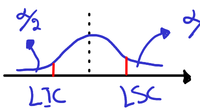
Estimativas: determinação de intervalos de confiança.

A confiabilidade do resultado será dentro de um interv. adotando um nível de significância (α) e uma representação amostral validada por uma quant. suficiente de dados.



LIC - Limite Inferior de Confiança
LSC - Limite Superior de Confiança

Estimativa para média populacional quando σ^2 é conhecida

n = tam. amostral
$\bar{x}$ = media amostral

$$P\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Estimativa para média populacional quando σ^2 é desconhecida

n = tam. amostral
$\bar{x}$ = media amostral

$$P\left(\bar{x} - t_{\alpha} \cdot \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha} \cdot \frac{s}{\sqrt{n}}\right) = 1-\alpha$$

t_α é obtido pela tabela t, adotando $\varphi = (n-1)$ e α

Estimativa para proporção populacional

tendo $f = \frac{x}{n}$; $P\left(f - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} \le p \le f + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}}\right) = 1-\alpha$

Estimativa para diferença de μ quando as σ^2 são conhecidas

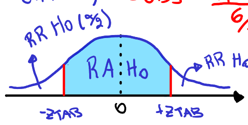
conhecidos $n_1, \bar{x}_1, \sigma_1^2$ e $n_2, \bar{x}_2, \sigma_2^2$ fazemos $d = \bar{x}_1 - \bar{x}_2$ e estimamos:

$$P\left(d - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le d + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1-\alpha$$

TH para média populacional (μ) quando σ^2 é conhecida

dados as hipóteses: $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ e conhecidos: $n, \bar{x}$ e σ^2

então, $Z_{OBS} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ e adotando α, teremos:



Se $Z_{OBS} \in [-Z_{TAB}, Z_{TAB}]$ aceitamos H_0
Se $Z_{OBS} \notin [-Z_{TAB}, Z_{TAB}]$ rejeitamos H_0

TH para média populacional (μ) quando σ^2 é conhecida

Se $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ e conhecidos: $n, \bar{x}$ e s^2 então determinamos: $t_{OBS} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

O intervalo de aceitação será dado por $[-T_{AB}, T_{AB}]$ sendo a estatística t, definida por α e $\varphi = n-1$

TH para proporção

Dados as hipóteses $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$ e conhecidos $f = \frac{x}{n}$

então $Z_{OBS} = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ dado um nível de significância α, teremos:

Se $Z_{OBS} \in [-Z_{TAB}, Z_{TAB}]$ aceitamos H_0
Se $Z_{OBS} \notin [-Z_{TAB}, Z_{TAB}]$ rejeitamos H_0

TH para a diferença de médias (μ) quando σ^2 são conhecidos

dados as hipóteses: $\begin{cases} H_0: \mu_A = \mu_B \rightarrow \mu_A - \mu_B = 0 \\ H_1: \mu_A \neq \mu_B \rightarrow \mu_A - \mu_B \neq 0 \end{cases}$

conhecidos $n_A, \bar{x}_A, \sigma^2_A, n_B, \bar{x}_B, \sigma^2_B$ então temos:

Estimativa para diferença de μ quando as σ^2 são desconhecidas

conhecidos $n_1, \bar{x}_1, s_1^2$ e $n_2, \bar{x}_2, s_2^2$ fazemos $d = \bar{x}_1 - \bar{x}_2$ e determinamos $S^2c = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}$

adotando $\varphi = n_1 - n_2 - 2$ e α, de modo que:

$$P\left(d - t_\alpha . S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le d + t_\alpha . S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1-\alpha$$

Estimativa para diferença de proporções

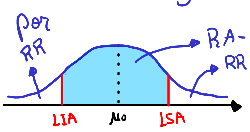
conhecidos $p_{s1} = \frac{x_1}{n_1}$, $p_{s2} = \frac{x_2}{n_2}$ e $\bar{p}_s = \frac{x_1 + x_2}{n_1 + n_2}$ temos $d = p_{s1} - p_{s2}$

$$P\left(d - z_{\frac{\alpha}{2}} \sqrt{\bar{p}_s(1-\bar{p}_s)\left(\frac{1}{n_1} + \frac{1}{n_1}\right)} \le p_1 - p_2 \le d + z_{\frac{\alpha}{2}} \sqrt{\bar{p}_s(1-\bar{p}_s)\left(\frac{1}{n_1} + \frac{1}{n_1}\right)}\right) = 1-\alpha$$

Teste de hipóteses: Para validar uma informação a respeito de uma medida populacional

Hipótese Nula (H_0): Medida que deseja validar
Hipótese Alternativa (H_1): Será considerada quando H_0 for rejeitada

Teremos a região de aceitação dada, no caso $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ por



LIA - Limitante Inferior de Aceitação
LSA - Limitante Superior de Aceitação

$Z_{OBS} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma^2_A}{n_A} + \frac{\sigma^2_B}{n_B}}}$ dado α, temos:

Se $Z_{OBS} \in [-Z_{TAB}, Z_{TAB}]$ aceitamos H_0
Se $Z_{OBS} \notin [-Z_{TAB}, Z_{TAB}]$ rejeitamos H_0

TH para a diferença de médias (μ) quando σ^2 são desconhecidos

Dados $\begin{cases} H_0: \mu_A = \mu_B \\ H_1: \mu_A \neq \mu_B \end{cases}$ e ainda, $n_A, \bar{x}_A, s^2_A, n_B, \bar{x}_B, s^2_B$ determinamos

$S^2c = \frac{(n_A-1).s^2_A + (n_A-1)s^2_B}{n_A+n_B-2}$; $t_{OBS} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{S_c . \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$

Adotando α e $\varphi = n_A + n_B - 2$ encontramos t_{TAB}, de forma que:
Se $t_{OBS} \in [-t_{TAB}, t_{TAB}]$ aceitamos H_0
Se $t_{OBS} \notin [-t_{TAB}, t_{TAB}]$ rejeitamos H_0

TH para diferença de proporção

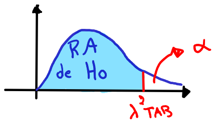
dados as hipóteses $\begin{cases} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{cases}$ e conhecidos $p_{s1} = \frac{x_1}{n_1}$; $p_{s2} = \frac{x_2}{n_2}$; $\bar{p}_s = \frac{x_1 + x_2}{n_1 + n_2}$

determinamos:

$Z_{OBS} = \frac{(P_{s1} + P_{s2}) - (P_1 - P_2)}{\sqrt{\bar{p}_s(1-\bar{p}_s)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ Segundo nível de significância α, teremos, Z_{TAB}, e ainda:

Se $Z_{OBS} \in [-Z_{TAB}, Z_{TAB}]$ aceitamos H_0
Se $Z_{OBS} \notin [-Z_{TAB}, Z_{TAB}]$ rejeitamos H_0

Teste pelo Qui-Quadrado: Quando desejarmos validar hipóteses na qual se analisam a forma de distribuição da ocorrência dos elementos do espaço amostral, considerando suas frequências, faremos o uso deste teste, podendo ser uma análise de aderência ou independência.



$\varphi = n-1$

Teste de Aderência

$\lambda^2_{OBS} = \sum \frac{(fo-fe)^2}{fe}$ e comparamos com λ^2_{TAB}

Teste de Independencia

$\lambda^2_{OBS} = \sum \frac{(fo-fe)^2}{fe}$ e comparamos com λ^2_{TAB} adotamos:

$fe = \frac{(\text{total linha})(\text{total coluna})}{\text{total geral}}$ e $\varphi = (\text{linha}-1)(\text{coluna}-1)$

TH para variância: validar se populações tem as mesmas variabilidades

Vamos testar $\begin{cases} H_0: \sigma^2_A = \sigma^2_B \\ H_1: \sigma^2_A \neq \sigma^2_B \end{cases}$ conhecidos $\bar{x}_A, n_A, s^2_A, x_B, n_B$ e s^2_B

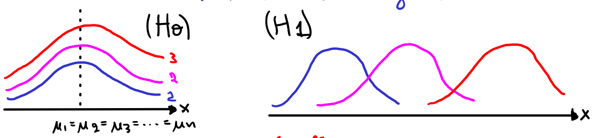
estabelecemos $F_{OBS} = \frac{S^2_N}{S^2_D}$ Se $S^2_A > S^2_B$, $S^2_N = S^2_A$ e $S^2_D = S^2_B$
Se $S^2_B > S^2_A$, $S^2_N = S^2_B$ e $S^2_D = S^2_A$

⚠ maior em cima.

Se $F_{OBS} < F_{TAB}$ aceitamos Ho, caso contrário, rejeitamos.

TH de variância para comparação de grupos (ANOVA).

Adotamos $\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n \\ H_1: \text{nem todos são iguais} \end{cases}$



Media Geral: $\bar{\bar{x}} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n_j} x_{ij}}{n}$

$SQE = \sum_{j=1}^{c} n_j(\bar{x}_j - \bar{\bar{x}})^2$ → Variação entre grupos

$SQD = \sum_{j=1}^{c}\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$ → Variação dentro dos grupos

Encontramos então: $MQE = \frac{SQE}{c-1}$ $MQD = \frac{SQD}{n-c}$

e obtemos a estatística F: $F_{OBS} = \frac{MQE}{MQD}$

dados α, φ_E, φ_D, obtemos F_{TAB}, se $F_{OBS} < F_{TAB}$, aceitamos Ho. Caso contrário rejeitamos.

Análise de Correção e Regressão: Validar a existência de relação entre variáveis, dada informações em pares (x,y), faremos:

$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$, onde $S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$

$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$ e $S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$

se $|r| = 1$ → Correlação perfeita
$0{,}7 \leq |r| < 1$ → Forte correlação
$0{,}5 \leq |r| < 0{,}7$ → Moderada correlação
$|r| < 0{,}5$ → fraca correlação

Por fim, fazemos $r^2 = (\%)$ → coef. de determinação

Regressão Linear: Expressar relação entre variáveis através de função de 1º grau.

$y = a + bx$ onde a e b são encontrados através de $\begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$

Regressão Quadrática

$y = b_0 + b_1x + b_2x^2$ → $\begin{bmatrix} n & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \\ \sum x^2y \end{bmatrix}$

Regressão exponencial

$y = ae^{bx}$ $\begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum \ln y \\ \sum x \ln y \end{bmatrix}$

Análise de Variância Residual

Para cada $P(x_i, y_i)$ determinamos a imagem da projeção x_i no gráfico $(\hat{y}_i)$

$e_i = (y_i - \hat{y}_i)^2$
↳ resíduo no ponto

$\sigma^2_R = \frac{\sum (y_i - \hat{y}_i)^2}{n}$
↳ escolhemos qual das regressões possui menor σ^2_R

Regressão Multipla

$y = f(x_1, x_2, x_3, \dots, x_k) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$

$$\begin{bmatrix} n & \sum x_1 & \sum x_2 & \cdots & \sum x_k \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & & \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \sum x_k & & & & \sum x_k^2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum x_1y \\ \sum x_2y \\ \vdots \\ \sum x_ky \end{bmatrix}$$