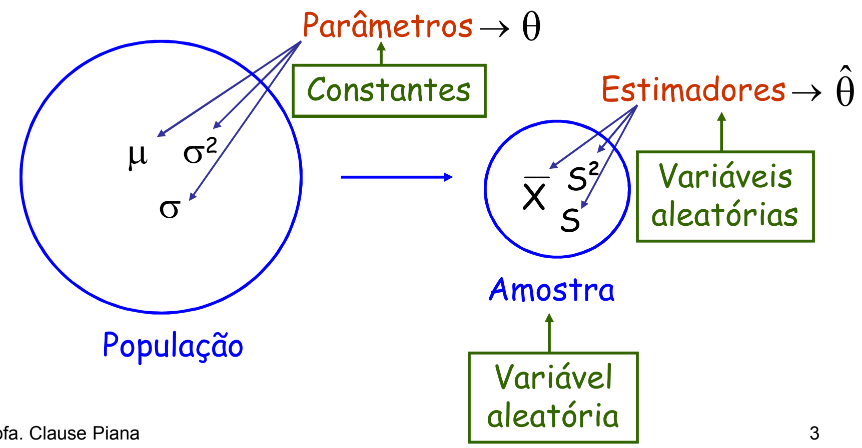
Unidade IV – Inferência estatística

4.1. Introdução e histórico
4.2. Conceitos fundamentais
4.3. Distribuições amostrais e Teorema central do limite
4.4. Estimação de parâmetros
4.5. Testes de hipóteses
4.6. Quebras das pressuposições no processo de inferência
4.7. Testes de qui-quadrado

Teoria da estimação de parâmetros

Estimação de parâmetros

Conceitos fundamentais



Parâmetros:

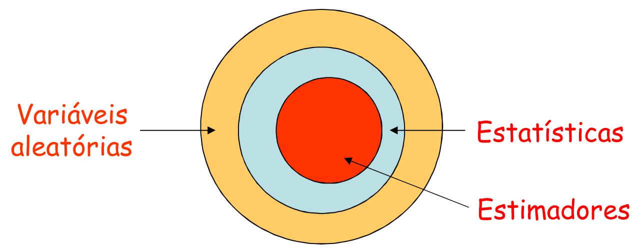
- constantes que caracterizam uma população (distribuição)
- geralmente são valores desconhecidos
- são representados, genericamente, pela letra grega teta (θ)

Exemplos: média da população (μ)
variância da população (σ^2)

Estimadores:

- valores (medidas) calculados na amostra com objetivo de fornecer informação sobre os parâmetros
- são estatísticas, portanto, são variáveis aleatórias
- são representados, genericamente, pela letra teta com um acento circunflexo $(\hat{\theta})$, onde se lê teta chapéu

Exemplos: média da amostra $(\bar{X}$ ou $\hat{\mu})$
variância da amostra $(S^2$ ou $\hat{\sigma}^2)$



⇨ O estimador é uma estatística e todas as estatísticas são variáveis aleatórias

Assim, o estimador é uma variável aleatória que pode assumir diferentes valores

Estimativa é um valor particular que o estimador assume

Exemplo:

Uma população é constituída por quatro valores ($N = 4$):

X = x	1	2	3	4
P(X=x)	0,2	0,3	0,3	0,2

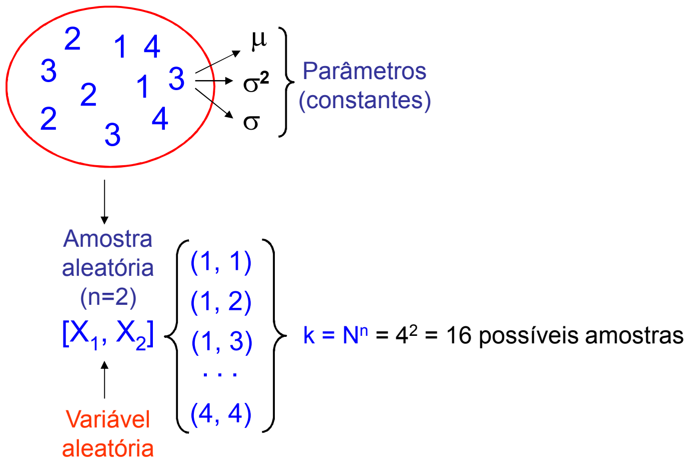
μ=2,5
σ^2=1,05

Desta população, retiramos uma amostra aleatória de tamanho dois ($n=2$) $\rightarrow$ $[X_1, X_2]$

Quantas e quais são as amostras de tamanho dois que podemos extrair desta população de tamanho quatro?

$k = N^n = 4^2 = 16$ possíveis amostras

População (X)



Profa. Clause Piana 7

Parâmetro		$\mu = 2,5$	$\sigma^2 = 1,05$
Estimador		$\bar{X} = \frac{\sum X_i}{n}$	$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$
Estimativas	Amostra 1: (1, 1)	$\bar{x}_1 = \frac{1+1}{2} = 1$	$s_1^2 = \frac{(1-1)^2 + (1-1)^2}{2-1} = 0$
	Amostra 2: (1, 2)	$\bar{x}_2 = \frac{1+2}{2} = 1,5$	$s_2^2 = \frac{(1-1,5)^2 + (2-1,5)^2}{2-1} = 0,5$
	Amostra 3: (1, 3)	$\bar{x}_3 = 2$	$s_3^2 = 2$
	Amostra 4: (1, 4)	$\bar{x}_4 = 2,5$	$s_4^2 = 4,5$
	Amostra 5: (2, 1)	$\bar{x}_5 = 1,5$	$s_5^2 = 0,5$
	Amostra 6: (2, 2)	$\bar{x}_6 = 2$	$s_6^2 = 0$
	Amostra 7: (2, 3)	$\bar{x}_7 = 2,5$	$s_7^2 = 0,5$
	Amostra 8: (2, 4)	$\bar{x}_8 = 3$	$s_8^2 = 2$
	Amostra 9: (3, 1)	$\bar{x}_9 = 2$	$s_9^2 = 2$
	Amostra 10: (3, 2)	$\bar{x}_{10} = 2,5$	$s_{10}^2 = 0,5$
	Amostra 11: (3, 3)	$\bar{x}_{11} = 3$	$s_{11}^2 = 0$
	Amostra 12: (3, 4)	$\bar{x}_{12} = 3,5$	$s_{12}^2 = 0,5$
	Amostra 13: (4, 1)	$\bar{x}_{13} = 2,5$	$s_{13}^2 = 4,5$
	Amostra 14: (4, 2)	$\bar{x}_{14} = 3$	$s_{14}^2 = 2$
	Amostra 15: (4, 3)	$\bar{x}_{15} = 3,5$	$s_{15}^2 = 0,5$
	Amostra 16: (4, 4)	$\bar{x}_{16} = 4$	$s_{16}^2 = 0$

Parâmetro: constante populacional desconhecida

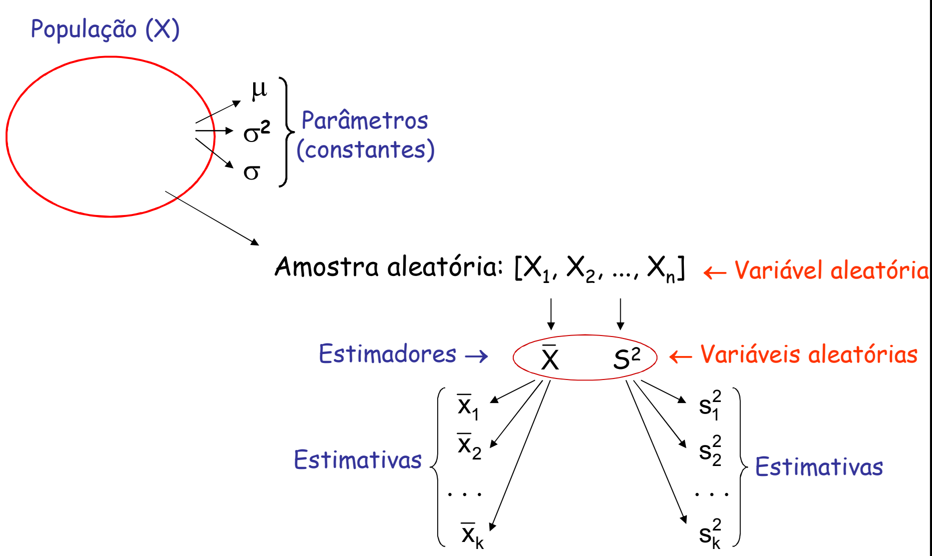
μ (média populacional)

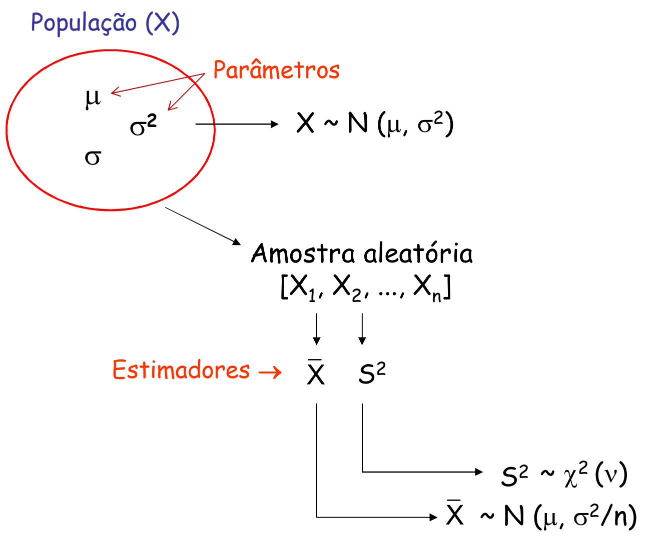
Estimador: expressão matemática da medida amostral

$$\bar{X} = \frac{\sum X_i}{n} \quad \text{(média amostral)}$$

Estimativa: valor numérico que o estimador assume na amostra

$$\bar{x}_1 = \frac{1+1}{2} = 1 \quad \text{(média amostral calculada)}$$





Todos os estimadores são variáveis aleatórias e têm uma distribuição de probabilidade.

Podem existir vários estimadores para um mesmo parâmetro.

Exemplos:

$$\left.\begin{aligned} \bar{X} &= \frac{\sum X_i}{n} \\ \bar{X}_p &= \frac{\sum X_i p_i}{\sum p_i} \\ X_1 & \end{aligned}\right\} \text{estimadores de } \mu$$

$$\left.\begin{aligned} S^2 &= \frac{\sum (X_i - \bar{X})^2}{n-1} \\ S_n^2 &= \frac{\sum (X_i - \bar{X})^2}{n} \end{aligned}\right\} \text{estimadores de } \sigma^2$$

O melhor estimador de um parâmetro será o que apresenta melhores propriedades estatísticas.

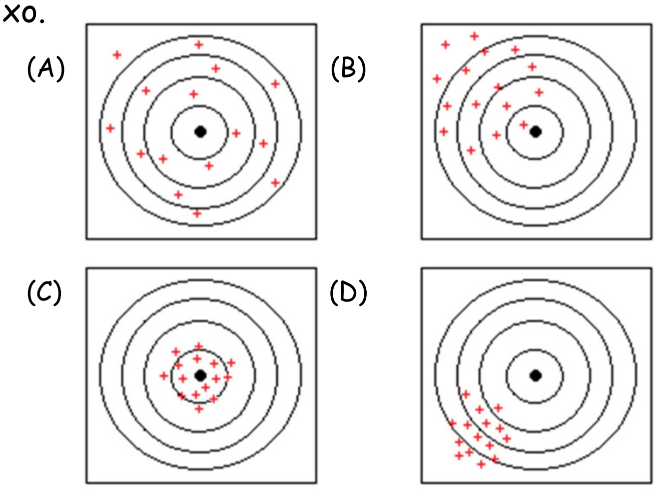
Algumas ideias básicas para compreender as propriedades estatísticas dos estimadores

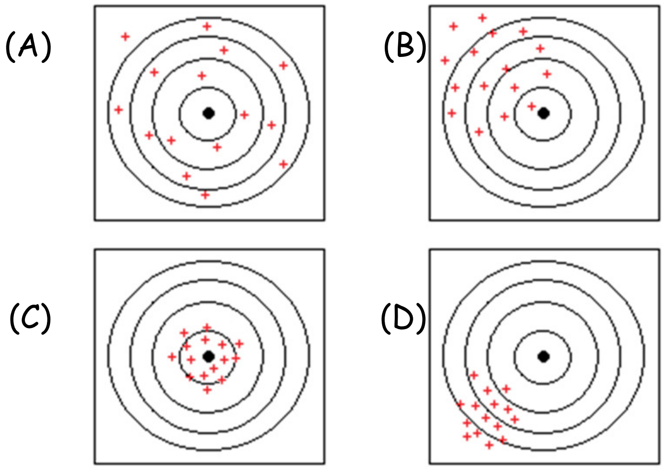
→ Acurácia

→ Precisão

→ Viés

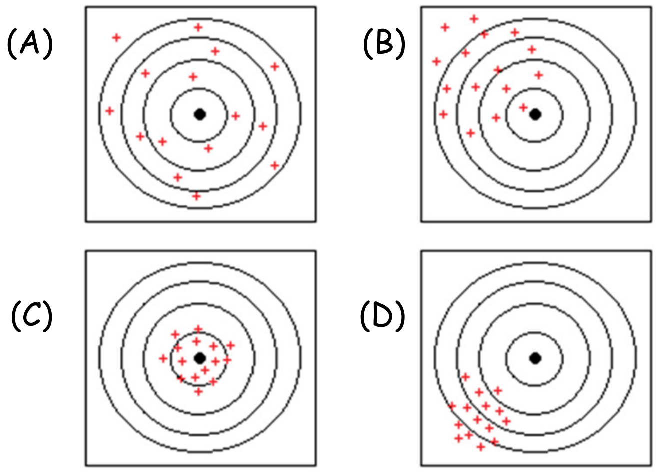
Exemplo: Desejamos comprar um rifle e, após algumas seleções, restaram quatro alternativas, que chamaremos de rifles A, B, C e D. Foi feito um teste com cada rifle que consistiu em fixá-lo num cavalete, mirar o centro de um alvo e disparar 15 tiros. Os resultados estão ilustrados na figura abaixo.





1º critério: em média acertar o alvo:

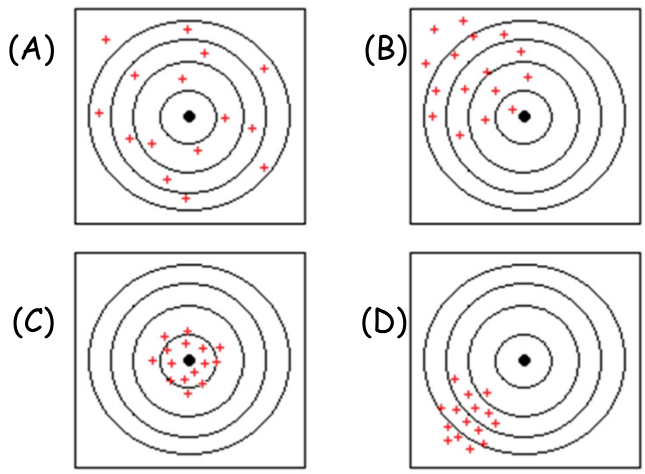
2º critério: não ser muito dispersivo:



Acurácia: mede a proximidade de cada observação do valor que se deseja atingir

Precisão: mede a proximidade de cada observação da média de todas as observações

Viés: diferença entre a média das observações e o valor que se deseja atingir



Arma A: não viesada, pouco acurada e tem baixa precisão
Arma B: viesada, pouco acurada e tem baixa precisão
Arma C: não viesada, muito acurada e tem alta precisão
Arma D: viesada, pouco acurada e tem alta precisão

Propriedades dos estimadores

1. **Imparcialidade ou não tendenciosidade**
2. **Eficiência ou variância mínima**
3. **Consistência**

1. Imparcialidade ou não tendenciosidade

Um estimador $\hat{\theta}$ é um estimador imparcial do parâmetro θ se o valor esperado de $\hat{\theta}$ for igual a θ.

$$E(\hat{\theta}) = \theta$$

Exemplos:

$\overline{X}$ é um estimador imparcial de μ, pois $E(\overline{X}) = \mu$

$\overline{X}_p$ é um estimador imparcial de μ, pois $E(\overline{X}_p) = \mu$

X_1 é um estimador imparcial de μ, pois $E(X_1) = \mu$

S^2 é um estimador imparcial de σ^2, pois $E(S^2) = \sigma^2$

S_n^2 não é um estimador imparcial σ^2, pois $E(S_n^2) = \frac{n-1}{n}\sigma^2$

Verificação da propriedade de imparcialidade da variância

Distribuição de probabilidade da população

$X = x$	40	45	50	Σ
$P(X = x)$	0,2	0,3	0,5	1

$E(X) = \mu = 46{,}5$

$V(X) = \sigma^2 = 15{,}25$

$n = 2 \rightarrow [X_1, X_2]$

Amostra	$[X_1, X_2]$	$P[X_1, X_2]$	$\overline{X}$	X_+	S^2	S_n^2		
1	(40, 40)	0,2 × 0,2 = 0,04	40	80	0	0		
2	(40, 45)	0,2 × 0,3 = 0,06	42,5	85	12,5	6,25		
3	(40, 50)	0,2 × 0,5 = 0,10	45	90	50	25		
4	(45, 40)	0,3 × 0,2 = 0,06	42,5	85	12,5	6,25		
5	(45, 45)	0,3 × 0,3 = 0,09	45	90	0	0		
6	(45, 50)	0,3 × 0,5 = 0,15	47,5	95	12,5	6,25		
7	(50, 40)	0,5 × 0,2 = 0,10	45	90	50	25		
8	(50, 45)	0,5 × 0,3 = 0,15	47,5	95	12,5	6,25		
9	(50, 50)	0,5 × 0,5 = 0,25	50	100	0	0		

Distribuição amostral de S^2 das amostras de tamanho 2

$S^2 = s^2$	0	12,5	50	Σ
$P(S^2 = s^2)$	0,38	0,42	0,2	1

$$E(S^2) = \mu_{S^2} = \sum s^2 p(s^2) = 0 \times 0,38 + 12,5 \times 0,42 + 50 \times 0,2 = 15,25$$

$$E(S^2) = \sigma^2 = 15,25$$

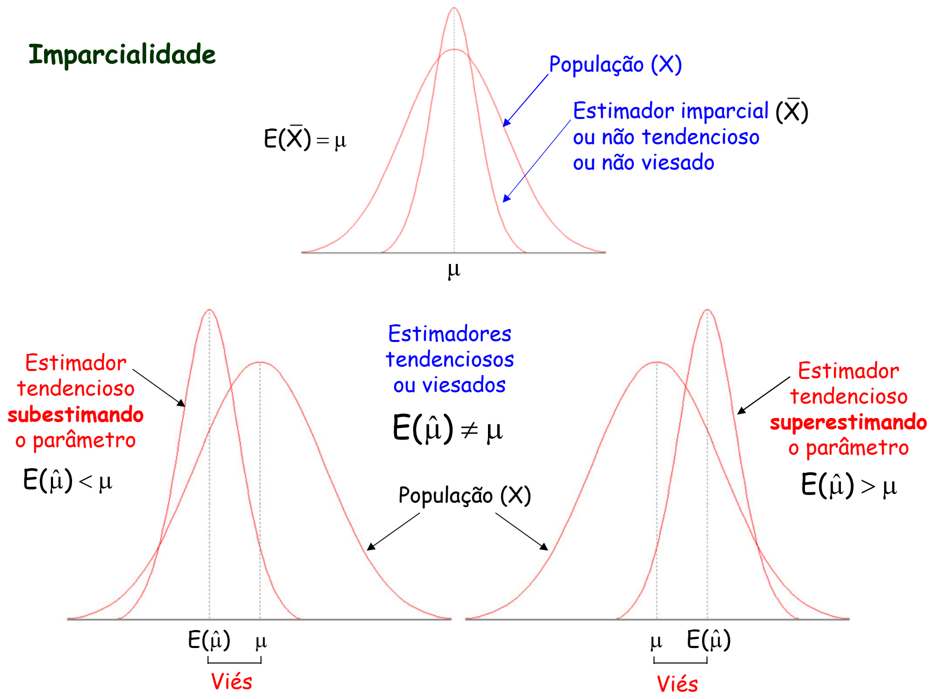
Distribuição amostral de S_n^2 das amostras de tamanho 2

$S_n^2 = s_n^2$	0	6,25	25	Σ
$P(S_n^2 = s_n^2)$	0,38	0,42	0,2	1

$$E(S_n^2) = \mu_{S_n^2} = \sum s_n^2 p(s_n^2) = 0 \times 0,38 + 6,25 \times 0,42 + 25 \times 0,2 = 7,625$$

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

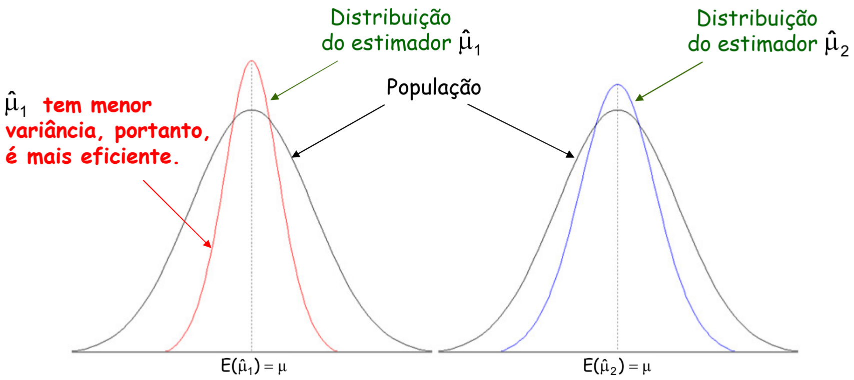
Imparcialidade



2. Eficiência ou variância mínima

Se dois ou mais estimadores de um mesmo parâmetro são imparciais, é mais eficiente aquele que possui a menor variância.

Dois estimadores imparciais de μ



Exemplo:

Dentre todos os estimadores imparciais de μ ($\overline{X}$, $\overline{X}_p$ e X_1), a média simples ($\overline{X}$) é o mais eficiente porque tem a menor variância.

Demonstração:

Consideremos uma amostra aleatória de tamanho três ($n = 3$) $\rightarrow$ $[X_1, X_2, X_3]$

Média simples: $\overline{X} = \dfrac{\sum X_i}{n} \longrightarrow \overline{X} = \dfrac{X_1 + X_2 + X_3}{3} = \dfrac{1}{3}(X_1 + X_2 + X_3)$

Média ponderada: $\overline{X}_p = \dfrac{\sum X_i p_i}{\sum p_i} \longrightarrow \overline{X}_p = \dfrac{1X_1 + 2X_2 + 1X_3}{4} = \dfrac{1}{4}(X_1 + 2X_2 + X_3)$

Comparação das variâncias

Média simples

$$\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$$

$$V(\bar{X}) = V\left[\frac{1}{3}(X_1 + X_2 + X_3)\right]$$

$$V(\bar{X}) = \frac{1}{9}[V(X_1 + X_2 + X_3)]$$

$$V(\bar{X}) = \frac{1}{9}[V(X_1) + V(X_2) + V(X_3)]$$

$$V(\bar{X}) = \frac{1}{9}(\sigma^2 + \sigma^2 + \sigma^2)$$

$$V(\bar{X}) = \frac{3\sigma^2}{9} = 0{,}33\sigma^2$$

Média ponderada

$$\bar{X}_p = \frac{1}{4}(X_1 + 2X_2 + X_3)$$

$$V(\bar{X}_p) = V\left[\frac{1}{4}(X_1 + 2X_2 + X_3)\right]$$

$$V(\bar{X}_p) = \frac{1}{16}[V(X_1) + V(2X_2) + V(X_3)]$$

$$V(\bar{X}_p) = \frac{1}{16}[V(X_1) + 4V(X_2) + V(X_3)]$$

$$V(\bar{X}_p) = \frac{1}{16}(\sigma^2 + 4\sigma^2 + \sigma^2)$$

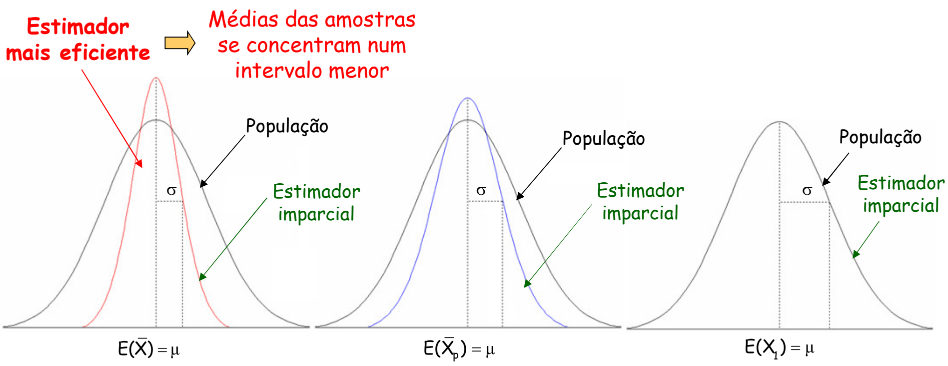
$$V(\bar{X}_p) = \frac{6\sigma^2}{16} = 0{,}38\sigma^2$$

Mais eficiente

$$V(\bar{X}) = 0{,}33\sigma^2 < V(\bar{X}_p) = 0{,}38\sigma^2 < V(X_1) = \sigma^2$$

Eficiência ou variância mínima

Estimadores imparciais de μ



$$V(\bar{X}) = 0{,}33\sigma^2 < V(\bar{X}_p) = 0{,}38\sigma^2 < V(X_1) = \sigma^2$$

3. Consistência

Um estimador é consistente se à medida que o tamanho da amostra aumenta o valor do estimador se aproxima do parâmetro.

$$\hat{\theta} \xrightarrow{n \to N} \theta$$

Exemplo:

S_n^2 é um estimador consistente de σ^2

Com base nessa propriedade, podemos concluir que:

→ Se a amostra for pequena, devemos utilizar S^2 para estimar σ^2

→ Se a amostra for grande, podemos utilizar S_n^2 ou S^2 para estimar σ^2

Melhores estimadores

Melhor estimador → $\boxed{\bar{X} = \dfrac{\sum X_i}{n}}$ (imparcial e mais eficiente)

$$\left.\begin{array}{l} \bar{X} = \dfrac{\sum X_i}{n} \\ \bar{X}_p = \dfrac{\sum X_i p_i}{\sum p_i} \\ X_1 \end{array}\right\} \text{estimadores de } \mu$$

Melhor estimador → $\boxed{S^2 = \dfrac{\sum (X_i - \bar{X})^2}{n-1}}$ (imparcial)

$$\left.\begin{array}{l} S^2 = \dfrac{\sum (X_i - \bar{X})^2}{n-1} \\ S_n^2 = \dfrac{\sum (X_i - \bar{X})^2}{n} \end{array}\right\} \text{estimadores de } \sigma^2$$

Profa. Clause Piana 28

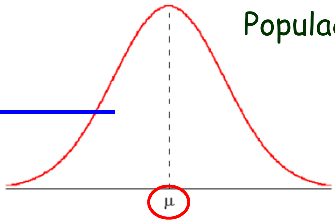
Significado prático das propriedades dos estimadores

Objetivo da pesquisa: conhecer o diâmetro médio das flores de uma nova espécie de hibisco

Variável em estudo: diâmetro da flor (cm)

A variável X tem distribuição normal com média μ e variância σ^2

População de diâmetros ← infinita



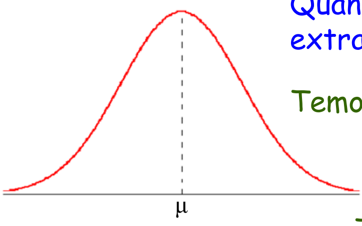
todos os possíveis diâmetros de todas as plantas desta espécie

Amostra n=50 → estimar μ

conhecer

Para estimar $\mu \rightarrow$ média aritmética simples da amostra: $\bar{X} = \frac{\sum X_i}{n}$

$X \sim N(\mu, \sigma^2)$

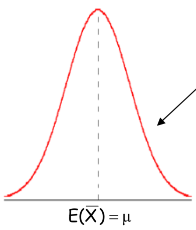


Quantas amostras de tamanho 50 é possível extrair desta população infinita?

Temos uma população infinita de amostras n=50

Cada amostra terá uma média diferente

Temos uma população infinita de médias de amostras n=50



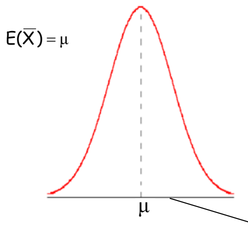
$\bar{X} \sim N(\mu, \sigma^2/50)$

Qual será a distribuição da média das amostras de tamanho 50?

A distribuição da média também será normal com:

$E(\bar{X}) = \mu \qquad V(\bar{X}) = \frac{\sigma^2}{50}$

$\overline{X} \sim N(\mu, \sigma^2/50)$



Ao escolhermos a amostra para estimar μ, passamos a ter apenas uma de todas as possíveis amostras de tamanho 50 e, portanto, apenas uma das possíveis médias.

$\bar{x} \rightarrow$ **estimativa de μ**

Objetivo: garantir que esta estimativa esteja o mais próximo possível de μ

Como a distribuição das médias é normal, é muito mais provável que escolhamos uma amostra que tenha uma média próxima de μ do que distante dela.

A imparcialidade garante que média das médias seja igual a média da população ou que as médias das amostras "flutuem" em volta da média da população.

A propriedade de eficiência garante que as médias das amostras se concentram num intervalo menor.

Processos de estimação

Existem duas maneiras de obter a estimativa de um parâmetro:

- **Estimação por ponto**
- **Estimação por intervalo**

Profa. Clause Piana 32

Processos de estimação

◻ Estimação por ponto (ou pontual)

É o processo através do qual obtemos um único ponto, ou seja, **um único valor** para estimar o parâmetro.

Exemplo: Amostra (1, 3, 2)

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1+3+2}{3} = 2$$ ← estimativa pontual de μ

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(1-2)^2 + (3-2)^2 + (2-2)^2}{3-1} = 1$$

estimativa pontual de σ^2

Parâmetro		$\mu = 2{,}5$	$\sigma^2 = 1{,}05$
Estimador		$\bar{X} = \frac{\sum X_i}{n}$	$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$
Estimativas pontuais	Amostra 1: (1, 1)	$\bar{x}_1 = \frac{1+1}{2} = 1$	$s_1^2 = \frac{(1-1)^2 + (1-1)^2}{2-1} = 0$
	Amostra 2: (1, 2)	$\bar{x}_2 = \frac{1+2}{2} = 1{,}5$	$s_2^2 = \frac{(1-1{,}5)^2 + (2-1{,}5)^2}{2-1} = 0{,}5$
	Amostra 3: (1, 3)	$\bar{x}_3 = 2$	$s_3^2 = 2$
	Amostra 4: (1, 4)	$\bar{x}_4 = 2{,}5$	$s_4^2 = 4{,}5$
	Amostra 5: (2, 1)	$\bar{x}_5 = 1{,}5$	$s_5^2 = 0{,}5$
	Amostra 6: (2, 2)	$\bar{x}_6 = 2$	$s_6^2 = 0$
	Amostra 7: (2, 3)	$\bar{x}_7 = 2{,}5$	$s_7^2 = 0{,}5$
	Amostra 8: (2, 4)	$\bar{x}_8 = 3$	$s_8^2 = 2$
	Amostra 9: (3, 1)	$\bar{x}_9 = 2$	$s_9^2 = 2$
	Amostra 10: (3, 2)	$\bar{x}_{10} = 2{,}5$	$s_{10}^2 = 0{,}5$
	Amostra 11: (3, 3)	$\bar{x}_{11} = 3$	$s_{11}^2 = 0$
	Amostra 12: (3, 4)	$\bar{x}_{12} = 3{,}5$	$s_{12}^2 = 0{,}5$
	Amostra 13: (4, 1)	$\bar{x}_{13} = 2{,}5$	$s_{13}^2 = 4{,}5$
	Amostra 14: (4, 2)	$\bar{x}_{14} = 3$	$s_{14}^2 = 2$
	Amostra 15: (4, 3)	$\bar{x}_{15} = 3{,}5$	$s_{15}^2 = 0{,}5$
	Amostra 16: (4, 4)	$\bar{x}_{16} = 4$	$s_{16}^2 = 0$

O problema da estimativa pontual é que não sabemos o quão próximo do parâmetro o seu valor está.

X = x	1	2	3	4
P(X=x)	0,2	0,3	0,3	0,2

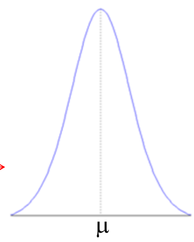
μ=2,5
σ^2=1,05

Distribuição amostral da média das amostras de tamanho 2

$\bar{X}=\bar{x}$	1	1,5	2	2,5	3	3,5	4	Σ
$P(\bar{X}=\bar{x})$	0,04	0,12	0,21	0,26	0,21	0,12	0,04	1

Se **n** é grande,

$\bar{X} \sim N(\mu, \sigma^2/n)$ →



Conhecendo a distribuição amostral do estimador podemos obter uma estimativa de melhor qualidade: o intervalo de confiança.

Estimação por intervalo

É um processo que permite obter os limites de um intervalo onde, com uma determinada probabilidade (nível de confiança), podemos esperar encontrar o verdadeiro valor do parâmetro.

dois valores

Limite inferior → $\underbrace{Li < \theta < Ls}_{\text{Intervalo de confiança}}$ ← Limite superior

As estimativas por intervalo são preferíveis porque indicam a precisão, estabelecendo limites que, com uma determinada probabilidade, devem compreender o parâmetro.

Intervalos de confiança para a média

Como construir intervalos de confiança?

1. Intervalo de confiança para a **média** de uma população (μ)

Duas situações:
- Grandes amostras ($n>30$) ou variância populacional conhecida
- Pequenas amostras ($n \leq 30$) ou variância populacional desconhecida

2. Intervalo de confiança para a **diferença entre médias** de duas populações ($\mu_1 - \mu_2$)

1.Intervalo de confiança para μ

Situação 1. Quando o tamanho da amostra é grande (**n>30**)

Parâmetro $\rightarrow$ μ (média da população X)

Qual é o melhor estimador de μ?

Estimador $\rightarrow$ $\bar{X}$ (média aritmética simples da amostra)

De acordo com o **TCL**:

Se **n** é grande,

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Padronizar a variável $\bar{X}$ $\rightarrow$ transformar $\bar{X}$ em Z

$X \sim N(\mu, \sigma^2)$

$\downarrow$

$Z \sim N(0,1)$

$$Z = \frac{X-\mu}{\sigma}$$ $\leftarrow$ Transformar uma variável X em Z

Transformar a variável $\bar{X}$ em Z

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\downarrow$

$Z \sim N(0,1)$

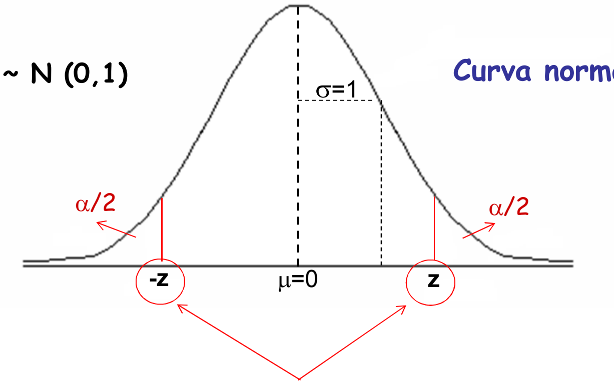
$$Z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}}$$, onde: $\mu_{\bar{X}} = E(\bar{X}) = \mu$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

então: $Z = \dfrac{\bar{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}} \sim N(0,1)$

Construção do intervalo de confiança para μ a partir de Z

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N\ (0,1)$$



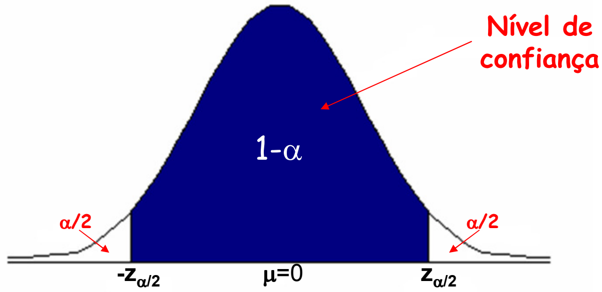
Curva normal padrão

Começamos a construir um intervalo de confiança para μ, a partir da curva normal padrão, estabelecendo dois limites simétricos para os valores da variável Z.

Esses valores delimitam uma área $\alpha/2$.

Construção do intervalo de confiança para μ a partir de Z

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$



$-z_{\alpha/2}$: valor de Z que delimita a área $\alpha/2$ a esquerda

$z_{\alpha/2}$: valor de Z que delimita a área $\alpha/2$ a direita

α: probabilidade de Z **não assumir** um valor entre $-z_{\alpha/2}$ e $z_{\alpha/2}$

$1-\alpha$: probabilidade de Z **assumir** um valor entre $-z_{\alpha/2}$ e $z_{\alpha/2}$

$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$ ⟵ **nível de confiança**

P(-$z_{\alpha/2}$<Z<$z_{\alpha/2}$)=1-α , onde: $Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$$P(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}) = 1-\alpha$$

$$P(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(-\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P\left[(-\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}})\times(-1)\right] = 1-\alpha$$

$$P(\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > \mu > \bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

Intervalo de confiança para μ

$$P(\underbrace{\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Limite inferior}} < \mu < \underbrace{\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{Limite superior}}) = 1-\alpha$$

Nível de confiança

IC (μ; 1–α): $\bar{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

A probabilidade de que os limites contenham (ou cubram) o verdadeiro valor de μ é 1-α.
Os limites são aleatórios.

onde:

: é o estimador de μ

$z_{\alpha/2}$: é o valor da variável Z que delimita a área $\alpha/2$

n : é o tamanho da amostra;

σ : é desvio padrão da população (parâmetro)

- Como as populações com as quais trabalhamos em geral são muito grandes, não conhecemos o parâmetro σ.
- Por isso usamos uma estimativa desse parâmetro que é o **s** (desvio padrão da amostra).

Em muitas situações, quando a amostra é grande, a estimativa é considerada suficientemente próxima do parâmetro, com base na propriedade de **consistência** dos estimadores.

Duas pressuposições para a utilização desta metodologia:

1. A variável em estudo tem distribuição normal, $X \sim N(\mu,\sigma^2)$
2. O tamanho da amostra é suficientemente grande para obtenção de uma estimativa aproximada da variação populacional (σ) $\rightarrow$ $n > 30$

Exercício proposto:

Um engenheiro de desenvolvimento de um fabricante de pneus está investigando a vida do pneu em relação a um novo componente de borracha. Ele fabricou 40 pneus e testou-os até o fim da vida em um teste na estrada. A média e o desvio padrão da amostra são 61.492 e 6.085 km, respectivamente. O engenheiro acredita que a vida média desse novo pneu supera 60.000 km.

Obtenha o intervalo de confiança, ao nível de 95%, para a vida média do pneu e conclua a respeito da suposição do engenheiro.

Resolução:

Variável em estudo: X = vida (durabilidade) do pneu (km rodados)

Pressuposições:

1. A variável em estudo tem distribuição normal.
2. A amostra é grande ($n \geq 30$).

$$IC(\mu; 1-\alpha): \overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Estimativas:

$\bar{x}$ = 61.492 km
s = 6.085 km
n = 40 pneus

$\alpha/2 = 0{,}025$
Área entre 0 e $z_{\alpha/2} = 0{,}475$
$z_{\alpha/2} = z_{,975} = 1{,}96$
$\uparrow$ quantil

Construção do intervalo:

$$IC(\mu; 1-\alpha): \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$IC\ (\mu;\ 0{,}95):\ 61.492 \pm 1{,}96 \times \frac{6.085}{\sqrt{40}}$$

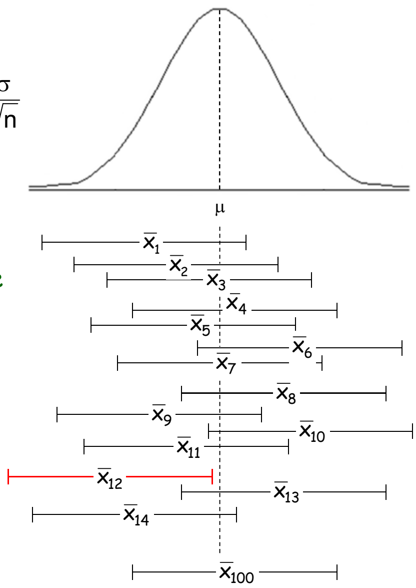
$$IC\ (\mu;\ 0{,}95):\ 61.492 \pm 1.886$$

Limite inferior: 61.492 – 1.886 = 59.606
Limite superior: 61.492 + 1.886 = 63.378

Conclusão: O intervalo de confiança, ao nível de 95%, para a verdadeira durabilidade média do novo pneu é de 59.606 a 63.378 km. Como o valor 60.000 km está coberto pelo intervalo, a durabilidade média do novo pneu não supera este valor.

Significado de um IC para μ, com nível de confiança **$1-\alpha=0{,}99$** e σ^2 **conhecida**

$$IC(\mu;\ 1-\alpha): \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

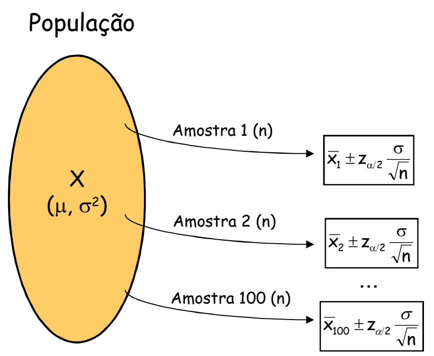


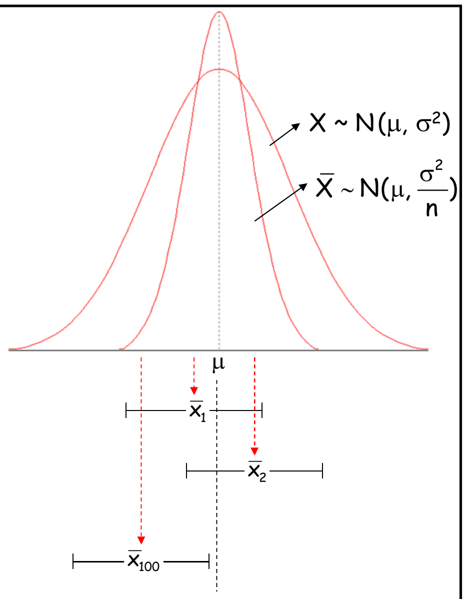
Todos os intervalos com mesmo nível de confiança terão a mesma amplitude:

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

Se 100 amostras de tamanho n são retiradas da população, **99** dos intervalos construídos cobrirão μ e **um** deles não cobrirá

Significado de um IC para μ, com nível de confiança $\mathbf{1-\alpha=0{,}95}$ e σ^2 **conhecida**





Todos os intervalos com mesmo nível de confiança terão a mesma amplitude:

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

95% dos intervalos contêm μ

Situação 2. Quando o tamanho da amostra é pequeno **(n<30)**

⇨ Quando a amostra é pequena, não podemos supor que o desvio padrão da amostra (**s**) seja uma estimativa suficientemente aproximada do parâmetro σ.

$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

desconhecido

⇨ Nesse caso, em vez de **Z**, utilizamos a variável **T** que não tem distribuição normal, mas tem distribuição **t de Student**, com parâmetro ν.

$$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(\nu)$$

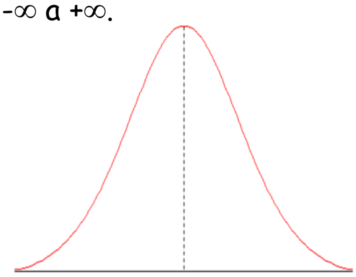
letra grega "ni"

William Gosset
(1857- 1936)

Em 1908, o inglês William Gosset, descobriu a distribuição t no intuito de resolver problemas relativos a pequenas amostras. Gosset trabalhava, na época, numa cervejaria na Irlanda e estava ciente de que seus empregadores não queriam que funcionários publicassem o que quer que fosse, com receio de que segredos industriais caíssem no domínio público e, principalmente, nas mãos da concorrência. Por isso, Gosset ao descobrir uma nova distribuição de probabilidades publicou seus trabalhos sob o pseudônimo de Student.

A distribuição t de Student

A distribuição **t** tem formato de campânula, é simétrica em torno da média ($\mu=0$), localizada no centro da distribuição, e varia de $-\infty$ a $+\infty$.



Função densidade de probabilidade

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

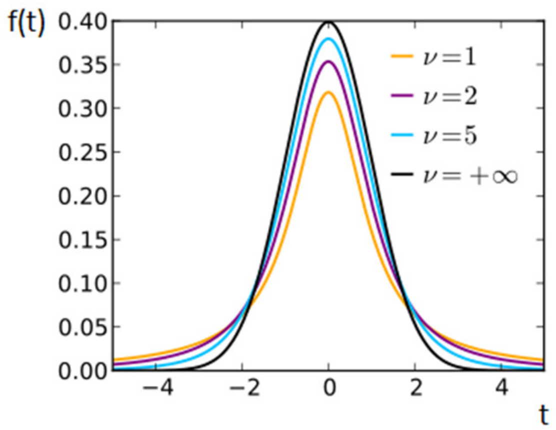
⇨ Se assemelha à curva da distribuição normal padrão, sendo um pouco mais achatada no centro.

⇨ **Parâmetro** → número de graus de liberdade: $\nu=n-1$

⇨ Quando muda o tamanho da amostra (n), o formato da curva da **distribuição t** se altera.

A distribuição t de Student

Quando muda ν, o formato da curva da **distribuição t** se altera.

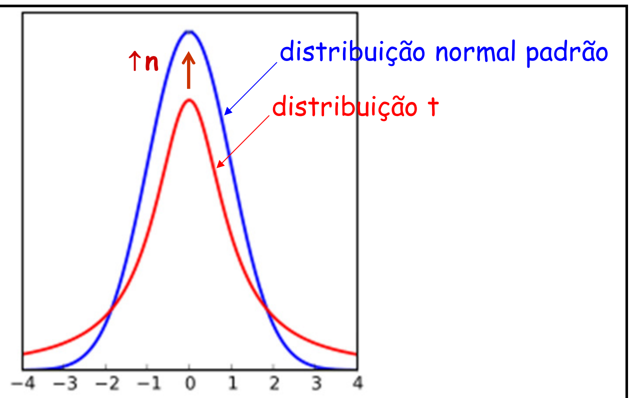


$T=\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ $Z=\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$n \to N$

$S \to \sigma$

$T \to Z$



A **distribuição t** se aproxima da normal padrão à medida que o **n** cresce

Tamanho da amostra se aproxima do tamanho da população (**n**→**N**)

⇩

Estimador **S** se aproxima do parâmetro **σ** (**S**→**σ**)

⇩

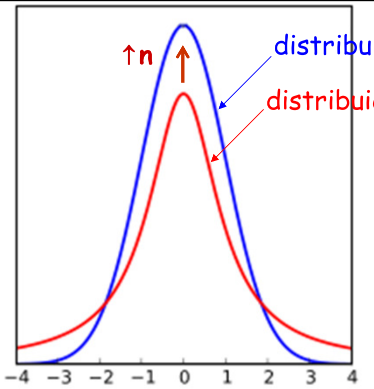
Estatística **T** se aproxima da variável **Z** (**T**→**Z**)

$T=\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ $Z=\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$n \rightarrow N$

$S \rightarrow \sigma$

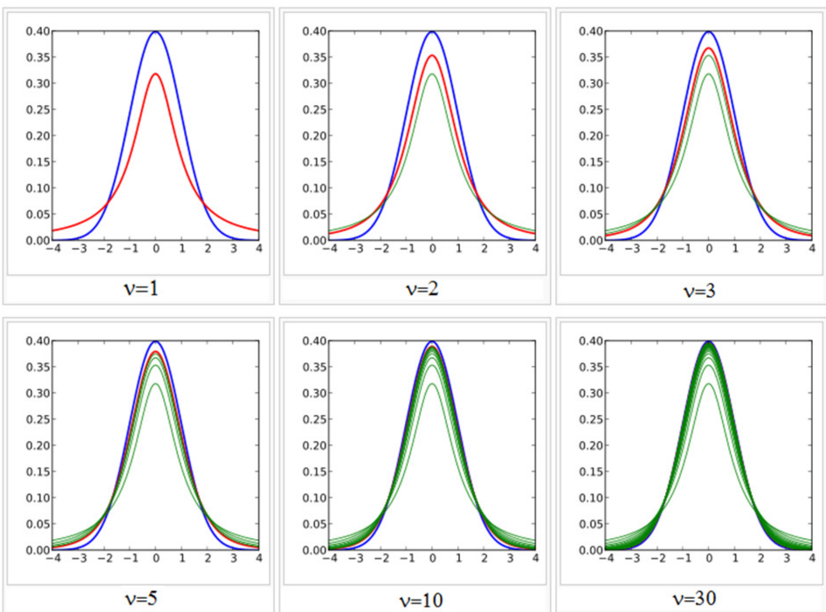
$T \rightarrow Z$



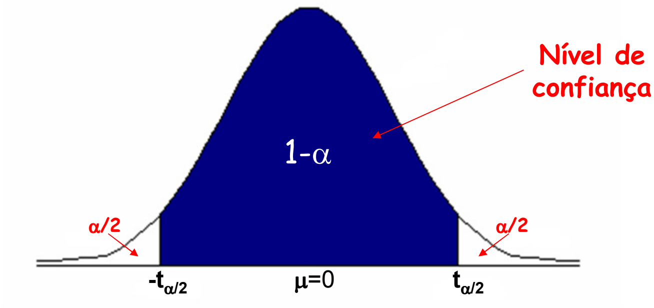
A **distribuição t** se aproxima da normal padrão à medida que o **n** cresce

Na prática:
ν = 30 → **distribuição t** é aproximadamente igual à **normal padrão**
ν = 120 → **distribuição t** é exatamente igual à **normal padrão** (curvas se sobrepõem)

ν = 30 → **distribuição t** é aproximadamente igual à **normal padrão**



Construção do intervalo de confiança para μ a partir de T



$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1-\alpha$$

Sabendo que $T = \dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ e fazendo a substituição, temos:

$$P\left(-t_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < t_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-t_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < t_{\alpha/2}\right) = 1-\alpha$$

Os limites (aleatórios) têm probabilidade 1-α de conter (ou cobrir) a média.

$$P\left(-t_{\alpha/2}\frac{S}{\sqrt{n}} < \overline{X}-\mu < t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

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

$$P\left[(-\overline{X}-t_{\alpha/2}\frac{S}{\sqrt{n}} < -\mu < -\overline{X}+t_{\alpha/2}\frac{S}{\sqrt{n}})(-1)\right] = 1-\alpha$$

$$P\left(\overline{X}+t_{\alpha/2}\frac{S}{\sqrt{n}} > \mu > \overline{X}-t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\underbrace{\overline{X}-t_{\alpha/2}\frac{S}{\sqrt{n}}}_{\text{Limite inferior}} < \mu < \underbrace{\overline{X}+t_{\alpha/2}\frac{S}{\sqrt{n}}}_{\text{Limite superior}}\right) = 1-\alpha$$

Nível de confiança

Variável T $\rightarrow T = \dfrac{\overline{X} - \mu}{\dfrac{S}{\sqrt{n}}} \sim t\ (\nu)$

onde:

$\overline{X}$: é a média da amostra (estimador de μ)

S : é desvio padrão da amostra

n : é o tamanho da amostra

ν = n–1 : é o número de graus de liberdade de S^2

Generalização $\rightarrow T = \dfrac{\hat{\theta} - \theta}{S(\hat{\theta})} \sim t\ (\nu)$

onde:

θ é o parâmetro que está sendo estimado

$\hat{\theta}$ é o estimador do parâmetro

$S(\hat{\theta})$ é o estimador do desvio (ou erro) padrão de $\hat{\theta}$

Intervalo de confiança para μ

$$IC\ (\mu;\ 1-\alpha):\ \overline{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

onde: $t_{\alpha/2}$ é o valor da estatística **T** que delimita a área $\alpha/2$

Este valor é encontrado na tabela da **distribuição t** de Student (Tabela II do Apêndice), a partir dos valores de ν e de α.

Generalizando a expressão, temos

$$IC\ (\ \theta\ ;\ 1-\alpha):\ \hat{\theta} \pm t_{\alpha/2}\ S(\hat{\theta})$$

Pressuposição

Para a utilização desta metodologia a seguinte pressuposição deve ser atendida:

A variável em estudo tem distribuição normal.

$$X \sim N(\mu, \sigma^2)$$

Devido à aproximação com a distribuição normal padrão quando **n>30**, a variável T, que tem distribuição t de Student, poderá ser utilizada para construir intervalos de confiança para a média, também quando a amostra for grande.

Mas a **distribuição normal padrão** só pode ser utilizada quando a amostra é grande.

Exercício proposto:

Através da amostra de tamanho 15 que segue, procura-se estimar a verdadeira potência média de aparelhos eletrônicos de alta sensibilidade medida em microwatts: 26,7; 25,8; 24,0; 24,9; 26,4; 25,9; 24,4; 21,7; 24,1; 25,9; 27,3; 26,9; 27,3; 24,8 e 23,6. Construa o intervalo de confiança, ao nível de 99%, para a verdadeira potência média dos aparelhos.

Resolução:

Variável em estudo: X = potência de aparelhos eletrônicos (microwatts)

Pressuposição: A variável em estudo tem distribuição normal.

Estimativas pontuais:

$\bar{x}$ =25,31	α=0,01
s=1,589	ν=15 – 1=14
n=15	$t_{\alpha/2(14)}$ = 2,763

Construção do intervalo:

$IC(\mu;1-\alpha): \bar{X} \pm t_{\alpha/2}\dfrac{S}{\sqrt{n}}$

IC (μ; 0,99): $25,31 \pm 2,763 \times \dfrac{1,589}{\sqrt{15}}$

IC (μ; 0,99): $25,31 \pm 1,13$

Limite inferior: 25,31 – 1,13 = 24,18

Limite superior: 25,31 + 1,13 = 26,44

Conclusão: Os limites de confiança, ao nível de 99%, para a verdadeira potência média dos aparelhos eletrônicos são 24,42 e 26,19 microwatts.

Profa. Clause Piana 62

2. Intervalo de confiança para a **diferença entre médias** de duas populações ($\mu_1 - \mu_2$)

Para utilizar a estatística **T** no estudo de uma variável X em duas populações distintas, três **pressupostiçoes** devem ser atendidas:

1. A variável em estudo tem distribuição normal:

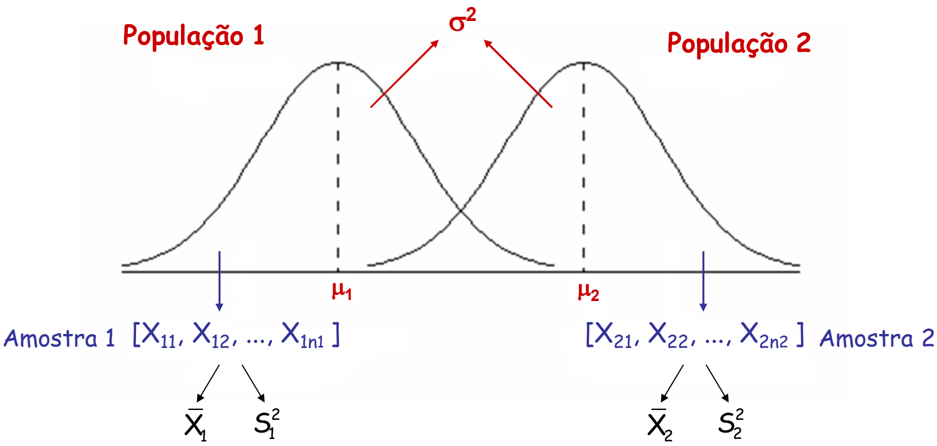
$$X \sim N(\mu, \sigma^2)$$

2. As variâncias das populações são iguais ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

3. As amostras retiradas das populações são **independentes**

Profa. Clause Piana 63

Pressupostições



Profa. Clause Piana 64

Atendidas as pressuposições, desejamos comparar as médias das populações, estimando por intervalo, o parâmetro θ.

Utilizamos, então, a variável T

$$T = \frac{\hat{\theta} - \theta}{S(\hat{\theta})} \sim t\,(\nu)$$

onde:

$$\theta = \mu_1 - \mu_2 \longleftarrow \boxed{\text{Parâmetro estimado}}$$

$$\hat{\theta} = \overline{X}_1 - \overline{X}_2 \longleftarrow \boxed{\text{Estimador do parâmetro}}$$

$$S(\hat{\theta}) = S\left(\overline{X}_1 - \overline{X}_2\right) \longleftarrow \boxed{\text{Estimador do erro padrão do } \hat{\theta}}$$

$$\nu = (n_1 - 1) + (n_2 - 1)$$

Grau de liberdade combinado

Como obter a estimativa do erro padrão de $\hat{\theta}$?

$$\hat{\theta} = \overline{X}_1 - \overline{X}_2$$

$$S\left(\overline{X}_1 - \overline{X}_2\right) = \sqrt{S^2\left(\overline{X}_1 - \overline{X}_2\right)}$$

$$= \sqrt{S^2\left(\overline{X}_1\right) + S^2\left(\overline{X}_2\right)}$$

$$= \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$= \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}$$

$$S\left(\overline{X}_1 - \overline{X}_2\right) = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S^2}$$

Variância da média

$$V(\overline{X}) = \frac{\sigma^2}{n}$$

Estimador da variância da média

$$S^2\left(\overline{X}\right) = \frac{S^2}{n}$$

Variância combinada das amostras

$$S^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}$$

Profa. Clause Piana 66

$$S(\hat{\theta}) = S(\overline{X}_1 - \overline{X}_2) = \boxed{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S^2}}$$ ← Estimador do erro padrão do $\hat{\theta}$

$$S^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}$$ ← Variância combinada

$$\nu = (n_1 - 1) + (n_2 - 1)$$ ← Grau de liberdade combinado

$$IC\ (\theta\ ;\ 1-\alpha):\ \hat{\theta} \pm t_{\alpha/2}\ S(\hat{\theta})$$

$$IC\ (\mu_1 - \mu_2\ ;\ 1-\alpha): \overline{X}_1 - \overline{X}_2 \pm t_{\alpha/2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S^2}$$

Exercício proposto:

Um pesquisador da área de computação está investigando a utilidade de duas diferentes linguagens de programação (A e B) na melhoria das tarefas computacionais.

Trinta programadores experientes, familiarizados com ambas as linguagens, foram divididos aletoriamente em dois grupos. Cada grupo codificou uma função padrão em uma das linguagens e os tempos de codificação da função (em minutos) foram registrados.

As medidas de cada grupo são apresentadas no quadro abaixo.

	n	Média	Variância
Linguagem A	15	17,5	2,31
Linguagem B	15	20,5	3,02

Utilizando um intervalo de confiança, ao nível de 95%, verifique se, em média, o tempo de codificação da função padrão difere entre as linguagens.

Resolução:

Variável em estudo: X = tempo de codificação da função padrão (minutos)

Pressupposições: 1. A variável em estudo tem distribuição normal.
2. As variâncias das populações são iguais ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).
3. As amostras retiradas das populações são independentes.

Estimativas pontuais:

Linguagem A	Linguagem B
$\bar{X}_1 = 17{,}5$	$\bar{X}_2 = 20{,}5$
$s_1^2 = 2{,}31$	$s_2^2 = 3{,}02$
$n_1 = 15$	$n_2 = 15$

$\alpha = 0{,}05$

$\nu = (15 - 1) + (15 - 1) = 28$

$t_{\alpha/2(28)} = 2{,}048$

$$s^2 = \frac{2{,}31 + 3{,}02}{2} = 2{,}67$$

Construção do intervalo:

$$IC(\mu_1 - \mu_2; 1-\alpha): \bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s^2}$$

$$IC(\mu_1 - \mu_2; 0{,}95): 20{,}5 - 17{,}5 \pm 2{,}048\sqrt{\left(\frac{1}{15} + \frac{1}{15}\right)2{,}67}$$

$$IC(\mu_1 - \mu_2; 0{,}95): 3 \pm 1{,}22$$

Limite inferior: 3 - 1,22 = 1,78 min

Limite superior: 3 + 1,22 = 4,22 min

$$IC(\mu_1 - \mu_2; 0{,}95): [1{,}78;\ 4{,}22]$$

Conclusão: Concluímos com 95% de confiança que a verdadeira diferença entre tempos médios de programação das duas linguagens é coberta pelos limites 1,78 e 4,22 minutos. Como o valor zero não está compreendido no intervalo, concluímos que as médias diferem significativamente entre si.

Intervalo de confiança

$$IC(\theta; 1-\alpha): \hat{\theta} \pm t_{\alpha/2}\ S(\hat{\theta})$$

- Para uma média

$$IC(\mu; 1-\alpha): \bar{X} \pm t_{\alpha/2}\ \frac{S}{\sqrt{n}}$$

- Para a diferença entre duas médias

$$IC(\mu_1 - \mu_2; 1-\alpha): \bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)s^2}$$

Intervalos de confiança - **Exemplos resolvidos**

Intervalo de confiança para μ (grandes amostras)

Exemplo resolvido:

Um engenheiro de desenvolvimento de um fabricante de pneus está investigando a vida do pneu em relação a um novo componente de borracha. Ele fabricou 40 pneus e testou-os até o fim da vida em um teste na estrada. A média e o desvio padrão da amostra são 61.492 e 6.085 km, respectivamente. O engenheiro acredita que a vida média desse novo pneu está em excesso em relação a 60.000 km. Obtenha o intervalo de confiança, ao nível de 95%, para a vida média do pneu e conclua a respeito da suposição do engenheiro.

Resolução:

Variável em estudo: X = durabilidade do pneu (km rodados)

Pressuposições:

1. A variável em estudo tem distribuição normal.
2. A amostra é grande ($n \geq 30$).

Estimativas:

$\bar{x}$ = 61.492 km
s = 6.085 km
n = 40 pneus

$\alpha/2 = 0{,}025$
Área entre 0 e $z_{\alpha/2} = 0{,}475$
$z_{\alpha/2} = z_{.975} = 1{,}96$
quantil

Construção do intervalo:

$$IC(\mu; 1-\alpha): \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$IC(\mu; 0{,}95): 61.492 \pm 1{,}96 \times \frac{6.085}{\sqrt{40}}$$

$$IC(\mu; 0{,}95): 61.492 \pm 1.886$$

Limite inferior: 61.492 – 1.886 = 59.606
Limite superior: 61.492 + 1.886 = 63.378

Conclusão: O intervalo de confiança, ao nível de 95%, para a verdadeira durabilidade média do novo pneu é de 59.606 a 63.378 km. Como o valor 60.000 km está coberto pelo intervalo, a durabilidade média do novo pneu não supera este valor.

Intervalo de confiança para μ (pequenas amostras)

Exemplo resolvido:

Através da amostra de tamanho 15 que segue, procura-se estimar a verdadeira potência média de aparelhos eletrônicos de alta sensibilidade medida em microwatts: 26,7; 25,8; 24,0; 24,9; 26,4; 25,9; 24,4; 21,7; 24,1; 25,9; 27,3; 26,9; 27,3; 24,8 e 23,6. Construa o intervalo de confiança, ao nível de 99%, para a verdadeira potência média dos aparelhos.

Resolução:

Variável em estudo: X = potência de aparelhos eletrônicos (microwatts)

Pressuposição: A variável em estudo tem distribuição normal.

Estimativas pontuais:

$\bar{x}$ =25,31 $\quad\alpha$=0,01
s=1,589 $\quad\nu$=15 – 1=14
n=15 $\quad t_{\alpha/2}$ = 2,977

Construção do intervalo:

$$IC(\mu; 1-\alpha): \bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

$$IC(\mu; 0{,}99): 25{,}31 \pm 2{,}997 \times \frac{1{,}589}{\sqrt{15}}$$

$$IC(\mu; 0{,}99): 25{,}31 \pm 1{,}22$$

Limite inferior: 25,31 – 1,21 = 24,1
Limite superior: 25,31 + 1,21 = 26,52

Conclusão: Os limites de confiança, ao nível de 99%, para a verdadeira potência média dos aparelhos eletrônicos são 24,1 e 26,52 microwatts.

Intervalo de confiança para a diferença entre médias ($\mu_1 - \mu_2$)

Exemplo resolvido:

Na fabricação de semicondutores o ataque químico por via úmida é freqüentemente usado para remover silicone da parte posterior das pastilhas antes da metalização. A taxa de ataque é uma característica importante nesse processo e é sabido que ela segue uma distribuição normal. Duas soluções diferentes para ataque químico são comparadas, usando duas amostras aleatórias de pastilhas. As taxas observadas de ataque (10^{-3} polegadas/min) são dadas a seguir:

Solução 1	9,9	9,4	9,3	9,6	10,2	10,6	10,3	10,0	10,3	10,1
Solução 2	10,2	10,6	10,7	10,4	10,5	10,0	10,7	10,4	10,3	-

Os dados justificam a afirmação de que a taxa média de ataque seja a mesma para ambas as soluções? Considere que ambas as populações têm variâncias iguais, construa o intervalo de confiança, ao nível de 95%, para a diferença entre as médias e conclua.

Resolução:

Variável em estudo: X = taxa de ataque químico por via úmida (10^{-3} polegadas/min)

Pressuposições: 1. A variável em estudo tem distribuição normal.
2. As variâncias das populações são iguais ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).
3. As amostras retiradas das populações são independentes.

Estimativas:

Solução 1	Solução 2
$\bar{X}_1 = 9,97$	$\bar{X}_2 = 10,42$
$s_1^2 = 0,178$	$s_2^2 = 0,054$
$n_1 = 10$	$n_2 = 9$

$\alpha = 0,05$

$\nu = (10 - 1) + (9 - 1) = 17$

$t_{\alpha/2} = 2,11$

$$S^2 = \frac{0,178 \times (10-1) + 0,054 \times (9-1)}{(10-1)+(9-1)} = 0,1196$$

Construção do intervalo:

$$IC(\mu_1 - \mu_2; 1-\alpha): \bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)S^2}$$

$$IC(\mu_1 - \mu_2; 0,95): (9,97 - 10,42) \pm 2,11\sqrt{\left(\frac{1}{10} + \frac{1}{9}\right) \times 0,1196}$$

$$IC(\mu_1 - \mu_2; 0,95): -0,45 \pm 0,335$$

Limite inferior: -0,45 – 0,335 = -0,785

Limite superior: –0,45 + 0,335 = -0,115

Conclusão: A confiança de que a verdadeira diferença entre as taxas médias de ataque químico das duas soluções seja coberta pelo intervalo de -0,785 a $-0,115 \times 10^{-3}$ polegadas/min é de 95%. Como o valor zero não está compreendido no intervalo, concluímos que as médias diferem significativamente entre si.

Bibliografia

BUSSAB, W.O.; MORETTIN, P.A. **Estatística Básica**. São Paulo: Saraiva. 2006. 526p.

FERREIRA, D.F. **Estatística Básica**. Lavras: Editora UFLA, 2005, 664p.

MLODINOW, L. **O andar do bêbado. Como o acaso determina nossas vidas**. Rio de Janeiro: Editora Zahar, 2009, 264p.

SILVEIRA JÚNIOR, P., MACHADO, A.A., ZONTA, E.P., SILVA, J.B. da Curso de Estatística v.1, Pelotas: Universidade Federal de Pelotas, 1989. 135p.

Sistema Galileu de Educação Estatística. Disponível em: http://www.galileu.esalq.usp.br/topico.html