INFERÊNCIA ESTATÍSTICA: INTERVALO DE CONFIANÇA

Intervalo de Confiança

Intervalo de confiança é o primeiro tópico da inferência estatística.

Um intervalo de confiança fornece uma estimativa de um parâmetro populacional (μ ou p) utilizando uma amostra para gerar um intervalo que contém o parâmetro, com um certo nível de confiança.

Intervalo de Confiança

- Intervalo de confiança para a proporção
- Intervalo de confiança para a média
- Nível de confiança
- Distribuição t de student
- Margem de erro e amplitude
- Interpretação
- Simulador
- Comparando dois intervalos
- Tamanho da amostra

INTERVALO DE CONFIANÇA

Intervalo de Confiança

Um intervalo de confiança dá uma estimativa do parâmetro da população usando o resultado de uma amostra para gerar um intervalo.

Esse intervalo deve conter, com um certo nível de confiança, o parâmetro.

1. Estimar a proporção populacional ρ
2. Estimar a média populacional μ quando σ é conhecido
3. Estimar a média populacional μ quando σ é desconhecido

Estimativa do Intervalo

Um intervalo de confiança tem a forma:

$\boxed{\text{Estimativa do parâmetro}} \pm \boxed{\text{Margem de Erro}}$

com um nível de confiança (normalmente 90% or mais).

Nível de confiança

O nível de confiança é geralmente 90%, 95% ou 99%.

O nível de significância é o alfa (α), que é o % de erro que você está disposto a assumir.

O nível de confiança é 1 - α, ou seja, se você assume 5% de erro, seu nível de confiança é de 95% (1- 0,05 = 0,95).

Estimativa do Parâmetro

A estimativa do parâmetro é a estatística da amostra:

- Estimativa da média populacional μ: media amostral $\bar{x}$
- Estimativa da proporção populacional ρ: proporção amostral $\bar{p}$

Margem de Erro

A margem de erro é o que dá a largura ou a amplitude do intervalo.

É calculada pelo nível de confiança multiplicado pelo erro padrão.

Nível de confiança: $z_{\alpha/2}$ ou $t_{\alpha/2}$

Erro padrão da média: $\mathrm{EP}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

Erro padrão da proporção: $\mathrm{EP}(\bar{P}) = \sqrt{\frac{\rho\ (1-\rho\ )}{n}}$

INTERVALO DE CONFIANÇA PARA A PROPORÇÃO

Proporção Populacional

O parâmetro ρ representa a proporção de sucesso na população, onde sucesso é definido pelo que você quer analisar.

A amostra da população, $\bar{p}$ é uma estimativa do parâmetro da proporção da população ρ:

$$\bar{p} = \frac{x}{n}$$

Onde x é o número de sucessos e
n é o tamanho da amostra.

Proporção

O valor esperado e o erro padrão de $\bar{P}$ são:

$$E(\bar{P}) = \bar{p}$$

$$\mathrm{EP}(\bar{P}) = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Intervalo de Confiança da Proporção

Um intervalo com um percentual de confiança para a proporção populacional é

Estimativa do parâmetro ± Margem de Erro

$$E(\bar{P}) \pm z \cdot \mathrm{EP}(\bar{P})$$

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

IC da proporção amostral

Um intervalo com um percentual de confiança para a proporção populacional é

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

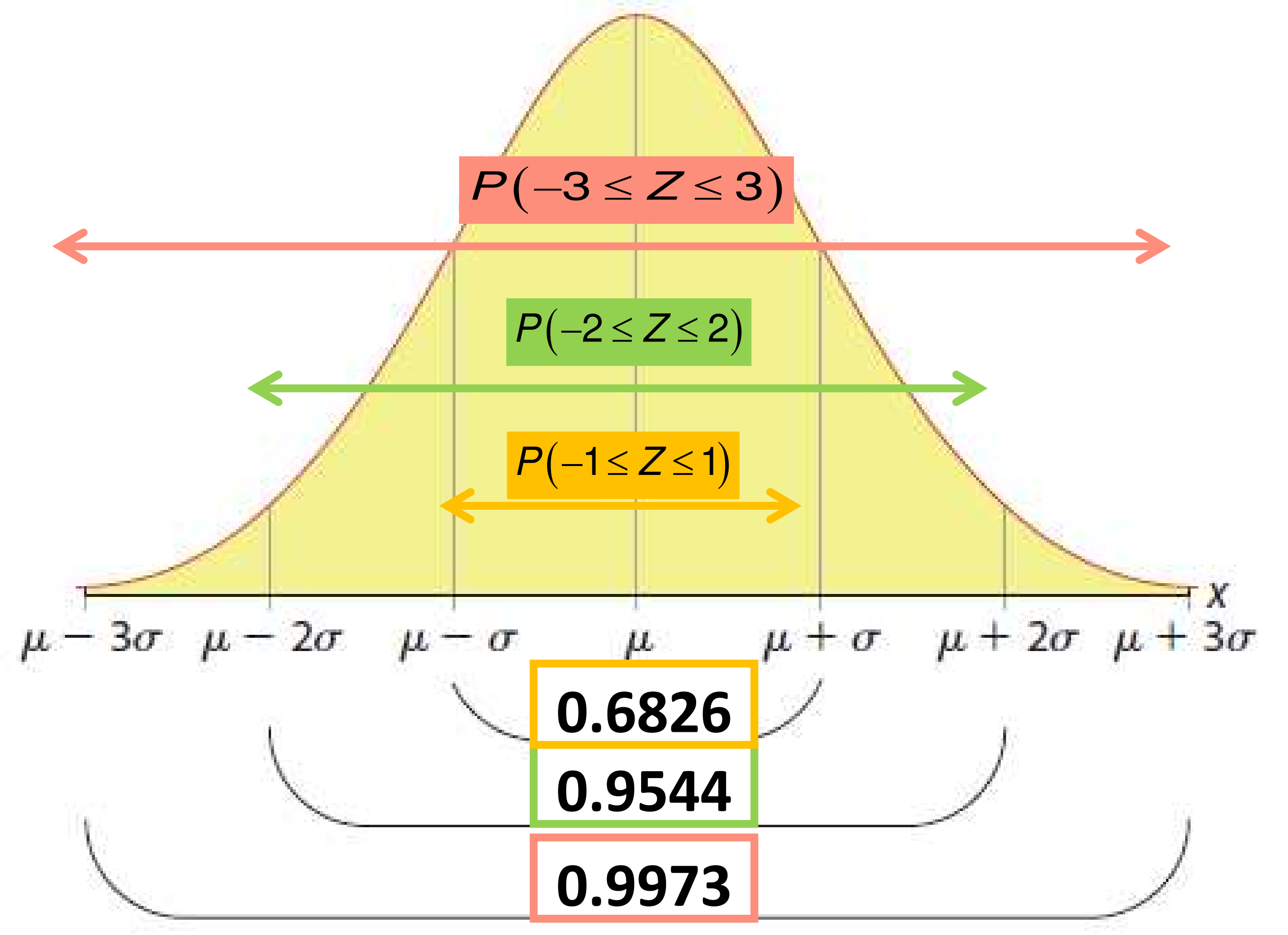
Ou o equivalente:

$$\left[\bar{p} - z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

Nível de confiança e z

O valor de $z_{\alpha/2}$ depende do nível de confiança:

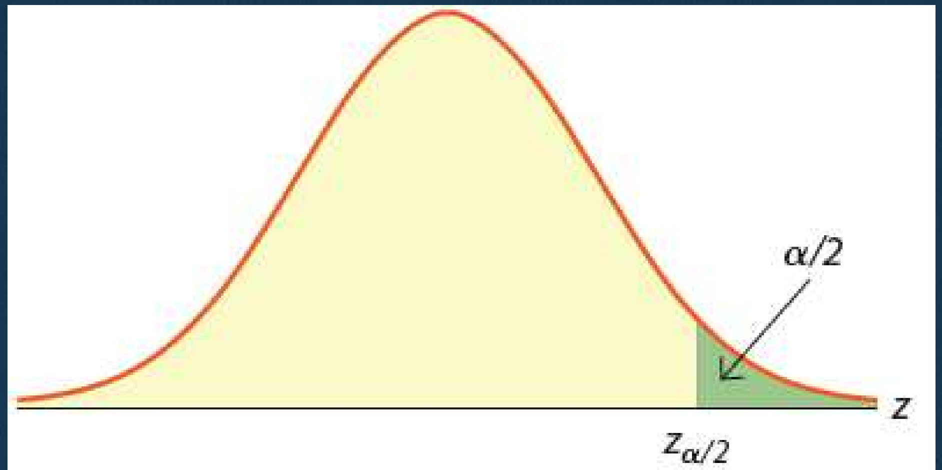
- 90%, $z_{\alpha/2}$ = 1,645
- 95%, $z_{\alpha/2}$ = 1,96
- 99%, $z_{\alpha/2}$ = 2,576



Nível de confiança e z

A fórmula do intervalo de confiança só é valida se a distribuição de $\overline{X}$ for normal.

$z_{\alpha/2}$ é o valor z associado com a probabilidade de $\alpha/2$ na cauda direita:



$P(Z > z_{\alpha/2}) = \alpha/2$

Suposições

A amostra é aleatória e representativa da população.

A fórmula do intervalo de confiança só é válida se a distribuição de $\bar{\mathrm{P}}$ for normal:

- A população segue uma distribuição normal, ou
- n$\bar{p}$ > 10 e n(1 – $\bar{p}$) > 10

Exemplo: cartão

Antes de oferecer um novo cartão de crédito para alunos de uma universidade, um banco gostaria de saber a proporção de clientes que vão aceitar a oferta.

Eles enviam uma oferta para 1000 pessoas e 140 delas aceitaram a proposta.

Qual o intervalo de confiança de 95% de alunos que vão aceitar a oferta?

Exemplo: cartão

(a) Qual o tamanho da amostra n?

n = 1000

(b) Qual é o $\bar{p}$?

$$\bar{p} = \frac{x}{n} = \frac{140}{1000} = 0{,}14$$

(c) Qual é o $z_{\alpha/2}$?

$$z_{\alpha/2} = 1{,}96$$

Exemplo: cartão

(d) Qual o intervalo de confiança?

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= 0{,}14 \pm 1{,}96\ \sqrt{\frac{0{,}14\ (1-0{,}14)}{1000}} = 0{,}14 \pm 1{,}96\ (0{,}01097) = [0{,}1185;\ 0{,}1615]$$

Podemos concluir com 95% de confiança que entre 12% e 16% dos alunos vão aceitar a proposta do cartão de crédito.

INTERVALO DE CONFIANÇA PARA A MÉDIA

Intervalo de Confiança

Um intervalo de confiança dá uma estimativa do parâmetro da população usando o resultado de uma amostra para gerar um intervalo.

Esse intervalo deve conter, com um certo nível de confiança, o parâmetro.

1. Estimar a proporção populacional ρ
2. Estimar a média populacional μ quando σ é conhecido
3. Estimar a média populacional μ quando σ é desconhecido

Média

O valor esperado e erro padrão de $\overline{\mathrm{X}}$ são:

$$E\left(\overline{\mathrm{X}}\right) = \bar{x}$$

$$EP(\overline{\mathrm{X}}) = \frac{\sigma}{\sqrt{n}}$$

ESTIMAR A MÉDIA QUANDO O DP É CONHECIDO

Estimar μ quando σ é conhecido

Um intervalo de confiança da média populacional μ quando o desvio padrão populacional σ é conhecido é:

Estimativa do parâmetro ± Margem de Erro

$$E(\bar{X}) \pm z \cdot \mathrm{EP}(\bar{X})$$

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Estimar μ quando σ é conhecido

Um intervalo de confiança da média populacional μ quando o desvio padrão populacional σ é conhecido é:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Ou o equivalente:

$$\left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

ESTIMAR A MÉDIA QUANDO O DP É DESCONHECIDO

Estimar μ quando σ é desconhecido

Geralmente, não sabemos qual é o desvio padrão da população σ.

Quando σ é desconhecido, usamos uma aproximação pelo desvio padrão da amostra s.

Diferença: não podemos usar o z-score. Ao invés da distribuição normal z, usamos uma “aproximação”, a distribuição t de student.

Estimar μ quando σ é desconhecido

Um intervalo de confiança da media populacional μ quando o desvio padrão populacional σ é desconhecido é:

Estimativa do parâmetro ± Margem de Erro

$$E(\bar{X}) \pm z \cdot \mathrm{EP}(\bar{X})$$

$$\bar{x} \pm \mathrm{t}_{\alpha/2} \frac{s}{\sqrt{n}}$$

Estimar μ quando σ é desconhecido

Um intervalo de confiança da media populacional μ quando o desvio padrão populacional σ é desconhecido é:

$$\bar{x} \pm \mathrm{t}_{\alpha/2} \frac{s}{\sqrt{n}}$$

Ou o equivalente:

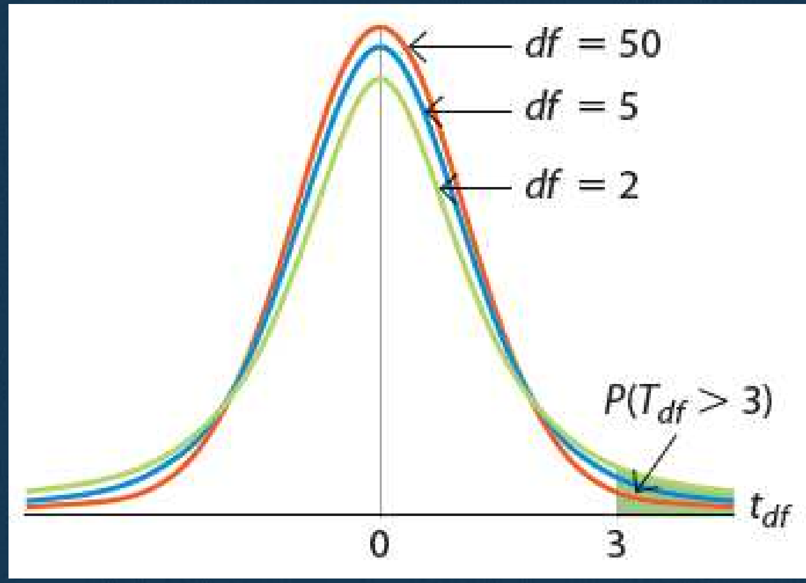
$$\left[\bar{x} - \mathrm{t}_{\alpha/2} \frac{s}{\sqrt{n}} ; \bar{x} + \mathrm{t}_{\alpha/2} \frac{s}{\sqrt{n}}\right]$$

Distribuição t de student

É usada porque substituímos s por σ no erro padrão estimado.

O parâmetro graus de liberdade (gl= n-1) controla a forma da distribuição.

À medida que n aumenta, a distribuição t se assemelha à distribuição normal.



Graus de liberdade

A definição seria o número de valores no cálculo de uma estatística que podem variar livremente.

Para um teste t, um grau de liberdade é usado estimando a média e os n - 1 graus de liberdade restantes estimam a variabilidade.

Ou seja, usamos o número de observações menos 1 (n - 1)
para o cálculo do t.

Unilateral α	0,25	0,10	0,05	0,025	0,01	0,005
Bilateral α	0,50	0,20	0,10	0,05	0,02	0,01
c	0,50	0,80	0,90	0,95	0,98	0,99
G.L						
1	1,000	3,078	6,314	12,706	31,821	63,657
2	0,816	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,694	1,350	1,771	2,160	2,650	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,690	1,337	1,746	2,120	2,583	2,921
17	0,689	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
∞	0,674	1,282	1,645	1,960	2,326	2,576

Para 90%, 95%, 99%:

- n=15, 90%, $t_{\alpha/2,14}$ = 1,761
- n=20, 95%, $t_{\alpha/2,19}$ = 2,093
- n=30, 99%, $t_{\alpha/2}$ = 2,756

Suposições

A amostra é aleatória e representativa da população.

A fórmula do intervalo de confiança só é valida de a distribuição de $\overline{X}$ é normal:

- A população segue uma distribuição normal, ou
- O tamanho da amostra é $n \geq 30$ (Teorema do Limite Central)

Exemplo: cartão

Agora que o banco estimou a proporção de alunos que tem interesse no cartão de crédito, queremos estimar o saldo médio mensal daqueles que aceitaram a oferta.

Lembrando que o banco enviou uma oferta para uma amostra de 1000 pessoas e 140 delas aceitaram a proposta.

Entre esses, a média do saldo é R$1990,50
e o desvio padrão é 2833,33.

Exemplo: cartão

Qual o intervalo de confiança de 95% do saldo médio mensal desses alunos?

$$\bar{x} \pm t\,\frac{s}{\sqrt{n}}$$

$\bar{x}$ = 1990,50 $\frac{s}{\sqrt{n}} = \frac{2833{,}33}{\sqrt{140}}$ = 239,46 t = 1,977

O interval de confiança de 95% é:
1990,50 ± 1,977(239,46) =
[R$1.517, R$2.464]

MARGEM DE ERRO E AMPLITUDE

Amplitude do IC

Margem de erro: $z\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

Amplitude do Intervalo de Confiança: $2\left(z\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right)$

Precisão do IC

$$\bar{p} \pm z\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Quanto menor a margem de erro/amplitude do intervalo de confiança, mais precisa é a estimação.

A margem de erro e a amplitude do intervalo de confiança são influenciadas:

- Tamanho da amostra n: quanto maior o tamanho da amostra, menor a margem de erro.
- Nível de confiança: quanto maior o nível de confiança, maior a margem de erro.

Interpretando um IC

<u>Incorreto</u>: A probabilidade que μ esteja no intervalo é 95% (isso implicaria que μ é uma variável, o que é errado).

<u>Correto</u>: Temos 95% de confiança que a média populacional μ vai estar no intervalo.

Já que podemos ter a possibilidade de muitas amostras, estaremos corretos 95% do tempo, então nos dando 95% de confiança (vamos estar errados 5% do tempo, em média!)

Interpretando um IC

Incorreto: Existe uma probabilidade de 95% do saldo médio mensal dos alunos ser entre R$1.517 e R$2.464.

Correto: O saldo médio mensal dos alunos está entre R$1.517 e R$2.464, com 95% de confiança.

Simulação

https://rpsychologist.com/d3/ci/

COMPARANDO DOIS INTERVALOS

Comparando 2 intervalos

Washington Post: “Hillary Clinton mantém uma estreita vantagem de quatro pontos sobre Donald Trump na Washington Post-ABC News Tracking Poll,[...] A pesquisa, que revela que 47% dos prováveis eleitores apoiam Clinton, enquanto 43% apoiam Trump, foi realizada com 2.220 eleitores [...] A pesquisa teve uma margem de 2,5 pontos percentuais no erro de amostragem.”

Comparando 2 intervalos

Com base nessa pesquisa, temos 95% de confiança de que Clinton estava à frente naquele ponto?

Clinton: 47 ± 2,5: [44,5; 49,5]

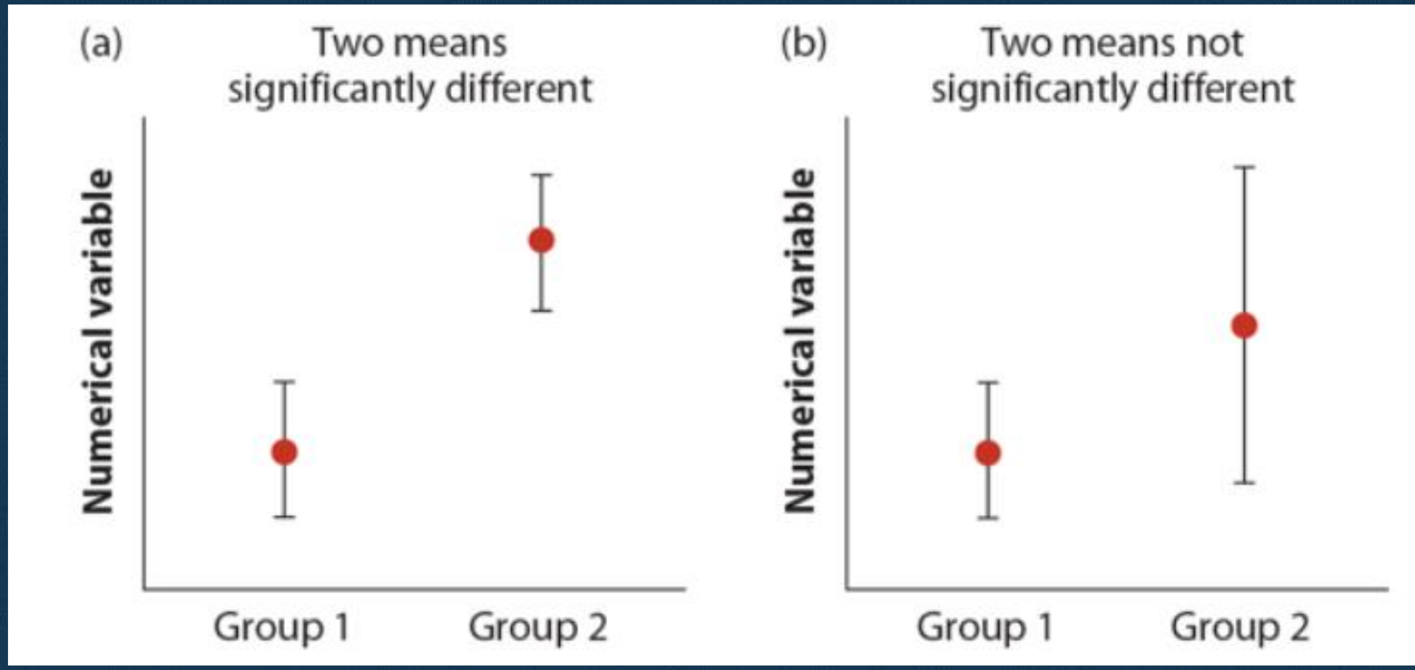
Trump: 43 ± 2,5: [40,5; 45,5]

Washington Post: “A vantagem de Clinton na pesquisa Post-ABC não atinge significância estatística dada a margem de 2,5 pontos percentuais da pesquisa no erro de amostragem em torno do apoio de cada candidato.”

Comparando 2 intervalos

Se não houver interseção: há uma diferença significativa entre as duas médias.

Há interseção: A diferença é inconclusiva.



TAMANHO DA AMOSTRA

Calculando o tamanho da amostra

Partindo da margem de erro que quero obter, escolhemos o nível de confiança.

<u>Exemplo</u>: 95% IC com 2% ME

$$1{,}96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0{,}02 \quad \Rightarrow \quad n = \left(\sqrt{\bar{p}(1-\bar{p})} \times \frac{1{,}96}{0{,}02}\right)^2$$

Sendo $\bar{p}$ = 0,5 ➡ n = 2.401

Calculando o tamanho da amostra

https://pt.surveymonkey.com/mp/sample-size-calculator/