

Ciência de Dados com R

Aula 5 | Ajuste de Modelos

Izabel Nolau

nolau@dme.ufrj.br

Exemplo 1: Inflação e Emprego

Ano	x	y
1947	83.0	60323
1948	88.5	61122
1949	88.2	60171
1950	89.5	61187
1951	96.2	63221
1952	98.1	63639
1953	99.0	64989
1954	100.0	63761

Ano	x	y
1955	101.2	66019
1956	104.6	67857
1957	108.4	68169
1958	110.8	66513
1959	112.6	68655
1960	114.2	69564
1961	115.7	69331
1962	116.9	70561

Exemplo 1: Inflação e Emprego

Parece haver associação entre inflação (x) e emprego (y)?

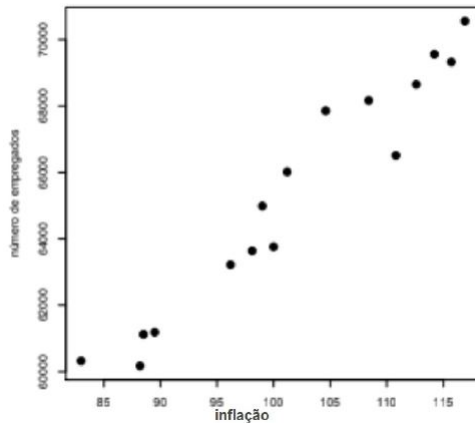
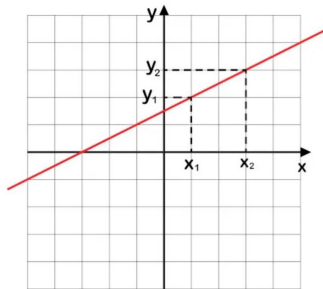


Figure 1: Diagrama de dispersão do exemplo 1.

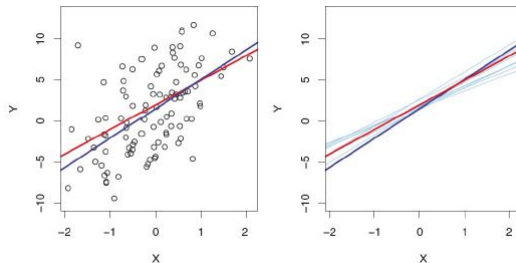
Equação da Reta



- A equação da reta é: $y = \beta_0 + \beta_1 x$.
- β_0 é o intercepto. Quando $x = 0$, $y = \beta_0$.
- β_1 é o coeficiente angular. Quando aumentamos uma unidade de x , aumentamos β_1 unidades em y .

Qual reta escolher?

Inúmeras retas passam pela nuvem de pontos.



Qual delas usar para explicar y a partir de x ?

Quais são os valores de β_0 e β_1 mais adequados à descrição da relação entre y e x ?

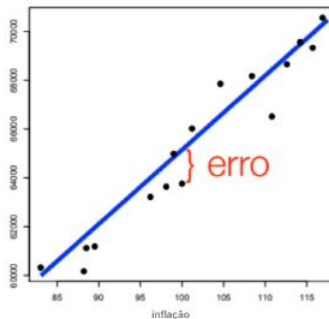
Devemos estimar β_0 e β_1 , a partir dos dados observados e aprender com eles sobre a associação entre x e y .

Modelo de Regressão Linear Simples

A relação entre x e y não é exata!

Consideramos um modelo estocástico para cada unidade i :

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad \text{para } i = 1, \dots, n.$$



Teste de Hipóteses sobre os Coeficientes do Modelo

Após obter estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$, devemos testar a significância estatística dos componentes do modelo.

Perguntas importantes:

- O intercepto é significativo ou um modelo mais simples com $\beta_0 = 0$ seria mais adequado?
- A variável x é realmente importante no modelo ou um modelo mais simples com $\beta_1 = 0$ seria mais adequado?

Se o p-valor do teste é menor que 0.05, temos que o coeficiente é significativo para a regressão!

Nesse caso, o modelo estocástico para cada unidade i é:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i,$$

para $i = 1, \dots, n$.

O problema de regressão linear simples é um caso particular, quando $p = 1$!

Bons estudos! 😊