Estatística Aplicada à Pesquisa

Aula 6 - Introdução à Inferência Estatística

Dra. Agatha Rodrigues

Outubro de 2020

agatha.srodrigues@gmail.com

Inferência Estatística

Inferência Estatística

O que fazer com as observações que coletamos?

Primeira etapa

Análise descritiva

Resumir os dados;

Entender os dados.

Como fazer?

- Tabelas de frequências.
- Medidas descritivas ou resumo
  - Medidas de posição
  - Medidas de dispersão

Depende do tipo de variável

Segunda etapa

Análise inferencial

Inferir certos fatos acerca da população a partir de resultados observados na amostra, controlando erros.

Como fazer?

Estimamos parâmetros populacionais

Testamos hipóteses

População e Amostra

A partir da amostra, podemos fazer INFERÊNCIA sobre as características da população.

Inferência Estatística:

é um conjunto de técnicas que objetiva estudar uma população através de evidências fornecidas por uma **amostra**.

População e Amostra

Exemplo 1:

Uma turma de um dado curso é composta por 120 alunos. O **interesse** consiste em saber qual é a altura média dos alunos dessa turma.

- Quem é a população de interesse?
- Qual é a variável de interesse?
- Quem é a medida de interesse?

População e Amostra

Exemplo 1:

Uma turma de um dado curso é composta por 120 alunos. O **interesse** consiste em saber qual é a altura média dos alunos dessa turma.

- Quem é a população de interesse?
- Qual é a variável de interesse?
- Quem é a medida de interesse?

Vamos imaginar dois cenários:

1. O professor mede a altura de todos os alunos e descobre que a altura média dos 120 alunos é de 1,68 m.
2. Por questão de tempo, o professor só consegue medir a altura de um subconjunto de 10 alunos. De forma apropriada, seleciona os 10 alunos e a média da altura desses alunos é 1,65 m.

Quais as diferenças entre esses dois cenários?
Faz sentido pensar em Inferñcia para o cenário 1?
O que significa a média de 1,68 m? E a média de 1,65 m?

Alguns Conceitos

População

Conjunto de todos os elementos sob investigação.

Amostra

Um subconjunto da população.

Variável

É uma característica de interesse da população.

Parâmetro

É uma medida usada para descrever uma característica da população.
Exemplos: Proporção, média, mediana, mínimo, máximo, etc.

Estimativa

É uma medida obtida da amostra com a finalidade de estimar um parâmetro populacional.

População e Amostra

Muitas vezes nõ é possível acessar toda a população para estudarmos as características de interesse.

Qual a solução?

Inferência estatística

Conjunto de técnicas que objetiva estudar a população através de evidências fornecidas por uma amostra, controlando erros.

População e Amostra

Muitas vezes nõ é possível acessar toda a população para estudarmos as características de interesse.

Qual a solução?

Inferência estatística

Conjunto de técnicas que objetiva estudar a população através de evidências fornecidas por uma amostra, controlando erros.

Como fazer a Inferência?

- Estimação: Pontual e Intervalar.
- Teste de Hipóteses.

Inferência Estatística: Estimativa pontual

Estimação Pontual

Para a amostra observada, as estimativas pontuais apresentam um único valor para o parâmetro.

Exemplo 2: IMC dos alunos do curso de Estatística da turma de 2019

- O **interesse** consiste em saber qual é o IMC médio dos alunos.
- Quem é parâmetro de interesse?

Estimação Pontual

Para a amostra observada, as estimativas pontuais apresentam um único valor para o parâmetro.

Exemplo 2: IMC dos alunos do curso de Estatística da turma de 2019

- O **interesse** consiste em saber qual é o IMC médio dos alunos.
- Quem é parâmetro de interesse? IMC médio
- Não foram consultados todos os alunos e uma amostra de tamanho n=10 é observada.
- A média amostral é uma estimativa pontual para a média populacional (parâmetro de interesse).

Estimação pontual

Uma estimativa para média populacional : Média amostral ($\bar{x}$)

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

em que x_i é a i-ésima observação, $i = 1, 2, \cdots, n$.

Exemplo 2: IMC dos alunos de 2019

x_i corresponde ao valor do IMC da i-ésima observação, $i = 1, 2, \cdots, 10$

IMC dos 10 alunos: 17,8 21,8 18,0 24,8 21,7 20,3 20,7 19,3 23,7 25,9

$$\bar{x} = \frac{17,8 + 21,8 + 18,0 + \cdots + 23,7 + 25,9}{10} = 21,4$$

Exemplo 3: Fetos gemelares sem RCF

- Deseja-se estudar a discordância de peso fetal (PF) de fetos gemelares sem restrição de crescimento fetal (RCF).
- A discordância é dada por:

$$\frac{(\text{PF feto maior} - \text{PF feto menor})}{(\text{PF feto maior})} * 100(\%)$$

- Para isso, uma amostra de n=150 gestações foi obtida.

Tabela 1: Medidas resumo da discordância fetal (%) obtidas da amostra (n=150).

	média	mediana	desvio-padrão	mínimo	máximo
Discordância	15,024	11,888	12,099	0,092	30,345

O que podemos dizer sobre os valores da Tabela 1?

Estimação pontual

Exemplo 3: Fetos gemelares sem RCF

- Deseja-se estudar a discordância de peso fetal (PF) de fetos gemelares sem restrição de crescimento fetal (RCF).
- A discordância é dada por:

$$\frac{(\text{PF feto maior} - \text{PF feto menor})}{(\text{PF feto maior})} * 100(\%)$$

- Para isso, uma amostra de n=150 gestações foi obtida.

Tabela 1: Medidas resumo da discordância fetal (%) obtidas da amostra (n=150).

	média	mediana	desvio-padrão	mínimo	máximo
Discordância	15,024	11,888	12,099	0,092	30,345

O que podemos dizer sobre os valores da Tabela 1?

São estimativas PONTUAIS da média, mediana, desvio-padrão, mínimo e máximo da discordância populacional, respectivamente.

Estimação pontual: Diferentes "n"

Pergunta: há relação da estimativa com o tamanho da amostra?

IMC dos alunos

Cenário I: n= 4	$\rightarrow$	estimativa pontual do IMC médio: 24,29.
Cenário II: n= 10	$\rightarrow$	estimativa pontual do IMC médio: 21,40.
Cenário III: n= 20	$\rightarrow$	estimativa pontual do IMC médio: 23,44.

Verdadeira média do IMC: 23,04

Erro absoluto:

Cenário I: n= 4	$\rightarrow$	$\vert 24,29 - 23,04\vert = 3,43.$
Cenário II: n= 10	$\rightarrow$	$\vert 21,40 - 23,04\vert = 1,64.$
Cenário III: n= 20	$\rightarrow$	$\vert 23,44 - 23,04\vert = 0,40.$

Estimação pontual

Há relação da estimativa com o tamanho da amostra?

É esperado que, quanto maior o n, mais próxima será a estimativa do valor do parâmetro.

Inferência Estatística: Estimativa intervalar

Estimação intervalar

- Para uma amostra observada, as estimativas pontuais apresentam um único valor numérico para o parâmetro.

Estimação intervalar

- Para uma amostra observada, as estimativas pontuais apresentam um único valor numérico para o parâmetro.

Estimação intervalar

Construir intervalos de confiança que incorporem à estimativa pontual informações a respeito de sua variabilidade (erro amostral).

Estimação intervalar

- Um intervalo $[LI; LS]$, em que LI é o limite inferior e LS é o limite superior, é definido como um intervalo de confiança para um parâmetro θ (qualquer) com coeficiente de confiança $1-\alpha$ e será denotado por:

$$IC(\theta; 1-\alpha) = [LI; LS].$$

- Em geral, um intervalo de confiança tem a forma:

$$IC(\theta; 1-\alpha) = [\text{Estimativa pontual} - \epsilon; \text{Estimativa pontual} + \epsilon].$$

em que ϵ é a margem de erro. Ou seja, $LI = \text{Estimativa pontual} - \epsilon$ e $LS = \text{Estimativa pontual} + \epsilon$.

Estimação intervalar

Intervalo de confiança para a média

- Vamos considerar a média populacional (μ) como o parâmetro de interesse.
- Um intervalo com coeficiente de confiança de 95% para μ é dado por:

$$IC(\mu; 95\%) = \left[\underbrace{\bar{x} - 1,96 \times \frac{s}{\sqrt{n}}}_{LI}; \underbrace{\bar{x} + 1,96 \times \frac{s}{\sqrt{n}}}_{LS}\right]$$

em que $\bar{x}$ é a média amostral;

s é o desvio padrão;

n é o tamanho da amostra;

$z = 1,96$ é o percentil de ordem $(1+\gamma)/2$ $(\gamma = 0,95)$ da distribuição normal padrão.

Margem de erro

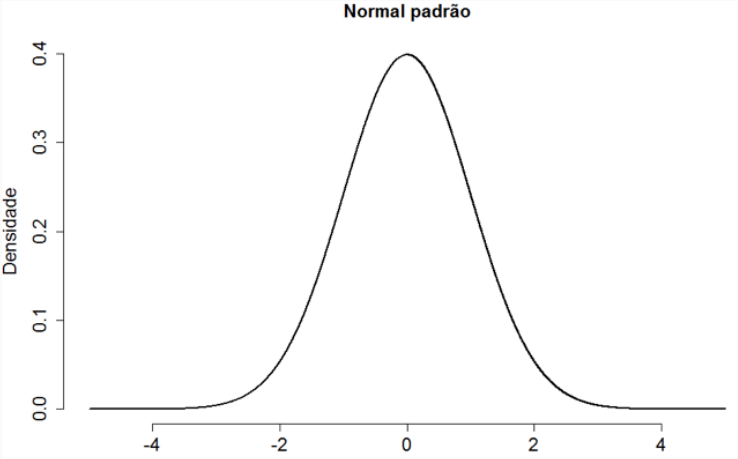
$$\epsilon = \underbrace{1,96}_{?} \times \underbrace{\frac{s}{\sqrt{n}}}_{\text{Erro padrão}}$$

Estimação intervalar

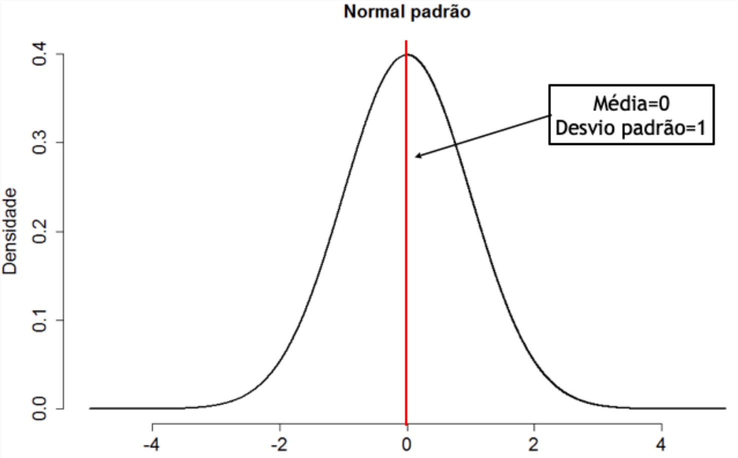
O que é uma Distribuição Normal Padrão?

Estimação intervalar

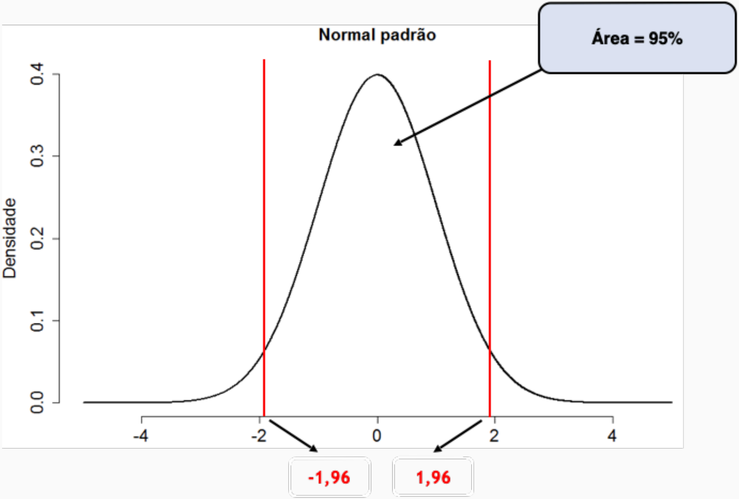
O que é uma Distribuição Normal Padrão?



Estimação intervalar



Estimação intervalar



Voltando ao Exemplo 3...

Exemplo 3: IMC dos alunos 2018 - 2019

IMC de $n = 10$ alunos: 17,8 21,8 18,0 24,8 21,7 20,3 20,7 19,3 23,7 25,9

$$\bar{x} = 21,40 \quad n = 10 \quad s = 2,75$$

$$IC(\mu, 95\%) = \left[21,40 - 1,96 \times \frac{2,75}{\sqrt{10}}; 21,40 + 1,96 \times \frac{2,75}{\sqrt{10}}\right] = [\underbrace{19,7}_{LI}; \underbrace{23,1}_{LS}]$$

Interpretação de um IC com 95% de confiança

- Se pudéssemos realizar o mesmo estudo em diferentes amostras da população, todas de tamanho n, e construir intervalos da forma $[LI; LS]$, 95% deles conteriam o verdadeiro valor do parâmetro.

- Observada a nossa amostra, o intervalo de 95% de confiança obtido pode conter o verdadeiro valor do parâmetro ou não, mas pelo exposto acima temos 95% de confiança que contenha.

Estimação intervalar

Exercício

Vejam o site...

`https://jitalo.shinyapps.io/inter/`

Estimação intervalar

Exemplo: IMC dos alunos

IMC dos 10 alunos: $\bar{x} = 21,40 \quad n = 10 \quad \sigma = 2,75$

Vamos calcular intervalos de confiança para o IMC médio (μ) com 3 diferentes níveis de confiança:

$$
\begin{aligned}
&\text{Caso 1: } 90\% \text{ de confiança} \quad \Rightarrow \quad IC(\mu, 90\%) = [20,0; 22,8] \\
&\text{Caso 2: } 95\% \text{ de confiança} \quad \Rightarrow \quad IC(\mu, 95\%) = [19,7; 23,1] \\
&\text{Caso 3: } 99\% \text{ de confiança} \quad \Rightarrow \quad IC(\mu, 99\%) = [19,1; 23,6]
\end{aligned}
$$

Estimação intervalar

Exemplo: IMC dos alunos

IMC dos 10 alunos: $\bar{x} = 21,40 \quad n = 10 \quad \sigma = 2,75$

Vamos calcular intervalos de confiança para o IMC médio (μ) com 3 diferentes níveis de confiança:

$$\begin{array}{lcl} \text{Caso 1: 90\% de confiança} & \Rightarrow & IC(\mu, 90\%) = [20,0; 22,8] \\ \text{Caso 2: 95\% de confiança} & \Rightarrow & IC(\mu, 95\%) = [19,7; 23,1] \\ \text{Caso 3: 99\% de confiança} & \Rightarrow & IC(\mu, 99\%) = [19,1; 23,6] \end{array}$$

O que falar da amplitude dos três intervalos?

Amplitude do intervalo: $LS - LI$

$$\begin{array}{lcl} \text{Caso 1: 90\% de confiança} & \Rightarrow & \text{Amplitude} = 22,8 - 20,0 = 2{,}9 \\ \text{Caso 2: 95\% de confiança} & \Rightarrow & \text{Amplitude} = 23,1 - 19,7 = 3{,}4 \\ \text{Caso 3: 99\% de confiança} & \Rightarrow & \text{Amplitude} = 23,6 - 19,1 = 4{,}5 \end{array}$$

Estimação intervalar

Exemplo: IMC dos alunos

Agora, vamos fixar o grau de confiança em 95% e calcular o intervalo de confiança para o IMC médio com 3 diferentes tamanhos amostrais.

$$
\begin{aligned}
\text{Caso 1: } n = 4 \quad &\Rightarrow \quad IC(\mu, 95\%) = [19,6; 29,0] \\
\text{Caso 2: } n = 10 \quad &\Rightarrow \quad IC(\mu, 95\%) = [19,7; 23,1] \\
\text{Caso 3: } n = 20 \quad &\Rightarrow \quad IC(\mu, 95\%) = [21,7; 24,5]
\end{aligned}
$$

Estimação intervalar

Exemplo: IMC dos alunos

Agora, vamos fixar o grau de confiança em 95% e calcular o intervalo de confiança para o IMC médio com 3 diferentes tamanhos amostrais.

$$\begin{array}{rcl} \text{Caso 1: } n = 4 & \Rightarrow & IC(\mu, 95\%) = [19,6; 29,0] \\ \text{Caso 2: } n = 10 & \Rightarrow & IC(\mu, 95\%) = [19,7; 23,1] \\ \text{Caso 3: } n = 20 & \Rightarrow & IC(\mu, 95\%) = [21,7; 24,5] \end{array}$$

O que falar da amplitude dos três intervalos?

Amplitude do intervalo: $LS - LI$

$$\begin{array}{rcl} \text{Caso 1: } n = 4 & \Rightarrow & \text{Amplitude} = 29,0 - 19,6 = 9{,}4 \\ \text{Caso 2: } n = 10 & \Rightarrow & \text{Amplitude} = 23,1 - 19,7 = 3{,}4 \\ \text{Caso 3: } n = 20 & \Rightarrow & \text{Amplitude} = 24,5 - 21,7 = 2{,}8 \end{array}$$

Inferência Estatística: Teste de hipóteses

Teste de hipóteses

O que é uma Hipótese?

Teste de hipóteses

O que é uma Hipótese?

Hipótese

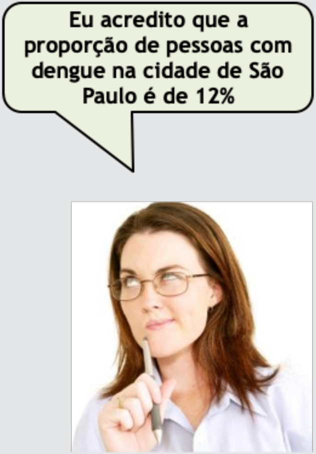
É uma determinada afirmação sobre a população de interesse.
Usualmente sobre um parâmetro dessa população.

- Desejamos saber se os resultados provenientes de uma amostra contrariam ou não tal afirmação.

Teste de hipóteses

Exemplo:

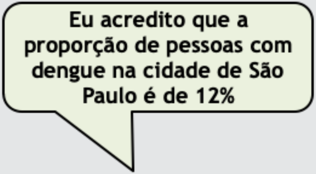
Quem é o parâmetro da afirmação?

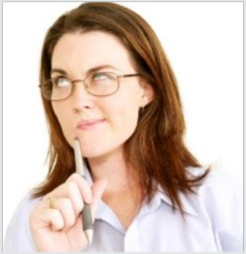


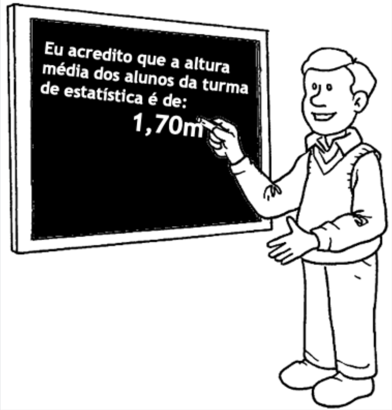
Teste de hipóteses

Exemplo:

Quem é o parâmetro da afirmação?







Teste de hipóteses

Estimação

- Qual é a proporção de pacientes da Obstetrícia com diabetes?
- Qual é a proporção de adultos acima do peso?
- Qual é a proporção de pessoas com Dengue na cidade de São Paulo?
- Qual a proporção de votos que o candidato A terá na próxima eleição?
- Qual é a altura média dos alunos de Estatística?

Teste de Hipóteses

- A proporção de pacientes com diabetes é de 20%?
- A proporção de adultos acima do peso é de 18%?
- A proporção de pessoas com Dengue é menor do que 12%?
- O candidato A vencerá a eleição?
- A altura média dos alunos é igual a 1,70m?

Teste de hipóteses

Definição das hipóteses

Hipótese nula (H0)

Afirmação sobre o parâmetro, geralmente relacionado a um valor de referência, ou a uma especificação padrão ou histórica.

Hipótese alternativa (H1)

A hipótese que aceitamos quando a hipótese nula é rejeitada.
Afirmação que suspeitamos ser verdadeira.

Observações:

- O termo Hipótese Nula é usado para ver se alguma hipótese estabelecida inicialmente pode ser rejeitada ou não.
- A ideia é exatamente o que é feito em processos criminais, onde um acusado (réu) é dito ser inocente até que se prove o contrário.
- A pressuposição de inocência é uma hipótese nula.

Definição das hipóteses

Exemplo 1: Altura média dos alunos de Estatística

Queremos avaliar se a altura média da turma é menor do que 1,70m

Hipótese nula: altura média é maior ou igual a 1,70m

contra

Hipótese alternativa: altura média é inferior a 1,70m

Seja μ a média populacional. Estatisticamente:

$$H_0: \mu \geq 1,70\text{m}$$
$$H_1: \mu < 1,70\text{m}$$

Definição das hipóteses: teste bilateral

Teste bilateral

Se as hipóteses de um teste de hipóteses são da forma:

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

em que μ_0 é o valor do parâmetro sob nossa conjectura.

Exemplo 1: Altura média dos alunos de Estatística

$$H_0: \mu = 1,70\text{m}$$
$$H_1: \mu \neq 1,70\text{m}$$

em que $\mu_0 = 1,70$ é a constante que estou testando (nossa conjectura).

Definição das hipóteses: testes unilaterais

Teste unilateral à esquerda

Se as hipóteses de um teste de hipóteses são da forma:

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

em que μ_0 é o valor do parâmetro sob nossa conjectura.

Exemplo 1: Altura média dos alunos de Estatística

$$H_0\colon\ \mu \geq 1,70\text{m}$$
$$H_1\colon\ \mu < 1,70\text{m}$$

em que $\mu_0 = 1,70$m.

Definição das hipóteses: testes unilaterais

Teste unilateral à direita
Se as hipóteses de um teste de hipóteses são da forma:

$$H_0\colon\ \mu \leq \mu_0$$
$$H_1\colon\ \mu > \mu_0$$

em que μ_0 é o valor do parâmetro sob nossa conjectura.

Exemplo 1: Altura média dos alunos de Estatística

$$H_0\colon\ \mu \leq 1,70\text{m}$$
$$H_1\colon\ \mu > 1,70\text{m}$$

em que $\mu_0 = 1,70$m.

Como decidimos?

Teste de hipóteses: Região de rejeição

Testar uma hipótese estatística

é estabelecer uma regra que nos permita, com base na informação de uma amostra, decidir pela rejeição ou não de H_0.

Teste de hipóteses: Região de rejeição

Testar uma hipótese estatística

é estabelecer uma regra que nos permita, com base na informação de uma amostra, decidir pela rejeição ou não de H_0.

Região de rejeição ou região critica (RC)

Conjunto de valores observados na amostra que levam à rejeição da hipótese nula H_0.

Será que a decisão está correta?

Teste de hipóteses: Erros

Quais são os erros associados a nossa decisão?

Decisão	Situação	
	H0: verdadeira	H0: Falsa
Rejeitar H0	*Erro Tipo I*	*Decisão correta*
Não Rejeitar H0	*Decisão correta*	*Erro Tipo II*

Teste de hipóteses: Erros

Quais são os erros associados a nossa decisão?

Decisão	Situação	
	H0: verdadeira	H0: Falsa
Rejeitar H0	*Erro Tipo I*	*Decisão correta*
Não Rejeitar H0	*Decisão correta*	*Erro Tipo II*

Erro Tipo I
Rejeita-se a hipótese nula H_0 quando H_0 é verdadeira.

Erro Tipo II
Não rejeita a hipótese nula H_0 quando H_0 é falsa.

Teste de hipóteses: Erros

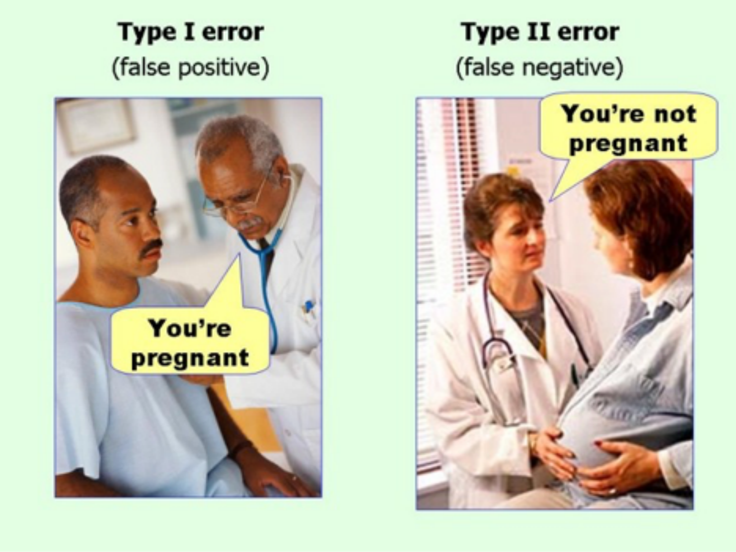
Quais são os erros associados a nossa decisão?

Exemplo 1: Altura média dos alunos de Estatística

$$H_0\text{: } \mu = 1,70\text{m}$$
$$H_a\text{: } \mu \neq 1,70\text{m}$$

Erro Tipo I: Concluir que a altura média não é 1,70m quando na verdade é 1,70m.
Erro Tipo II: Concluir que a altura média é 1,70m quando na verdade não é.

Teste de hipóteses: Erros



Teste de hipóteses: Erros

Exemplo 2: Diabetes gestacional

$\boldsymbol{H_0}$: A proporção de gestantes com diabetes gestacional é menor o igual a 20%.

$\boldsymbol{H_1}$: A proporção de gestantes com diabetes gestacional é maior que 20%.

$$H_0 = p \leq 0,20$$
$$H_1 = p > 0,20$$

Erro Tipo I (Rejeitar H_0 quando H_0 é verdadeira):

Concluir que a proporção de diabetes gestacional é maior que 20% quando na verdade não é.

Erro Tipo II (Não rejeitar H_0 quando H_0 é falsa):

Concluir que a proporção de diabetes gestacional é 20% quando na verdade é maior que 20%.

Teste de hipóteses: Probabilidades de Erros

Dados os erros, definimos suas probabilidades

$\alpha = $ P(Erro Tipo I) $=$ P(Rejeitar H_0 quando H_0 é verdadeira)

$\boldsymbol{\alpha}$: **Nível de significância do teste**

$\beta = $ P(Erro Tipo II) $=$ P(Não rejeitar H_0 quando H_0 é falsa)

$1 - \beta = $ P(Rejeitar H_0 quando H_0 é falsa)

$\mathbf{1 - \boldsymbol{\beta}}$: **Poder do teste**

Teste de hipóteses: Probabilidades de Erros

Dados os erros, definimos suas probabilidades

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

$\boldsymbol{\alpha}$: **Nível de significância do teste**

$\beta = \text{P(Erro Tipo II)} = \text{P(Não rejeitar } H_0 \text{ quando } H_0 \text{ é falsa)}$

$1 - \beta = \text{P(Rejeitar } H_0 \text{ quando } H_0 \text{ é falsa)}$

$\mathbf{1 - \boldsymbol{\beta}}$: **Poder do teste**

Observações:

- α e β tem uma relação inversa.
- Em geral, podemos controlar (fixar) apenas um dos erros.
- Usualmente o nível de significancia α é fixado em 5%, no entanto, podem se fixar outros valores (1%, 10%).

Teste de hipóteses: Procedimento

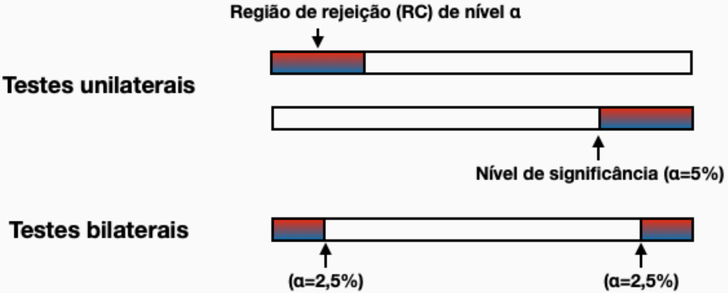
Como decidir se rejeita ou não H_0?

- A decisão basicamente é tomada ao comparar uma quantidade da amostra, chamada de **estatística do teste**, com a região de rejeição ou região critica (RC) da hipótese nula (H_0).
- A região de rejeição é construída com base no nível de significância fixado α, no *"sinal"* da hipótese alternativa e na teoria Estatística adequada para o teste em questão.

Teste de hipóteses: Procedimento

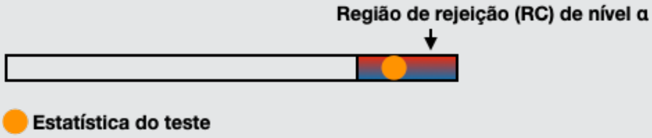
Como decidir se rejeita ou não H_0?

- A decisão basicamente é tomada ao comparar uma quantidade da amostra, chamada de **estatística do teste**, com a região de rejeição ou região critica (RC) da hipótese nula (H_0).
- A região de rejeição é construída com base no nível de significância fixado α, no *"sinal"* da hipótese alternativa e na teoria Estatística adequada para o teste em questão.



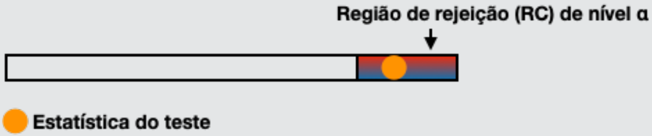
Teste de hipóteses: Decisão

Situação 1:



Teste de hipóteses: Decisão

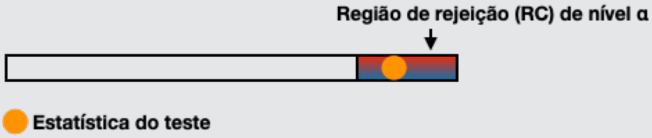
Situação 1:



Decisão? $\Longrightarrow$ Rejeitar H_0

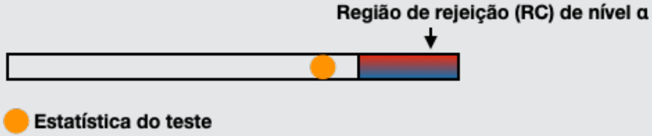
Teste de hipóteses: Decisão

Situação 1:



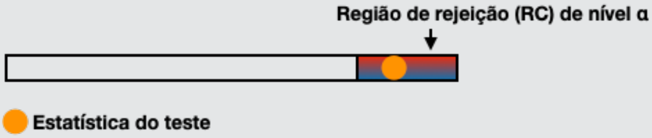
Decisão? $\Longrightarrow$ Rejeitar H_0

Situação 2:



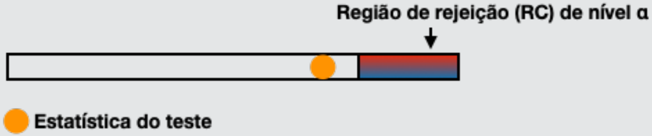
Teste de hipóteses: Decisão

Situação 1:



Decisão? $\Longrightarrow$ Rejeitar H_0

Situação 2:



Decisão? $\Longrightarrow$ Não Rejeitar H_0

Teste de hipóteses

Exemplo 1: Altura dos alunos de Estatística

Queremos testar a hipótese de que a altura média dos alunos é diferente de 170cm.

$$H_0:\ \mu = 170 \text{ cm}$$
$$H_1:\ \mu \neq 170 \text{ cm}$$

O que observamos a partir da amostra de n=20 alunos:

n	Média	Desvio padrão	Intervalo 95% de confiança para média
20	172 cm	3,1 cm	[170,55 ; 173,45]

Teste de hipóteses

Exemplo 1: Altura dos alunos de Estatística

Queremos testar a hipótese de que a altura média dos alunos é diferente de 170cm.

$$H_0:\ \mu = 170 \text{ cm}$$
$$H_1:\ \mu \neq 170 \text{ cm}$$

O que observamos a partir da amostra de n=20 alunos:

n	Média	Desvio padrão	Intervalo 95% de confiança para média
20	172 cm	3,1 cm	[170,55 ; 173,45]

Antes de realizarmos o teste de hipóteses.
O que se pode concluir ao analisar o IC 95%?

Teste de hipóteses

Exemplo 1: Altura dos alunos de Estatística

Cálculo da estatística do teste (T)

n	Média	Desvio padrão (s)
20	172 cm	3,1 cm

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{172 - 170}{\frac{3,1}{20}} = 2,88$$

Exemplo 1: Altura dos alunos de Estatística

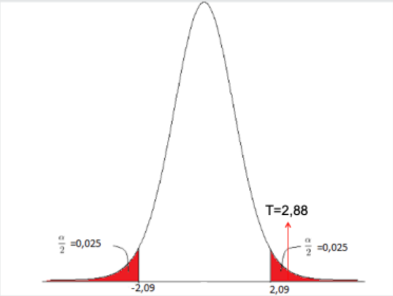
Cálculo da estatística do teste (T)

n	Média	Desvio padrão (s)
20	172 cm	3,1 cm

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{172 - 170}{\frac{3,1}{20}} = 2,88$$

Região de rejeição (RC)

$\alpha = 0,05$



Teste de hipóteses

Exemplo 1: Altura dos alunos de Estatística

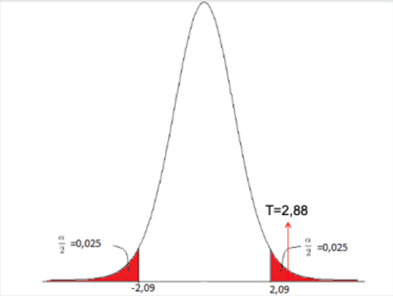
Cálculo da estatística do teste (T)

n	Média	Desvio padrão (s)
20	172 cm	3,1 cm

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{172 - 170}{\frac{3,1}{20}} = 2,88$$

Região de rejeição (RC)

$\alpha = 0,05$



Decisão? $\Longrightarrow$ Rejeitar H_0

A altura média dos alunos é diferente de 170cm.

Teste de hipóteses: Valor p

Nível descritivo = Valor *p*

é o menor nível de significância para o qual o resultado observado é significante, ou seja, conduz à rejeição da hipótese nula H_0.

Teste de hipóteses: Valor p

Nível descritivo = Valor *p*

é o menor nível de significância para o qual o resultado observado é significante, ou seja, conduz à rejeição da hipótese nula H_0.

Nível descritivo = Valor *p*

Se o valor p for menor que nível de significância α, então a região de rejeição de nível α contém o valor observado e, portanto, rejeitamos H_0.

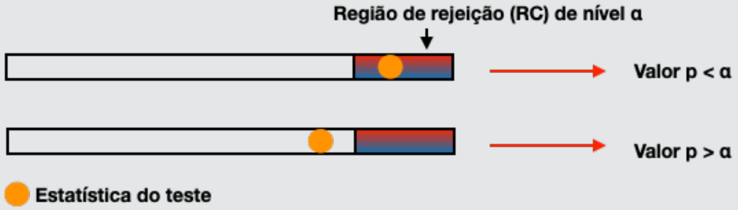
Portanto, para um α fixado,

$$p < \alpha \Longrightarrow \text{Rejeitamos } H_0.$$
$$p > \alpha \Longrightarrow \text{Não rejeitamos } H_0.$$

Teste de hipóteses: Valor p

Nível descritivo = Valor *p*

Relação valor p com procedimento do teste de hipóteses



Portanto, ao invés de calcular a região de rejeição e estatística ou região crítica do teste (RC), posso calcular apenas o valor p e comparar com nível de significância (fixado).

Teste de hipóteses: Valor p

Nível descritivo = Valor *p*

Observações:

- O teste de hipóteses é construído assumindo que a hipótese nula é verdadeira.
- O valor p pode ser interpretado como a probabilidade de ter efeito maior que o observado sob a hipótese nula (H_0) verdadeira.

Teste de hipóteses: Valor p

Nível descritivo = Valor *p*

Observações:

- O teste de hipóteses é construído assumindo que a hipótese nula é verdadeira.
- O valor p pode ser interpretado como a probabilidade de ter efeito maior que o observado sob a hipótese nula (H_0) verdadeira.

Exemplo 1: Altura média dos alunos de Estatística

- O p-valor é calculado ao assumir que a altura média é de 170cm.
- Essa é uma característica contra-intuitiva dos testes: Se você quer provar que a altura média é diferente de 170cm, você faz isso ao mostrar que os dados são inconsistentes com a altura média de 170cm.

Teste de hipóteses: Valor p

Exemplo 1: Altura média dos alunos de Estatística

Testando a hipótese por meio da medida de evidência (valor p).

$$H_0\text{: } \mu = 170\text{cm}$$
$$H_a\text{: } \mu \neq 170\text{cm}$$

n	Média	Desvio padrão	Intervalo 95% de confiança para média	valor p
20	172 cm	3,1 cm	[170,55 ; 173,45]	0,009

Teste de hipóteses: Valor p

Exemplo 1: Altura média dos alunos de Estatística

Testando a hipótese por meio da medida de evidência (valor p).

$$H_0: \mu = 170\text{cm}$$
$$H_a: \mu \neq 170\text{cm}$$

n	Média	Desvio padrão	Intervalo 95% de confiança para média	valor p
20	172 cm	3,1 cm	[170,55 ; 173,45]	0,009

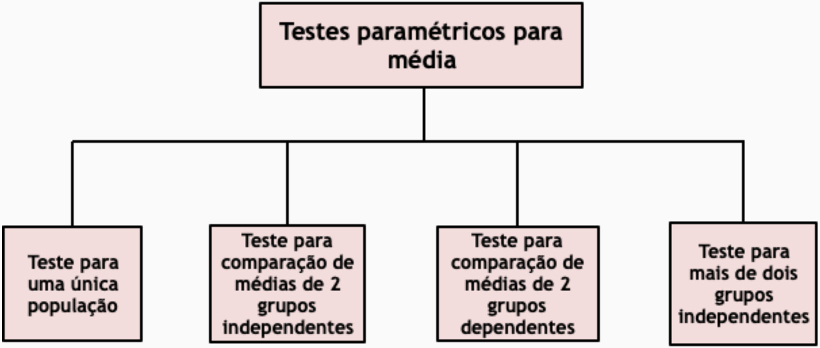
O que se pode concluir ao considerar α=5% ?

valor $p < \alpha \implies$ Rejeitamos H_0

Ou seja, segundo a evidência amostral, há indícios para rejeitar a hipótese de que a altura média é 170cm.

Teste de hipóteses para a média

Teste de hipóteses para a média



Teste de hipóteses para média de uma população

Teste para média de 1 população

- É o teste que realizamos no exemplo de testar se a altura média (parâmetro) dos alunos de Estatística (população) é 170cm (valor de referência).
- O teste de hipóteses para a média (única) de uma população é conhecido como **teste t de amostra única.**

Teste de hipóteses para média de uma população

Teste para média de 1 população

- É o teste que realizamos no exemplo de testar se a altura média (parâmetro) dos alunos de Estatística (população) é 170cm (valor de referência).
- O teste de hipóteses para a média (única) de uma população é conhecido como **teste t de amostra única.**

Exemplo 1: Idade gestacional do parto

- Um obstetra brasileiro acredita que a idade gestacional média do parto de gestantes brasileiras com determinada comorbidade é menor que a média americana, conhecida ser de 38 semanas.
- Como formular as hipóteses para testar a conjectura do obstetra?
- Qual parâmetro? Qual valor de referência? Qual o sinal da hipótese alternativa?

Teste de hipóteses para média de uma população

Exemplo 1: Idade gestacional do parto

Seja μ a média da idade gestacional do parto da população brasileira.

$$H_0 : \mu \geq 38 \text{ semanas.}$$
$$H_1 : \mu < 38 \text{ semanas.}$$

Um estudo foi realizado no Brasil com uma amostra de $n = 100$ gestantes com a determinada comorbidade.

n	Média	Desvio padrão
100	36,5	2,9

Do teste t de única amostra: valor $p < 0,001$

O que se pode concluir?