

Resumo Prova 2

↳(revisão)

Sophia Carrozza

Inferência Estatística:

fazer generalizações sobre uma população com base em uma amostra

- confiança = probabilidade de acerto na estimação ($1-\alpha$)
- risco = probabilidade de erro na estimação (α)
 ↳ significância

estimação pontual vs estimação intervalar

Fórmulas da estatística básica (média, variância e desvio padrão).

Depende de + fatores, como o módulo da probabilidade, variabilidade, confiança, etc.

Estimação Intervalar: determina um intervalo de valores que tem a maior probabilidade de conter o verdadeiro valor de um parâmetro populacional, c/ um certo nível de confiança.

→ Esse intervalo é chamado de Intervalo de Confiança (IC)

a amplitude da estimação intervalar é determinada pelo qntd. de desvios padrão e a confiança desejada.

↳ representada por Z (normal) ou t (Student)

Nível de Confiança	Nível de significância	Valores de Z
90%	0,10	1,65
95%	0,05	1,96
99%	0,01	2,58

Média (μ) (populacional)
quando σ é conhecido:

$$IC(1-\alpha) = \bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Proporção:

$$IC(1-\alpha) = \hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Média (μ)
quando σ é conhecido: (amostral)

$$IC(1-\alpha) = \bar{x} \pm t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}}$$

Amostragem aleatória:

- ↳ aleatória simples = todas amostras de mesmo tamanho têm mesma prob.
- ↳ sistemática = a partir de um ponto inicial aleatório, seleciona-se amostras em intervalos regulares
- ↳ estratificada = a população é dividida em subgrupos homogêneos, chamados estratos e é feita uma amostragem aleatória dentro desses estratos
- ↳ por conglomerado = a população é dividida em grupos maternais ou conglomerados aleatórios (a variabilidade entre os conglomerados é baixa)

baseados em
características
semelhantes

Teste de Hipóteses: verifica se os dados amostrais trazem evidências suficientes p/ apoiar ou não uma hipótese estatística

► hipótese nula (H_0) = símbl de igualdade

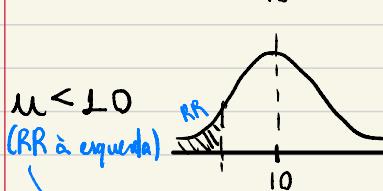
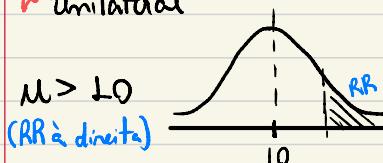
► hipótese alternativa (H_1) = símbl de desigualdade

► tipos de teste:

► bilateral



► unilateral



↳ Região de Rejeição
da hipótese nula

► passos dos testes:

1- estabelecer as hipóteses

2- estabelecer as regiões (RA; RR)

3- calcular a estatística do teste

4- concluir o teste e calcular o valor P

↳ valor p: menor valor de alfa que levaria a rejeição da hipótese nula

$P(Z \leq \text{valor de } z \text{ encontrado})$
na tabela

↳ subtraia de 1 p/ ter a probab. na cauda direita

↳ p/ um teste bilateral, multiplique por 2.

Média (μ)

desvio padrão é conhecido:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \quad \left. \begin{array}{l} \text{valor que estamos} \\ \text{comparando} \end{array} \right\} SE$$

(desvio padrão!)

Média (μ)

σ não é conhecido:

$$T = \frac{\bar{x} - \mu_0}{S / \sqrt{n}}$$

Proporções:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

\bar{x} = média amostral

μ_0 = média populacional sob H_0

σ = desvio padrão populacional

n = tamanho da amostra

Possíveis erros e acertos de uma decisão a partir de um teste de hipótese.

		Realidade	
		H_0 verdadeira	H_0 falsa
Decisão	Aceitar H_0	Decisão correta ($1 - \alpha$)	Erro tipo II (β)
	Rejeitar H_0	Erro tipo I (α)	Decisão correta ($1 - \beta$)

↳ Depois de calcular Z (ou T), que significa o Z_{calc} , verificamos o $Z_{crítico}$, que é a significância metabola (não a confiança) e vemos se:

$\frac{\alpha}{2}$ se forbi et neforum

* pois estamos interessados na cauda esquerda

$\mu < \mu_0 \quad Z_{calc} \leq \ominus Z_{crítica}$

$\mu > \mu_0 \quad Z_{calc} > + Z_{crítica}$

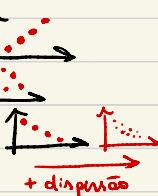
$\mu = \mu_0 \quad Z_{calc} \leq + Z_{crítica}$

Correlação e Regressão:

► Correlação → mede o grau de associação em relação linear entre duas variáveis

(o coeficiente varia entre -1 e $+1$)

- $r = 1 \rightarrow$ correlação perfeita positiva
- $r = 0 \rightarrow$ nenhuma correlação linear
- $r = -1 \rightarrow$ correlação perfeita negativa



+ dispersão

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} \quad \text{ou} \quad r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

(adicional)
+ reloções quadrática

► Regressão linear → usada p/ modelar a relação entre uma variável dependente (y) e uma independente (x)

equação da reta { $y = a + bx$ coeficiente angular

↳ inturupto

$$a = \bar{y} - b\bar{x}$$

↳ intercepto é o valor de y quando $x = 0$.

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

↳ coeficiente angular é a taxa de variação y em relação a x .

Outras Informações:

Como saber se o desvio padrão é populacional (σ) ou amostral (s)?

descreve a dispersão de uma variável dentro de toda a população
 - menciona que é o σ
 é populacional

é calculado a partir de uma amostra da pop., serve como uma estimativa do σ (tem + incerteza).
 - não diz que é populacional
 - menciona que foi obtida de uma amostra
 - o tamanho da amostra é pequeno ($n < 30$)

$$\text{grau de liberdade} = n - 1$$

$$\text{significância} = 1 - \text{confiança} (\alpha)$$

p/ estimação de intervalo,
sempre vai ser bilateral!

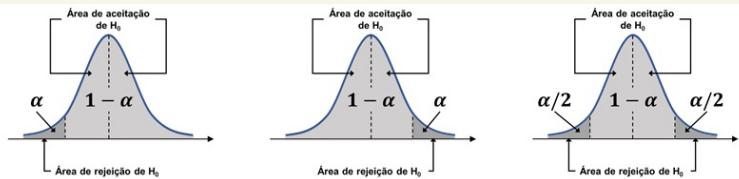
► Como olhar na tabela Z (normal)

- > determinar a significância ($1 - \text{confiança}$)
- > p/ testes bilaterais $\rightarrow \alpha/2$
- > p/ testes unilaterais \rightarrow use α diretamente
- > na tabela da Demirz, é para olhar o valor de $Z_{\frac{\alpha}{2}}$ ou só Z (não é unilateral).

p/ teste da hipótese, depende ($>, <, \neq$)

► Como olhar na tabela t (t Student)

- > determinar a significância
- > p/ testes bilaterais $\rightarrow \alpha/2$
- > p/ testes unilaterais $\rightarrow \alpha$
- > verifique o grau de liberdade na tabela da Demirz, este de acordo com o $\alpha/2$ e ± 1 (ou de ± 1)



	Teste da Cauda Inferior	Teste da Cauda Superior	Teste Bicaudal
Hipóteses	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
Estatística de Teste	$Z = \frac{u - \mu(u)}{\sigma(u)}$		
Regras de Rejeição de H_0			
Valor Crítico z	Rejeitar se $Z \leq -z_\alpha$	Rejeitar se $Z \geq z_\alpha$	Rejeitar se $Z \leq -z_{\alpha/2}$ ou se $Z \geq z_{\alpha/2}$
Valor Crítico t	Rejeitar se $Z \leq -t_\alpha$	Rejeitar se $Z \geq t_\alpha$	Rejeitar se $Z \leq -t_{\alpha/2}$ ou se $Z \geq t_{\alpha/2}$
Valor p	Rejeitar se $p_value \leq \alpha$		

Interpretacão de um Intervalo de Confiança: "Pode-se dizer, com $X\%$ de confiança, que o \bar{x} é o \bar{x} médio do \bar{x} e que \bar{x} está entre $\bar{x} - E$ e $\bar{x} + E$."

$$n = \left(\frac{Z \cdot \sigma}{E} \right)^2$$

m. de observações

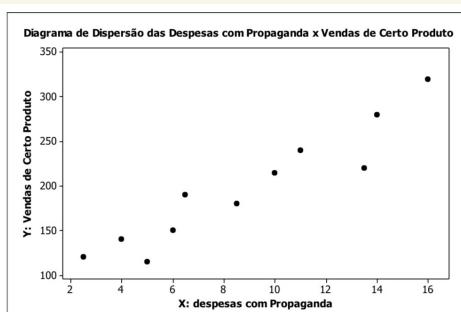
↓

GRPO

(sempre arredonda p/ cima)

$$\text{Média Populacional} = Z \cdot E R R O$$

(Amplitude total)



O gráfico sugere relação linear direta entre as variáveis despesas com propaganda e vendas de certo produto.