Universidade Federal de São Paulo

Intervalo de confiança

Professor

Julio Cezar

Vimos na aula anterior que:

Um processo de indução, na qual usamos dados extraídos de uma amostra para produzir inferência sobre a população é chamado de **estimação**.

Tipos de estimação de parâmetros

- Estimação pontual.
- Estimação intervalar.

ESTIMAÇÃO PONTUAL

Quando temos o interesse de estudar um parâmetro de uma população, para isso, retira-se uma amostra, de tamanho n, dessa população que estamos trabalhando, e através desta amostra, **estima-se** o **parâmetro populacional** (μ, σ^2 e p) através dos **estimadores**:

- $\mathbf{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$ (o estimador para estimar μ).
- $\mathbf{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ (é um estimador não viesado para estimar σ^2).
- $\mathbf{\hat{p} = \frac{\%\ com\ característica}{n}}$ (é o estimador para a proporção p. Note que, $\hat{p}$ é a proporção amostral).

Assim, $\bar{X}$, S^2e $\hat{p}$ produzem, para a amostra selecionada, as **estimativas pontuais** (valor único para cada amostra selecionada).

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Estimação pontual

Então, os estimadores (estatísticas) discutidos até aqui são estimadores pontuais, pois como já dito, fornecem como estimativa um único valor numérico para o parâmetro de interesse.

Tabela de estimadores pontuais.

	Parâmetro (População)	Estimador (amostra)
Tamanho	N	n
Média	μ	$\hat{\mu} = \bar{X}$
Variância	σ^2	$\hat{\sigma}^2 = S^2$
Desvio padrão	σ	$\hat{\sigma} = S$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

ESTIMAÇÃO INTERVALAR

Mas podemos apresentar uma estimativa mais informativa para o parâmetro. Então, uma outra maneira de se calcular uma estimativa de um parâmetro, é construir um intervalo de confiança, ou seja, ao invés de apresentar uma estimativa pontual podemos apresentar um **intervalo de confiança** para o parâmetro.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

ESTIMAÇÃO INTERVALAR

Portanto, **na aula de hoje**:

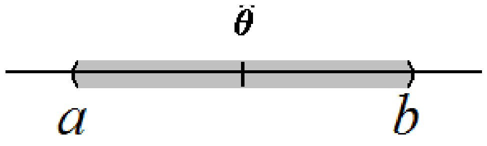
- Intervalo de Confiança para média.

- Intervalo de Confiança para proporção.

- Intervalo de Confiança para variância.

ESTIMAÇÃO INTERVALAR

- No processo de "investigação" de um **parâmetro $\boldsymbol{\theta}$**, necessitamos ir além da **estimativa pontual $\hat{\boldsymbol{\theta}}$**. Questionamentos quando não se conhece θ:
  - Quão próximo estamos do valor real do parâmetro θ quando obtemos sua estimativa? Depende da precisão (ou variância) do estimador.
  - Uma maneira de contornar tal questionamento consiste em se encontrar um intervalo em torno da **estimativa pontual $\hat{\boldsymbol{\theta}}$** que tenha alta probabilidade de englobar o **parâmetro $\boldsymbol{\theta}$**.



P(do intervalo [a,b] englobar o parâmetro θ) =γ

- O intervalo acima é construído com base na amostra dos dados, ou seja, a partir dos dados e da distribuição amostral associada a **estimativa pontual $\hat{\boldsymbol{\theta}}$**.

Estimação intervalar

Seja $(X_1, X_2,\ldots, X_n)$ uma amostra aleatória de tamanho n de uma população e **θ o parâmetro de interesse**. Sejam $\hat{\theta}_1$ e $\hat{\theta}_2$ **estatísticas** tais que:

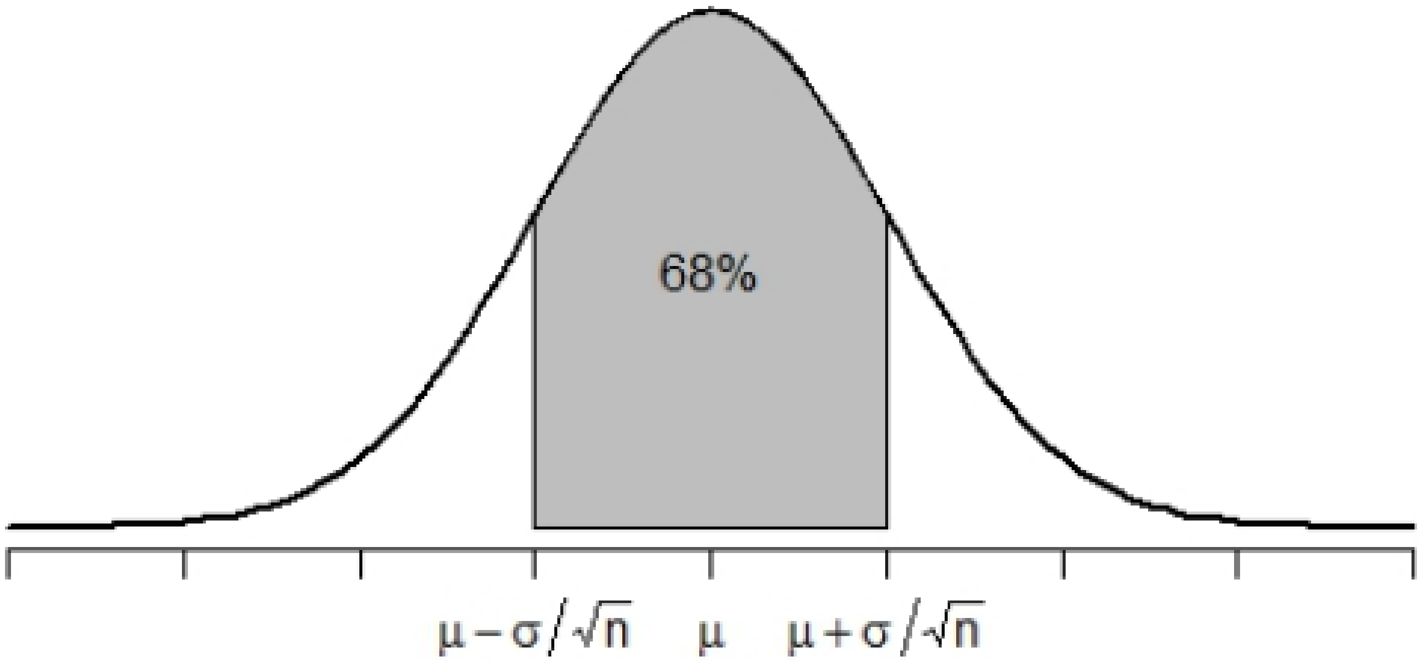
$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) - 1 - \alpha$$

Então o intervalo $\widehat{(\theta_1; \hat{\theta}_2)}$ é chamado **intervalo de confiança** de nível $\mathbf{(1-\alpha)100\%}$ para o parâmetro θ. Usualmente toma-se o **nível de confiança** $(1-\alpha)$ igual à $0{,}95$ ou $0{,}99$. **Interpretação:** De todos os possíveis intervalos que possam ser construídos, espera-se que $(1-\alpha)100\%$ deles contenham o verdadeiro valor do parâmetro θ. (**Obs: α** é o **nível de significância** utilizado para calcular o **nível de confiança**, em outras palavras, um **nível de significância** **α** de **0,05** indica um **nível de confiança** de **95%**.)

UNIFESP
Universidade Federal de São Paulo
1933

ESTIMAÇÃO INTERVALAR: INTERPRETAÇÃO

Distribuição Normal

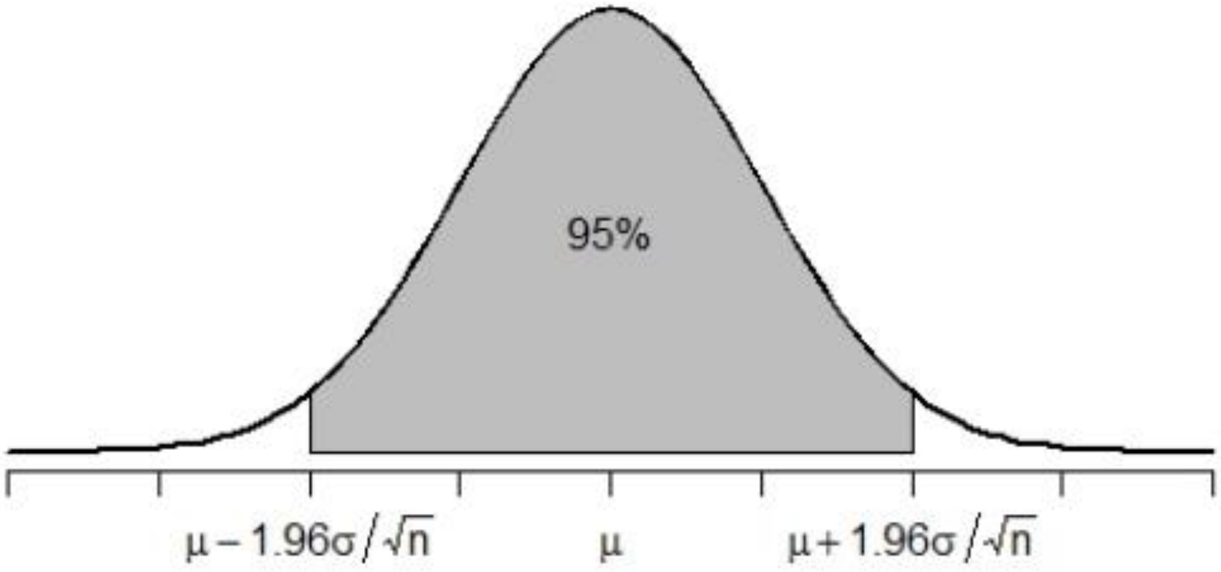


Podemos dizer que 68% dos possíveis valores da média de uma amostra aleatória simples de tamanho n não se afastam mais do que $\sigma/\sqrt{n}$ (**Erro Padrão**) .

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Estimação intervalar: interpretação

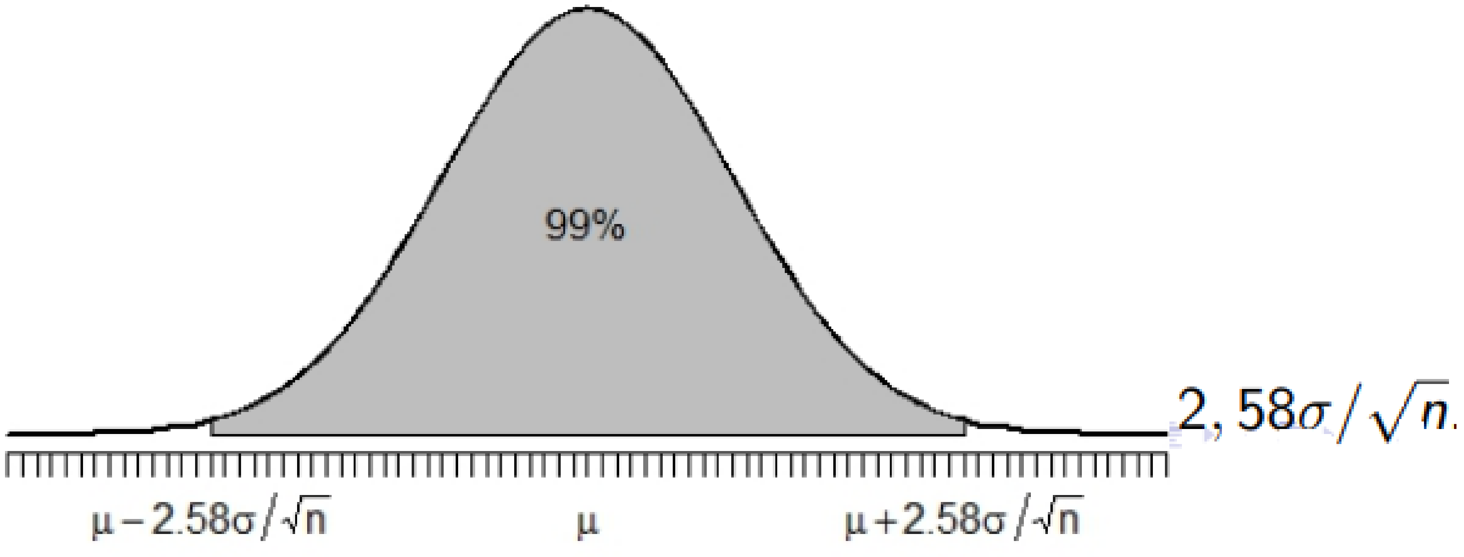
Distribuição Normal



Podemos dizer que 95% dos possíveis valores da média de uma amostra aleatória simples de tamanho n não se afastam mais do que $1{,}96\sigma/\sqrt{n}$.

Estimação intervalar: interpretação

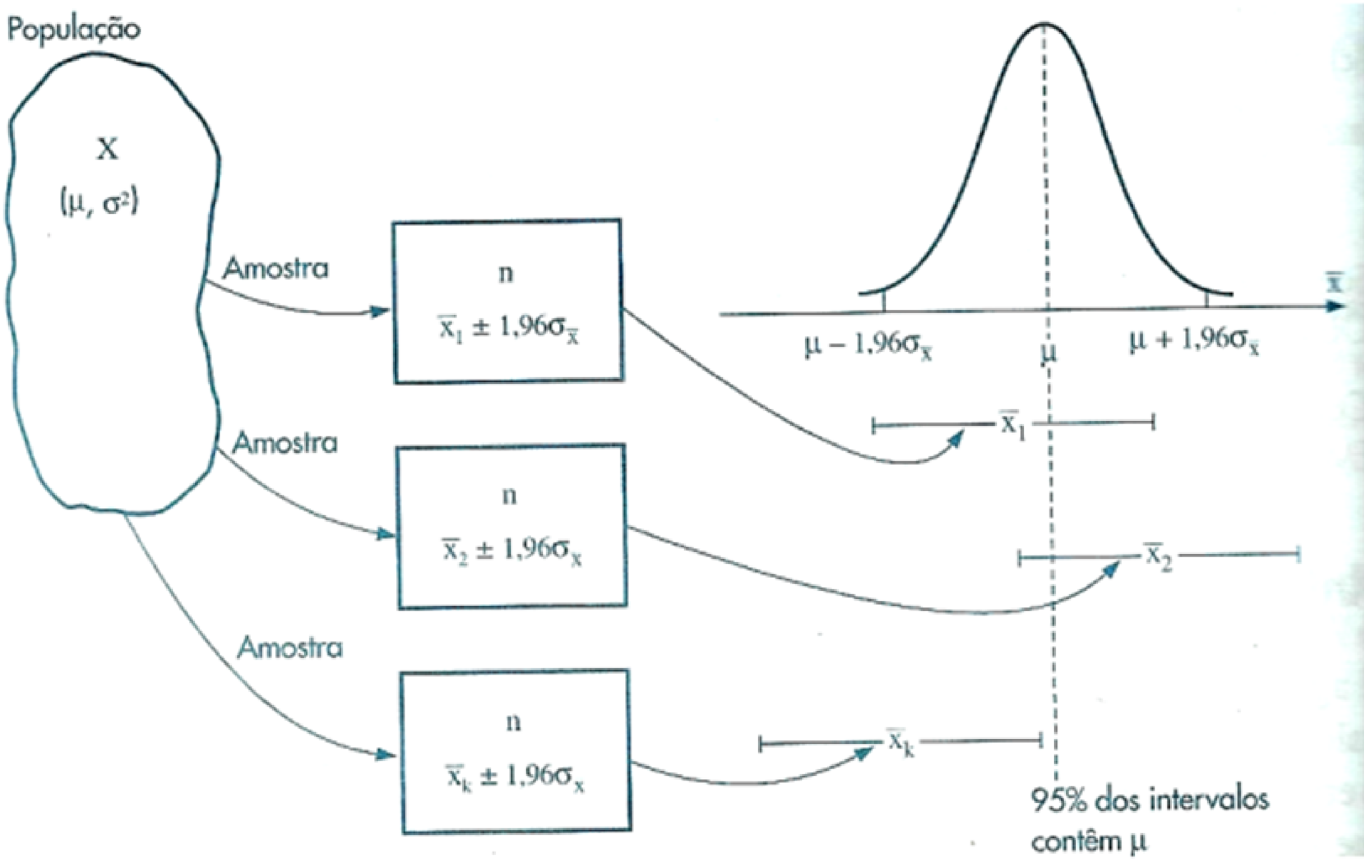
Distribuição Normal



Podemos dizer que 99% dos possíveis valores da média de uma amostra aleatória simples de tamanho n não se afastam mais do que $2{,}58\sigma/\sqrt{n}$.

ESTIMAÇÃO INTERVALAR: INTERPRETAÇÃO

Significado de um IC para μ, com 95% de confiança e σ^2 conhecido

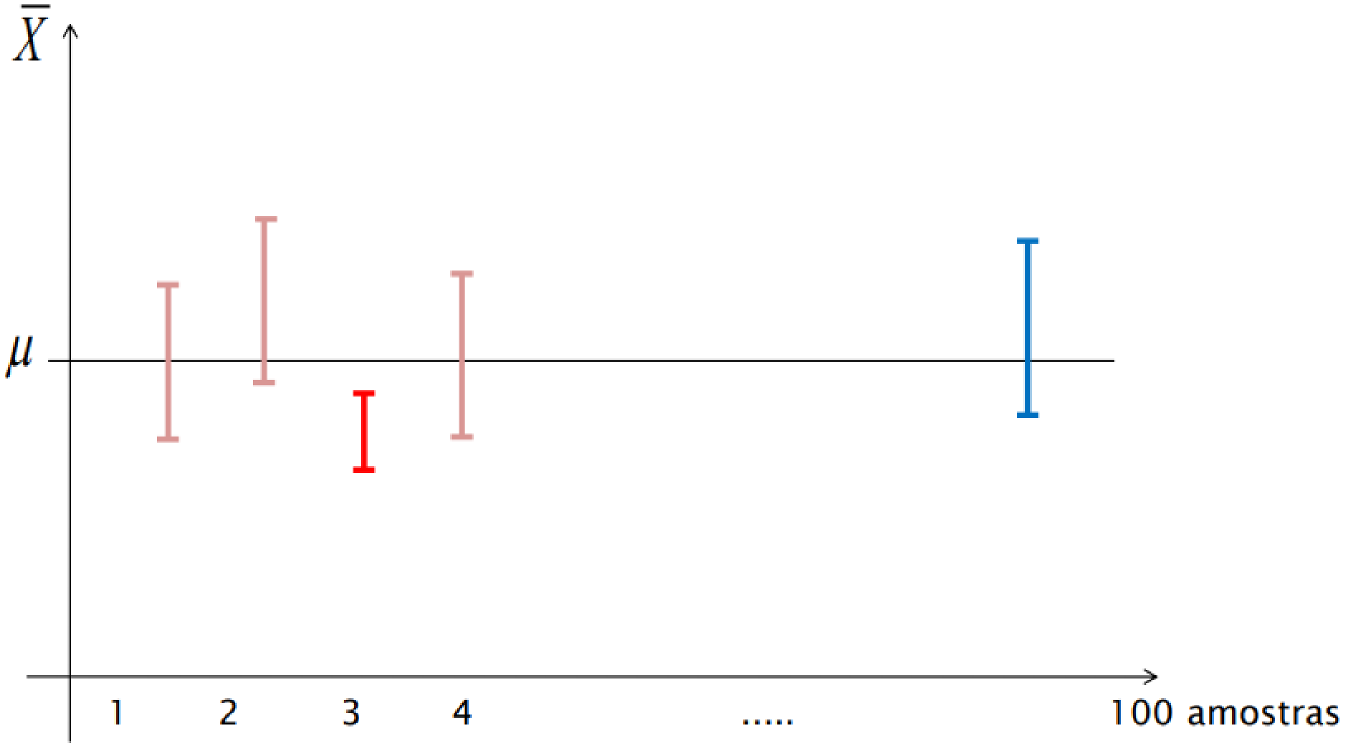


Fonte: Bussab W De O, Morettin PA. Estatística Básica. São Paulo: Saraiva, 2006.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Estimação intervalar: interpretação

Ou seja, se obtivermos um intervalo de confiança para o parâmetro θ, por exemplo, para cada uma dentre 100 amostras aleatórias da população, somente 5, em média destes intervalos de confiança não conterão θ (**parâmetro de interesse**).



UNIFESP
Universidade Federal de São Paulo
1933

Composição de um intervalo de confiança

1) Estimativa pontual do parâmetro.

$$\bar{x}, \quad \hat{p}, \quad s^2$$

2) Percentil crítico da distribuição de probabilidade.

$$z_{\alpha/2}, t_{\alpha/2}$$

3) Erro padrão da estimativa.

$$EP = \sigma/\sqrt{n}$$

4) Margem de erro.

ME = percentil crítico × erro padrão

UNIFESP
Universidade Federal de São Paulo
1933

Composição de um intervalo de confiança

1) Estimativa pontual do parâmetro.

$$\bar{x}, \quad \hat{p}, \quad s^2$$

2) Percentil crítico da distribuição de probabilidade.

$$z_{\alpha/2}, t_{\alpha/2}$$

3) Erro padrão da estimativa.

$$EP = \sigma/\sqrt{n}$$

4) Margem de erro.

ME = percentil crítico × erro padrão

IC(θ; 1 – α) = Estimativa por ponto ± Margem de erro.

UNIFESP
Universidade Federal de São Paulo
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Seja $X_1, \dots, X_n$ uma amostra de uma variável aleatória X com média μ (**desconhecida**) e variância $\boldsymbol{\sigma^2}$ **conhecida.** Para construir um intervalo de confiança para a média μ devemos considerar a distribuição da média amostral de $\bar{X}$.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right) \Rightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

O intervalo de confiança $\gamma = (1 - \alpha)100\%$ para μ, temos que obter as constantes a e b, tal que $P(\mathbf{a} \leq \mu \leq \mathbf{b}) = (1 - \alpha)$. Lembre-se que, a probabilidade **γ = (1-α)** é o **nível de confiança** do intervalo e **α** é o **nível de significância**.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Assim, fixando o **nível de confiança (1-α)**, pode-se determinar o intervalo de confiança da seguinte forma:

Sabemos que $X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

como $P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = \gamma$

então substituindo Z, tem-se $P\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < Z_{\frac{\alpha}{2}}\right) = \gamma$

Quantidade Pivotal

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Assim,

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right) = \gamma$$

$$P\left(-Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = \gamma$$

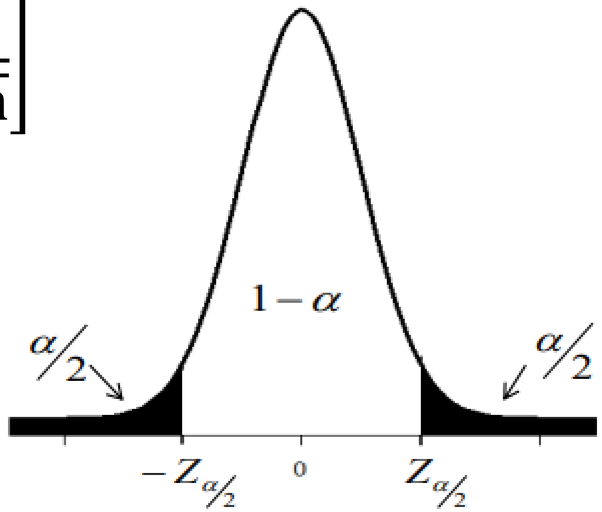
$$P\left(\underbrace{\overline{X} - Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}}_{a \text{ limite inferior}} < \mu < \underbrace{\overline{X} + Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}}_{b \text{ limite superior}}\right) = \gamma$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Portanto, um intervalo de confiança $(1-\alpha)100\%$ para μ com σ^2 conhecido, é dado por:

$$IC_{(1-\alpha)\%}(\mu) = \left[\bar{X} - Z_{\frac{\alpha}{2}} \text{ x } \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_{\frac{\alpha}{2}} \text{ x } \frac{\sigma}{\sqrt{n}}\right]$$



Se $\alpha = 0{,}05$, então $\alpha/_2 = 0{,}025$, assim $Z_{0,025} = -1{,}96$, logo um

I.C. 95% para μ, com σ^2 conhecido, é dado por:

$$\left[\bar{X} - 1{,}96 \text{ x } \frac{\sigma}{\sqrt{n}} ; \bar{X} + 1{,}96 \text{ x } \frac{\sigma}{\sqrt{n}}\right]$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança de 44 < μ < 49.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança de $44 < \mu < 49$.

Interpretação 1

Temos 95% de confiança de que a verdadeira média populacional μ se encontra entre 44 e 49.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança de 44 < μ < 49.

Interpretação 1

Temos 95% de confiança de que a verdadeira média populacional μ se encontra entre 44 e 49.

Interpretação 2

Temos 95% de confiança de que o intervalo entre 44 e 49 realmente contém a verdadeira média populacional μ.

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Interpretação de um intervalo de confiança

Suponha que obtivemos um intervalo de 95% de confiança de $44 < \mu < 49$.

Interpretação 1 — ERRADA

Temos 95% de confiança de que a verdadeira média populacional μ se encontra entre 44 e 49.

Interpretação 2 — CERTA

Temos 95% de confiança de que o intervalo entre 44 e 49 realmente contém a verdadeira média populacional μ.

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Como o intervalo de confiança é calculado a partir de uma **amostra aleatória**, este intervalo **também é aleatório**!

Isso significa que para cada amostra aleatória que tivermos, um intervalo **diferente** será calculado.

Como o valor de μ é fixo, é o intervalo que deve conter o valor de μ, e não o contrário.

Isso significa que se pudéssemos obter 100 amostras diferentes, e calcularmos um intervalo de confiança de 95% para cada uma das 100 amostras, esperaríamos que 5 destes intervalos **não** contenham o verdadeiro valor da média populacional μ.

COMPOSIÇÃO DE UM INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

1) Estimativa pontual do parâmetro.

$$\bar{x}$$

2) Percentil crítico da distribuição de probabilidade.

$$z_{\frac{\alpha}{2}}$$

3) Erro padrão da estimativa.

$$EP = \frac{\sigma}{\sqrt{n}}$$

4) Margem de erro.

$$ME = z_{\frac{\alpha}{2}} \, x \, \frac{\sigma}{\sqrt{n}}.$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

COMPOSIÇÃO DE UM INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

1) Estimativa pontual do parâmetro.

$$\bar{x}$$

2) Percentil crítico da distribuição de probabilidade.

$$z_{\frac{\alpha}{2}}$$

3) Erro padrão da estimativa.

$$EP = \frac{\sigma}{\sqrt{n}}$$

4) Margem de erro.

$ME = z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$. Então $IC(\mu; 1 - \alpha) = \bar{x} \pm ME$.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 CONHECIDA OU n > 30)

Exemplo: Um pesquisador obteve a partir de uma amostra uma média $\bar{x}$ = 180cm para altura de um determinado grupo de pessoas utilizando uma amostra n = 40, sabe-se que a variância populacional da altura é de σ^2=100 cm^2. Qual o intervalo de confiança de 95% para a média populacional?

DISTRIBUIÇÃO t-STUDENT

William Gosset (1876-1937) foi pioneiro no projeto e análise experimental de pequenas amostras com uma abordagem econômica para a lógica da incerteza. Gosset publicou sob o pseudônimo de Student e desenvolveu a mais famosa distribuição t-Student.

William Gosset (1876-1937).

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Distribuição t-Student

A variável aleatória T é dada por $T = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} \sim t_{(n-1)}$

Distribuição t-Student é:

- Simétrica em relação ao zero;
- Semelhante à distribuição normal padrão, porém com “caudas mais pesadas”;
- Para $n \to \infty$ ($n > 30$) a distribuição t tende para a normal padrão;

υ

- Existe uma curva para cada tamanho de amostra (n) e o valor $\nu = n - 1$ (número de graus de liberdade) é usado para obtenção de valores na tabela.

Distribuição t-Student

Uma distribuição t-Student com $\nu = n - 1$ graus de liberdade e, a função de densidade é dada por:

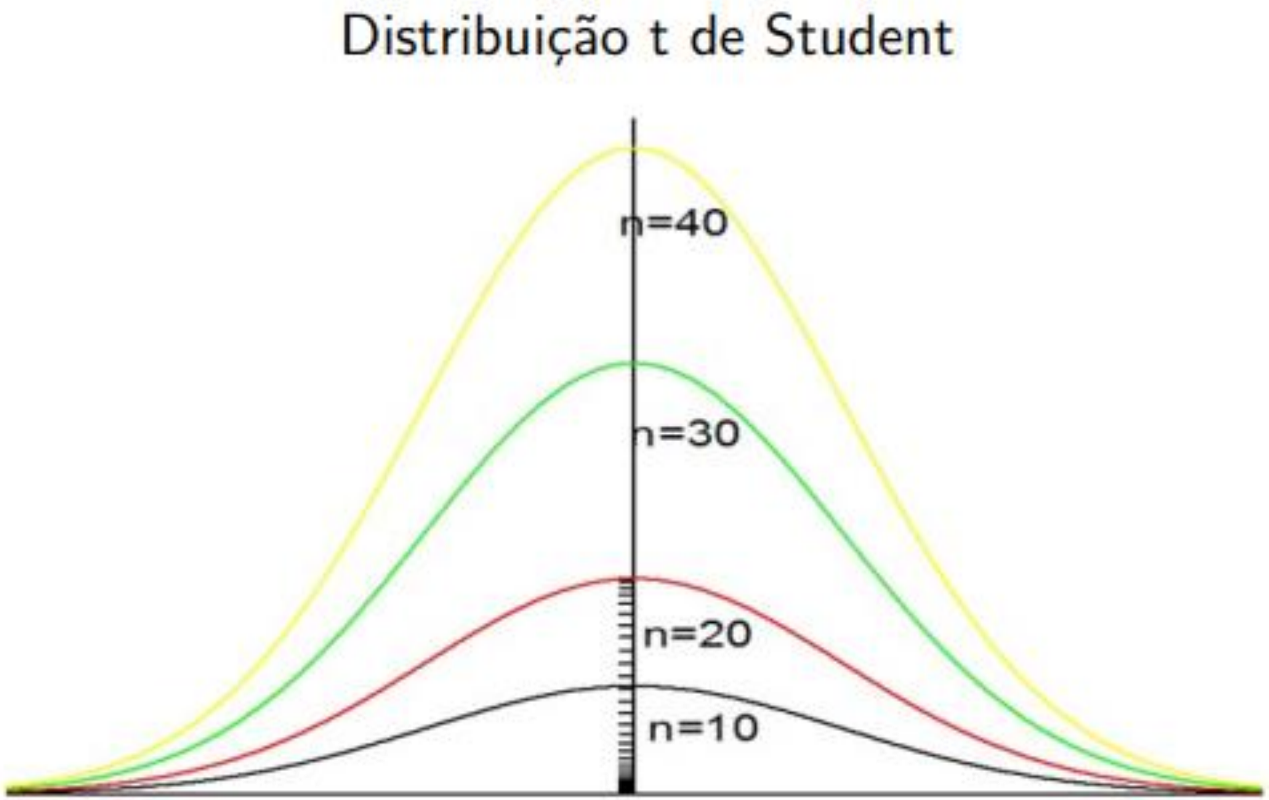
$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

Graus de liberdade

- Uma restrição matemática ao estimar uma estatística de uma outra estimativa.
- O número de informações independentes da amostra resulta no número de graus de liberdade.

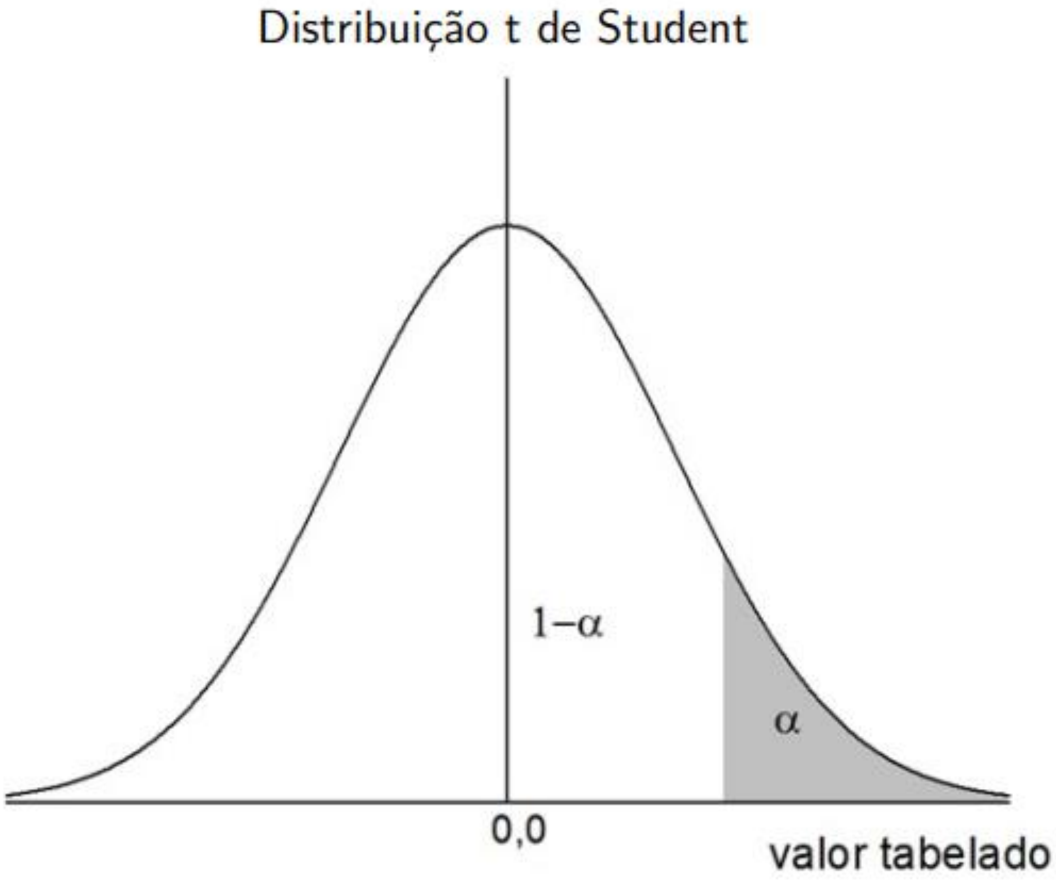
DISTRIBUIÇÃO t-STUDENT

Observe que, como já mencionado, a medida que n cresce a distribuição t se aproxima da distribuição normal padrão.



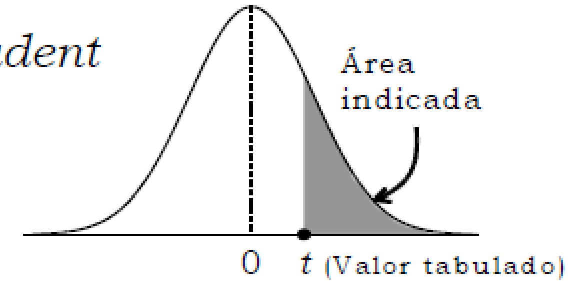
Distribuição t-student

Valores de probabilidade de t são obtidos em tabelas. A tabela de t informa o valor acima do qual se encontra a área α.



DISTRIBUIÇÃO t-STUDENT

Tabela 5 Distribuição ***t*** de *Student*



gl	Área na cauda superior								
	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

DISTRIBUIÇÃO t-STUDENT

Exemplo: Seja uma amostra n = 15. Qual é o valor de t acima do qual tem-se 5% de probabilidade. Graus de liberdade:

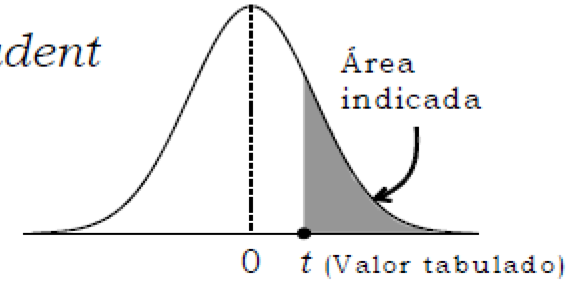
$$\nu = \mathrm{n} - 1$$

$$\nu = 15 - 1 = 14$$

Probabilidade: $\alpha = 0{,}05$.

DISTRIBUIÇÃO t-STUDENT

Tabela 5 Distribuição ***t*** de *Student*



gl	Área na cauda superior								
	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Distribuição t-student

Portanto, tem-se

$$t_{\alpha;n-1} = t_{0,05;14} = 1,761$$

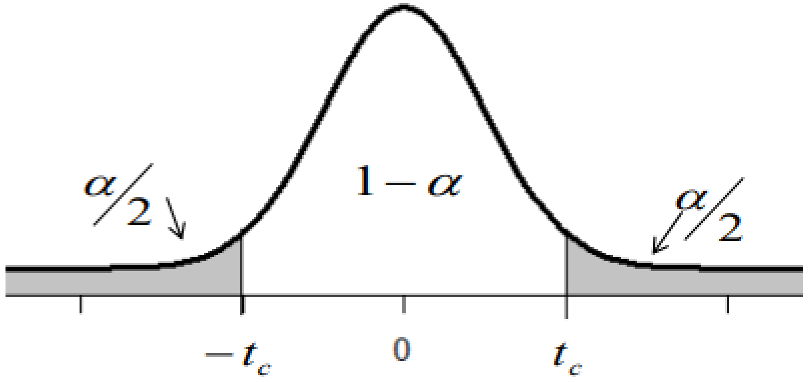
INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 DESCONHECIDA E $n \leq 30$)

Na prática quando não se conhece a média μ também não se conhece a variância, nesse caso, utilizamos o intervalo de confiança

$$P\left(\bar{X} - t_{\frac{\alpha}{2},n-1} x \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2},n-1} x \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$$IC_{(1-\alpha)\%}(\mu) = \left[\bar{X} - t_{\frac{\alpha}{2},n-1} \; x \; \frac{S}{\sqrt{n}} ; \bar{X} + t_{\frac{\alpha}{2},n-1} \; x \; \frac{S}{\sqrt{n}}\right]$$

Valores críticos



Composição de um Intervalo de confiança para a média (σ^2 desconhecida e $n \leq 30$)

1) Estimativa pontual do parâmetro.

$$\bar{x}$$

2) Percentil crítico da distribuição de probabilidade.

$$t_{\frac{\alpha}{2},n-1}$$

3) Erro padrão da estimativa.

$$EP = \frac{S}{\sqrt{n}}$$

4) Margem de erro.

$$ME = t_{\frac{\alpha}{2},n-1} \text{ x } \frac{S}{\sqrt{n}}.$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Composição de um Intervalo de confiança para a média (σ^2 desconhecida e $n \leq 30$)

1) Estimativa pontual do parâmetro.

$$\bar{x}$$

2) Percentil crítico da distribuição de probabilidade.

$$t_{\frac{\alpha}{2},n-1}$$

3) Erro padrão da estimativa.

$$EP = \frac{S}{\sqrt{n}}$$

4) Margem de erro.

$$ME = t_{\frac{\alpha}{2},n-1} \; x \, \frac{S}{\sqrt{n}}. \text{ Então } IC(\mu; 1-\alpha) = \bar{x} \pm ME.$$

UNIFESP

Universidade Federal de São Paulo

1933

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 DESCONHECIDA E $n \leq 30$)

Exemplo: Em uma determinada indústria para verificar a qualidade dos rolamentos esféricos produzidos foi tomado uma amostra ao acaso um lote de 15 peças, fornecendo um diâmetro médio de 240 cm com desvio padrão de 15 cm. Encontre um intervalo de confiança de 95% para o diâmetro.

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 DESCONHECIDA)

E se σ^2 desconhecida e $n > 30$?

INTERVALO DE CONFIANÇA PARA A MÉDIA (σ^2 DESCONHECIDA)

E se σ^2 desconhecida e n > 30?

Sabe-se que quando n > 30 a distribuição t-Student aproxima-se de uma distribuição normal.

$$IC_{(1-\alpha)\%}(\mu) = \left[\overline{X} - Z_{\frac{\alpha}{2}} \text{ x } \frac{S}{\sqrt{n}} ; \overline{X} + Z_{\frac{\alpha}{2}} \text{ x } \frac{S}{\sqrt{n}}\right]$$

em que, $Z_{\frac{\alpha}{2}}$ é tal que,

$$P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = \gamma$$

para $\gamma = (1 - \alpha)$ e $Z \sim N(0,1)$.

Intervalo de confiança para a proporção

Como a proporção p é de fato a média amostral de uma a.a. cuja v.a. tem distribuição de Bernoulli(p), para se construir intervalos de confiança para p devemos seguir os mesmos procedimentos anteriores.

Considerando que o estimador da **proporção $\hat{\mathbf{p}}$** tem valor esperado p e variância $\frac{p(1-p)}{n}$, ou seja,

$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ com distribuição $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$

Portanto, um intervalo de confiança $(1-\alpha)100\%$ para p, é dado por

$$IC_{(1-\alpha)\%}(p) = \left[\hat{p} - Z_{\frac{\alpha}{2}} \, x \sqrt{\frac{p(1-p)}{n}} ; \hat{p} + Z_{\frac{\alpha}{2}} \, x \sqrt{\frac{p(1-p)}{n}}\right]$$

UNIFESP

Universidade Federal de São Paulo

1933

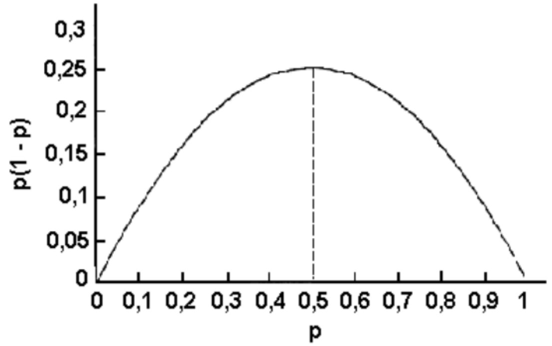
INTERVALO DE CONFIANÇA PARA A PROPORÇÃO

Em geral não temos conhecimento sobre p, assim podemos construir intervalos de confiança para a proporção substituindo p e $(1-p)$ por $\hat{p}$ e $(1-\hat{p})$, respectivamente. Neste caso um intervalo de confiança $(1-\alpha)100\%$ para p, é dado por

$$IC_{(1-\alpha)\%}(p) = \left[\hat{p} - Z_{\frac{\alpha}{2}} \, x \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + Z_{\frac{\alpha}{2}} \, x \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Outra possibilidade é considerar o fato de que $p(1-p) \leq {}^{1}/_{4}$ e construir um I.C. conservador para p assumindo $p = {}^{1}/_{2}$. Assim, $\frac{p(1-p)}{n} = \frac{1}{4n}$ com I.C. conservador para p

$$\left[\hat{p} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{4n}} ; \hat{p} + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{4n}}\right]$$



Composição de um Intervalo de confiança para a proporção

1) Estimativa pontual do parâmetro.

$$\bar{x}$$

2) Percentil crítico da distribuição de probabilidade.

$$z_{\frac{\alpha}{2}}$$

3) Erro padrão da estimativa.

$$EP = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

4) Margem de erro.

$$ME = z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

UNIFESP
Universidade Federal de São Paulo
1933

Composição de um Intervalo de confiança para a proporção

1) Estimativa pontual do parâmetro.

$$\bar{x}$$

2) Percentil crítico da distribuição de probabilidade.

$$z_{\frac{\alpha}{2}}$$

3) Erro padrão da estimativa.

$$EP = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

4) Margem de erro.

$$ME = z_{\frac{\alpha}{2}} \text{ x } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \text{ Então } IC(\hat{p}; 1-\alpha) = \hat{p} \pm ME.$$

UNIFESP
Universidade Federal de São Paulo
1933

INTERVALO DE CONFIANÇA PARA A PROPORÇÃO

Exemplo: Em uma linha de produção de certa mecânica, colheu-se uma amostra de 100 itens, constatando-se que 4 peças eram defeituosas. Construa um IC para a proporção das peças defeituosas ao nível de confiança de 95%.

RESUMO

- **IC para a média quando a variância (σ^2) é conhecida ou n > 30:**

$$IC_{(1-\alpha)\%}(\mu) = \left[\bar{X} - Z_{\frac{\alpha}{2}} \text{ x } \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_{\frac{\alpha}{2}} \text{ x } \frac{\sigma}{\sqrt{n}}\right].$$

- **IC para a média quando a variância (σ^2) é desconhecida e n ≤ 30:**

$$IC_{(1-\alpha)\%}(\mu) = \left[\bar{X} - t_{\frac{\alpha}{2},n-1} \text{ x } \frac{S}{\sqrt{n}} ; \bar{X} + t_{\frac{\alpha}{2},n-1} \text{ x } \frac{S}{\sqrt{n}}\right].$$

- **IC para proporção:**

$$IC_{(1-\alpha)\%}(p) = \left[\hat{p} - Z_{\frac{\alpha}{2}} \text{ x } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + Z_{\frac{\alpha}{2}} \text{ x } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE μ E p

Se imaginarmos a amplitude do intervalo como a especificação de sua precisão ou acurácia, então seu nível de confiança (ou confiabilidade) estará inversamente relacionado à sua precisão.

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE μ E p

Se imaginarmos a amplitude do intervalo como a especificação de sua precisão ou acurácia, então seu nível de confiança (ou confiabilidade) estará inversamente relacionado à sua precisão. Uma estimativa altamente confiável do intervalo pode ser imprecisa, quando os pontos finais do intervalo estiverem muito distantes, enquanto um intervalo preciso pode exigir confiabilidade relativamente baixa.

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE μ E p

Se imaginarmos a amplitude do intervalo como a especificação de sua precisão ou acurácia, então seu nível de confiança (ou confiabilidade) estará inversamente relacionado à sua precisão. Uma estimativa altamente confiável do intervalo pode ser imprecisa, quando os pontos finais do intervalo estiverem muito distantes, enquanto um intervalo preciso pode exigir confiabilidade relativamente baixa. Dessa forma, **não se pode dizer inequivocamente que um intervalo de 99% será preferível a um intervalo de 95%**; o ganho na confiabilidade exige uma perda na precisão. Uma estratégia interessante é especificar o nível de confiança desejado e a amplitude do intervalo e então **determinar o tamanho necessário da amostra**.

UNIFESP

UNIVERSIDADE FEDERAL DE SÃO PAULO

1933

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE μ E *p*

- **O erro máximo de estimação na estimação de μ é dado por**

$$\epsilon = Z_{\frac{\alpha}{2}} \text{ x } \frac{\sigma}{\sqrt{n}} \longrightarrow n = \frac{Z_{\frac{\alpha}{2}}^2 \text{ x } \sigma^2}{\epsilon^2}$$

Ou seja, o erro máximo de estimação é a metade da largura do intervalo desejado.

Por ex.: ±1mm, ±5kg, ±0,1MPa

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE μ E *p*

- **O erro máximo de estimação na estimação de *p* é dado por**

$$\epsilon = Z_{\frac{\alpha}{2}} \text{ x} \sqrt{\frac{p\,(1-p)}{n}} \longrightarrow n = \frac{Z_{\frac{\alpha}{2}}^{2} \text{ x } p(1-p)}{\epsilon^{2}}$$

No caso onde não temos informação de p

$$n = \frac{Z_{\frac{\alpha}{2}}^{2} \text{ x} \frac{1}{4}}{\epsilon^{2}}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE *p*

- No caso onde não temos informação sobre *p* e de estarmos usando nível de confiança de 95% ($Z_{0,025}$ =1,96 ≅2), então:

$$n_0 = \frac{1}{\epsilon^2}$$

Muito utilizada no planejamento de pesquisa de levantamento, com o objetivo de estimar várias proporções, por exemplo:

- **Pesquisa eleitoral**
- **Pesquisa de mercado**

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE μ

Exemplo 1: De experiências passadas, sabe-se que o desvio padrão da altura de crianças de 5ª série do 1º grau é 5cm.

(a) Colhendo uma amostra de 36 dessas crianças observou-se a média de 150cm. Qual o intervalo de confiança de 95% para a média populacional?

(b) Que tamanho deve ter uma amostra para que o intervalo 150 ± 0,98 tenha 95% de confiança?

Cálculo do tamanho amostral para estimação de *p*

Exemplo 2: Suponha que numa pesquisa de mercado estima-se que no mínimo 60% das pessoas entrevistadas preferirão a marca A de um produto. Essa informação é baseada em pesquisas anteriores. Se quisermos que o erro am$\hat{p}$ stral de seja menor do que $\epsilon = 0{,}03$, com probabilidade de 95%, teremos n igual a?

CÁLCULO DO TAMANHO AMOSTRAL PARA ESTIMAÇÃO DE p

Exemplo 3: O serviço social de um município deseja determinar a proporção de famílias com uma renda familiar inferior a R$200,00. Estudos anteriores indicam que esta proporção é de 20%.

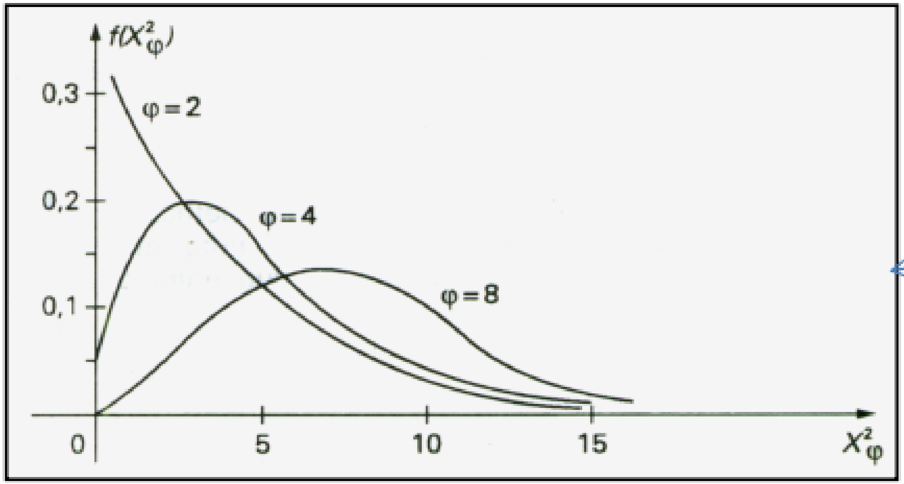
(a) Que tamanho de amostra se requer para assegurar uma confiança de 95% que o erro máximo de estimação desta proporção não ultrapasse 0,05?

(b) Em quanto variará o tamanho da amostra se o erro máximo permissível é reduzido a 0,01?

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Teorema: Seja $(X_1, X_2, \dots, X_n)$ uma amostra aleatória de tamanho n de uma população $N(\mu, \sigma^2)$. A variável aleatória Q tem distribuição chi-quadrado com $\upsilon = n - 1$ graus de liberdade, ou seja,

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$



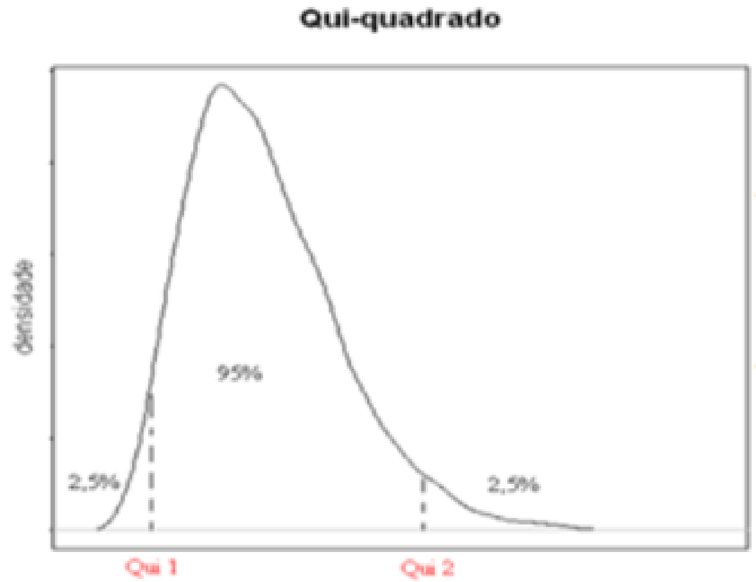
A distribuição χ^2 é assimétrica e positiva para qualquer grau de liberdade.

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Para construir um I.C. a $(1-\alpha)100\%$ para σ^2 parte-se de

$$P\left(\chi_1^2 \le \frac{(n-1)S^2}{\sigma^2} \le \chi_2^2\right) = 1-\alpha = \gamma$$

Dado que a distribuição de Q **não é simétrica** é preciso determinar os valores tabelados, tal que:



$$P\left(Q \ge \chi_2^2\right) = \frac{\alpha}{2}$$

$$P\left(Q \le \chi_1^2\right) = \frac{\alpha}{2} \Rightarrow P\left(Q > \chi_1^2\right) = 1 - \frac{\alpha}{2}$$

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Assim,

$$P(Q \geq \chi_2^2) = \frac{\alpha}{2}$$

$$P(Q \leq \chi_1^2) = \frac{\alpha}{2} \quad \longrightarrow \quad P(Q > \chi_1^2) = 1 - \frac{\alpha}{2}$$

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Portanto, um Intervalo de Confiança a $(1-\alpha)100\%$ para σ^2 é

$$IC_{(1-\alpha)\%}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi_2^2}; \frac{(n-1)S^2}{\chi_1^2}\right]$$

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Exemplo 1: Numa linha de produção, é muito importante que o tempo gasto numa determinada operação não varie muito de empregado para empregado. Considere a seguinte amostra:

125, 135, 115, 120, 150, 130, 125, 145, 125, 140 e 130.

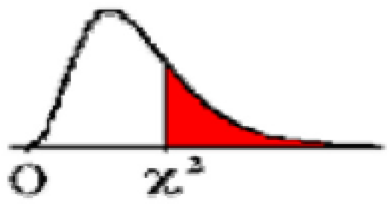
Determine:

a) Que parâmetro poderia ser usado para avaliar esse fato?

b) Encontre a estimativa pontual e o intervalo de confiança de 95%.

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Distribuição Qui-Quadrado



A tabela fornece os valores "c" tais que
$P(\chi^2 > c) = p$
onde "n" é o número de graus de liberdade e "p" é a probabilidade de sucesso.

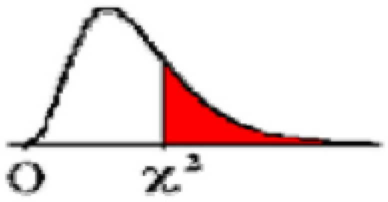
gl	0,995	0,990	0,975	0,950	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	$\boxed{3{,}247}$	3,940	4,865	9,342	15,987	18,307	$\boxed{20{,}483}$	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	12,340	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Exemplo 2: As notas dos alunos da UNIFESP são normalmente distribuídas. O desvio-padrão das notas de oito estudantes escolhidos aleatoriamente do BCT é 2,4. Determine um intervalo de 90% de confiança para o desvio padrão populacional.

INTERVALO DE CONFIANÇA PARA A VARIÂNCIA σ^2 DE POPULAÇÃO NORMAL

Distribuição Qui-Quadrado



A tabela fornece os valores "c" tais que
$P(\chi^2 > c) = p$
onde "n" é o número de graus de liberdade e "p" é a probabilidade de sucesso.

gl	0,995	0,990	0,975	0,950	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	$\boxed{2{,}167}$	2,833	6,346	12,017	$\boxed{14{,}067}$	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	12,340	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718

Referências

ANDRADE, D. F.; OGLIARI, P. J. **Estatística para as ciências agrárias e biológicas com noções de experimentação**. Editora da UFSC, Florianópolis, 2007.

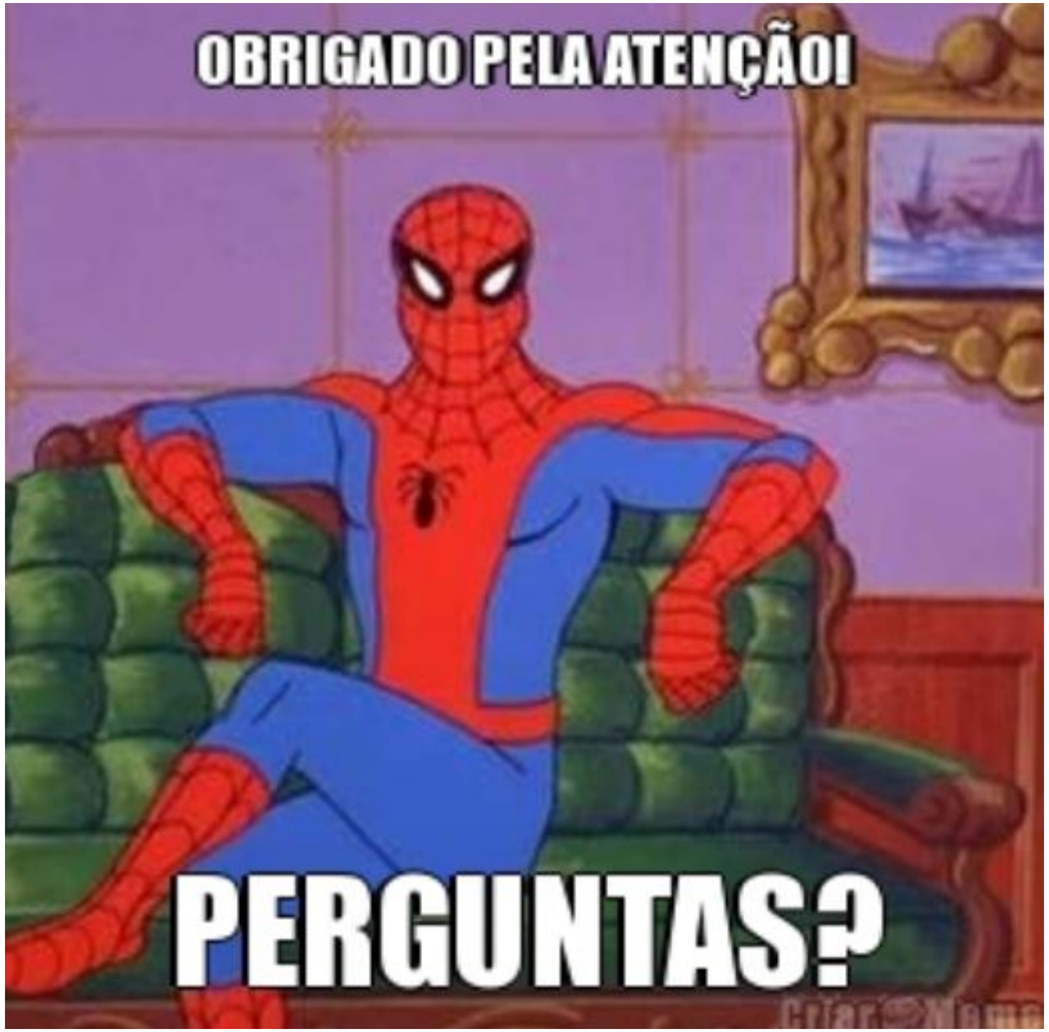
MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 6 ed. Rio de Janeiro: LTC, 2018. 628p.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. 9 ed. São Paulo: Saraiva, 2017. 554p.

MOORE, D. S. NOTZ, W. I.; FLIGNER, M. A. **A estatística básica e sua prática**. 7 ed. Rio de Janeiro: LTC, 2017. 628p.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Class finished



UNIFESP
Universidade Federal de São Paulo
1933