

MODELO DE REGRESSÃO LINEAR SIMPLES

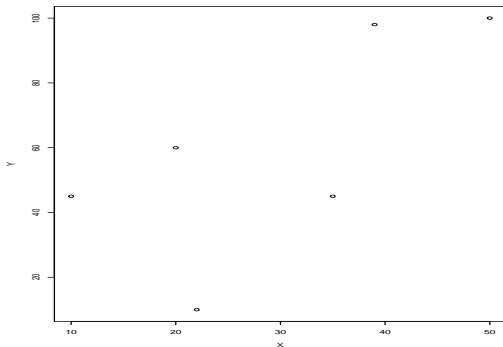
O modelo de regressão linear simples considera apenas uma variável explicativa e a função de regressão é linear. O modelo é definido por:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad (1)$$

em que:

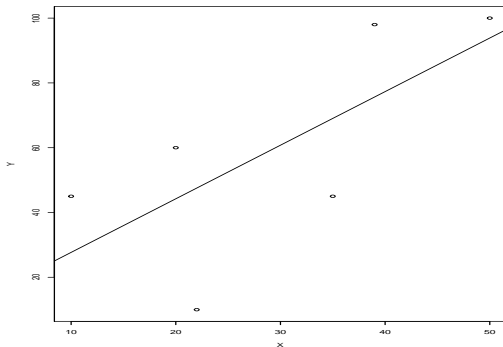
- Y_i é o valor da variável resposta para a i -ésima observação,
- β_0 é o intercepto e β_1 é o coeficiente angular, ambos são parâmetros desconhecidos,
- X_i é uma constante conhecida, o valor da variável explicativa para a i -ésima observação. X_i é uma variável fixa (ou sem erro ou determinística),
- ε_i é o erro aleatório.

DIAGRAMA DE DISPERSÃO



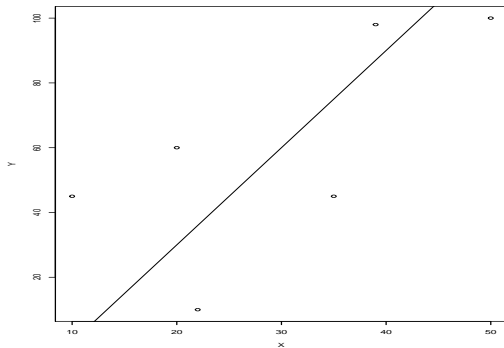
→ Como se pode traçar uma reta passando por esses pontos?

DIAGRAMA DE DISPERSÃO



→ Como determinar qual reta se "ajusta melhor"?

DIAGRAMA DE DISPERSÃO



→ Como determinar qual reta se "ajusta melhor"?

O MÉTODO DO MÍNIMOS QUADRADOS

- Os parâmetros β_0 e β_1 são desconhecidos e precisam ser estimados.
- Considere uma amostra aleatória de tamanho n de pares de dados, $(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$.
- O método de mínimos quadrados consiste no desvio de Y_i do seu respectivo valor esperado:

$$Y_i - (\beta_0 + \beta_1 X_i), \quad (2)$$

em particular, esse método considera a soma de quadrados do erro (ou desvio):

$$SQ(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_i))^2. \quad (3)$$

O MÉTODO DO MÍNIMOS QUADRADOS

Os valores de β_0 e β_1 que minimizam a equação (3) são os estimadores de mínimos quadrados, $\hat{\beta}_0$ e $\hat{\beta}_1$. São obtidos por:

$$\frac{\partial SQ(\beta_0, \beta_1)}{\partial \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1} = 0$$

e

$$\frac{\partial SQ(\beta_0, \beta_1)}{\partial \beta_1} \Big|_{\hat{\beta}_0, \hat{\beta}_1} = 0$$

O MÉTODO DO MÍNIMOS QUADRADOS

Ao resolver o sistema de equações normais, obtém-se:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

e

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}}$$

em que, $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$ e $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$.

O MÉTODO DE MÁXIMA VEROSSIMILHANÇA

- **Ideia do método:** baseado nos resultados obtidos pela amostra, qual é a distribuição, entre todas aquelas definidas pelos possíveis valores de seus parâmetros, com maior possibilidade de ter gerado tal amostra?

Ou seja, se a distribuição da variável em estudo é Normal, para cada combinação diferente de μ e σ , tem-se diferentes distribuições Normais. O estimador de máxima verossimilhança escolhe aquele par de μ e σ que melhor explique a amostra observada.

O MÉTODO DE MÁXIMA VEROSSIMILHANÇA

- Ao especificar a distribuição de probabilidade dos erros, podemos estimar os parâmetros do modelo de regressão linear simples usando o método de Máxima Verossimilhança.
- O modelo de regressão linear simples:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

em que: ε_i são independentes e $N(0, \sigma^2)$.

O MÉTODO DE MÁXIMA VEROSSIMILHANÇA

- Considere uma amostra aleatória de tamanho n de pares de dados, $(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$. A função de verossimilhança é dada por:

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{\varepsilon_i^2}{2\sigma^2} \right],$$

em que $\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_i)$

ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2} \right],$$

$$L(\beta_0, \beta_1, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \right], \quad (4)$$

ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA

- Os estimadores de máxima verossimilhança da equação (4) são dados por:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

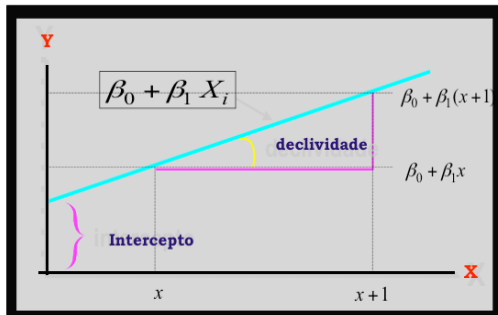
e

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n}$$

PROPRIEDADES DOS ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA

- Desde que os estimadores de máxima verossimilhança são iguais aos estimadores de mínimo quadrados, eles têm as propriedades dos estimadores de mínimo quadrados:
 - não viesados, e
 - possuem variância mínima entre todos os estimadores lineares não viesados.
- Os estimadores de máxima verossimilhança de β_0 e β_1 para o modelo de regressão com erros normais, também têm outras propriedades:
 - são consistentes, isto é, $\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \geq \kappa) = 0$ para qualquer κ ,
 - são suficientes, isto é, a distribuição de probabilidade conjunta condicional das observações da amostra dado $\hat{\theta}$ não depende de θ ,
 - São estimadores não tendenciosos de variâncias mínima, isto é, eles têm variância mínima na classe dos estimadores não tendenciosos (lineares ou não).

INTERPRETAÇÃO DOS PARÂMETRO DO MODELO



INTERPRETAÇÃO DOS PARÂMETRO DO MODELO

- Coeficiente angular ou declividade: β_1

$$\text{Para } X = x \rightarrow E(Y_i) = \beta_0 + \beta_1 x$$

$$\text{Para } X = x + 1 \rightarrow E(Y_i) = \beta_0 + \beta_1(x + 1)$$

\Rightarrow Representa o quanto varia o valor médio de Y para cada aumento de uma unidade de X .

- Intercepto: β_0

\Rightarrow Valor médio de Y quando $X = 0$, representa o ponto onde a reta corta o eixo das ordenadas.