L.EIC
#2.2

Métodos Estatísticos

Resumos

1. Estatística Descritiva
2. Probabilidades
3. Variáveis Aleatórias
4. Intervalos de confiança
5. Testes de hipótese paramétricos
6. Proporções
7. Outros testes de hipótese

1. Estatística Descritiva

Classificação de dados

Qualitativo
- Nominal / Categórico (sem ordem)
- Ordinal (com noção de ordem)

Quantitativo
- D Discreto (contar)
- C Contínuo (medir)

Organização de dados

- Freq. relativa

$$f_r = \frac{f_a}{n}$$

f_a: Freq. absoluta (valor em si)

- Freq. absoluta acumulada

(somar f_a anteriores)

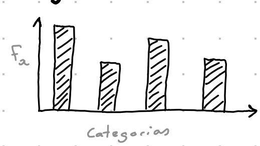
- Densidade

$$= \frac{f_r}{a}$$

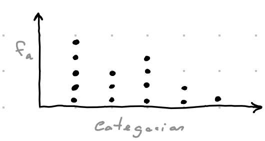
a: amplitude de uma classe

Representações gráficas

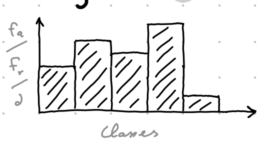
- Diagrama de barras D



- Diagrama de pontos D



- Histograma C

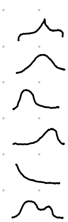


→ Fórmula de Sturges:

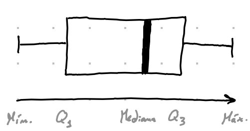
$$k \simeq 1 + 3{,}322 \log_{10} n$$

→ Formas:

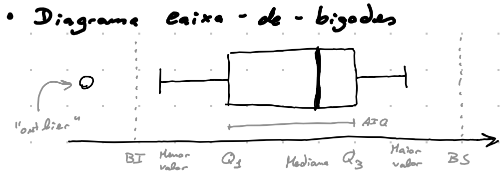
↳ Simétrica
↳ Simétrica em forma de sino
↳ Enviesada à direita
↳ Enviesada à esquerda
↳ Exponencial
↳ Bimodal



- Diagrama de extremos e quartis



- Diagrama caixa-de-bigodes

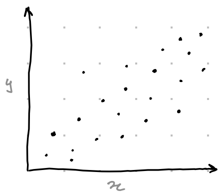


→ $AIQ = Q_3 - Q_1$

→ $BI = Q_1 - 1{,}5\ AIQ$

→ $BS = Q_3 + 1{,}5\ AIQ$

- Diagrama de dispersão



↳ Coeficiente de correlação (ρ)

$$r = \frac{\rho_{x,y}}{s_x \, s_y}$$

desvios padrão

↳ Covariância amostral

$$\rho_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{y}}{n-1}$$

Parâmetros

- **Média** ($\mu_X = E(X)$)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- **Mediana**

→ n ímpar | Observação central

→ n par | Média das duas

- **Quantis / Quartis**

$$Q_p = \begin{cases} \dfrac{x_{np} + x_{np+1}}{2} & \text{se } np \text{ for inteiro} \\ x_{k+1} & \text{se } np \text{ não for inteiro} \end{cases}$$

parte inteira de np

- **Moda**

Valor(es) com maior frequência

- **Coeficiente de variação**

$$CV = \frac{s}{\bar{x}}$$

↓ % = ↑ estabilidade

- **Desvio padrão** (σ_X)

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

- **Variância** ($V(X)$)

$$s^2 = \frac{\sum_{i=1}^{n} (x_i^2) - n \bar{x}^2}{n-1}$$

2. Probabilidades

Combinações

$$^{m}C_{p} = \frac{m!}{p(m-p)!}$$

Acontecimentos independentes

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \mid B) = P(A)$$

Probabilidade condicional

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Propriedades

- $\overline{A \cap B} = \overline{A} \cup \overline{B}$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $P(\overline{A}) = 1 - P(A)$
- $P(A \cap B) = P(A) + P(B) - P(A \cup B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A) = P(A \cap B) + P(A \cap \overline{B})$

3. Variáveis Aleatórias

Discreta

$f_X(x) = P(X = x)$

$F_X(x) = P(X \leq x)$

$\mu_X(x) = \sum_i x_i \cdot P(X = x_i)$

$\sigma^2_X(x) = \sum_i x_i^2 \cdot P(X = x_i) - \mu_X^2$

Distribuição

Binomial

Um conjunto de n experiências de Bernoulli idênticas, independentes e de resultado binário

$X \sim Bi(n, p)$

$f(x) = C^n_x \, p^x (1-p)^{n-x}$

$E(x) = np$

$V(x) = np(1-p)$

Contínua

$P(a < X < b) = \int_a^b f(x)\, dx$

* $P(X = x) = 0$

$= \int x \, f(x) \, dx$

$= \int (x - \mu_x)^2 f(x) \, dx$

Normal / Gauss

$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$X \sim N(\mu, \sigma^2)$

Normal Padrão

$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

- $P(Z > x) = 1 - P(Z \leq x)$
- $P(Z > -x) = P(Z \leq x)$
- $P(x < Z < y) = P(Z < y) - P(Z < x)$

- **Condições de aproximação $Bi(n, p) \rightarrow N(np, np(1-p))$**

  $\rightarrow n > 25$ $\rightarrow np > 5$ $\rightarrow n(1-p) > 5$

- **Regra 68 / 95 / 99** ($\pm 1\sigma, \pm 2\sigma, \pm 3\sigma$) para avaliar a normalidade

4. Intervalos de confiança

Erro padrão da média

$$se = \frac{s}{\sqrt{n}}$$

- Média ponderada de variâncias

$$s_p^2 = \frac{(n_x - 1)\, s_x^2 + (n_y - 1)\, s_y^2}{n_x + n_y - 2}$$

- Erro padrão ponderado

$$se_p = \sqrt{s^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}$$

IC para médias

$$IC_{(1-\alpha)} =$$

(*)

α	0,1	0,05	0,01
$z_{\frac{\alpha}{2}}$	1,65	1,96	2,58

σ conhecida

$$= \bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

(*) $P(|\bar{x} - \mu| < a) = 1 - \alpha$

σ desconhecida

$$= \bar{x} \pm t_{(\frac{\alpha}{2},\, n-1)} \cdot se$$

População não normal

$$= \bar{x} \pm z_{\frac{\alpha}{2}} \cdot se$$

(*) $n \geq 30$

IC para diferença de médias

$$E(\bar{X} - \bar{Y}) = \mu_X - \mu_Y$$

$$V(\bar{X} - \bar{Y}) = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}$$

$$se_{(\bar{X} - \bar{Y})} = \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

$= se_p$ (se $\sigma_x = \sigma_y$)

σ conhecida

$$= \bar{x} - \bar{y} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

σ desconhecida

$$= \bar{x} - \bar{y} \pm t_{(\frac{\alpha}{2},\, n_x + n_y - 2)} \cdot se_{(\bar{X} - \bar{Y})}$$

População não normal

$$= \bar{x} - \bar{y} \pm z_{\frac{\alpha}{2}} \cdot se_{(\bar{X} - \bar{Y})}$$

(*) Em **amostras emparelhadas**, cria-se uma nova população D, em que $D_i = X_i - Y_i$.

ME

5. Testes de hipótese paramétricos

1. Formular hipótese com a intenção de rejeitar H_0

Ex//

- H_0 | Hipótese nula — $H_0 : \mu = 75$
- H_1 | Hipótese alternativa — $H_1 : \mu \neq 75$

2. Calcular estatística de teste t_s

- **Comparar μ com um valor específico**

$t_s = \frac{\bar{x} - VE}{se}$, sendo VE o valor específico $\quad g.l. = n - 1$

- **Comparar duas amostras independentes**

$t_s = \frac{\bar{x}_1 - \bar{x}_2}{se} \quad g.l. = n_x + n_y - 2$

3. Encontrar valor-p das tabelas t-student

$t_{s\left(\frac{valor-p}{2},\ g.l.\right)}$

4. Tomar a decisão se H_0 deve ser rejeitada ou não

- **valor-p < α** | Rejeitar hipótese H_0 / Tem evidências para aceitar H_1
- **valor-p > α** | Não é possível rejeitar hipótese H_0 / Não há evidência de H_1

Tipos de erro:

- **I** | Rejeitar H_0 quando é verdadeira (α é a probabilidade deste erro)
- **II** | Não rejeitar H_0 quando é falso

ME

6. Proporções

$\hat{p} = \frac{X}{n} \quad (X \sim Bi(n, p)) \qquad E(\hat{p}) = p \qquad V(\hat{p}) = \frac{p(1-p)}{n} \qquad se_{\hat{p}} \equiv s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$\tilde{p} = \frac{x+2}{n+4}$

Método de Wald

$$I.C. = \hat{p} \pm z_{\frac{\alpha}{2}} \cdot se_{\hat{p}}$$

- Usado em amostras grandes e quando o p não é próximo de 0 ou 1

Método de Agresti - Coull

$$I.C. = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

- Método mais usado e com resultado mais alargado.

7. Outros testes de hipótese

1. Calcular a Frequência esperada

- **Ajustamento** (g.l. = nº categorias − 1)
  (Probabilidade dada pelo rácio)
- **Independência / Homogeneidade** (g.l. = (nº linhas − 1) × (nº colunas − 1))
  $P(A \cap B) = P(A) \times P(B)$ $\quad e_i = \dfrac{\text{total linha} \times \text{total coluna}}{\text{total global}}$

2. Calcular $\chi^2_1 = \sum_i \frac{(o_i - e_i)^2}{e_i}$

3. Comparar com $\chi^2_{(gl, \alpha)}$

→ $\chi^2_1 > \chi^2_{(gl, \alpha)}$ => valor-p < α => Rejeita-se H_0
→ $\chi^2_1 < \chi^2_{(gl, \alpha)}$ => valor-p > α => Sem evidências de H_1

LC