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).
  - Nº 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:
  - Dados Quantitativos Contínuos (em classes).

Medidas de Localização

→ Central:

- → Moda (maior n_i | mais repetido)
- → Média (R → mean() | $\left(\sum_{i=1}^{n} x_i\right)/n = \bar{x}$)
- → 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:

$Q_1 \to nf = n \times 0.25$	$Q_2 \to \tilde{x}$	$Q_3 \to nf = n \times 0.75$

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:

→ $f(x) \geq 0, \ \forall x$

→ $\sum f(x) = 1$

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

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

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

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

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

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

$P(a \cap \bar{b}) = P(a - b) = P(a) - P(a \cap b)$

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

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

$\cup$ → ou
$\cap$ → e
$|$ → Sabendo que → se

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

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

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

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

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

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

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

→ Variância

$\sqrt{V[X]} = \sigma$ → Desvio Padrão

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

V. Al. Contínuas

$X - \ldots\ldots$

$f(x) = P(X = x)$ → f. Densidade de Probabilidade

↳ $f(x) = F'(x)$

↳ Apenas é f.D.P. se:

→ $f(x) \geq 0, \forall x$

→ $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

Primitivas Importantes

$$P(x^k) = \frac{x^{k+1}}{k+1}, \quad k \neq -1$$

$$P(k) = kx$$

$$F(x) = P(X \leq x) = P(X < x) = \int_{-\infty}^{x} f(t)\, dt$$ → f. Distribuição

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

Valor esperado

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

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

$$E[a] = a$$

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

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

$$V[a] = \emptyset$$

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

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

↳ Desvio Padrão

Exemplo de uma Integral:

$$\int_2^4 \left(1 - \frac{x}{4}\right) dx = \int_2^4 1\, dx - \int_2^4 \frac{x}{4}\, dx$$

$$= \int_2^4 1\, dx - \frac{1}{4}\int_2^4 x\, dx$$

$P(1)$ $\qquad \frac{1}{4}P(x)$

$$= [x]_2^4 - \frac{1}{4}\left[\frac{x^2}{2}\right]_2^4$$

$$= (4-2) - \frac{1}{4}\left(\frac{4^2}{2} - \frac{2^2}{2}\right)$$

$$= 2 - \frac{1}{4}\left(\frac{16}{2} - \frac{4}{2}\right) = 2 - \frac{12}{8}$$

$$= 4/8$$

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).

$X \sim U_{(n)}$

$n = b - a + 1 \Rightarrow$ N.º 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:

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

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

$$D_x = \{a, a+1, a+2, \ldots, b\}$$

a → 1º val. do domínio

b → 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: $Y = X_1 + X_2 \sim B(n, p)$

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

→ 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, λ)

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

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

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

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.

→ 1º val. do intervalo

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

↳ último val. do intervalo

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

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

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

→ **exponencial:**

→ Tempo entre eventos.

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

$X \sim Exp(\theta)$ → $\theta > 0$ É Assimétrica à Direita!

X → Tempo/Distância entre ocorrências sucessivas.

Falta de Memória

$P(X \geq a+b \mid X \geq a) = P(X \geq b)$

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

$F(x) \to \underline{\text{pexp}}(x, \frac{1}{\theta})$

$V[X] = \theta^2$

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

→ Normal:

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

$$f(x) \to \underline{\text{dnorm}}(x, \mu, \sigma)$$

$$F(x) \to \underline{\text{pnorm}}(x, \mu, \sigma)$$

$$\text{Inversa} \to \underline{\text{qnorm}}(\text{prob}, \mu, \sigma)$$

$$E[X] = \mu$$

$$V[X] = \sigma^2$$

Aditividade da Normal:

$$\hookrightarrow S = X_1 + X_2 \sim N(\mu, \sigma)$$

$\mu \to \mu_{X_1} + \mu_{X_2}$

$\sigma \to \sigma_{X_1} + \sigma_{X_2}$

→ Normal Standard / Reduzida:

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

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

Nota:

$$\to \Phi(-z) = 1 - \Phi(z)$$

$$f(z) \to \underline{\text{dnorm}}(z)$$

$$\underbrace{F(z)}_{\Phi(z)} \to \underline{\text{pnorm}}(z)$$

$$F^{-1}(z) \to \underline{\text{qnorm}}(\text{prob})$$

→ 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:

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

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
$\sigma = \sigma_\varnothing$	$\sigma \neq \sigma_\varnothing$	Bilateral
	$\sigma > \sigma_\varnothing$	Unilateral Direito
	$\sigma < \sigma_\varnothing$	Unilateral Esquerdo
$\sigma \geq \sigma_\varnothing$	$\sigma < \sigma_\varnothing$	Unilateral Esquerdo
$\sigma \leq \sigma_\varnothing$	$\sigma > \sigma_\varnothing$	Unilateral Direito