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 $\left(\frac{n_i}{n}\right)$ (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

> No caso de V. nominais apenas se contabiliza a moda!

→ Central:

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

> $nf = n \times 0.5$:
> - Se nf é inteiro: $\tilde{x} = \frac{x_{(nf)} + x_{(nf+1)}}{2}$
> - Senão: $\tilde{x} = x_{([nf]+1)}$ Mediana

→ Não central:

→ Quantis:

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

Medidas de Dispersão

→ Dispersão Absoluta:

- → Amplitude Total: (max() − min())
- → Amplitude Interquartis: $AIQ = Q_3 - Q_1$ = IQR()
- → Variância: σ^2 = Var()
- → Desvio Padrão: σ = sd() = sqrt(σ^2)

→ Dispersão Relativa:

- → Coeficiente de Variação: $CV = \frac{\sigma}{\bar{x}} \times 100\%$

Medidas de Simetria

→ b_1 = skewness() $= \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^3$

- $b_1 = 0$ → Simétrica
- $b_1 > 0$ → Assimétrica positiva / para a direita
- $b_1 < 0$ → Assimétrica negativa / para a esquerda

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

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

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

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

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

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

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

$E[X^2] = \sum x^2 . f(x)$

$E[aX + b] = a . E[X] + b$

$E[a] = a$

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

$V[X] = E[X^2] - E^2[X]$

$V[a] = 0$

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

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

Variância

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

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

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

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

$P(k) = kx$

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

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

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

$E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx$

$E[a] = a$

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

Exemplo de uma Integral:

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

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

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

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

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

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

$= 4/8$

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

$V[a] = 0$

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

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

↳ 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 val. discretos.

$X \sim U(n)$

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

$$f(x) = P(X = x) = \frac{1}{n} = \begin{cases} 1/n \,, & \forall x \in D_x \\ 0 \,, & \text{c.c.} \end{cases}$$

• D_x = Inteiros Consecutivos:

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

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

• $D_x \neq$ Inteiros Consecutivos:

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

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

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

→ Binomial:

→ Para situações em que há 2 resultados possíveis e interessa calcular a probabilidade do nº de (in)sucessos.

$X \sim B(n, \underbrace{p}_{P(\text{Sucesso})})$

$E[X] = n.p$

$V[X] = n.p.(1-p)$

X → Nº de Sucessos em n provas.

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

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

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

$D_X = \{0, 1, 2, \ldots\}$

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

$\boxed{n = n_{X_1} + n_{X_2}}$

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

→ 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.

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

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

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

$D_X = \{0, 1, 2, \ldots\}$

$\boxed{\lambda = \lambda_{X_1} + \lambda_{X_2}}$

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.

$X \sim U(a, b)$

$D_x = [a, b]$

$a \to$ 1º do domínio

$b \to$ último do domínio

$f(x) \to$ dunif (x, a, b)

$F(x) \to$ punif $(x, a, b) \to P(X \leq x) = \frac{x-a}{b-a}$

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

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

→ Exponencial:

→ Tempo/Dist. entre ocorrências sucessivas.

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

$X \sim Exp(\theta)$

$D_x = [0, +\infty[$

$E[X] = \theta$

$V[X] = \theta^2$

$\boxed{\text{Até a 1ª ocorrência, entre 2 ocorrências}}$

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

$f(x) \to$ dexp $(x, 1/\theta)$

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

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

→ Normal:

→ Usada para modelar dados distribuídos simetricamente em torno de uma média.

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

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

$\mu \to$ Média
$\sigma \to$ Desvio

$F(x) \to \underline{\text{pnorm}}(x, \mu, \sigma) \to P(X \leq x)$ / $P(X < x)$

$F^{-1}(x) \to \underline{\text{qnorm}}(prob, \mu, \sigma)$

$E[X] = \mu$

$V[X] = \sigma^2$

$D_X = \mathbb{R}$

→ Normal Standard / Reduzida:

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

→ $P(Z \leq x) = P(Z < x) = \Phi(x)$

$\Phi(-x) = 1 - \Phi(x)$

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

→ $\Phi(x) \to \underline{\text{pnorm}}(x)$

$\Phi^{-1}(x) = z_{(x)} \to \underline{\text{qnorm}}(prob)$

$\left| \begin{array}{l} P(X \leq k) = P(X < k) = \alpha \\ \Phi(k) = \alpha \end{array} \right| \Leftrightarrow k = \Phi^{-1}(\alpha) = z_{(\alpha)}$

- Exemplo entre Normal e Normal Reduzida:

Exemplo 25 dos slides

Dado no Enunciado:

X - Altura a que crescem pinheiros, em m.

$\mu = ?$ $\quad X \sim N(\mu, 1.1) \quad P(X \geq 16) = 0.9$

$\sigma = 1.1$

$E[X] = ? = \mu = 17.41\,m$

$P(X \geq 16) = 0.9 \Leftrightarrow 1 - P(X < 16) = 0.9 \Leftrightarrow$

$F(16) = 0.1 \Leftrightarrow$ $\longrightarrow$ $Z = \frac{x - \mu}{\sigma} \sim N(0,1)$

$P\left(Z < \frac{16 - \mu}{\sigma}\right) = 0.1 \Leftrightarrow$

$\Phi\left(\frac{16 - \mu}{1.1}\right) = 0.1 \Leftrightarrow$ $\longrightarrow$ $\Phi(x)$ é a $F(x)$ de uma normal 0 1.

$\frac{16 - \mu}{1.1} = \Phi^{-1}(0.1) = \underline{qnorm}(0.1) = -1.282 \Leftrightarrow$

$16 - \mu = -1.4102 \Leftrightarrow \mu = 17.41\,m$ //

- Exemplo entre Normal e Binomial:

X - Comprimento de um salto dado por um atleta.

$X \sim N(7.23, 0.33)$

Qual a probabilidade de, em 5 saltos, haver 2 com mais de 7.5m?

$Y \sim B(n, p) \Leftrightarrow Y \sim B(5, 0.2066)$

$n = 5$ saltos

$p = P(\text{Sucesso}) = P(X > 7.5) = 1 - P(X \leq 7.5)$

$= 1 - \text{pnorm}(7.5, 7.23, 0.33) = 0.2066$ //

$P(Y = 2) = \text{dbinom}(2, 5, 0.2066) = 0.2132$ //

- Propriedades da Normal (V. Independentes):

- Aditividade da Normal:

$$Y = \sum_{i=1}^{k} X_i \sim N\left(\sum_{i=1}^{k} \mu_i , \sqrt{\sum_{i=1}^{k} \sigma_i^2}\right)$$

- Combinação Linear da Normal:

$$Y = \sum_{i=1}^{k} a_i X_i \sim N\left(\sum_{i=1}^{k} a_i \mu_i , \sqrt{\sum_{i=1}^{k} a_i^2 \sigma_i^2}\right)$$

- Teorema do Limite Central:

$$T = \sum_{i=1}^{n} X_i \sim N\left(n\mu, \sqrt{n\sigma^2}\right)$$

$(=)$

$$T = \sum_{i=1}^{n} X_i \sim N\left(n\mu, \sigma\sqrt{n}\right)$$

É Parecido à Aditividade, mas é para Distribuições Desconhecidas!

Para $n \geq 30$!

→ Aprox. da Binomial pela Normal:

$$X \sim B(n,p) \xrightarrow{n \to +\infty} X \sim (np, \sqrt{npq})$$

→ Aprox. da Poisson pela Normal:

$$X \sim P(\lambda) \xrightarrow{\lambda \to +\infty} X \sim N(\lambda, \sqrt{\lambda})$$

- **Exemplos das Propriedades:**
  - **Exemplo 26:**

X_1 - Pontuação do jogador 1

X_2 - Pontuação do jogador 2

$X_1 \sim N(84, 5) \quad X_2 \sim N(85, 8)$

1) Aditividade

$P(X_1 + X_2 > 176)$

$Y = X_1 + X_2 \sim N(84 + 85, \sqrt{5^2 + 8^2})$

$Y \sim N(169, \sqrt{89})$

$P(Y > 176) = 1 - F(176)$

$= 1 - pnorm(176, 169, sqrt(89))$

$= 0.229$

2) Combinação Linear

$P(X_2 > X_1) \Leftrightarrow P(X_1 - X_2 < 0)$

$W = X_1 - X_2 \sim N(84 - 85, \sqrt{5^2 - (-1)^2 8^2})$

$W \sim N(-1, \sqrt{89})$

$P(W < 0) = pnorm(0, -1, sqrt(89))$

$= 0.5422$

- **Exemplo 28: T. Limite Central**

X_i - Conteúdo, em l., da garrafa i.

$i = [1, 500] \quad \mu_i = 1 \quad \sqrt{V[X]} = 0.0201$

T - Capacidade de um recipiente com 500 garrafas, em l..

$$T = \sum_{i=1}^{500} X_i \sim N(500,\ 0.0201\sqrt{500})$$

$$P(T > 500.1) = 1 - F_T(500.1) =$$

1 - pnorm(500.1, 500, 0.0201 * sqrt(500)) = 0.412 //

→ **Qui-Quadrado:**

$X \sim X^2(n)$

$f(x)$ → dchisq(x, n)

$F(x)$ → pchisq(x, n)

$F^{-1}(x)$ → qchisq(prob, n)

→ **T-Student:**

$X \sim t(n)$

$f(x)$ → dt(x, n)

$F(x)$ → pt(x, n)

$F^{-1}(x)$ → qt(prob, n)

→ **F de Snedecor:**

$X \sim F(m, n)$

$f(x)$ → df(x, m, n)

$F(x)$ → pf(x, m, n)

$F^{-1}(x)$ → qf(prob, 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.

$\theta \to$ 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 — θ

X - N.º de avarias em 8h.

$X \sim P(2)$

Y - Tempo entre avarias consecutivas.

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

- $Exp \to P$: (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

→ Diferença de Medianas:

- → Wilcoxon
- → Mann-Whitney

• Escolha do Tipo de Teste

H_0	H_1	Tipo de Teste
	$\sigma \neq \sigma_0$	Bilateral
$\sigma = \sigma_0$	$\sigma > \sigma_0$	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

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 (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)