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

Profª. Clause Piana

4.2.1 Processos de estimação

Existem duas maneiras de obter a estimativa de um parâmetro:

- Estimação por ponto (ou pontual)
- Estimação por intervalo

4.2.1 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

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

$\mu=2,5$

$\sigma^2=1,05$

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

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)	$\bar{x}_{16} = 4$	$s_{16}^2 = 0$

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

$\mu=2{,}5$

$\sigma^2=1{,}05$

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

$P(2 \leq X \leq 3) = 0{,}68$

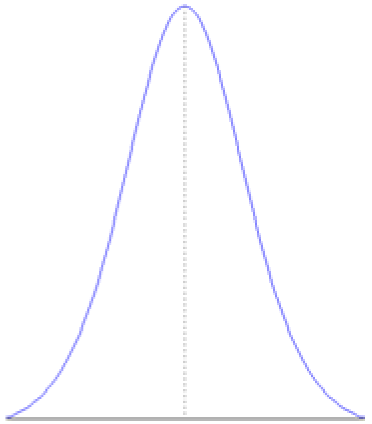
Distribuição amostral da média das amostras de tamanho 3

$\bar{X} = \bar{x}$	1	1,33	1,67	2	2,33	2,67	3	3,33	3,67	4	Σ
$P(\bar{X} = \bar{x})$	0,008	0,044	0,09	0,159	0,207	0,207	0,159	0,09	0,036	0,008	1

$P(2 \leq X \leq 3) = 0{,}72$

Se **n** é grande,

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



$E(\bar{X}) = \mu$

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.

Exemplo 1

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 95%, para a verdadeira potência média dos aparelhos.

Exemplo 2

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.

Intervalos de confiança para a média

Profª. Clause Piana

Como construir intervalos de confiança?

4.2.1.1 Intervalo de confiança para a média de uma população (μ)

Duas situações

- Grandes amostras ($n>30$) ou variância populacional conhecida
- Pequenas amostras ($n\leq30$) ou variância populacional desconhecida

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

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

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

$Z \sim N(0,1)$

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

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

$Z \sim N(0,1)$

Transformar a variável $\bar{X}$ em Z

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

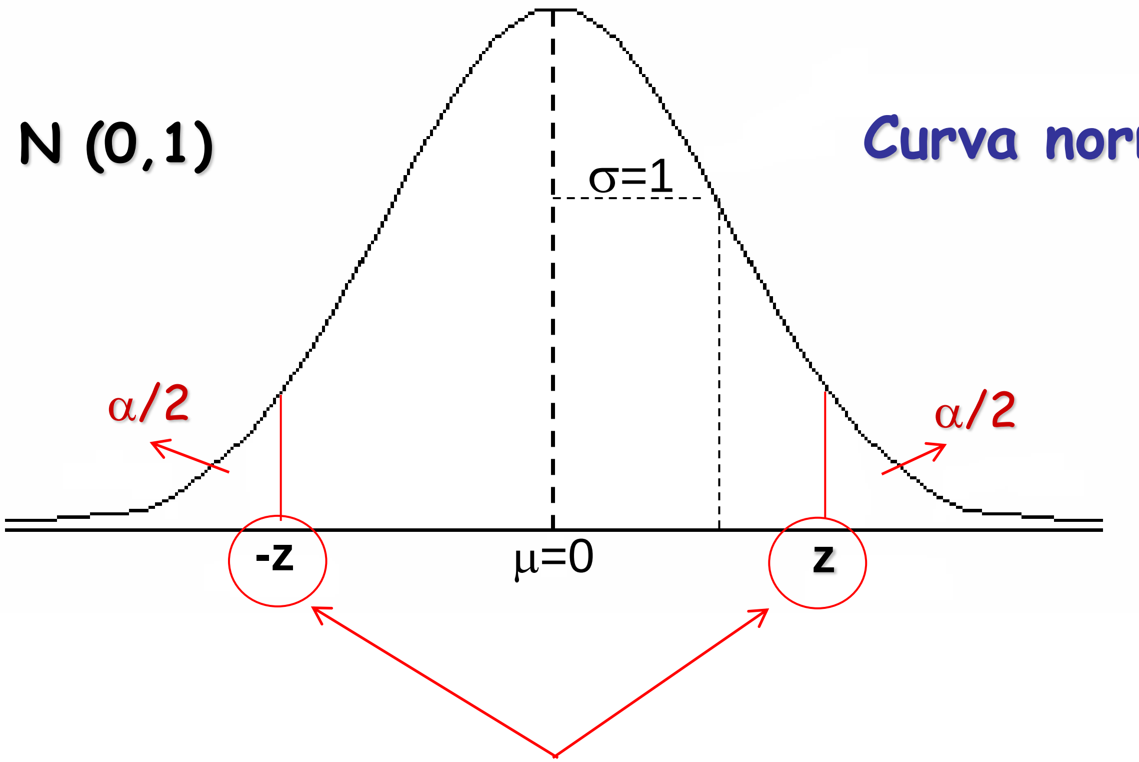
$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

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

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

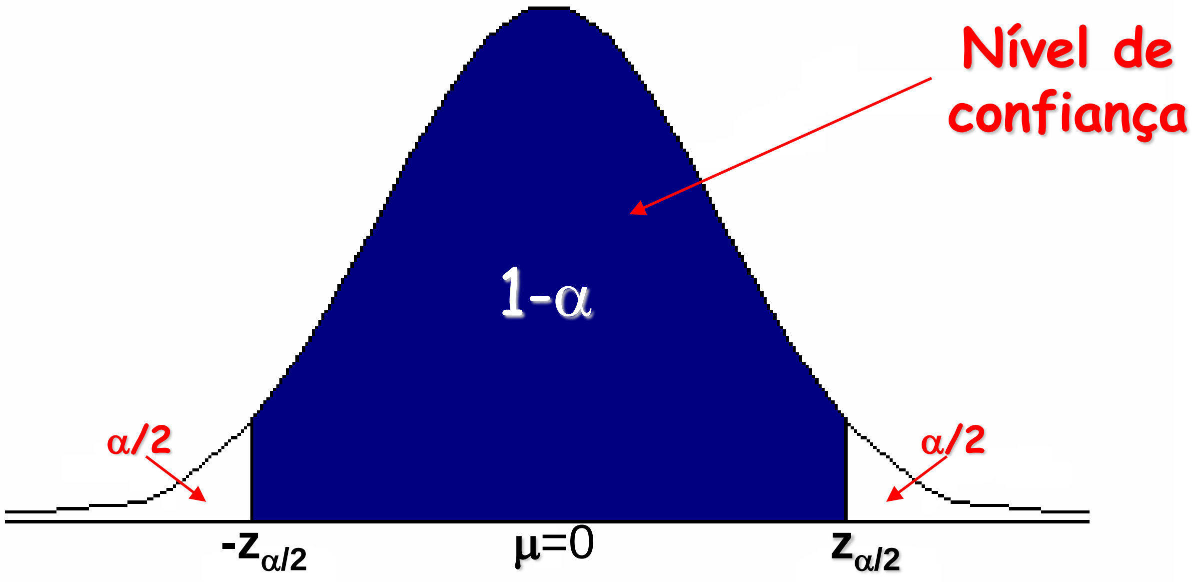


Começamos a construir um intervalo de confiança para μ, a partir da curva normal padrão, estabelecendo dois limites simétricos para os valores da variável Z.

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

Construção do intervalo de confiança para μ a partir de Z

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



$-z_{\alpha/2}$: 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 $z_{\alpha/2}$

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

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

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

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

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

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

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

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

Intervalo de confiança para μ

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

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

A probabilidade de que os limites contenham (ou cubram) o verdadeiro valor de μ é $1-\alpha$.
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 ($n \geq 30$).

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

Estimativas:

$\bar{x}$ = 61.492 km

s = 6.085 km

n = 40 pneus

$\alpha/2 = 0{,}025$

Área entre 0 e $z_{\alpha/2} = 0{,}475$

$z_{\alpha/2} = z_{,975} = 1{,}96$

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

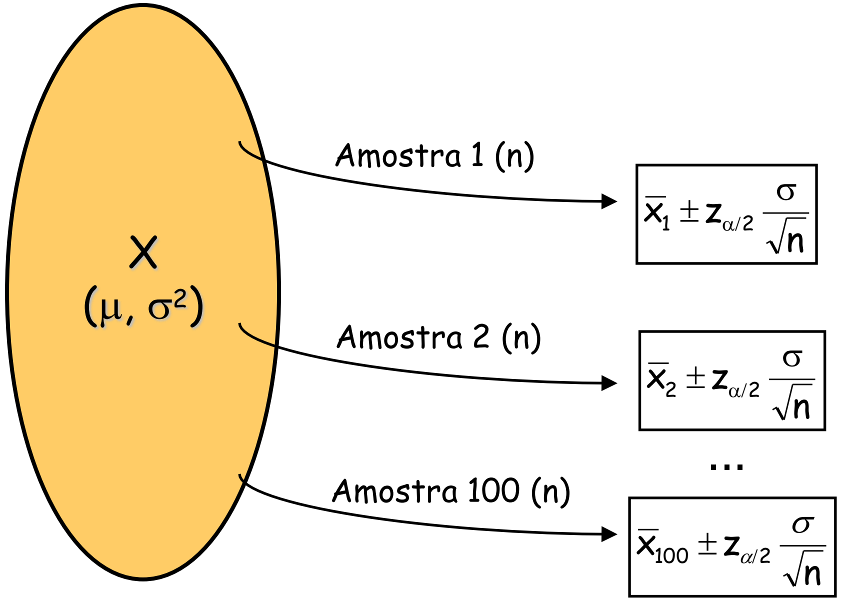
Limite superior: 61.492 + 1.886 = 63.378

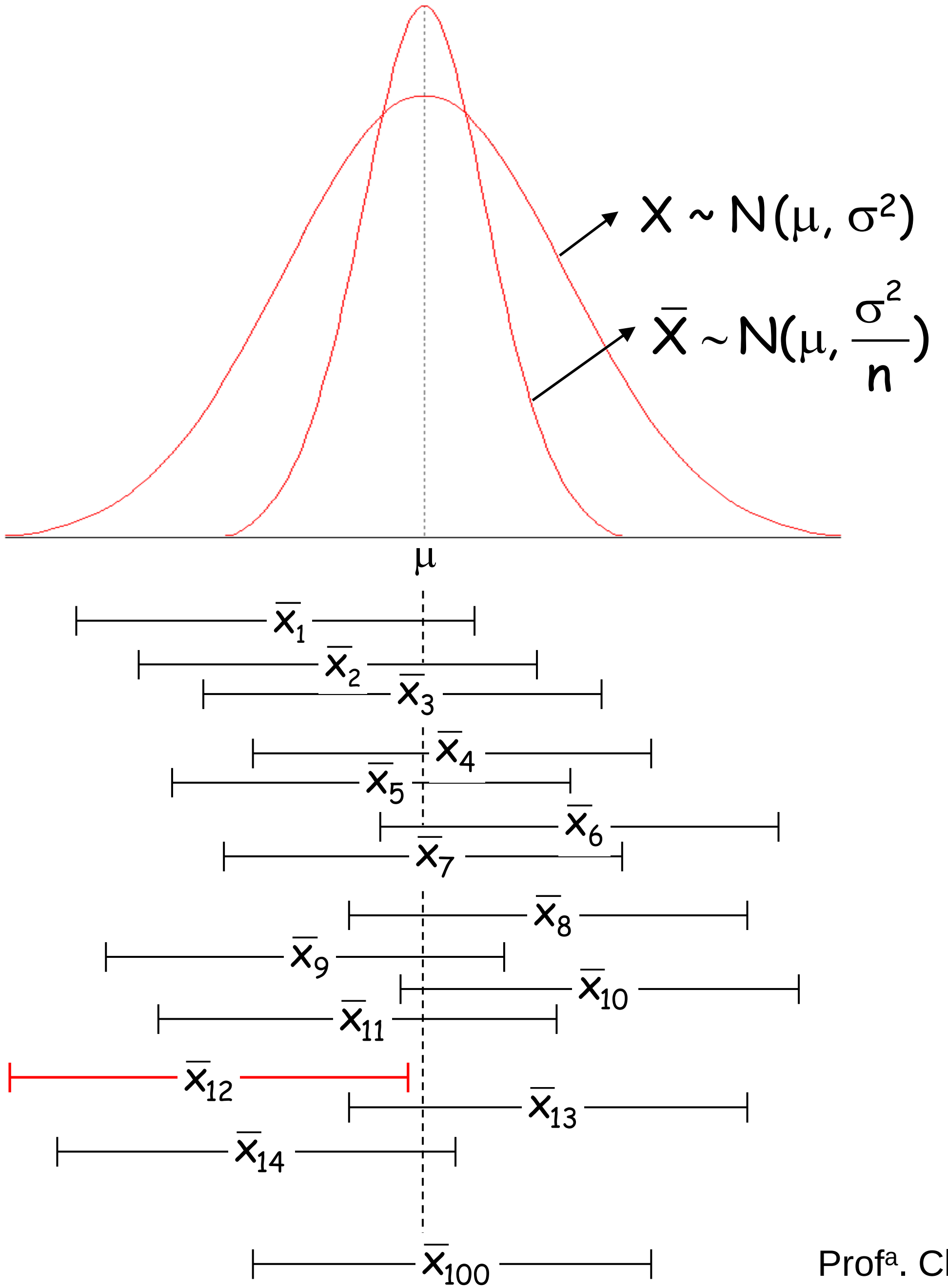
Conclusão: O intervalo de confiança, ao nível de 95%, para a verdadeira durabilidade média do novo pneu é de 59.606 a 63.378 km. Como o valor 60.000 km está coberto pelo intervalo, a durabilidade média do novo pneu não supera este valor.

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

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

População





Todos os intervalos com mesmo nível de confiança terão a mesma amplitude: $2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

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

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**.

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

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.

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

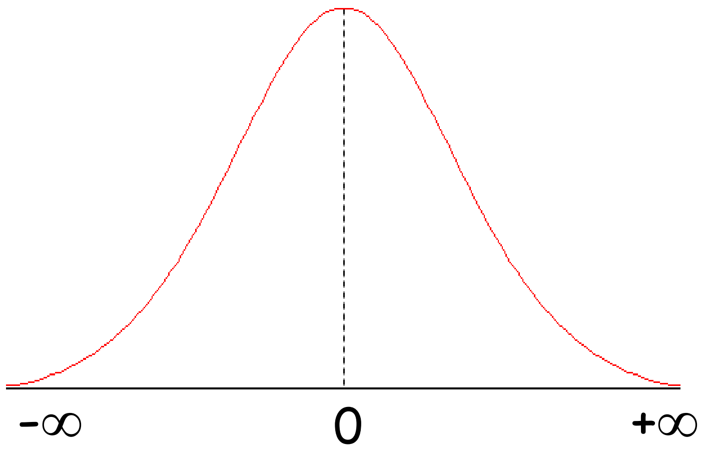
letra grega "ni"

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

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

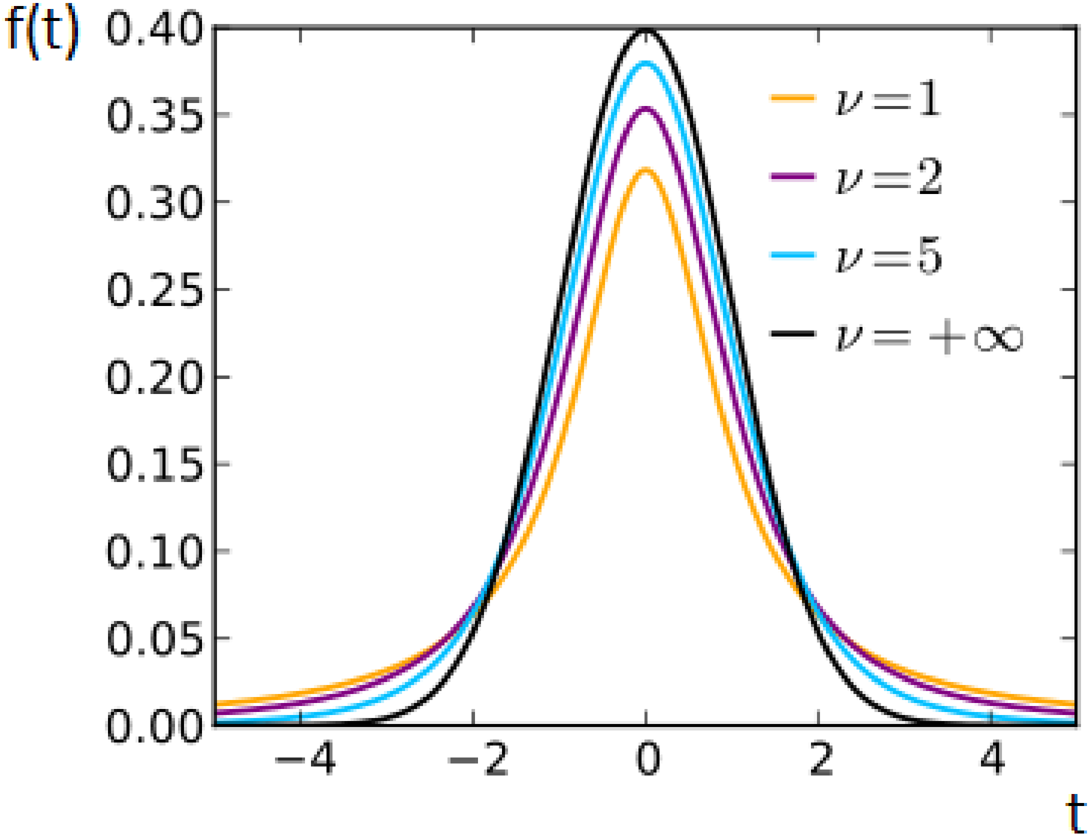
⇨ Se assemelha à curva da distribuição normal padrão, sendo um pouco mais achatada no centro.

⇨ **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.

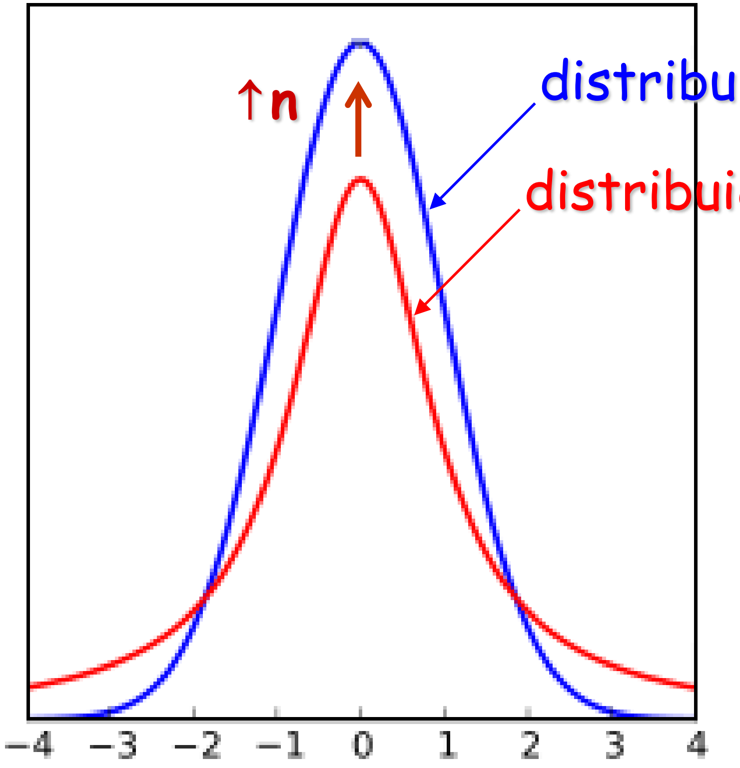


$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \qquad Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$n \rightarrow N$

$S \rightarrow \sigma$

$T \rightarrow Z$



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

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

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

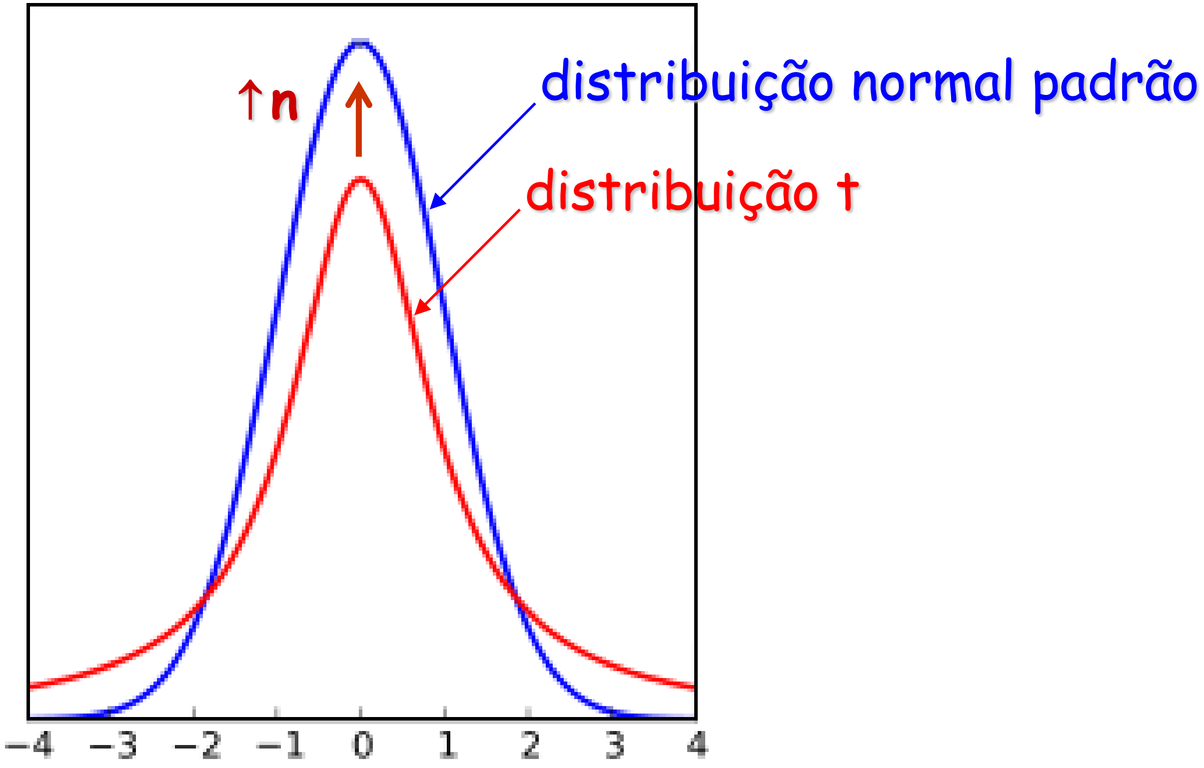
Estatística **T** se aproxima da variável **Z** (**T**→**Z**)

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

$n \rightarrow N$

$S \rightarrow \sigma$

$T \rightarrow Z$



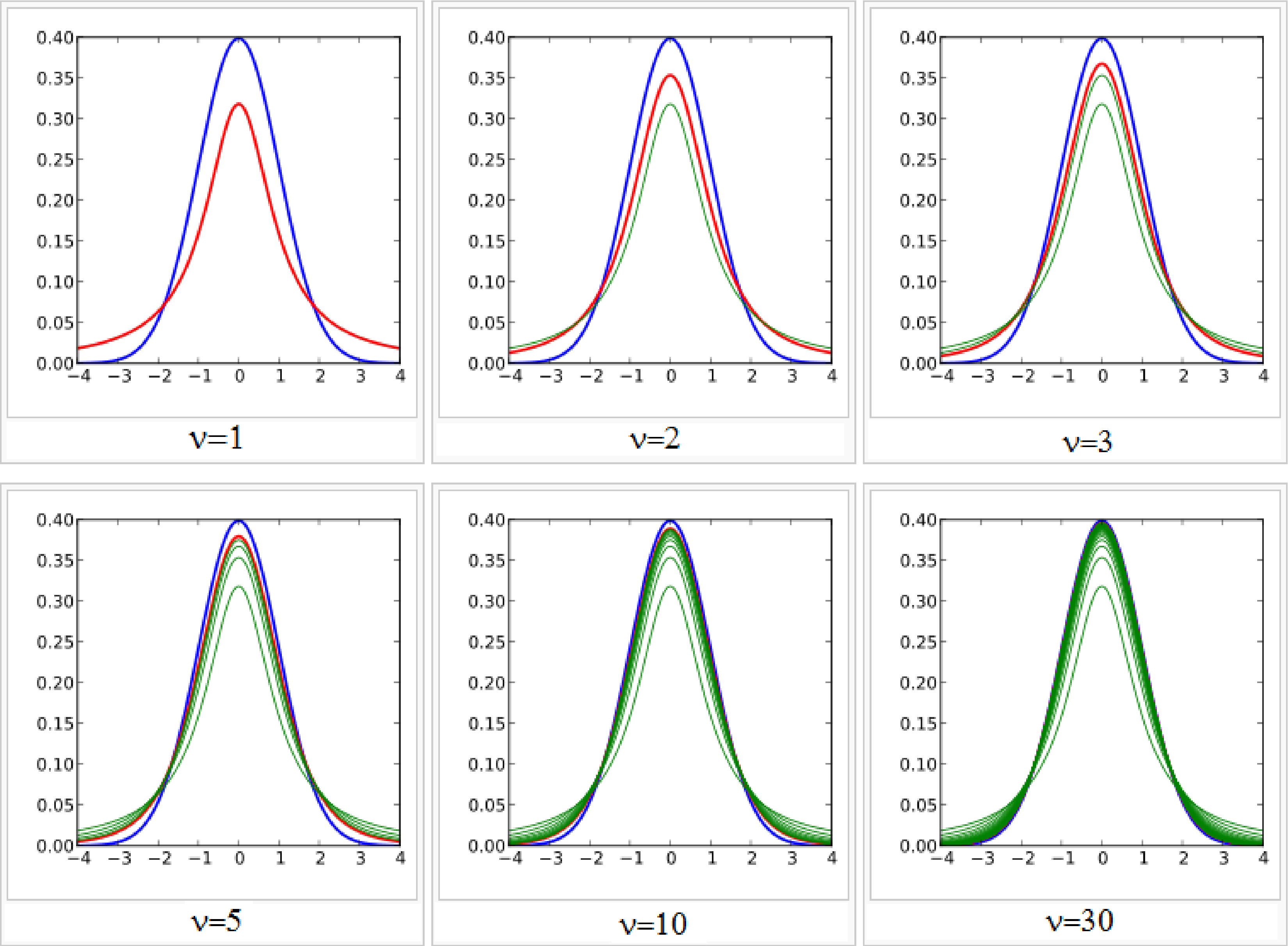
A **distribuição t** se aproxima da normal padrão à medida que o **n** cresce

Na prática:

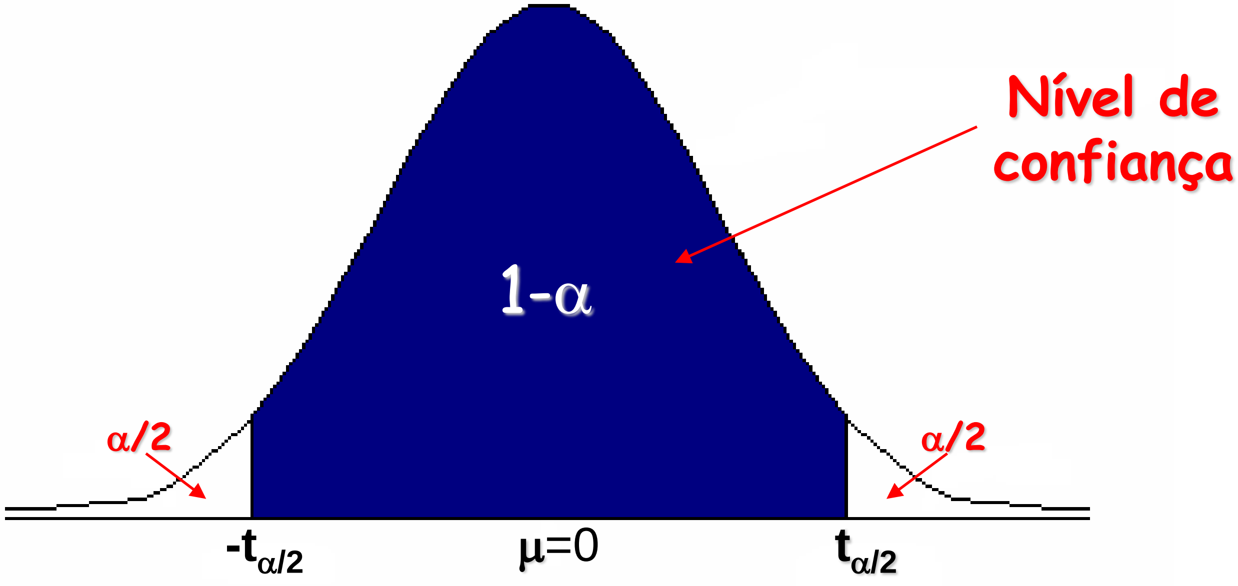
$\nu = 30 \rightarrow$ **distribuição t** é aproximadamente igual à **normal padrão**

$\nu = 120 \rightarrow$ **distribuição t** é exatamente igual à **normal padrão** (curvas se sobrepõem)

$\nu = 30 \rightarrow$ **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 $T = \dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ e fazendo a substituição, temos:

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

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

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

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

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

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

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

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

Nível de confiança

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

onde:

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

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

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

Graus de Liberdade (ν)	**Limites bilaterais**: $P(\lvert t \rvert > t_{\alpha/2})$							
	Nível de Significância (α)							
	0,50	0,20	0,10	0,05	0,025	0,02	0,01	0,005
1	1,000	3,078	6,314	12,706	25,542	31,821	63,657	127,320
2	0,816	1,886	2,920	4,303	6,205	6,965	9,925	14,089
3	0,715	1,638	2,353	3,183	4,177	4,541	5,841	7,453
4	0,741	1,533	2,132	2,776	3,495	3,747	4,604	5,598
5	0,727	1,476	2,015	2,571	3,163	3,365	4,032	4,773
6	0,718	1,440	1,943	2,447	2,969	3,143	3,707	4,317
7	0,711	1,415	1,895	2,365	2,841	2,998	3,500	4,029
8	0,706	1,397	1,860	2,306	2,752	2,896	3,355	3,833
9	0,703	1,383	1,833	2,262	2,685	2,821	3,250	3,690
10	0,700	1,372	1,813	2,228	2,634	2,764	3,169	3,581
11	0,697	1,363	1,796	2,201	2,503	2,718	3,106	3,497
12	0,695	1,356	1,782	2,179	2,560	2,681	3,055	3,428
13	0,694	1,350	1,771	2,160	2,533	2,650	3,012	3,373
14	0,692	1,345	1,761	2,145	2,510	2,624	2,977	3,326
15	0,691	1,341	1,753	2,132	2,490	2,602	2,947	3,286
16	0,690	1,337	1,746	2,120	2,473	2,583	2,921	3,252
17	0,689	1,333	1,740	2,110	2,458	2,567	2,898	3,223
18	0,688	1,330	1,734	2,101	2,445	2,552	2,878	3,197
19	0,688	1,328	1,729	2,093	2,433	2,539	2,861	3,174
20	0,687	1,325	1,725	2,086	2,423	2,528	2,845	3,153
21	0,686	1,323	1,721	2,080	2,414	2,518	2,831	3,135
22	0,686	1,321	1,717	2,074	2,406	2,508	2,819	3,119
23	0,685	1,319	1,714	2,069	2,398	2,500	2,807	3,104
24	0,685	1,318	1,711	2,064	2,391	2,492	2,797	3,091
25	0,684	1,316	1,708	2,060	2,385	2,485	2,787	3,078
26	0,684	1,315	1,706	2,056	2,379	2,479	2,779	3,067
27	0,684	1,314	1,703	2,052	2,373	2,473	2,771	3,057
28	0,683	1,313	1,701	2,048	2,369	2,467	2,763	3,047
29	0,683	1,311	1,699	2,045	2,364	2,462	2,756	3,038
30	0,683	1,310	1,697	2,042	2,360	2,457	2,750	3,030
40	0,681	1,303	1,684	2,021	2,329	2,423	2,705	2,971
60	0,679	1,296	1,671	2,000	2,299	2,390	2,660	2,915
120	0,677	1,289	1,658	1,980	2,270	2,358	2,617	2,860
...	0,674	1,282	1,645	1,960	2,241	2,326	2,576	2,807
Graus de Liberdade (ν)	0,25	0,10	0,05	0,025	0,0125	0,01	0,005	0,0025
	Nível de Significância (α)							
	Limites unilaterais: $P(\lvert t \rvert > t_{\alpha})$							

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 95%, 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:

$\bar{x}=25,31$ $\qquad \alpha=0,05$

$s=1,589$ $\qquad \nu=15-1=14$

$n=15$ $\qquad t_{\alpha/2(14)}=2,145$

Construção do intervalo:

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

$$IC\ (\mu;\ 0,95):\ 25,31 \pm 2,145 \times \frac{1,589}{\sqrt{15}}$$

$$IC\ (\mu;\ 0,95):\ 25,31 \pm 1,13$$

Limite inferior: $25,31 - 0,88 = 24,43$

Limite superior: $25,31 + 0,88 = 26,19$

Conclusão: Os limites de confiança, ao nível de 95%, para a verdadeira potência média dos aparelhos eletrônicos são 24,43 e 26,19 microwatts.

Exercício:

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:

$\bar{x} = 25{,}31$ $\quad\alpha = 0{,}01$

$s = 1{,}589$ $\quad\nu = 15 - 1 = 14$

$n = 15$ $\quad t_{\alpha/2(14)} = 2{,}997$

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.

4.2.1.2 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 **pressuposições** devem ser atendidas:

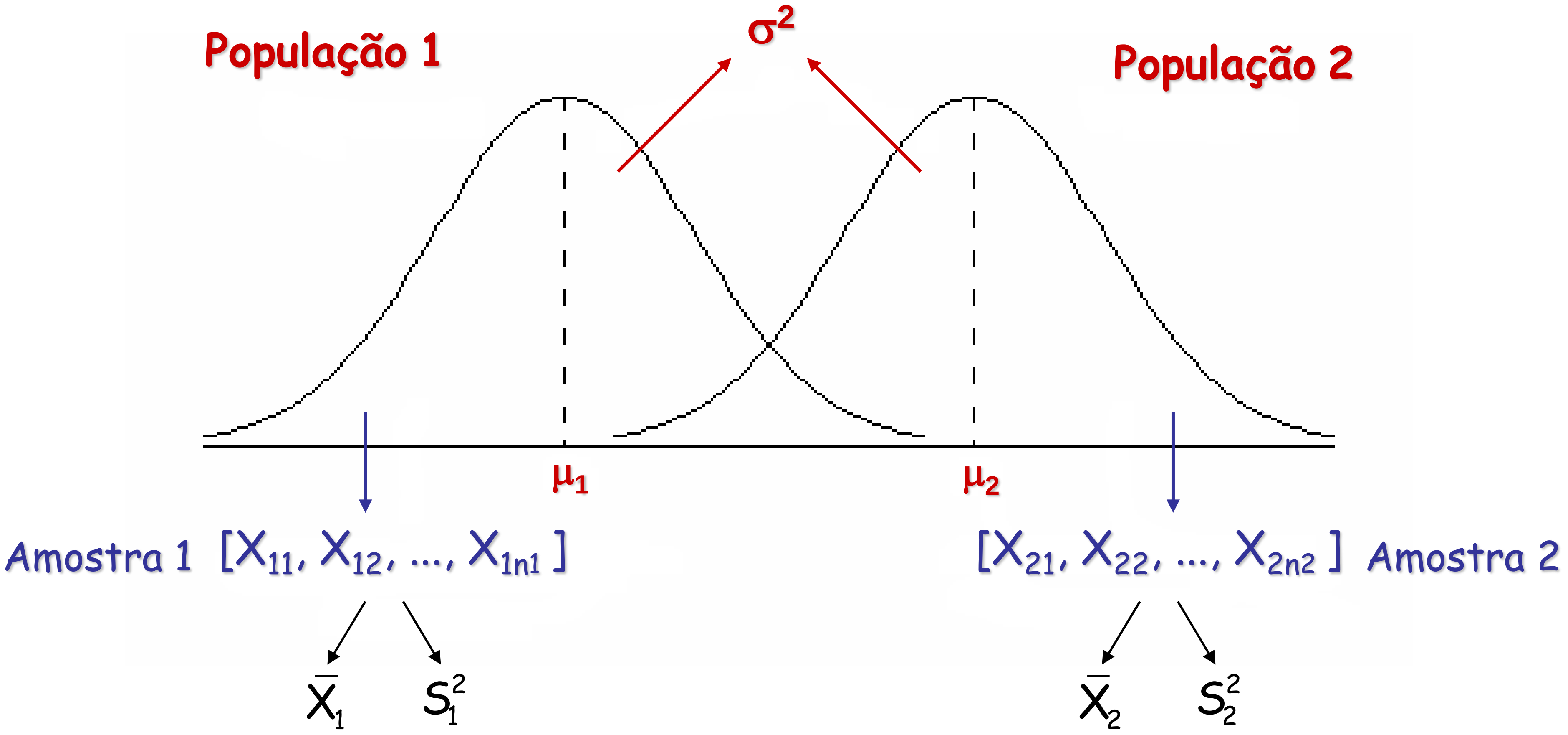
1. A variável em estudo tem distribuição normal:

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

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

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

Pressuposições



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:

$\theta = \mu_1 - \mu_2$ ← Parâmetro estimado

$\hat{\theta} = \overline{X}_1 - \overline{X}_2$ ← Estimador do parâmetro

$S(\hat{\theta}) = S\left(\overline{X}_1 - \overline{X}_2\right)$ ← 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 $\hat{\theta}$?

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

$$s(\bar{X}_1 - \bar{X}_2) = \sqrt{s^2(\bar{X}_1 - \bar{X}_2)}$$

$$= \sqrt{s^2(\bar{X}_1) + s^2(\bar{X}_2)}$$

$$= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

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

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

Variância da média

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

Estimador da variância da média

$$s^2(\bar{X}) = \frac{s^2}{n}$$

Variância combinada das amostras

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

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

Estimador do erro padrão do $\hat{\theta}$

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

Variância combinada

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

Grau de liberdade combinado

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

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

Exercícios propostos:

1. 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)

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 pontuais:

Linguagem A

$\bar{x}_1 = 17,5$

$s_1^2 = 2,31$

$n_1 = 15$

Linguagem B

$\bar{x}_2 = 20,5$

$s_2^2 = 3,02$

$n_2 = 15$

$\alpha=0,05$

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

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

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

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.

2. Na fabricação de semicondutores o ataque químico por via úmida é frequentemente 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 pontuais:

Solução 1

$\bar{x}_1 = 9,97$

$s_1^2 = 0,178$

$n_1 = 10$

Solução 2

$\bar{x}_2 = 10,42$

$s_2^2 = 0,054$

$n_2 = 9$

$\alpha = 0,05$

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

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

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

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 $-0,115\times10^{-3}$ polegadas/min é de 95%. 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

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

- Para a diferença entre duas médias

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

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