# Resumo

21 de abril de 2023 00:01

## Variável Estatística

→ Qualitativa:

→ Nominal:

→ Nomes sem uma ordem de categorias.

→ Ordinal:

→ Nomes com uma ordem de categorias.

→ Quantitativa:

→ Discreta (Domínio = conj. (in)finito de inteiros):

→ Associada a contagens.

→ Contínua (Domínio = conj. de nºs reais):

→ Associada a medições.

## Tabelas de frequências

| $i$ | $x_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ |
|---|---|---|---|---|---|
| Nº Linha | Valor ou Classe | Freq. Absoluta (contagem) | Freq. Absolutas Acumuladas | Freq. Relativas ($\frac{n_i}{n}$) (0-1) | Freq. Relativas Acumuladas |

→ Sem classes:

→ V. Qualitativas;

→ V. Quantitativas Discretas (Pouca Distinção).

→ Com classes:

→ V. Quantitativas contínuas;

→ V. Quantitativas Discretas (Muita Distinção).

→ N.º de classes: $k = 1 + \log_2 n = 1 + \frac{\ln n}{\ln 2}$

→ Amplitude das classes: $h = \frac{\max(x_i) - \min(x_i)}{k}$

## Gráficos

→ Barras:

→ Dados qualitativos;

→ Dados Quantitativos Discretos.

→ Circulares:

→ Dados qualitativos;

→ Dados Quantitativos Discretos (Menos Usado).

→ Histogramas: (Mais para contínuos)

→ Dados Quantitativos em classes.

# Medidas de Localização

→ Central:

→ Moda (maior $n_i$ | mais repetido)

→ Média (R → mean() | $(\sum_{i=1}^{n} x_i)/n = \bar{x}$)

→ Mediana (R → median()):

→ $nf = n \times 0.5$:

- Se $nf$ é inteiro: $\tilde{x} = \dfrac{x_{(nf)} + x_{(nf+1)}}{2}$
- Senão: $\tilde{x} = x_{([nf]+1)}$

→ Não central:

→ Quantis:

| | | |
|---|---|---|
| $Q_1$ → $nf = n \times 0.25$ | $Q_2$ → $\tilde{x}$ | $Q_3$ → $nf = n \times 0.75$ |

# Medidas de Dispersão

→ Dispersão Absoluta:

→ Amplitude Total: (R → max() − min())

→ Amplitude Interquartis: $AIQ = Q_3 - Q_1$ (R → IQR())

→ Variância: $s^2$ (R → var())

→ Desvio Padrão: $s$ (R → sd() ou sqrt(var()))

→ Dispersão Relativa:

→ Coeficiente de Variação: $CV = \dfrac{s}{\bar{x}} \times 100\%$

# V. Al. Discretas

$X$ - Nº de ...... (contagens)

$f(x) = P(X = x)$ -> função Probabilidade

↳ Nota:

| $x$ | Min | ... | Max |
|---|---|---|---|
| $f(x)$ | ... | ... | ... |

-> $f(x) \geq 0, \ \forall x$

-> $\sum f(x) = 1$

$F(x) = P(X \leq x)$ -> função Distribuição

$$F(x) = \begin{cases} 0, & x \leq Min \\ \ldots, & Min < x < Max \\ 1, & x \geq Max \end{cases}$$

$P(a < X \leq b) = F(b) - F(a)$

$P(\bar{a}) = 1 - P(a)$

**Intervalo de Resultados:** $0 \leq P(a) \leq 1$

$P(a \cup b) = P(a) + P(b) - P(a \cap b)$

$P(a \cap b) = P(a - b) = P(a) - P(a \cap b)$

$P(a|b) = \dfrac{P(a \cap b)}{P(b)}$

$\cup$ -> ou
$\cap$ -> e
$|$ -> Sabendo que
↳ se

$E[X] = \mu = \sum x \cdot f(x)$ -> Valor esperado

$E[X^2] = \sum x^2 \cdot f(x)$ $\quad E[a] = a$

$E[aX + b] = a \cdot E[X] + b$

$E[X + Y] = E[X] + E[Y]$

$V[X] = E[X^2] - E^2[X]$ $\quad V[a] = 0$

$\sigma^2 = V[X]$

$= E[(X - \mu)^2]$

$= \sum_x (x - \mu)^2 \times f(x)$

↳ Variância

$\sqrt{V[X]} = \sigma$

↓ Desvio Padrão

$V[aX + b] = a^2 \cdot V[X]$

# V. Al. Contínuas

$f(x)$ → função Densidade de Probabilidade

$F(x) = P(X \,|\leq, <|\, x)$ → função Distribuição

$P(a \,|<, \leq|\, X \,|\leq, <|\, b) = F(b) - F(a)$

Nota:

→ $f(x) = \emptyset$

→ $\int f(x)\, dx = 1$

$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx = \mu$ → Valor esperado

$E[X^2] = \int_a^b x^2 \cdot f(x)\, dx$

$E[a] = a$

$E[aX + b] = a \cdot E[X] + b$

$V[X] = E[X^2] - E^2[X] = \sigma^2$ → Variância

$V[a] = \emptyset$

$V[aX + b] = a^2 \cdot V[X]$

$\sigma = \sqrt{\sigma^2} = \sqrt{V[X]}$

→ Desvio Padrão

# Distribuições Teóricas Conhecidas

- ## V. Al. Discretas:

### → Uniforme Discreta:

→ Para situações em que todos os valores possíveis têm a mesma probabilidade de ocorrer.

→ Aplica-se apenas a conjuntos de valores discretos (individuais).

$X \sim U_{(n)}$

$n = b - a + 1 \Rightarrow$ Nº de elementos do domínio.

$$f(x) = P(X = x) = \frac{1}{n} = \begin{cases} \frac{1}{n}, & \forall_x \in D_x \\ \emptyset, & \text{c.c.} \end{cases}$$

- Se estiver definida num conjunto de inteiros consecutivos:

$$E[X] = \frac{a+b}{2}$$

$$V[X] = \frac{(b-a+1)^2 - 1}{12}$$

- Senão:

$$E[X] = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$V[X] = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - (E[X])^2$$

$$D_x = \{\underbrace{a}_{\text{1º val. do domínio}}, a+1, a+2, \dots, \underbrace{b}_{\text{último val. do domínio}}\}$$

## → Binomial:

→ Para situações em que há 2 resultados possíveis e interessa calcular a probabilidade do nº de (in)sucessos. $E[x] = n.p$

$X \sim B(n, p)$ → $p$ (Sucesso) $V[x] = n.p.(1-p)$

$X$ → Nº de Sucessos em $n$ provas.

$f(x) = P(X = x) \Rightarrow$ dbinom $(x, n, p)$

$F(x) = P(X \leq x) \Rightarrow$ pbinom $(x, n, p)$

$F(k) = prob \Leftrightarrow k = F^{-1}(prob) \Rightarrow$ qbinom $(prob, n, p)$

$D_X = \{0, 1, 2, \dots\}$ $\boxed{n = n_{X_1} + n_{X_2}}$

Aditividade da Binomial: $Y = X_1 + X_2 \sim B(n, p)$

$\boxed{p = \text{em todos}}$

## → Poisson:

→ Ocorrência de eventos raros num intervalo de tempo.

→ É calcular a probabilidade de ocorrência de eventos raros num intervalo de tempo.

$X \sim P(\lambda)$ $\lambda = E[x] = V[x]$

$X$ → Nº de ocorrências num intervalo de medida.

$f(x) = P(X = x) \Rightarrow$ dpois $(x, \lambda)$

$F(x) = P(X \leq x) \Rightarrow$ ppois $(x, \lambda)$

$F(x) = prob \Leftrightarrow x = F^{-1}(prob) \Rightarrow$ qpois $(prob, \lambda)$

$D_X = \{0, 1, 2, \dots\}$ $\boxed{\lambda = \lambda_{X_1} + \lambda_{X_2}}$

Aditividade da Poisson: $Y = X_1 + X_2 \sim P(\lambda)$

- V. Al. Contínuas:

→ Uniforme Contínua:

→ É igual à Uniforme Discreta, mas para intervalos contínuos.

$X \sim U_{(a,\, b)}$ (a → 1º val. do intervalo; b → último val. do intervalo)

$f(x) \to \underline{\text{dunif}}(x, a, b)$

$F(x) = P(X \le / < x) \to \underline{\text{punif}}(x, a, b)$

$F^{-1}(x) \to \underline{\text{qunif}}(\text{prob}, a, b)$

$$E[X] = \frac{a+b}{2} \qquad V[X] = \frac{(b-a)^2}{12}$$

→ Exponencial:

→ Tempo entre eventos.

→ É usada para calcular a probabilidade de um evento ocorrer num intervalo de tempo.

$X \sim Exp(\theta)$, $\theta > \phi$ — É Assimétrica à Direita!

X → Tempo/Distância entre ocorrências sucessivas.

$$\underbrace{P(X \geq a+b \mid X \geq a) = P(X \geq b)}_{\text{Falta de Memória}}$$

$f(x) \to \underline{\text{dexp}}\left(x, \frac{1}{\theta}\right)$ $\qquad E[X] = \theta$

$F(x) \to \underline{\text{pexp}}\left(x, \frac{1}{\theta}\right)$ $\qquad V[X] = \theta^2$

$F^{-1}(x) \to \underline{\text{qexp}}\left(\text{prob}, \frac{1}{\theta}\right)$ $\qquad D_X = [0, +\infty[$

→ Normal:

$$X \sim N(\overset{\text{Média}}{\mu}, \underset{\text{Desvio Padrão}}{\sigma}) \qquad D_X = \mathbb{R}$$

$f(x) \to$ <u>dnorm</u>$(x, \mu, \sigma)$

$F(x) \to$ <u>pnorm</u>$(x, \mu, \sigma)$

Inversa → <u>qnorm</u>(prob, $\mu$, $\sigma$)

$E[X] = \mu$

$V[X] = \sigma^2$

Aditividade da Normal:

$$\hookrightarrow S = X_1 + X_2 \sim N(\overset{\mu_{X_1} + \mu_{X_2}}{\mu}, \underset{\sigma_{X_1} + \sigma_{X_2}}{\sigma})$$

→ Normal Standard/Reduzida:

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

$$P(Z \underset{<}{\leq} z) = \Phi(z)$$

Nota:

→ $\Phi(-z) = 1 - \Phi(z)$

$f(z) \to$ <u>dnorm</u>$(z)$

$\underbrace{F(z)}_{\Phi(z)} \to$ <u>pnorm</u>$(z)$

$F^{-1}(z) \to$ <u>qnorm</u>(prob)

→ Qui-Quadrado:

$X \sim \chi^2_{(n)}$

→ T-Student:

$X \sim t_{(n)}$

→ F de Snedecor:

$X \sim F_{(m,n)}$

# Relações Entre Distribuições

- **Exponencial e Poisson**

Se

$X \sim P(\lambda)$

$X \rightarrow$ N.º de ocorrências num intervalo de tempo $t$

$\lambda \rightarrow \mu$: Médio de ocorrências num intervalo de tempo $t$

e

$Y \sim Exp(\theta)$

$Y \rightarrow$ Tempo de espera entre ocorrências sucessivas

$\theta \rightarrow$ Tempo de espera médio entre ocorrências sucessivas

então

$$\theta = \frac{t}{\lambda} \Leftrightarrow \lambda = \frac{t}{\theta}$$

# Testes

- **Paramétricos:**

Servem para testar:

$\rightarrow \mu \rightarrow$ Média

$\rightarrow \sigma^2 \rightarrow$ Variância

$\rightarrow p \rightarrow$ Proporção

$\rightarrow \mu_1 - \mu_2 \rightarrow$ Diferença de Médias

$\rightarrow \sigma_1^2 / \sigma_2^2 \rightarrow$ Quociente de Variâncias

$\rightarrow p_1 - p_2 \rightarrow$ Diferença de Proporções

## • Não Paramétricos:

- Testes Ajustamento:
  - Qui-quadrado
  - Kolmogorov-Smirnov
  - Lillie-fors
  - SW
- Diferenças de Medianas:
  - Wilcoxon
  - Mann-Whitney

## Passos Para os Testes de Hipóteses:

1º - Formular hipóteses;

2º - Definir a distribuição amostral e a estatística de teste;

3º - Tomar a Decisão:

Rejeita-se $H_0$ se:

- Pela Região Crítica:

  ET pertencer à RC.

- Pelo P-Value:

  $P\text{-Value} \leq \alpha$

4º - Fazer a conclusão

# Testes para Diferenças em Populações

- Testes para Diferenças de Médias:
  → Há normalidade / independência?

Vê-se pelos testes de ajustamento:
- Kolmogorov-Smirnov;
- Lilliefors;
- Shapiro-Wilk.

→ Sim (cap. 5):
Teste Paramétrico ($\mu_1 - \mu_2$)

→ Não: $n \geq 30$ → Não (cap 6.2): $n < 30$

Teste Não Paramétrico:
- Wilcoxon (vars. emparelhadas)
- Mann-Whitney (vars. Independentes)

## escolha do Tipo de Teste

| $H_0$ | $H_1$ | Tipo de Teste |
|---|---|---|
| $\sigma = \sigma_0$ | $\sigma != \sigma_0$ | Bilateral |
| | $\sigma > \sigma_0$ | Unilateral Direito |
| | $\sigma < \sigma_0$ | Unilateral esquerdo |
| $\sigma \geq \sigma_0$ | $\sigma < \sigma_0$ | Unilateral esquerdo |
| $\sigma \leq \sigma_0$ | $\sigma > \sigma_0$ | Unilateral Direito |