Estatística para Metrologia

Aula 10

Mônica Barros, D.Sc.

Junho de 2008

Testes de Hipóteses

- Muitos problemas práticos exigem que a gente decida aceitar ou rejeitar alguma afirmação a respeito de um parâmetro de interesse.
- Esta afirmação é chamada de hipótese estatística e o procedimento de tomada de decisão é um teste de hipóteses.
- Muitos problemas reais podem ser formulados naturalmente como testes de hipóteses.
- Existe uma conexão muito próxima entre Intervalos de Confiança e Testes de Hipóteses.

Testes de Hipóteses

- Objetivo geral
  - Inferir sobre os parâmetros desconhecidos de uma população usando uma amostra (de tamanho possivelmente reduzido).
  - Testar hipóteses é um problema que envolve a tomada de uma decisão. Eventualmente, após "recolhermos" (ou processarmos) a informação contida numa amostra, devemos chegar a uma conclusão sobre parâmetros não observáveis relacionados à população que gerou aquela amostra.

Testes de Hipóteses

- Qual o teste ideal?
  - É aquele que sempre toma a decisão correta. É claro que isso é uma abstração, e não existe na realidade.
- Na prática ...
  - Procuraremos limitar a probabilidade de um certo tipo de erro, mas não se pode descartá-lo totalmente.

Testes de Hipóteses

- O Teste de Hipóteses é um procedimento em que procuramos testar uma hipótese inicial contra uma alternativa.
- A primeira hipótese (hipótese inicial) é denominada hipótese nula e representada por H_0.
- A segunda hipótese é chamada hipótese alternativa e representada por H_a ou H_1.
- Em geral a hipótese alternativa representa uma conjectura nova a ser testada, e a hipótese nula representa a situação usual, o "status quo".

Testes de Hipóteses

- A partir dos dados observados, como podemos decidir sobre qual hipótese (nula ou alternativa) deverá ser rejeitada?
- A rejeição da hipótese nula implica na aceitação da hipótese alternativa e vice-versa.
- Não é possível aceitar (ou rejeitar) ambas as hipóteses simultaneamente.

Testes de Hipóteses

- O que é um teste de hipóteses?
  - É qualquer regra usada para nos levar à decisão sobre qual hipótese devemos aceitar.
  - Podemos criar um número infinito de testes de hipóteses, o problema é identificar quais são os bons testes, e tentar obter um "algoritmo" para criar bons testes em diversas situações.
- Aqui estaremos concentrados em obter testes de hipóteses para a média de distribuições.

Construção de um Teste de Hipóteses

- Teste
  - Rejeitar H_0 se T(x), uma função apropriada dos X_i's da amostra, está numa região especificada R.
  - Do contrário, se T(x) *não está* na região R, *não rejeitamos* a hipótese nula.
  - A região R é chamada de região de rejeição ou região crítica.

Erros do Tipo I e II

- A partir do que foi observado na amostra podemos tomar a decisão de aceitar ou rejeitar H_0 e esta decisão não é necessariamente correta, como mostra a tabela a seguir.

Decisão tomada → Estado da realidade ↓	Aceitar H_0 (Rejeitar H_1)	Rejeitar H_0 (Aceitar H_1)
H_0 é verdadeira (H_1 é falsa)	DECISÃO CORRETA	Erro do tipo I (α)
H_1 é verdadeira (H_0 é falsa)	Erro do tipo II (β)	DECISÃO CORRETA

Erros do Tipo I e II

- A eficiência do teste pode ser medida através das probabilidades dos erros de tipo I e II.
- Idealmente gostaríamos que a probabilidade de incorrermos em qualquer tipo de erro fosse zero, mas isto não é possível .
- Para um tamanho de amostra fixo também não é possível fixarmos ambos os erros de tipo I e II.

Erros do Tipo I e II

- α = Probabilidade de erro do tipo I
  - α = Pr{ rejeitar H_0 | H_0 é verdadeira }
  - α = Pr{ T(x) na região crítica | H_0 é verdadeira }
  - α é chamado de tamanho do teste ou nível de significância do teste.
- β = Probabilidade de erro do tipo II
  - β = Pr{ aceitar H_0 | H_0 é falsa }
  - β = Pr{ T(x) fora da região crítica| H_0 é falsa }

Potência de um Teste

- Potência do teste (ou poder do teste)
  - 1- β = 1- Probabilidade de erro do tipo II
  - 1- β = Pr{ rejeitar H_0 | H_0 é falsa }
- Ou seja, a potência do teste é a probabilidade de uma decisão correta!
- Idealmente, a potência de um teste seria sempre alta, mas isso não é sempre verdade.

Função Potência

- Define-se a função potência como:
- K(θ) = Pr{ rejeitar H_0 | o valor do parâmetro é θ}

- O que é uma "boa" função potência?
- Se θ está na região da hipótese nula, a função potência deve ser pequena (pois não queremos rejeitar H_0 quando ela é verdadeira). Ao contrário, se θ estiver na região onde a hipótese alternativa é válida, gostaríamos que a função potência fosse alta.

Função Característica de Operação (OCC)

- É definida como:
- J(θ) = 1 – K(θ) = Pr{ aceitar H_0 | o valor do parâmetro é θ }

- Note que, ambas K(θ) e J(θ) são probabilidades, e portanto limitadas ao intervalo [0,1].

- A OCC é muito utilizada em Controle de Qualidade, mas não falaremos mais dela aqui neste curso.

Testes de Hipóteses - intuição

- Suponha que temos uma amostra de tamanho 25 de uma Normal com variância conhecida 100 e desejamos testar as seguintes hipóteses:

$$H_0 : \mu = 2$$
$$H_1 : \mu > 2$$

- O que a nossa intuição nos diz? A média amostral, $\overline{X}$, é um bom estimador de μ, e portanto deve trazer evidência sobre qual hipótese (H_0 ou H_1) é verdadeira. Imagine que observamos $\overline{X}$ = 50000. Dados os parâmetros (n = 25 e variância 100), este parece um número bem exagerado, e então H_0 deve ser falsa. Logo, a nossa intuição parece apontar para a seguinte regra de decisão:

Testes de Hipóteses - intuição

- Devemos rejeitar H_0 se $\overline{X}$ é grande.
- Ou seja, a região crítica tem a forma:

$$R = \{\overline{X} \geq k\}$$

- A questão que surge agora é: como escolher a constante k? Uma possibilidade é arbitrar o máximo erro do tipo I, ou seja, a maior probabilidade de rejeitar H_0 quando H_0 é verdadeira.

Testes de Hipóteses - intuição

- Mas, esta probabilidade pode ser escrita em termos da função potência. Suponha que FIXAMOS α, a probabilidade do erro do tipo I, isto é:

$$\alpha = \Pr\{\text{Rejeitar } H_0 | H_0 \text{ é Verdadeiro}\} =$$

$$= \Pr\{\overline{X} \geq k | \mu = 2\} = \Pr\left\{\frac{\overline{X}-2}{\sqrt{100/25}} \geq \frac{k-2}{\sqrt{100/25}}\right\} =$$

$$= 1 - \Phi\left(\frac{k-2}{2}\right)$$

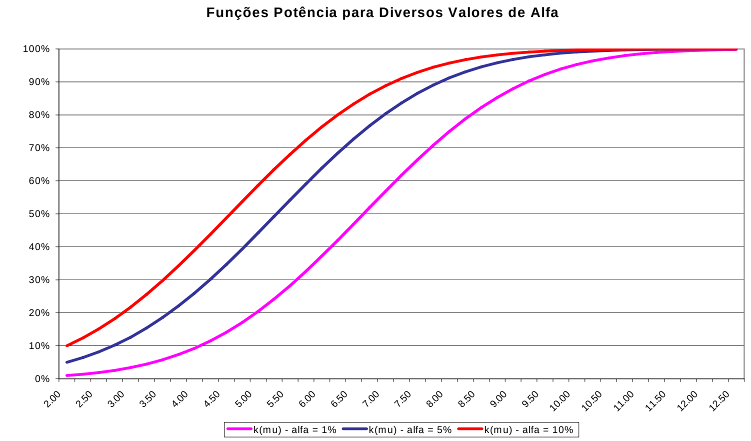
- Por exemplo:

α	k	z (da N(0,1))
1%	6.65	2.33
5%	5.29	1.64
10%	4.56	1.28

- k na tabela ao lado é o valor a partir do qual rejeita-se a hipótese nula

Testes de Hipóteses - intuição

- Vamos ver a função potência em cada um dos casos anteriores



Testes de Hipóteses - intuição

- Conclusões
- Se α é muito pequeno (erro do tipo I muito pequeno, p.ex., 1%), a região de rejeição "exige" um valor de k muito grande para rejeitar a hipótese nula neste caso, e a função potência "demora muito" a crescer.
- À medida que passamos a aceitar erros do tipo I maiores (por ex, 5% ou 10%, a função potência já começa a rejeitar a hipótese nula "com mais facilidade", pois o valor de k diminui.

Testes de Hipóteses - intuição

- Conclusões
- A título de exemplo, se o valor de μ fosse 5 (e a hipótese nula decididamente falsa!), as probabilidades de rejeição usando as funções potência do exemplo seriam 20.4%,44.8% e 58.7% respectivamente.

Testes de Hipóteses uni-caudais

- E agora, o que acontece se estendemos o nosso teste de hipótese para:

$$H_0 : \mu \le 2$$
$$H_1 : \mu > 2$$

- A resposta é: basicamente nada! Por que?
- Considere a função potência para valores de μ dentro da hipótese nula – estes valores estarão abaixo do α especificado.
- O próximo gráfico mostra esta idéia para a função potência com k = 5.29 (isto é, α = 5%)

Testes de Hipóteses uni-caudais

Testes de Hipóteses uni-caudais

- Generalização
- Suponha que desejamos testar as seguintes hipóteses:

$$H_0 : \mu \le \mu_0$$
$$H_1 : \mu > \mu_0$$

- O teste tem exatamente a forma descrita antes, em que a região de rejeição é:

$$R = \left\{ \overline{X} \ge k \right\}$$

Testes de Hipóteses uni-caudais

- Como escolher k?
- Através do erro do tipo I (α), previamente especificado, que leva às seguintes "receitas de bolo":
- Rejeitar H_0 se:

$$\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha \Leftrightarrow \overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

σ conhecido

$$\frac{\overline{X} - \mu_0}{s / \sqrt{n}} > z_\alpha \Leftrightarrow \overline{X} > \mu_0 + z_\alpha \frac{s}{\sqrt{n}}$$

σ desconhecido e usando a hipótese de uma amostra grande, o que possibilita o uso da Normal

Testes de Hipóteses uni-caudais

- O **nível de significância** de um teste (α) é definido como a maior probabilidade de rejeição de H_0 quando H_0 é verdadeira.
- Ou seja, **o nível de significância é o maior erro do tipo I cometido pelo teste**.
- No exemplo anterior, α é apenas o valor da função potência em $\mu = 2$.

Testes de Hipóteses uni-caudais

- Suponha que agora desejamos testar as seguintes hipóteses:

$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

- Pelos mesmos argumentos que no teste anterior, faz sentido rejeitar a hipótese nula quando a média amostral for "pequena".
- O que é um valor "pequeno"? Vai depender do nível de significância especificado para o teste, ou seja, do erro máximo do tipo I.

Testes de Hipóteses uni-caudais

- "Receita de Bolo"
  - Rejeitar H_0 se:

$$\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} < -z_\alpha \Leftrightarrow \overline{X} < \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \quad \sigma \text{ conhecido}$$

$$\frac{\overline{X} - \mu_0}{s / \sqrt{n}} < -z_\alpha \Leftrightarrow \overline{X} < \mu_0 - z_\alpha \cdot \frac{s}{\sqrt{n}} \quad \sigma \text{ desconhecido}$$

- Estes testes são válidos para **amostras Normais com variância conhecida** ou para amostras não necessariamente Normais de tamanho GRANDE e σ desconhecido. Note que estamos usando z_α, que é um ponto obtido da tabela N(0,1).

Testes de Hipóteses uni-caudais

- Exemplo
- Uma empresa produz café em pó em embalagens de 1 kg. O gerente de produção deseja saber se as embalagens realmente possuem em média 1 kg do produto e decidiu realizar um teste.
- Ele retirou uma amostra de 50 embalagens e obteve uma um peso médio de 0,985 kg de produto.
- Informações anteriores a respeito da quantidade de produto por embalagem indicaram um desvio-padrão de 0,075 kg. O gerente deseja saber, com um nível de significância de 1% se o conteúdo de cada embalagem é de, no mínimo, 1 kg.

Teste de Hipóteses uni-caudais

- **Solução: As hipóteses nula e alternativa para o teste são:**

$$H_0 : \mu \geq 1$$
$$H_a : \mu < 1$$

- **Para α = 1%, o valor de z_α é (a partir da tabela normal) de 2.33. A região de rejeição é:**

$$\text{Rejeitar } H_0 \text{ se } \overline{X} \leq \mu_0 - z_\alpha . \frac{\sigma}{\sqrt{n}} \Leftrightarrow \overline{X} \leq 1\text{-}2.33\frac{0.075}{\sqrt{50}} = 0.975$$

- **Como a média amostral (0,985) não é menor que 0.975, a hipótese nula não pode ser rejeitada.**
- **Se σ não fosse conhecido, deveríamos utilizar o desvio-padrão da amostra s.**

Teste de Hipóteses uni-caudais

- **A região crítica do teste anterior é:**

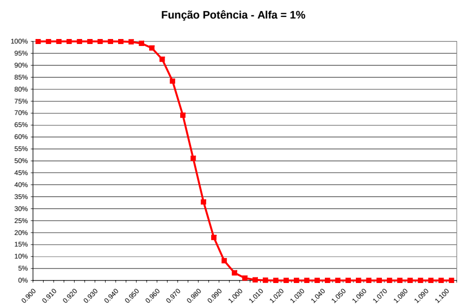
$$\text{Rejeitar } H_0 \text{ se } \overline{X} \leq 0.975$$

- **A função potência deste teste é então:**

$$K(\mu) = \Pr(\text{Rejeitar H}_0 | \text{a média é } \mu) =$$
$$= \Pr(\overline{X} < 0.975 | \mu) = \Pr\left(\sqrt{50}\left(\frac{\overline{X}-\mu}{0.075}\right) < \sqrt{50}\left(\frac{0.975-\mu}{0.075}\right)\right) =$$
$$= \Pr\left(Z < \sqrt{50}\left(\frac{0.975-\mu}{0.075}\right)\right) = \Phi\left(\sqrt{50}\left(\frac{0.975-\mu}{0.075}\right)\right) = \Phi\left(\sqrt{50}\left(\frac{0.975-\mu}{0.075}\right)\right) =$$
$$= \Phi(94.28(0.975-\mu))$$

Teste de Hipóteses uni-caudais

- **Pela magnitude dos números envolvidos (tamanho da amostra grande e desvio padrão pequeno) é intuitivo perceber que qualquer pequena variação na média amostral levará a grandes oscilações da função potência, o que pode ser confirmado no próximo gráfico.**



Valor de p ("p value")

- **Muitos softwares estatísticos calculam e exibem o "p-value", que é a probabilidade de que a estatística de teste tenha valor pelo menos tão extremo (muito grande ou muito pequeno) quanto o valor encontrado na amostra.**

- **O "valor-p" (p-value) indica o menor nível de significância que levaria à rejeição da hipótese nula.**

- **Por exemplo, se o p-value é 0.04, a hipótese H_0 seria rejeitada com nível 5%, mas não com nível 1%.**

Testes de Hipóteses uni-caudais

- Uma outra forma para realizar o teste de hipótese é através do "p value".

- A hipótese nula é rejeitada se essa probabilidade ("p value") for menor que o nível de significância definido para o teste

$$\text{Rejeitar } H_0 \text{ se "p value"} < \alpha$$

Teste de hipótese bi-caudal

- Agora vamos desenvolver um teste de hipótese bi-caudal, para uma amostra grande ($n \geq 30$) e σ da população conhecido

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

- **O nível de significância α é tal que, se a hipótese nula for falsa, queremos ter uma probabilidade máxima $\underline{\alpha}$ de aceitá-la, isto é, queremos cometer uma probabilidade especificada de erro do tipo I.**

Teste de hipótese bi-caudal

- Mas, intuitivamente, qual a "cara" da região crítica? Devemos rejeitar H_0 quando estivermos "longe" de μ_0, ou seja, quando o módulo da média amostral estiver muito distante de μ_0.

- Suponha inicialmente que a hipótese nula seja verdadeira. Para uma amostra grande, podemos considerar a distribuição da média amostral como praticamente Normal (pelo teorema central do limite).

Teste de hipótese bi-caudal

- Agora, dado um nível de significância α, devemos considerar dois valores de Z
  - Um, abaixo do qual há uma probabilidade $\alpha/2$ da média de uma amostra estar localizada ($-z_{\alpha/2}$)
  - Outro, acima do qual há uma probabilidade $\alpha/2$ da média de uma amostra estar localizada ($z_{\alpha/2}$)
- A regra de rejeição é:

$$\text{Rejeitar } H_0 \text{ se } Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} < -z_{\alpha/2} \text{ ou } Z > z_{\alpha/2}$$

Teste de hipótese bi-caudal

- Ou seja, em termos da média amostral, a região crítica pode ser descrita como:

$$\text{Rejeitar } H_0 \text{ se } \overline{X} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ ou se } \overline{X} > \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Isto é, **rejeita-se a hipótese nula se a média amostral estiver "longe" de μ_0.**
- Note que, analogamente ao teste uni-caudal, α é o nível de significância do teste, isto é, o maior erro do tipo I.

Teste de hipótese bi-caudal

- Mas, como aqui rejeita-se a hipótese nula dos dois lados, o ponto usado da Normal é $z_{\alpha/2}$ e não z_{α} (que era usado nos testes uni- caudais), de tal forma que $\Pr(Z > z_{\alpha/2}) = \alpha/2$.
- "Receita de Bolo" – pontos percentuais da distribuição N(0,1) para testes bi-caudais

Testes Bi-caudais	
α	z
1%	2.576
5%	1.960
10%	1.645

Teste de hipótese bi-caudal

- **Exemplo**
- Um fabricante de autopeças utiliza esferas de aço na fabricação de rolamentos. Essas esferas devem ter um diâmetro de 12 mm, caso contrário os rolamentos não atingem as especificações exigidas.
- Uma amostra de 30 rolamentos escolhidos ao acaso forneceu um diâmetro médio de 11,45 mm e um desvio-padrão de 1 mm.
- Pode-se dizer que o diâmetro médio dos rolamentos utilizados é igual a 12 mm com um nível de significância de 5%?

Teste de hipótese bi-caudal

- Solução: este é um teste de hipótese bi-caudal, com α = 0.05, onde:

$$H_0 : \mu = 12$$
$$H_a : \mu \neq 12$$

- Para α = 0,05, $z_{\alpha/2}$ = 1,96
- Para $\overline{X}$ = 11,45mm, temos: $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{11.45 - 12}{1/\sqrt{30}} = -3.01$

E portanto podemos rejeitar a hipótese nula. Note que rejeitar a hipótese nula para Z < - 1,96 é completamente equivalente à rejeitá-la para:

$$\overline{X} < 12 - 1.96\frac{1}{\sqrt{30}}$$

Teste de hipótese bi-caudal

- A região crítica neste exemplo é:

$$\text{Rejeitar } H_0 \text{ se } \bar{X} < 12 - 1.96\frac{1}{\sqrt{30}} \text{ ou se } \bar{X} > 12 + 1.96\frac{1}{\sqrt{30}}$$

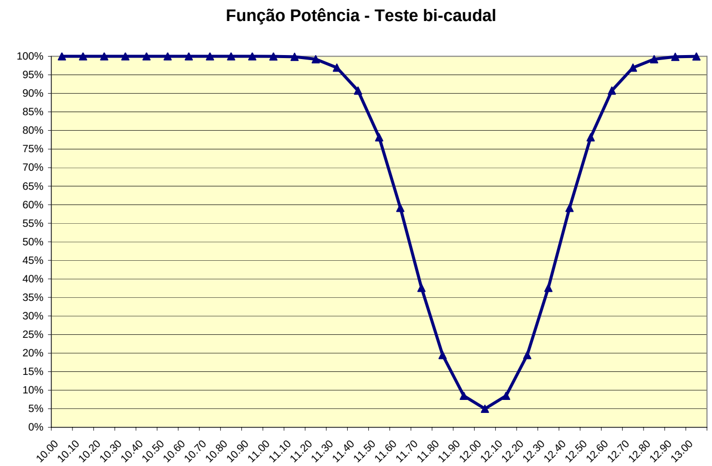
$$\text{Isto é, rejeitar } H_0 \text{ se : } \bar{X} < 11.64 \text{ ou } \bar{X} > 12.36$$

- A função potência é, neste caso (verifique!):

$$K(\mu) = \Phi\left(\frac{\sqrt{30}}{1}(12-\mu) - 1.96\right) + 1 - \Phi\left(\frac{\sqrt{30}}{1}(12-\mu) + 1.96\right)$$

- O gráfico desta função potência é mostrado na próxima página. Note que a potência (probabilidade de rejeitar a hipótese nula) cresce à medida que nos afastamos de μ = 12 e, em μ = 12, o valor da função potência é exatamente o erro do tipo I, estipulado em 5%.

Teste de hipótese bi-caudal



Teste de hipótese – amostra pequena

- Até o momento, consideramos o caso de uma amostra grande ($n \geq 30$)

- Para $n < 30$ existem as seguintes possibilidades:
  - A população é Normal e σ é conhecido: utilizamos o mesmo procedimento que para o caso de $n \geq 30$, com σ conhecido (use a distribuição Normal)
  - A população é normalmente distribuída e σ não é conhecido: utilizamos o mesmo procedimento que para o caso de $n \geq 30$, utilizando s como estimador de σ e a distribuição t ao invés da Normal
  - A população não é normalmente distribuída: aumentamos o tamanho da amostra pois não é possível usar uma aproximação Normal.

Teste de hipótese – amostra pequena

- Exemplo
- Uma revista decidiu realizar uma pesquisa sobre a qualidade de serviço em grandes aeroportos ao redor do mundo.
- O nível de serviço de um aeroporto é considerado superior se a nota obtida é igual ou superior a 7. Para o aeroporto de Heathrow, em Londres, foram entrevistadas 12 pessoas que atribuíram as seguintes notas: 7, 8, 10, 8, 6, 9, 6, 7, 7, 8, 9, e 8.
- Determine, com um nível de significância de 5%, se o serviço do aeroporto de Heathrow pode ser considerado superior. Suponha que a população é normalmente distribuída.

Teste de hipótese – amostra pequena

- **Solução: As hipóteses nula e alternativa para o teste são:**

$$H_0 : \mu < 7$$
$$H_a : \mu \geq 7$$

- **Com uma população normal, n < 30 e σ desconhecido, utilizaremos s e a distribuição t com 11 graus de liberdade para o teste.**
- **A média da amostra é 7,75 e s = 1,215**
- **O valor de t_α para o teste é 1,796**
- **A regra de rejeição é:**

$$\text{Rejeita } H_0 \text{ se } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_\alpha \Rightarrow t = \frac{7{,}75 - 7}{1{,}215/\sqrt{12}} = 2{,}14 > t_\alpha = 1{,}796$$

- **Logo, existe evidência para rejeitar a hipótese nula!**

Teste de hipótese – amostra pequena

- **Exemplo**
- **Historicamente, a comissão média paga por pessoas físicas para operações em bolsa de valores através de internet numa corretora é R$15.**
- **Neste mês você fez uma pesquisa com 16 clientes da corretora e notou que a comissão média foi de R$ 10 e o desvio padrão R$ 6. Com nível de significância 10%, há evidência para afirmar que a comissão neste mês foi mais baixa que historicamente? E com nível de significância 5%? Suponha que os valores pagos são Normalmente distribuídos.**

Teste de hipótese – amostra pequena

- **Desejamos testar as hipóteses:**

$$H_0 : \mu = 15$$
$$H_1 : \mu < 15$$

- **A região crítica tem a forma: rejeitar H_0 se a média amostral é pequena, e como o tamanho da amostra é pequeno devemos usar a distribuição t de Student.**

$$\text{Rejeitar } H_0 \text{ se } \bar{X} \leq \mu_0 - t_{n-1,\alpha} \cdot \frac{s}{\sqrt{n}}$$

Ou seja :

$$\text{Rejeitar } H_0 \text{ se } \bar{X} \leq 15 - t_{n-1,\alpha} \cdot \frac{6}{\sqrt{16}}$$

$$\text{Rejeitar } H_0 \text{ se } \bar{X} \leq 15 - t_{15,\alpha}(1.5)$$

Teste de hipótese – amostra pequena

- **Do Excel:**
  - **Para o teste com nível 5% usamos $t_{15,0.05}$ = INVT(0.10,15) = 1.753**
  - **Para o teste com nível 1% usamos $t_{15,0.01}$ = INVT(0.02,15) = 2.602**
  - **Região crítica a 5%:**

$$\text{Rejeitar } H_0 \text{ se } \bar{X} \leq \mu_0 - t_{n-1,\alpha} \cdot \frac{s}{\sqrt{n}}$$

Ou seja :

$$\text{Rejeitar } H_0 \text{ se } \bar{X} \leq 15 - t_{n-1,\alpha} \cdot \frac{6}{\sqrt{16}}$$

$$\text{Rejeitar } H_0 \text{ se } \bar{X} \leq 15 - t_{15,\alpha}(1.5)$$