

# Teste de Hipóteses para média populacional com variância populacional desconhecida

# Objetivos de Aprendizagem

Os alunos devem ser capazes de:

- ✓ Estender a metodologia de teste de hipóteses que aborda média populacional, mas agora com  $\sigma^2$  desconhecido;
- ✓ Buscar estatística de teste adequada e usá-la para tomada de decisão via Região Crítica e via valor-p.

**Acompanhe, previamente, o PLANO DE AULA  
no BLACKBOARD!**

# Exemplo 1

O número médio de pontos em um exame de inglês tem sido historicamente igual a 80.

Foram sorteados 10 estudantes que fizeram recentemente esse exame e observadas as notas:

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 65 | 70 | 76 | 86 | 59 | 81 | 75 | 72 | 81 | 83 |
|----|----|----|----|----|----|----|----|----|----|

Especialistas desconfiam que o rendimento médio dos alunos diminuiu e desejam testar essa afirmação por meio de um teste de hipóteses, com nível de significância de 5%.

Fazendo as suposições necessárias, qual seria a conclusão do teste?

|             |       |
|-------------|-------|
| $\bar{x} =$ | 74,80 |
| $s =$       | 8,48  |

# Passos para Construção de um Teste de Hipóteses (via RC)

**1º.Passo:** Fixe qual a hipótese nula,  $H_0$ , a ser testada e qual a hipótese alternativa ( $H_A$ ).

**2º.Passo:** Defina a *estatística de teste* sob  $H_0$ . Não se esqueça de levantar as propriedades dessa estatística.

❑ Uma **estatística** é qualquer função da amostra que não depende de parâmetros desconhecidos.

# Estadística de Teste

## Caso 2

(variância populacional desconhecida)

## Relembrando...

Vamos considerar a seguinte hipótese nula:

$$H_0: \mu = \mu_0$$

Vimos que um estimador com boas propriedades para o parâmetro  $\mu$  é  $\bar{X}$ .

Sob algumas suposições, temos que

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

## Relembrando...

Sob a hipótese nula, vem que

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

Porém, a quantidade anterior não pode ser usada como *estatística de teste* pois  $\sigma$  é parâmetro desconhecido.

Ainda,

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0; 1)$$

também não pode ser considerada uma *estatística de teste*, uma vez que  $\sigma$  é desconhecido.

# FATO

Se o desvio padrão populacional  $\sigma$  for desconhecido, o desvio padrão amostral,  $S$ , é usado para estimar  $\sigma$ .

Entretanto, a padronização da média amostral  $\bar{X}$  utilizando o desvio padrão amostral segue uma distribuição de probabilidades conhecida como *distribuição t-Student*, desde que uma a.a.s. tenha sido coletada de uma população em que  $X \sim \text{Normal}$ .

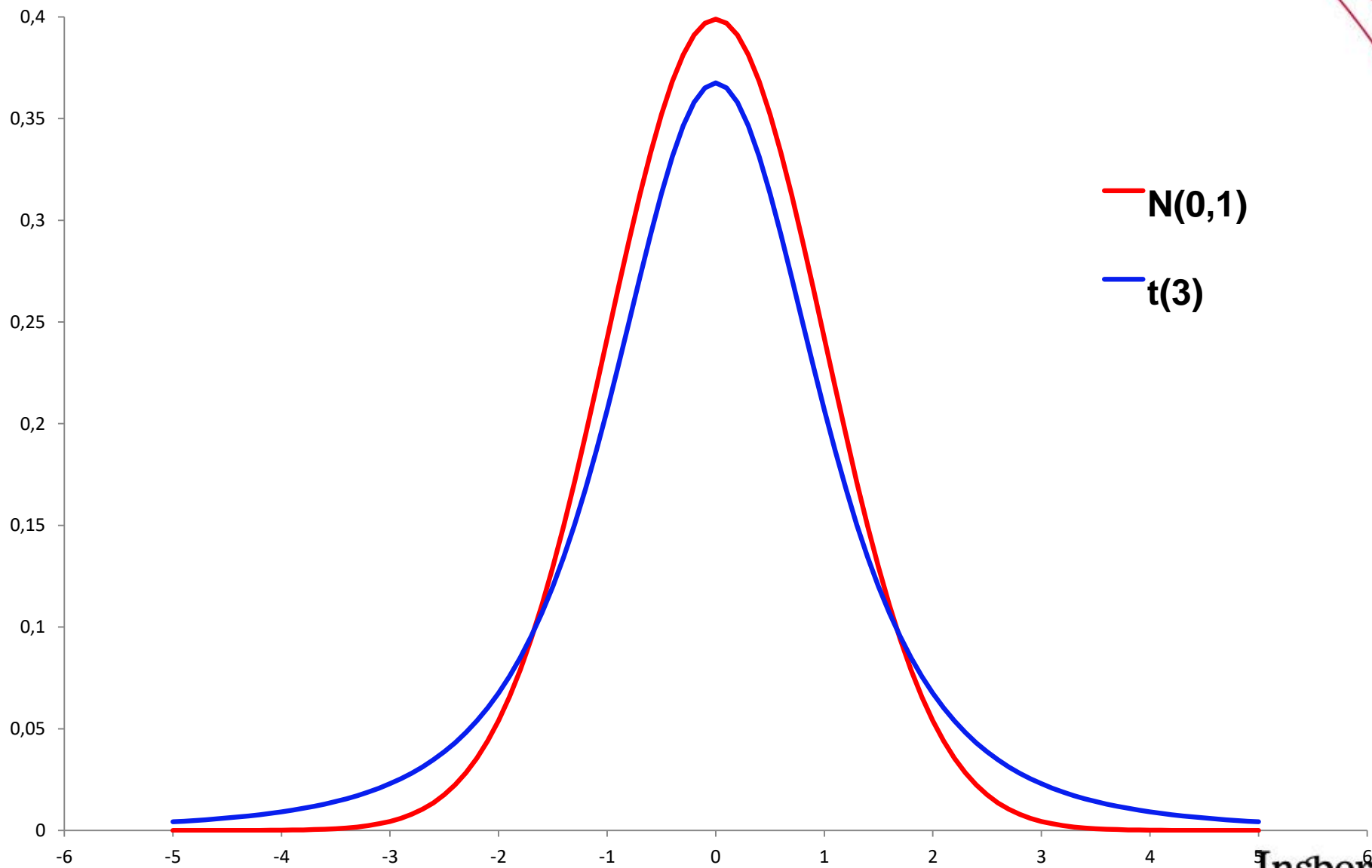


Assim, utilizando o estimador  $S^2$  para  $\sigma^2$  e supondo que a a.a.s. foi coletada de uma população cuja variável de interesse seja normalmente distribuída, temos, sob  $H_0$ , que

$$t = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \sim t_{(n-1)}$$

pode ser considerada uma estatística de teste.

# $t_{(3)}$ versus Normal Padrão



# Passos para Construção de um Teste de Hipóteses (via RC)

**1º.Passo:** Fixe qual a hipótese nula,  $H_0$ , a ser testada e qual a hipótese alternativa ( $H_A$ ).

**2º.Passo:** Use a teoria estatística e as informações disponíveis para decidir qual *estatística de teste* será usada sob  $H_0$ . Não se esqueça de levantar as propriedades dessa estatística.

**3º.Passo:** Fixe a probabilidade  $\alpha$  de cometer erro de rejeitar  $H_0$ , sob  $H_0$  verdadeiro, e use este valor para *construir a região crítica RC*. Lembre que esta região é construída para a estatística definida no segundo passo, usando o valor hipotetizado em  $H_0$ .

**4º.Passo:** Use as informações fornecidas pela amostra para encontrar o *valor observado da estatística de teste*.

**5º.Passo:** Se o *valor observado da estatística de teste* pertencer à região crítica, rejeite  $H_0$ ; caso contrário, não rejeite.

# Valor-p do Teste

**Valor-p** é o menor nível de  
significância que leva à rejeição de  
 $H_0$  com base na amostra.

# Passos para Construção de um Teste de Hipóteses (via valor-p)

**1º.Passo:** Fixe qual as hipóteses  $H_0$  e  $H_A$ .

**2º.Passo:** Use a teoria estatística e as informações disponíveis para decidir qual *estatística de teste* será usada sob  $H_0$ . Não se esqueça de levantar as propriedades dessa estatística.

**3º.Passo:** Use as informações fornecidas pela amostra para encontrar o *valor observado da estatística de teste*.

**4º.Passo:** Use o valor observado da *estatística de teste* para *encontrar o valor-p*, ou seja, a probabilidade de encontrar valores tão ou mais desfavoráveis à  $H_0$  quanto a *estatística de teste* observada pela amostra.

**5º.Passo:** Se o *valor-p* for menor do que algum  $\alpha$  fixado, *rejeite  $H_0$* ; caso contrário, não rejeite.

# Teste unilateral ou unicaudal à direita

$$H_0 : \mu = \mu_0$$

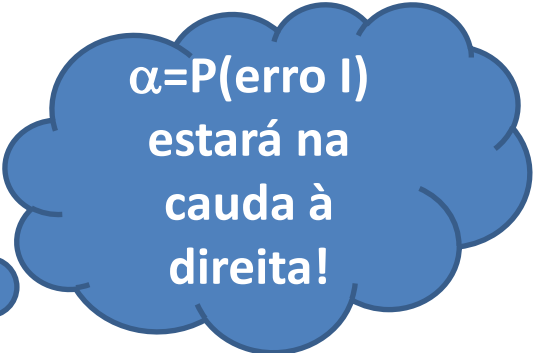
$$H_A : \mu > \mu_0$$

Estatística do teste observada (sob  $H_0$ ):

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s / \sqrt{n}}$$

Regra de rejeição, ao nível  $\alpha$  de significância:

**Rejeito  $H_0$  se**  $t_{obs} > t_{(n-1)}^{(\alpha)} = t_c$



$\alpha = P(\text{erro I})$   
estará na  
cauda à  
direita!

# Teste unilateral ou unicaudal à esquerda

$$H_0 : \mu = \mu_0$$

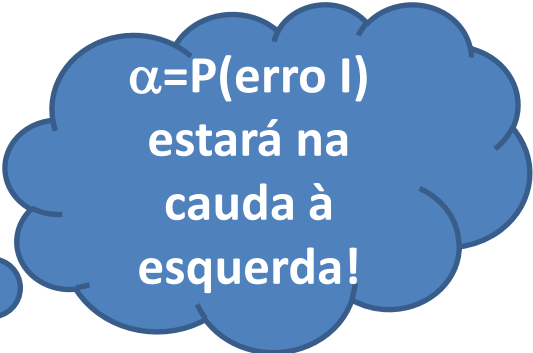
$$H_A : \mu < \mu_0$$

Estatística do teste observada (sob  $H_0$ ):

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s / \sqrt{n}}$$

Regra de rejeição, ao nível  $\alpha$  de significância:

**Rejeito  $H_0$  se**  $t_{obs} < -t_{(n-1)}^{(\alpha)} = t_c$



$\alpha = P(\text{erro I})$   
estará na  
cauda à  
esquerda!

# Teste bilateral ou bicaudal

$$H_0 : \mu = \mu_0$$

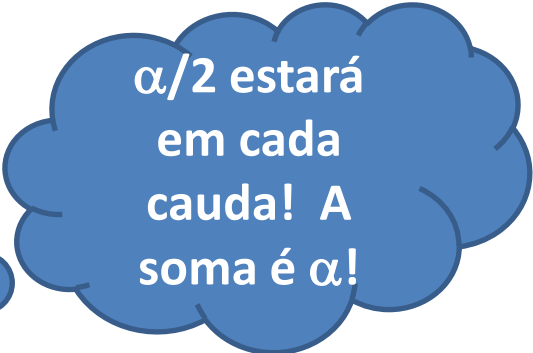
$$H_A : \mu \neq \mu_0$$

Estatística do teste observada (sob  $H_0$ ):

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s / \sqrt{n}}$$

Regra de rejeição, ao nível  $\alpha$  de significância:

**Rejeito  $H_0$  se**  $|t_{obs}| > t_{(n-1)}^{(\alpha/2)} = t_c$



$\alpha/2$  estará  
em cada  
cauda! A  
soma é  $\alpha$ !



## Exemplo 2

As latas de certa marca de refrigerante apresentam em seu rótulo o volume de 350 ml.

O fabricante deseja testar se o **conteúdo médio das latas é igual a 350 ml**, como anunciado no rótulo. Isto equivale a verificar se a máquina está regulada para colocar 350 ml, ou não, nas latas.

Para averiguar a afirmação do fabricante, foi coletada uma amostra de 36 latas do refrigerante em pontos de comercialização e mediu-se o conteúdo destas latas.

Os resultados obtidos na amostra foram:  $\bar{x} = 347$  ml e  $s = 10,5$  ml

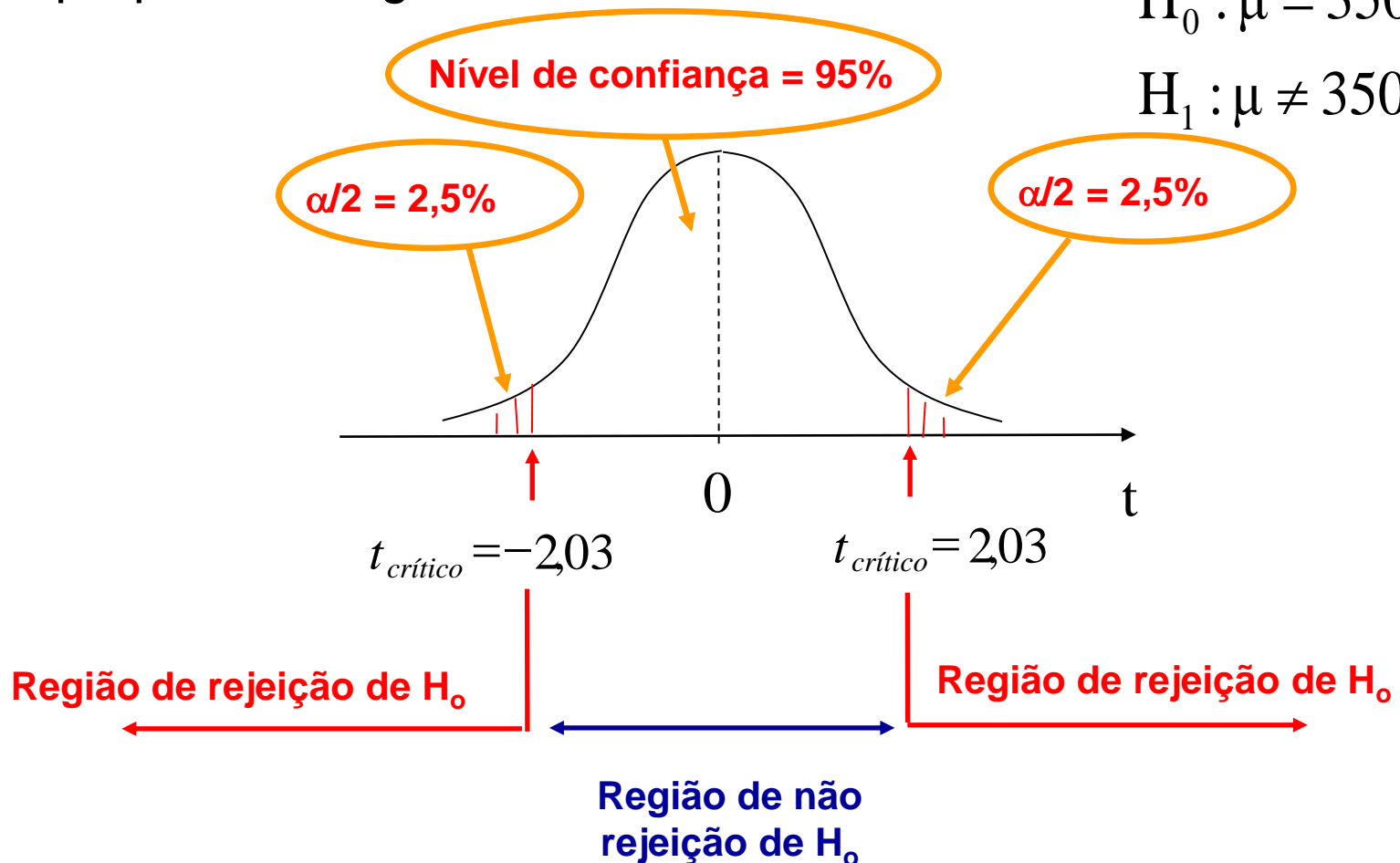
**Será que as latas contêm 350 ml de líquido com 95% de confiança?**

# Voltando ao exemplo 2

Com base nas hipóteses do fabricante, rejeita-se a hipótese nula para valores pequenos ou grandes.

$$H_0 : \mu = 350 \text{ ml}$$

$$H_1 : \mu \neq 350 \text{ ml}$$



$$t_c = \text{stats.t.ppf}(1-0.025, \text{df}=35) = 2.030$$

# Voltando ao exemplo 2

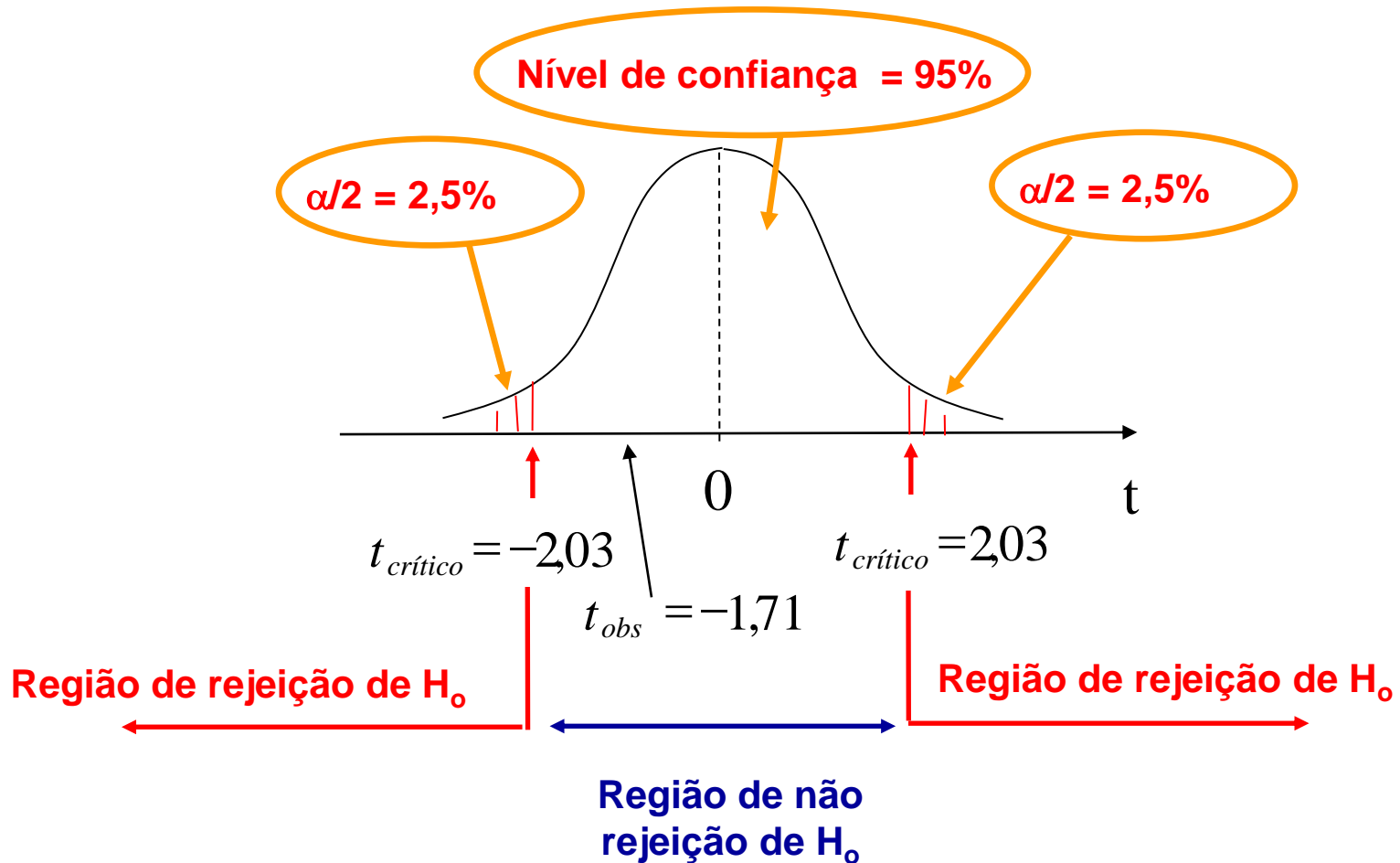
**Região crítica:** se o valor da estatística  $t_{obs}$  for menor que  $-2,03$  ou maior que  $2,03$ , então rejeita-se a hipótese nula (o produto não está de acordo com as especificações do fabricante).

## **Estatística do Teste (obtida da amostra)**

Padronização dos dados amostrais sob a hipótese nula ( $H_0$ ), ou seja  $\mu_0 = 350$ .

$$t_{obs} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \quad \begin{array}{c} \text{amostra} \\ \xrightarrow{\hspace{1cm}} \\ \bar{x} = 347 \text{ ml} \\ s = 10,5 \text{ ml} \end{array} \quad t_{obs} = \frac{347 - 350}{10,5 / \sqrt{36}} = -1,71$$

# Voltando ao exemplo 2



**Conclusão:** Não rejeitamos a hipótese nula, isto é, não existem evidências estatísticas de que o conteúdo das latas esteja fora das especificações do fabricante, ao nível de significância de 5% (ou com 95% de confiança).

# Voltando ao exemplo 2

Para calcular o valor-p para testes bicaudais devemos multiplicar por 2 o valor da probabilidade calculada com a estatística do teste, já que rejeitamos a hipótese nula tanto para pequenos como para grandes valores amostrais.

Dessa forma:  $t_{obs} = (347 - 350)/(10,5/6) = -1,71$   
 $valor - p = 2 * stats.t.cdf(tobs, df = 35)$

ou:  $valor - p = 2 * stats.t.cdf(347, df = 35, loc = 350, scale = 10.5/6)$

Portanto, **não rejeitamos a hipótese nula** (pois valor-p = 0,0962 > 0,05 =  $\alpha$ ), isto é, não existem evidências estatísticas de que o conteúdo das latas esteja fora das especificações do fabricante.

# Exemplo 3

Supondo as hipóteses a seguir, calcule e interprete o valor-p.

$$H_0 : \mu = \mu_0$$

$$H_A : \mu > \mu_0$$

$$n = 30; \quad s = 6; \quad \mu_0 = 30; \quad \bar{x} = 31,87; \quad \alpha = 0,05$$

Resp.:

$$\text{valor} - p = P(t_{(29)} > 1,707) = 4,92\% < 5\% \Rightarrow \text{rejeita } H_0!$$

# Exemplo 4

Supondo as hipóteses a seguir, calcule e interprete o valor-p.

$$H_0 : \mu = \mu_0$$

$$H_A : \mu < \mu_0$$

$$n = 30; \quad s = 6; \quad \mu_0 = 30; \quad \bar{x} = 28,13; \quad \alpha = 0,05$$

Resp.:

$$\begin{aligned} \text{valor} - p &= P(t_{(29)} < -1,707) = P(t_{(29)} > 1,707) = \\ &= 4,92\% < 5\% \Rightarrow \text{rejeita } H_0! \end{aligned}$$

# Exemplo 5

Supondo as hipóteses a seguir, calcule e interprete o valor-p.

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0$$

$$n = 30; \quad s = 6; \quad \mu_0 = 30; \quad \bar{x} = 31,87; \quad \alpha = 0,05$$

Resp.:

$$\begin{aligned} \text{valor} - p &= 2 * P(t_{(29)} > 1,707) = 2 * 4,92\% = \\ &= 9,84\% > 5\% \Rightarrow \text{NÃO rejeita } H_0! \end{aligned}$$



# Exercícios

# Exercício 1

O índice de poluição no município de Curitiba segue uma distribuição normal com média e variância desconhecidas. O departamento ambiental deseja estimar o índice médio de poluição no município. Para isso, ele medirá a poluição em uma amostra de dias escolhidos aleatoriamente.

a) Pretende-se extrair, em Curitiba, uma amostra aleatória de 16 dias. Em uma cidade com características similares, verificou-se que o índice médio de poluição é de 90 u.m..

Construa uma regra de decisão para concluir se Curitiba é ou não mais poluída do que a outra cidade. Adote um nível de significância de 10%.

# Exercício 1 (cont.)

- b) Interprete os erros do tipo I e II relacionados ao teste acima, em termos do problema em questão.
- c) Extraída uma amostra aleatória de 16 dias verificou-se, em Curitiba, um índice médio amostral de poluição de 95 u.m., com desvio padrão amostral igual a 10 u.m.. Conclua o T.H. por meio da construção da R.C..

# Exercício 1 (cont.)

- d) Através do cálculo do valor-p, conclua o teste de hipóteses. Interprete o valor-p.
- e) Descreva as suposições necessárias para as conclusões acima serem confiáveis.
- f) Um técnico resolveu medir a poluição em 16 dias consecutivos. A amostra obtida satisfaz as suposições necessárias para a realização do teste? Por quê?

# Exercício 2

O volume diário de negócios da corretora K. B. Sashata, em reais, segue uma distribuição normal. O diretor da corretora deseja fazer inferências sobre o volume médio negociado por ela diariamente.

a) Numa corretora de mesmo porte verificou-se que, em média, o volume negociado diariamente é de R\$ 116.000,00. Formule as hipóteses de um teste para verificar essas duas corretoras apresentam, diariamente, o mesmo volume de negociações.

b) Interprete os erros tipo I e tipo II relacionados ao teste acima, em termos do problema em questão.

c) Extraída uma amostra de 25 dias, verificou-se que o volume médio negociado diariamente na corretora K. B. Sashata é igual a R\$ 115.000,00, com desvio padrão amostral igual a R\$2.000,00. Conclua o teste descrito no item (a) com base no cálculo do valor-p.

# Exercício 3

Em 2012, antes dos supermercados deixarem de distribuir sacolas plásticas gratuitamente para os consumidores, o consumo mensal de sacolas plásticas na cidade de São Paulo por adulto era, em média, de 25 sacolas plásticas, com desvio padrão de 6 unidades (valores populacionais).

No mesmo ano, os estabelecimentos passaram a cobrar pelo uso das mesmas e depois voltaram atrás. Alguns anos se passaram e uma ONG da área de sustentabilidade deseja avaliar se essas idas e vindas deixaram o consumidor um pouco mais consciente com o meio ambiente.

Para avaliar se consumo está mais consciente, verifique se o consumo de sacolas plásticas reduziu, em média, em mais do que 15% após todo esse período, ao nível de significância de 1%?

A ONG monitorou, por um mês, 100 adultos moradores da cidade e observou que o consumo médio amostral de sacolas plásticas entre eles foi de 20,01 unidades.