Teste de Hipóteses

Para populações

Prof.: Wagner Pinheiro
wagner2235@gmail.com

Sumário

Introdução

O teste de Hipóteses é uma regra decisória que nos permite reajustar ou não uma hipótese estatística com base nos resultados de uma amostra

De outro modo: são suposições que fazemos para testar a fixação de decisões, que poderão ser verdadeiras ou não.

i) Parâmetro: em função de valores populacionais, é desconhecido.

Ex.: Parâmetros da distribuição Normal

$$\begin{cases} \mu = E(X) \\ \sigma^2 = V(X) \end{cases}$$

i) Estimador: Em função das observações da amostra aleatória. Representa uma forma de cálculo que fornecerá valores diferentes conforme a amostra selecionada.

Ex.:

- Estimador da média: $\bar{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
- Estimador da variância: $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$

iii) Estimativa: valor obtido pelo estimador.

Ex.: $\bar{X} = 10,42$; $s^2 = 4,67$

É uma suposição quanto ao valor de um parâmetro populacional que será verificada para um teste paramétrico.

1. A média populacional da altura dos brasileiros adultos é 1,65m, isto é, $\mu = 1,65m$;
2. A distribuição dos pesos dos alunos do CEFET-MG é normal;
3. A proporção de indivíduos com doença X é 3%, ou seja, $p = 0,03$.

Hipótese Nula (H_0): a ser validada pelo teste. Hipótese Alternativa (H_1 ou H_a): complementar a H_0.

Assim, o teste poderá aceitar ou rejeitar a hipótese nula, sendo que no último caso implicaria na aceitação da hipótese alternativa.

Hipótese Nula

A hipótese H_0 é formulada com o "expresso propósito de ser rejeitada" e os teste são construídos sob a pressuposição de H_0 ser verdadeira.

Ex.: O fabricante afirma que a durabilidade média de suas lâmpadas é de $6.000h$, para esse caso a hipótese nula seria formulada da seguinte maneira: $H_0 : \mu = 6000$.

Hipótese alternativa

É contrária à H_0, formulada com base no conhecimento prévio do problema.

Ex.: A durabilidade média das lâmpadas é diferente da afirmativa do fabricante, para esse caso a hipótese alternativa seria formulada da seguinte maneira: $H_a : \mu \neq 6000$.

Para o caso onde se busca comparar a média, a H_a fica

$$H_{a_1} : \mu_1 < \mu_2 \quad \text{ou} \quad H_{a_2} : \mu_1 > \mu_2 \quad \text{ou} \quad H_{a_3} : \mu_1 \neq \mu_2$$

nesse caso H_{a_1} e H_{a_2} são unilaterais e H_{a_3} é bilateral.

Região Crítica

É a faixa de valores que nos leva à rejeição de H_0. Isto é, caso o valor observado da estatística de teste (Z, t, χ^2, F) pertença a região crítica, rejeita-se H_0, caso contrário não rejeita-se H_0.

Qualquer decisão implica na possibilidade de cometer basicamente dois tipos de erros, são eles:

DECISÃO	**REALIDADE**	
	H_0 Verdadeira	H_0 Falsa
Aceita H_0	Decisão Correta (1- α)	Erro Tipo II β
Rejeita H_0	Erro Tipo I α	Decisão Correta (1 - β)

Região Crítica

Unilateral à direita:

Ho: $\mu = 50$

H1:: $\mu > 50$

Unilateral à esquerda:

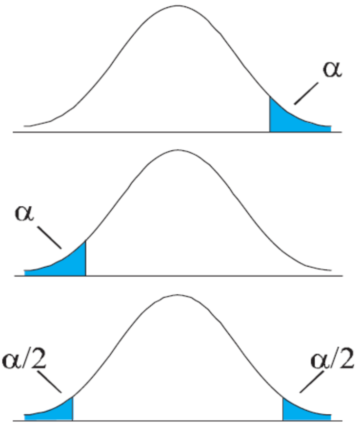
Ho: : $\mu = 50$

H1: : $\mu < 50$

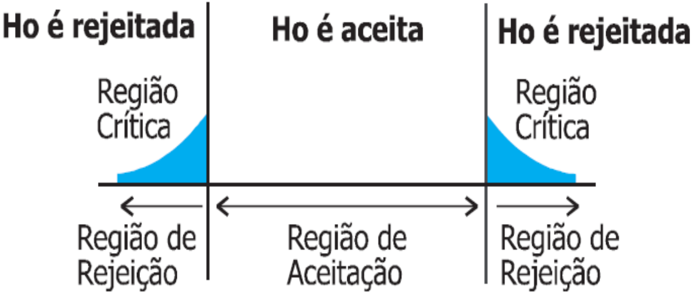
Bilateral:

Ho: : $\mu = 50$

H1:: $\mu \neq 50$



Região Crítica



Procedimento para realização do teste de hipótese

1. Enunciar as hipóteses H_0 e H_a.
2. Fixar o nível de significância α e identificar a estatística de teste.
3. determinar a região crítica e a região de aceitação em função do nível α pelas tabelas estatísticas.
4. Com base na amostra, calcular a estatística do teste.
5. concluir pela rejeição ou não de H_0.

Para uma média

Caso em que X é normalmente distribuída e com variância conhecida.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Usando a variável padronizada Z, temos

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Exemplo 1

De uma população normal com variância 36, toma-se um amostra aleatória de tamanho 16, obtendo-se $\bar{X} = 43$. Ao nível de 10%. Pede-se: testar as seguintes hipóteses,

$$H_0 : \mu = 45 \qquad \text{versus} \qquad H_0 : \mu \neq 45$$

Exemplo 2

Uma instituição de ensino alega que a média de seus alunos em provas de vestibulares de universidades de primeira linha é igual a 7,60. Uma amostra aleatória formada por 60 alunos revelou uma média igual a 6,80. Sabendo que o desvio padrão populacional é igual a 2,40 e assumindo $\alpha = 5\%$, teste a alegação da instituição.

Para duas médias

Usaremos o teste t para duas sitações:

1. Para variâncias homogêneas

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{s^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} \sim t_{(n_x+n_y-2)} g.l$$

onde

$$s^2 = \frac{(n_x - 1)s_x^2 - (n_y - 1)s_y^2}{(n_x + n_y - 2)}$$

Para duas médias

2. Variâncias não homogêneas

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)}} \sim t_{n^*} g.l$$

onde

$$n^* = \frac{\left(\frac{s_{n_x}^2}{n_x} + \frac{s_{n_y}^2}{n_y}\right)^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y - 1}}$$

Exemplo 1

Suponhamos que um pesquisador diante de duas técnicas de memorização X e Y tenha o interesse de compará-las, medindo-se a eficiência pelo tempo exigido para decorar certo tipo de material. O mesmo material foi apresentado a $n_x = 18$ e $n_y = 13$ pessoas que o decoraram através das técnicas X e Y, respectivamente. O pesquisador pretende verificar se há diferença significativa entre as duas técnicas de memorização, para isso adotou $\alpha = 5\%$. Os resultados foram: $\bar{X} = 20min$, $\sigma_x^2 = 20min^2$, $n_x = 18$, $\bar{Y} = 17min$, $\sigma_x^2 = 15min^2$, $n_x = 13$. O pesquisador considerou, por meio do teste F, as variâncias estatisticamente iguais. Pede-se: verifique se a suspeita do pesquisador confirma a diferença entre os métodos de memorização.

Exemplo 2

Deseja-se saber se duas rações alimentares A e B para determinada raça de suíno, são equivalentes. Ou se a ração A é superior a ração B no sentido de causar maior aumento de peso nos suínos. Para 11 animais sorteados ao acaso foi dada a ração A, e a outros 19 a ração B. Os resultados foram: $\bar{A} = 66kg$, $\sigma_A^2 = 17kg^2$, $n_A = 11$, $\bar{B} = 63kg$, $\sigma_B^2 = 16kg^2$, $n_B = 19$. As variâncias foram consideradas homogêneas. A que conclusão podemos chegar se adotarmos um nível de significância $\alpha = 2\%$?