MODELOS REGRESSIVOS & ANÁLISES MULTIVARIADAS

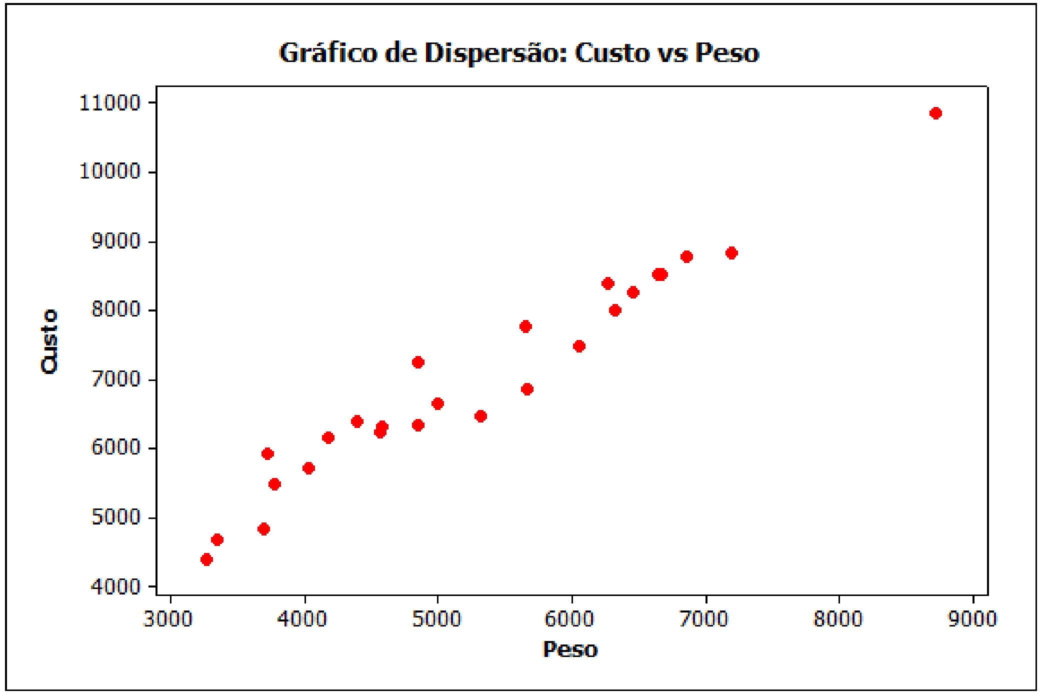
Correlação e Regressão linear simples

Profs. Atsler Luana Lehun e Lidiany Doreto Cavalcanti

Maringá, 2025

ANÁLISE DE CORRELAÇÃO

A correlação é usada para explorar a relação entre duas variáveis, ou seja ela mede a força de associação entre duas variáveis e a direção da relação.



DIREÇÃO

POSITIVA

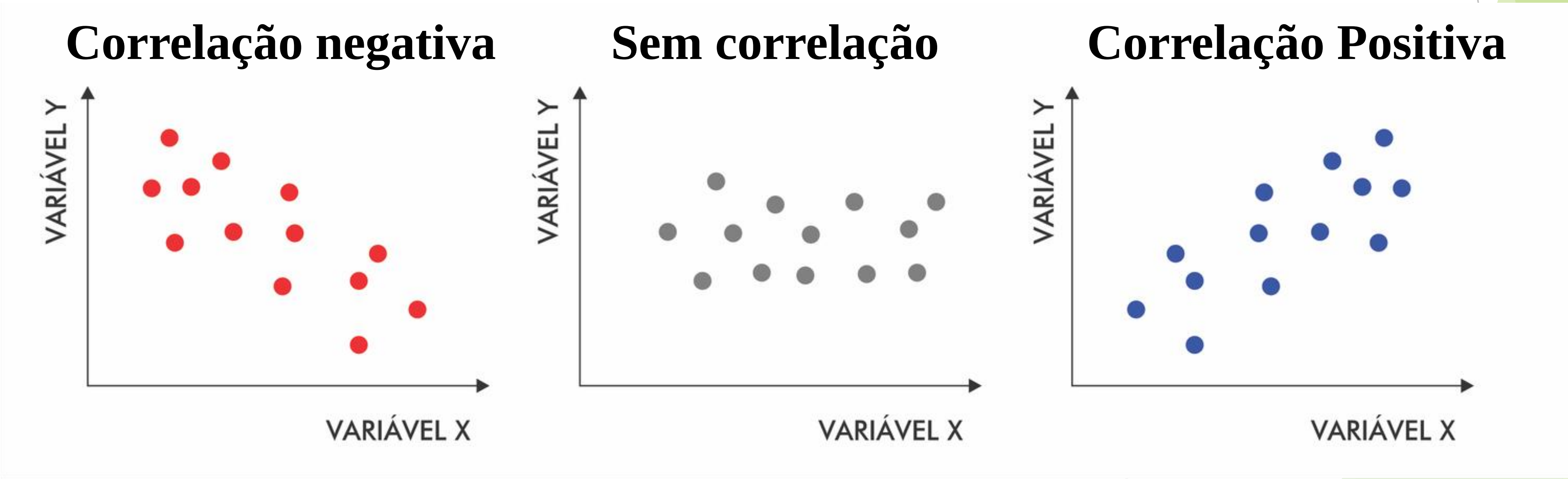
NEGATIVA

O que observar?

VALORES DA CORRELAÇÃO -1 ⇔ 1

VALOR DE P

INSPEÇÃO VISUAL DO GRÁFICO



ANÁLISE DE CORRELAÇÃO

Pressupostos

O número de observações devem ser igual entre o eixo X e eixo Y

Correlação de Pearson → **Distribuição Normal**

Correlação de Spearman → **Distribuição Não-normal**

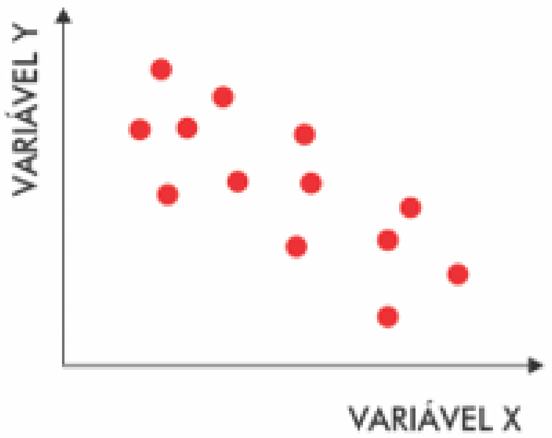
Interpretação

```
        Shapiro-Wilk normality test

data:  dados$mpg
W = 0.94756, p-value = 0.1229
```

Normalidade

```
        Pearson's product-moment correlation

data:  dados$wt and dados$mpg
t = -9.559, df = 30, p-value = 1.294e-10
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9338264 -0.7440872
sample estimates:
       cor
-0.8676594
```

Correlação negativa



Exemplo de Correlação

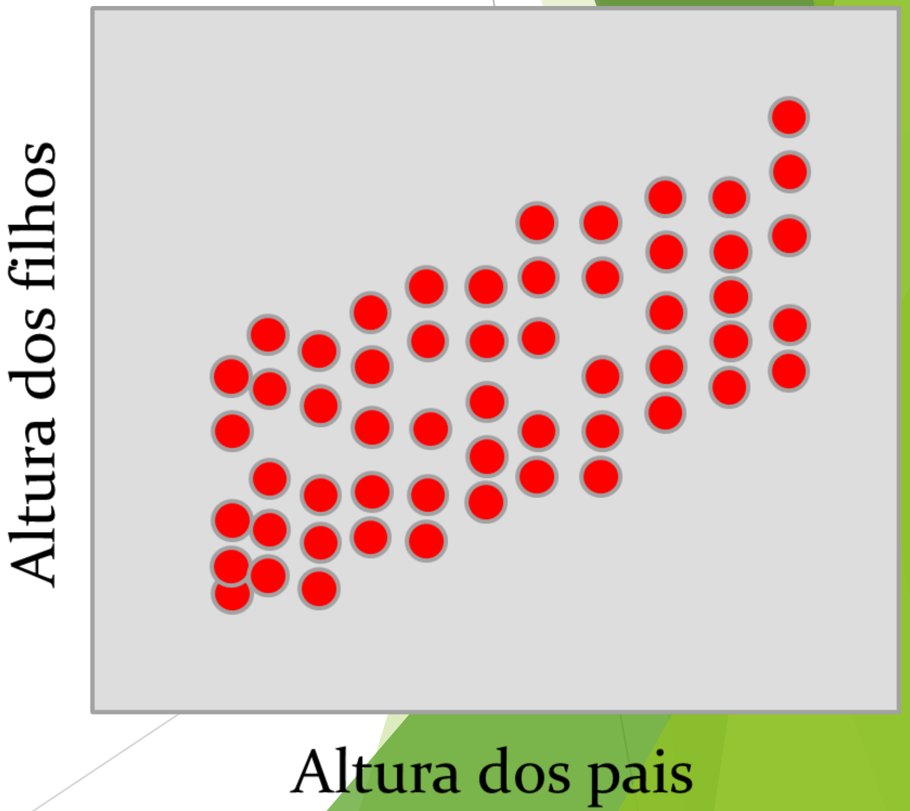
Pacote mtcars

mpg: Milhas por galão (eficiência de combustível)
wt: Peso do carro (em milhares de libras)

REGRESSÃO LINEAR SIMPLES

- Fornece uma equação que descreve o comportamento linear de uma das variáveis em função do comportamento da outra variável.

Através da formula, nós conseguimos prever o comportamento de y em função da alteração no x para variáveis continuas.



REGRESSÃO LINEAR SIMPLES

Será que o X explica a variação em Y?
Toda a regressão tem a seguinte equação:

$$\mathbf{Y = a + bX}$$

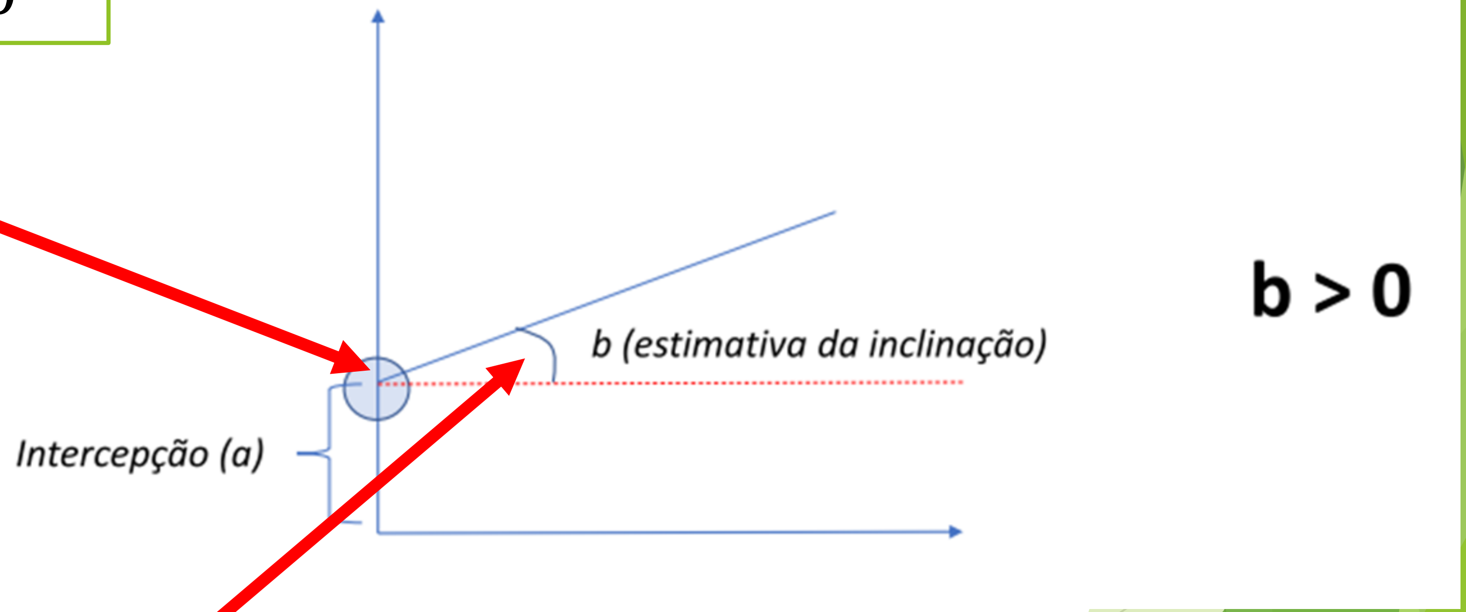
a ➔ intercepto (o ponto que a reta toca o eixo y quando x=0) → onde a reta ajustada dos mínimos quadrados cruza

b ➔ representa a **inclinação** da reta. Caso b é positivo a reta é ascendente, caso negativo a reta é descendente

REGRESSÃO LINEAR SIMPLES

a (intercepto): o ponto que a reta toca o eixo y quando x=0

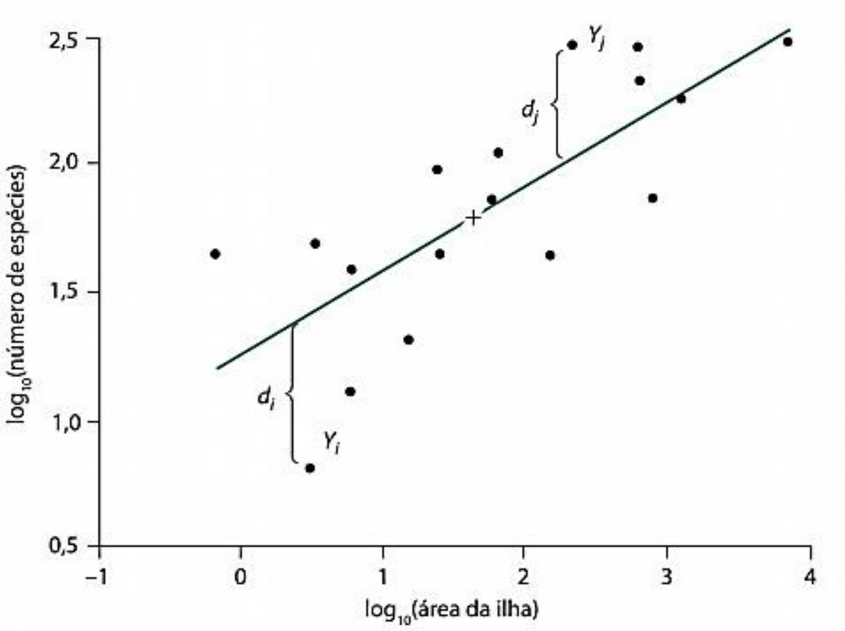
$$y = \alpha + \beta x$$

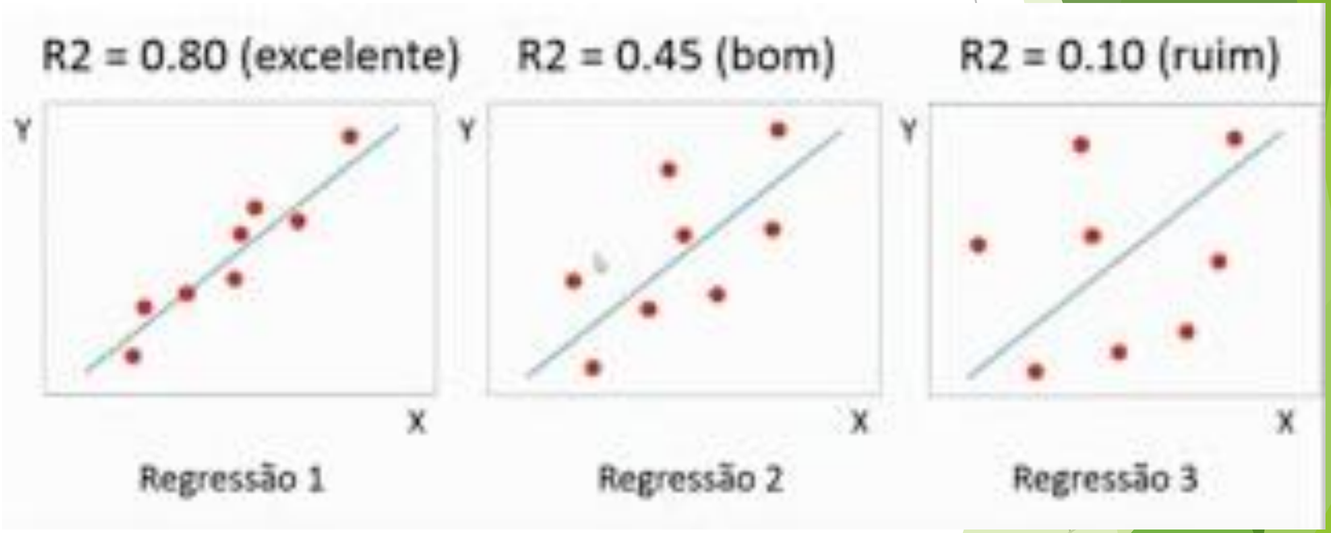


b: representa a **inclinação** da reta

REGRESSÃO LINEAR SIMPLES

Método dos mínimos quadrados



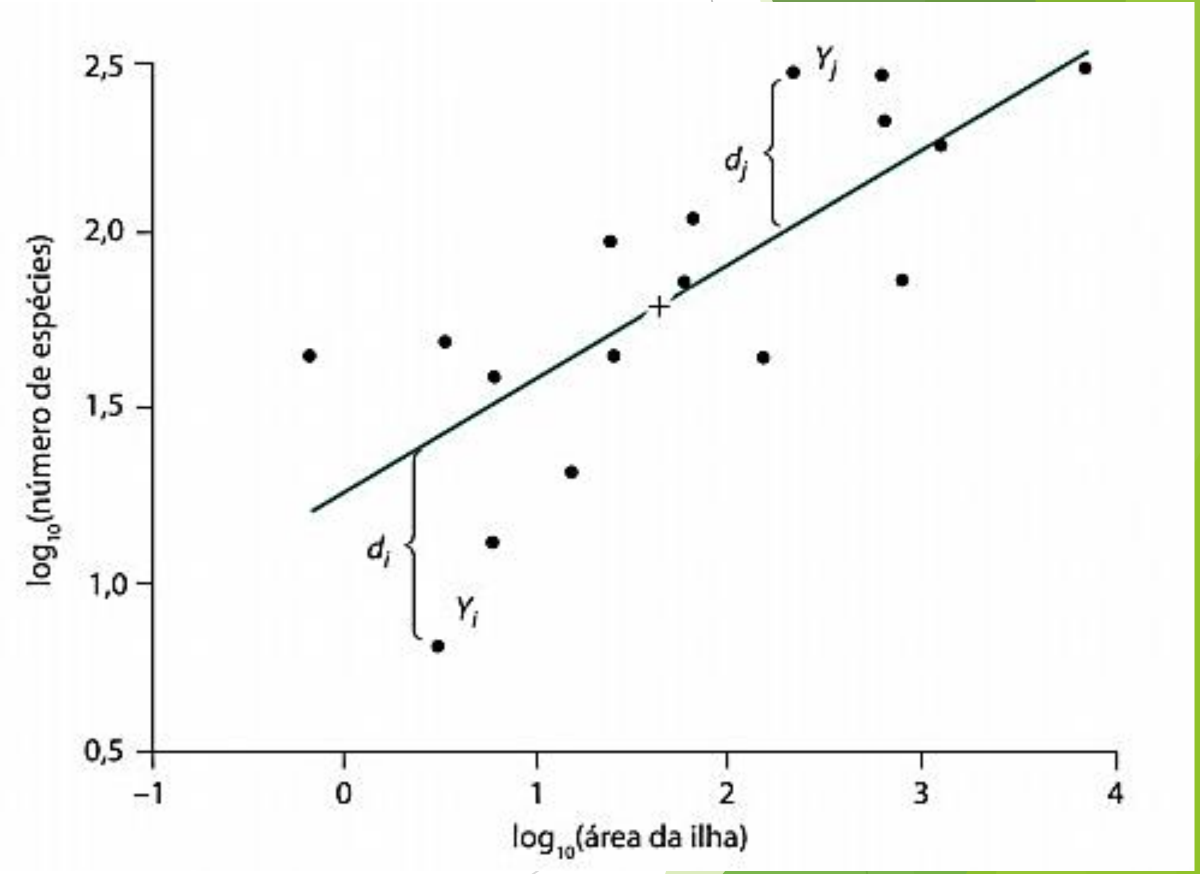


A melhor reta é aquela que minimiza o desvio padrão médio
(ou seja, a que minimiza os resíduos)
Quanto maior o valor do erro ➔ pior o x prediz o y

- **Resumindo**: a linha da regressão de "melhor ajuste" ➔ minimiza a soma dos quadrados dos resíduos

Assim, garantimos que a linha da regressão resulte na menor diferença média entre cada valor de:

y observado do valor de **y predito** pelo **modelo da regressão.**



O "famoso" R^2

- R^2 ➔ coeficiente de determinação: informa qual a proporção da variação de y pode ser atribuída a x

- O R^2 varia de 0 até 1 (quanto maior valor, menor é o erro e mais os dados sobrepõem a reta ajustada)

- Estatística F: F = efeito erro

Quanto maior o valor de F, maior vai ser o valor do R^2

Interpretação

```
Call:
lm(formula = altura ~ idade)

Residuals:
    Min      1Q  Median      3Q     Max 
-26.884 -11.907  -1.964   9.816  26.540 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 105.8698     6.3568  16.655 5.88e-14 ***
idade         1.7153     0.2443   7.021 4.80e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.66 on 22 degrees of freedom
Multiple R-squared:  0.6914,	Adjusted R-squared:  0.6774 
F-statistic:  49.3 on 1 and 22 DF,  p-value: 4.798e-07
```

Usar esses valores para montar a equação

Intercepto (a = 105,86)
Inclinação da reta (b = 1,71)

Y = 105,870 + 1,71*x

R^2adj = 0,67
F = 49,3
Gl = 22
p=<0,05

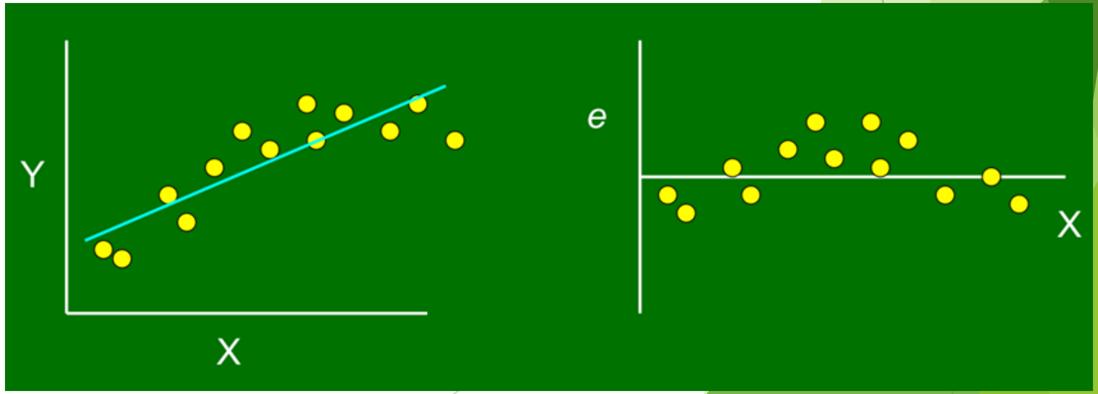
REGRESSÃO LINEAR SIMPLES

PRESSUPOSTOS:

Os resíduos representam os erros da previsão do modelo: a diferença entre os valores observados e os valores ajustados.

Independência: As observações em cada amostra devem ser independentes.

Linearidade: entre a variável dependente e a variável independente.



REGRESSÃO LINEAR SIMPLES

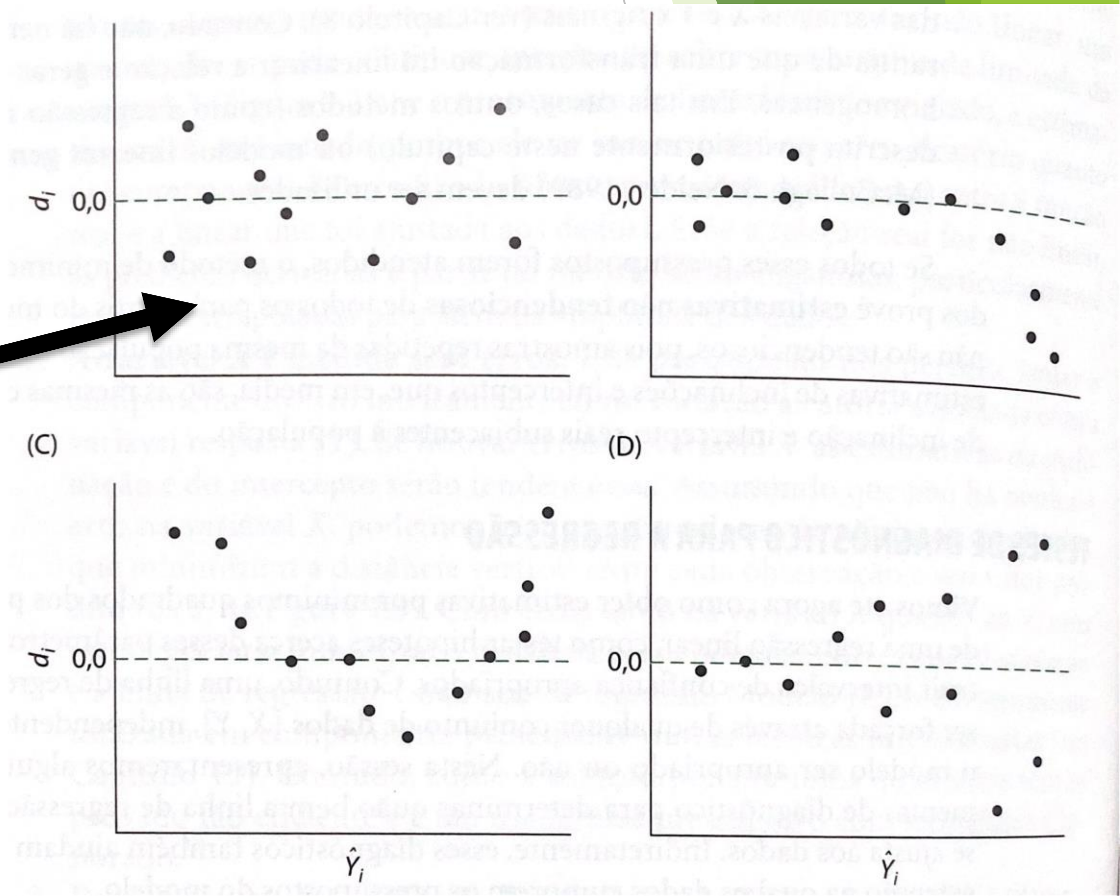
PRESSUPOSTOS:

Normalidade: Os dados devem seguir uma distribuição normal.

shapiro.test (rstudent (modelo))

```
        Shapiro-Wilk normality test

data:  rstudent(modelo)
W = 0.96438, p-value = 0.5325
```

Tem que ser assim para uma distribuição normal e relação linear entre as variáveis

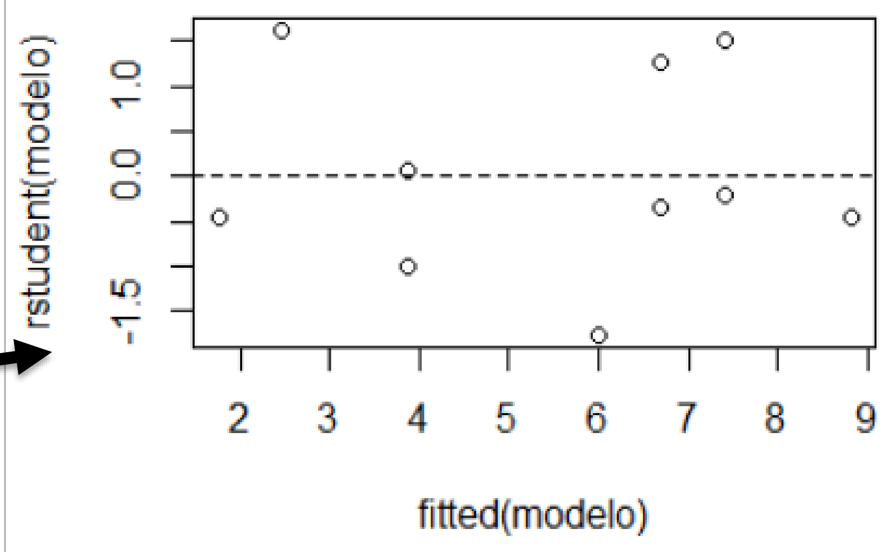


REGRESSÃO LINEAR SIMPLES

PRESSUPOSTOS:

Homocedasticidade: A variância dos erros e, condicionada aos valores das variáveis explanatórias, será constante.

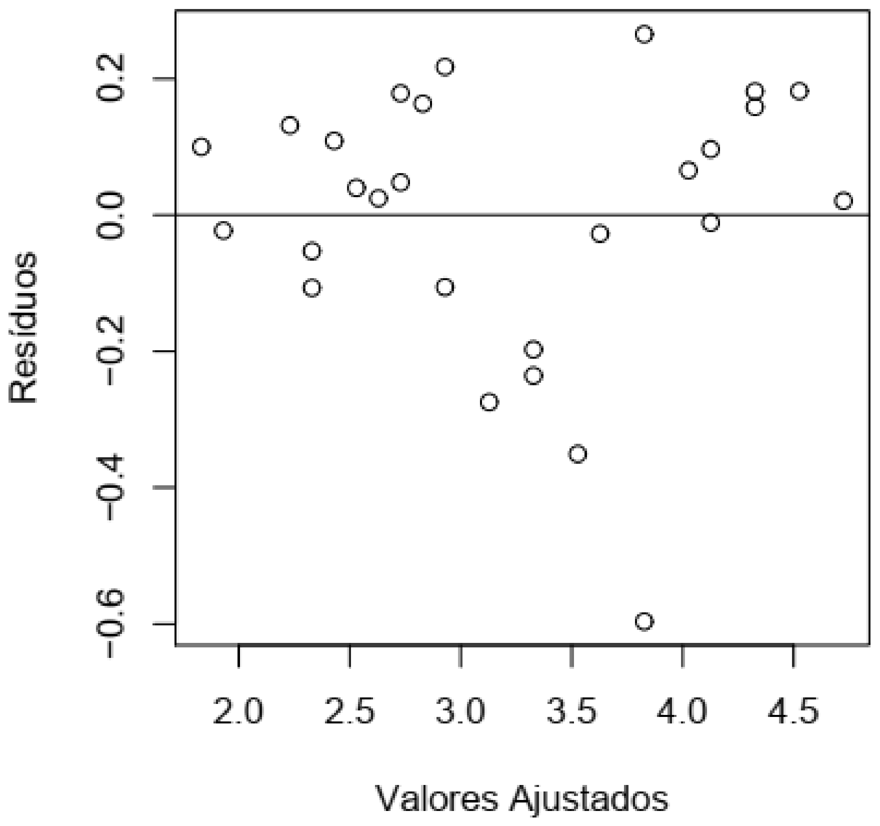
Os pontos devem ser distribuídos igualmente em torno do 0



REGRESSÃO LINEAR SIMPLES

PRESSUPOSTOS:

Homocedasticidade



Variância heterogênea

Exemplo de regressão linear simples

Referências