CAPÍTULO 6 - Testes de significância

Introdução

Testes de significância (também conhecidos como Testes de Hipóteses) correspondem a uma regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra.

Obs.: essas hipóteses são, em geral, sobre parâmetros populacionais e a realização do teste se baseia na distribuição amostral dos respectivos estimadores.

Exemplos: Foi discutido em aula: Parâmetro vs estimador.

<u>Hipótese Estatística</u>

É uma suposição quanto ao valor de um parâmetro populacional, ou uma afirmação quanto à natureza da população.

Exemplos: discutido em aula

<u>Hipótese de Nulidade e Hipótese Alternativa</u>

- Hipótese de Nulidade (H_o)
  É a hipótese a ser testada.
- Hipótese Alternativa (H_a)
  É uma hipótese que contraria H_o. É formulada com base no conhecimento prévio do problema, informações de pesquisa, etc.

Ex:

$H_o : \mu$ = 6.000 horas (durabilidade de lâmpadas)

$H_a : \mu > 6000$; ou $H_a : \mu < 6000$; ou $H_a : \mu \neq 6000$

hipóteses unilaterais hipótese bilateral

Após a realização do teste concluimos por uma das hipóteses dadas acima.

Qualquer decisão tomada implica na possibilidade de cometer basicamente dois tipos de erros: Erro tipo I e Erro tipo II.

Obs: P(erro tipo I) = α ou nível de significância do teste.

P(erro tipo II) = β

O quadro abaixo facilita o entendimento.

Decisão	Realidade	
	H_o é verdadeira	H_o é falsa
Rejeitar H_o	α	$1-\beta$
Aceitar H_o	$1-\alpha$	β

então:
$\alpha = \text{P(rej.} H_o \ / \ H_o \text{ é verd.)}$
$\beta = P(\text{aceitar } H_o \ / \ H_o \text{ é falsa})$

Procedimentos para a realização de um Teste de Hipótese

1. Enunciar as hipóteses H_o e H_a;
2. Fixar o nível de significância α e identificar a estatística do teste;
3. Determinar a região crítica (faixa de valores que nos levam à rejeição da hipótese H_o) e a região de aceitação em função do nível α pelas tabelas estatísticas apropriadas;
4. Baseado na amostra, calcular o valor da estatística do teste;
5. Concluir: Se estatística do teste $\in$ região crítica $\Rightarrow$ rej. H_o caso contrário não rej. H_o.

Na INF 162 veremos apenas:

- Teste z
- Teste t
- Teste de Qui-quadrado (teste χ^2)
- Teste F

Teste z

Teste z (veremos apenas o teste para uma média populacional)

Obs: assume-se que a variável em estudo tenha distribuição normal com variância populacional conhecida.

a) <u>Teste z para 1 média</u>

A estatística do teste é baseada na média amostral $\overline{X}$. Pode ser demonstrado que a média amostral tem distribuição aproximadamente normal com média μ e variância σ^2/n, onde n é o tamanho da amostra.

Testamos as hipóteses: $H_o : \mu = \mu_0$ versus H_a : uma alternativa conveniente.

A estatística do teste z para 1 média é: $z_{\text{calc}} = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$

De acordo com o nível de significância e a hipótese alternativa definidas obtemos o valor tabelado de z na tabela apropriada.

A regra de decisão será:

- Se $|z_{\text{calc}}| \geq z_{\text{tab}} \Rightarrow$ rejeitamos H_0;
- caso contrário não rejeitamos H_0.

Exercício:

Uma máquina automática de encher pacotes de café enche-os segundo uma distribuição normal, com média μ e variância 400 g^2. O valor de μ pode ser fixado num mostrador situado numa posição um pouco inacessível dessa máquina. A máquina foi regulada para μ = 500 g. Desejamos, de meia em meia hora, colher uma amostra de 16 pacotes e verificar se a produção está sob controle, isto é, se μ = 500 g ou não. Se uma dessas amostra apresentasse uma média $\bar{x}$ = 492 g, você pararia ou não a produção para verificar se o mostrador está na posição correta? Usar α = 1%.

Problema proposto:

Uma companhia de cigarros anuncia que o índice médio de nicotina dos cigarros que fabrica apresenta-se abaixo de 23 mg por cigarro. Um laboratório realiza 6 análises desse índice, obtendo: 27, 24, 21, 25, 26, 22. Sabe-se que o índice de nicotina se distribui normalmente, com variância igual a 4,86 mg^2. Pode-se considerar a afirmativa do fabricante verdadeira, ao nível de 10% de probabilidade?

Teste t

Teste t (p/ 1 média populacional ou p/ ≠ entre médias populacionais)

Obs: assume-se que a variável em estudo tenha distribuição normal com variância desconhecida.

a) Teste t para 1 média

Exemplo:

Determinada firma desejava comprar cabos tendo recebido do fabricante a informação de que a tensão média de ruptura é 8000 kgf. Para analisar se a afirmação do fabricante é verdadeira, efetuou-se um teste de hipótese unilateral. Se um ensaio com 6

cabos forneceu uma tensão média de ruptura de 7750 kgf, com desvio padrão de 145 kgf , a qual conclusão chegar, usando um nível de significância de 5%?
Resposta:

$$H_o : \mu = 8000 \text{ kgf}$$
$$H_a : \mu < 8000 \text{ kgf}$$

A estatística do teste t para 1 média é: $t_{\text{calc}} = \dfrac{\overline{X} - \mu}{\dfrac{s}{\sqrt{n}}}$

No exemplo teríamos: $t_{\text{calc}} = \dfrac{7750 - 8000}{\dfrac{145}{\sqrt{6}}} = -4{,}22$

A conclusão será:

b) <u>Teste t para duas médias</u> (2 amostras independentes)

Obs: Pressupõe-se normalidade dos dados.
Sejam X e Y normalmente distribuídos com variâncias desconhecidas.
Desejamos testar:

$$H_o : \mu_X = \mu_Y \quad \text{contra}$$
$$H_a : \begin{cases} \mu_X > \mu_Y & \text{ou} \\ \mu_X < \mu_Y & \text{ou} \\ \mu_X \neq \mu_Y \end{cases}$$

Exemplo: dois métodos de execução de determinada tarefa. X e Y seriam os tempos gastos com cada método.

Outra pressuposição (apenas para efeito de nosso curso de INF 162) seria a homogeneidade das variâncias populacionais σ_X^2 e σ_Y^2 (desconhecidas).

Portanto assumimos: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, isso quer dizer que s_X^2 e s_Y^2 são estimativas de um mesmo valor σ^2.

Portanto podemos combinar s_X^2 e s_Y^2 a fim de obter um melhor estimador para σ^2.

Então: $s^2 = \dfrac{SQD_X + SQD_Y}{(n_X - 1) + (n_Y - 1)} = \dfrac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$

Utilizamos para o teste, a variável aleatória:

(ou t_{calc}) $t = \dfrac{\overline{X} - \overline{Y}}{\sqrt{s^2\left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$

que tem distribuição t de Student com $n_X + n_Y - 2$ graus de liberdade.

Decisão:

Se $|t_{\text{calc}}| \geq t_{\text{tab}} \Rightarrow$ rej. H_o

Exercício:

Suponhamos que duas técnicas de memorização X e Y deverão ser comparadas 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 usando as técnicas X e Y respectivamente. Verificar se há diferença significativa entre as duas técnicas de memorização, adotando $\alpha = 5\%$. Os resultados foram:

$\overline{X} = 20$ min $\qquad \overline{\text{Y}} = 17$min

$\text{s}_\text{X}^2 = 12\text{min}^2 \qquad s_Y^2 = 15\text{min}^2$

$n_X = 18 \qquad n_Y = 13$

Resposta: $H_o: \mu_X = \mu_Y$

$H_a: \mu_X \neq \mu_Y$

Teste t para 2 médias; α = 5%

$$s^2 = \frac{(18-1).12 + (13-1).15}{18+13-2} = 13{,}24$$

$$t_{\text{calc}} = \frac{20-17}{\sqrt{13{,}24\left(\dfrac{1}{18} + \dfrac{1}{13}\right)}} = 2{,}27$$

$$t_{\frac{5\%}{2}(29)} = 2{,}045$$

$|t_{\text{calc}}| > t_{\text{tab}} \Rightarrow$ rej. H_o ao nível de significância de 5%.

Teste de Qui-quadrado (χ^2)

Os dois últimos testes a serem apresentados a seguir (Teste F e teste t) são usados para testar hipóteses referentes a um parâmetro populacional ou mesmo à comparação de dois parâmetros. O teste de Qui-quadrado faz parte dos chamados "testes não-paramétricos", ou seja, que não dependem dos parâmetros populacionais, nem de suas respectivas estimativas.

O teste de Qui-quadrado pode ser usado principalmente como:

i) Teste de aderência

ii) Teste de independência

iii) Teste de homogeneidade

Veremos, a princípio, apenas o teste de aderência, sendo os demais testes filosoficamente (e até mesmo "mecanicamente") similares, mas aplicáveis quando queremos estudar a relação entre duas ou mais variáveis de classificação. Se o tempo permitir será apresentado também pelo menos mais um dos outros testes de qui-quadrado.

<u>Teste de Aderência</u>

Existe apenas uma variável e o que se testa é o padrão hipotético de frequências ou a distribuição da variável.

A estatística do teste é dada por: $\chi^2_{calc} = \sum \frac{(O_i - E_i)^2}{E_i}$, onde

$O_i =$ frequência observada da categoria (evento) i ;

$E_i =$ frequência esperada da categoria (evento) i.

Obs: A expressão acima nos dá um valor sempre positivo e tanto menor quanto maior for o acordo entre as frequências observadas e as frequências esperadas, calculadas com base em H_o.

Obs: A hipótese H_o afirmará não haver discrepâncias entre as frequências observadas e as frequências esperadas, ou H_o será colocada em termos de distribuição de probabilidade que vamos por à prova.

O valor de χ^2_{calc} é comparado com o $\chi^2_{tabelado}$.

Se $\chi^2_{calc} \geq \chi^2_{tab} \Rightarrow$ rejeita-se H_o.

Obs: Para obter o χ^2_{tab} precisamos conhecer o nível de significância (α) do teste e o número de graus de liberdade *v*, onde $v = k - 1 - r$, onde *k* é o número de categorias em que foi dividida a amostra; e r é o número de parâmetros estimados para o cálculo das E_i.

Exemplo:

Em 100 lances de uma moeda, observaram-se 65 coroas e 35 caras. Testar a hipótese de a moeda ser honesta, adotando-se $\alpha = 5\%$.

Solução: (passo a passo)

1) H_o : A moeda é honesta (ou, H_o : proporção cara:coroa = 1:1)
   H_a : não H_o

2) α = 5%. Verifica-se que existem 2 categorias (cara e coroa). Então k = 2. Nenhum parâmetro foi calculado, então r = 0. Logo v = 2 – 1 – 0 = 1.

3)

4)

Categorias	Cara	Coroa
freq. observadas	35	65
freq. esperadas	50	50

logo $\chi^2_{\text{calc}} = \dfrac{(35-50)^2}{50} + \dfrac{(65-50)^2}{50} = ... = 9$

5) Como $\chi^2_{\text{calc}} > \chi^2_{\text{tab}}$, rejeita-se H_o, concluindo-se, com risco de 5%, que a moeda não é honesta.

Teste F

Teste F (teste para comparação de variâncias)

Exemplo: Na aplicação de dois métodos A e B, obteve-se os resultados abaixo. Testar a hipótese de igualdade das variâncias ao nível de 5% de probabilidade.

Método	s^2	n
A	40	11
B	16	19

Resposta: $H_o : \sigma_A^2 = \sigma_B^2$

$H_a : \sigma_A^2 > \sigma_B^2$

$F_{5\%(10;\ 18)} = 2{,}41$

$$F_{calc} = \frac{\text{maior } s^2}{\text{menor s}^2} = \frac{s_A^2}{s_B^2} = \frac{40}{16} = 2{,}50$$

Conclusão: $F_{\text{calc}} > F_{\text{tab}} \Rightarrow \text{rej}: H_o$ ao nível de 5% de significância.

$\therefore$ para α = 5% as variâncias não seriam consideradas iguais.

LISTA DE EXERCÍCIOS

INF 161 – Iniciação à Estatística e INF 162 – Estatística I
Testes de Hipóteses

OBS.: Como o teste z para duas médias e o teste de Qui-quadrado para independência não foram apresentados formalmente nesse período, os exercícios de número **9, 11 e 12 não precisam ser resolvidos**. No entanto a existência desses exercícios é importante pois os mesmos poderão ser discutidos em sala de aula pelo professor caso necessário.

1) Sabe-se que o consumo mensal per capita de um determinado produto tem distribuição normal, com desvio padrão 2 kg. A diretoria de uma firma que fabrica esse produto resolveu que retiraria o produto da linha de produção se a média de consumo per capita fosse menor que 8 kg. Caso contrário, continuaria a fabricá-lo. Foi realizada uma pesquisa de mercado, tomando-se uma amostra de 25 indivíduos, e verificou-se que a soma dos valores coletados foi de 180 kg.
a) Utilizando um nível de significância de 5%, e com base na amostra colhida determine a decisão a ser tomada pela diretoria.
b) Utilizando um nível de significância de 1 %, a decisão seria a mesma? (Justifique a sua resposta.)

2) Estamos desconfiados de que a média das receitas municipais per capita das cidades pequenas (0 - 20.000 habitantes) é maior do que a das receitas do estado, que é de 1229 unidades monetárias. Para comprovar ou não esta hipótese, sorteamos dez cidades pequenas, e obtivemos os seguintes resultados: 1230; 582; 576; 2093; 2621; 1045; 1439; 717; 1838; 1359. A que conclusão chegar a um nível de 5% de probabilidade?

3) Uma grande cadeia de magazines está interessada em saber se o valor médio das compras é maior em suas lojas do centro da cidade do que no "Shopping center" de certa localidade. O desvio padrão populacional para ambos os casos é de $10,00. Teste a afirmação de que ambas são iguais, contra a alternativa de que ambas não são iguais, ao nível de 0,01. Uma amostra aleatória das transações nos dois locais deu os seguintes dados:

	Centro	"Shopping center"
média	$45,00	$43,50
Tamanho da amostra	100	100

4) Uma fábrica de embalagens para produtos químicos está estudando dois processos para combater a corrosão de suas latas especiais. Para verificar o efeito dos tratamentos, foram usadas amostras cujos resultados estão no quadro abaixo. Qual seria a conclusão sobre os dois tratamentos, ao nível de 5% de significância ?

Método	Amostra	Média	Desvio padrão
A	15	48	10
B	12	52	15

5) Suponhamos que um pesquisador, desejando colocar à prova a hipótese de que a idade da mãe tem certa influência sobre o nascimento de criança prematura, verificou que, dentre 90 casos de prematuridade, 40 envolviam mães com idade inferior a 18 anos; 15 envolviam mães de 18 a 35 anos e 35 mães com idade acima de 35 anos. Isto leva o pesquisador a manter sua hipótese? Use nível de significância de 0,01.

6) No decurso de um ano, determinada firma teve 50 acidentes. Um dos aspectos de uma investigação levada a efeito pelo engenheiro de segurança diz respeito ao dia de ocorrência do acidente. Pelos dados que seguem abaixo, pode-se dizer que o dia da semana tenha alguma influência? Teste a hipótese nula, de que os dias são igualmente prováveis, ao nível de 10% de probabilidade.

DIA	Segunda	Terça	Quarta	Quinta	Sexta
Nº de acidentes	15	6	4	9	16

7) A associação dos proprietários de indústrias metalúrgicas está muito preocupada com o tempo perdido com acidentes de trabalho, cuja média, nos últimos tempos, tem sido da ordem de 60 horas/homem por ano e desvio padrão de 20 horas/homem. Tentou-se um programa de prevenção de acidentes, após o mesmo, tomou-se uma amostra de 9 indústrias e mediu-se o número médio de horas/homem perdidas por acidente, que foi 50 horas. Você diria, ao nível de 5%, que há evidência de melhoria?

8) Uma firma de produtos farmacêuticos afirma que o tempo médio para certo remédio fazer efeito é de 24 minutos. Numa amostra de 19 casos, o tempo médio foi de 25 minutos, com desvio padrão de 2 minutos. Teste a alegação, contra a alternativa de que o tempo médio é superior a 24 minutos, a um nível de significância de 1%.

9) Uma máquina automática enche latas com base no peso líquido, com variabilidade praticamente constante e independente dos ajustes, dada por um desvio padrão de 5 g. Duas amostras retiradas em dois períodos de trabalho consecutivos, de 10 e de 20 latas, forneceram pesos líquidos médios de, respectivamente, 184,6 e 188,9 gramas. Desconfia-se que a regulagem da máquina quanto ao peso médio fornecido possa ter sido modificada no período entre a coleta das duas amostras. Qual a conclusão ?
a) ao nível de 5% de significância ?
b) ao nível de 1% de significância ?

10) Num estudo comparativo do tempo médio de adaptação, uma amostra aleatória, de 50 homens e 50 mulheres de um grande complexo industrial, produziu os seguintes resultados:

Estatísticas	Homens	Mulheres
Médias	3,2 anos	3,7 anos
Desvios padrões	0,8 anos	0,9 anos

Que conclusões você poderia tirar para a população de homens e mulheres desta indústria, ao nível de 5% de significância ?

11) 125 proprietários de certa marca de automóvel foram entrevistados acerca do desempenho e do consumo de combustível de seus carros. O resultado da pesquisa de opiniões é resumido na seguinte tabela:

	DESEMPENHO		
CONSUMO	PÉSSIMO	REGULAR	BOM
ALTO	29	27	42
BAIXO	4	6	17

Verificar, ao nível de 5% de significância, se devemos considerar que, no consenso geral, desempenho e consumo não guardam relação entre si.

12) Uma pesquisa sobre a qualidade de certo produto foi realizada enviando-se questionários a donas-de-casa através do correio. Suspeitando-se que os respondentes voluntários tenham um particular vício de respostas, fizeram-se mais duas tentativas com os não respondentes. Os resultados estão indicados abaixo. Você acha que existe relação entre a opinião e o número de tentativas? (Utilize o nível de significância de 5 %)

	NÚMERO DE RESPONDENTES (Donas-de-casas)		
OPINIÃO	TENTATIVAS		
	1ª	2ª	3ª
EXCELENTE	62	36	12
SATISFATÓRIO	84	42	14
INSATISFATÓRIO	24	22	24

13) Uma das maneiras de medir o grau de satisfação dos empregados de uma mesma categoria quanto à política salarial é através do desvio padrão de seus salários. A fábrica A diz ser mais coerente na política salarial do que a fábrica B. Para verificar essa afirmação, sorteou-se uma amostra de 10 funcionários não especializados de A, e 15 de B, obtendo-se os desvios padrões s_A = 1,0 SM e s_B = 1,6 SM. Qual seria a sua conclusão, ao nível de 1% ?

RESPOSTAS

1. $z_c = -2{,}00$ $H_o : \mu = 8$ vs $H_a : \mu < 8$

a) $z_{5\%} \cong -1{,}64$, Rejeita-se H_o

b) $z_{1\%} \cong -2{,}33$, Não se rejeita H_o

2. $H_o : \mu = 1229$ vs $H_a : \mu > 1229$

$t_c = 0{,}566$; $t_{5\%}(9) = 1{,}833$; Não se rejeita H_o

3. $H_o : \mu_1 = \mu_2$ vs $H_a : \mu_1 > \mu_2$

$z_c = 1{,}06$; $z_{1\%} = 2{,}33$; Não se rejeita H_o

4. $H_o : \mu_A = \mu_B$ vs $H_a : \mu_A \neq \mu_B$
$|t_c| = 0{,}829$; $t_{2,5\%}(25) = 2{,}060$; Não se rejeita H_o

5. H_o : Proporção = 1:1:1 vs H_a : Proporção $\neq$ 1:1:1
$\chi^2_c = 11{,}667$; $\chi^2_{1\%}(2) = 9{,}210$; Rejeita-se H_o

6. H_o : Proporção = 1:1:1:1:1 vs H_a : Proporção $\neq$ 1:1:1:1:1
$\chi^2_c = 11{,}400$; $\chi^2_{10\%}(4) = 7{,}779$; Rejeita-se H_o

7. $H_o : \mu = 60$ vs $H_a : \mu < 60$
$z_c = -1{,}5$ $z_{5\%} = -1{,}64$, Não se rejeita H_o

8. $H_o : \mu = 24$ vs $H_a : \mu > 24$
$t_c = 2{,}179$; $t_{1\%}(18)=2{,}552$; Não se rejeita H_o

9. $H_o : \mu_1 = \mu_2$ vs $H_a : \mu_1 \neq \mu_2$
$z_c = -2{,}22$
a) $z_{2,5\%} = 1{,}96$; Rejeita-se H_o
b) $z_{0,5\%} \cong 2{,}57$; Não se rejeita H_o

10. $H_o : \mu_M = \mu_H$ vs $H_a : \mu_M \neq \mu_H$
$|t_c| = 2{,}926$; $1{,}98 < t_{2,5\%}(98) < 2{,}00$; Rejeita- se H_o

11. H_o: Desempenho e consumo são independentes vs H_a: "não H_o"
$\chi^2_c = 3{,}791$; $\chi^2_{5\%}(2) = 5{,}991$; Não se rejeita H_o

12. H_o:Opinião e número de tentativas são independentes vs H_a:"não H_o"
$\chi^2_c = 26{,}2$; $\chi^2_{5\%}(4) = 9{,}488$; Rejeita-se H_o

13. a) $H_o : \sigma^2_B = \sigma^2_A$ vs $H_a : \sigma^2_B > \sigma^2_A$
$F_c = 2{,}56$; $F_{1\%}\ (14,9) = 5{,}00$; Não se rejeita H_o