ME414 - Estatística para Experimentalistas

Parte 18

Teste de Hipóteses

Relembrando: Teste de Hipóteses Passo-a-Passo

- **Passo 1**: Suposições
- **Passo 2**: Hipóteses
- **Passo 3**: Estatística do Teste
- **Passo 4**: p-valor
- **Passo 5**: Conclusões

Teste de Hipótese para uma proporção

Suponha que temos uma população e uma hipótese sobre a proporção p de indíviduos com certa característica.

Hipóteses:

$$H_0 : p = p_0 \quad \text{vs} \quad H_A : \begin{aligned} &p \neq p_0 \text{ (bilateral)} \\ &p < p_0 \text{ (unilateral à esquerda)} \\ &p > p_0 \text{ (unilateral à direita)} \end{aligned}$$

Estatística do teste: Baseada na distribuição amostral de $\hat{p}$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{H_0}{\sim} N(0,1)$$

Condição: $np_0 \geq 10$ e $n(1-p_0) \geq 10$ para aproximação normal.

Teste de Hipótese para uma proporção

p-valor

- $H_A : p \neq p_0$ (bilateral): p-valor=$P(|Z| \geq |z_{obs}|)$
- $H_A : p < p_0$ (unilateral à esquerda): p-valor=$P(Z \leq z_{obs})$
- $H_A : p > p_0$ (unilateral à direita): p-valor=$P(Z \geq z_{obs})$

Conclusão: Para um nível de significância α

- Se p-valor $\leq \alpha$: rejeitamos H_0
- Se p-valor > α: não rejeitamos H_0

Experimento da Coca vs Coca Zero

Os alunos dizem que conseguem distinguir entre Coca-Cola normal e Coca-Cola Zero.

Um experimento foi feito em sala de aula para comprovar se a afirmação é verdadeira.

Um aluno experimentou, em ordem aleatória, 20 amostras (ao acaso era Coca normal ou zero) e anotou-se a quantidade de acertos.

Cada tentativa, X_i, é uma Bernoulli(p), em que p é a probabilidade de acerto.

Veja que $T = \sum_{i=1}^{20} X_i \sim Bin(20, p)$, onde T é o número de acertos.

Dos 20 testes, o aluno acertou 19! Temos então uma proporção amostral de acertos $\hat{p} = 19/20 = 0.95$.
Isso mostra que o aluno realmente sabe a diferença?

Experimento da Coca vs Coca Zero

Vamos testar o seguinte:

$$H_0 : p = 0.50 \quad \text{vs} \quad H_A : p > 0.50$$

Podemos testar essas hipóteses de duas maneiras:

- Usando a aproximação normal para a proporção de acertos, como vimos na última aula, já que as condições $np_0 \geq 10$ e $n(1 - p_0) \geq 10$ são satisfeitas.
- Usando a distribuição exata do número total de acertos.

Vamos revisar o que vimos na aula passada e também fazer o teste com a distribuição exata de T.

Experimento da Coca vs Coca Zero

Usando a distribuição exata do número de acertos em 20 tentativas.

Hipóteses: $H_0 : p = 0.50 \quad \text{vs} \quad H_A : p > 0.50$

Hipóteses: $H_0 : \text{Acertos} = 10 \quad \text{vs} \quad H_A : \text{Acertos} > 10$

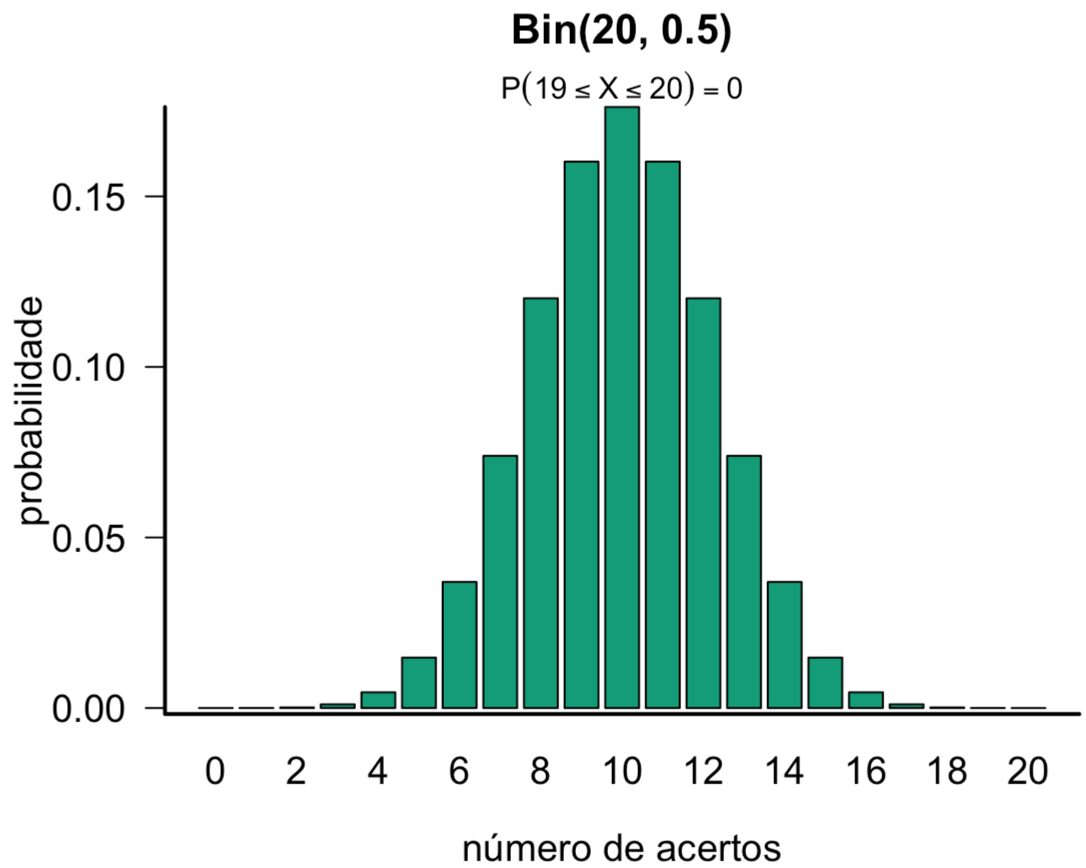
Estatística do teste: $T = \sum_{i=1}^{20} X_i \overset{H_0}{\sim} Bin(20, 0.5)$

O valor observado da estatística do teste é $t_{obs} = 19$, ou seja, o número total de acertos.

p-valor = $P(T \geq 19) = 0.00002$

Conclusão: Fixando $\alpha = 0.05$, rejeitamos a hipótese de que $p = 0.5$ e, portanto, acreditamos que a probabilidade de acertos é maior que 50%.

Experimento da Coca vs Coca Zero



Tipos de Erro

Quando realizamos um teste de hipóteses, podemos cometer 2 tipos de erros:

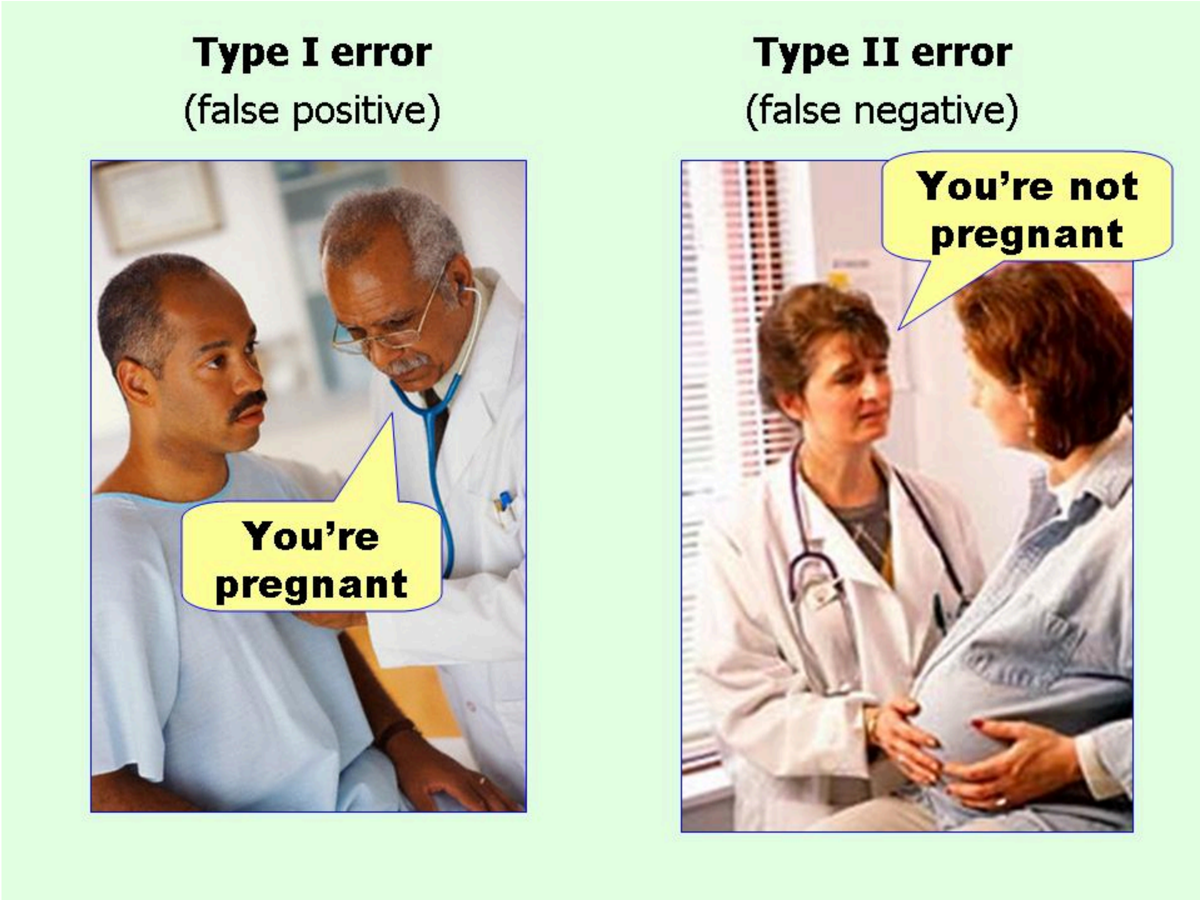
1- **Erro Tipo I:** Rejeitar a hipótese H_0, quando tal hipótese é verdadeira

2- **Erro Tipo II:** Não rejeitar a hipótese H_0, quando tal hipótese é falsa

Decisão	**Ho**	
	Verdadeira	**Falsa**
Rejeitar H_0	Erro Tipo I	OK ✓
Não Rejeitar H_0	OK ✓	Erro Tipo II

Erro Tipo I: erro mais grave

Tipos de Erro



H_0 : você não está grávida(o)

H_A : você está grávida(o)

Tipos de Erro

Podemos calcular as probabilidades dos dois tipos de erro, chamadas de α e β:

$$\alpha = P(\text{Erro Tipo I}) = P(\text{Rejeitar } H_0 | H_0 \text{ verdadeira})$$

$$\beta = P(\text{Erro Tipo II}) = P(\text{Não Rejeitar } H_0 | H_0 \text{ falsa})$$

Na situação ideal, ambas as probabilidades de erro, α e β, seriam próximas de zero. Entretanto, à medida que diminuímos α, a probabilidade β tende a aumentar.

Levando isso em conta, em teste de hipóteses tentamos controlar a probabilidade do erro do tipo I, já que esse é o erro mais grave.

A probabilidade α é chamada de **nível de significância**, que geralmente fixamos em 5%.

Tipos de Erro

No experimento da Coca-Cola tivemos 19 acertos em 20 tentativas e decidimos rejeitar H_0.

Mas e se tivéssemos observado 14 acertos? Ou 12?

Existe um valor, t_c, de maneira que se observarmos algo igual ou maior que ele decidimos rejeitar H_0?

Esse valor é chamado de **valor crítico** e vamos denotá-lo por t_c.

Tipos de Erro

No experimento da Coca-Cola: $H_0 : p = 0.5 \quad \text{vs} \quad H_A : p > 0.5$

Seja T o número de acertos em uma amostra de tamanho $n = 20$. Então $T \sim Bin(20, p)$.

Vamos considerar o seguinte valor crítico: $t_c = 12$.

Lembrando que T pode assumir os valores $0, 1, 2, \ldots, 20$.

O valor crítico t_c determina as probabilidades de cometer os erros tipo I e II.

Tipos de Erro

Considerando $t_c = 12$

$$\begin{aligned} P(\text{Erro Tipo I}) &= P(\text{Rejeitar } H_0 | H_0 \text{ verdadeira}) \\ &= P(T \geq t_c | p = 0.5) \\ &= \sum_{x=12}^{20} P(T = x | p = 0.5) \approx 0.25 \end{aligned}$$

$$\begin{aligned} P(\text{Erro Tipo II}) &= P(\text{Não Rejeitar } H_0 | H_0 \text{ falsa}) \\ &= P(T < t_c | p = 0.7) \\ &= \sum_{x=0}^{11} P(T = x | p = 0.7) \approx 0.11 \end{aligned}$$

Tipos de Erro

Observando a relação entre os erros tipo I e II, e t_c: $H_0 : p = 0.5$ vs $H_A : p = 0.7$

t_c	P(Erro Tipo I)	P(Erro Tipo II)
12	0.25	0.11
13	0.13	0.23
14	0.06	0.39
15	0.02	0.58

Veja que à medida que $\alpha =$P(Erro Tipo I) diminui, $\beta =$P(Erro Tipo II) aumenta.

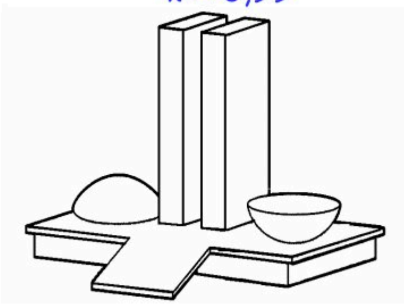
Então, optamos por controlar $\alpha =$P(Erro Tipo I), que é considerado o erro mais grave. Geralmente fixamos $\alpha = 0.05$ e rejeitamos H_0 se p-valor $< \alpha$.

Teste de hipóteses para média (σ conhecido)

Teste de hipóteses: proporção ou média

Proporção de políticos honestos

$\pi = 0{,}99$



Quantidade média de carboidratos

$\mu = 13g$

QUANTIDADE POR EMBALAGEM	
VALOR ENERGÉTICO	117kcal = 491kJ
CARBOIDRATOS	13g
PROTEINAS	10g
GORDURAS TOTAIS	2,8g
GORDURAS SATURADAS	1,9g

Exemplo: Café

Vamos voltar no problema da máquina que enche pacotes de café. Digamos que o peso nominal do pacote de café seja de 500g. Assume-se que o desvio padrão é conhecido ($\sigma = 10$).

Retiraram uma amostra de 25 pacotes e observaram um peso médio de 485g.

Isso nos traz evidência de que os pacotes têm menos de 500g?

Já calculamos o IC de 95% para esse problema:

$$IC(\mu, 0.95) = [481.08; 488.92]$$

Vamos agora testar as hipóteses:

$$H_0 : \mu = 500 \quad \text{vs} \quad H_A : \mu \neq 500$$

Exemplo: Café

Suposições: Seja X_i o peso do i-ésimo pacote de café. Sabemos que $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2$. Coletou-se uma amostra de tamanho $n = 25$. Pelo TCL:

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

Hipóteses: $H_0 : \mu = \mu_0 = 500 \quad \text{vs} \quad H_A : \mu \neq \mu_0 = 500$

Estatística do teste:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$$

Considerando a amostra obtida:

$$z_{obs} = \frac{485 - 500}{10/5} = -7.5$$

Teste de hipóteses para média (σ conhecido)

Como medir se -7.5 é evidência contra H_0?

O teste é bilateral, portanto o p-valor é calculado como:

p-valor: $P(|Z| \geq 7.5) = 2P(Z \geq 7.5) \approx 0$

Conclusão: Como o p-valor é praticamente zero, rejeitamos H_0, ou seja, rejeitamos a hipótese de que a média é 500g.

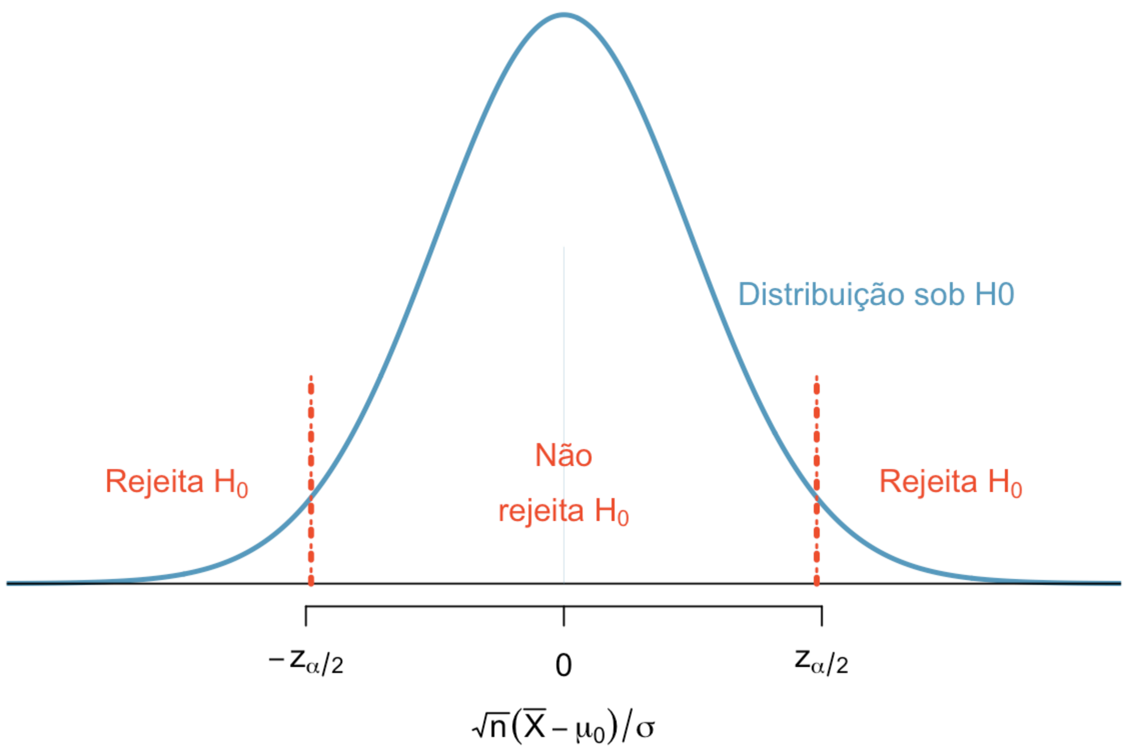
Região Crítica (Região de Rejeição)

Outra forma de decidirmos se a evidência encontrada nos dados é forte o suficiente para rejeitar H_0 é determinando a **região crítica** ou **região de rejeição**.

Região Crítica: conjunto de valores da estatística do teste para os quais a hipótese nula é rejeitada.

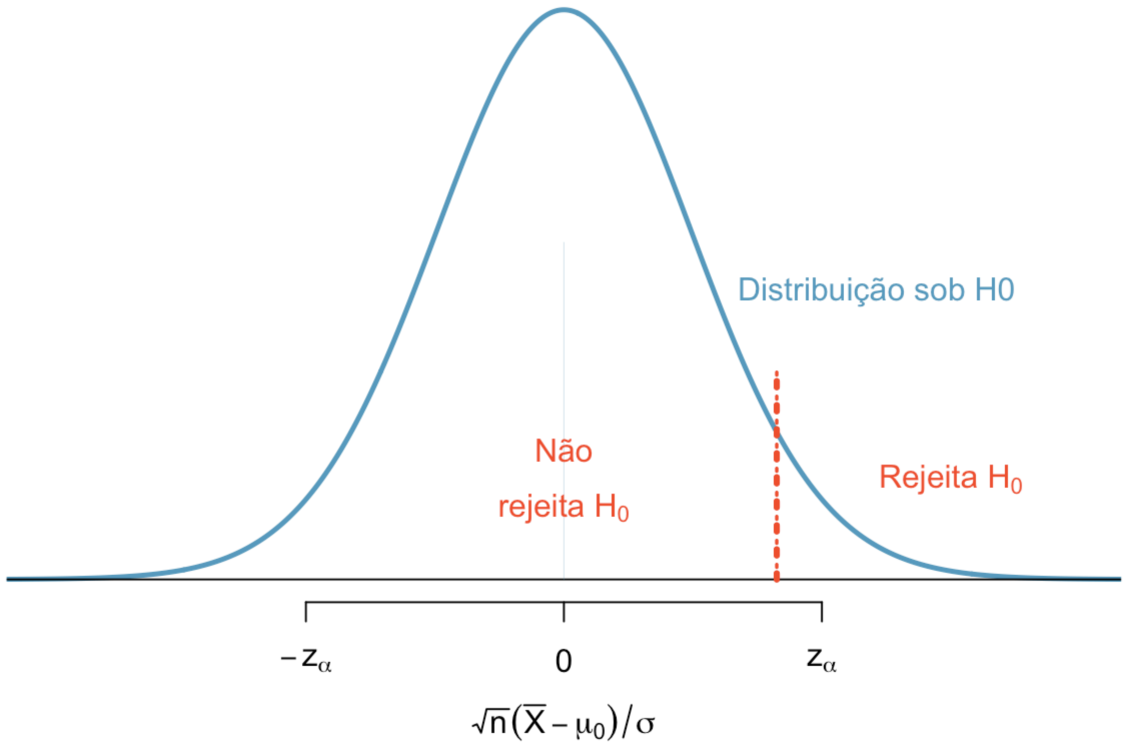
Região crítica: teste bilateral

Hipóteses $H_0 : \mu = \mu_0$ vs $H_A : \mu \neq \mu_0$
Para um **nível de significância** α, definimos a **região crítica** do teste:



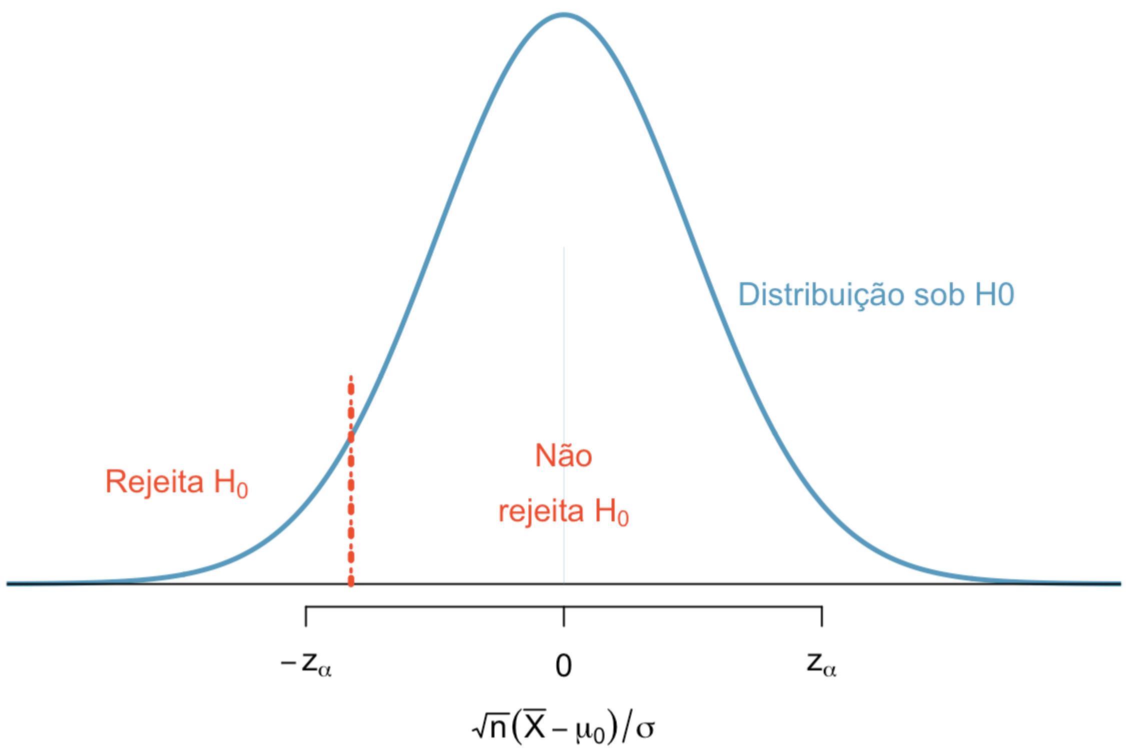
Região crítica: teste unilateral à direita

Hipóteses $H_0 : \mu = \mu_0$ vs $H_A : \mu > \mu_0$
Para um **nível de significância** α, definimos a **região crítica** do teste:



Região crítica: teste unilateral à esquerda

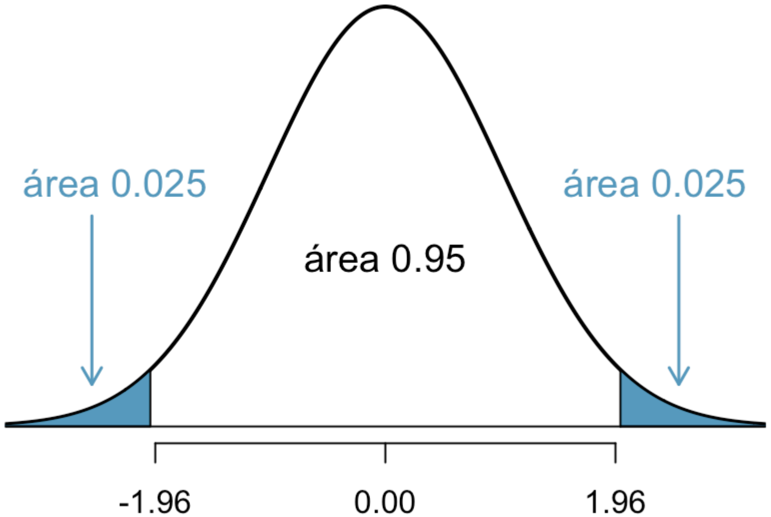
Hipóteses $H_0 : \mu = \mu_0$ vs $H_A : \mu < \mu_0$
Para um **nível de significância** α, definimos a **região crítica** do teste:



Região Crítica: teste bilateral

Quando o teste for bilateral: $H_0 : \mu = 500 \quad \text{vs} \quad H_A : \mu \neq 500$

A região critíca, para $\alpha = 0.05$, é a área em azul na figura abaixo:

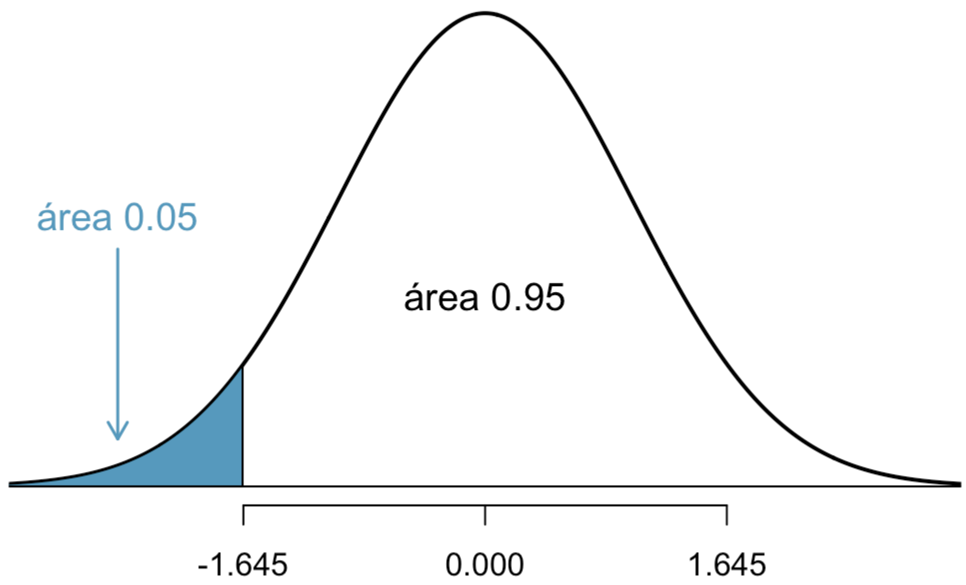


Decisão: Rejeitamos H_0 se $z_{obs} < -1.96$ ou $z_{obs} > 1.96$.
No nosso exemplo, $z_{obs} = -7.5$. Portanto, rejeitamos H_0.

Região Crítica: teste unilateral à esquerda

Quando o teste for unilateral à esquerda: $H_0 : \mu = \mu_0 \quad \text{vs} \quad H_A : \mu < \mu_0$

A região critíca, para $\alpha = 0.05$, é a área em azul na figura:

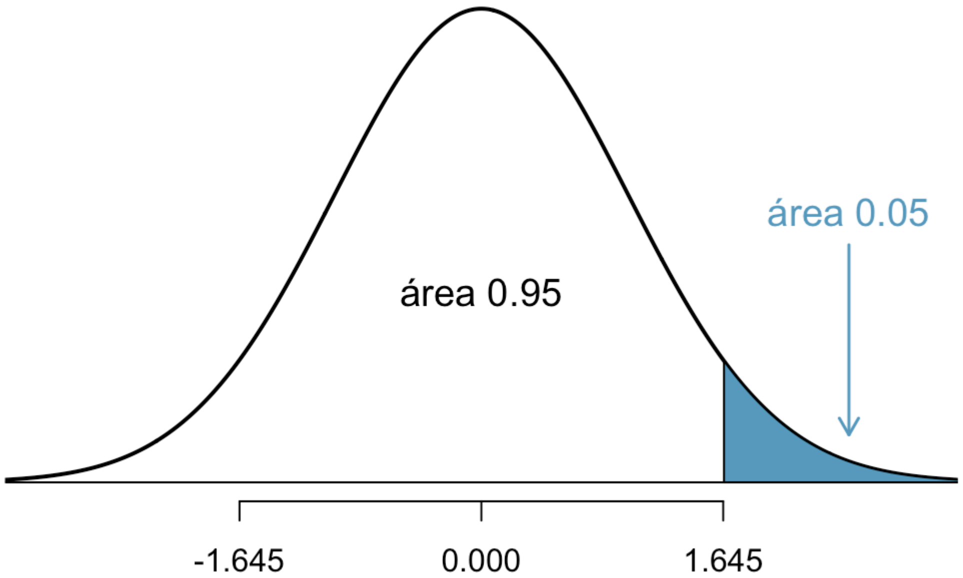


Decisão: Rejeitamos H_0 se $z_{obs} < -1.645$.

Região Crítica: teste unilateral à direita

Quando o teste for unilateral à direita: $H_0 : \mu = \mu_0 \quad \text{vs} \quad H_A : \mu > \mu_0$

A região critíca, para $\alpha = 0.05$, é a área em azul na figura:



Decisão: Rejeitamos H_0 se $z_{obs} > 1.645$.

Teste de hipóteses para média (σ desconhecido)

Teste de hipóteses para média (σ desconhecido)

No caso de testar

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_A : \mu \neq \mu_0$$

quando σ é desconhecido e a amostra é pequena ($n < 30$) devemos utilizar a distribuição t.

Estatística do teste:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$$

p-valor: $P(|t_{n-1}| \geq |t_{obs}|) = 2P(t_{n-1} \geq |t_{obs}|)$
Para as hipóteses unilaterais, o raciocínio é semelhante ao que foi feito anteriormente quando σ é conhecido.

Teste de hipóteses para média (σ desconhecido)

No nosso exemplo, suponha que não sabemos o valor de σ, mas o desvio padrão da amostra é 7.1g. Queremos testar

$$H_0 : \mu = 500 \quad \text{vs} \quad H_A : \mu \neq 500$$

Estatística do teste:

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}} = \frac{485 - 500}{7.1/5} = -10.56$$

p-valor: $P(|t_{24}| \geq 10.56) = 2P(t_{24} \geq 10.56) \approx 0$

Conclusão: Rejeitamos a hipótese de que a média é 500g.

valor crítico: para nível de significância $\alpha = 0.05$ e teste bilateral, t_{crit} é tal que $P(t_{24} > t_{crit}) = P(t_{24} < -t_{crit}) = 0.025$. De maneira que $t_{crit} = 2.06$. Portanto, se $|t_{obs}| > t_{crit}$, rejeita-se H_0.

Exemplo: Dieta LowCarb

- 41 pacientes obesos, selecionados aleatoriamente, foram submetidos a uma dieta com baixa quantidade de carboidratos.
- Pesquisadores responsáveis pelo estudo acreditam que essa dieta faz com que os pacientes apresentem uma redução de peso.
- Após 16 semanas, a diferença média de peso foi -9.7kg, com desvio padrão 3.4 kg.
- O que podemos concluir deste estudo?

Detalhes do estudo: Effect of 6-month adherence to a very low carbohydrate diet program.

Exemplo: Dieta LowCarb

Suposições: X_i é a diferença entre peso inicial e final do i-ésimo obeso.

Sabemos que $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2$.

Coletou-se uma amostra de tamanho $n = 41$.

Pelo TCL: $\bar{X} \sim N(\mu, \sigma^2/n)$

Hipóteses: $H_0 : \mu = 0 \quad \text{vs} \quad H_A : \mu < 0$

Ou seja, estamos testando se não há diferença no peso após a dieta versus a hipótese que há redução no peso após a dieta.

Exemplo: Dieta LowCarb

Estatística do teste: Como $n = 41$, podemos usar a aproximação normal

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}} = \frac{-9.7 - 0}{3.4/\sqrt{41}} = -18.3$$

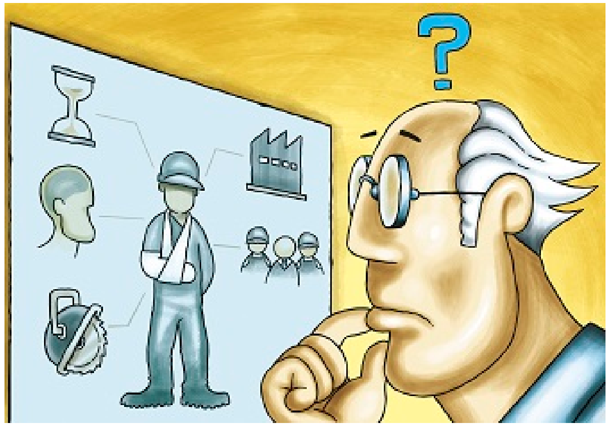
p-valor: Como o teste é unilateral à esquerda

$$\text{p-valor} = P(Z < -18.3) \approx 0$$

Conclusão: Como o p-valor é bem pequeno (<0.05) rejeitamos H_0, ou seja, rejeitamos a hipótese de que a dieta não produz diferença no peso.

Exemplo: Acidentes de trabalho

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 qual foi tomada uma amostra de nove indústrias e medido o número de horas/homens perdidos por acidentes, que foi de 50 horas.

Você diria, a um nível de significância de 5%, que há evidência de melhoria?

Fonte: Morettin & Bussab, Estatística Básica 5^a edição, pág. 334.

Exemplo: Acidentes de trabalho

Queremos testar a hipótese que μ, o número médio de horas perdidas com acidentes de trabalho, tenha permanecido o mesmo. Ou seja,

$$H_0 : \mu = 60 \quad \text{vs} \quad H_A : \mu < 60$$

Estatística do teste:

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}} = \frac{50 - 60}{20/3} = -1.5$$

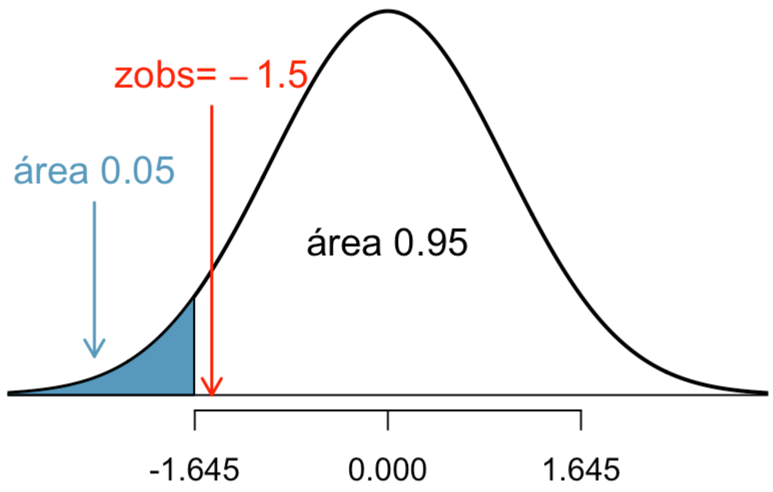
p-valor: $P(Z \leq -1.5) = 0.067$

Conclusão: Como o p-valor é maior que 0.05, não rejeitamos a hipótese de que a média é 60. Ou seja, não há evidência contra da hipótese de que o número médio de horas perdidas tenha se mantido o mesmo.

Exemplo: Acidentes de trabalho

Podemos também determinar a região crítica.

Como temos um teste unilateral à esquerda, para um nível de significância de 5%, rejeitamos H_0 se $z_{obs} < -z_{0.05} = -1.645$.



Como $z_{obs} = -1.5 > -1.645$, então não rejeitamos H_0.

Resumo: Teste de hipóteses para média

σ conhecido	σ desconhecido
$H_0: \mu = \mu_0$ vs	$H_0: \mu = \mu_0$ vs
$H_a: \mu \neq \mu_0$ ou	$H_a: \mu \neq \mu_0$ ou
$\mu > \mu_0$ ou $\mu < \mu_0$	$\mu > \mu_0$ ou $\mu < \mu_0$
Estatística do teste:	Estatística do teste:
$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$	$t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$
valor-de-p=	valor-de-p=
$P(\lvert Z\rvert \geq \lvert z_{obs}\rvert)$ se $H_a: \mu \neq \mu_0$	$P(\lvert t_{n-1}\rvert \geq \lvert t_{obs}\rvert)$ se $H_a: \mu \neq \mu_0$
$P(Z \geq z_{obs})$ se $H_a: \mu > \mu_0$	$P(t_{n-1} \geq t_{obs})$ se $H_a: \mu > \mu_0$
$P(Z \leq z_{obs})$ se $H_a: \mu < \mu_0$	$P(t_{n-1} \leq t_{obs})$ se $H_a: \mu < \mu_0$

Leituras

- Ross: capítulo 9.
- OpenIntro: seção 5.1.
- Magalhães: capítulo 8.

Slides produzidos pelos professores:

- Samara Kiihl
- Tatiana Benaglia
- Larissa Matos
- Benilton Carvalho