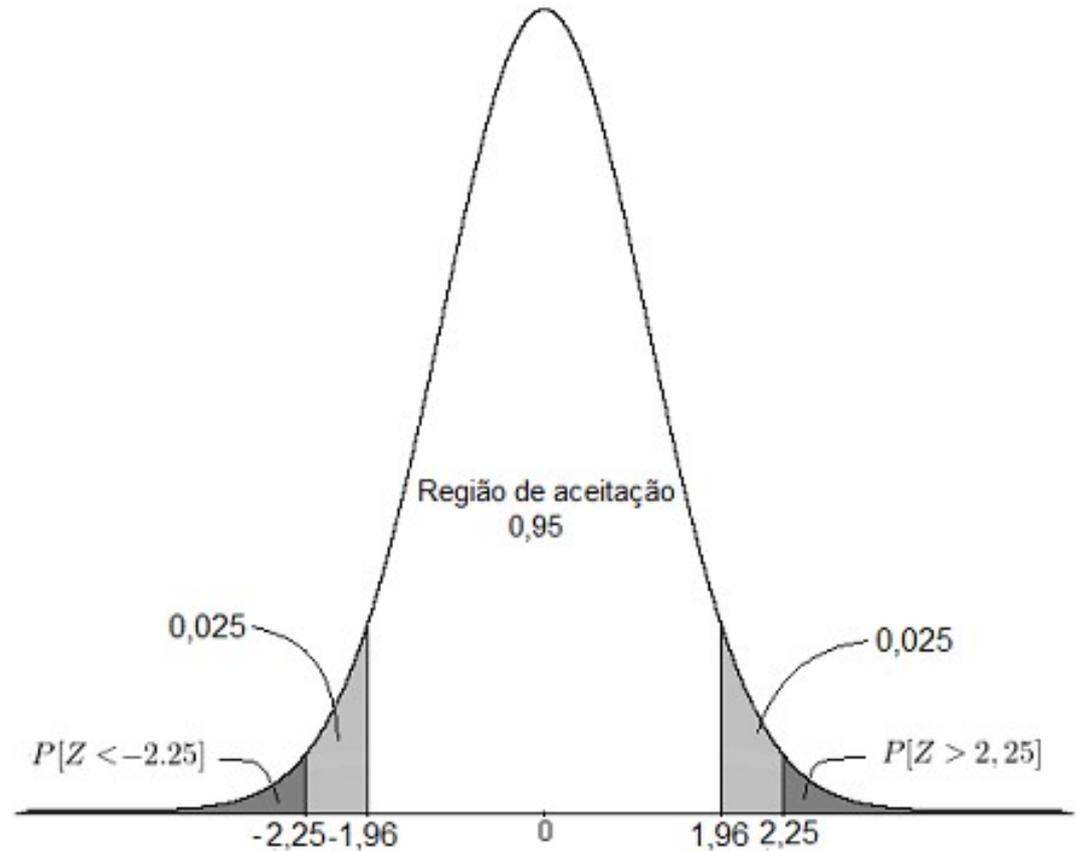


Tópicos em Gestão da Informação II

Aula 08 – Inferência Estatística – Teste de Significância



Prof. Dalton Martins
dmartins@gmail.com

Gestão da Informação
Faculdade de Informação e Comunicação
Universidade Federal de Goiás

Hipótese

- Uma hipótese é uma afirmação sobre uma população;
- Ela geralmente é uma previsão na qual um parâmetro que descreve uma característica de uma variável assume um valor numérico particular ou está em certo intervalo de valores.
- Um teste de significância usa dados para resumir a evidência sobre uma hipótese, comparando as estimativas por pontos dos parâmetros aos valores previstos pela hipótese.

5 etapas de um teste de significância

- Todos os testes têm 5 partes:
 - Suposição
 - Hipóteses
 - Estatística-teste
 - Valor-p
 - Conclusão.

Etapa 1 - Suposições

- Cada teste faz certas suposições ou têm certas condições para ser válido:
 - **Tipo de dados:** cada teste se aplica tanto a dados quantitativos quanto para os dados categóricos;
 - **Aleatorização:** um teste assume que os dados foram obtidos usando uma amostra aleatória;
 - **Distribuição da população:** para alguns testes, assume-se que a variável tenha uma distribuição particular, como a normal;
 - **Tamanho da amostra:** a validade de muitos testes melhora à medida que o tamanho da mostra aumenta.

Etapa 2 - Hipóteses

- Cada teste de significância tem duas hipóteses sobre o valor de um parâmetro:
 - **Hipótese nula** é uma afirmação de que o parâmetro assume um valor em particular;
 - **Hipótese alternativa** declara que o parâmetro está em um intervalo alternativo de valores;
 - Geralmente o valor na hipótese nula corresponde, em certo sentido, a sem efeito;
 - Os valores na hipótese alternativa, então, representam um efeito de certo tipo.

Etapa 3 – Estatística teste

- A estatística teste resume quão longe essa estimativa está do valor suposto do parâmetro em H_0 ;
- Geralmente a diferença é expressa em números de erros padrão que a estimativa dista do valor de H_0 ;

Etapa 4 – Valor p

- Valor-p é a probabilidade de que a estatística teste seja igual ou mais extrema que o valor observado na direção prevista pela H_a ;
- Ele é calculado presumindo que H_0 seja verdadeira.
- Um valor p pequeno (como $p = 0,01$) significa que os dados observados seriam incomuns, se H_0 fosse verdadeira;
- Quanto menor o valor-p, mais forte a evidência contra H_0 .

Etapa 5 – Conclusão

- Se o valor p for suficientemente pequeno, nós rejeitamos H_0 e aceitamos H_a ;
- Muitos estudos requerem valores p bem pequenos, como $p \leq 0,05$ para rejeitar H_0 .
 - Nesse caso, os resultados são ditos serem significativos no nível 0,05.
 - Isso quer dizer que, se H_0 fosse verdadeira, a chance de obter resultados tão extremos como nos dados amostrais não seria maior do que 0,05.

Teste de Significância para uma Média

- Para as variáveis quantitativas, os testes de significância geralmente se referem à média da população μ ;
- Vejamos a seguir as 5 etapas de um teste de significância para uma média....

Etapa 1 - Suposições

- O teste assume que os dados são obtidos por meio de uma amostra aleatória;
- A variável quantitativa é presumida ter uma distribuição populacional normal.

Etapa 2 - Hipóteses

- A hipótese nula para uma média populacional y tem a forma: $H_0: y = y_0$, onde y_0 é um valor particular para a média da população;
 - Essa hipótese geralmente se refere a sem efeito ou sem mudança comparada a um padrão;
- A hipótese alternativa contém os valores alternativos ao parâmetro em H_0 . A hipótese alternativa mais comum é: $H_a: y \neq y_0$, como por exemplo $y \neq 0$.
 - Essa hipótese é chamada de bilateral, porque contém valores tanto acima como abaixo do valor fixado em H_0 .

Etapa 3 – Estatística Teste

- A média amostral estima a média da população;
- Sob a suposição de que $H_0: y = y_0$ é verdadeira, o centro da distribuição amostral será também y_0 .
- Um valor da média amostral que esteja bem longe do centro em uma das caudas fornece forte evidência contra H_0 , por que seria incomum se verdadeiramente $y = y_0$.
- A evidência sobre H_0 é resumida pelo número de erros padrão que a média amostral está distante do valor da hipótese nula y_0 .

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

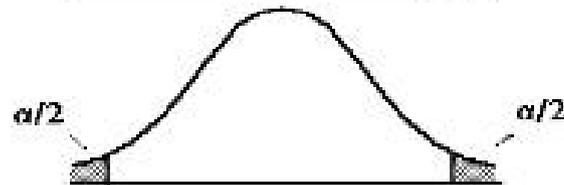
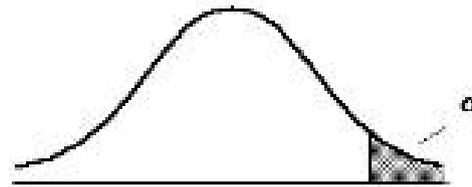
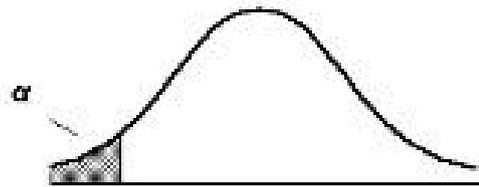
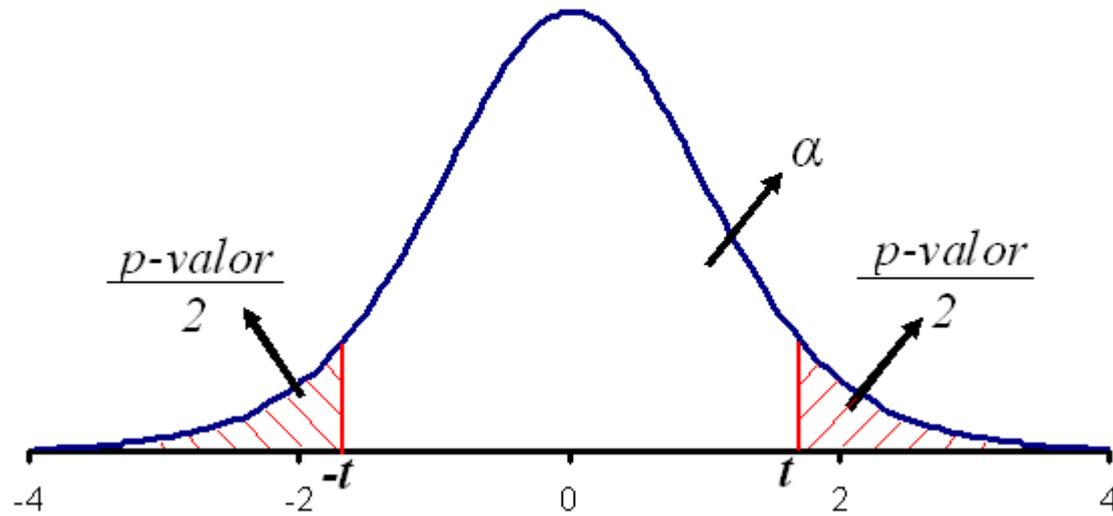
- A estatística teste é o score-t
 - Quanto mais longe a média amostral está de y_0 maior será o valor absoluto da estatística teste t;
 - Quanto maior for o valor de t, mais forte a evidência contra H_0 .

Etapa 4 – Valor-p

- Para $H_a: \mu \neq \mu_0$, os valores t mais extremos são aqueles mais distantes nas caudas da distribuição t;
- O valor p é a soma das probabilidades de duas caudas de que a estatística teste t seja, pelo menos, tão grande em valor absoluto quanto a estatística teste observada
 - Para calcular o valor p precisamos saber o valor t, calculado pela fórmula apresentada no slide anterior;
 - E também precisamos calcular a probabilidade das duas caudas, ou seja, para $\mu > \mu_0$ e $\mu < \mu_0$.

Etapa 4 – Valor p

Função Densidade de Probabilidade
t de Student



Etapa 5 – Conclusão

- Quanto menor for o valor de p , maior a evidência contra H_0 e em favor de H_a .

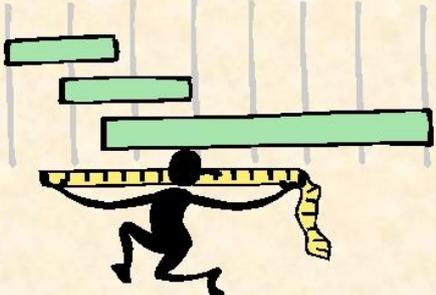
Exemplo

- Em um teste para saber o posicionamento político de uma população, criou-se uma categorização que permitia pontuar esse posicionamento. Um valor em torno de 4,0 significava uma posição moderada. $H_0: y = 4,0$ e $H_a: y \neq 4,0$.
 - Foram feitas 186 observações;
 - A média amostra foi de 4,075 e desvio padrão 1,512.
 - O valor da estatística teste é $t = 0,68$ para $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$
 - A média amostral está a 0,68 erros padrão acima do valor da hipótese nula de uma média igual a 4,0.
 - O valor do grau de liberdade no escore-t é $gl = n - 1 = 185$;
 - Esse n aproxima a distribuição t da normal e podemos obter o valor do escore-z.
 - O escore-z para $t = 0,68$ é 0,25 → soma das duas caudas, teremos $p = 0,50$.
 - Conclusão: o valor p não é pequeno assim ele não contradiz H_0 .

Teste de significância para uma proporção

- Valem as mesmas etapas e suposições do que para uma média, a única diferença é a fórmula de cálculo da estatística teste:

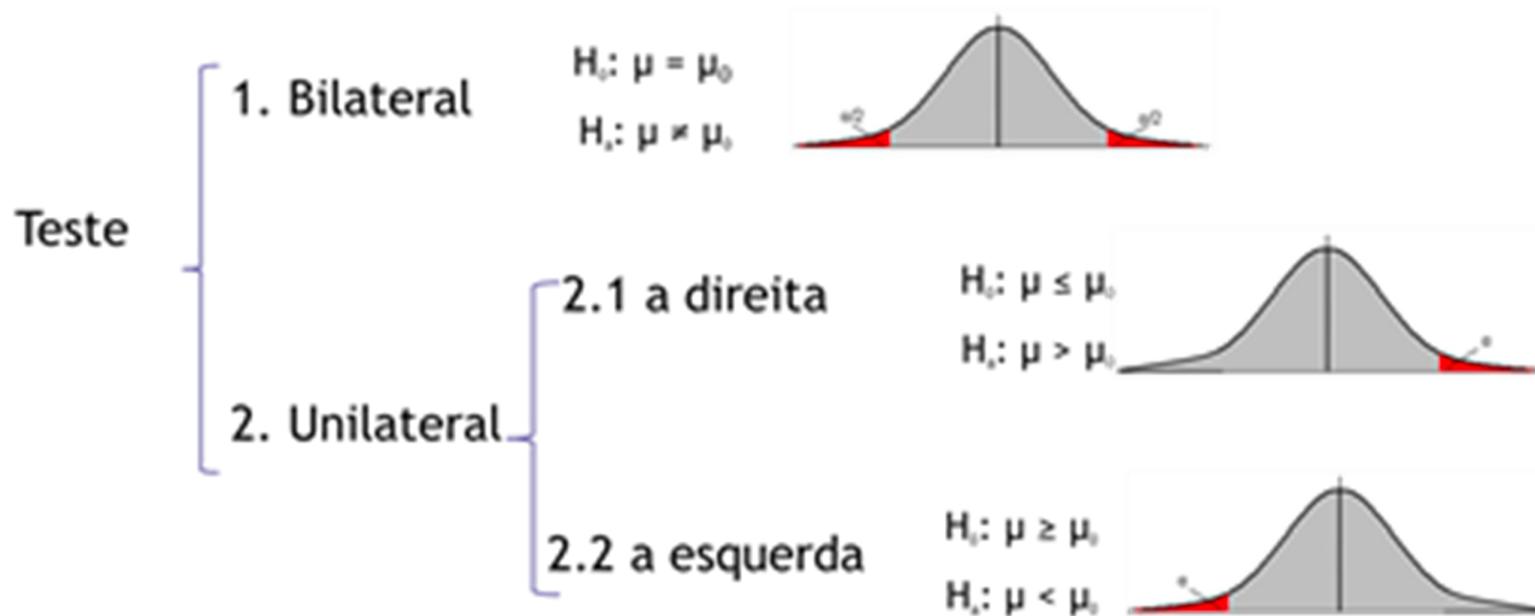
Teste para a Proporção



• Estatística:
$$Z = \frac{\bar{p} - \pi_0}{\sqrt{\frac{\pi_0 (1 - \pi_0)}{n}}}$$

\bar{p} - proporção observada na amostra
 π_0 - proporção considerada na hipótese nula
 n - tamanho da amostra

Testes bilaterais e unilaterais



Exercícios

- Para um teste $H_0: y=0$ contra $H_a: y \neq 0$ com $n=1000$, a estatística teste é igual a 1,04.
 - Encontre o valor p e interprete-o (Obs: você pode usar a normal para aproximar a distribuição t);
 - Suponha que $t = -2,5$ em vez de 1,04. Encontre o valor p. Isto fornece evidência mais forte ou mais fraca contra a hipótese nula. Explique.
- Encontre e interprete o valor p para testar $H_0: y=100$ contra $H_a: y \neq 100$, se uma amostra tem:
 - $N=400$, média amostral 103 e desvio padrão = 40;
 - $N=1600$, média amostral 103 e desvio padrão = 40;
 - Comente o efeito de n nos resultados do teste de significância.
- Segundo a declaração de um sindicato, a renda média para todos os trabalhadores da categoria superior da linha de montagem de uma grande empresa é igual a R\$500,00 por semana. Um representante de um grupo de mulheres da cidade decide analisar se a renda média y para funcionárias se equipara a esse valor. Para uma amostra aleatória de nove mulheres funcionárias, a média amostral foi R\$410,00 e desvio padrão R\$90,00.
 - Teste se a renda média das funcionárias difere de R\$ 500,00 por semana. Inclua as suposições, as hipóteses, a estatística-teste e o valor p. Interprete.
 - Determine o valor p para $H_a: y < 500$. Interprete.
 - Determine o valor p para $H_a: y > 500$. Interprete.

Exercícios

- Para um teste de proporção, temos $H_0: \pi = 0,50$, a estatística teste z é igual a 1,04.
 - Encontre o valor p para $H_a: \pi > 0,50$;
 - Encontre o valor p para $H_a: \pi \neq 0,50$;
 - Encontre o valor p para $H_a: \pi < 0,50$;
 - Os valores p nos itens anteriores fornecem forte evidência contra H_0 ? Explique.
- Para um teste de $H_0: \pi=0,50$, a proporção amostral é de 0,35 baseado em uma amostra de tamanho 100.
 - Mostre que a estatística teste é $z=-3,0$
 - Encontre e interprete o valor p para $H_a: \pi < 0,50$
 - Para um nível de significância de 0,05, qual a sua decisão?
- O casamento entre pessoas do mesmo sexo foi legalizado em todo o Canadá pela Lei de Casamento Civil de 2005. Esta lei é apoiada pela maioria ou minoria canadense? Uma pesquisa de opinião pública realizada por um jornal com 1000 canadenses perguntou se essa lei deveria permanecer ou ser derrubada. As respostas foram 55% para deveria permanecer, 39% para derrubar e 6% não sabiam. Seja π a representação populacional de adultos canadenses que acreditam que ela deveria permanecer. Para testar $H_0: \pi = 0,50$ contra $H_a: \pi \neq 0,50$:
 - Encontre o erro padrão e interprete.
 - Encontre a estatística teste e interprete.
 - Encontre o valor p e interprete.