# MÉTODOS ESTATÍSTICOS - L.EIC

EXERCÍCIOS – 2022/2023

## Folha 4

(Intervalos de Confiança para a Média e Diferença de Médias)

1. [Adaptado de (*)]

   Um farmacologista mediu a concentração cerebral de dopamina numa amostra de ratos. A concentração média obtida foi de $1\,269\,ng/g$ e o desvio padrão de $145\,ng/g$.

   Qual o erro padrão obtido para a média, considerando:

   (a) Uma amostra de 8 ratos?

   $$se = 51,26524$$

   (b) Uma amostra de 30 ratos?

   $$se = 26,47326$$

2. Considerar o exercício anterior e suponha que a variável em estudo é normalmente distribuída. Determinar para cada um dos casos (amostra de 8 e 30 ratos) um intervalo de confiança a 95% para a média. Comparar os dois intervalos e comentar o resultado.

   (a) Uma amostra de 8 ratos?

   $X \sim N\left(\mu;\sigma^2\right)$ ; $\overline{X} \sim N\left(\mu;\sigma^2/n\right)$

   População normal com $\sigma^2$ desconhecida $\Rightarrow$ $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ ; $n = 8$

   $$\mathbf{I.C.}_{(95\%)} = ]1147; 1391[$$

   (b) Uma amostra de 30 ratos?

   $X \sim N\left(\mu;\sigma^2\right)$ ; $\overline{X} \sim N\left(\mu;\sigma^2/n\right)$

   População normal com $\sigma^2$ desconhecida $\Rightarrow$ $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ ; $n = 30$

   $$\mathbf{I.C.}_{(95\%)} = ]1214; 1324[$$

3. [Adaptado de (*)]

   Supor que estamos a planear uma experiência para testar o efeito de uma dieta no aumento de peso de uma população de perus. Seja $Y$ a variável que representa o aumento de peso em 3 semanas relativo a essa dieta. Experiências anteriores sugerem que o desvio padrão de $Y$ é aproximadamente 80 g.

   Determinar quantos perus deverão constituir a amostra da experiência se se pretender que o erro padrão da média seja não superior a 20 g.

$$se_{\overline{Y}} = \frac{s_Y}{\sqrt{n}} \leq 20 \quad \Rightarrow \quad n \geq 16$$

4. Sabe-se, por experiência passada, que o desvio padrão relativo ao tempo necessário para a recuperação total das pessoas que sofrem de uma determinada doença, é de 15 dias. Depois de analisar uma amostra de 16 doentes obteve-se um tempo médio de 85 dias para recuperar um doente.

   Estimar, com um nível de confiança de 95%, o tempo médio (em dias), necessário para um doente que sofra da doença recuperar totalmente.

   $X \sim N\left(\mu; 15^2\right)$

   População normal com $\sigma^2$ conhecida $\quad \Rightarrow \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\text{I.C.}_{(95\%)} = ]77; 93[$$

5. [Adaptado de (*)]

   Numa experiência integrada num estudo sobre o desenvolvimento da glândula timo, os investigadores pesaram as glândulas de cinco embriões de frangos após 14 dias de incubação. Os pesos da *glândula timo* (em mg) foram os seguintes:

   29.6 21.5 28.0 34.6 44.9

   Para estes dados, a média é 31.7 mg, e o desvio padrão 8.7 mg.

   (a) Calcular o erro padrão da média.

   $se_{\overline{X}} = 3.890758$

   (b) Construir um intervalo de confiança a 90% para a média da população.

   Como a amostra é pequena... assumimos que a população é normal!

   $X \sim N\left(\mu; \sigma^2\right) \quad ; \quad \overline{X} \sim N\left(\mu; \sigma^2/n\right)$

   População normal com $\sigma^2$ desconhecida $\quad \Rightarrow \quad T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1) \quad ; \quad n = 5$

$$\text{I.C.}_{(90\%)} = ]23.4; 40.0[$$

6. Considerar os dados seguintes (exercício 4 da Folha 1), referentes a uma amostra das medições do perímetro cefálico (em cm) de 35 recém nascidos do sexo masculino.

| | | | | | | |
|---|---|---|---|---|---|---|
| 33.1 | 33.4 | 34.8 | 33.8 | 34.7 | 34.3 | 35.6 |
| 34.5 | 34.6 | 34.1 | 33.9 | 33.6 | 34.6 | 35.2 |
| 33.7 | 35.8 | 34.2 | 34.0 | 34.7 | 35.2 | 34.3 |
| 33.4 | 36.0 | 34.5 | 36.1 | 35.1 | 35.1 | 34.6 |
| 33.7 | 34.9 | 34.2 | 34.2 | 34.2 | 35.3 | 34.2 |

Dados auxiliares: $\left[\sum_{i=1}^{35} x_i = 1207.6 \quad \sum_{i=1}^{35} x_i^2 = 41684.32\right]$

Determinar um intervalo de confiança a 95% para o perímetro cefálico médio da população.

Usando o software R:

```
PC=read.csv2("DadosFolha1.csv")
PC <- PC$Ex4
PC<-PC[!is.na(PC)]
```

```
Dados<-read.csv("DadosFicha1.csv")
PC<-Dados$Ex4
PC<-PC[!is.na(PC)]
```

```
# 1o CASO:
# caso de não se assumir uma distribuição específica para a população

mean(PC)-qnorm(0.975,0,1)*sd(PC)/sqrt(length(PC)) # 34.25736
mean(PC)+qnorm(0.975,0,1)*sd(PC)/sqrt(length(PC)) # 34.74835

]34.25, 34.75[
```

```
# no 1o caso:
mean(PC)-qnorm(0.975,0,1)*sd(PC)/sqrt(length(PC))
```

```
## [1] 34.25736
```

```
mean(PC)+qnorm(0.975,0,1)*sd(PC)/sqrt(length(PC))
```

```
## [1] 34.74835
```

```
# 2o CASO:
# assumindo, após avaliação da normalidade, que a população é
# normalmente distribuída

mean(PC)-qt(0.975,length(PC)-1)*sd(PC)/sqrt(length(PC)) # 34.24831
mean(PC)+qt(0.975,length(PC)-1)*sd(PC)/sqrt(length(PC)) # 34.75741

# Alternativa para o 2o CASO:

t.test(PC,conf.level = 0.95) # 34.24831 34.75741

]34.24, 34.76[
```

```r
# no 2o caso
mean(PC)-qt(0.975,34)*sd(PC)/sqrt(length(PC))
```

```
## [1] 34.24831
```

```r
mean(PC)+qt(0.975,34)*sd(PC)/sqrt(length(PC))
```

```
## [1] 34.75741
```

7. O nível de ácido úrico $X$ (mg/dl) numa determinada população de homens adultos e saudáveis tem uma distribuição normal de valor esperado $\mu$ mg/dl e desvio padrão $\sigma = 1$ mg/dl. Recolheu-se uma amostra de dimensão 100, cujos níveis de ácido úrico apresentaram valor médio igual a 5.5 mg/dl.

   (a) Determinar um intervalo de confiança a 99% para $\mu$.

   $X \sim N(\mu, 1)$

   $$I.C._{(99\%)} = ]5.5 - \Delta, 5.5 + \Delta[ = 5.24, 5.76[$$

   (b) Admitir agora que se tem $X \sim N(5.5, 1)$. Calcular a dimensão da amostra de modo a que seja pelo menos igual a 0.9, a probabilidade da média da amostra se situar entre 5 e 6 mg/dl.

   $X \sim N(5.5, 1)$

   $$n \geq 11$$

8. A distribuição dos diâmetros dos caules nas plantas de determinada espécie tem distribuição normal. Uma amostra de 5 plantas apresentou os seguintes diâmetros (em mm):

   25.4  25.2  25.3  25.0  25.4

   Construir um intervalo de confiança a 98% para o diâmetro médio dos caules das plantas da espécie em causa.

$X \sim N\left(\mu; \sigma^2\right)$

$$\bar{x} = \frac{1}{5}\sum x_i = \mathbf{25.26}$$

$$s^2 = \frac{1}{4}\sum (x_i - \bar{x})^2 = \frac{1}{4}\left[(25.4 - 25.26)^2 + (25.2 - 25.26)^2 + (25.3 - 25.26)^2 + (25.0 - 25.26)^2 + (25.4 - 25.26)^2\right]$$

$$s^2 = \frac{1}{4}\sum (x_i - \bar{x})^2 = \frac{1}{4}[0.112] = 0.028$$

$$s = \mathbf{0.167332}$$

População normal com $\sigma^2$ desconhecida $\quad T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \quad ; \quad n = 5$

$$gl = n - 1 = \mathbf{4}$$

$$1 - \alpha = 0.98 \quad \Rightarrow \quad \alpha/2 = 0.01$$

$$t_{\alpha/2,gl} \equiv t_{0.01,4} = \mathbf{3.7469}$$

$$\bar{x} \pm t_{\alpha/2,gl} \times \frac{s}{\sqrt{n}} = 25.26 \pm 3.7469 \times \frac{0.167332}{\sqrt{5}}$$

$$\text{I.C.}_{(98\%)} = ]24.97961, 25.54039[$$

Este intervalo tem que estar contido no intervalo depois de arredondado, e então, com 2 casas decimais,

$$\text{I.C.}_{(98\%)} = ]24.97, 25.55[$$

```
D<-c(25.4,25.2,25.3,25.0,25.4)

mean(D)-qt(0.99,4)*sd(D)/sqrt(length(D))
```

```
## [1] 24.9796
```

```
mean(D)+qt(0.99,4)*sd(D)/sqrt(length(D))
```

```
## [1] 25.5404
```

9. Uma amostra aleatória retirada de uma população normal, produziu os intervalos de confiança seguintes, com graus de confiança distintos para a média dessa população:

$$]28.84, 31.10[ \qquad ]28.48, 31.46[$$

   (a) Qual o valor da média amostral?

   Os intervalos de confiança são centrados na média amostral. Assim, apenas temos que calcular o ponto médio dos intervalos. Nos dois casos obtemos o valor 29.97, que é portanto o valor da média amostral.

   (b) Qual dos dois intervalos corresponde a um menor grau de confiança?

   A amplitude do primeiro intervalo é menor do que a do segundo, e assim, é aquele que corresponde a um menor grau de confiança.

10. Numa amostra de doentes com uma determinada doença, verificou-se que o tempo médio de vida, após o diagnóstico, foi de 7 anos.

    Admitindo que o tempo de vida médio destes doentes segue uma distribuição normal de média $\mu$ e desvio padrão $\sigma = 1$ ano, indicar o tamanho da amostra para ter pelo menos 95% de confiança de que o erro de estimação seja inferior a 0.05.

$$n \geq 1536.64\ldots \quad (\text{n} \geq 1537)$$

11. Dois grupos de frangos, escolhidos aleatoriamente e de modo independente, foram submetidos a duas dietas diferentes. Após 2 semanas, observou-se o aumento de peso dos frangos, obtendo-se os seguintes valores (em gramas):

| | Dieta 1 | Dieta 2 |
|---|---|---|
| $\bar{x}$ | 165 | 180 |
| $s$ | 42 | 56 |

Determinar o erro padrão e o erro padrão ponderado de $\overline{X}_1 - \overline{X}_2$, considerando as dimensões das amostras tal como indicado, e comentar os resultados obtidos:

(a) $n_1 = 10$ e $n_2 = 15$

Erro padrão:

$$se = 19.63330504$$

Erro padrão ponderado:

$$se_p = 20.81317866$$

(b) $n_1 = 25$ e $n_2 = 25$.

Erro padrão:

$$se = 14$$

Erro padrão ponderado:

$$se_p = 14$$

12. Considerar os dados do exercício anterior e supor que o aumento de peso para cada dieta pode ser considerado como tendo uma distribuição normal.

Construir um intervalo de confiança a 95% para a diferença entre os aumentos de peso médios das populações correspondentes.

(a) $n_1 = 10$ e $n_2 = 15$

Não foi utilizado o erro padrão ponderado, pois não se assumiram variâncias iguais. Como as amostras são pequenas, assumimos populações normais.

```r
xb1=165
xb2=180
n1=10
n2=15
s1=42
s2=56
se1=sqrt(s1^2/n1)
se2=sqrt(s2^2/n2)
se=sqrt(se1^2 + se2^2)
gl=(se1^2+se2^2)^2/(se1^4/(n1-1)+se2^4/(n2-1))
gla=n1+n2-2
#
# IC com quantil calculado com gl não aproximado no R
IC = c(xb1-xb2 - qt(0.975,gl)*se, xb1-xb2 + qt(0.975,gl)*se)
IC
```

```
## [1] -55.65613  25.65613
```

```r
#
# IC com quantil calculado com gl aproximado (gla) no R
ICa = c(xb1-xb2 - qt(0.975,gla)*se, xb1-xb2 + qt(0.975,gla)*se)
ICa
```

```
## [1] -55.61459  25.61459
```

```r
#
# IC com quantil visto nas tabelas utilizando gl aproximado (gla)
ICt = c(xb1-xb2 - 2.0687*se, xb1-xb2 + 2.0687*se)
ICt
```

```
## [1] -55.61542  25.61542
```

Portanto, pelas tabelas, obtém-se:

$$\text{I.C.}_{(95\%)} = ]-55.62, 25.62[$$

(b) $n_1 = 25$ e $n_2 = 25$.

Assumimos populações normais, pois as amostras são pequenas.

```r
xb1=165
xb2=180
n1=25
n2=25
s1=42
s2=56
se1=sqrt(s1^2/n1)
se2=sqrt(s2^2/n2)
se=sqrt(se1^2 + se2^2)
gl=(se1^2+se2^2)^2/(se1^4/(n1-1)+se2^4/(n2-1))
gla=n1+n2-2
#
# IC com quantil calculado com gl não aproximado no R
IC = c(xb1-xb2 - qt(0.975,gl)*se, xb1-xb2 + qt(0.975,gl)*se)
IC
```

```
## [1] -43.20601  13.20601
```

```r
#
# IC com quantil calculado com gl aproximado (gla) no R
ICa = c(xb1-xb2 - qt(0.975,gla)*se, xb1-xb2 + qt(0.975,gla)*se)
ICa
```

```
## [1] -43.14889  13.14889
```

```r
#
# IC com quantil visto nas tabelas utilizando gl aproximado (gla)
ICt = c(xb1-xb2 - 2.0687*se, xb1-xb2 + 2.0687*se)
ICt
```

```
## [1] -43.9618  13.9618
```

Portanto, pelas tabelas, obtém-se:

$$\text{I.C.}_{(95\%)} = ]-43.97, 13.97[$$

13. Foram medidos os níveis de destruição dos pulmões em 9 indivíduos não fumadores e em 12 indivíduos fumadores, tendo-se obtido os resultados seguintes:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| fumadores ($x$) | 18.1 | 6.0 | 10.8 | 11.0 | 7.7 | 17.9 | 8.5 | 13.0 | 18.9 | | | |
| não fumadores ($y$) | 16.6 | 13.9 | 11.3 | 26.5 | 17.4 | 15.3 | 15.8 | 12.3 | 18.6 | 12.0 | 24.1 | 16.5 |

Construir um intervalo de confiança a 95% para a diferença das médias dos níveis de destruição dos pulmões nos dois grupos. O que se pode concluir?

Utilizando o **R**:

```r
F<-c(18.1,6.0,10.8,11.0,7.7,17.9,8.5,13.0,18.9)
NF<-c(16.6,13.9,11.3,26.5,17.4,15.3,15.8,12.3,18.6,12.0,24.1,16.5)

ep=sqrt(sd(F)^2/length(F)+sd(NF)^2/length(NF))
mean(F)-mean(NF)-qt(0.975,length(F)+length(NF)-2)*ep
```

```
## [1] -8.649664
```

```r
mean(F)-mean(NF)+qt(0.975,length(F)+length(NF)-2)*ep
```

```
## [1] 0.1329978
```

$$\text{I.C.}_{(95\%)} = ]-8.65, 0.14[$$

Como zero (0) pertence ao intervalo de confiança, não podemos concluir, com uma confiança de 95%, que os níveis médios de destruição dos pulmões sejam diferentes nas populações de fumadores e não fumadores.

14. Pretende-se testar se a altura média $\mu_1$, dos pais numa dada população difere significativamente da altura média $\mu_2$ dos respetivos filhos.

    Para tal, foi selecionada uma amostra de 12 pais e respetivos filhos adultos, tendo-se registado as seguintes alturas (em cm):

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| altura do pai ($x_i$) | 190 | 184 | 183 | 182 | 181 | 178 | 175 | 174 | 170 | 168 | 165 | 164 |
| altura do filho ($y_i$) | 189 | 186 | 180 | 179 | 187 | 182 | 183 | 171 | 170 | 178 | 174 | 165 |

    Dados auxiliares:

$$\sum_{i=1}^{12} x_i = 2114 \quad \sum_{i=1}^{12} y_i = 2144 \quad \sum_{i=1}^{12} x_i^2 = 373160 \quad \sum_{i=1}^{12} y_i^2 = 383666 \quad \sum_{i=1}^{12} (x_i - y_i)^2 = 330$$

    (a) Construir um intervalo de confiança a 95% para $\mu_1 - \mu_2$. O que se pode concluir?

    Amostras emparelhadas. Amostras pequenas.
    Assumimos normalidade na diferença de alturas entre pais e filhos.
    Utilizando o **R**:

$$\text{I.C.}_{(95\%)} = ]-5.56, 0.56[$$

    Uma vez que $0 \in$ I.C. não é possível concluir, com 95% de confiança, que a altura média dos pais seja significativamente diferente da altura média dos respetivos filhos.

    (b) Como se poderia obter a partir desta amostra um intervalo de confiança com o dobro da precisão? As conclusões seriam as mesmas?

    Para se ter o dobro da precisão, o intervalo de confiança teria que ter metade da amplitude, isto é, utilizando o **R**, seria:

$$\text{I.C.} = ]-4.03, -0.97[$$

    o que diminuiria o grau de confiança (para cerca de 70%).

Neste caso a conclusão seria diferente uma vez que este intervalo não contém 0. Agora poder-se-ia concluir, com uma confiança inferior a 95% (cerca de 70%), que a média das alturas dos filhos é superior à dos pais (I.C. negativo).

```r
Pais<-c(190,184,183,182,181,178,175,174,170,168,165,164)
Filhos<-c(189,186,180,179,187,182,183,171,170,178,174,165)

D=Pais-Filhos
ep=sqrt(sd(D)^2/length(D))

# (a)
mean(D)-qt(0.975,length(D)-1)*ep
```

```
## [1] -5.559146
```

```r
mean(D)+qt(0.975,length(D)-1)*ep
```

```
## [1] 0.5591462
```

```r
#(b)
mean(D)-0.5*qt(0.975,length(D)-1)*ep
```

```
## [1] -4.029573
```

```r
mean(D)+0.5*qt(0.975,length(D)-1)*ep
```

```
## [1] -0.9704269
```

(*) *Statistics for the Life Sciences*, Samuels, Witmer & Schaffner, PrenticeHall, (2012)