

Estatística para Ciência de Dados 2024

Exemplos

- Análise Exploratória de Dados

- Tipos de dados

quantitativos
qualitativos

contínuos
discretos

Normal
Exponencial

Bernoulli
binomial
Poisson

nominais - Sexo

- Inadimplência

ordinais - Grau de

escolaridade

- Probabilidades destaque: TLC $\bar{X} \stackrel{n \rightarrow \infty}{\sim} N$

- Inferência Estatística $\left\{ \begin{array}{l} \text{Clássica ou Frequentista} \\ \text{Bayesiana} \end{array} \right.$

- Estimação pontual e intervalar

Exemplo
• $\hat{p} = 0,24$

• $IC(p) = [0,20; 0,28]$
95%

- Testes de Hipóteses $\begin{cases} H_0: p = 0,3 \text{ contra} \\ H_1: p < 0,30 \end{cases} \dots \text{Rejeito } H_0$

- Modelagem estatística • Principais vantagens:

- Interpretabilidade
- Inferência sobre
os coeficientes

- Modelos lineares

- Modelos lineares generalizados MLGs ou GLM

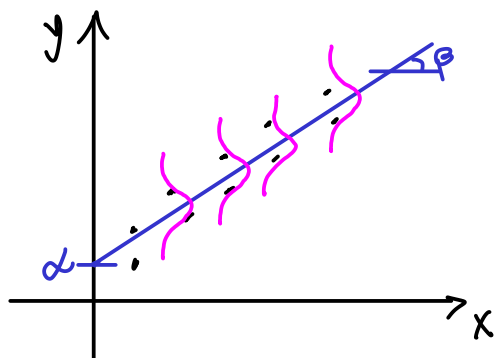
- Regressão com resposta Bernoulli (p)

- Regressão com resposta Poisson (λ)

- Análise de diagnóstico

- O modelo é adequado?

- As suposições fazem sentido?



y_i : Resposta, x_i : v. preditora
 α, β : parâmetros

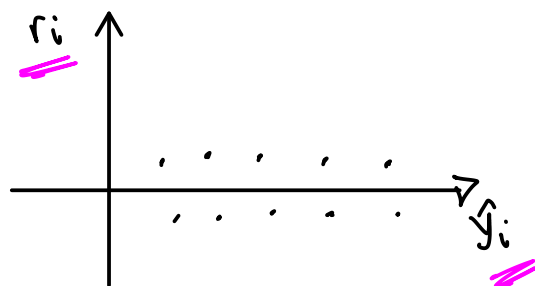
$$y_i = \alpha + \beta x_i + \epsilon_i, i = 1, \dots, n$$

$$\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

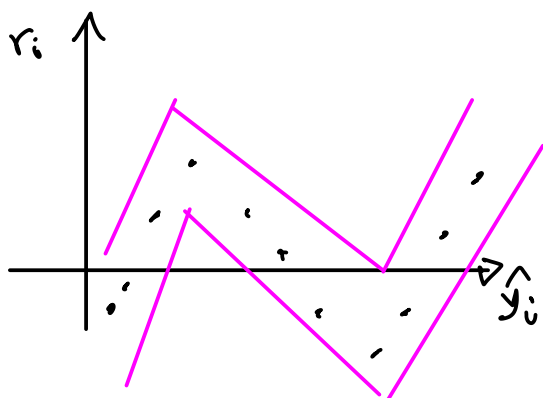
Após o ajuste, $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i, i = 1, \dots, n$

Resíduos: $r_i = \hat{y}_i - y_i$
 ("erros observados")

$\hat{\alpha}$
 $\hat{\beta}$ } estimadores
 dos
 parâmetros



E se



Aprendizado dinâmico

Possível modelo: SARIMAX

$$y_i = \alpha + \beta x_i + \epsilon_i$$

ϵ_i SARIMA

