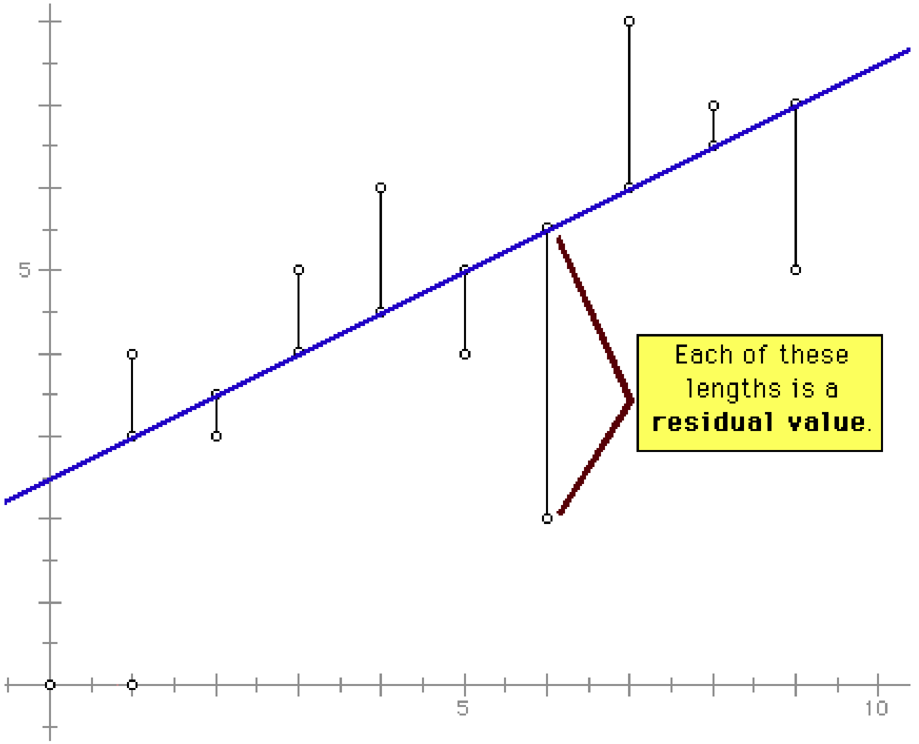
XMAC02
Métodos Matemáticos para Análise de Dados

Aula 21 – Equações Lineares

Regressão Linear

- Sempre que traçamos uma linha de regressão linear, haverá um valor residual em cada x
  - Trata-se da diferença vertical entre o valor real de y e o valor $\hat{y}$ (previsão)
- A equação da linha é dada por : $\hat{y} = mx + b$
  - m é a inclinação da linha
  - b é o ponto de interceptação no eixo y



Regressão Linear

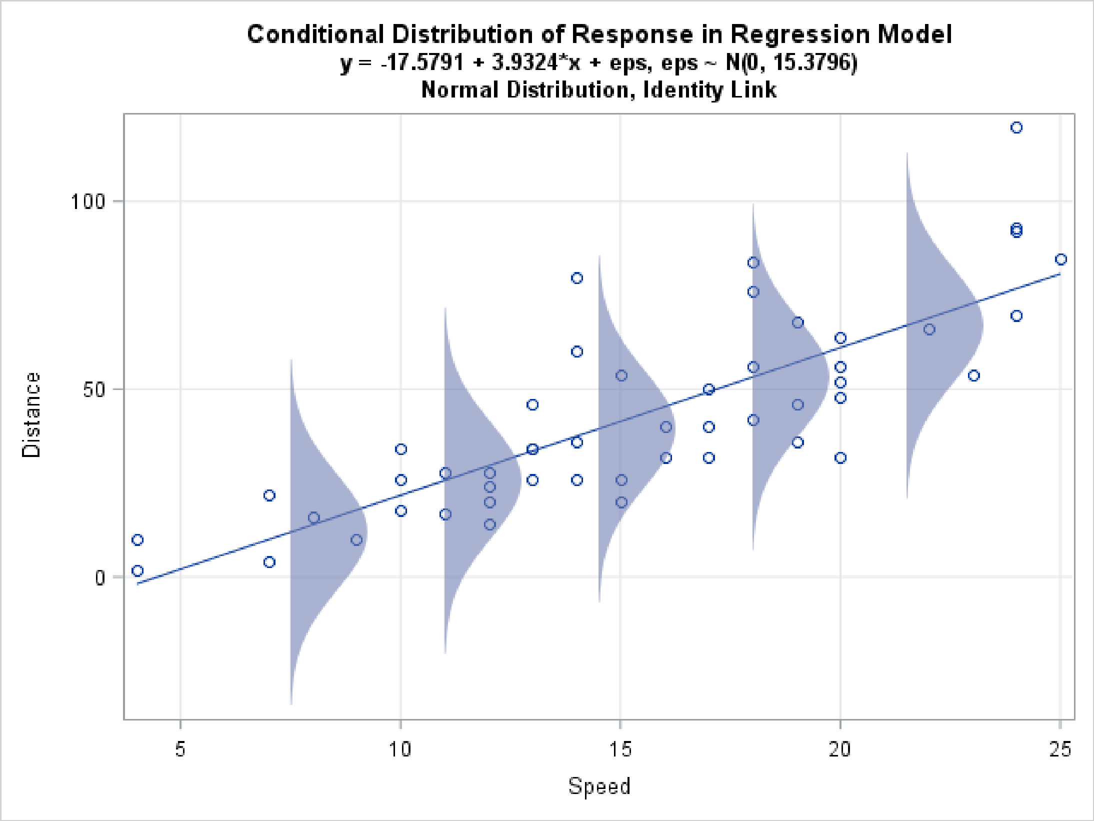
- Ao invés de $\hat{y} = mx + b$, podemos escrever a equação da linha da seguinte forma:

$$y = \beta_0 + \beta_1 x + e$$

  - β_0 é o ponto de interceptação no eixo y
  - β_1 é a inclinação da reta
  - e (erro) representa um desvio aleatório de um valor observado (real) de y e a linha de regressão da população $y = \beta_0 + \beta_1 x$
- Suposições a respeito de e :
  - é normalmente distribuído
  - seu desvio padrão σ não depende de x

Regressão Linear

- Para cada valor de x, o valor de e é normalmente distribuído
- A média de e é 0
- e tem o mesmo desvio padrão ao longo de toda reta
- σ = desvio padrão de y



Regressão Linear Múltipla

- A variável y pode ser linearmente dependente de múltiplas variáveis independentes $x_1, x_2, \ldots x_k$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + e$$

- β_is são chamados coeficientes de regressão populacional
- Na fórmula acima, cada variável independente x_i representa uma lista de n valores aleatórios

Regressão Linear Múltipla

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + e$$

- Para obter os valores dos coeficientes de regressão de uma população é necessário resolver um sistema normal de equações lineares
- Exemplo: considerando $k = 2$ e n = número de triplas x_1, x_2, y temos:

$$\begin{aligned}
n\beta_0 \quad &+ \beta_1 \sum x_1 &+ \beta_2 \sum x_2 &= \sum y \\
\beta_0 \sum x_1 &+ \beta_1 \sum x_1^2 &+ \beta_2 \sum x_1 x_2 &= \sum x_1 y \\
\beta_0 \sum x_2 &+ \beta_1 \sum x_1 x_2 &+ \beta_2 \sum x_2^2 &= \sum x_2 y
\end{aligned}$$

Regressão Linear Múltipla

Exemplo: Preços de casas

Número de quartos x_1	Número de banheiros x_2	\$ Preço y
3	2	48,800
2	1	44,300
4	3	53,800
2	1	44,200
3	2	49,700
2	2	44,900
5	3	58,400
4	2	52,900

$n = 8$

$\sum x_1 = 25$

$\sum x_2 = 16$

$\sum y = 397{,}000$

$\sum x_1 y = 1{,}281{,}100$

$\sum x_2 y = 817{,}700$

$\sum x_1^2 = 87$

$\sum x_2^2 = 55$

Regressão Linear Múltipla

Exemplo: Preços de casas

$n = 8$

$\sum x_1 = 25$

$\sum x_2 = 16$

$\sum y = 397{,}000$

$\sum x_1 y = 1{,}281{,}100$

$\sum x_2 y = 817{,}700$

$\sum x_1^2 = 87$

$\sum x_2^2 = 55$

$$
\begin{aligned}
n\beta_0 &+ \beta_1 \sum x_1 + \beta_2 \sum x_2 = \sum y \\
\beta_0 \sum x_1 &+ \beta_1 \sum x_1^2 + \beta_2 \sum x_1 x_2 = \sum x_1 y \\
\beta_0 \sum x_2 &+ \beta_1 \sum x_1 x_2 + \beta_2 \sum x_2^2 = \sum x_2 y
\end{aligned}
$$

$$
\begin{aligned}
8\beta_0 + 25\beta_1 + 16\beta_2 &= 397{,}000 \\
25\beta_0 + 87\beta_1 + 55\beta_2 &= 1{,}281{,}100 \\
16\beta_0 + 55\beta_1 + 36\beta_2 &= 817{,}700
\end{aligned}
$$

Solução:

$$
\begin{aligned}
\beta_0 &= 35{,}191.67 \\
\beta_1 &= 4{,}133.33 \\
\beta_2 &= 758.33
\end{aligned}
$$

Regressão Linear Múltipla
Exemplo: Preços de casas

$$\boxed{\begin{aligned} \beta_0 &= 35{,}191.67 \\ \beta_1 &= 4{,}133.33 \\ \beta_2 &= 758.33 \end{aligned}}$$

$$\begin{aligned} \hat{y} &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\ &= 35{,}191.67 + 4{,}133.33 x_1 + 758.33 x_2 \end{aligned}$$

$\hat{y}$ pois o erro e foi desconsiderado

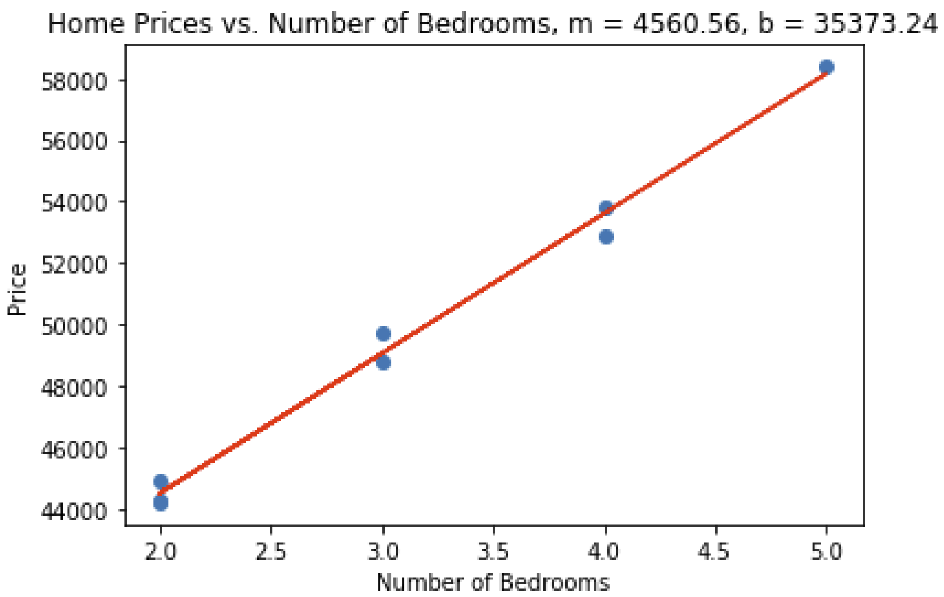
- Calcular valores estimados de casas:
  - $x_1 = 2$ quartos e $x_2 = 1$ banheiros

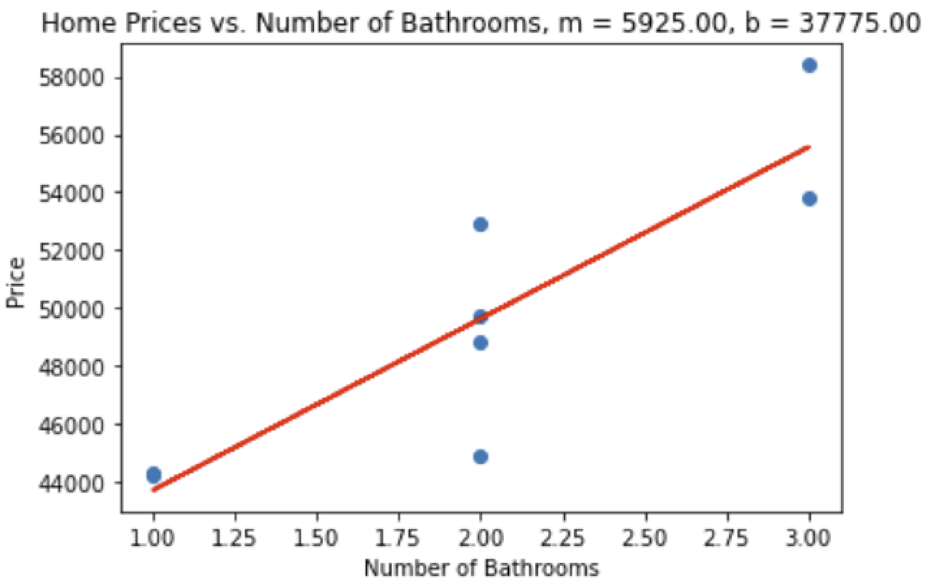
    $$35{,}191.67 + 4{,}133.33(2) + 758.33(1) = \$44{,}216.67$$

  - $x_1 = 5$ quartos e $x_2 = 3$ banheiros

    $$35{,}191.67 + 4{,}133.33(5) + 758.33(3) = \$58{,}133.33$$

Regressão Linear Múltipla
Exemplo: Preços de casas





Créditos

- Este conteúdo é uma tradução do original em inglês produzido pelo Prof. Ronald Mak (SJSU).