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:

→ Dados Quantitativos Contínuos (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()):
    - -> $np = n \times 0.5$:
      - Se np é inteiro: $\tilde{x} = \dfrac{x_{(np)} + x_{(np+1)}}{2}$
      - Senão: $\tilde{x} = x_{([np]+1)}$
- -> Não central:
  - -> Quantis:

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

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 - N: de ...... (contagens)

$f(x) = P(X = x)$ → função Probabilidade

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

→ Nota:

→ $f(x) \geq 0, \forall x$

→ $\sum f(x) = 1$

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

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

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

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

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

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

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

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

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

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

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

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

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

$\sigma^2 \quad = \sum_x (x - \mu)^2 \times f(x)$ $\quad \sqrt{V[X]} = \sigma$

→ Variância

→ Desvio Padrão

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

V. Al. Contínuas

$X - \ldots\ldots$

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

$\hookrightarrow f(x) = F'(x)$

$\hookrightarrow$ Apenas é F.D.P. se:

$\rightarrow f(x) \geq 0, \forall x$

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

Primitivas Importantes

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

$P(k) = kx$

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

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

Valor esperado

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

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

$E[a] = a$

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

Exemplo de uma Integral:

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

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

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

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

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

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

$= 4/8$

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

$V[a] = \emptyset$

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

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

$\hookrightarrow$ 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:

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

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

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

↳ 1º val. do domínio

último val. do domínio

→ Binomial:

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

$X \sim B(n, p) \rightarrow P(\text{Sucesso})$ $\quad V[X] = n.p.(1-p)$

$X \rightarrow$ Nº de Sucessos em n provas.

$f(x) = P(X = x) \Rightarrow \underline{\text{dbinom}}(x, n, p)$

$F(x) = P(X \leq x) \Rightarrow \underline{\text{pbinom}}(x, n, p)$

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

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

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.

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

$X \rightarrow$ Nº de ocorrências num intervalo de medida.

$f(x) = P(X = x) \Rightarrow \underline{\text{dpois}}(x, \lambda)$

$F(x) = P(X \leq x) \Rightarrow \underline{\text{ppois}}(x, \lambda)$

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

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

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)

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

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

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

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

→ Exponencial:

→ Tempo/Dist. entre ocorrências sucessivas.

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

Até a 1ª ocorrência, entre 2 ocorrências

$X \sim Exp(\sigma)$ $\quad D_X = [0, +\infty[$

$E[X] = \sigma$

$V[X] = \sigma^2$

$\sigma > 0$

Assimetria à Direita

Falta de Memória: $P(X \geq a+b \mid X \geq a) = P(X \geq b)$

$f(x) \to \underline{dexp}(x, 1/\sigma)$

$F(x) \to \underline{pexp}(x, 1/\sigma) \to P(X \leq / < x) = 1 - e^{-\frac{x}{\sigma}}, x \geq 0$

$F^{-1}(x) \to \underline{qexp}(prob, 1/\sigma)$

→ Normal:

$X \sim N(\mu, \sigma)$ — μ → Média; σ → Desvio Padrão

$D_x = \mathbb{R}$

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

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

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

$E[X] = \mu$

$V[X] = \sigma^2$

Aditividade da Normal:

↳ $S = X_1 + X_2 \sim N(\mu, \sigma)$

μ → $\mu_{X_1} + \mu_{X_2}$

σ → $\sigma_{X_1} + \sigma_{X_2}$

→ Normal Standard / Reduzida:

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

$$P\left(Z \begin{matrix} \leq \\ < \end{matrix} z\right) = \Phi(z)$$

Nota:

→ $\Phi(-z) = 1 - \Phi(z)$

$f(z)$ → $\underline{dnorm}(z)$

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

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

→ 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

$X \sim P(\lambda)$

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

$\lambda \to$ N.º Médio de ocorrências num intervalo de tempo t.

e

$Y \sim Exp(\theta)$

$Y \to$ Tempo de espera entre ocorrências sucessivas.

$\sigma \to$ Tempo de espera médio entre oc. sucessivas.

então

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

Exemplos com a Regra de 3 Simples:

- $P \to Exp$: (Ex. 21)

2 Avarias — 8h.
1 Avaria — σ

X - N.º de avarias em 8h.

$X \sim P(2)$

Y - Tempo entre avarias consecutivas.

$Y \sim Exp(\sigma) = Exp\left(\frac{1 \times 8}{2}\right) = Exp(4)$

- $Exp \to P$: (Ex. 22)

1 — 90 s = 1.5 m
λ — 6 m

X - Tempo entre chegadas consecutivas, em s.

$X \sim Exp(90)$

Y - N.º de utentes que chegam em 6 minutos.

$Y \sim P(\lambda) = P\left(\frac{1 \times 6}{1.5}\right) = P(4)$

Testes

- **Paramétricos:**

Servem para testar:

- ⊳ μ ⊳ Média
- ⊳ σ^2 ⊳ Variância
- ⊳ p ⊳ Proporção
- ⊳ $\mu_1 - \mu_2$ ⊳ Diferença de Médias
- ⊳ σ_1^2/σ_2^2 ⊳ Quociente de Variâncias
- ⊳ $p_1 - p_2$ ⊳ 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-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 (cf. 5):

Teste Paramétrico ($\mu_1 - \mu_2$)

→ Não: $n \geq 30$ → Não (cf 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