Métodos Estatísticos – L.EIC

Semana 9

Aula 7

16 de maio de 2022

Métodos Estatísticos – L.EIC

Aula 7

Inferência Estatística

Intervalo de Confiança (IC) para μ

População NÃO Normal mas Amostra Grande

IC para a Diferença de Médias

Amostras Independentes (AI)

Populações Normais: Variância Conhecida / Desconhecida

Populações NÃO Normais: Caso de Amostras Grandes

Amostras Emparelhadas (AE)

5

INFERÊNCIA ESTATÍSTICA

– Intervalos de Confiança –

Intervalo de Confiança para μ

População NÃO Normal

IC para μ – **População NÃO Normal**

Amostra de dimensão elevada

Se a população não tiver distribuição normal (e **variância desconhecida**) então só podemos construir um **intervalo de confiança** para a média da população, **se o tamanho da amostra for grande** ($n \geq 30$).

Este facto é justificado pelo Teorema do Limite Central.

IC para μ – **População NÃO Normal**

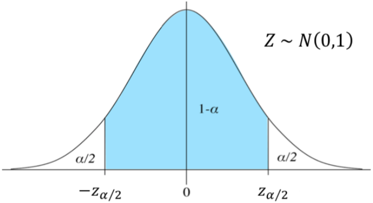
Amostra de dimensão elevada

O IC é dado neste caso por:

$$\bar{x} \pm z_{\alpha/2}\, se$$

onde *se* é o erro padrão da média,

$$se = \frac{s}{\sqrt{n}}$$



IC para μ – **População NÃO Normal**

Amostra de dimensão elevada

Deste modo, para uma amostra concreta com média μ e desvio padrão s,

$$(\overline{x} \pm 1,96 \cdot s/\sqrt{n})$$

é um IC, a aproximadamente 95%, para a média μ.

Notar que o valor de n, deve ser **suficientemente elevado** para que a aproximação anterior seja válida.

Se a **distribuição** da população é **quase simétrica** e com uma forma somente **moderadamente afastada da normal**, um tamanho moderado $n \geq 30$, **deverá ser suficiente**.

IC para μ – **População NÃO Normal – Exemplo**

Foi medida a concentração de cálcio (em mg/l) em 100 amostras aleatórias de água retirada de um lago. A **média e variância** amostrais obtidas, foram 0.65 e 0.12, respetivamente.

Construir um intervalo de confiança a 95% para a concentração média de cálcio nesse lago.

Sem informação da normalidade, mas amostra grande

intervalo de confiança: $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

$$\bar{x} = 0.65 \quad s = \sqrt{0.12} \approx 0.35 \quad n = 100 \quad \alpha = 0.05 \quad z_{\alpha/2} = 1.96$$

intervalo de confiança: 0.65 ± 0.07 (aprox. por excesso)

IC e **Dimensão da Amostra**

Exemplo:

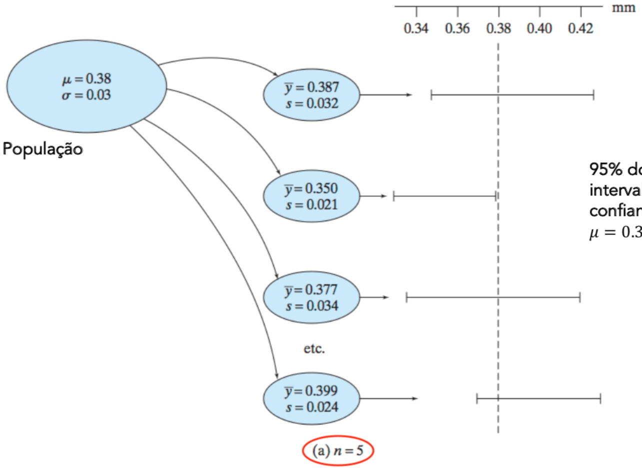
Suponha-se que, numa grande população de ovos, se pode considerar que a espessura da casca tem distribuição normal de média 0,38 mm e desvio padrão 0,03 mm.

A seguir estão representadas algumas amostras típicas desta população ($n = 5$ e $n = 20$), e os respetivos IC a 95% para a média.

Nota: Entre todos os intervalos possíveis de obter com amostras de dimensão n, 95% deles conteriam a média da população.

Inferência Estatística
Intervalo de Confiança

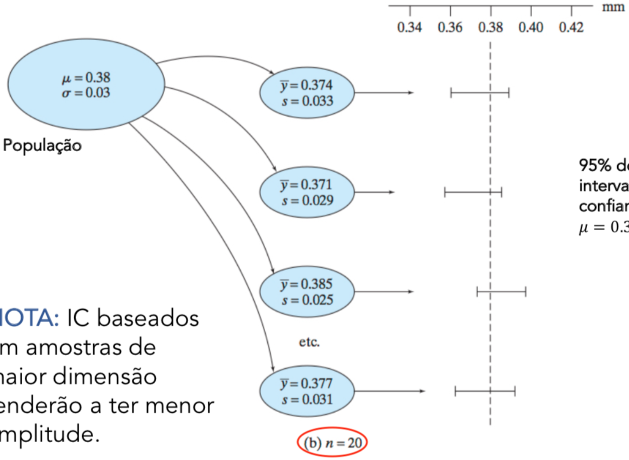
IC e **Dimensão da Amostra**



95% dos intervalos de confiança contêm $\mu = 0.38$

Inferência Estatística
Intervalo de Confiança

IC e **Dimensão da Amostra**



95% dos intervalos de confiança contêm $\mu = 0.38$

NOTA: IC baseados em amostras de maior dimensão tenderão a ter menor amplitude.

IC para μ – **Resumo**

Consideremos uma população com uma média μ e um desvio padrão σ, e uma amostra para a qual se tem:

$\overline{x}$ - média da amostra

s_x - desvio padrão da amostra

n - tamanho da amostra

Se queremos construir um intervalo de confiança para a média μ, com um grau de confiança $1-\alpha$, os casos possíveis estão resumidos a seguir.

IC para μ – **Resumo**

População Normal

- variância $\left(\sigma_x^2\right)$ conhecida

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \longrightarrow N(0,1)$$

- variância $\left(\sigma_x^2\right)$ desconhecida

$$\overline{x} \pm t_{\alpha/2,n-1} se \longrightarrow t_{(n-1)}; \quad \left(se = \frac{s_x}{\sqrt{n}}\right)$$

População NÃO Normal $(n \geq 30)$

$$\overline{x} \pm z_{\alpha/2}\frac{s_x}{\sqrt{n}} \longrightarrow N(0,1)$$

$$P\left(Z > z_{\alpha/2}\right) = \alpha/2 \quad Z \sim N(0,1)$$
$$P\left(T > t_{\alpha/2,n-1}\right) = \alpha/2 \quad T \sim t(n-1)$$

Intervalos de Confiança para a Diferença de Médias

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Duas Populações**

Até aqui considerámos a análise de um único parâmetro populacional.

Em muitas situações práticas, a análise estatística planeada, envolve a comparação de duas amostras de duas populações distintas.

Vamos então considerar a comparação das médias de duas amostras de duas populações distintas:

População 1 – μ_x

População 2 – μ_y

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Duas Populações**

População 1 – v.a. X

$E(X) = \mu_x; V(X) = \sigma_x^2$
$(X_1, \ldots, X_{n_x})$ – amostra aleatória (v.a. i.i.d.)
Média amostral – $\overline{X}$
Desvio padrão amostral – S_X

População 2 – v.a. Y

$E(Y) = \mu_y \quad ; V(Y) = \sigma_y^2$
$(Y_1, \ldots, Y_{n_y})$ – amostra aleatória (v.a. i.i.d.)
Média amostral – $\overline{Y}$
Desvio padrão amostral – S_Y

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Duas Populações**

Vamos usar $\overline{X} - \overline{Y}$ para estimar $\mu_x - \mu_y$.

Recordar que:

$$E(\overline{X}) = \mu_x \text{ e } V(\overline{X}) = \frac{\sigma_x^2}{n_x}$$

$$E(\overline{Y}) = \mu_y \text{ e } V(\overline{Y}) = \frac{\sigma_y^2}{n_y}$$

Assim,

$$E(\overline{X} - \overline{Y}) = \mu_x - \mu_y$$

e

$$V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

Inferência Estatística
Intervalo de Confiança
IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Duas Populações**

No caso dos desvios padrão das populações, σ_x e σ_y, serem conhecidos, podemos calcular o desvio padrão de $\overline{X} - \overline{Y}$, da forma seguinte:

$$\sigma_{(\overline{X}-\overline{Y})} = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

Na situação mais comum de não se conhecerem os valores dos desvios padrão, podemos usar como estimativa os desvios padrão amostrais s_x e s_y, e definir o erro padrão de $\overline{X} - \overline{Y}$:

$$se_{(\overline{X}-\overline{Y})} = \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

Inferência Estatística
Intervalo de Confiança

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Duas Populações**

Exemplo:

Foi realizado um estudo para comparar um procedimento novo, na cirurgia às amígdalas, em crianças com necessidade de as remover.

Uma das variáveis em estudo era **avaliar o 'nível' de dor**, numa escala de 0 a 10, quatro dias após a cirurgia.

Na tabela seguinte estão resumidos os resultados para os dois grupos do estudo: 49 crianças submetidas a um tipo de cirurgia, e 52 ao outro.

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Duas Populações**

Nível de Dor		
	Tipo de Cirurgia	
	Convencional	Inovadora
Média	4,3	1,9
Desvio Padrão	2,8	1,8
Tamanho	49	52

Como comparar as 2 cirurgias no que respeita à dor?

Comparando, por exemplo, as médias das variáveis.

Um dos modos de comparação das médias μ_x e μ_y das duas populações, consiste em estimar a sua diferença $\mu_x - \mu_y$.

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Duas Populações**

Uma estimativa pontual para $(\mu_x - \mu_y)$ é $(\bar{x} - \bar{y})$

- $\bar{x}$ – estimativa para μ_x obtida a partir de uma amostra aleatória de tamanho n_x e desvio padrão s_x de uma das populações
- $\bar{y}$ – estimativa para μ_y obtida a partir de uma amostra aleatória de tamanho n_y e desvio padrão s_y da outra população.

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Erro Padrão**

Nível de Dor		
	Tipo de Cirurgia	
	Convencional	Inovadora
Média	4,3	1,9
Desvio Padrão	2,8	1,8
Tamanho	49	52

Cálculo do erro padrão para cada amostra

- Cirurgia Convencional: $se_x \equiv s_{\overline{x}} = 2.8/\sqrt{49} \simeq 0.40$
- Cirurgia Inovadora: $se_y \equiv s_{\overline{y}} = 1.8/\sqrt{52} \simeq 0.25$

Inferência Estatística
Intervalo de Confiança

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Erro Padrão**

Nível de Dor		
	Tipo de Cirurgia	
	Convencional	Inovadora
Média	4,3	1,9
Desvio Padrão	2,8	1,8
Tamanho	49	52

μ_x: média do "Nível de dor com a cirurgia Convencional"

μ_y: média do "Nível de dor com a cirurgia Inovadora"

Então a estimativa de $\mu_x - \mu_y$ é: $\bar{x} - \bar{y} = 4.3 - 1.9 = 2.4$

e o erro padrão da estimativa é: $se = \sqrt{\frac{2.8^2}{49} + \frac{1.8^2}{52}} \approx 0.47$

Inferência Estatística
Intervalo de Confiança

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Erro Padrão**

O erro padrão de $\overline{X} - \overline{Y}$ está relacionado com o erro padrão individual de $\overline{X}$ e $\overline{Y}$.

Na verdade, note-se que:

$$se_{(\overline{X}-\overline{Y})} = \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} = \sqrt{se_x^2 + se_y^2}$$

onde

$$se_x \equiv s_{\overline{X}} = \frac{s_x}{\sqrt{n_x}}$$

e

$$se_y \equiv s_{\overline{Y}} = \frac{s_y}{\sqrt{n_y}}$$

Inferência Estatística
Intervalo de Confiança

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Erro Padrão**

O erro padrão para a diferença das duas médias amostrais, é dado por:

$$se_{(\overline{X}-\overline{Y})} = \sqrt{se_x^2 + se_y^2}$$

Como vimos, no caso anterior

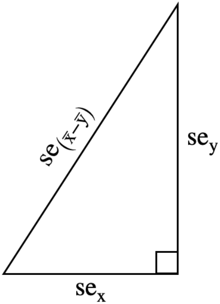
$$se_{(\overline{X}-\overline{Y})} = \sqrt{0,40^2 + 0,25^2} \approx 0,47$$

É fácil verificar que o erro padrão da diferença $\overline{X} - \overline{Y}$, dado por $se_{(\overline{x}-\overline{y})}$, é maior do que qualquer erro padrão individual (se_x e se_y), mas menor do que a sua soma.

IC – Diferença de Médias $(\mu_x - \mu_y)$ – **Erro Padrão**

Na verdade, o erro padrão da diferença das médias é dado por uma relação pitagórica:

$$se^2_{(\overline{X}-\overline{Y})} = se^2_x + se^2_y$$



IC para $(\mu_x - \mu_y)$ – **Erro Padrão Ponderado**

Em certas situações, usa-se **uma estimativa diferente para a variância**.

Se $\sigma_x^2 = \sigma_y^2 = \sigma^2$, a variância comum das populações pode ser estimada através de S_p^2, que é a **média ponderada das variâncias amostrais** S_x^2 e S_y^2:

$$S_p^2 = \frac{(n_x - 1)\, S_x^2 + (n_y - 1)\, S_y^2}{n_x + n_y - 2}$$

Inferência Estatística
Intervalo de Confiança

IC para $(\mu_x - \mu_y)$ – **Erro Padrão Ponderado**

$$S_p^2 = \frac{(n_x - 1)\, S_x^2 + (n_y - 1)\, S_y^2}{n_x + n_y - 2}$$

Usando agora S_p^2, obtém-se uma outra estimativa do desvio padrão de $\overline{X} - \overline{Y}$, designada por **erro padrão ponderado** para a diferença de médias SE_p,

$$SE_p = \sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y}\right)} = \sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}}$$

IC para $(\mu_x - \mu_y)$ – **Erro Padrão Ponderado**

Como S_p^2 é uma média ponderada das variâncias amostrais S_x^2 e S_y^2, o **erro padrão ponderado** SE_p só deve ser utilizado no caso em que é razoável admitir que $\sigma_x = \sigma_y$.

Neste caso os valores obtidos através dos dois métodos não são muito diferentes.

Nota: se as dimensões das amostras são iguais, ou se o desvio padrão das duas amostras são iguais, obtém-se o mesmo valor para o erro padrão de $\overline{X} - \overline{Y}$, com qualquer dos métodos.

IC para $(\mu_x - \mu_y)$ – **Erro Padrão de** $\overline{X} - \overline{Y}$

Temos então duas expressões para o erro padrão da diferença de médias $\overline{X} - \overline{Y}$:

- Admitindo $\sigma_x^2 \neq \sigma_y^2$

$$SE_{(\overline{X}-\overline{Y})} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}$$

- Admitindo $\sigma_x^2 = \sigma_y^2 = \sigma^2$

$$SE_{(\overline{X}-\overline{Y})} \equiv SE_p = \sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}}$$

Amostras Independentes (AI)

Populações Normais e Variâncias Conhecidas

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Conhecidas**

Vamos estimar $(\mu_X - \mu_Y)$ através de um intervalo de confiança, admitindo populações normais e amostras independentes.

- Construímos um IC para a diferença das médias de duas populações normais de variâncias conhecidas.
- Esta situação é uma 'extensão' do caso que já tratámos de um IC para a média de uma população normal com variância conhecida.

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Conhecidas**

Vamos considerar que as amostras aleatórias das populações normais X e Y são independentes, e têm dimensões, respetivamente, n_X e n_Y:

$$X \sim N\left(\mu_X, \sigma_X^2\right); \quad Y \sim N\left(\mu_Y, \sigma_Y^2\right)$$

Então

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)$$

e

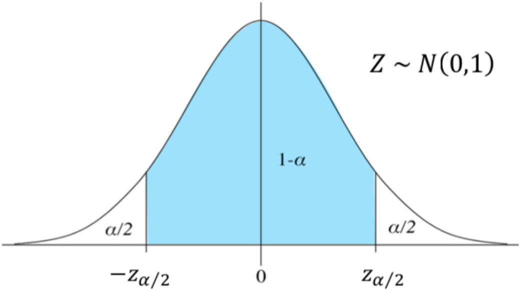
$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim N(0,1)$$

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Conhecidas**

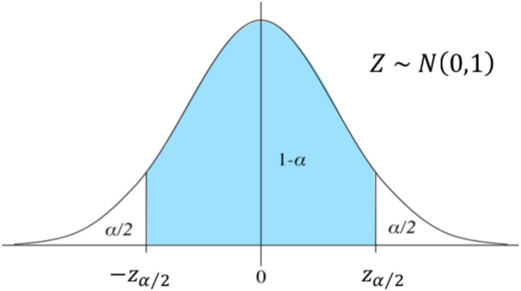
Nesta situação, com um **grau de confiança** $1 - \alpha$, temos:

$$P\left(\left| \overline{X} - \overline{Y} - (\mu_X - \mu_Y) \right| \leq z_{\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \right) \approx 1 - \alpha$$



Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Conhecidas**



O intervalo aleatório para $\mu_1 - \mu_2$, com grau de confiança $1-\alpha$, é dado por:

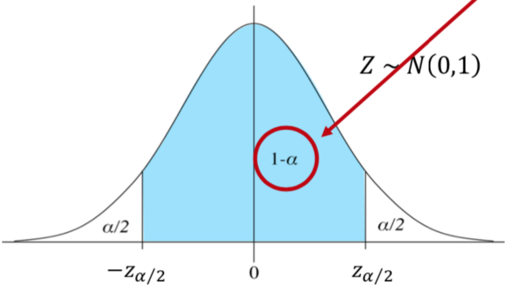
$$\overline{X} - \overline{Y} \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Conhecidas**

Caso $1 - \alpha = 0.95$ (IC a 95%):

$$Z \sim N(0,1) \Rightarrow P(-1.96 \leq Z \leq 1.96) = 0.95$$



Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Conhecidas**

Caso $1 - \alpha = 0.95$ (IC a 95%):

O IC a 95% para a diferença das médias de duas populações normais com desvios padrão conhecidos, baseado em amostras aleatórias independentes de dimensões n_X e n_Y, é da forma:

$$\overline{x} - \overline{y} \pm 1.96 \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

onde $\overline{x}$ e $\overline{y}$ são as médias amostrais observadas.

Amostras Independentes (AI)

Populações Normais e Variâncias Desconhecidas

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Usualmente não se conhecem os valores das variâncias das populações.

Neste caso o procedimento é também análogo ao já efetuado para o caso de um IC para a média de uma população normal com variância desconhecida.

O intervalo de confiança para $(\mu_X - \mu_Y)$, com grau de confiança $1 - \alpha$, é dado por:

$$\overline{x} - \overline{y} \pm t_{\alpha/2,gl} \cdot \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

O valor de $t_{\alpha/2,gl}$ é determinado através da distribuição t **de Student** com um número de graus de liberdade dado por gl,

$$gl = \frac{\left(se_X^2 + se_Y^2\right)^2}{\frac{se_X^4}{(n_X-1)} + \frac{se_Y^4}{(n_Y-1)}}$$

Contudo, para facilitar os cálculos, podemos utilizar a expressão aproximada

$$gl = n_X + n_Y - 2$$

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 1

Num estudo do crescimento das plantas Brassica campestris, 7 foram tratadas com a substância Ancymidol e comparadas com um grupo de controlo constituído por 8 plantas que receberam somente água.

A altura de todas as plantas foi medida depois de 14 dias de crescimento.

Os valores registados estão resumidos na tabela seguinte

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 1

	Grupo de controlo (1)	Grupo de Tratamento 'Ancy' (2)
	10,0	13,2
	13,2	19,5
	19,8	11,0
	19,3	5,8
	21,2	12,8
	13,9	7,1
	20,3	7,7
	9,6	
n	8	7
$\bar{y}$	15,9	11,0
s	4,8	4,7
SE	1,7	1,8

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 1

Neste caso, com 95% de confiança, temos:

$$15.9 - 11.0 \pm t_{0.025} \cdot \sqrt{\frac{4.8^2}{8} + \frac{4.7^2}{7}}$$

O valor de $t_{0.025}$ é obtido pela distribuição t de Student com os seguintes graus de liberdade

$$gl = \frac{\left(1.7^2 + 1.8^2\right)^2}{\frac{1.7^4}{7} + \frac{1.8^4}{6}} = 12.8$$

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 1

$$15.9 - 11.0 \pm t_{0.025} \cdot \sqrt{\frac{4.8^2}{8} + \frac{4.7^2}{7}}$$

No **R**, o valor de $t_{0.025}$ é dado por:

```
t = qt(0.975, df=12.8)
[1] 2.163805 (≈ 2.164)
```

Substituindo e efetuando os cálculos, temos então o IC a 95% para $(\mu_X - \mu_Y)$, dado por:

$$IC_{(95\%)} =]-0.4; 10.2[$$

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 2

Os biólogos teorizaram que as borboletas Monarca macho, tinham em média um tórax maior que as fêmeas.

Numa amostra de 7 machos e 8 fêmeas Monarca, foram obtidos os dados registados na tabela que se mostra a seguir.

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 2

	Machos	Fêmeas
	67	73
	73	54
	85	61
	84	63
	78	66
	63	57
	80	75
		58
n	7	8
$\bar{y}$	75,7	63,4
s	8,4	7,5
SE	3,2	2,7

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 2

Com um **grau de confiança de** 95%, o IC é:

$$\bar{x} - \bar{y} \pm t_{0.025} \cdot \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

$$\Leftrightarrow 75.7 - 63.4 \pm t_{0.025} \cdot \sqrt{\frac{8.4^2}{7} + \frac{7.5^2}{8}}$$

e os graus de liberdade são

$$gl = \frac{\left(3.2^2 + 2.7^2\right)^2}{\frac{3.2^4}{6} + \frac{2.7^4}{7}} = 12,3$$

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. Normais – **Variâncias Desconhecidas**

Exemplo 2

$$75.7 - 63.4 \pm t_{0.025} \cdot \sqrt{\frac{8.4^2}{7} + \frac{7.5^2}{8}}$$

No **R**, o valor de $t_{0.025}$ é dado por:

```
t = qt(0.975, df=12.3)
[1] 2.172933 (≈ 2,173)
```

Substituindo, temos:

$$IC_{(95\%)} =]3.3; 21.3[$$

Temos 95% de confiança em que o peso médio do tórax da população de machos é superior ao das fêmeas, por uma quantidade entre 3.3 e 21.3 mg

Amostras Independentes (AI)

Populações NÃO Normais mas Amostras Grandes

Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. NÃO Normais – **Amostras Grandes**

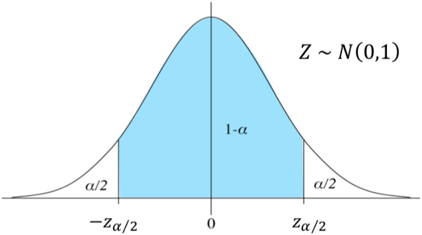
A hipótese de normalidade das populações não é muito importante quando as amostras são grandes.

No caso de **populações não normais** mas **amostras grandes**, o intervalo de confiança para a diferença $(\mu_X - \mu_Y)$ com grau de confiança (aproximado) $1-\alpha$, é dado por:

$$\overline{x} - \overline{y} \pm z_{\alpha/2} se$$

sendo

$$se = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$



Inferência Estatística
Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. NÃO Normais – **Amostras Grandes**

Exemplo

Nível de Dor		
	Tipo de Cirurgia	
	Convencional	Inovadora
Média - $\bar{y}_1$,$\bar{y}_2$	4,3	1,9
Desvio Padrão - s_1, s_2	2,8	1,8
Dimensão - n_1, n_2	49	52

μ_X: média do "Nível de dor com a cirurgia Convencional"

μ_Y: média do "Nível de dor com a cirurgia Inovadora"

Então a estimativa de $\mu_X - \mu_Y$ é: $\bar{x} - \bar{y} = 4.3 - 1.9 = 2.4$

e o erro padrão da estimativa é: $se = \sqrt{\frac{2.8^2}{49} + \frac{1.8^2}{52}} \approx 0.47$

Inferência Estatística

Intervalo de Confiança

$(\mu_X - \mu_Y)$ – Pop. NÃO Normais – **Amostras Grandes**

Exemplo

Nível de Dor		
	Tipo de Cirurgia	
	Convencional	Inovadora
Média - $\bar{y}_1, \bar{y}_2$	4,3	1,9
Desvio Padrão - s_1, s_2	2,8	1,8
Dimensão - n_1, n_2	49	52

Não se conhecem as distribuições populacionais, mas as **amostras são grandes**. O intervalo de confiança a 95% para $(\mu_X - \mu_Y)$, é:

$$\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

$$IC_{(95\%_0)} \equiv 2.4 \pm 1.96 \times 0.47 \approx]1.47, 3.33[$$

$(\mu_X - \mu_Y)$ – Amostras Independentes – **Resumo**

Populações Normais

- variâncias conhecidas

$$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \quad \longleftarrow \quad N(0,1)$$

- variâncias desconhecidas

$$\bar{x} - \bar{y} \pm t_{\alpha/2, n_X+n_Y-2}\, se \quad \longleftarrow \quad t_{(n_X+n_Y-2)}$$

Populações NÃO Normais – Amostras Grandes $(n \geq 30)$

$$\bar{x} - \bar{y} \pm z_{\alpha/2}\, se \quad \longleftarrow \quad N(0,1)$$

com $se = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$

Amostras Emparelhadas (AE)

$(\mu_X - \mu_Y)$ – Amostras Emparelhadas

Os **intervalos de confiança** construídos **atrás** para $(\mu_X - \mu_Y)$ aplicam-se quando as **amostras são independentes**.

Esta independência significa que cada elemento de uma das amostras não está relacionado com qualquer outro elemento da outra amostra.

Contudo, existem situações em que cada uma das observações de **uma das amostras** está relacionada com uma observação na **outra amostra**, de tal forma que **os dados aparecem aos pares**.

$(\mu_X - \mu_Y)$ – Amostras Emparelhadas

Exemplo 1

(mesmos indivíduos em instantes diferentes)

Considerar, por exemplo, um tratamento durante o qual é feita uma amostragem em n indivíduos em dois instantes diferentes:

$(X_1, \ldots, X_n)$ níveis de glicose antes do tratamento

$(Y_1, \ldots, Y_n)$ níveis de glicose depois do tratamento

$(\mu_X - \mu_Y)$ – **Amostras Emparelhadas**

Exemplo 2

(mesmos indivíduos e características diferentes)

É feita uma amostragem em n indivíduos:

$(X_1, \ldots, X_n)$ comprimento do braço esquerdo

$(Y_1, \ldots, Y_n)$ comprimento do braço direito

Temos nestes casos, amostras aleatórias **emparelhadas**:

$$(X_1, Y_1)\,,\ (X_2, Y_2)\,,\ \ldots,\ (X_n, Y_n)$$

$(\mu_X - \mu_Y)$ – Amostras Emparelhadas

Supor que pretendemos construir um IC para a diferença das médias $\mu_X - \mu_Y$, de duas populações X e Y através de uma amostra emparelhada,

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$$

Façamos

$$D_i = X_i - Y_i$$

$(D_1, D_2, \ldots, D_n)$ é então uma amostra da população

$$D = X - Y$$

e

$$E(D) = E(X - Y) = E(X) - E(Y) = \mu_X - \mu_Y$$

$(\mu_X - \mu_Y)$ – **Amostras Emparelhadas**

Um intervalo de confiança para $\mu_X - \mu_Y$ é um intervalo de confiança para a média (ou valor esperado) de D.

Aplicam-se agora os resultados anteriores para determinar um intervalo de confiança para a média de uma só população.

Neste caso, a partir de um IC para a média das diferenças, obtemos um IC para a diferença das médias.

$(\mu_X - \mu_Y)$ – Amostras Emparelhadas

Exemplo

10 pessoas participaram num estudo sobre o efeito de uma dada dieta na redução do colesterol no sangue.

Os resultados foram os seguintes:

	Teor de colesterol		Diferença
Indivíduo	Antes da dieta (x_i)	Depois da dieta (y_i)	x_i-y_i
1	201	200	1
2	231	236	-5
3	221	216	5
4	260	233	27
5	228	224	4
6	237	216	21
7	326	296	30
8	235	195	40
9	240	207	33
10	267	247	20

$(\mu_X - \mu_Y)$ – Amostras Emparelhadas

Exemplo

Amostra emparelhada: (x_i, y_i)

Amostra simples: $d_i = x_i - y_i$

$$\overline{d} = 17.6; \quad s = 15.3782$$

A amostra é pequena.

Assumimos a normalidade da população (D).

IC a 95% (aproximado) com base na amostra:

$$\overline{d} \pm t_{0.025,9} \frac{s}{\sqrt{10}} = 17.6 \pm 11.1$$

Diferença
x_i-y_i
1
-5
5
27
4
21
30
40
33
20