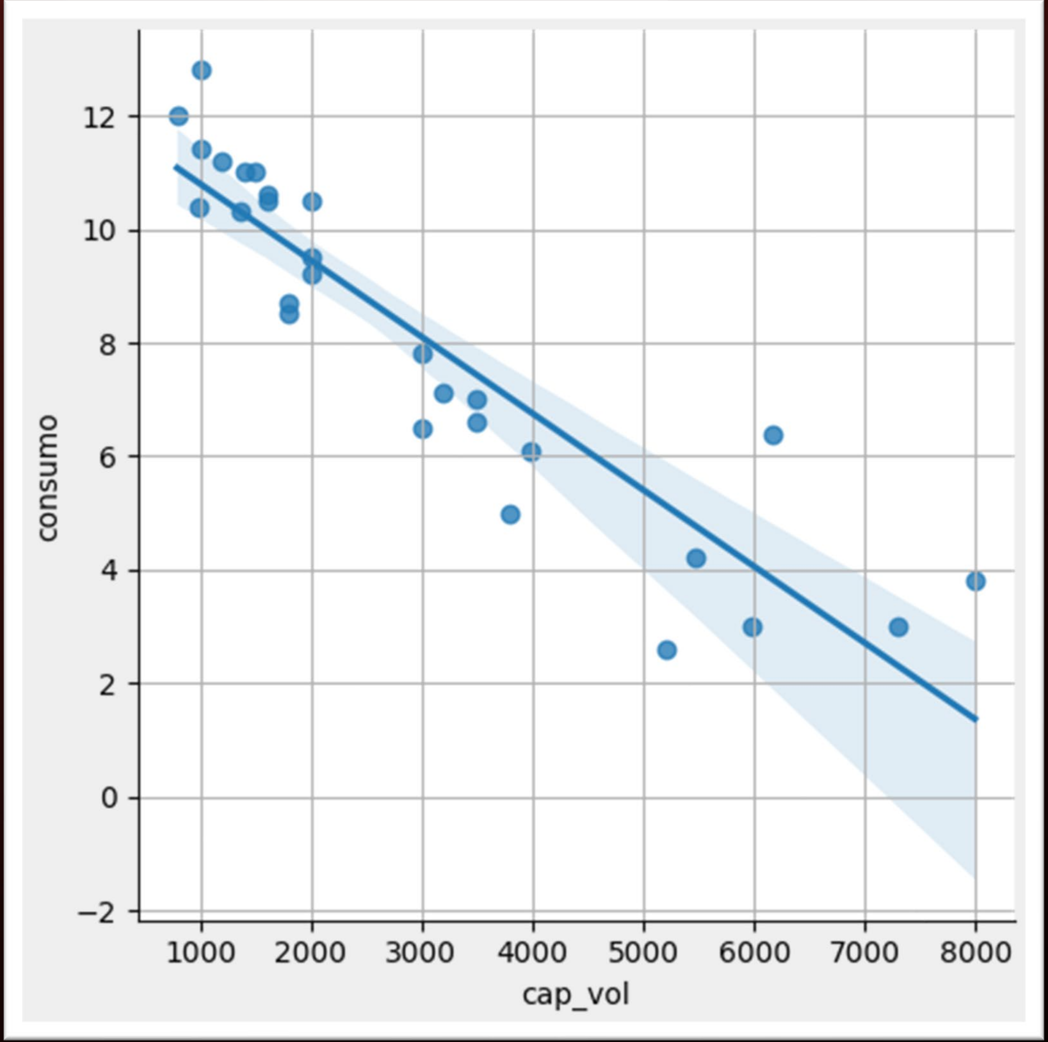
Coeficiente de determinação, proporção da variação do *y* observado que pode ser explicada pelo modelo de regressão linear simples.

$$r^2 = 1 - \frac{SQE}{SQT}$$

Valores altos de r^2 → o modelo de regressão linear simples pode ser usado para explicar a variação de *y*.

Valores baixos de r^2 → procurar outro modelo.

INFERÊNCIAS SOBRE β_1



INFERÊNCIAS SOBRE β_1

Intervalo de confiança para β_1

$$P\left(-t_{\alpha/2,n-2} < \frac{\hat{\beta}_1 - \beta_1}{S_{\beta_1}} < t_{\alpha/2,n-2}\right) = 1 - \alpha$$

Um IC de 100(1 –α)% para a inclinação β_1 da reta de regressão verdadeira é

$$\hat{\beta}_1 = \beta_1 \pm \boldsymbol{t}_{\alpha/2,n-2}\ s_{\hat{\beta}_1}$$

MODELAGEM E INFERÊNCIA ESTATÍSTICA

Análise de ajustes