Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipóteses e Decisão Estatística

Prof. Dr. Rogério de Melo Costa Pinto

Curso de Pós-Graduação

Especialização em Estatística Empresarial

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

É uma técnica para fazer inferência. A partir de uma amostra fazemos inferência sobre a população.

Formular hipótese quanto ao valor do parâmetro populacional. Fazer um teste que indicará a aceitação ou rejeição da hipótese

É uma proposição cuja veracidade pode ser colocada em dúvida, ou da qual não se tem total certeza.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

OBJETIVOS:

O objetivo de um teste de hipótese é fornecer ferramentas que nos permita tomar decisões acerca de parâmetros baseados nas estatísticas obtidas das amostras.

Exemplos:

Julgar se uma nova vacina é eficiente na cura de uma doença.

Verificar se um processo de produção é melhor do que outro

Julgar se um determinado índice de tecnologia alcançado por uma indústria é significativo nos lucros da empresa

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Tipos de Hipóteses

- Hipótese de nulidade (H_o) - é a hipótese a ser testada. Ex.: $\mu = 80$ Kg

- Hipótese alternativa (H_1 ou H_a) - é uma hipótese que se contrapõe à hipótese de nulidade

Ex.: Ha : $\mu \neq 80$ (teste bilateral)
Ha : $\mu > 80$ (teste unilateral a direita)
Ha : $\mu < 80$ (teste unilateral a esquerda)

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Região de ACEITAÇÃO de Ho (RA) ou de REJEIÇÃO de Ho (RC)

Região de aceitação (RAHo):

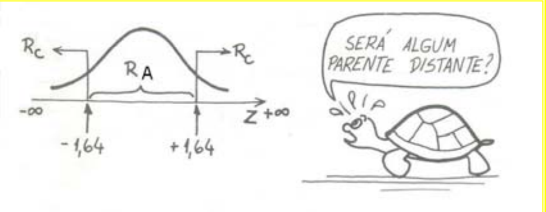
É a região na qual aceitamos a hipótese Ho. Esta região será determinada em função do tipo de teste que será realizado (bilateral ou unilateral)

Região Crítica (RC ou RRHo)

É a região que nos levará a rejeição da hipótese Ho

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Região de ACEITAÇÃO de Ho (R_A) ou de REJEIÇÃO de Ho (R_C)



Teste de Hipótese

S

distribuição desconhecida
e/ou
parâmetros desconhecidos

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese

inferir certas características
da população

amostra

S

distribuição desconhecida
e/ou
parâmetros desconhecidos

$\bar{X} \xrightarrow{\text{estimar}} \mu$

$s^2 \xrightarrow{\text{estimar}} \sigma^2$

$\hat{p} \xrightarrow{\text{estimar}} p$

$\bar{X}_1 - \bar{X}_2 \xrightarrow{\text{estimar}} \mu_1 - \mu_2$

$\vdots$

Intervalo
de
Confiança

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese

S

amostra

$\bar{X} = 107{,}56$

$\mu = 100$

Uma população com média $\mu = 100$ conhecida poderia produzir uma amostra com média $\bar{X} = 107{,}56$?

Hipóteses

$H_0: \mu = 100$ (hipótese nula)

$H_1: \mu \neq 100$ (hipótese alternativa)

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Se H_0 é verdadeira, então

$$z = \frac{\bar{X} - 100}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese

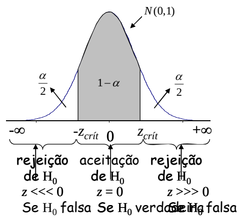
Hipóteses

$H_0: \mu = 100$
$H_1: \mu \neq 100$

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Se H_0 é verdadeira, então

$$z = \frac{\bar{X} - 100}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$



Região Crítica:

- aceito H_0 se $-z_{crít} < z < z_{crít}$ $\rightarrow P(-z_{crít} < z < z_{crít}) = 1 - \alpha$
- rejeito H_0 caso contrário $\rightarrow P(|z| > z_{crít}) = \alpha$

Conclusão (sempre associada a um nível de significância)

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese

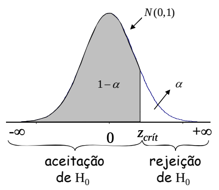
Hipóteses

$H_0: \mu = 100$
$H_1: \mu > 100$ (teste unilateral)

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Se H_0 é verdadeira, então

$$z = \frac{\bar{X} - 100}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$



Região Crítica:

- aceito H_0 se $z < z_{crít}$ $\rightarrow P(z < z_{crít}) = 1 - \alpha$
- rejeito H_0 caso contrário $\rightarrow P(z > z_{crít}) = \alpha$

Conclusão (sempre associada a um nível de significância)

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

As hipóteses nem sempre são verdadeiras; Além disso, o cientista nem sempre decide acertadamente.

A hipótese pode ser / O cientista pode decidir	(V) Verdadeira	(F) Falsa
Rejeitar a hipótese	Aqui ele ERRA	Aqui ele ACERTA
Não Rejeitar a hipótese	Aqui ele ACERTA	Aqui ele ERRA

Teste de Hipóteses e Decisão Estatística

Conceitos

Tipos de Erros

Tipo I - Rejeição de uma hipótese H_0 verdadeira

$Prob_{\text{erro tipo I}} = \alpha$ (nível de significância)

Tipo II - Aceitação de uma hipótese H_0 falsa

$Prob_{\text{erro tipo II}} = \beta$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese

Exemplo: A tensão de ruptura de cabos produzidos por um fabricante apresenta média (μ) de 1800 kg e desvio padrão (σ) de 100 kg. Mediante nova técnica de produção, proclamou-se que a tensão de ruptura pode ter aumentado. Para testar essa declaração, selecionou-se uma amostra de 50 cabos, chegando-se a uma média amostral de 1830 kg. Pode-se confirmar a declaração ao nível de significância de 1%?

$H_0: \mu = 1800$
$H_1: \mu > 1800$

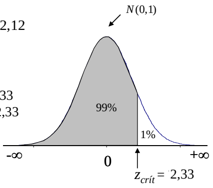
$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Se H_0 é verdadeira, então

$$z = \frac{\bar{X} - 1800}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$z = \frac{1830 - 1800}{\frac{100}{\sqrt{50}}} = 2{,}12$$

aceito H_0 se $z < 2{,}33$
rejeito H_0 se $z > 2{,}33$



Conclusão:
Aceito H_0, ou seja, a nova técnica não melhora significativamente (a 1%) a tensão de ruptura

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese – valor-P (*p-value*)

Toda conclusão de um teste de hipótese está associada a um nível de significância.

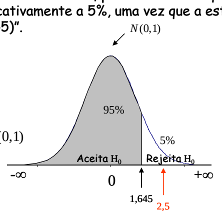
Por exemplo: "Com base num teste z unilateral, pôde-se concluir que as médias μ_1 e μ_2 são diferentes significativamente a 5%, uma vez que a estatística z obtida foi de 2,5 ($z_{\text{crítico}}$ = 1,645)".

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 > 0$

Se H_0 é verdadeira, então

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$z = 2,5$



As médias μ_1 e μ_2 continuariam ser significativamente diferentes caso fosse adotado um nível de significância de 1%?

Teste de Hipótese – valor-P (*p-value*)

Toda conclusão de um teste de hipótese está associada a um nível de significância.

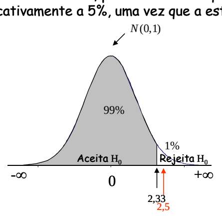
Por exemplo: "Com base num teste z unilateral, pôde-se concluir que as médias μ_1 e μ_2 são diferentes significativamente a 5%, uma vez que a estatística z obtida foi de 2,5".

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 > 0$

Se H_0 é verdadeira, então

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$z = 2,5$



Para que valores de nível de significância, as médias μ_1 e μ_2 poderiam ser consideradas iguais?

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese – valor-P (*p-value*)

Toda conclusão de um teste de hipótese está associada a um nível de significância.

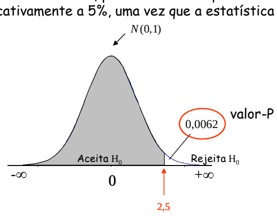
Por exemplo: "Com base num teste z unilateral, pôde-se concluir que as médias μ_1 e μ_2 são diferentes significativamente a 5%, uma vez que a estatística z obtida foi de 2,5".

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 > 0$

Se H_0 é verdadeira, então

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$z = 2,5$



Para que valores de nível de significância, as médias μ_1 e μ_2 poderiam ser consideradas iguais?

Pode-se aceitar H_0 para qualquer nível de significância (α) menor que 0,0062.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese – valor-P (*p-value*)

Exemplo: Foram coletadas amostras (50 pontos) em mapas a fim de avaliar sua exatidão. Procedeu-se o teste z para verificar quais deles possuíam exatidão (p) de 0,90. A tabela abaixo apresenta a exatidão estimada, o resultado do teste (estatística z) e o valor-P de cada mapa.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

	$\hat{p}$	z	valor-P
Mapa 1	0,87	-0,707	0,2397
Mapa 2	0,62	-6,600	2,07e-11
Mapa 3	0,82	-1,886	0,0297
Mapa 4	0,84	-1,414	0,0786

Quais mapas possuem exatidão menor que 0,90, com 5% de significância?

Mapas 2 e 3

Quais mapas possuem exatidão menor que 0,90, com 1% de significância?

Somente Mapa 2

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Significância

Como proceder nos testes:

1) Fazer as hipóteses H_0 e H_1
2) Fixar a probabilidade de erro α
3) Usar as tabelas estatísticas de probabilidade
4) Calcular a variável do teste
5) Aceitar ou rejeitar as hipóteses conforme a região em que a variável teste caiu sobre a curva;
Região de aceitação de Ho (RA) ou de rejeição de Ho (RC)

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes de Médias

a) População infinita, normal ou aproximadamente normal, variância populacional conhecida (amostra grande ou pequena)

Ho	H_1	R. CRITICA
$\mu=\mu_o$	$\mu<\mu_o$ $\mu>\mu_o$ $\mu \neq \mu_o$	$z<-z_\alpha$ $z>z_\alpha$ $z<-z_{\alpha/2}$ e $z>z_{\alpha/2}$

$$Z = \frac{\overline{X} - \mu_o}{\sigma_x / \sqrt{n}}$$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes de Médias

1) Fazer as hipóteses H_0 e H_1

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$ (ou > ou < conforme o caso)

2) Fixar a probabilidade de erro α

α = 5% (cuidado que $\alpha/2$ = 2,5%) bicaudal (ou monocaudal conforme o caso)

3) Usar as tabelas estatísticas de probabilidade

Testes de Médias

4) Calcular a estatística do teste

$$z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

onde:

$\bar{x}$ **= média da amostra**

μ **= valor da hipótese nula (média da população)**

σ **= desvio padrão da população**

n = tamanho da amostra

5) Aceitar ou rejeitar as hipóteses

Se Z_{cal} pertence a RA, não se pode rejeitar H_0

Se Z_{cal} pertence a RC, rejeita-se H_0

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

A Superior Shields, empresa fabricante de pára-brisas para automóveis, estava enfrentando um problema. Ela precisava fabricar um produto que pudesse obter uma avaliação de qualidade de 5.000 pontos (a média da competitividade). Todavia, qualquer melhora nessa pontuação acarretaria um aumento significativo dos custos de produção e acabaria resultando em uma desvantagem competitiva. Com base em experiências anteriores, o fabricante sabia que o desvio-padrão na avaliação da qualidade é de 250. Para verificar se os produtos de sua empresa atingem o padrão de qualidade para produtividade, ele selecionou uma amostra aleatória de 100 clientes industriais e descobriu que a média da avaliação da qualidade nessa amostra era de 4.960. Tendo por base essa amostra, os responsáveis pela Superior Shields queriam saber se seus produtos atingiam o padrão de competitividade, não sendo nem mais baixos, nem mais altos. Eles queriam testar sua hipótese em um nível de significância de 0,05.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

O Sr. James Ginter, chefe de compras de uma grande fábrica de automóveis, quer comprar pára-brisas que tenham avaliação de qualidade de pelo menos 5.000, e não se importa em pagar o preço de um produto de alta qualidade. Todavia, Sr. Ginter está cético quanto à afirmação da Superior Shields de que seus produtos atingem o nível de qualidade de competitividade. Para convencê-lo, a Superior Shields ofereceu-se para pagar uma enquete com uma amostra de 50 consumidores. A avaliação da qualidade na amostra revelou uma média de 4.970 pontos. A questão agora enfrentada pelo Sr. Ginter é aceitar ou não a afirmação da Superior Shields com base nessa média amostral. Além disso, ele quer que a probabilidade de ocorrência de um erro tipo II seja a mais baixa possível; para tal, o Sr. Ginter quer que o teste seja realizado com um nível de significância de 0,01.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes de Médias

b) População infinita, normal ou aproximadamente normal, variância populacional desconhecida e amostra pequena

Ho	H_1	R. CRITICA
$\mu=\mu_o$	$\mu<\mu_o$ $\mu>\mu_o$ $\mu \neq \mu_o$	$t<-t_{\alpha}$ $t>t_{\alpha}$ $t<-t_{\alpha/2}$ e $t>t_{\alpha/2}$

$$t = \frac{\overline{X} - \mu_o}{S_x / \sqrt{n}}$$

GL= n-1

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

Vamos supor agora que a Superior Shields ofereça ao Sr. Ginter uma amostra de apenas 25 consumidores para o exemplo apresentado, e que a média da avaliação da qualidade na amostra tenha sido de 4.962 pontos. Além disso, suponhamos que a Superior Shields não tenha conhecimento prévio do desvio-padrão da população, e que o desvio-padrão da amostra seja de 245. O Sr. Ginter deseja que o teste seja realizado com um nível de significância de 0,01.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes para Diferenças entre Médias

a) População infinita, normal ou aproximadamente normal, variância populacional conhecida (amostras grandes ou pequenas)

Ho	H_1	R. CRITICA
$\mu_1 - \mu_2 = d_O$	$\mu_1-\mu_2<d_o$ $\mu_1-\mu_2>d_o$ $\mu_1 - \mu_2 \neq d_o$	$z<-z_{\alpha}$ $z>z_{\alpha}$ $z<-z_{\alpha/2}$ e $z>z_{\alpha/2}$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}}$$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

Um prefeito quer testar se os salários diários pagos aos empregados do sexo masculino e feminino, pelas grandes organizações de sua cidade, são os mesmos para as mesmas funções. Para testar essa hipótese, uma amostra aleatória de 400 homens e 576 mulheres foi selecionada e registradas as médias salariais. A média e o desvio-padrão para os salários dos homens eram $105,70 e $5,00, respectivamente, enquanto que para as mulheres esses números foram $112,80 e $4,80. Teste a hipótese a um nível de significância de 0,01.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes para Diferenças entre Médias

b) Populações aproximadamente normais, amostras pequenas, variâncias populacionais desconhecidas e estatisticamente iguais

Ho	H_1	R. CRITICA
$\mu_1 - \mu_2 = d_O$	$\mu_1 - \mu_2 < d_o$ $\mu_1 - \mu_2 > d_o$ $\mu_1 - \mu_2 \neq d_o$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ e $t > t_{\alpha/2}$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{1/n_1 + 1/n_2}}$$

$$GL = n_1 + n_2 - 2$$

$$sp^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

As amostras aleatórias seguintes são medidas da capacidade de gerar calor (em milhões de calorias por toneladas) de amostras de carvão de duas minas:

Mina 1: 8.400 8.230 8.380 7.860 7.930

Mina 2: 7.510 7.690 7.720 8.070 7.660

Ao nível de 0,05 de significância, teste se a diferença entre as médias dessas duas amostras é significativa.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes para Diferenças entre Médias

c) **Populações aproximadamente normais, amostras pequenas, variâncias populacionais desconhecidas e estatisticamente desiguais**

Ho	H_1	R. CRITICA
$\mu_1 - \mu_2 = d_O$	$\mu_1-\mu_2<d_o$ $\mu_1-\mu_2>d_o$ $\mu_1 - \mu_2 \neq d_o$	$t<-t_\alpha$ $t>t_\alpha$ $t<-t_{\alpha/2}$ e $t>t_{\alpha/2}$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2 / n_1 + s_2^2 / n_2}}$$

$$v = \frac{(s_1^2 / n1 + s_2^2 / n_2)^2}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}}$$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes para Observações Emparelhadas

d) **Teste para observações emparelhadas ou seja, amostras dependentes (ex: antes, depois)**

Ho	H_1	R. CRITICA
$\mu_D=d_o$	$\mu_D<d_o$ $\mu_D>d_o$ $\mu \neq d_o$	$t < -t_\alpha$ $t > t_\alpha$ $t<-t_{\alpha/2}$ e $t>t_{\alpha/2}$

$$t = \frac{\overline{D} - d_O}{S_D / \sqrt{n}}$$

GL= n-1

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

A academia de ginástica Instant Fit faz uma propaganda que, em média, seus clientes perdem peso nos primeiros 30 dias de freqüência. Um executivo de uma empresa, preocupado com a questão da saúde, quer oferecer a seus funcionários a utilização gratuita da academia como parte do programa de benefícios aos empregados da organização, mas o gerente financeiro está um tanto cético em relação à publicidade da Instant Fit. Em uma tentativa de conquistar o gerente financeiro, foram fornecidos a ele dados sobre o peso "antes e depois" de 10 de seus clientes. O gerente financeiro quer testar essa propaganda em um nível de significância de 0,05.

Antes: 237 135 183 225 147 146 214 157 157 144

Depois: 153 114 181 186 134 166 189 113 188 111

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Testes de Proporções

Ho	H_1	R. CRITICA
$p=p_o$	$p<p_o$ $p>p_o$ $p \neq p_o$	$z<-z_\alpha$ $z>z_\alpha$ $z<-z_{\alpha/2}$ e $z>z_{\alpha/2}$

$$Z = \frac{p - p_o}{\sqrt{p_o q_o / n}}$$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

No departamento de controle de qualidade da empresa Lumiar (fabricante de lâmpadas), o gerente do departamento, com base em sua experiência, garante que 95% das lâmpadas fabricadas pela empresa não apresentam nenhum defeito. O presidente da empresa, um indivíduo preocupado com qualidade, verifica uma amostra aleatória de 225 lâmpadas e descobre que apenas 87% delas não apresentam defeitos. Ele, então, decide testar a hipótese (em um nível de significância de 0,05) de que 95% das lâmpadas fabricadas por sua empresa não apresentam defeitos.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Teste de Hipótese para Diferença Proporções

Ho	H_1	R. CRITICA
$p_1-p_2=p_o$	$p_1-p_2<p_o$ $p_1-p_2>p_o$ $p_1 - p_2 \neq p_o$	$z<-z_\alpha$ $z>z_\alpha$ $z<-z_{\alpha/2}$ e $z>z_{\alpha/2}$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}_1\hat{q}_1 / n_1 + \hat{p}_2\hat{q}_2 / n_2}}$$

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Exemplo

John e Linda, executivos de vendas de uma grande empresa de computadores, são finalistas da competição anual entre os vendedores da empresa. Eles têm desempenhos idênticos, e o vencedor da competição será selecionado com base em seu "índice de conversão" (ou seja, o número de negócios potenciais convertidos em vendas). O gerente de vendas pegou aleatoriamente 100 dos contatos de John e descobriu que 84 deles haviam-se convertido em clientes. No caso de Linda, esse número foi de 82 para 100. O gerente de vendas precisa saber (com α = 0,05) se existe diferença no índice de conversão, com base nas proporções das amostras.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

TESTE DE QUI-QUADRADO (χ^2)

O teste de χ^2 mede a discrepância existente entre freqüências observadas e freqüências esperadas em um conjunto de dados, podendo ser utilizado como teste de aderência e de independência

Teste de aderência

Utilizado para verificar se as diferenças entre as freqüências esperadas e observadas são estatisticamente significativas

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

Passos:

1) Determinar o modelo teórico

2) Calcular as freqüências esperadas (fe) para cada classe da variável aleatória.

3) compara as fe com as observadas (fo) através do seguinte teste:

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(fo_i - fe_i)^2}{fe_i}$$

onde k: número de classes

4) Se $\chi^2_{\alpha_{k-1g.l}} < \chi_c^2$ então o modelo teórico não se ajusta a distribuição observada

Os graus de liberdade são calculados por:

GL = k-1 quando as freqüências esperadas são calculadas sem que se necessite realizar estimativas de parâmetros.

K é o número de classes em que foi dividida a amostra.

Exemplo

Um fabricante de automóveis, ao planejar a produção do modelo do próximo ano, quer definir quantas cores diferentes de carro serão produzidas, e quantas unidades de cada uma delas. Os dados de anos anteriores indicam que as cores de maior aceitação são vermelho, verde, preto e branco, e que, de cada 100 carros, foram vendidos 30 vermelhos, 25 verdes, 25 pretos e 20 brancos. Além disso, dos 2.500 veículos do modelo atual vendidos até agora, 680 eram vermelhos, 520 verdes, 675 pretos e 625 brancos. Com base nesta amostra de 2.500 carros, o gerente de produção concluiu que houve uma mudança substancial nas preferências dos consumidores em relação às cores, e que o modelo do próximo ano não deve seguir a proporção dos outros anos, ou seja, 30:25:25:20. O gerente quer testar sua hipótese em um nível de significância de 0,05.

Teste de independência

Neste caso H_o: variável linha independe da variável coluna

$$fe = \frac{(total\ linha).(total\ coluna)}{total\ geral}$$

$$\lambda_c^2 = \sum_{i=1}^{k} \frac{(fo_i - fe_i)^2}{fe_i}$$

onde k: número de classes

Neste caso os graus de liberdade são dados por:

a) GL = k-1 nas tabelas de simples entrada

b) GL = (h-1) . (k-1) nas tabelas com h linhas e k colunas

Exemplo

A tabela a seguir mostra os resultados de uma enquete com 200 assinantes de temporadas musicais, aos quais foi perguntado com que freqüência costumavam assistir aos concertos em uma cidade vizinha. A freqüência do comparecimento aos espetáculos foi dividida nas categorias nunca, ocasionalmente e frequentemente. Também foi perguntado aos respondentes se eles percebiam o local dos concertos como convenientes ou inconvenientes. Teste a hipótese de que a freqüência de comparecimento aos concertos independe do local, com um nível de significância de 0,05.

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

	Local Conveniente	inconveniente	Total Linha
Freqüente (mais de 6 vezes por temporada)	22	18	40
Ocasionalmente	48	52	100
Nunca	10	50	60
Total coluna	80	120	200

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

CASE

A campanha de propaganda de um refrigerante de marca conhecida teria sofrido alteração se menos de 30% dos consumidores a tivesse aprovado.

a) Formule a hipóteses nula e alternativa.

b) Discuta os erros tipo I e tipo II que podem ocorrer no teste de hipóteses.

c) Que teste estatístico utilizaria? Por que?

d) Pesquisada uma amostra aleatória de 300 consumidores, 84 entrevistados afirmaram ter gostado da campanha. Deve-se modificar a campanha? Por que?

Universidade Federal de Uberlândia
Faculdade de Matemática
Curso de Especialização em Estatística Empresarial

CASE

Foi realizado um estudo para determinar a relação entre utilização de uma biblioteca e a idade dos usuários. Uma amostra de 40 indivíduos foi levantada e gerou-se a tabulação cruzada apresentada a seguir.

a) Complete a Tabela.

b) Interprete a expressão E_1=17,8.

c) Calcule o valor de χ^2.

d) O valor de χ^2 é significante? Em que nível? Qual é exatamente a hipótese nula?

e) Esses dados comprovam que a utilização da biblioteca difere de acordo com a faixa etária. Verdadeiro ou falso? Por que?

		Idade dos Usuários da Biblioteca				
		Menos de 25	25 a 45	Mais de 45	Total linha	Prop.
Utilização	Freqüente	26,2% (21) E_1 = 17,8	19,5% (41) E_4 =	24,5% (27) E_7 =	22,3% (89)	0,223
da	Média	32,5% (26) E_2 =	18,1% (38) E_5 =	31,8% (35) E_8 =		
Biblioteca	Ocasional	41,3% (33) E_3 =	62,4% (131) E_6 =	43,6% (48) E_9 =		
	Total Coluna	100% (80)			100% (400)	1,00
	Proporção	0,2			1,00	

CASE

Um novo produto foi testado em Fresno com um cupom de desconto de 25 centavos, e em Tulsa, com um de 50 centavos. Foi contatada uma amostra de 100 pessoas em cada cidade. Um total de 40% daqueles contatados em Tulsa experimentaram o produto, enquanto que em Fresno apenas 30% o fizeram; uma diferença de 10%. Antes de tomar uma decisão sobre qual cupom utilizar no programa de marketing, foi sugerida uma hipótese.

a) Qual seria a hipótese nula?

b) Qual seria a hipótese alternativa?

c) A probabilidade de obter esse resultado sob a hipótese nula, ou seja, que o nível de experimentação do produto em Tulsa foi 10% mais alto que em Fresno, foi determinada em 0,06. Qual é o valor de p?.

d) Este resultado é significante em um nível de 0,10? E em 0,05? Você rejeitaria a hipótese nula em um nível de 0,10? E em 0,05?

e) A hipótese mostra que haverá mais experimentação do produto com o cupom de desconto de 50 centavos? Você acha que esse cupom deveria ser utilizado?