Universidade Federal do Paraná
Campus Avançado de Jandaia do Sul
Estatística

Estimativas e Tamanhos Amostrais

Dr. Landir Saviniec

E-mail: landir.saviniec@gmail.com
Homepage: github.com/lansaviniec

Outubro de 2018

Visão geral

Nesta parte, queremos **estimar os parâmetros média e variância de uma população a partir de uma amostra**.

Tipos de estimativas

Estimativa pontual: é um único valor (ou ponto) usado para estimar um parâmetro populacional. Ex.: a estatística de uma amostra.

Intervalo de confiança (ou estimativa intervalar): é um intervalo de valores usado para estimar um parâmetro populacional.

Valores críticos

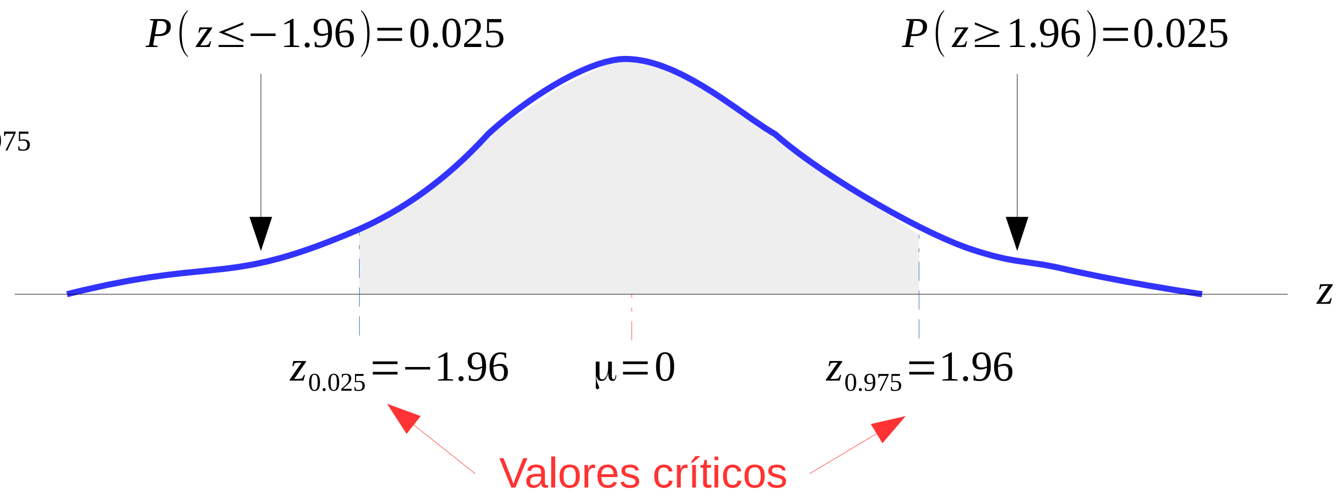
Em uma distribuição normal padrão, **um valor crítico** $z_{\alpha/2}$ é um escore z que **separa os valores** que tem ocorrência **provável, daqueles** que tem ocorrência **improvável**. Geralmente trabalhamos com $\alpha=0.01$, $\alpha=0.05$ e $\alpha=0.10$.

Exemplo: Em uma distribuição normal padrão, determine os valores de z que separam os 5% menos prováveis dos dados, dos 95% mais prováveis.

$\alpha=0.05$

$z_{\alpha/2}=z_{0.025}$

$z_{1-\alpha/2}=z_{0.975}$



Nível de confiança

Nível de confiança é a probabilidade (1 - α) de que o intervalo de confiança realmente contenha o parâmetro populacional. Geralmente trabalhamos com:

$$\alpha=0.1 \quad e \quad nível\, de\, confiança=0.9$$

$$\alpha=0.05 \quad e \quad nível\, de\, confiança=0.95$$

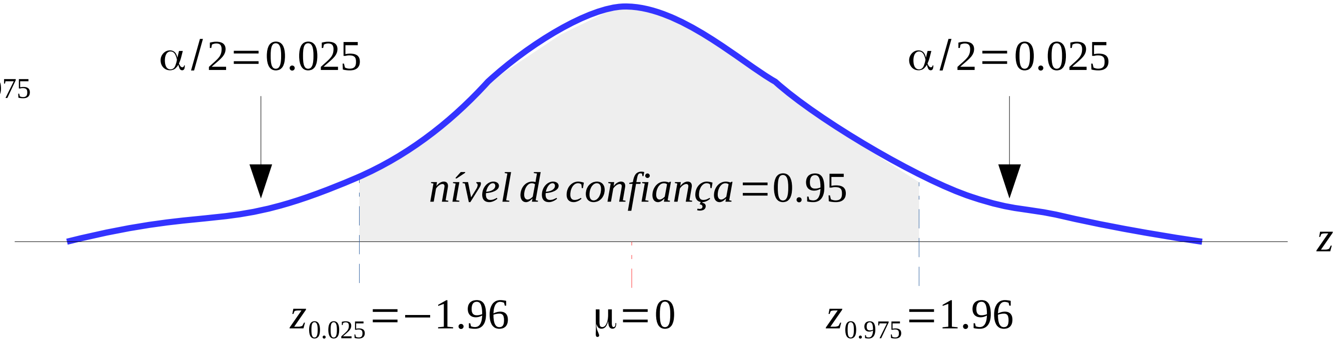
$$\alpha=0.01 \quad e \quad nível\, de\, confiança=0.99$$

Exemplo: Determinar os valores críticos para um nível de confiança de 95%.

$$\alpha=0.05$$

$$z_{\alpha/2}=z_{0.025}$$

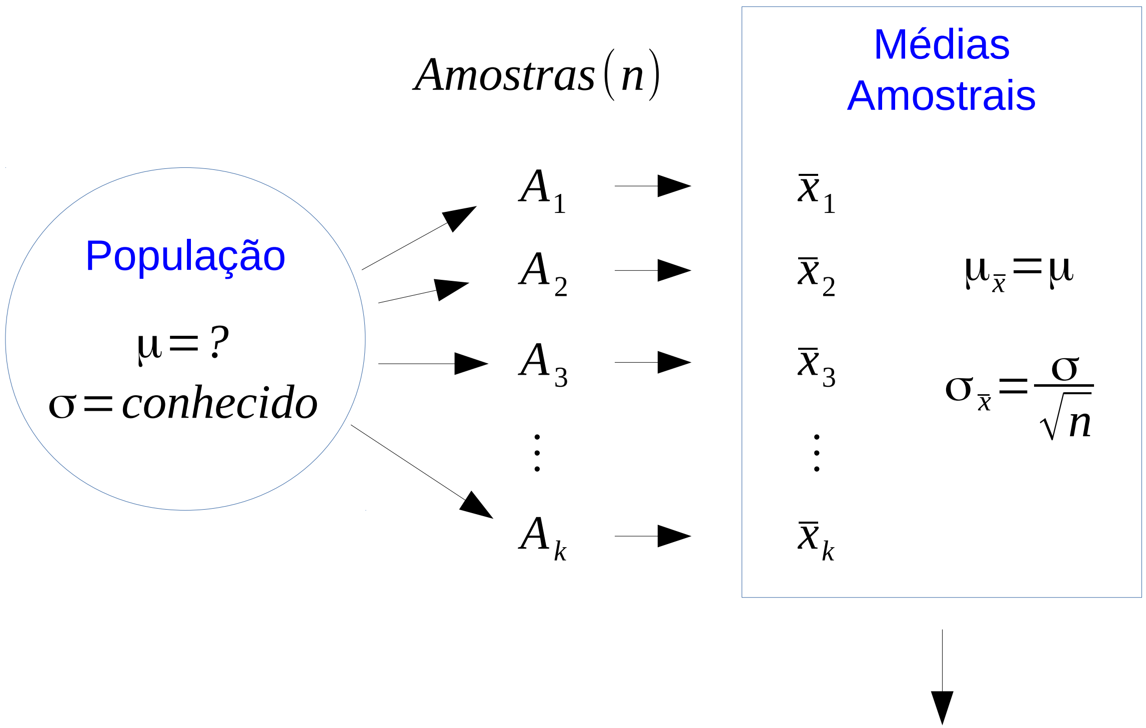
$$z_{1-\alpha/2}=z_{0.975}$$



Estimação da Média Populacional: desvio padrão conhecido

Objetivo

Estimar um intervalo de confiança para a média populacional a partir de uma amostra de tamanho n, com desvio-padrão populacional conhecido.



Normal se a população original é normal, ou se n>30.

Objetivo

Estimar um intervalo de confiança para a média populacional a partir de uma amostra de tamanho n, com desvio-padrão populacional conhecido.

Médias Amostrais

$$\bar{x}_1 \quad \bar{x}_2 \quad \bar{x}_3 \quad \vdots \quad \bar{x}_k$$

$$\mu_{\bar{x}} = \mu$$

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

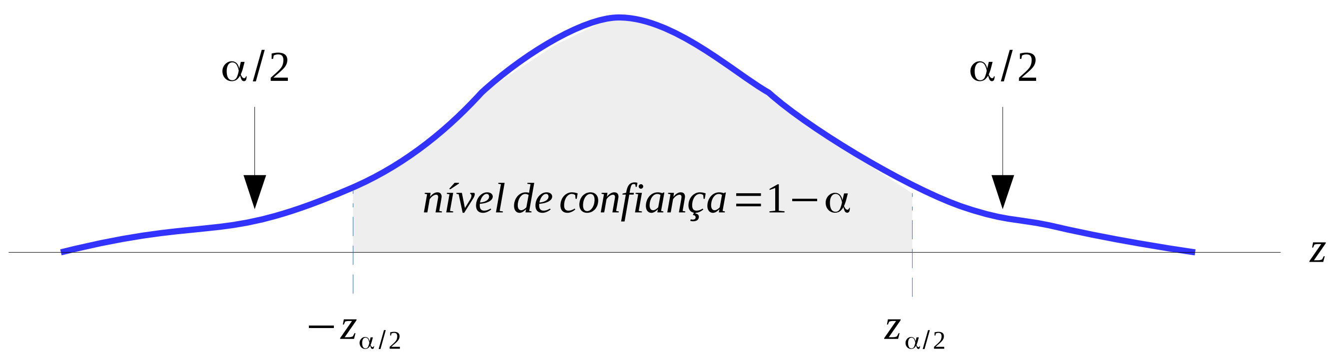
Convertendo para NP, temos: $z = \dfrac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

Para:

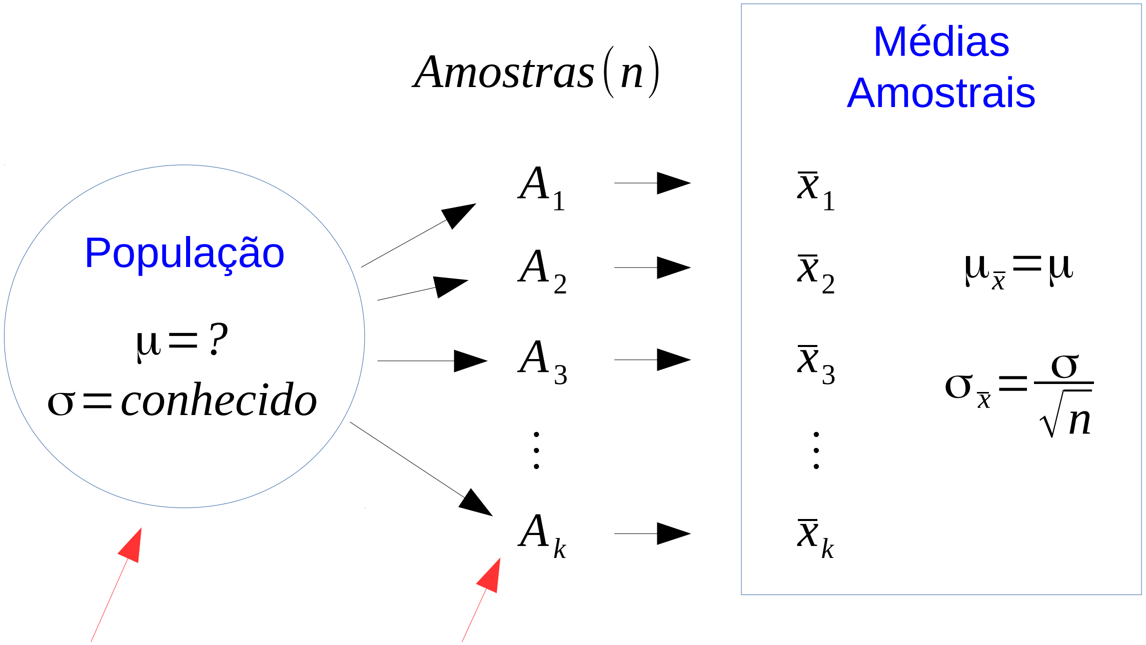
$$-z_{\alpha/2} \leq z \leq z_{\alpha/2} \Rightarrow -z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha/2}$$

$$-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow -\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



Intervalo de confiança para a média populacional



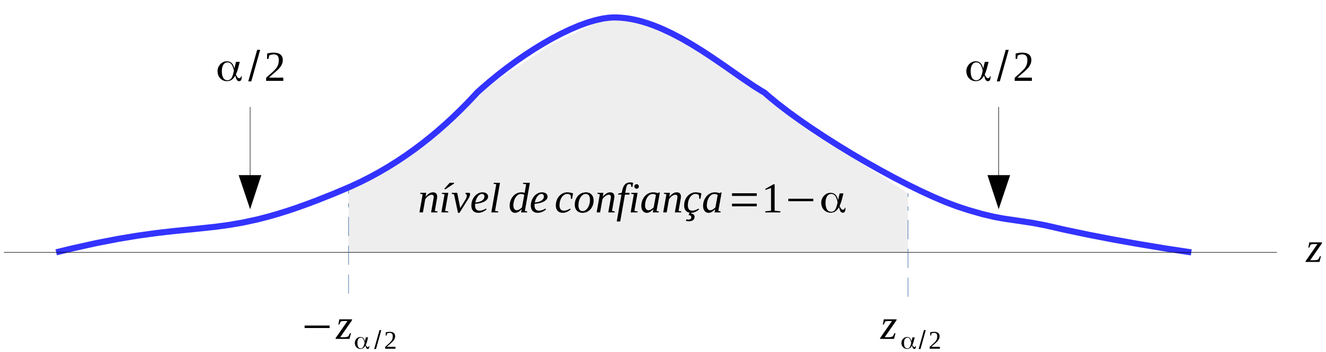
Requisito: Normal ou n>30.

Intervalo de confiança:

$$\bar{x} - E \;\leq\; \mu \;\leq\; \bar{x} + E$$

$Onde:$

$$E \;=\; z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



Exemplo 1

Considere os dados de alturas de pessoas na tabela abaixo como uma **população finita**.

Altura (m)	Altura (m)	Altura (m)	Altura (m)
1.84	1.75	1.81	17
1.88	1.63	1.6	1.8
1.76	1.62	1.7	1.65
1.65	1.81	1.83	1.56
1.74	1.9	1.89	1.7
1.9	1.82	1.8	1.89
1.86	1.74	1.64	1.68
1.7	1.86	1.76	1.75
1.6	1.86	1.66	1.6
1.58	1.97	1.67	1.74
1.8	1.87	1.5	1.8
1.65	1.73	1.59	1.92
1.71	1.82	1.8	1.8
1.84	1.74	1.6	1.64
1.6	1.85	1.76	1.82

Desvio padrão populacional conhecido

$\sigma = 1.96$

Exemplo 1

Os dados a seguir são uma **amostra aleatória simples** de tamanho (n=35) da população anterior.

Altura (m)	Altura (m)	Altura (m)
17	1.89	1.81
1.7	1.7	1.7
1.83	1.62	1.85
1.5	1.7	1.86
1.56	1.86	1.82
1.6	1.64	
1.82	1.66	
1.6	1.5	
1.83	1.62	
1.8	1.82	
1.83	1.8	
1.89	1.63	
1.86	1.56	
1.7	1.8	
1.82	1.58	

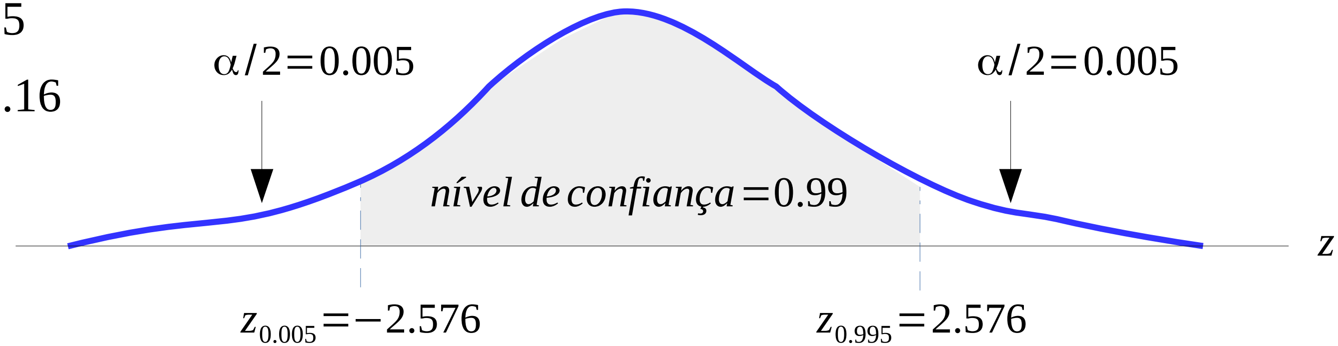
$$média\, amostral \quad \bar{x} = 2.16$$

Exemplo 1

Com base na amostra anterior e sabendo-se que o desvio padrão da população é 1.96. **Construa o intervalo de confiança para o valor da média da população** com um **nível de confiança de 99%**.

$\sigma = 1.96 \qquad n = 35$

$\alpha = 0.01 \qquad \bar{x} = 2.16$



$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$E = 2.576 \cdot \frac{1.96}{\sqrt{35}} = 0.853$$

$$\bar{x} - E \le \mu \le \bar{x} + E$$

$$2.16 - 0.853 \le \mu \le 2.16 + 0.853 \quad \Rightarrow \quad 1.307 \le \mu \le 3.013$$

Logo, temos 95% de confiança que o intervalo (1.307, 3.013) contem a média populacional. De fato, a média populacional é 2.00.

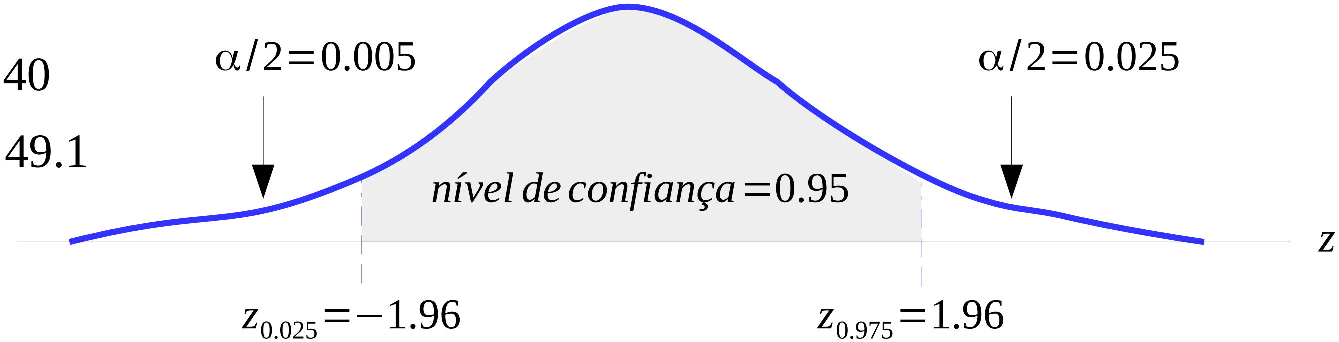
Exemplo 2

Uma **máquina de empacotar cimento** opera **regulada** para **empacotar sacos de 50 kg**, com um **desvio padrão de 0.3 kg**. Após passar por manutenção, foi coletada uma amostra de 40 sacos de cimento embalados pela máquina e obteve-se uma média amostral de 49.1 kg. Há evidências de que a máquina está desregulada?

Vamos estimar um intervalo de confiança para a média da máquina, com um nível de confiança de 95%.

$\sigma = 0.3 \qquad n = 40$

$\alpha = 0.05 \qquad \bar{x} = 49.1$



$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$E = 1.96 \cdot \frac{0.3}{\sqrt{40}} = 0.093$$

$$\bar{x} - E \le \mu \le \bar{x} + E$$

$$49.1 - 0.093 \le \mu \le 49.1 + 0.093 \quad \Rightarrow \quad 49.01 \le \mu \le 49.19$$

Exemplo 2

Uma **máquina de empacotar cimento** opera **regulada** para **empacotar sacos de 50 kg**, com um **desvio padrão de 0.3 kg**. Após passar por manutenção, foi coletada uma amostra de 40 sacos de cimento embalados pela máquina e obteve-se uma média amostral de 49.1 kg. Há evidências de que a máquina está desregulada?

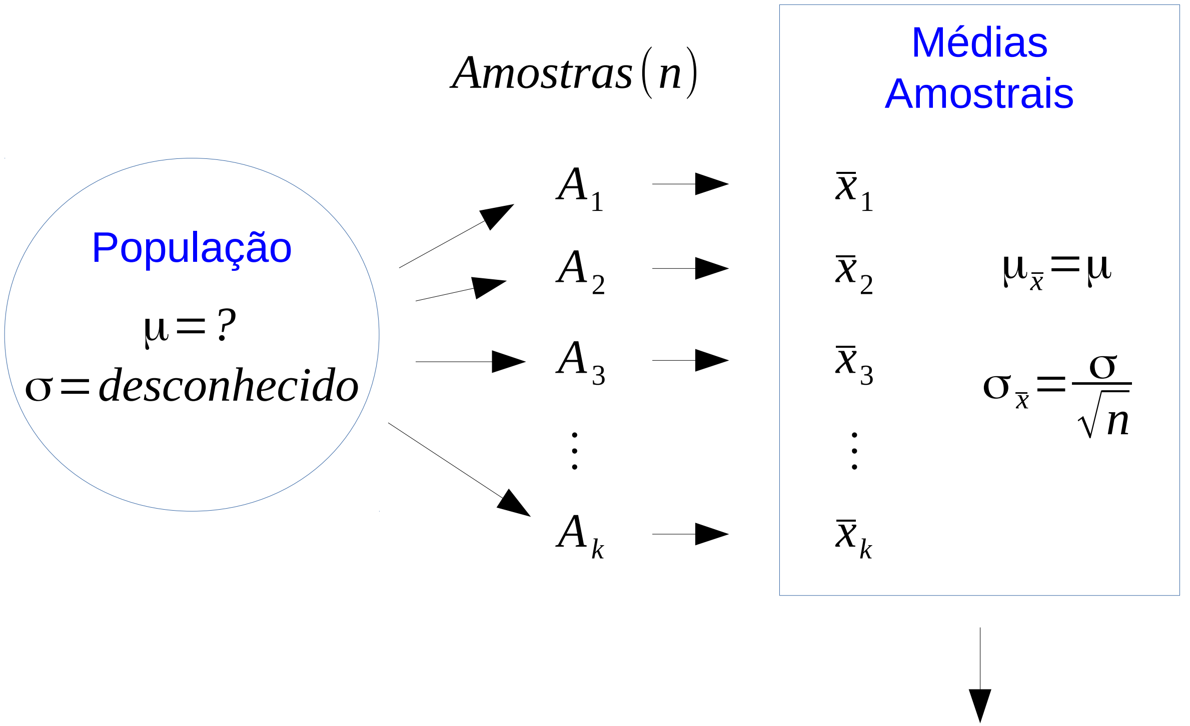
$$49.01 \leq \mu \leq 49.19$$

Logo. A amostra sugere, com 95% de confiança, que a média da máquina é menor que 49.2. Portanto, apresentando fortes evidências de que a máquina está desregulada.

Estimação da Média Populacional: desvio padrão desconhecido

Objetivo

Estimar um intervalo de confiança para a média populacional a partir de uma amostra de tamanho n, com desvio-padrão populacional desconhecido.



Normal se a população original é normal, ou se n>30.

Distribuição t de Student

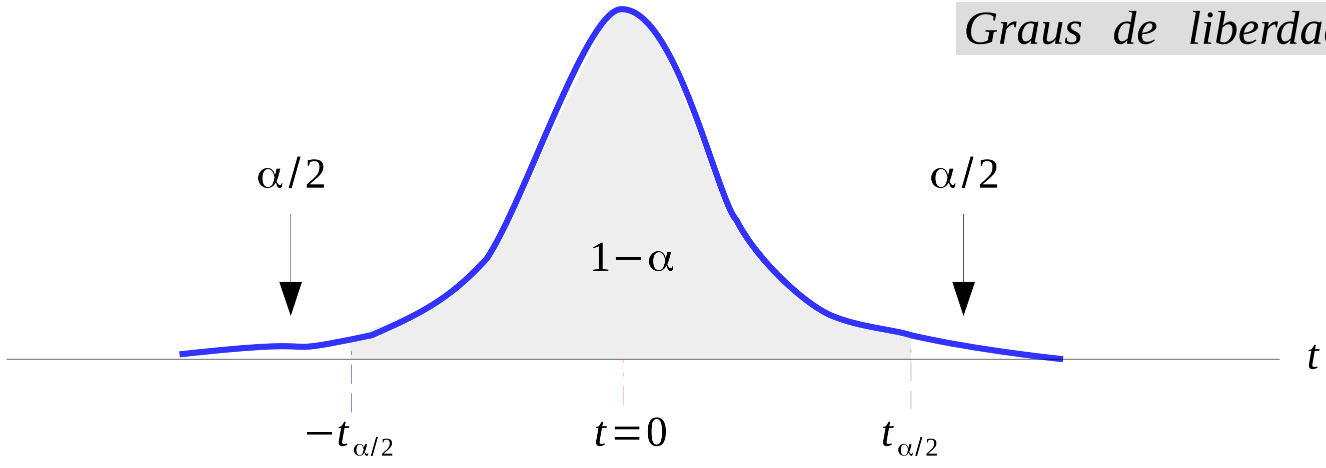
Se uma população tem distribuição normal, então a distribuição:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Ao contrário da distribuição Normal Padrão. Não depende do desvio padrão populacional ser conhecido.

é uma **distribuição t de Student** com média $\mu = 0$ e desvio padrão $\sigma > 1$ para todas as amostras de tamanho n.

$$Graus \;\; de \;\; liberdade \; = \; n-1$$



Graus de liberdade

O número de **graus de liberdade** para uma amostra é o número de valores amostrais que podem variar depois que certas restrições tiverem sido impostas aos dados amostrais.

Exemplo:

Se **10 estudantes tem escores de testes com média 80**, podemos livremente atribuir valores aos 9 primeiros escores, mas o 10º escore estará então determinado. A soma dos 10 escores deve ser 800. De modo que o 10º escore deve ser 800 menos a soma dos 9 primeiros escores. Como esses **9 primeiros escores podem ter valores escolhidos livremente**, então dizemos que há **9 graus de liberdade**.

Objetivo

Estimar um intervalo de confiança para a média populacional a partir de uma amostra de tamanho n, com desvio-padrão populacional desconhecido.

Médias Amostrais

$\bar{x}_1$

$\bar{x}_2$ $\qquad \mu_{\bar{x}} = \mu$

$\bar{x}_3$ $\qquad \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$\vdots$

$\bar{x}_k$

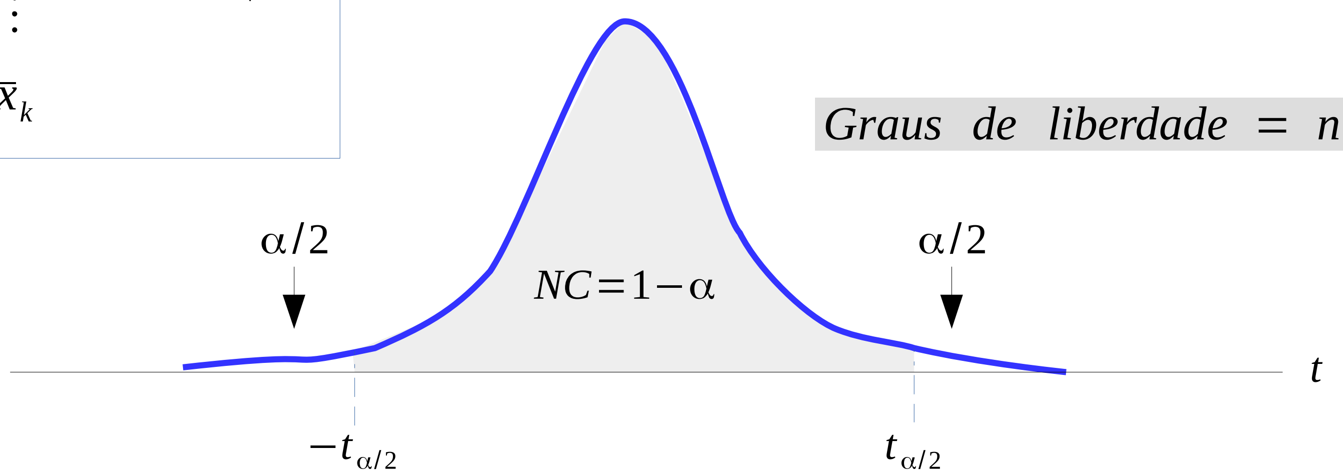
Convertendo para distribuição t, temos: $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$

Para:

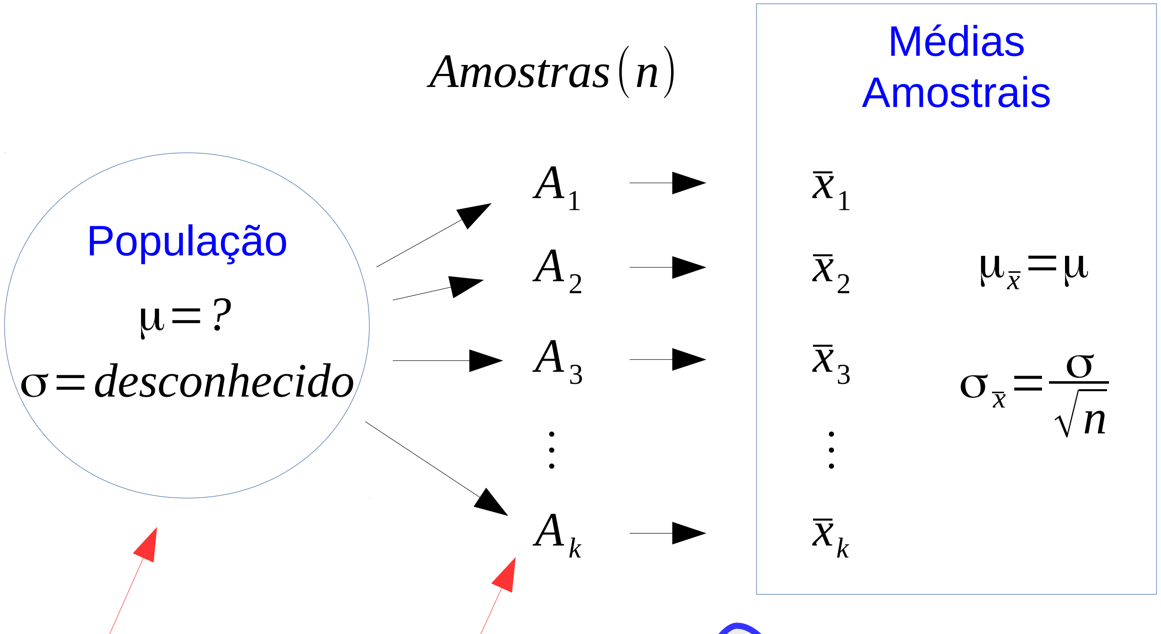
$$-t_{\alpha/2} \leq t \leq t_{\alpha/2} \Rightarrow -t_{\alpha/2} \leq \frac{\bar{x}-\mu}{\dfrac{s}{\sqrt{n}}} \leq t_{\alpha/2} \Rightarrow$$

$$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$Graus \ de \ liberdade = n - 1$



Intervalo de confiança para a média populacional



Requisito: Normal ou n>30.

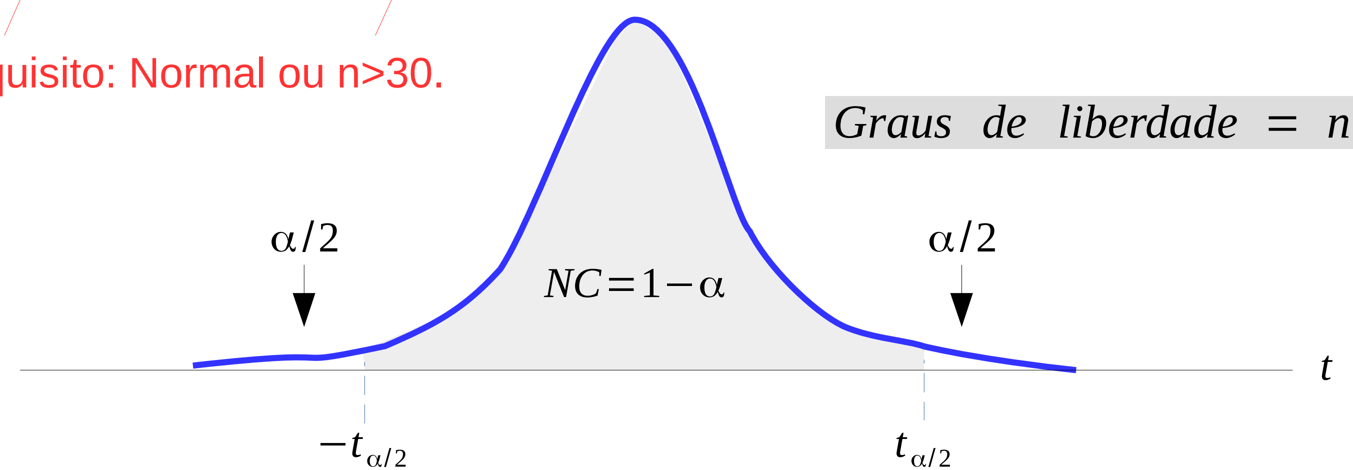
Intervalo de confiança:

$$\bar{x} - E \;\le\; \mu \;\le\; \bar{x} + E$$

Onde:

$$E \;=\; t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$Graus \;\; de \;\; liberdade \;=\; n-1$$



Exemplo 3

Considere os dados de alturas de pessoas na tabela abaixo como uma **população finita**.

Altura (m)	Altura (m)	Altura (m)	Altura (m)
1.84	1.75	1.81	17
1.88	1.63	1.6	1.8
1.76	1.62	1.7	1.65
1.65	1.81	1.83	1.56
1.74	1.9	1.89	1.7
1.9	1.82	1.8	1.89
1.86	1.74	1.64	1.68
1.7	1.86	1.76	1.75
1.6	1.86	1.66	1.6
1.58	1.97	1.67	1.74
1.8	1.87	1.5	1.8
1.65	1.73	1.59	1.92
1.71	1.82	1.8	1.8
1.84	1.74	1.6	1.64
1.6	1.85	1.76	1.82

Desvio padrão populacional desconhecido

$\sigma = ?$

Exemplo 3

Os dados a seguir são uma **amostra aleatória simples** de tamanho (n=35) da população anterior.

Altura (m)	Altura (m)	Altura (m)
17	1.89	1.81
1.7	1.7	1.7
1.83	1.62	1.85
1.5	1.7	1.86
1.56	1.86	1.82
1.6	1.64	
1.82	1.66	
1.6	1.5	
1.83	1.62	
1.8	1.82	
1.83	1.8	
1.89	1.63	
1.86	1.56	
1.7	1.8	
1.82	1.58	

$$média\, amostral \quad \bar{x} = 2.16$$

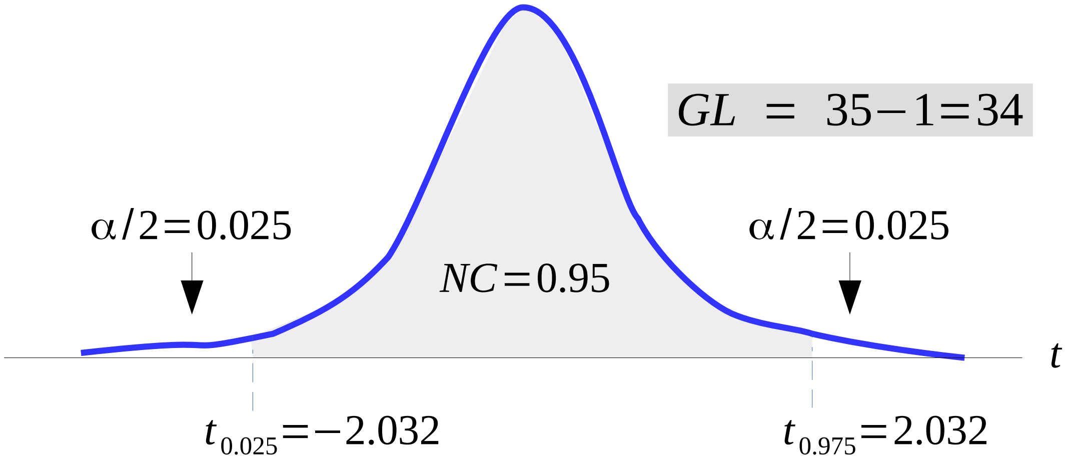
$$desvio\; padrão\; amostral \quad s = 2.58$$

Exemplo 3

Com base na amostra anterior e sem conhecer o desvio padrão da população. **Construa o intervalo de confiança para o valor da média da população** com um **nível de confiança de 95%**.

$\alpha = 0.05 \qquad n = 35$

$\bar{x} = 2.16 \qquad s = 2.58$



$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$E = 2.032 \cdot \frac{2.58}{\sqrt{35}} = 0.886$$

$$\bar{x} - E \le \mu \le \bar{x} + E$$

$$2.16 - 0.886 \le \mu \le 2.16 + 0.886 \quad \Rightarrow \quad 1.27 \le \mu \le 3.05$$

Logo, temos 95% de confiança que o intervalo (1.27, 3.05) contem a média populacional. De fato, a média populacional é 2.00.

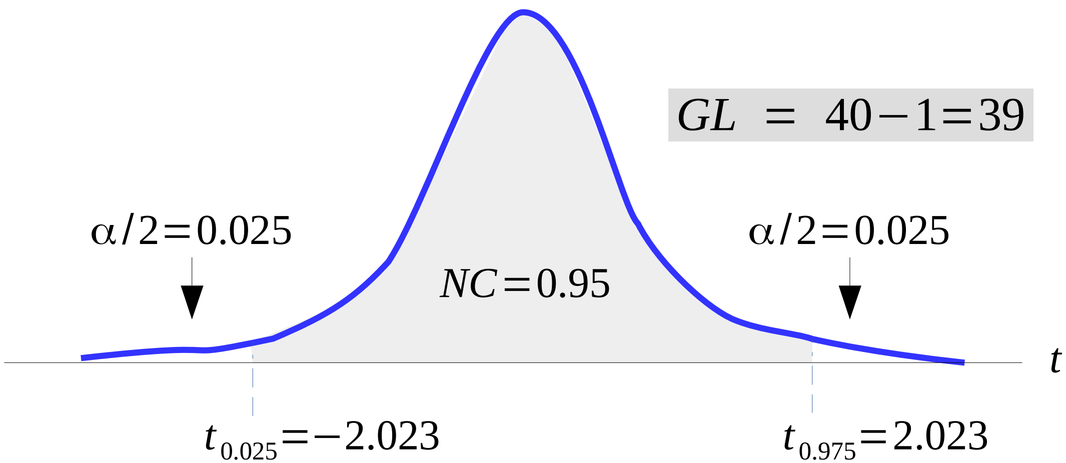
Exemplo 4

Um **fabricante de bebidas** afirma que sua marca de **cerveja Z possui um teor alcoólico de 5.8% de alcool**. Uma **amostra de 40 latas** dessa cerveja apresentou uma **média de 5.2%**, com **desvio padrão de 0.4%**. Há evidências de que a informação do fabricante é correta?

Vamos estimar um intervalo de confiança para a média do verdadeiro teor alcoólico da cerveja Z, com um nível de confiança de 95%.

$\alpha = 0.05 \qquad n = 40$

$\bar{x} = 5.2 \qquad s = 0.4$



$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$E = 2.023 \cdot \frac{0.4}{\sqrt{40}} = 0.128$$

$$\bar{x} - E \leq \mu \leq \bar{x} + E$$

$$5.2 - 0.128 \leq \mu \leq 5.2 + 0.128 \quad \Rightarrow \quad 5.07 \leq \mu \leq 5.33$$

Exemplo 4

Um **fabricante de bebidas** afirma que sua marca de **cerveja Z possui um teor alcoólico de 5.8% de alcool**. Uma **amostra de 40 latas** dessa cerveja apresentou uma **média de 5.2%**, com **desvio padrão de 0.4%**. Há evidências de que a informação do fabricante é correta?

$$5.07 \leq \mu \leq 5.33$$

Logo. A amostra sugere, com 95% de confiança, que o verdadeiro teor alcoólico médio da cerveja Z é menor que 5.4. Portanto, apresentando fortes evidências de que a informação do fabricante está errada.

Determinação do tamanho amostral necessário para estimar a média populacional

Para população com desvio padrão conhecido

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad E \cdot \sqrt{n} = z_{\alpha/2} \cdot \sigma \quad \Rightarrow \quad \sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E} \quad \Rightarrow \quad n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

Para população com desvio padrão desconhecido

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \Rightarrow \quad E \cdot \sqrt{n} = t_{\alpha/2} \cdot s \quad \Rightarrow \quad \sqrt{n} = \frac{t_{\alpha/2} \cdot s}{E} \quad \Rightarrow \quad n = \left(\frac{t_{\alpha/2} \cdot s}{E}\right)^2$$

$$n = \left(\frac{t_{\alpha/2} \cdot s}{E}\right)^2$$

Exemplo 5

Determine o **tamanho da amostra necessária** para se obter um intervalo de confiança de 95% com erro E=0.5 no Exemplo 3.

$$n = \left(\frac{t_{\alpha/2} \cdot s}{E}\right)^2$$

Podemos usar o desvio padrão de uma amostra qualquer.

$s = 2.58$ $\qquad$ $t_{0.025} = 2.032$

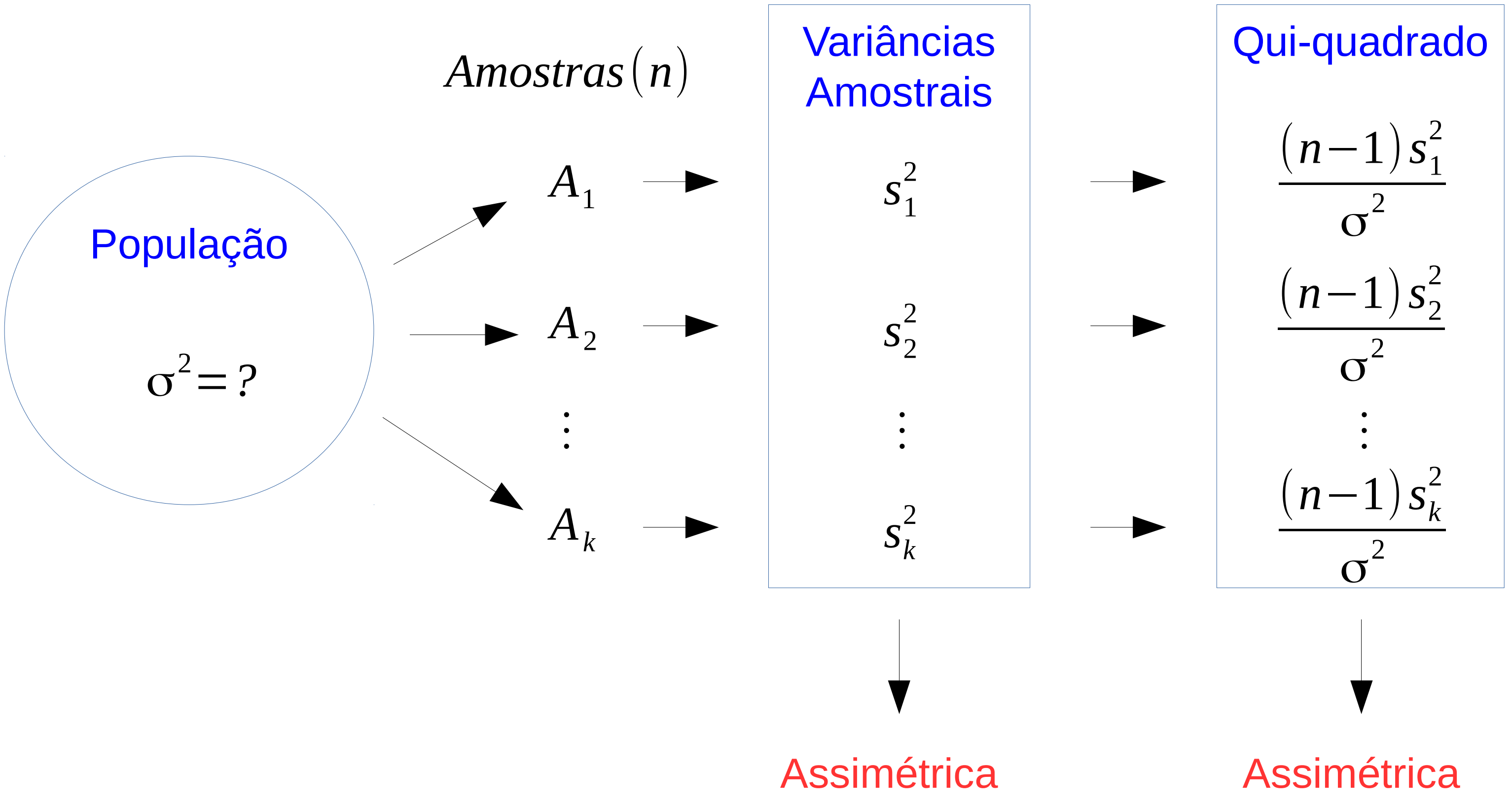
$$n = \left(\frac{2.032 \cdot 2.58}{0.5}\right)^2 \approx 110$$

Logo, precisariamos de uma amostra de tamanho n=110 para se obter um intervalo de confiança com erro E=0.5.

Estimação da Variância Populacional

Objetivo

Estimar um intervalo de confiança para a variância populacional a partir de uma amostra de tamanho n.

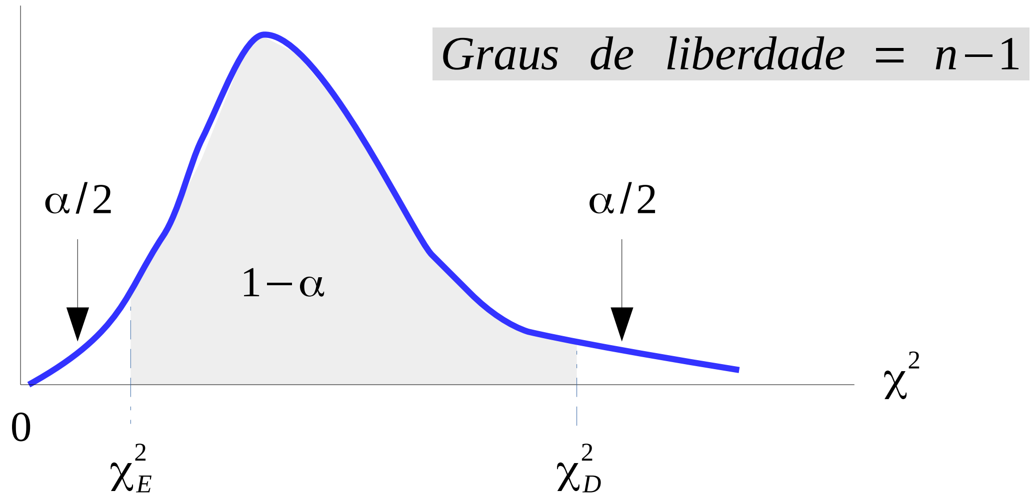


Distribuição Qui-Quadrado

Se uma população tem distribuição normal, então a distribuição:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

é uma **distribuição Qui-Quadrado**.



Objetivo

Estimar um intervalo de confiança para a variância populacional a partir de uma amostra de tamanho n.

Qui-quadrado

$$\chi_1^2 = \frac{(n-1)\,s_1^2}{\sigma^2}$$

$$\chi_2^2 = \frac{(n-1)\,s_2^2}{\sigma^2}$$

$$\vdots$$

$$\chi_k^2 = \frac{(n-1)\,s_k^2}{\sigma^2}$$

Seja a distribuição Qui-Quadrado: $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$

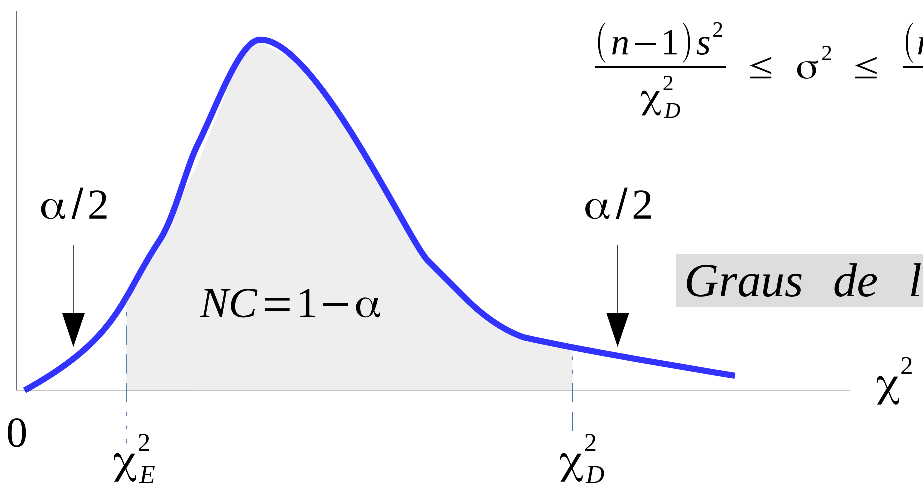
Para:

$$\chi_E^2 \leq \chi^2 \leq \chi_D^2 \quad \Rightarrow \quad \chi_E^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_D^2$$

$$\chi_E^2 \leq \frac{(n-1)s^2}{\sigma^2} \quad \Rightarrow \quad \sigma^2 \chi_E^2 \leq (n-1)s^2 \quad \Rightarrow \quad \sigma^2 \leq \frac{(n-1)s^2}{\chi_E^2}$$

$$\frac{(n-1)s^2}{\sigma^2} \leq \chi_D^2 \quad \Rightarrow \quad (n-1)s^2 \leq \sigma^2 \chi_D^2 \quad \Rightarrow \quad \frac{(n-1)s^2}{\chi_D^2} \leq \sigma^2$$

$$\frac{(n-1)s^2}{\chi_D^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_E^2}$$



Graus de liberdade $= n-1$

Intervalo de confiança para a variância populacional

População

$\sigma^2 = ?$

Requisito: Normal

$Amostras\,(n)$

A_1

A_2

$\vdots$

A_k

Qui-quadrado

$$\chi_1^2 = \frac{(n-1)\,s_1^2}{\sigma^2}$$

$$\chi_2^2 = \frac{(n-1)\,s_2^2}{\sigma^2}$$

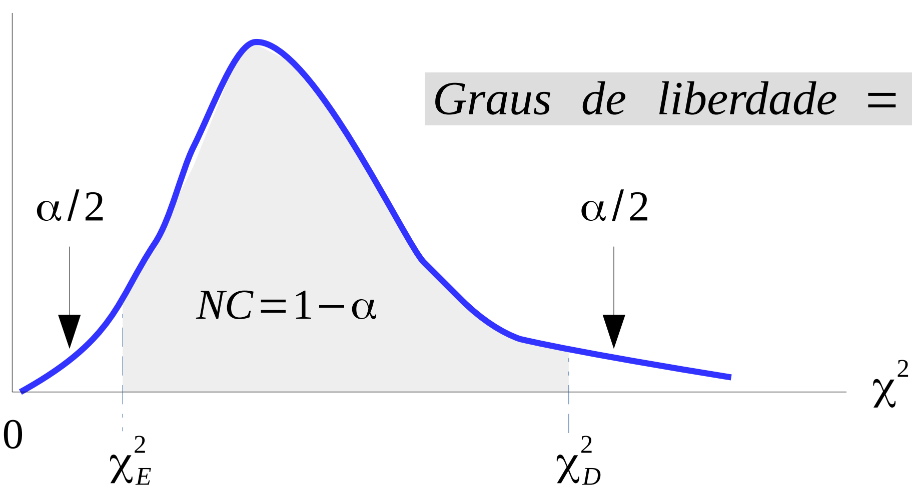
$$\vdots$$

$$\chi_k^2 = \frac{(n-1)\,s_k^2}{\sigma^2}$$

Intervalo de confiança:

$$\frac{(n-1)\,s^2}{\chi_D^2} \le \sigma^2 \le \frac{(n-1)\,s^2}{\chi_E^2}$$

$$Graus\ \ de\ \ liberdade\ = \ n-1$$



Exemplo 5

Considere os dados de alturas de pessoas na tabela abaixo como uma **população finita**.

Altura (m)	Altura (m)	Altura (m)	Altura (m)
1.84	1.75	1.81	17
1.88	1.63	1.6	1.8
1.76	1.62	1.7	1.65
1.65	1.81	1.83	1.56
1.74	1.9	1.89	1.7
1.9	1.82	1.8	1.89
1.86	1.74	1.64	1.68
1.7	1.86	1.76	1.75
1.6	1.86	1.66	1.6
1.58	1.97	1.67	1.74
1.8	1.87	1.5	1.8
1.65	1.73	1.59	1.92
1.71	1.82	1.8	1.8
1.84	1.74	1.6	1.64
1.6	1.85	1.76	1.82

Exemplo 5

Os dados a seguir são uma **amostra aleatória simples** de tamanho (n=35) da população anterior.

Altura (m)	Altura (m)	Altura (m)
17	1.89	1.81
1.7	1.7	1.7
1.83	1.62	1.85
1.5	1.7	1.86
1.56	1.86	1.82
1.6	1.64	
1.82	1.66	
1.6	1.5	
1.83	1.62	
1.8	1.82	
1.83	1.8	
1.89	1.63	
1.86	1.56	
1.7	1.8	
1.82	1.58	

$$desvio\ padrão\ amostral \quad s = 2.58$$

Exemplo 5

Com base na amostra anterior e sem calcular o desvio padrão da população. **Construa o intervalo de confiança para o valor da variância da população** com um **nível de confiança de 99%**.

$\alpha = 0.01 \qquad n = 35$

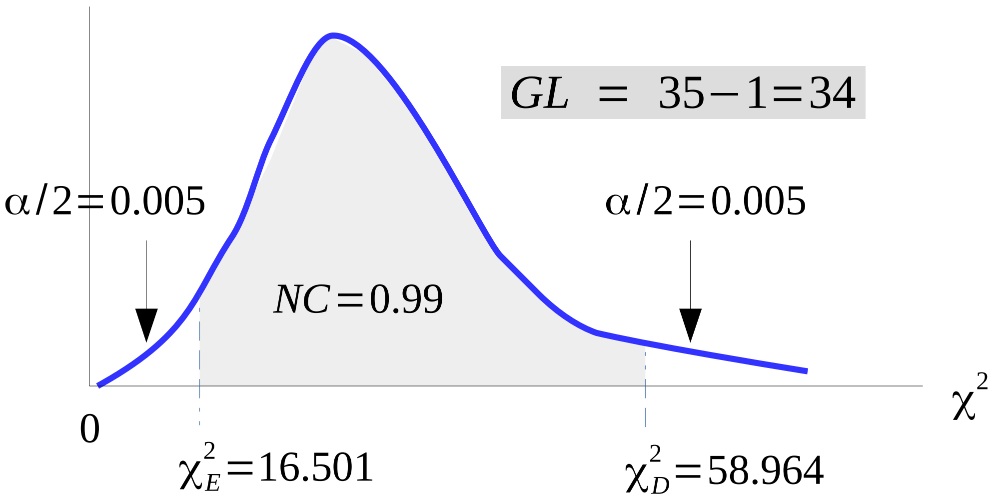
$s = 2.58$

$$\frac{(n-1)s^2}{\chi_D^2} \le \sigma^2 \le \frac{(n-1)s^2}{\chi_E^2}$$

$$\frac{34 \cdot 2.58^2}{58.964} \le \sigma^2 \le \frac{34 \cdot 2.58^2}{16.501}$$

$$3{,}84 \le \sigma^2 \le 13.72$$

$$\sqrt{3.84} \le \sigma \le \sqrt{13.72} \quad \Rightarrow \quad 1.96 \le \sigma \le 3.70$$



Logo, temos 99% de confiança que o intervalo (3.84, 13.72) contem a variância populacional. De fato, a variância populacional é 1.96^2 = 3.84.

Exemplo 6

Um **engenheiro elétrico** está **testando duas marcas de estabilizadores** de **voltagem 127V** para saber **qual é o mais estável**. Para isto, o engenheiro selecionou 10 estabilizadores de cada marca e os submeteu a testes. Cada estabilizador foi submetido a um teste recebendo mais que 127V e mediu-se a voltagem estabilizada, obtendo-se os seguintes resultados:

A={127.5, 127.6, 127.2, 126.9, 127.6, 127, 126.7, 127, 127.1, 126.8} e
B={127.9, 128, 126.5, 126.9, 127.9, 127.5, 126.4, 126.3, 127.7, 126.9}.

Há evidências de que um dos estabilizadores é mais estável que o outro?

Exemplo 6

Para fazer esta análise, vamos comparar a variância populacional da voltagem estabilizada dos dois tipos de estabilizadores. Portanto, vamos construir um intervalo de confiança de 95% para a variância da voltagem estabilizada de ambos tipos de estabilizadores.

$\alpha = 0.05 \qquad n = 10$

$s_A = 0.327 \qquad s_B = 0.673$

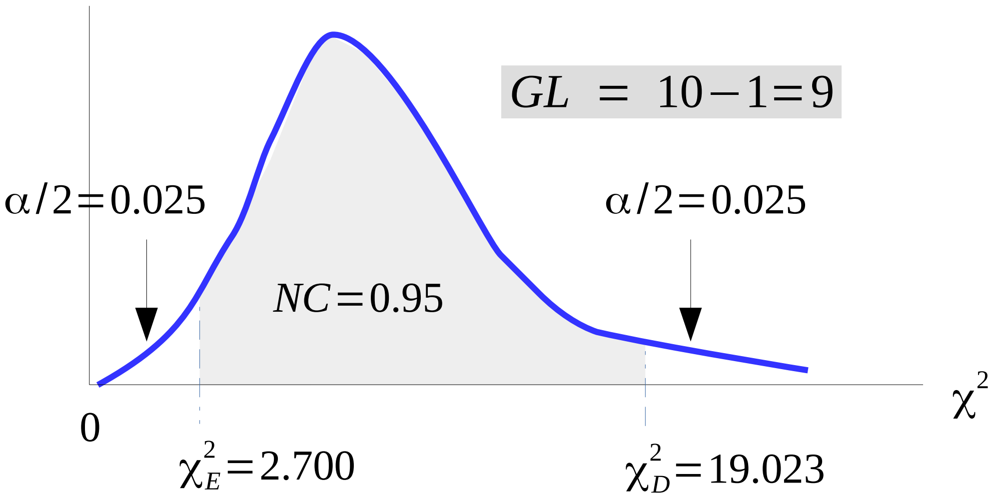
$$\frac{(n-1)s^2}{\chi^2_D} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_E}$$

$$\frac{9 \cdot 0.327^2}{19.023} \leq \sigma^2_A \leq \frac{9 \cdot 0.327^2}{2.7}$$

$$0.05 \leq \sigma^2_A \leq 0.36$$

$$\frac{9 \cdot 0.673^2}{19.023} \leq \sigma^2_B \leq \frac{9 \cdot 0.673^2}{2.7}$$

$$0.21 \leq \sigma^2_B \leq 1.51$$



Logo. Como os intervalos se sobrepõe, não podemos fazer qualquer afirmativa. Ou seja, não temos muita certeza se o estabilizador A é mais estável que o B.