INSTITUTO DO CÂNCER DO ESTADO DE SÃO PAULO

OCTAVIO FRIAS DE OLIVEIRA

Mestrado Profissional em Inovação e Avaliação de Tecnologias em Cancerologia

MCM5940 Métodos Estatísticos Aplicados na Área de Saúde

Aula 5

Dra. Rossana V. Mendoza López

Conteúdo da Aula 5

- **Intervalos de confiança: definição e interpretação**

Intervalo de confiança para a média

Intervalo de confiança para a proporção

Intervalo de confiança para a diferença de duas médias

Intervalo de confiança para a diferença de duas proporções

- **Testes de hipóteses**

Definição de teste de hipóteses: hipótese nula (H_0), hipótese alternativa (H_1)

Nível de significância e valor de p (*p-value*)

Exercício prático. Lista de exercícios

Aula 5

Estimação por intervalos

Aula 5

Intervalo de confiança

É um intervalo calculado a partir da amostra e que muito provavelmente contém o parâmetro populacional

Estimativa por intervalo

Aula 5

Intervalo de confiança → tentamos “cercar” um parâmetro

Teste de hipóteses → hipótese em quanto ao valor do parâmetro

Amostra → Aceitar/Rejeitar a hipótese

Média medida de tendência central

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Estimação pontual

Aula 5

Proporção Frequência relativa

$$\textbf{proporção} = \frac{a}{n}$$

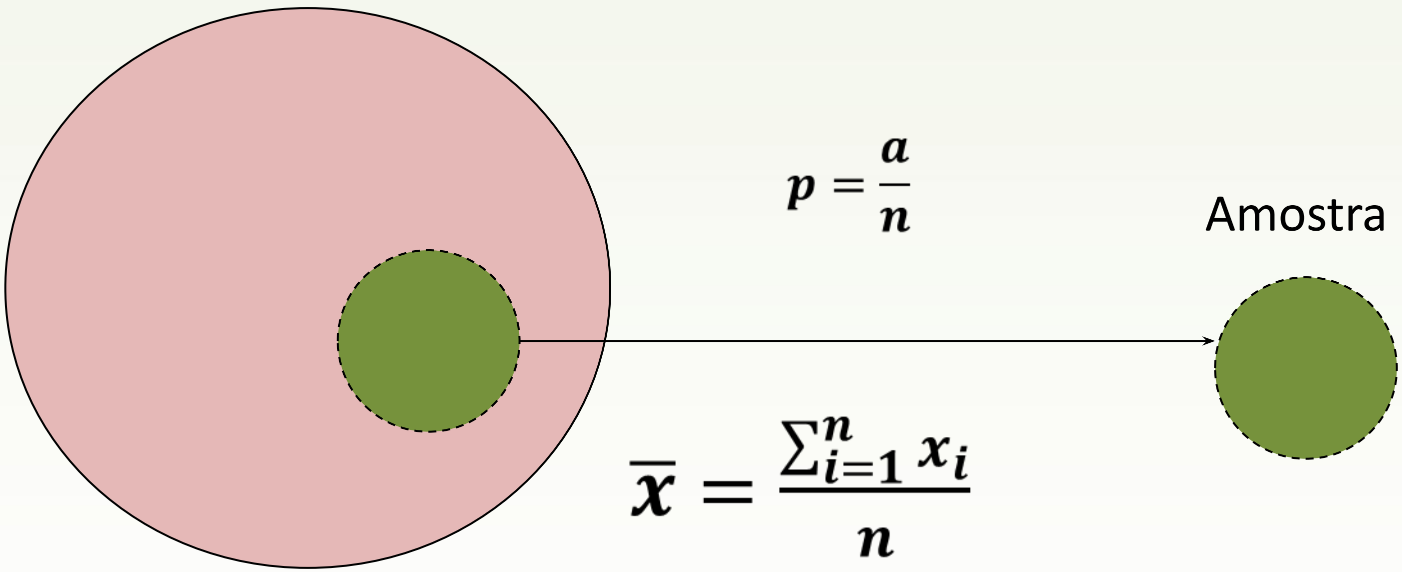
Estimação pontual

Aula 5

População de estudo

$$p = \frac{a}{n}$$

Amostra



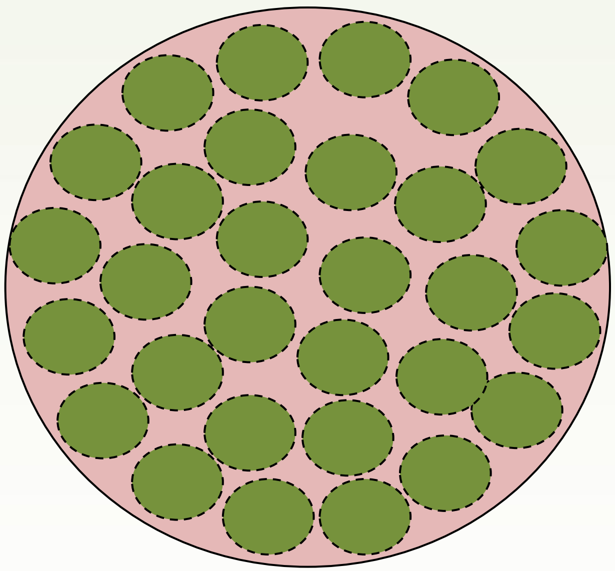
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$\boldsymbol{\theta} \rightarrow ?$

$\hat{\boldsymbol{\theta}}$ Valor conhecido

Aula 5

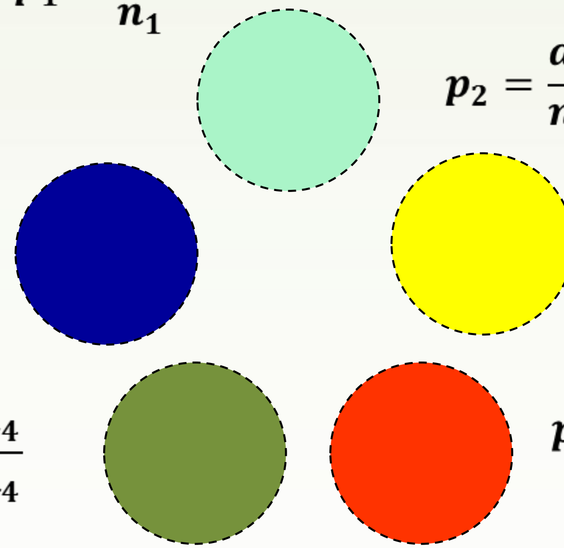
População de estudo



$$p_1 = \frac{a_1}{n_1}$$

$$p_2 = \frac{a_2}{n_2}$$

$$p_5 = \frac{a_5}{n_5}$$



$$p_4 = \frac{a_4}{n_4}$$

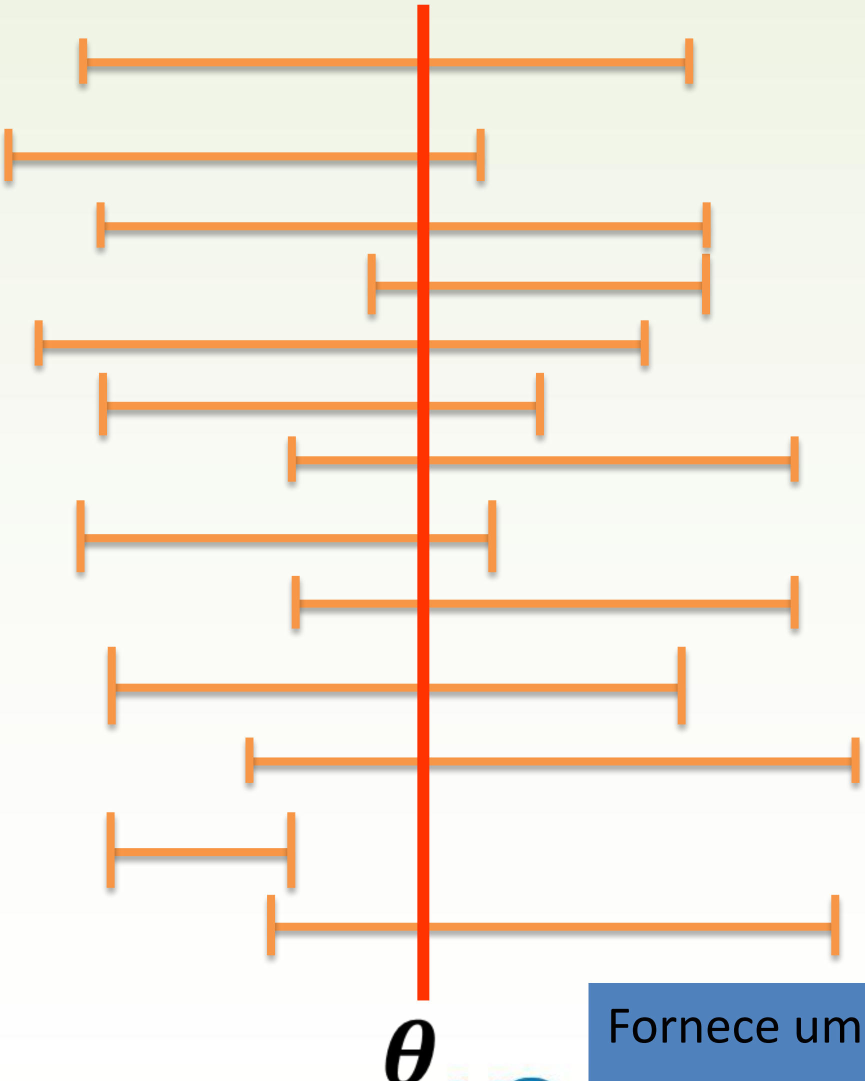
$$p_3 = \frac{a_3}{n_3}$$

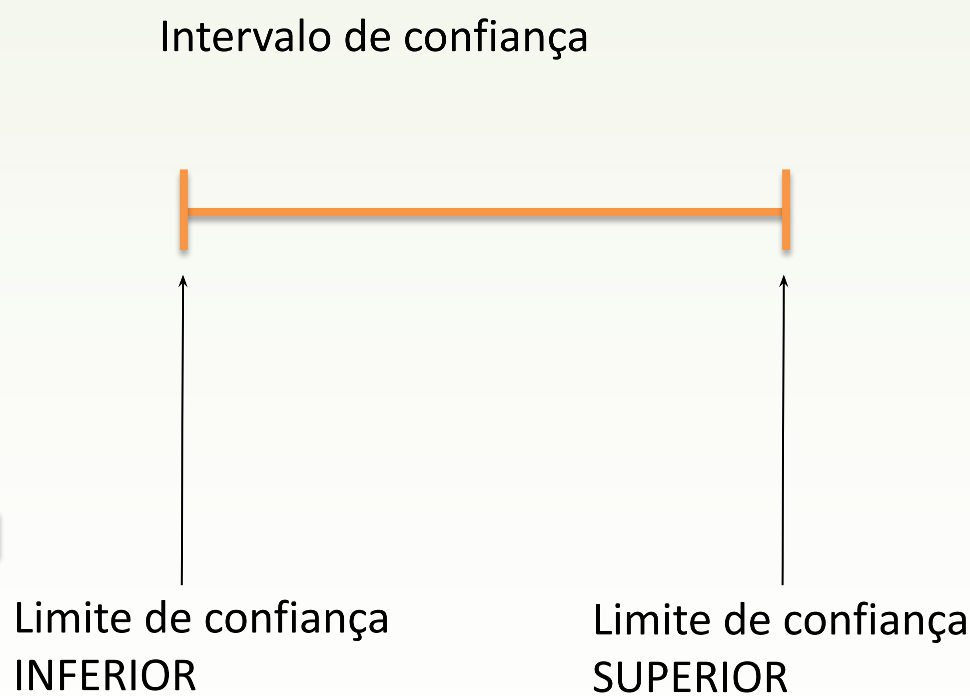
$$\boldsymbol{\theta} \rightarrow ?$$

$$\widehat{\boldsymbol{\theta}}$$ Valor conhecido

$$p_i = \frac{a_i}{n_i}$$

Aula 5





Fornece um intervalo de valores razoável no qual se presume que esteja o parâmetro de interesse com certo grau de confiança

Aula 5

Cálculo do intervalo de confiança de uma estatística

Em geral

Intervalo de confiança = Estatística da amostra ± Percentil crítico na distribuição de probabilidades x Erro padrão

Aproximadamente

IC95% = estatística ± 2 EP

Aula 5

Cálculo do intervalo de confiança de uma estatística

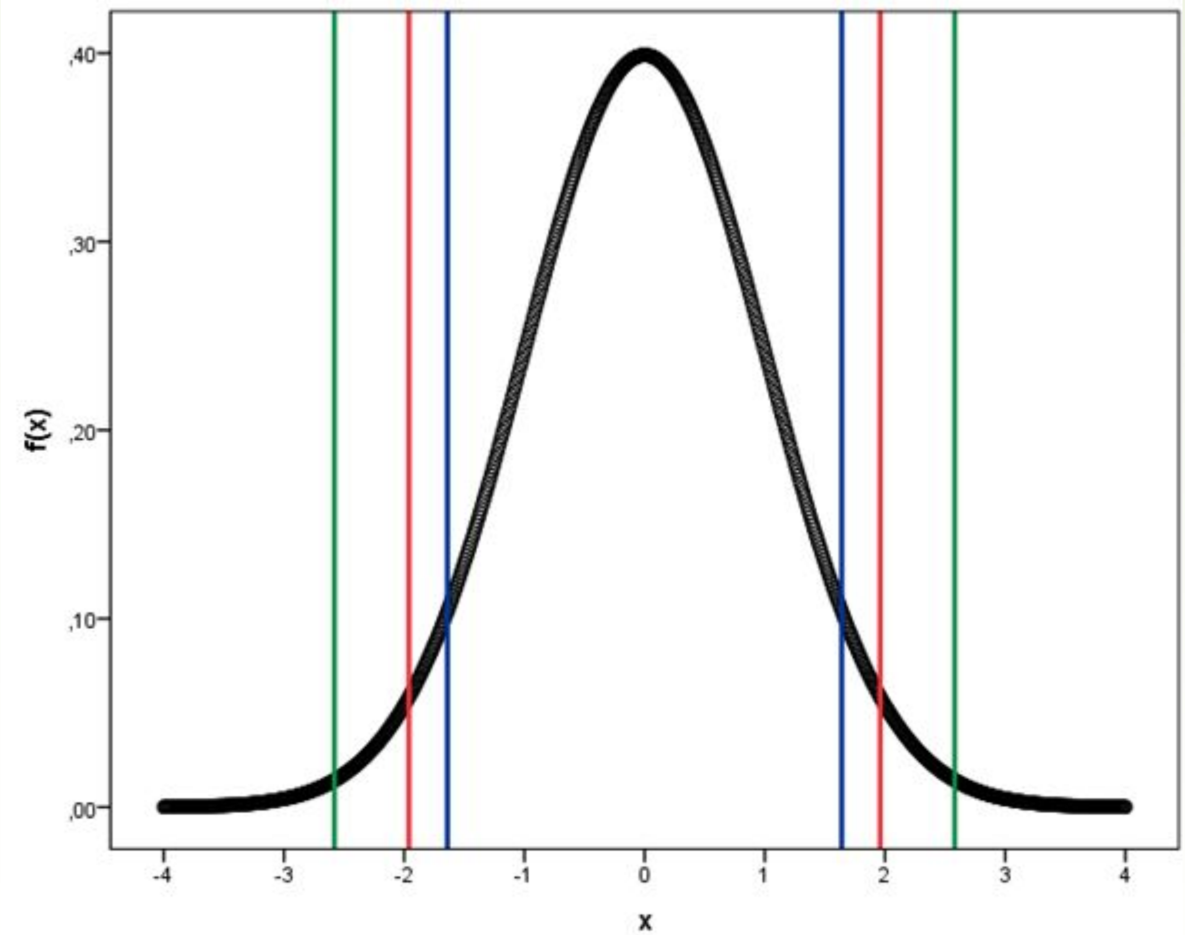
Em geral

Intervalo de confiança = Estatística da amostra ± Percentil crítico na distribuição de probabilidades x Erro padrão

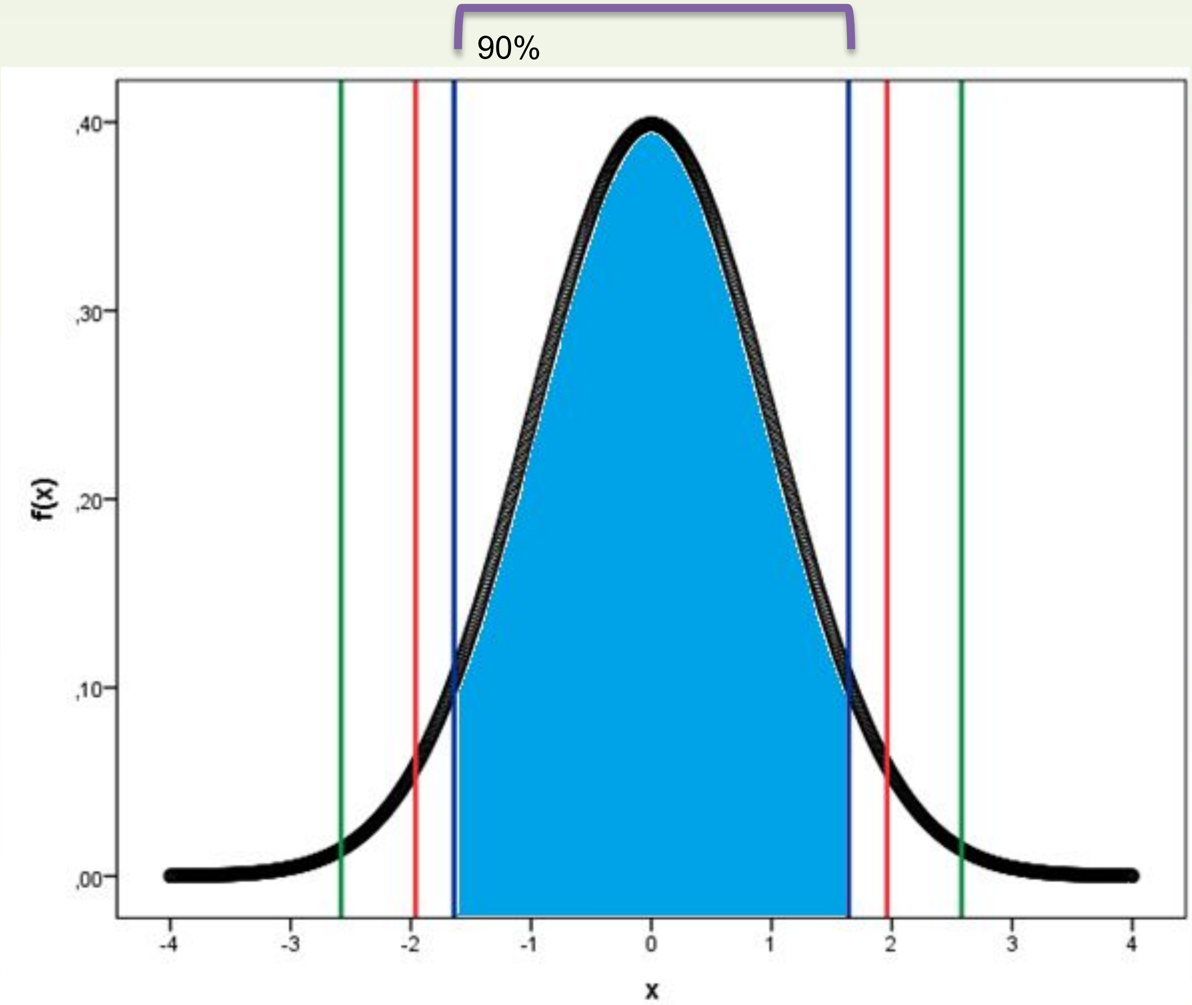
Cuidado com a interpretação!

Se espera com um 95% de confiança que o intervalo [... ; ...] contenha o verdadeiro valor do parâmetro

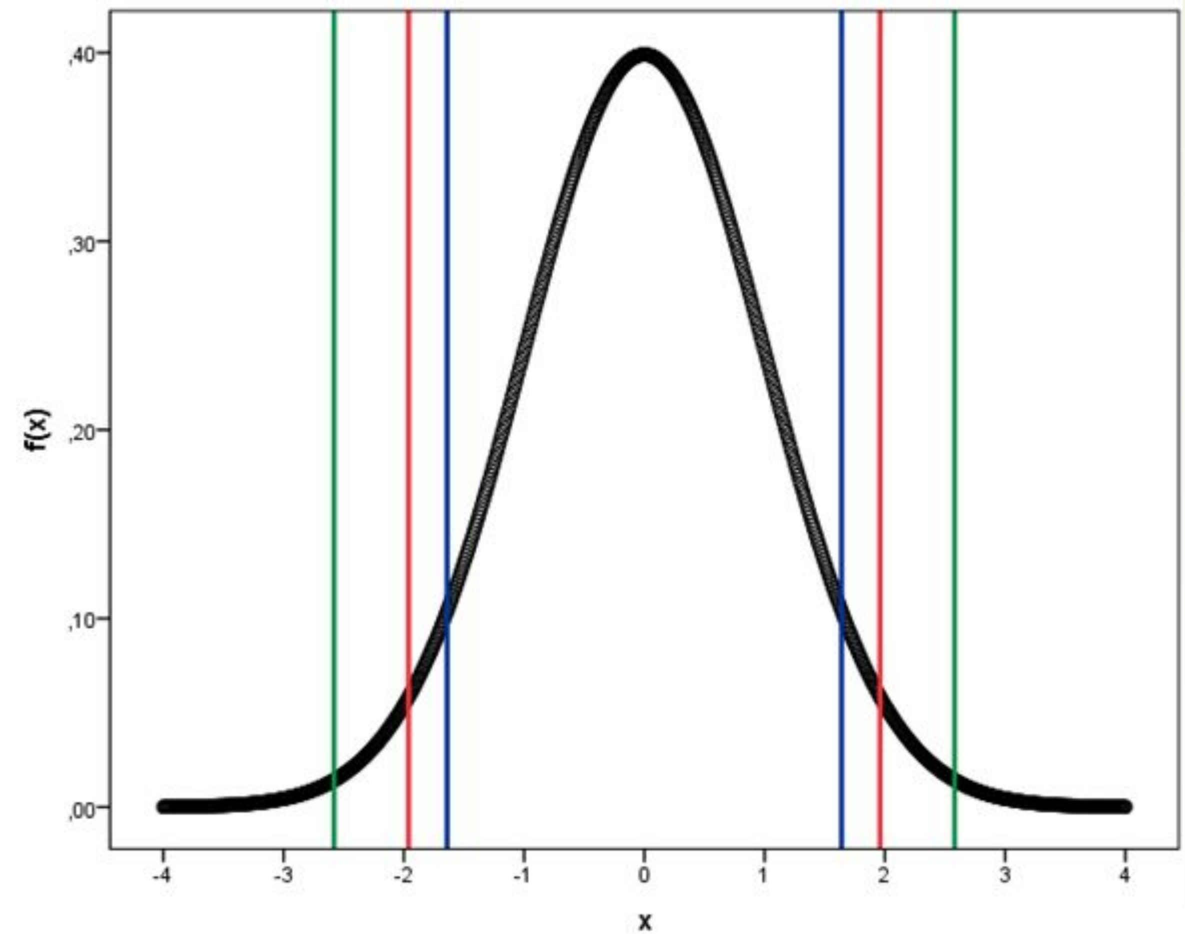
Aula 5



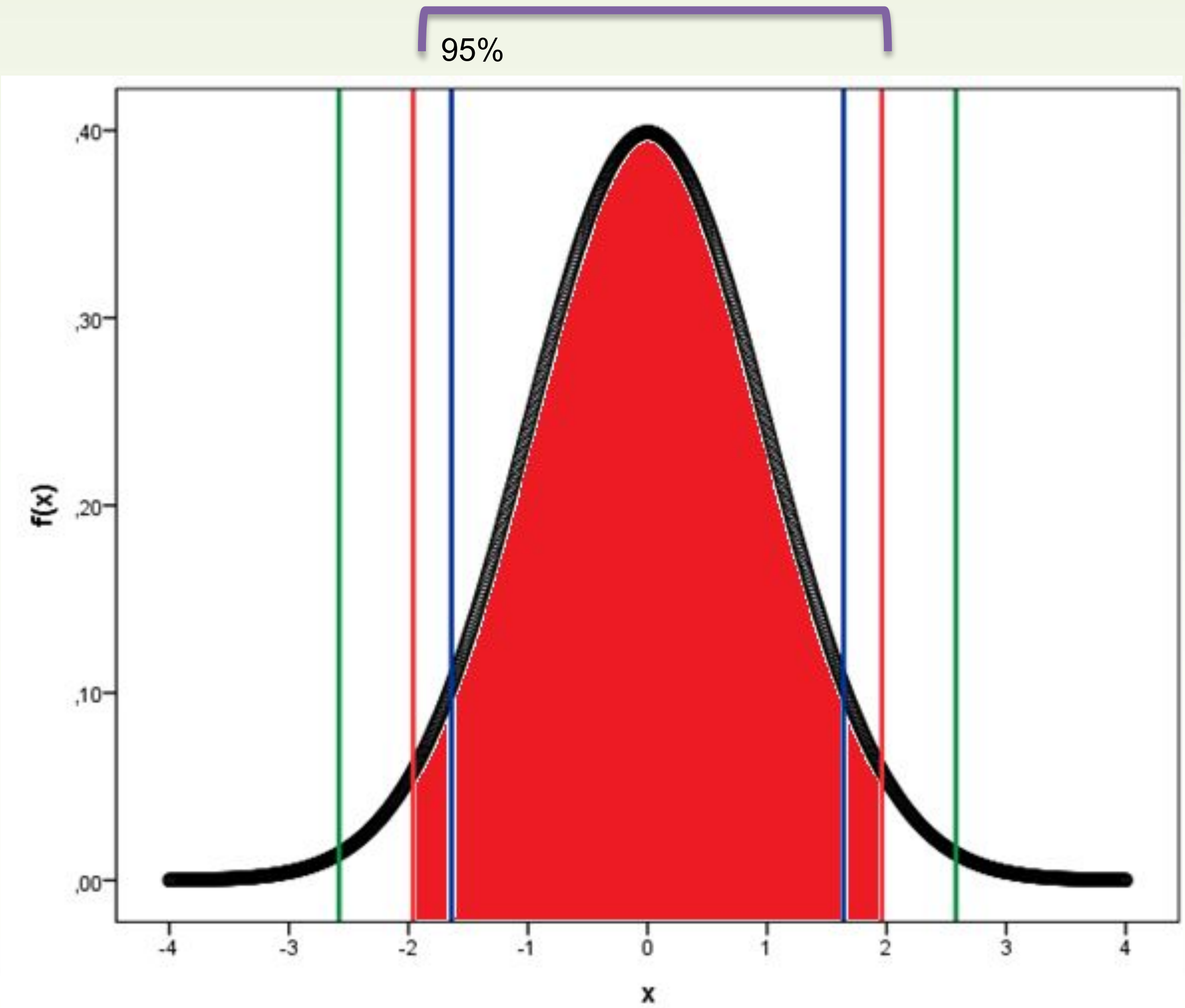
Aula 5



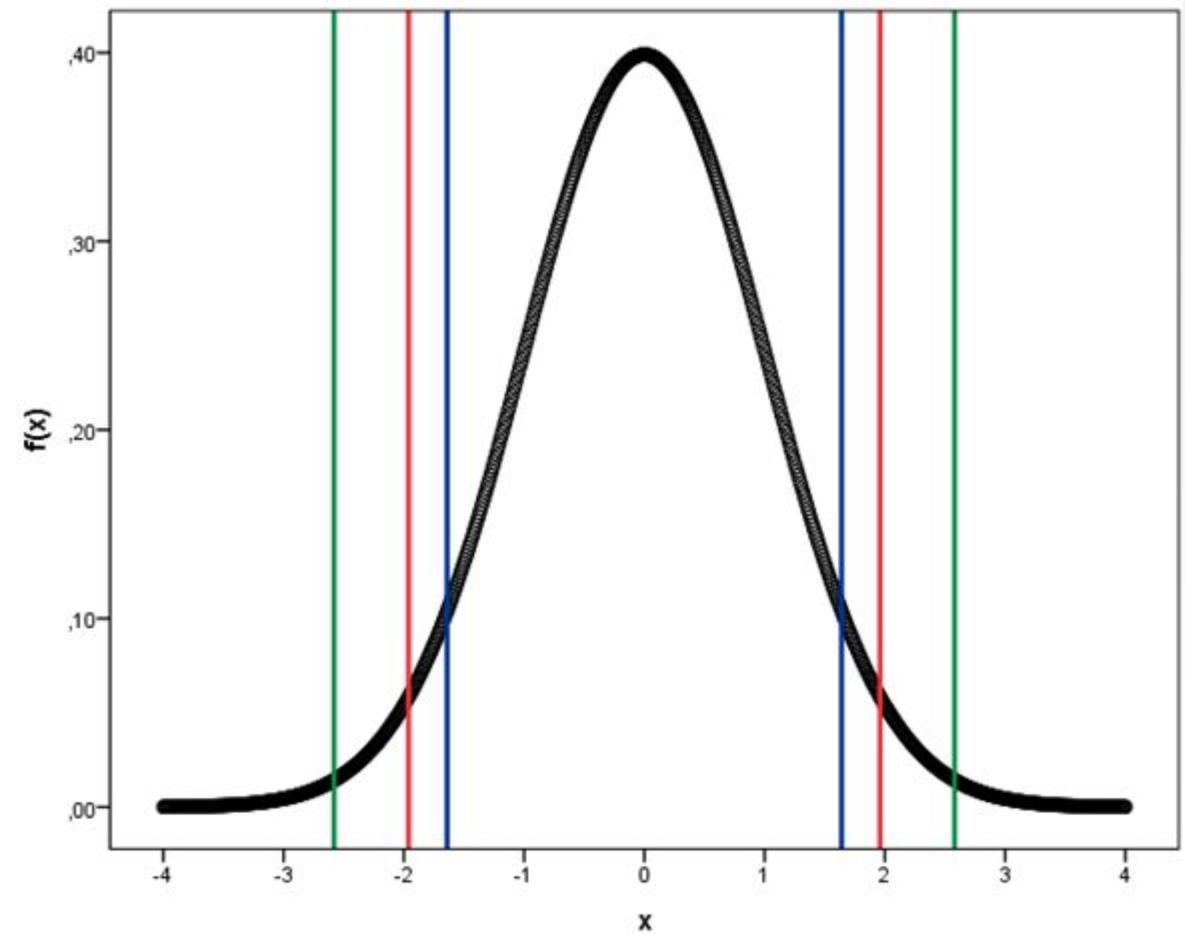
Aula 5



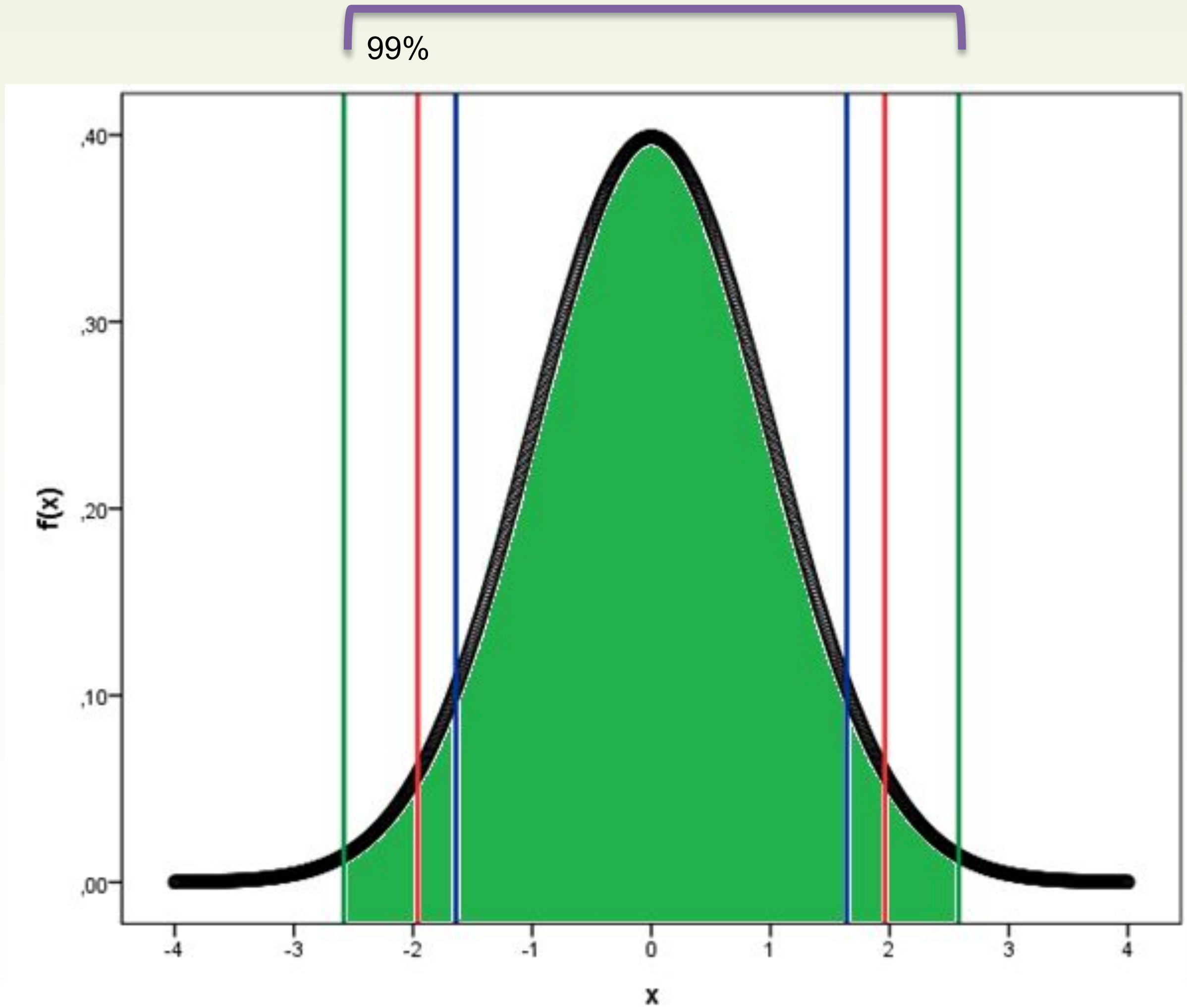
Aula 5



Aula 5



Aula 5

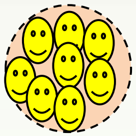


Aula 5

IC(μ)

O que precisamos para calcular um intervalo de confiança para uma média?

Amostra



Estimativa de μ

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Estimativa de erro padrão

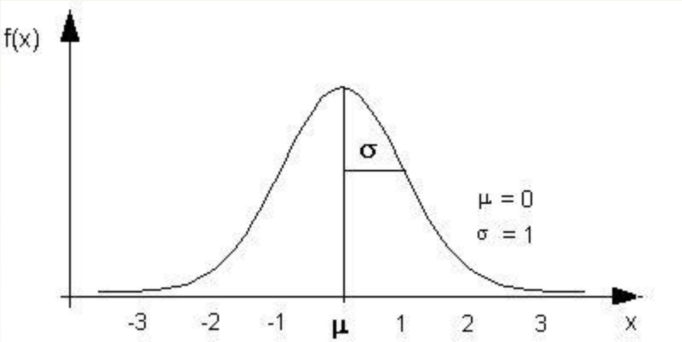
$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$s = \sqrt{s^2}$$

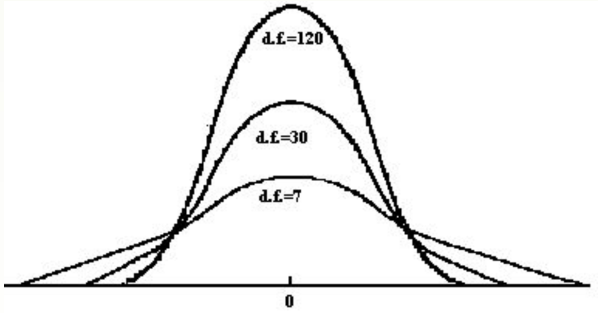
Aula 5

IC(μ)

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$



$$IC(\mu) = \bar{x} \pm t_{n-1}\frac{s}{\sqrt{n}}$$



Aula 5

IC(μ)

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

$$IC(\mu) = \bar{x} \pm t_{n-1}\frac{s}{\sqrt{n}}$$

$\bar{x}$=média

IC(μ)

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

erro padrão

$$IC(\mu) = \bar{x} \pm t_{n-1}\frac{s}{\sqrt{n}}$$

Lembrando que:
Erro padrão=desvio padrão/√n

IC(μ)

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

distribuição z

Quando n>120, z ~ t

$$IC(\mu) = \bar{x} \pm t_{n-1}\frac{s}{\sqrt{n}}$$

distribuição t

Exemplo

250 crianças com problemas de desnutrição participaram de um programa para recuperação e educação nutricional em São Paulo

A média do tempo de permanência no programa foi de 39,65 meses, e o desvio padrão de 30,17 meses

Calcular o IC95% para o tempo médio de permanência

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \qquad\qquad n = 250$$

$$z = 1,96$$

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

Instituto do Câncer do Estado de São Paulo Octavio Frias de Oliveira

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \quad n = 250$$

$$z = 1,96$$

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

				Limite inferior	Limite superior
39,65	30,17				

Aula 5

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \quad n = 250$$

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

$$z = 1,96$$

				Limite inferior	Limite superior
39,65	30,17	30,17/√250			

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \quad n = 250$$

$$z = 1,96$$

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

				Limite inferior	Limite superior
39,65	30,17	30,17/√250	1,96		

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \quad n = 250$$

$$z = 1,96$$

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

				Limite inferior	Limite superior
39,65	30,17	30,17/√250	1,96	39,65-1,96*1,91	

Aula 5

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \quad n = 250$$

$$z = 1,96$$

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

				Limite inferior	Limite superior
39,65	30,17	30,17/√250	1,96	35,91	

Aula 5

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \quad n = 250$$

$$z = 1,96$$

$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

				Limite inferior	Limite superior
39,65	30,17	30,17/√250	1,96	35,91	39,65+1,96*1,91

Exemplo

$$\bar{x} = 39,65$$

$$s = 30,17 \quad n = 250$$

$$z = 1,96$$

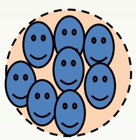
$$IC(\mu) = \bar{x} \pm z.\frac{s}{\sqrt{n}}$$

				Limite inferior	Limite superior
39,65	30,17	30,17/√250	1,96	35,91	43,40

Aula 5

IC(p)

O que precisamos para calcular um intervalo de confiança para uma proporção?

Amostra 

Estimativa de p $$\hat{p} = \frac{a}{n}$$

Estimativa de erro padrão $$EP(\hat{p}) = \sqrt{\frac{p\,(1-p)}{n}}$$

IC(p)

$$IC(p) = p \pm z\sqrt{\frac{p\,(1-p)}{n}}$$

p=proporção

IC(p)

$$IC(p) = p \pm z\sqrt{\frac{p\,(1-p)}{n}}$$

Distribuição z

Para 90% z=1,64

Para 95% z=1,96

Para 99% z=2,58

Aula 5

IC(p)

$$IC(p) = p \pm z\sqrt{\frac{p\,(1-p)}{n}}$$

erro padrão

Variação amostral considerando o tamanho da amostra

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	$\sqrt{[p(1-p)/n]}$	$z_{95\%}$	Limite inferior	Limite superior
G/G	253						
G/C	150						
C/C	33						

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	√[p(1-p)/n]	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58					
G/C	150	0,34					
C/C	33	0,08					

Aula 5

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	√[p(1-p)/n]	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58	0,42				
G/C	150	0,34	0,66				
C/C	33	0,08	0,92				

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	√[p(1-p)/n]	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58	0,42	0,0236			
G/C	150	0,34	0,66	0,0227			
C/C	33	0,08	0,92	0,0130			

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	√[p(1-p)/n]	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58	0,42	0,0236	1,96		
G/C	150	0,34	0,66	0,0227	1,96		
C/C	33	0,08	0,92	0,0130	1,96		

Aula 5

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	√[p(1-p)/n]	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58	0,42	0,0236	1,96	0,58-1,96*0,0236	
G/C	150	0,34	0,66	0,0227	1,96	0,34-1,96*0,0227	
C/C	33	0,08	0,92	0,0130	1,96	0,08-1,96*0,0130	

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	$\sqrt{[p(1-p)/n]}$	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58	0,42	0,0236	1,96	0,53	
G/C	150	0,34	0,66	0,0227	1,96	0,30	
C/C	33	0,08	0,92	0,0130	1,96	0,05	

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	√[p(1-p)/n]	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58	0,42	0,0236	1,96	0,53	0,58+1,96*0,0236
G/C	150	0,34	0,66	0,0227	1,96	0,30	0,34+1,96*0,0227
C/C	33	0,08	0,92	0,0130	1,96	0,05	0,08+1,96*0,0130

Exemplo

Frequência alélica do polimorfismo IL6 -174 em pacientes com câncer de cabeça e pescoço

IL6 -174	n=436	p	1-p	$\sqrt{[p(1-p)/n]}$	$z_{95\%}$	Limite inferior	Limite superior
G/G	253	0,58	0,42	0,0236	1,96	0,53	0,63
G/C	150	0,34	0,66	0,0227	1,96	0,30	0,38
C/C	33	0,08	0,92	0,0130	1,96	0,05	0,11

Espera-se com 95% de confiança que o intervalo [0,53-0,63] contenha o verdadeiro valor da proporção de pacientes com câncer de cabeça e pescoço que apresentam a frequência alélica G/G do IL6 -174

Aula 5

Comparação entre dois grupos independentes

Variáveis quantitativas

Parâmetro	Grupo 1	Grupo 2
Média		
Desvio padrão		
Variância		

Aula 5

Comparação entre dois grupos independentes

Variância

- Variâncias iguais entre os grupos: Homocedasticidade
- Variâncias diferentes entre os grupos: Heterocedasticidade

Comparação entre dois grupos independentes

- Variâncias iguais entre os grupos: Homocedasticidade

A variável de interesse tem distribuição normal, e podemos assumir que a variação entre os grupos é igual

$$X_1 \sim N(\mu_1; \sigma_1^2)$$

$$X_2 \sim N(\mu_2; \sigma_2^2)$$

Comparação entre dois grupos independentes

- Variâncias iguais entre os grupos: Homocedasticidade

A variável de interesse tem distribuição normal, e podemos assumir que a variação entre os grupos é igual

Podemos organizar os dados da seguinte forma

Grupo	Tamanho da amostra	Observações	Média	Desvio padrão
1				
2				

Aula 5

Comparação entre dois grupos independentes

- Variâncias iguais entre os grupos: Homocedasticidade

Variância combinada dos dois grupos

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Comparação entre dois grupos independentes

- Variâncias iguais entre os grupos: Homocedasticidade

Intervalo de confiança (1-α)100% para a diferença das médias

$$IC = \left[(\bar{x}_1 - \bar{x}_2) - t_{n_1+n_2-2;1-\alpha/2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,;\, (\bar{x}_1 - \bar{x}_2) + t_{n_1+n_2-2;1-\alpha/2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

Comparação entre dois grupos independentes

- Exemplo: Dados de um ensaio clínico com o uso de uma nova droga vs. placebo. A seguir os dados sobre eficácia:

Grupo	Tamanho da amostra	Média	Desvio padrão
Placebo			
Tratamento			

Comparação entre dois grupos independentes

- Exemplo: Dados de um ensaio clínico com o uso de uma nova droga vs. placebo.

1. Estimativa do efeito da droga: $\bar{x}_1 - \bar{x}_2 = 20{,}53 - 11{,}37 = 9{,}16$
2. Estimativa para a variância comum: $s_p^2 = \frac{(15-1)(11{,}09)^2+(16-1)(7{,}26)^2}{15+16-2}$

$$s_p^2 = (9{,}31)^2 = 86{,}68$$

3. IC95% para a diferença das médias

$$\left[9{,}16 - (2{,}0452)\sqrt{86{,}68}\sqrt{\frac{1}{15}+\frac{1}{16}};\, 9{,}16 + (2{,}0452)\sqrt{86{,}68}\sqrt{\frac{1}{15}+\frac{1}{16}}\right]$$

Aula 5

Comparação entre dois grupos independentes

- Exemplo: Dados de um ensaio clínico com o uso de uma nova droga vs. placebo.

3. IC95% para a diferença das médias

$$\left[9{,}16 - (2{,}0452)\sqrt{86{,}68}\sqrt{\frac{1}{15}+\frac{1}{16}}\,;\, 9{,}16 + (2{,}0452)\sqrt{86{,}68}\sqrt{\frac{1}{15}+\frac{1}{16}}\right]$$

$$[2{,}32; 16{,}00]$$

Interpretação:

Aula 5

Comparação entre dois grupos independentes

Variáveis qualitativas

Parâmetro	Grupo 1	Grupo 2
Proporção		
Desvio padrão		
Variância		

Aula 5

Comparação entre dois grupos independentes

Variáveis qualitativas

Exposição	Desfecho		Total
	Presente	Ausente	
Presente	a	b	a+b
Ausente	c	d	c+d
Total	a+c	b+d	a+b+c+d

Aula 5

Comparação entre dois grupos independentes

Comparação de proporções para amostras grandes

$$IC = \left[(\bar{p}_1 - \bar{p}_2) - z_{1-\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}\,;\, (\bar{p}_1 - \bar{p}_2) + z_{1-\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \right]$$

Comparação entre dois grupos independentes

Exemplo: Objetivo do estudo é testar se 365mg de aspirina (65% de um comprimido usual, que é de 0,50g) reduzem a mortalidade devida a doenças cardiovasculares. Foi realizado um ensaio clínico que envolveu 22071 médicos americanos, com idades entre 40 e 84 anos que não tinham história de infarto do miocárdio, AVC ou ataque isquêmico transitório, não usavam regularmente aspirina nem apresentavam contraindicações ao seu uso.

Aula 5

Comparação entre dois grupos independentes

Exemplo:
Grupo Experimental: 11037
Grupo Placebo/Controle: 11034

Tempo de seguimento: média de 57 meses, foram observados 139 infartos entre aqueles que tomavam aspirina e 239 entre aqueles que tomaram placebo

Aula 5

Comparação entre dois grupos independentes

Exemplo:
Grupo Experimental: 11037
Grupo Placebo/Controle: 11034

Como calculamos o intervalo de confiança de 95% para a diferença das proporções de infarto entre os grupos?

Comparação entre dois grupos independentes

Exemplo:
Grupo Experimental: 11037
Grupo Placebo/Controle: 11034

1. Proporções por grupos:
   - $\overline{p_c} = {}^{239}/_{11034} = 0{,}0217$ (Grupo controle)
   - $\overline{p_e} = {}^{139}/_{11037} = 0{,}0126$ (Grupo tratamento)

2. Diferença entre as proporções:
   - $\overline{p_c} - \overline{p_e} = 0{,}0217 - 0{,}126 = 0{,}0091$

Aula 5

Comparação entre dois grupos independentes

Exemplo:
Grupo Experimental: 11037
Grupo Placebo/Controle: 11034

3. Erro padrão da diferença:

$$-\quad \sqrt{\frac{0{,}0217(1-0{,}0217)}{11034}+\frac{0{,}0126(1-0{,}0126)}{11037}}=0{,}0017$$

4. Calculando o IC95%

$$IC=\left[(\bar{p}_1-\bar{p}_2)-z_{1-\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1}+\frac{\bar{p}_2(1-\bar{p}_2)}{n_2}};(\bar{p}_1-\bar{p}_2)+z_{1-\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1}+\frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}\right]$$

Comparação entre dois grupos independentes

Exemplo:
Grupo Experimental: 11037
Grupo Placebo/Controle: 11034

Calculando o IC95%

$$IC = [0{,}0123 - 1{,}96(0{,}0017); 0{,}0123 + 1{,}96(0{,}0017)]$$
$$IC = [0{,}90; 1{,}56]$$

Interpretação:

Aula 5

Para que medidas calculamos intervalos de confiança?

Incidência
Prevalência
Mediana
Percentil
Risco relativo
Odds ratio
Razão de prevalência
Hazard ratio
Probabilidade de sobrevida

$$IC95\% = estatística \pm 2\ EP$$

aproximadamente

Aula 5

Mas... nem todos os estimadores são fáceis de calcular

$$IC95\% = ln(\widehat{OR}) \pm z\alpha/_2 \sqrt{\hat{V}[ln(\widehat{OR})]}$$

IC para *odds ratio*

E para terminar o cálculo...

$$LI = e^{ln(\widehat{OR}) - z\alpha/_2 \sqrt{\hat{V}[ln(\widehat{OR})]}}$$

$$LS = e^{ln(\widehat{OR}) + z\alpha/_2 \sqrt{\hat{V}[ln(\widehat{OR})]}}$$

Aula 5

Até agora...

- **População**
  - População
  - População de estudo
- **Amostra**
  - Tipos de amostragem
- **Parâmetros e estimadores**
  - Estimativas dos parâmetros (estimação pontual)
- **Intervalos de confiança**

Aula 5

Estatística inferencial

É o ramo da estatística que fornece métodos para que o pesquisador possa tomar sua decisão a respeito de hipóteses formuladas, informando sobre o risco de erro que acompanha a decisão

Aula 5

Estatística inferencial

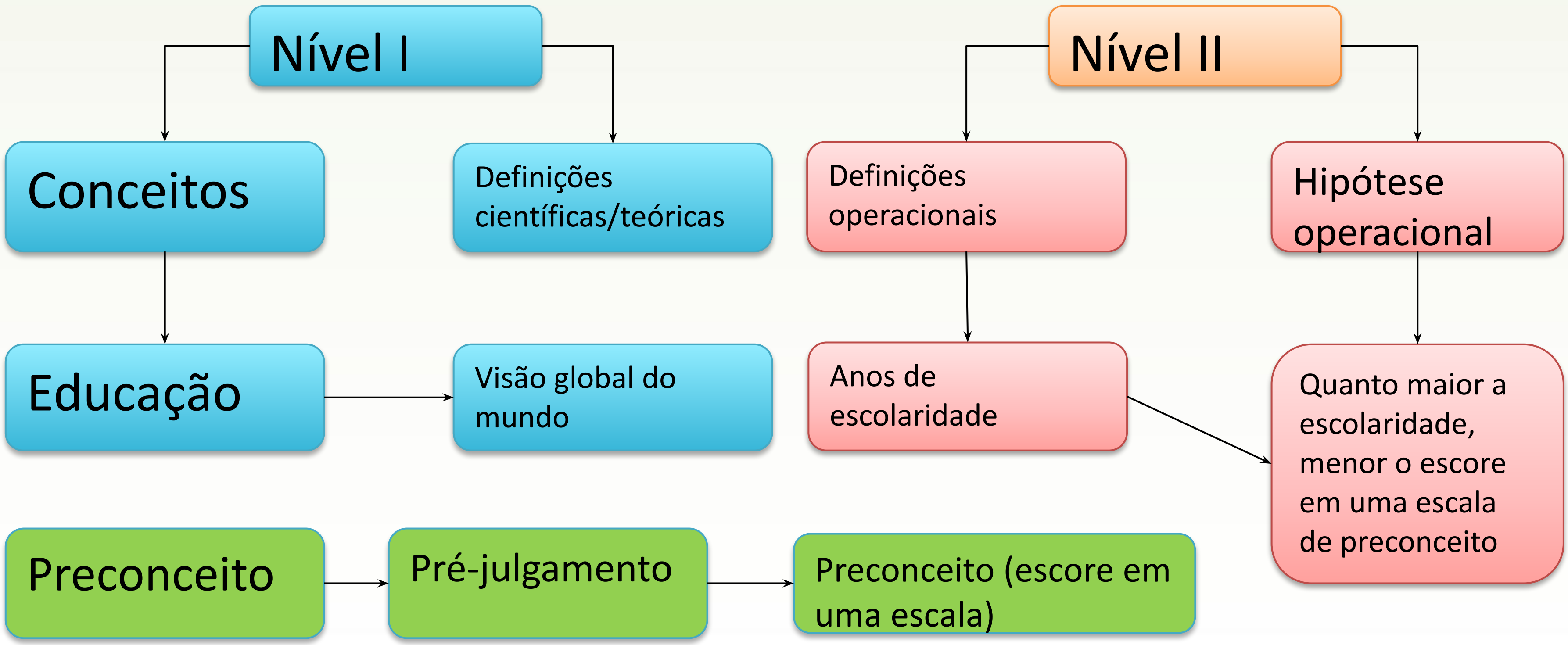
Nível I: Teórico (conceitos, hipóteses científicas)

Nível II: Operacional (hipótese estatística)

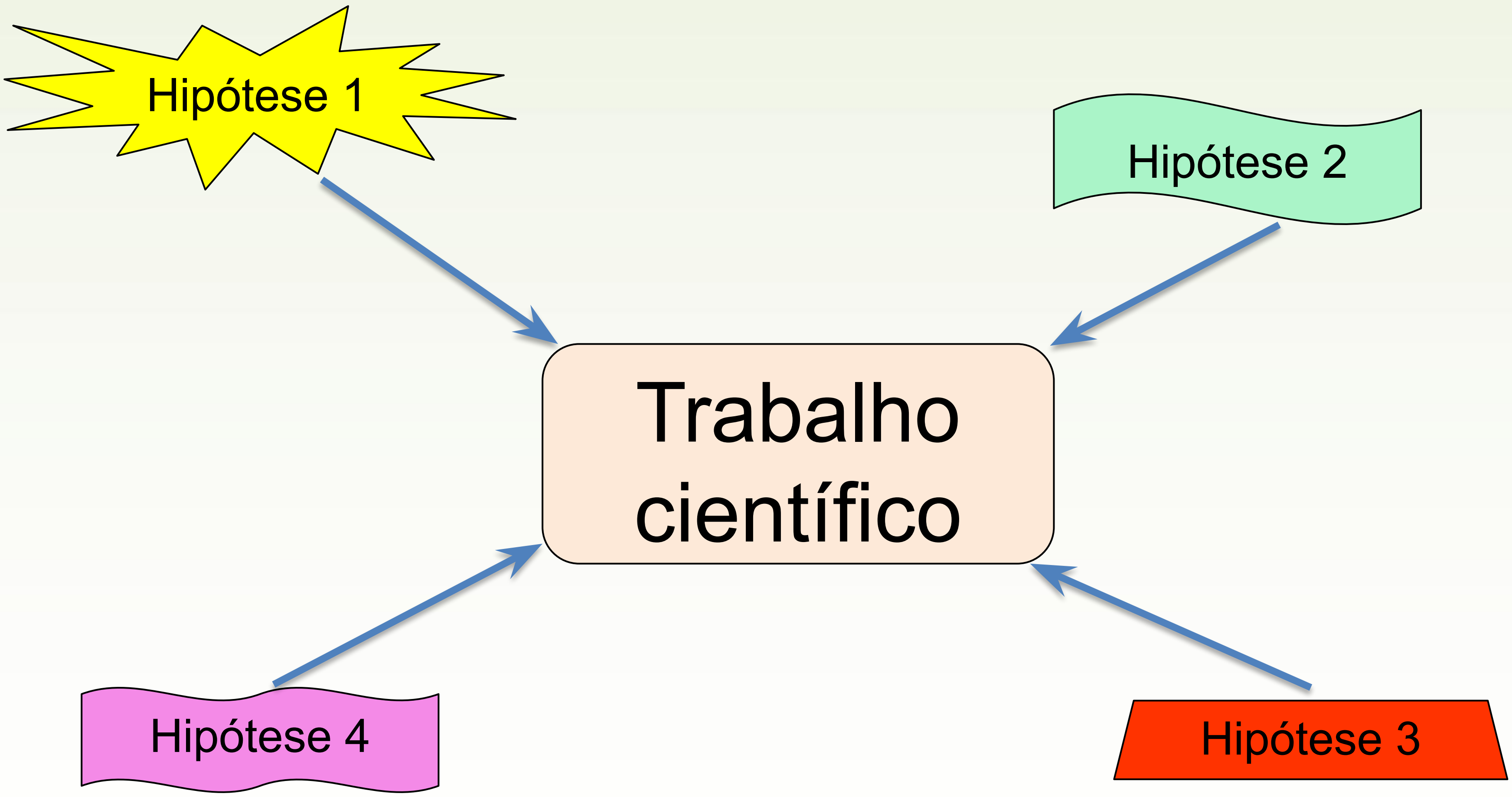
Aula 5

Situação

Quanto mais bem educada uma pessoa, menor o seu preconceito em aceitar certa campanha sanitária



Aula 5


Hipótese 1
Hipótese 2
Trabalho científico
Hipótese 4
Hipótese 3

Aula 5

Se quiser comparar o valor que obtive da minha amostra com um valor hipotético?

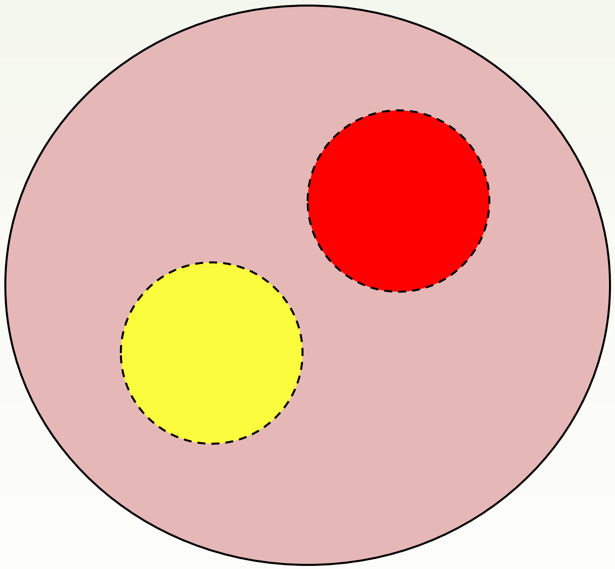
E se quiser comparar o valor da minha amostra com o valor estimado por outra amostra?

Será que o valor que achei é igual?

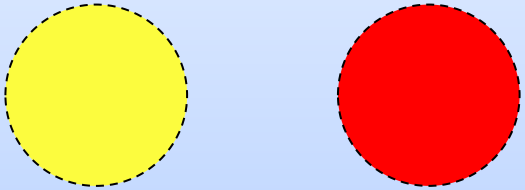
É maior?

É menor?

População de estudo



Amostras



$$p_1 = \frac{a_1}{n_1} \qquad p_2 = \frac{a_2}{n_2}$$

$$\overline{x_1} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} \qquad \overline{x_2} = \frac{\sum_{i=1}^{n_2} x_i}{n_2}$$

$$\theta \rightarrow ?$$

$$\widehat{\theta}$$

Valor conhecido

Hipótese

É uma forma de especulação relativa a um fenômeno estudado (qualquer que seja)

É qualquer afirmação sobre a distribuição de probabilidade de uma variável aleatória (afirmação sobre um parâmetro)

Hipótese estatística

É uma especulação feita em relação à uma proposição, porém relativa à uma população definida

Tipos de hipóteses estatísticas

Comparam sempre dois ou mais parâmetros

- Igual =
- Menor ou igual ≤
- Maior ou igual ≥

Tipos de hipóteses estatísticas

Hipótese nula (H_0)

Estabelece a ausência de diferença entre os parâmetros

É sempre a primeira a ser formulada

Tipos de hipóteses estatísticas

Hipótese alternativa (H_A ou H_1)

É a hipótese contrária à hipótese nula

Geralmente é aquela que o pesquisador quer confirmar

Verificação das hipóteses

Somente se dará com certeza se for estudada toda a população

Pesquisador desconhece o verdadeiro valor do **parâmetro**

Dados obtidos a partir de **amostras**

Decisões

Risco

Risco máximo admitido para o erro de afirmar que existe uma diferença, quando ela efetivamente existe

Aula 5

Teste de hipóteses

Procedimento estatístico pelo qual se rejeita ou não uma hipótese, associando à conclusão um risco máximo de erro

A hipótese testada é sempre H_0

H_0 rejeitada □ Aceitamos H_1

H_0 não é rejeitada □ Rejeitamos H_1

Aula 5

Intervalo de confiança → tentamos "cercar" um parâmetro

Teste de hipóteses → hipótese em quanto ao valor do parâmetro

Amostra → Aceitar/Rejeitar a hipótese

Erro do *tipo I*

Todo teste de hipótese tem sua conclusão sujeita a erro

O erro de afirmar que existe uma diferença quando ela efetivamente não existe (isto é, rejeitar incorretamente a hipótese nula) é chamado de erro do tipo I e tem uma probabilidade de ocorrer igual a **α (*alfa*)**

Aula 5

Erro do *tipo II*

Todo teste de hipótese tem sua conclusão sujeita a erro

Também é possível cometer-se o erro de aceitar H_0 quando não se deveria, isto é, afirmar uma igualdade quando o correto seria afirmar uma diferença erro do tipo II e tem uma probabilidade de ocorrer igual a β (*beta*)

Aula 5

Verdade	Conclusão do teste	
	Não se rejeita H_0	Rejeita-se H_0
H_0 é verdadeira	Decisão correta	Decisão errada: erro tipo I
	Probabilidade: $1-\alpha$	Probabilidade α
H_0 é falsa	Decisão errada: erro tipo II	Decisão correta
	Probabilidade β	Probabilidade $1-\beta$ (Poder do teste)

Aula 5

"Foi encontrada diferença significativa entre os grupos de mulheres com e sem linfedema em relação à mediana de linfonodos retirados (*p=0,02*); apresentação de trombose linfática superficial no braço homolateral à cirurgia (*p<0,01*)..."

Daniella Marta Ferreira de Paiva[1]
Isabel Cristina Gonçalves Leite[2]
Vivian de Oliveira Rodrigues[3]
Marcelle Goldner Cesca[3]

Fatores associados ao linfedema em pacientes com câncer de mama

Associated factors of lymphedema in breast cancer patients

Rev Bras Ginecol Obstet. 2011; 33(2):75-80

Aula 5

Table 1 - Risk Factors for Primary Cardiac Arrest among the Case and Control Patients Treated with Drugs for Hypertension

Risk Factor	Case Patients (N = 114)	Control Patients (N = 535)	P Value†
Age (yr)	65±9	63±9	0.07
Pretreatment blood pressure (mm Hg)			
Systolic	173±26	166±21	0.01
Diastolic	103±12	102±11	0.20
Pretreatment heart rate (beats/min)	85±15	81±13	0.03
Duration of hypertension (yr)	8±6	7±5	0.06
No. of clinic visits, previous yr	5±4	5±5	0.38
Male sex (%)	63	53	0.06
White race (%)	89	92	0.20
Current smoker (%)	45	27	<0.001
Diabetes mellitus (%)	18	9	0.01
Antihypertensive regimen (%)			0.65
Single drug	68	70	
Multiple drugs	32	30	

*Plus–minus values are means ±SD.

†Based on the two-sample t-test or the chi-square test. CI denotes confidence interval.

Diuretic Therapy for Hypertension And The Risk Of Primary Cardiac Arrest.

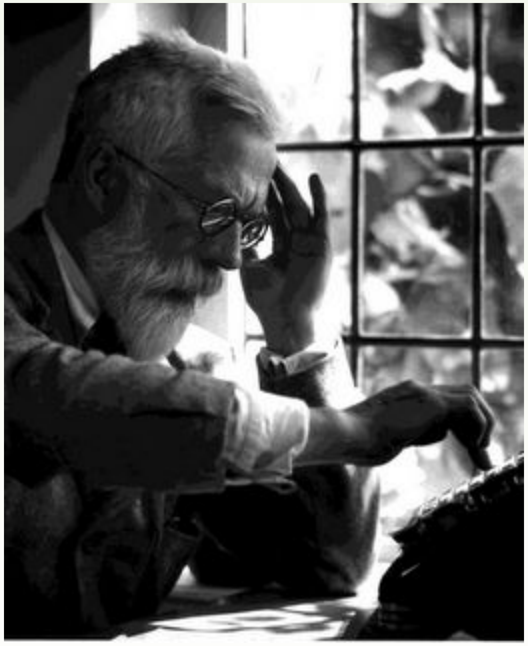
[Original Articles]

Siscovick, David S.; Raghunathan, T.E.; Psaty, Bruce M.; Koepsell, Thomas D.; Wicklund, Kristine G.; Lin, Xihong; Cobb, Leonard; Rautaharju, Pentti M.; Copass, Michael K.; Wagner, Edward H.

The NEW ENGLAND JOURNAL of MEDICINE

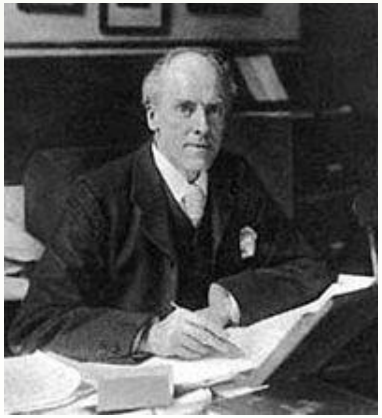
Valor de p
Abordagem de R.A. Fisher

- Se inicia a abordagem com uma proposição inicial H_0
- Se realiza o experimento e se calcula a probabilidade de ocorrência do valor observado ou de um valor mais extremo da estatística do teste, em uma curva de probabilidade especificada na proposição inicial
- Essa probabilidade calculada é conhecida como o valor de p (*p-value*)
- Valor de p pequeno ou valor de p grande escolha do pesquisador



R.A. Fisher

Nível de significância
Abordagem de Neyman-Pearson



- Nível de significância α
- Fixado antes de realizar o experimento
- Estabelecer a distribuição a ser utilizada
- Se compara o valor crítico com o valor da distribuição
- Decisão baseada na comparação

Valor de p e Nível de significância

Na prática as duas abordagens são “misturadas” e são consideradas iguais...

Valor de p

p

Valor de p

Nível descritivo do teste

O valor de p é uma probabilidade

Aula 5

H_0

Média do GRUPO A = Média do GRUPO B

H_1

Média do GRUPO A ≠ Média do GRUPO B

Média de idade:
Grupo A = 45 anos
Grupo B = 38 anos

Teste estatístico *adequado*
Valor de p = 0,572

Supondo que não há diferença nas médias de idade entre os grupos A e B, há **57,2%** de **probabilidade** da diferença observada ser explicada pelo acaso (variação amostral)

Aula 5

H_0

Proporção GRUPO A = Proporção GRUPO B

H_1

Proporção GRUPO A ≠ Proporção GRUPO B

Menopausa:
Grupo A = 30%
Grupo B = 42%

Teste estatístico *adequado*
Valor de p = 0,255

Supondo que não há diferença nas proporções entre os grupos A e B, há **25,5%** de **probabilidade** da diferença observada ser explicada pelo acaso (variação amostral)

Aula 5

H_0

Média do GRUPO A = Média do GRUPO B

H_1

Média do GRUPO A ≠ Média do GRUPO B

Média de idade:
Grupo A = 30 anos
Grupo B = 70 anos

Teste estatístico *adequado*
Valor de p = 0,012

Supondo que não há diferença nas médias de idade entre os grupos A e B, há **1,2%** de **probabilidade** da diferença observada ser explicada pelo acaso (variação amostral)

Aula 5

1º Hipóteses

H_0 : GRUPO A = GRUPO B

H_1 : GRUPO A ≠ GRUPO B

2º Estatística do teste

Teste estatístico *adequado*

3º Calcula-se o valor de p

Valor de p

4º Decisão

Quando iremos rejeitar H_0 e aceitar que a diferença observada não é simplesmente consequência do acaso (variação amostral)?

Quando a probabilidade (p) do acaso explicar a diferença for pequena

Hipótese nula H_0
Não há diferença entre os grupos

Hipótese alternativa H_1
Há diferença entre os grupos

Nível de significância α

É a **probabilidade** máxima de errar ao assumirmos que **há diferença entre os grupos** quando, na verdade, não há...

Escolhido antes de iniciar nosso "experimento"

Geralmente é 5%, mas pode ser 1%, 10%...

Hipótese nula H_0
Não há diferença entre os grupos

Hipótese alternativa H_1
Há diferença entre os grupos

Nível de significância
α

"Assumo a probabilidade máxima de **5%** (ou 1%, 10%...) de errar ao achar que há diferença entre os grupos quando, NA VERDADE, não há!"

Aula 5

1º Hipóteses

H_0 : GRUPO A = GRUPO B

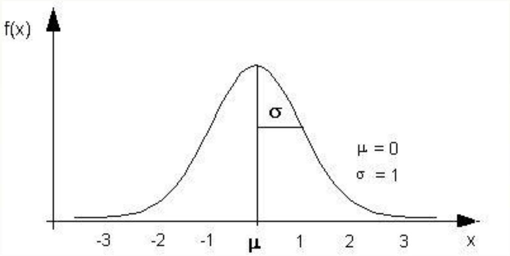
H_1 : GRUPO A ≠ GRUPO B

2º Nível de significância

α (5%, 1%, 10%)

3º Calcula a estatística do teste

Valor crítico: Distribuição z, t



4º Compara o valor crítico

Se compara o valor crítico com o valor da distribuição

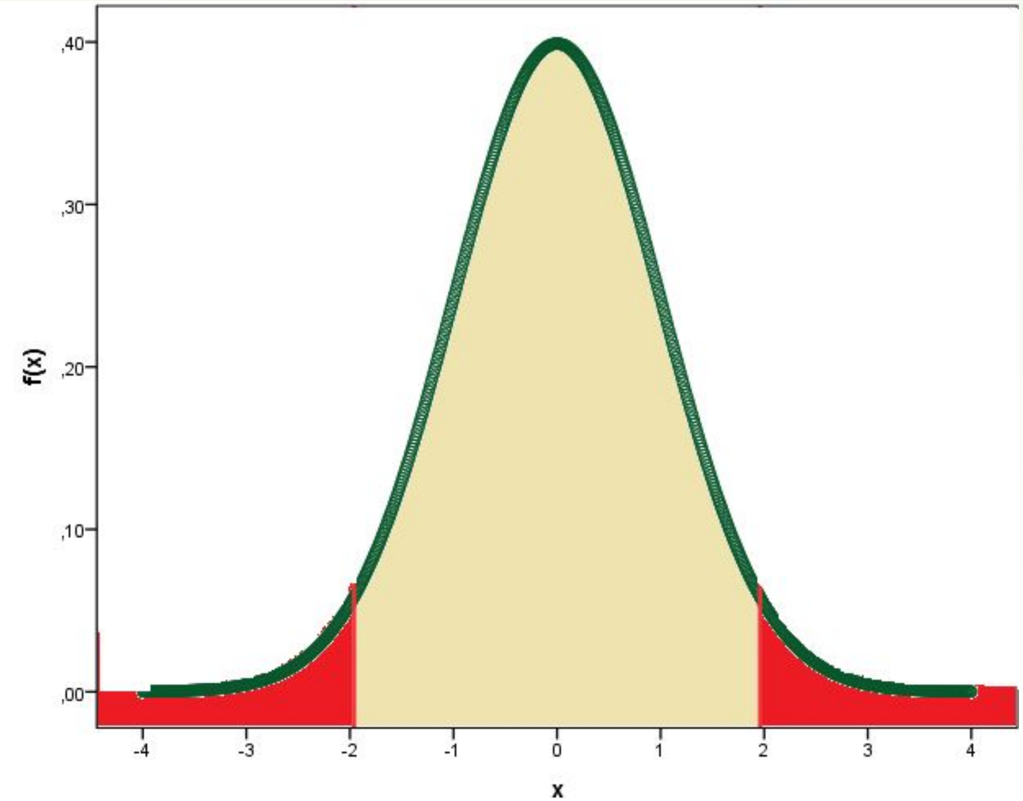
Aula 5

4º Compara o valor crítico

Se compara o valor crítico com o valor da distribuição

$z_{crítico}$

z_{α}



O valor de z_{α} é fixo, por exemplo, com nível de significância de 5% temos $z_{\alpha} = 1{,}96$

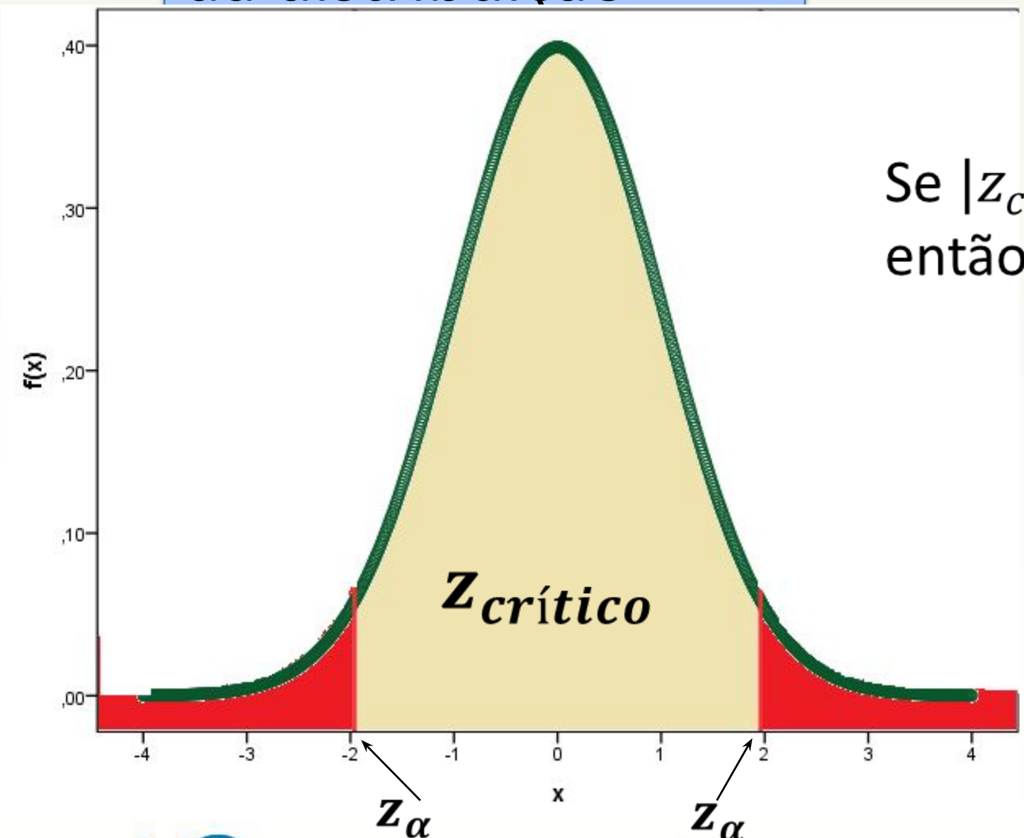
Aula 5

4º Compara o valor crítico

$z_{crítico}$

z_{α}

Se compara o valor crítico com o valor da distribuição



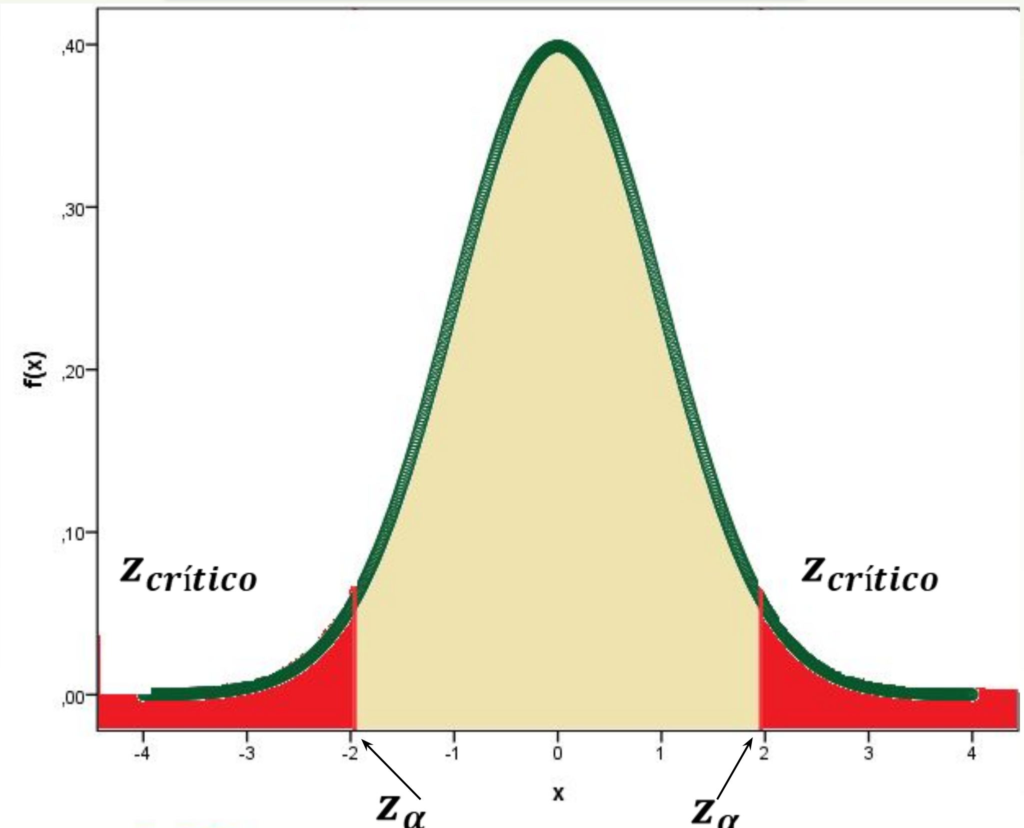
Se $|z_{crítico}| \leq z_{\alpha} = 1{,}96$ então não rejeitamos H_0

Aula 5

4º Compara o valor crítico

Se compara o valor crítico com o valor da distribuição

$z_{crítico}$



Se $|z_{crítico}| > z_\alpha = 1{,}96$ então rejeitamos H_0

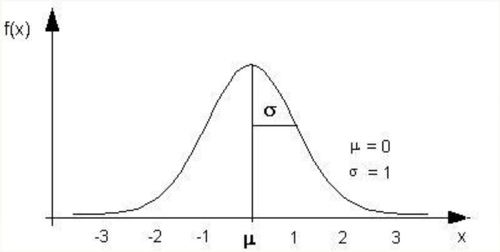
Aula 5

Testando hipóteses...

Aula 5

Estudo sobre níveis séricos de colesterol para a população de homens hipertensos e fumantes

Distribuição normal
Média desconhecida μ



Média do nível sérico de colesterol para a população geral de todos os homens entre 20 e 74 anos é 211 mg/100 ml

Aula 5

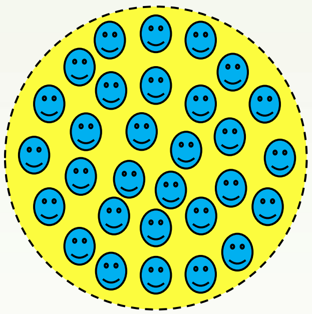
Estudo sobre níveis séricos de colesterol para a população de homens hipertensos e fumantes

Se quiséssemos testar se o nível médio sérico de colesterol da subpopulação de fumantes hipertensos é igual à média da população geral de homens de 20 a 74 anos então:

$$H_0: \mu = \mu_0 = 211 \text{ mg/100 ml}$$
$$H_1: \mu \neq 211 \text{ mg/100 ml}$$

Aula 5

Estudo sobre níveis séricos de colesterol para a população de homens hipertensos e fumantes



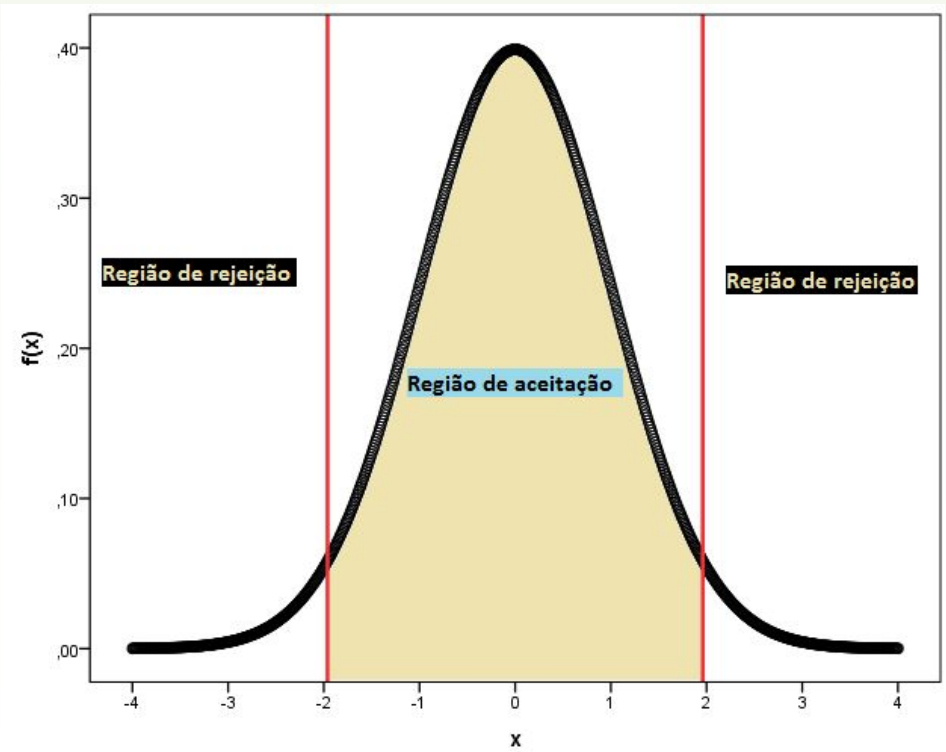
$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s^2$$

$$n$$

$$z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

Distribuição normal

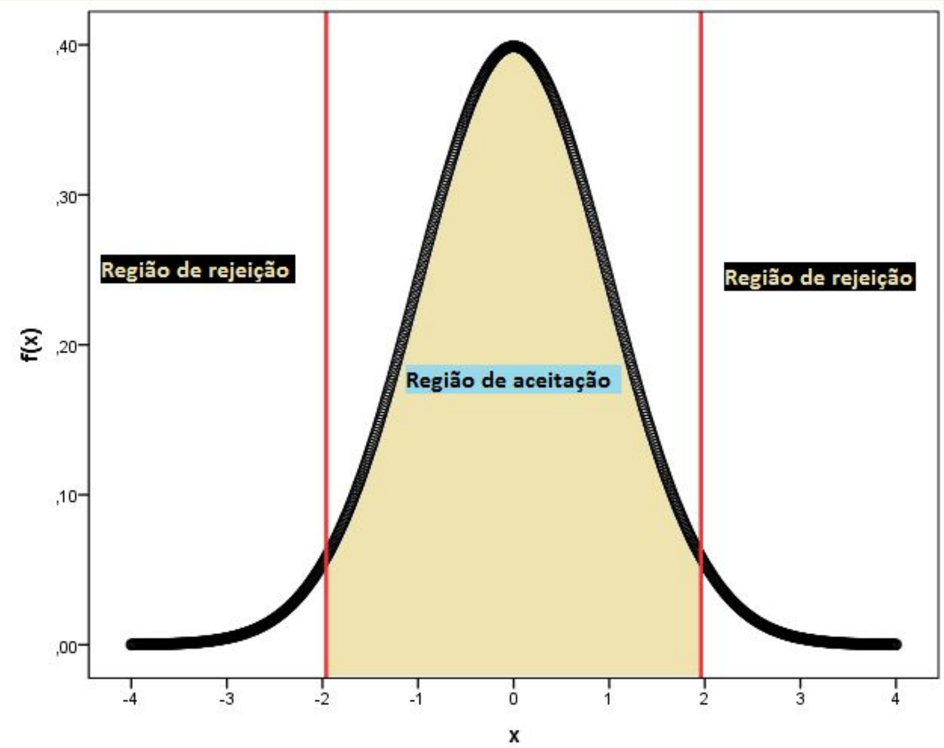


Aula 5

Estudo sobre níveis séricos de colesterol para a população de homens hipertensos e fumantes

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Distribuição normal



Tomamos uma decisão baseados no valor de z

Aula 5

Estudo sobre níveis séricos de colesterol para a população de homens hipertensos e fumantes

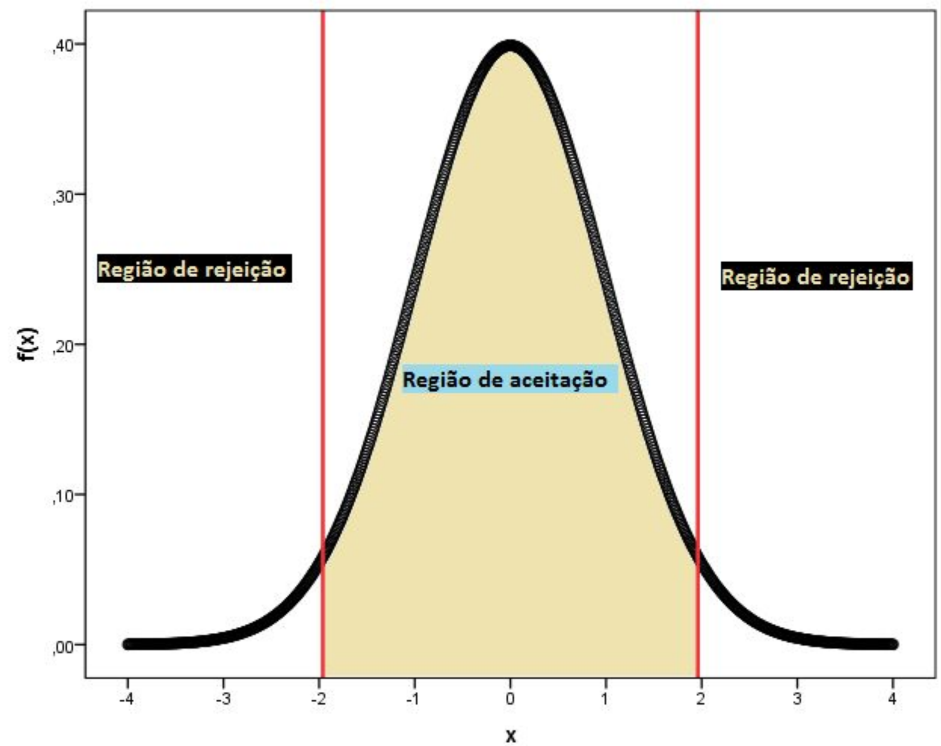
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Supondo que:

$\bar{x} = 217$ $\sigma = 46$ $n = 12$

$$z = \frac{217 - 211}{46 / \sqrt{12}}$$

$$z_c = 0,45$$



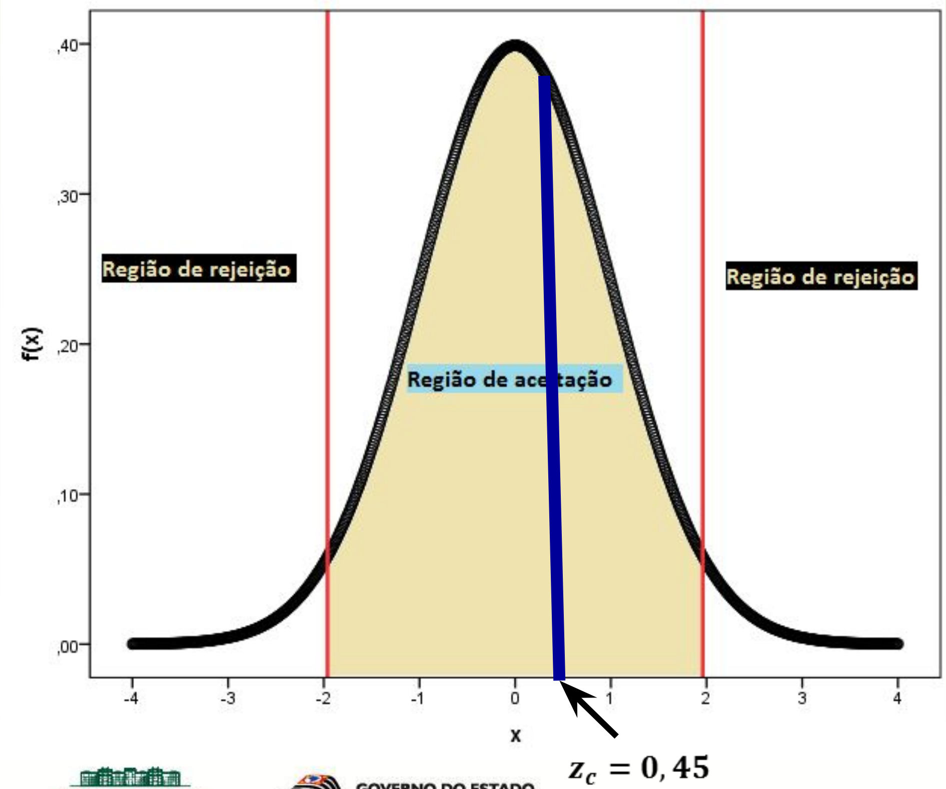
Aula 5

Estudo sobre níveis séricos de colesterol para a população de homens hipertensos e fumantes

$$z_c = 0,45$$

Podemos colocar esse valor de z_c no gráfico de distribuição normal

Ele cai dentro da nossa região de "aceitação"



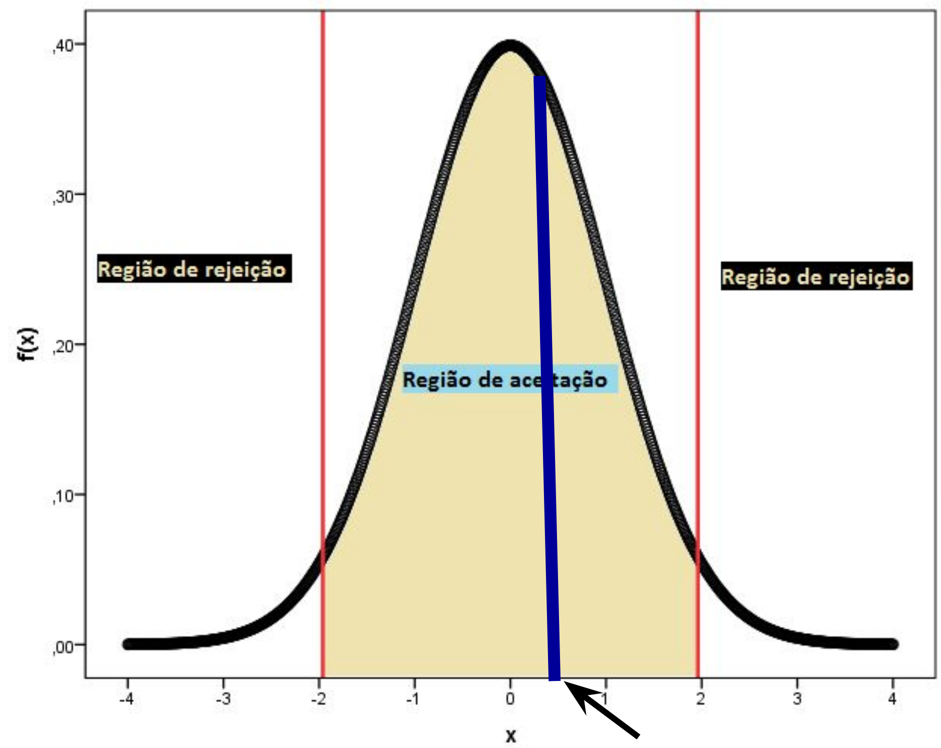
Aula 5

Estudo sobre níveis séricos de colesterol para a população de homens hipertensos e fumantes

$$z_c = 0,45$$

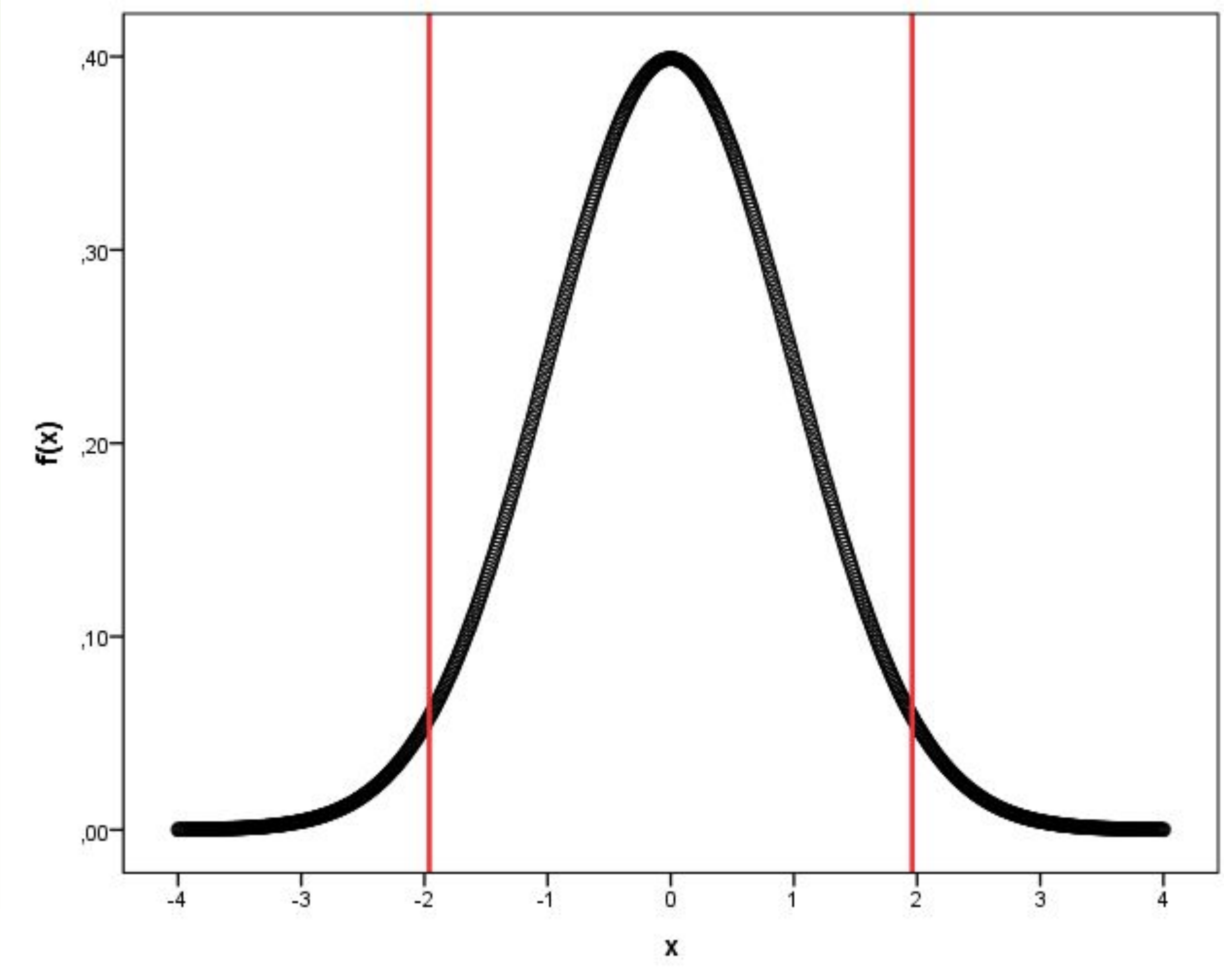
Não rejeitamos a hipótese nula

Com base nessa amostra, a evidência é insuficiente para concluirmos que o nível médio sérico de colesterol da população de fumantes hipertensos é diferente de 211 mg/100 ml



$$z_c = 0,45$$

Aula 5



Aula 5

Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

1) **Hipótese:** é possível elaborar a hipótese de que crianças que nascem com esta síndrome possuem **peso médio ao nascer menor** do que o peso médio ao nascer de crianças sadias

2) **Variável do estudo:** peso ao nascer (em gramas)

Peso ao nascer → Variável quantitativa contínua

Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

3) **Distribuição da variável do estudo:** baseado em conhecimento prévio (da literatura), sabe-se que a distribuição do peso ao nascer em crianças sadias segue uma distribuição normal com média 3000 gramas e desvio padrão 500 gramas

$$X \sim N(\mu_X = 3000;\ \sigma_X = 500)$$

Na linguagem estatística ~ significa "***segue uma distribuição...***"

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

Usando a abordagem de Neyman-Pearson

4) **Investigação da hipótese:** elaborando as hipóteses

$$H_0: \mu_{SDIG} = \mu_{Sadia}$$
$$H_1: \mu_{SDIG} < \mu_{Sadia}$$

ou

$$H_0: \mu_{SDIG} = 3000$$
$$H_1: \mu_{SDIG} < 3000$$

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

5) **Fixando-se o nível de significância:** $\alpha=0{,}05$

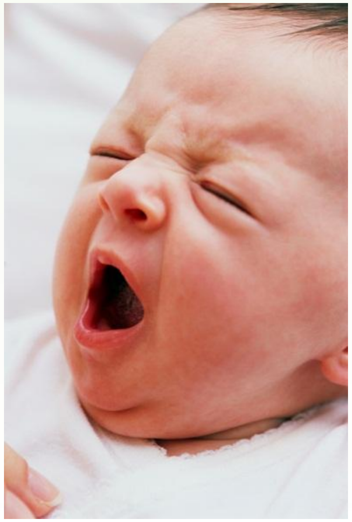
Atenção!
O nível de significância é escolhido antes de realizar a coleta dos dados

Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

6) **Tamanho de amostra:** supondo um tamanho de amostra n=50 recém nascidos com SDIG

Nossa hipótese será testada a partir da nossa amostra

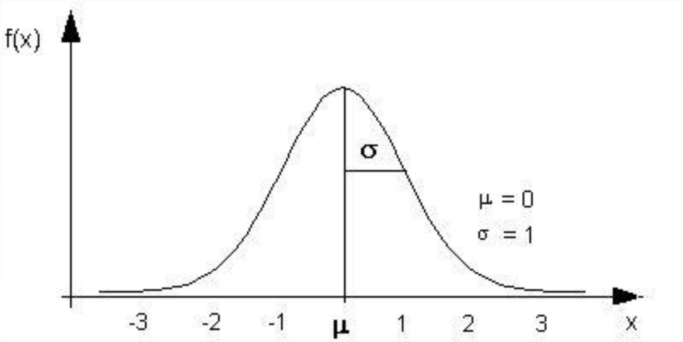


Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

7) **Distribuição de probabilidade:** a média populacional segue uma distribuição normal

Precisamos conhecer a distribuição da nossa variável aleatória

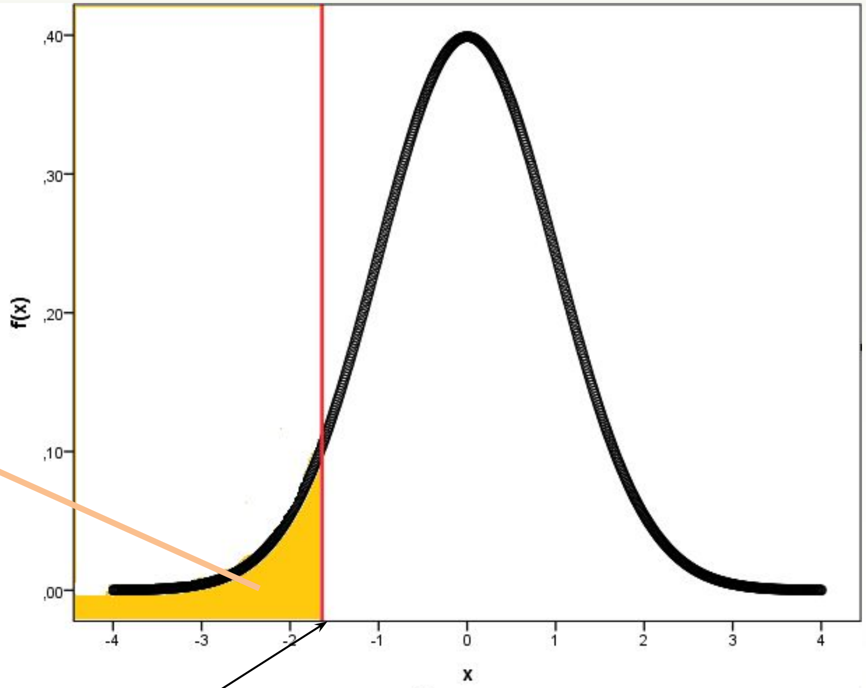


Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

8) **Determinar a região de rejeição e aceitação da hipótese H_0:**

Região de rejeição



x= -1,64

$$z_{crítico} = -1,64$$

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

9) **Cálculo do peso médio na amostra:** Supor que na amostra de 50 crianças, foi observado peso médio ao nascer igual a 2800 gramas

$$\bar{x} = 2800$$

Cálculo do peso médio observado em número de desvios $z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{2800 - 3000}{70,71} = -2,83$

Aula 5

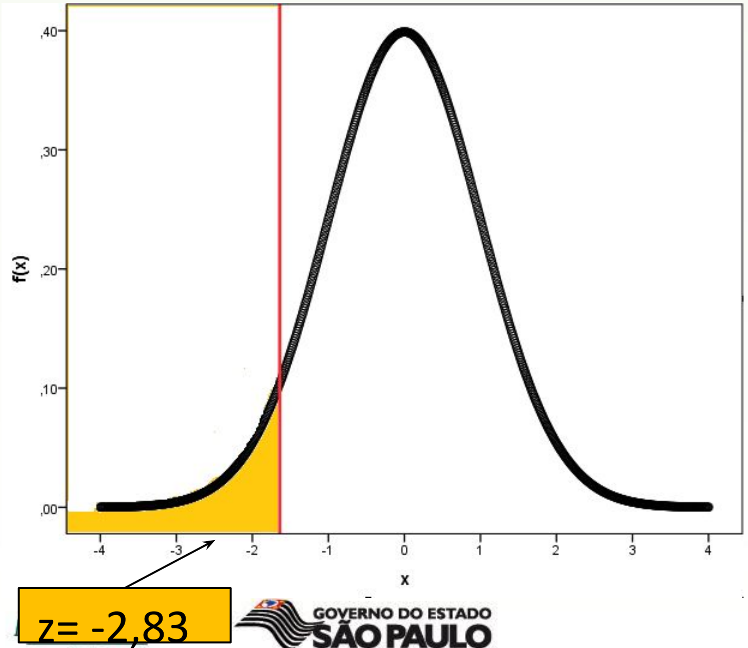
a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

10) **Confrontar o valor da estatística do teste com a região de rejeição e aceitação do H_0:**

$$z_{\bar{x}} = -2,83$$

$$z_{crítico} = -1,64$$

O valor calculado **(-2,83)** caiu dentro da região de rejeição

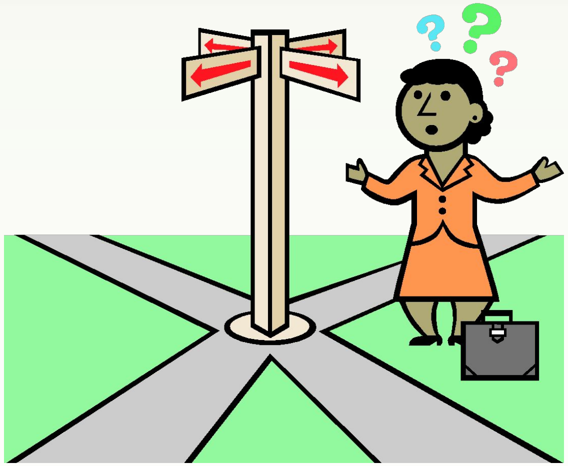


Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

11) **Decisão:**

Rejeita-se H_0



Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

12) **Conclusão:** Foi encontrada diferenças estatisticamente significante entre os pesos ao nascer de crianças sadias e com SDIG para nível de significância α=0,05. Crianças com SDIG nascem com peso menor do que crianças sadias.

Lembrando do H_0

$$H_0: \mu_{SDIG} = \mu_{Sadia}$$

$$\boxed{H_1: \mu_{SDIG} < \mu_{Sadia}}$$

Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

Usando a abordagem de Fisher

1) **Investigação da hipótese:** elaborando as hipóteses

$$H_0: \mu_{SDIG} = \mu_{Sadia}$$
$$H_1: \mu_{SDIG} < \mu_{Sadia}$$

ou

$$H_0: \mu_{SDIG} = 3000$$
$$H_1: \mu_{SDIG} < 3000$$

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

2) **Cálculo do peso médio na amostra:** Supor que na amostra de 50 crianças, foi observado peso médio ao nascer igual a 2800 gramas

$$\bar{x} = 2800$$

Cálculo do peso médio observado em número de desvios $z_{\bar{x}} = \dfrac{\bar{x} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \dfrac{2800 - 3000}{70,71} = -2,83$

Aula 5

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

3) **Calcular o valor da probabilidade dado o valor achado considerando a distribuição normal:**

$$z_{\bar{x}} = -2,83$$

Qual a probabilidade de ter valores menores que **-2,83** considerando a distribuição normal?

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

Distribuição normal:

$$z_{\overline{x}} = -2,83$$

Qual a probabilidade de ter valores menores que **-2,83** considerando a distribuição normal?

z-calculado	valor de p		
-2,83	**0,0023274**		
-1,96	0,0249979		
-1,64	0,0505026		

a. Recém nascidos com Síndrome de Desconforto Idiopático Grave (SDIG)

4) **Decisão:**

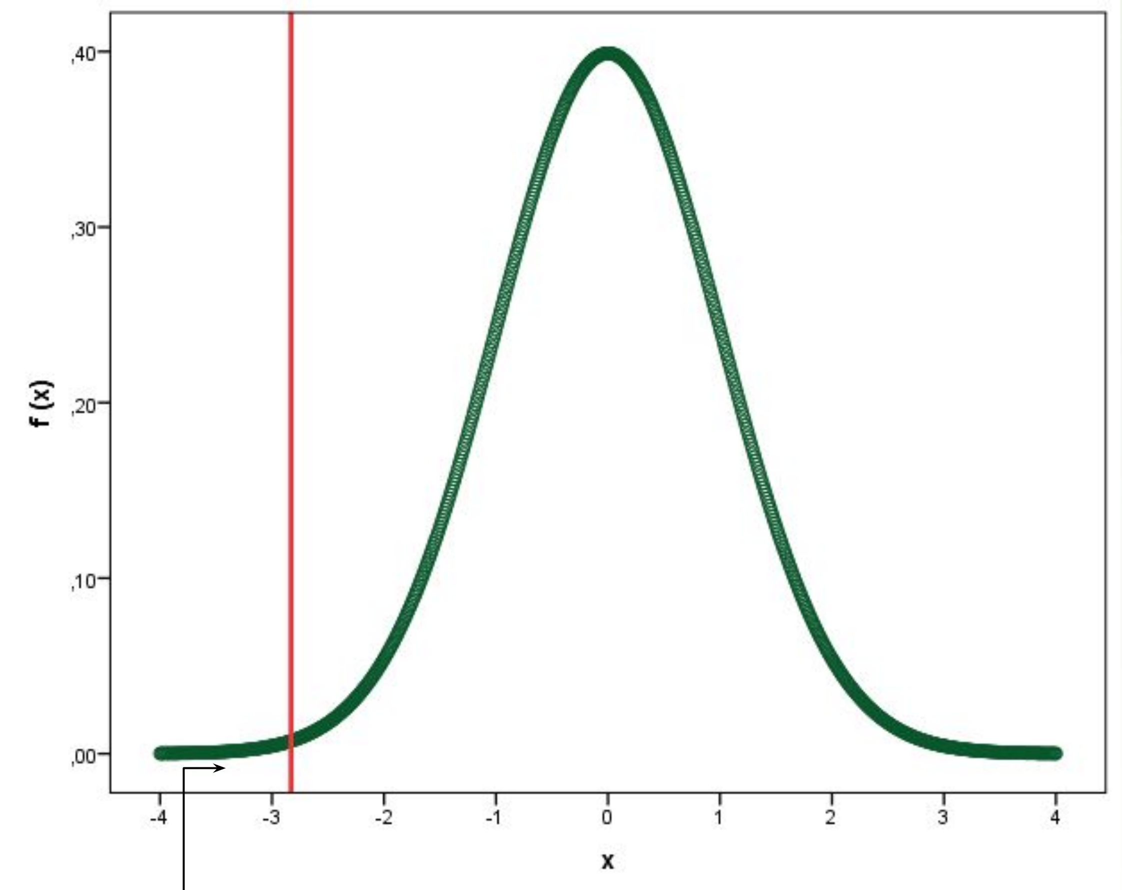
$$\boldsymbol{valor\ de\ p = 0,002}$$

O valor de p é uma probabilidade

Supondo que não há diferença na média dos pesos das crianças sadias e com SDIG...

há 0,2% de probabilidade da diferença observada ser explicada pelo acaso (variação amostral)

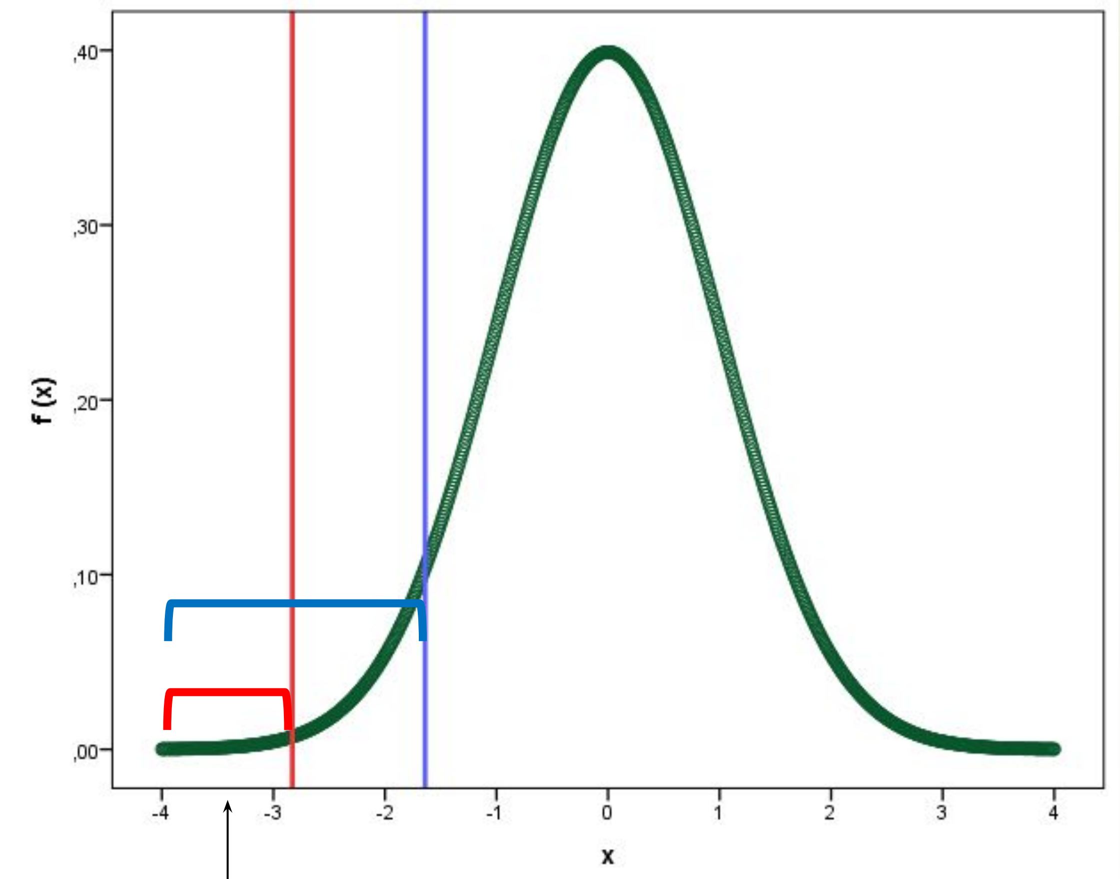
Na prática



$$valor\ de\ p = 0,002$$

Aula 5

Na prática



$$valor\ de\ p = 0,002$$

Aula 5

a. Quando testamos o valor da média com um valor fixo

$$H_0: \mu_X = \mu_0 \qquad H_0: \mu_X \leq \mu_0 \qquad H_0: \mu_X \geq \mu_0$$

$$H_1: \mu_X \neq \mu_0 \qquad H_1: \mu_X > \mu_0 \qquad H_1: \mu_X < \mu_0$$

Onde μ_0 é um valor fixo (um número)

Aula 5

a. Quando testamos o valor da média com um valor fixo

$$X \sim N(\mu; \sigma) \quad \Rightarrow \quad \overline{X} \sim N\left(\mu_{\overline{X}} = \mu_X; \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}\right)$$

O que é isso?

Teorema Central do Limite

Teste de hipóteses para a média

a. Quando testamos o valor da média com um valor fixo

Se conhecemos **σ (desvio padrão)** da população

$$Z_{\overline{X}_{observado}} = \frac{\overline{x}_{obs} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{x}_{obs} - \mu_{\overline{X}}}{\sigma_X / \sqrt{n}}$$

Se não conhecemos σ **(desvio padrão)** da população

$$t_{\overline{X}_{observado}} = \frac{\overline{x}_{obs} - \mu_{\overline{X}}}{s_{\overline{X}}} = \frac{\overline{x}_{obs} - \mu_{\overline{X}}}{s_X / \sqrt{n}}$$

Da nossa amostra podemos calcular todos esses valores sem problemas!

Teste de hipóteses para a média

a. Quando testamos o valor da média com um valor fixo

Decisão baseada na abordagem

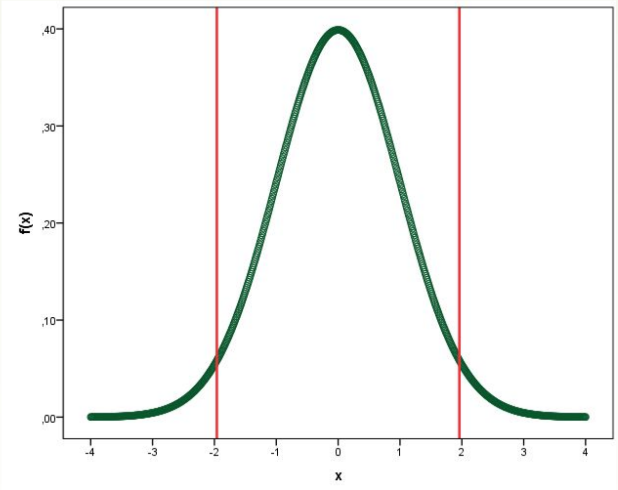
$$Z_{\overline{X}_{observado}}$$

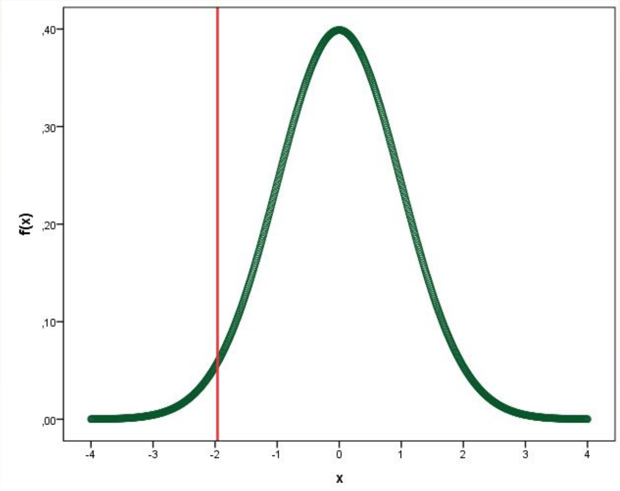
$$t_{\overline{X}_{observado}}$$

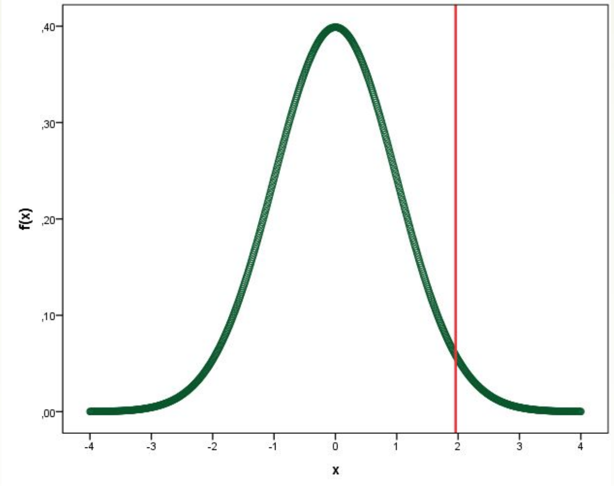
Aula 5

Teste de hipóteses para a média

a. Quando testamos o valor da média com um valor fixo







$H_1: \mu_X \neq \mu_0$ $H_1: \mu_X < \mu_0$ $H_1: \mu_X > \mu_0$

Teste de hipóteses para a média

b. Quando comparamos duas médias

$$H_0: \mu_A = \mu_B \qquad H_0: \mu_A \leq \mu_B \qquad H_0: \mu_A \geq \mu_B$$

$$H_1: \mu_A \neq \mu_B \qquad H_1: \mu_A > \mu_B \qquad H_1: \mu_A < \mu_B$$

Onde μ_A e μ_B são as médias de duas populações diferentes

Teste de hipóteses para a média

b. Quando comparamos duas médias

Se as variâncias nas duas populações são iguais σ^2

$$t_{calculado} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{s^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{(\bar{x}_A - \bar{x}_B)}{\sqrt{s^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}, pois\ (\mu_A - \mu_B) = 0$$

$$s^2 = \frac{(n_A - 1)s_A^2 \quad + (n_B - 1)s_B^2}{n_A + n_B - 2},$$

$onde\ s_A^2 \quad e\ s_B^2$

$são\ as\ variâncias\ das\ duas\ amostras\ estudadas$

Teste de hipóteses para a média

c. Quando testamos o valor de uma proporção

$$H_0 : p_A = p_0 \qquad H_0 : p_A \leq p_0 \qquad H_0 : p_A \geq p_0$$

$$H_1 : p_A \neq p_0 \qquad H_1 : p_A > p_0 \qquad H_1 : p_A < p_0$$

Onde p_0 é um valor fixo (um número)

Teste de hipóteses para a média

c. Quando testamos o valor de uma proporção

Calculamos a estatística do teste

$$z_{calculado} = \frac{(p - p_0)}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Aula 5

Não fique em pânico!

Aula 5

Sobrevivência de pacientes com câncer de pulmão

A distribuição de sobrevivência de cinco anos para indivíduos abaixo dos 40 anos diagnosticados com câncer de pulmão tem proporção da população *p* desconhecida. Estudo anterior encontrou que a proporção de pacientes que sobrevive 5 anos entre os que estão acima de 40 anos no momento do diagnóstico é de 8,2%

Aula 5

É possível que a proporção dos que sobrevivem na população abaixo dos 40 anos seja 0,082 também?

Para poder determinar se esse é o caso então conduzimos um teste estatístico de hipóteses

Aula 5

Começamos com uma afirmação sobre o valor da proporção da população *p*

Então elaborando o teste de hipóteses

$$H_0: p = 0,082$$

$$H_0: p \neq 0,082$$

Considerando um teste de hipótese bilateral

Aula 5

Extraímos uma amostra aleatória de observações dicotômicas da população original e encontramos a probabilidade de se obter uma proporção da amostra tão ou mais extrema que $\hat{p}$, sendo p a verdadeira proporção da população.

Aula 5

Calculamos a estatística do teste

$$z_{calculado} = \frac{(p - p_0)}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Considerando uma amostra de 52 pessoas abaixo de 40 anos que foram diagnosticadas com câncer de pulmão, temos $\hat{p} = 0{,}115$

$$z_{calculado} = \frac{(\hat{p} - p_0)}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,115 - 0,082}{\sqrt{\frac{0,082(1-0,082)}{52}}} = 0,87$$

Aula 5

Cuidado com as interpretações!

"p<0,001 : diferença foi muito significativa"

"p = 0,04 : diferença foi pouco significativa"

O valor de p é uma probabilidade!

A significância esta relacionada com o α e se determina antes

Em resumo

Na prática

p>0,05

α = 5%

Não há diferença estatisticamente significativa

H_0 não é rejeitada

Não há diferença entre os grupos (%, média, etc.)

Diferença é explicada pelo acaso (variação amostral)

p≤0,05

α = 5%

Diferença estatisticamente significativa

H_0 é rejeitada, H_1 é tida como possível

Há diferença entre os grupos (%, média, etc.)

Diferença não é explicada pelo acaso (variação amostral)

Outros exemplos de medidas de associação a serem testadas em estudos epidemiológicos

Estudos transversais (razão de prevalências) $\boldsymbol{H_0: RP = 1}$

Estudos caso-controle (*odds ratio*) $\boldsymbol{H_0: OR = 1}$

Estudos de coorte (risco relativo) $\boldsymbol{H_0: RR = 1}$

Tabelas de contingência → Teste de associação

H_0: Não há associação entre as variáveis

Aula 5

Outros exemplos de trabalhos publicados em estudos epidemiológicos

Revisemos algumas publicações...

Aula 5

Research Article

Risk and Protective Factors for Breast Cancer in Midwest of Brazil

Lívia Emi Inumaru,[1] Maíra Irineu Gomes Duarte Quintanilha,[2] Érika Aparecida da Silveira,[1] and Maria Margareth Veloso Naves[1]

TABLE 2: Socioeconomic, gynecological, and anthropometric variables of women with breast cancer (cases) and respective controls, from Midwest of Brazil (2008–2010).

Variable	Cases (n = 93)		Controls (n = 186)		P value*
	Mean	SD	Mean	SD	
Age (years)	51.93	10.07	51.77	9.73	0.897
Height (cm)	1.55	0.05	1.55	0.06	0.570
Current body mass index (BMI) (kg/m^2)	27.14	5.44	26.84	4.53	0.623
Waist circumference (cm)	86.94	11.86	86.62	10.15	0.819
Adult weight gain (kg)	16.94	11.65	15.65	10.51	0.416
	Median		Median		P value**
Per capita income (US$)	147.00		200.00		0.059
Body mass index (BMI) (kg/m^2) at 20 years old	21.91		21.35		0.369
Age at menarche (years)	13		13		0.283
Age at menopause (years)	50		48		0.115
Age at first full pregnancy (years)	21		20		0.099
Number of children	3		3		0.719
Total breastfeeding (months)	24		28		0.557

* Student's t-test. ** Mann-Whitney test. SD: standard deviation.

Aula 5

Daniella Marta Ferreira de Paiva[1]
Isabel Cristina Gonçalves Leite[2]
Vivian de Oliveira Rodrigues[3]
Marcelle Goldner Cesca[3]

Fatores associados ao linfedema em pacientes com câncer de mama

Associated factors of lymphedema in breast cancer patients

Tabela 3 - Orientação recebida pelas pacientes tratadas para câncer de mama a respeito do membro superior homolateral à cirurgia, segundo presença ou ausência de linfedema

Orientações recebidas	Mulheres com linfedema (Total = 112)		Mulheres sem linfedema (Total = 138)		Valor p
	Número	%	Número	%	
Evitar retirar cutícula					
Sim	45	40,2	70	50,7	0,09
Cuidados na depilação da axila					
Sim	39	34,8	60	43,5	0,16
Evitar exercícios repetitivos					
Sim	42	37,5	60	43,5	0,34
Cuidados com picadas de insetos					
Sim	41	36,6	60	43,5	0,27
Não carregar peso					
Sim	68	60,7	96	69,6	0,14
Observar sinais de inflamação e infecção					
Sim	35	31,3	55	39,9	0,16
Hidratar o braço					
Sim	35	31,3	49	35,5	0,48
Não tomar vacina/injeção ou tirar sangue no braço					
Sim	61	54,5	88	63,8	0,14
Cuidados para não machucar ou queimar o braço					
Sim	41	36,6	66	47,8	0,07

p: valor para comparação entre os dois grupos.

Aula 5

Exercício 5

Lista de exercícios sobre intervalos de confiança

Enviar o exercício até **terça feira** dia **13 de setembro às 18h por e-mail**

rossana.veronica@hc.fm.usp.br

alessandro.campolina@hc.fm.usp.br

Aula 5

Exercício 6

Lista de exercícios sobre intervalos de confiança

Enviar o exercício até **terça feira** dia **20 de setembro às 18h por e-mail**

rossana.veronica@hc.fm.usp.br

alessandro.campolina@hc.fm.usp.br

Obrigada

Rossana V. Mendoza López
Bioestatística e Epidemiologista
Centro de Investigação Translacional em Oncologia
ICESP
rossana.veronica@hc.fm.usp.br