Intervalos de Confiança para a diferença de médias de dois grupos independentes

Bruno Martins
Lucas Gomes
Rafael Fratoni

Comparação entre dois grupos

*A comparação de grupos é fundamental em vários estudos científicos. (Ex: teste de drogas, protocolos, hipóteses etc.)

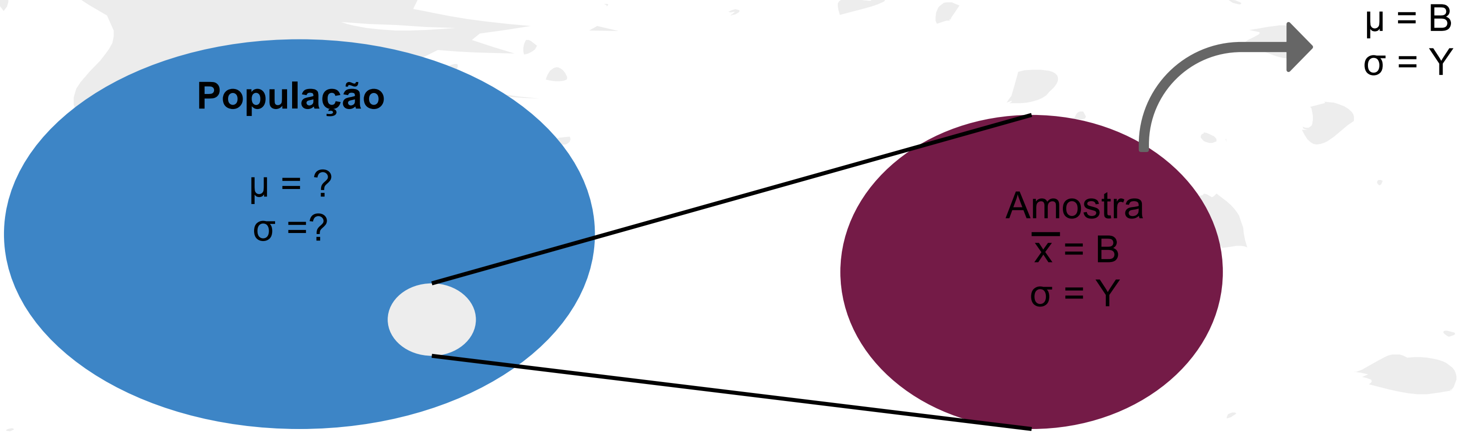
*As respostas individuais dos organismos dificultam os estudos, portanto devemos analisar as diferenças entre as médias dos grupos.

Como comparar dois grupos?

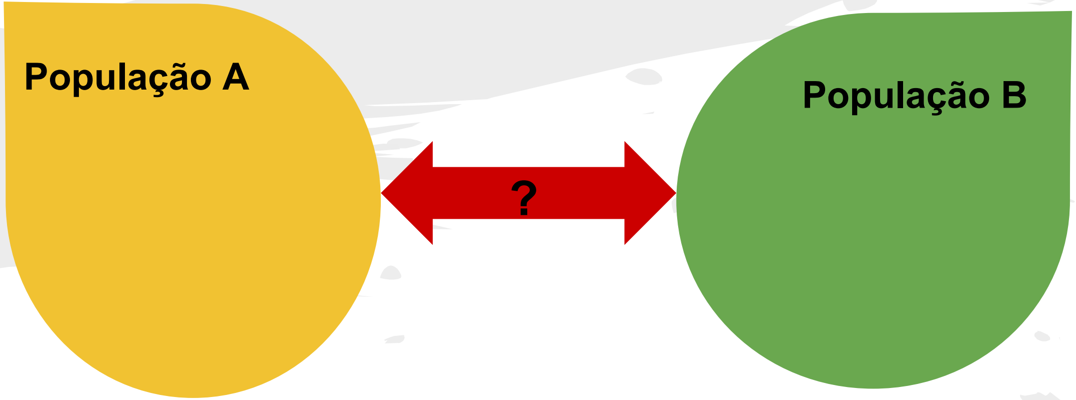
- Primeiro devemos lembrar:

*Estimamos uma média populacional (μ) à partir de uma amostra aleatória, com intervalo de confiança:

I.C. = $\bar{x} \pm t.SE$

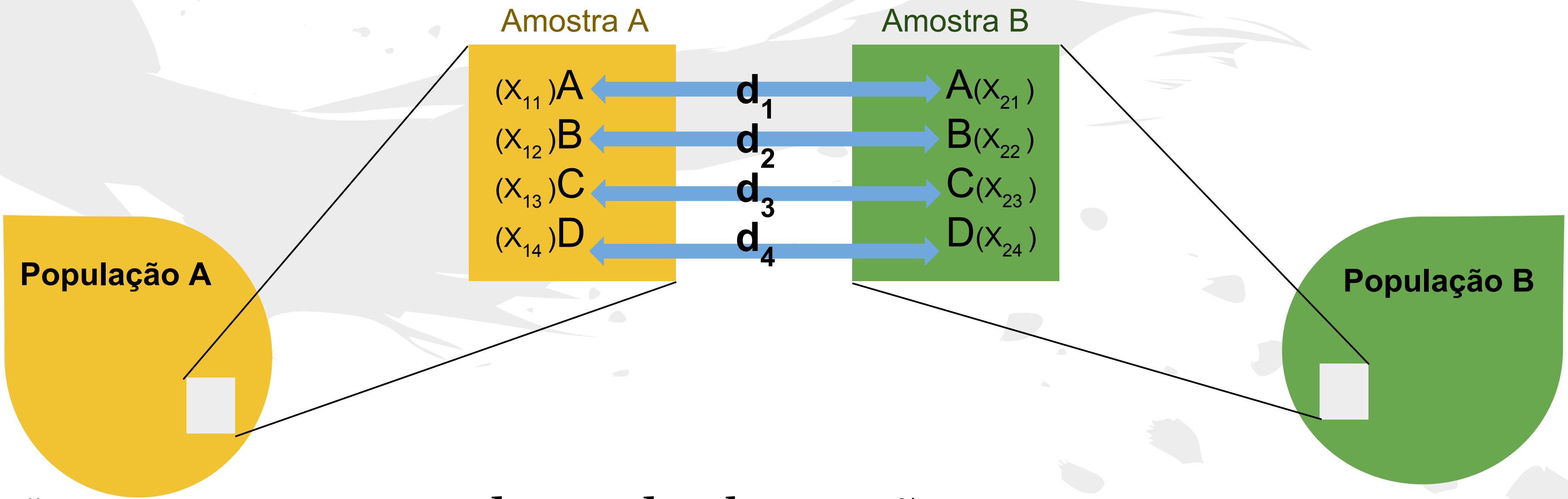


*Mas agora, a intensão é comparar dois grupos. Logo, devemos comparar a diferença entre as médias das duas populações.



*Para isso, temos dois caminhos: utilizar amostras pareadas (dependentes) ou amostras não-pareadas (independentes)

Amostras dependentes (Pareadas)



*Em amostras pareadas, cada observação em uma amostra, tem um equivalente na outra (Ex: observações de antes/depois, variáveis semelhantes)

***Exemplo de amostras dependentes:** para saber se a Aspirina A é mais eficiente do que a Aspirina B, medimos suas concentrações (em mg/ml) na urina de voluntários, após uma hora da ingestão do fármaco; com um intervalo de uma semana entre a ingestão de cada aspirina. Assim, temos a tabela:

Indivíduo	**mg/ml de Aspirina A**	**mg/ml de Aspirina B**	**Diferença**
1	15	13	2
2	26	20	6
3	17	17	0
4	7	5	2

*Com o pareamento das amostras A e B, temos um novo conjunto, formado pelas diferenças entre as amostras. Agora, a partir desses valores podemos realizar os cálculos de análise: $\bar{d}$, S, SE, I.C., teste de hipótese

Amostras independentes

*Não podem ser pareadas.

População 1
μ_1 = ?
σ_1 =?

Amostra
$\bar{x}_1$ = B
s_1 = Y
n_1= R

Amostra
$\bar{x}_2$ = B
s_2 = Y
n_1= R

População 2
μ_2 = ?
σ_2 =?

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2$$

Dois casos:

$\sigma_2 / \sigma_1 < 3$	$\sigma_2 / \sigma_1 > 3$
$\sigma_2 = \sigma_1$	$\sigma_2 \neq \sigma_1$

I. C. assumindo desvios padrão iguais

*Razão entre os desvios menor que 2 ou 3;

O I. C. entre duas médias é dado por:

$$\left((\bar{x}_1 - \bar{x}_2) - t \times \mathrm{SE}, \quad (\bar{x}_1 - \bar{x}_2) + t \times \mathrm{SE} \right)$$

Tabela de distribuição t: $n_1 + n_2 - 2$ = G. L.

I. C. assumindo desvios padrão iguais

Variância combinada: $$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

Desvio padrão: $S_p = \sqrt{s^2_p}$

Erro padrão das diferenças nas médias:

$$\mathrm{SE} = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

I. C. assumindo desvios padrão iguais

Exemplo: Intervalo de Confiança de 95% para os bicos de tentilhões de duas regiões distintas.

- Na região 1 foram coletados 22 espécimes, com média de 120 mm, variância 400 e desvio padrão de 20 mm.

- Na região 2 foram coletados 20 espécimes, com média de 100 mm, variância 225 e desvio padrão de 15 mm.



I. C. assumindo desvios padrão iguais

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

$$S_p^2 = \frac{21 \times 400 + 19 \times 225}{40} = 316.87$$

Lembre-se: por ser uma média ponderada, esse valor deve estar entre os valores de **variância** das populações. Neste caso, entre 225 e 400.

$S_p = \sqrt{S_p^2} = \sqrt{316.87} = 17.80$

$$\mathrm{SE} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

$\mathrm{SE} = 17.80\sqrt{0.095} = 5.52$

G.L = n1+n2-2 = 22 + 20 - 2 = 40

Tabela de distribuição t:

G.L	$p<0.05$
40	**2.021**

I. C. assumindo desvios padrão iguais

$$\left((\bar{x}_1 - \bar{x}_2) - t \times \mathrm{SE}, \quad (\bar{x}_1 - \bar{x}_2) + t \times \mathrm{SE}\right)$$

((120-100) - 2.021 x 5.52 , (120-100) + 2.021 x 5.52)

(8.85 , 31.15)

95% de confiança que, em média, as aves da população 1 tem entre 8.85 mm e 31.15 mm de comprimento de bico **a mais** do que a população 2.

I.C. para desvios padrão diferentes

Quando a razão do maior desvio pelo menor é maior que 2 ou 3 (depende do seu critério)

Não seguem distribuição t-student

Comparação de grandes amostras (ambas maiores que 30) é possível, pois a distribuição aproxima-se da Normal.

I.C. para desvios padrão diferentes

- Intervalo:

$$[(\bar{x}_1 - \bar{x}_2) - t \times SE \,,\, (\bar{x}_1 - \bar{x}_2) + t \times SE]$$

		p				
		0.20	0.10	0.05	0.01	0.001
	90	1.291	1.662	1.987	2.632	3.402
	100	1.290	1.660	1.984	2.626	3.390
	∞	1.282	1.645	1.960	2.576	3.291

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

I.C. para desvios padrão diferentes

Exemplo:

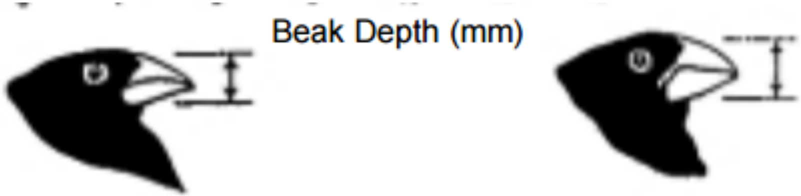
Populações de tentilhões em ilhas diferentes de Galápagos.

População 1:
$n = 48$
$\bar{x} = 9.5$ mm
$s^2 = 0.49$ mm
$s = 0.7$

População 2:
$n = 55$
$\bar{x} = 8.7$ mm
$s^2 = 0.04$ mm
$s = 0.2$

$s1/s2 = 0.7/0.2$

$s1/s2 = 3.5$



I.C. para desvios padrão diferentes

População 1:
n = 48
$\bar{x}$ = 9.5 mm
s^2 = 0.49 mm
s = 0.7

População 2:
n = 55
$\bar{x}$ = 8.7 mm
s^2 = 0.04 mm
s = 0.2

<u>Intervalo para 95% de confiança:</u>

$$((\bar{x}_1 - \bar{x}_2) - t \times SE, (\bar{x}_1 - \bar{x}_2) + t \times SE)$$

$$(0.8 - 1.96 \times SE, 0.8 + 1.96 \times SE)$$

$$(0.8 - 1.96 \times 0.10, 0.8 + 1.96 \times 0.10)$$

$$(0.60, 0.99)$$

Para estes tentilhões, há 95% de confiança de que a espessura do bico da população 1 mede entre 0.6 e 1 mm **a mais** que a espessura de bico da população 2

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

$$SE = \sqrt{\frac{0.49}{48} + \frac{0.04}{55}}$$

$$SE = 0.10$$