

Estatística Descritiva

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \bar{x} = \frac{1}{n} \sum_{k=1}^K n_k x_k; \bar{x} = \sum_{k=1}^K f_k x_k.$$

$$s_c^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2; s_c^2 = \frac{n}{n-1} \sum_{k=1}^K f_k (x_k - \bar{x})^2.$$

Nota: $s^2 = \frac{(n-1)s_c^2}{n}$.

Momentos de ordem θ ($\theta = 1, 2, \dots$) em relação à média:

$$m_\theta = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^\theta; m_\theta = \sum_{k=1}^K f_k (x_k - \bar{x})^\theta.$$

Coeficiente de assimetria (*skewness*): $\frac{n^2}{(n-1)(n-2)} \frac{m_3}{s_c^3}$.

Ordem da observação correspondente ao quantil ordem α : $(n+1)\alpha$.

Barreiras de *outliers*:

$$Q_1 - 1.5(Q_3 - Q_1) \text{ e } Q_3 + 1.5(Q_3 - Q_1);$$

$$Q_1 - 3.0(Q_3 - Q_1) \text{ e } Q_3 + 3.0(Q_3 - Q_1).$$

Teoria Elementar da Probabilidade

$$0 \leq P(A) \leq 1; P(\emptyset) = 0; P(\Omega) = 1;$$

$$P(\bar{A}) = 1 - P(A); P(A) \leq P(B) \text{ se } A \subset B;$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B);$$

$$P(A \cup B) = P(A) + P(B) \text{ se } A \text{ e } B \text{ disjuntos};$$

$$P(A \cap B) = P(A)P(B) \text{ se } A \text{ e } B \text{ independentes.}$$

Probabilidade condicionada

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)P(B);$$

$$P(A|B) = P(A) \text{ se } A \text{ e } B \text{ independentes};$$

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B).$$

Teoremas: Se $\Omega = A_1 \cup \dots \cup A_n$, $A_i \cap A_j = \emptyset$, $i \neq j$, então,

(i) **Bayes:** $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}$;

(ii) **Prob. Total:** $P(B) = \sum_{j=1}^n P(A_j)P(B|A_j)$.

Variáveis Aleatórias (v.a.'s)

Função massa de prob. (fmp), v.a. X discreta:

$$f(x) = P(X = x) \text{ é fmp sse (i) } 0 \leq f(x) \leq 1 \text{ e}$$

$$(ii) \sum x_i f(x_i) = 1.$$

Função densidade de prob. (fdp), v.a. X contínua:

$$f(x) \text{ é fdp sse (i) } f(x) \geq 0 \text{ e (ii) } \int_{-\infty}^{+\infty} f(x)dx = 1.$$

Função de distribuição (fd)

$$F(x) = P(X \leq x) = \begin{cases} \sum_{x_i \leq x} f(x_i), & X \text{ discreta;} \\ \int_{-\infty}^x f(t) dt, & X \text{ contínua;} \end{cases}$$

$F(x)$ é fd sse (i) $0 \leq F(x) \leq 1$, (ii) F é contínua (só à direita, se X discreta), (iii) F é não decrescente,

$$(iv) \lim_{x \rightarrow -\infty} F(x) = 0 \text{ e } \lim_{x \rightarrow +\infty} F(x) = 1.$$

Se X contínua, então $F'(x) = f(x)$.

Média ou valor esperado: $\mu = E[X]$

$$E(X) = \begin{cases} \sum_{x_i} x_i P(X = x_i), & X \text{ discreta} \\ \int_{-\infty}^{+\infty} x f(x) dx, & X \text{ contínua} \end{cases}$$

$$E[aX + b] = aE[X] + b; E[X + Y] = E[X] + E[Y];$$

Variância: $\sigma^2 = \text{Var}[X]$ (Desvio padrão $\sigma = \sqrt{\sigma^2}$)

$$\text{Var}(X) = \begin{cases} \sum_{x_i} (x_i - \mu)^2 P(X = x_i), & X \text{ discreta} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx, & X \text{ contínua} \end{cases}$$

$$\text{Var}[X] = E[X^2] - \mu^2; \text{Var}[aX + b] = a^2 \text{Var}[X];$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \text{ se } X \text{ e } Y \text{ independentes};$$

Das propriedades resulta que: Se X_1, \dots, X_n são v.a.'s i.i.d. com $E[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$, então $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ tem $E[\bar{X}] = \mu$ e $\text{Var}[\bar{X}] = \sigma^2/n$.

Quantil de ordem p

$$x_p = \begin{cases} x_p : p \leq F(x_p) \leq p + P(X = x_p), & X \text{ discreta} \\ x_p : F(x_p) = p, & X \text{ contínua} \end{cases}$$

(Exemplo: a mediana é o quantil de ordem $p = 1/2$.)

Distribuições e TLC

Bernoulli $X \sim Be(p)$

$X = 1$ ("sucesso" com prob. p) ou 0 (prob. $1 - p$);

$$P(X = x) = p^x(1 - p)^{1-x}, \quad x = 0, 1;$$

$$E[X] = p; \text{Var}[X] = p(1 - p).$$

Binomial $X \sim B(n, p)$

$X = \text{n}^\circ$ de "sucessos" em n provas independentes;

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, \dots, n$$

$$E[X] = np; \text{Var}[X] = np(1 - p);$$

Propriedade Se X_1, X_2, \dots, X_m são v.a.'s independentes com $X_i \sim B(n_i, p)$, $i = 1, 2, \dots, m$, então

$$\sum_{i=1}^m X_i \sim B\left(\sum_{i=1}^m n_i, p\right).$$

Geométrica (nº de provas) $X \sim G(p)$

$X = \text{n}^\circ$ de provas independentes até obter o primeiro "sucesso";

$$P(X = x) = (1 - p)^{x-1} p, \quad x = 1, 2, \dots;$$

$$E[X] = 1/p; \text{Var}[X] = (1 - p)/p^2.$$

Hipergeométrica $X \sim H(N, M, n)$

$X = \text{n}^\circ$ de "sucessos" em n provas dependentes;

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, \quad \forall x \text{ possíveis};$$

$$E[X] = np; \text{Var}[X] = np(1 - p) \frac{N-n}{N-1}; p = M/N.$$

Poisson $X \sim P(\lambda)$

Em geral, $X = \text{n}^\circ$ de ocorrências por intervalo de tempo ou regiões do espaço;

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, \dots; \lambda > 0;$$

$$E[X] = \lambda; \text{Var}[X] = \lambda;$$

Propriedade Se X_1, X_2, \dots, X_n são v.a.'s independentes com $X_i \sim P(\lambda_i)$, $i = 1, 2, \dots, n$, então

$$\sum_{i=1}^n X_i \sim P\left(\sum_{i=1}^n \lambda_i\right).$$

Uniforme $X \sim U(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases} \quad \begin{matrix} E[X] = (a+b)/2; \\ \text{Var}[X] = (b-a)^2/12. \end{matrix}$$

Exponencial $X \sim \text{Exp}(\lambda)$

Em geral, $X = \text{tempo de duração ou tempo entre ocorrências consecutivas};$

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \lambda > 0;$$

$$E[X] = 1/\lambda; \text{Var}[X] = 1/\lambda^2.$$

Normal $X \sim N(\mu, \sigma^2)$ **Normal Standard** $X \sim N(0, 1)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x, \mu \in \mathbb{R}, \sigma > 0;$$

$$E[X] = \mu; \text{Var}[X] = \sigma^2;$$

(I) Se $X \sim N(\mu, \sigma^2)$, então (i) $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ e

(ii) $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$;

(II) Se $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ são independentes, então $\sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$;

(III) Se $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$ são independentes, então $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \sigma^2/n)$.

Teorema do Limite Central (TLC)

Se $X_1, X_2, \dots, X_n, \dots$ é uma sucessão de v.a's i.i.d. com $E[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$, então

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{n \rightarrow +\infty} N(0, 1).$$

Do **TLC** resulta que (i) $\sum_{i=1}^n X_i \overset{\text{aprox}}{\sim} N(n\mu, n\sigma^2)$ e (ii) $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \overset{\text{aprox}}{\sim} N(\mu, \sigma^2/n)$, para n suficientemente elevado (em geral $n > 30$).

Estimação pontual

Método dos Momentos: (usualmente)

Um parâmetro: $E[X] = \bar{X}$ (ou $\text{Var}[X] = S^2$);

Dois parâmetros: $E[X] = \bar{X}$ e $\text{Var}[X] = S^2$.

Método da Máxima Verosimilhança:

(1) escrever a função de verosimilhança $L(\theta)$;

(2) considerar o logaritmo de $L(\theta)$, se necessário;

(3) obter o maximizante de $L(\theta)$ (ou $\ln L(\theta)$).

TH - noções e procedimentos

Nível de significância:

$\alpha = P(\text{Erro Tipo I}) = P(\text{rejeitar } H_0 | H_0 \text{ verdadeira});$

Potência do teste: $1 - \beta = 1 - P(\text{Erro Tipo II}) =$

$= P(\text{rejeitar } H_0 | H_1 \text{ verdadeira});$

Conversão p -value bilateral a unilateral:

Se a amostra aponta no sentido de H_1 , então

$p\text{-value}_{\text{uni}} = \frac{p\text{-value}_{\text{bil}}}{2}$; senão $p\text{-value}_{\text{uni}} = 1 - \frac{p\text{-value}_{\text{bil}}}{2}$.

Cálculo de p -value num teste bilateral:

$p\text{-value}_{\text{bil}} = 2P(T < t_{\text{obs}} | H_0)$, se t_{obs} é reduzido;

$p\text{-value}_{\text{bil}} = 2P(T > t_{\text{obs}} | H_0)$, se t_{obs} é elevado.

Considera-se t_{obs} reduzido (elevado) quando a estimativa que se obtém para o parâmetro a testar é inferior (superior) ao valor especificado em H_0 .

IC e TH em Populações Normais

Média populacional, μ , com σ^2 conhecida

Variável fulcral: $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1);$

$$\text{IC}_{1-\alpha}(\mu) = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right);$$

TH $\rightarrow H_0 : \mu = \mu_0$; usa-se ET $\underset{\text{sob } H_0}{\sim} N(0, 1)$.

(ET significa Estatística de Teste)

Média populacional, μ , com σ^2 desconhecida

Variável fulcral: $T = \frac{\bar{X} - \mu}{S_c/\sqrt{n}} \sim t_{n-1};$

$$\text{IC}_{1-\alpha}(\mu) = \left(\bar{X} - t_{1-\frac{\alpha}{2}, n-1} \frac{S_c}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \frac{S_c}{\sqrt{n}} \right);$$

TH $\rightarrow H_0 : \mu = \mu_0$; usa-se ET $\underset{\text{sob } H_0}{\sim} t_{n-1}$.

Diferença $\mu_X - \mu_Y$ em amostras emparelhadas com variâncias desconhecidas

Considerar $D_i = X_i - Y_i \sim N(\mu_D, \sigma_D^2)$, com $\mu_D = \mu_X - \mu_Y$.

Variável fulcral: $T = \frac{\bar{D} - \mu_D}{S_{cD}/\sqrt{n}} \sim t_{n-1};$

$$\text{IC}_{1-\alpha}(\mu_D) = \left(\bar{D} - t_{1-\frac{\alpha}{2}, n-1} \frac{S_{cD}}{\sqrt{n}}, \bar{D} + t_{1-\frac{\alpha}{2}, n-1} \frac{S_{cD}}{\sqrt{n}} \right)$$

TH $\rightarrow H_0 : \mu_D = \mu_0$; usa-se ET $\underset{\text{sob } H_0}{\sim} t_{n-1}$.

Diferença $\mu_X - \mu_Y$ em amostras independentes com variâncias conhecidas

Variável fulcral: $Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0, 1);$

$$\text{IC}_{1-\alpha}(\mu_X - \mu_Y) = \left(\bar{X} - \bar{Y} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \right);$$

TH $\rightarrow H_0 : \mu_X - \mu_Y = \mu_0$; usa-se ET $\underset{\text{sob } H_0}{\sim} N(0, 1)$.

Diferença $\mu_X - \mu_Y$ em amostras independentes com variâncias desconhecidas (consideradas iguais)

Variável fulcral: $T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S} \sim t_{n+m-2},$

$$\text{onde } S = \sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)S_{cX}^2 + (m-1)S_{cY}^2}{n+m-2}};$$

$$\text{IC}_{1-\alpha}(\mu_X - \mu_Y) = (\bar{X} - \bar{Y} \mp t_{1-\frac{\alpha}{2}, n+m-2} S);$$

TH $\rightarrow H_0 : \mu_X - \mu_Y = \mu_0$; usa-se ET $\underset{\text{sob } H_0}{\sim} t_{n+m-2}$.

Variância populacional, σ^2 , com μ conhecida

Variável fulcral: $V = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2;$

$$\text{IC}_{1-\alpha}(\sigma^2) = \left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}, n}^2}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}, n}^2} \right);$$

TH $\rightarrow H_0 : \sigma^2 = \sigma_0^2$; usa-se ET $\underset{\text{sob } H_0}{\sim} \chi_n^2$.

Variância populacional, σ^2 , com μ desconhecida

Variável fulcral: $V = \frac{(n-1)S_c^2}{\sigma^2} \sim \chi_{n-1}^2;$

$$\text{IC}_{1-\alpha}(\sigma^2) = \left(\frac{(n-1)S_c^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)S_c^2}{\chi_{\frac{\alpha}{2}, n-1}^2} \right);$$

TH $\rightarrow H_0 : \sigma^2 = \sigma_0^2$; usa-se ET $\underset{\text{sob } H_0}{\sim} \chi_{n-1}^2$.

IC e TH em outras Populações (invocando TLC)

Média populacional, μ , com σ^2 (des)conhecida

Variável fulcral: $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{aprox}}{\sim} N(0, 1);$

Variável fulcral: $Z = \frac{\bar{X} - \mu}{S_c/\sqrt{n}} \overset{\text{aprox}}{\sim} N(0, 1);$

$$\text{IC}_{1-\alpha}(\mu) \overset{\text{TLC}}{\approx} \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right);$$

$$\text{IC}_{1-\alpha}(\mu) \overset{\text{TLC}}{\approx} \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{S_c}{\sqrt{n}} \right);$$

TH $\rightarrow H_0 : \mu = \mu_0$; usa-se ET $\overset{\text{aprox}}{\underset{\text{sob } H_0}{\sim}} N(0, 1)$.

Proporção populacional, p

Estimador para p : $\hat{p} = \frac{X}{n} \overset{\text{aprox}}{\sim} N\left(p, \frac{p(1-p)}{n}\right).$

(X representa a quantidade de interesse na amostra.)

Variável fulcral: $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\text{aprox}}{\sim} N(0, 1);$

$$\text{IC}_{1-\alpha}(p) \overset{\text{TLC}}{\approx} \left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right);$$

TH $\rightarrow H_0 : p = p_0$; usa-se ET $\overset{\text{aprox}}{\underset{\text{sob } H_0}{\sim}} N(0, 1)$.