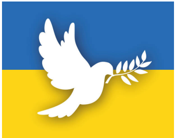
Estatística II

Aulas 16 & 17

Nelson Faustino [1]

[1] **Faculdade de Economia (FEUC)**
Universidade de Coimbra – Portugal

nelson@fe.uc.pt



Conteúdos

Intervalo aleatório e intervalo de confiança

[Murteira et al. (2015), pp. 428–429]

A partir de uma amostra casual $(X_1, X_2, \ldots, X_n)$ de uma v.a. contínua $X \sim f(x|\theta)$:

- **Estatísticas de Interesse:** Identificar $T_1 = T_1(X_1, X_2, \ldots, X_n)$ e $T_2 = T_2(X_1, X_2, \ldots, X_n)$ tais que $T_1 < T_2$.
- **Grau de Confiança:** Parâmetro

  $1-\alpha = (1-\alpha) \times 100\% \ (0 < \alpha < 1)$.
- **Intervalo Aleatório para θ:** $]T_1, T_2[$, com um grau de confiança $1-\alpha$, é determinado a partir da condição

$$P(T_1 < \theta < T_2) = 1-\alpha.$$

- **Intervalo de confiança para θ:** intervalo $]t_1, t_2[$ determinado a partir de estimativas para o intervalo aleatório de θ, i.e.
  - $t_1 = t_1(x_1, x_2, \ldots, x_n)$;
  - $t_2 = t_1(x_1, x_2, \ldots, x_n)$.

Intervalo aleatório e intervalo de confiança

[Murteira et al. (2015), pp. 428–429]

- **Determinar grau de confiança de $]T_1, T_2[$ (valor $1-\alpha$):** Envolve o cálculo da probabilidade

$$P(T_1 < \theta < T_2)$$

a partir de duas estatísticas conhecidas, T_1 e T_2, respectivamente.

- **Determinar o intervalo de confiança para θ:**
Para um valor $1-\alpha$ dado a priori (grau de confiança), o intervalo $]t_1, t_2[$ é tal que
  - $t_1 = t_1(x_1, x_2, \ldots, x_n)$;
  - $t_2 = t_1(x_1, x_2, \ldots, x_n)$.

são os quantis associados à condição $P(T_1 < \theta < T_2) = 1-\alpha$

Método da Variável Fulcral

[Murteira et al. (2015), p. 429 em diante]

- **O que é uma variável fulcral?** v.a. $Z = Z(X_1, X_2, \ldots, X_n, \theta)$ tal que $Z \sim f(x|\theta)$ é **independente do parâmetro** θ (não confundir com o conceito de estatística).
- **Para que serve?** Para determinar os quantis t_1, t_2 de $]t_1, t_2[$, com um **grau de confiança** de $(1-\alpha) \times 100\%$.
- **Como posso construir variáveis fulcrais?**
  - A partir de estatísticas associadas a **populações normais** $N(\mu, \sigma^2)$ ($\overline{X}, S^2$ e $(S')^2$).
  - A partir do **Teorema do Limite Central (TLC)**.
- **Qual é o intervalo de confiança de uma variável fulcral?**

$$]q_{\alpha/2}, q_{1-\alpha/2}[,$$

onde $q_{\alpha/2}$ e $q_{1-\alpha/2}$ são quantis de $Z \sim f(x|\theta)$.

Populações Normais

Estatísticas vs. Variáveis Fulcrais de Interesse

Populações Normais

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

Média Amostral ($\overline{X}$)

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

μ, σ^2 conhecidas
n conhecida/a estimar

Variância Amostral (S^2)

$\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

n conhecida
σ^2 conhecida/a estimar

Rácio de Student ($\overline{X}$ vs. $(S')^2$)

$\sqrt{n}\,\frac{\overline{X}-\mu}{S'} \sim t_{n-1}$

n conhecida
μ conhecida/a estimar

Variância Amostral Corrigida ($(S')^2$)

$\frac{(n-1)(S')^2}{\sigma^2} \sim \chi^2_{n-1}$

n conhecida
σ^2 conhecida/a estimar

Rácio de Student ($\overline{X}$ vs. S^2)

$\sqrt{n-1}\,\frac{\overline{X}-\mu}{S} \sim t_{n-1}$

n conhecida
μ conhecida/a estimar

Aproximações Assintóticas

Rácio de Student

$U \sim N(0,1)$

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

$V \sim \chi^2_{n-1}$

$$V = \frac{nS^2}{\sigma^2} \text{ ou } V = \frac{(n-1)(S')^2}{\sigma^2}.$$

Média vs. Variância da Amostra ($\overline{X}$ vs. S^2)

$$\frac{U}{\sqrt{\frac{V}{n-1}}} = \sqrt{n-1}\,\frac{\overline{X}-\mu}{S} \sim t_{n-1} \overset{n\to+\infty}{\Longrightarrow} \sqrt{n-1}\,\frac{\overline{X}-\mu}{S} \dot{\sim} N(0,1).$$

Média vs. Variância Corrigida da Amostra ($\overline{X}$ vs. $(S')^2$)

$$\frac{U}{\sqrt{\frac{V}{n-1}}} = \sqrt{n}\,\frac{\overline{X}-\mu}{S'} \sim t_{n-1} \overset{n\to+\infty}{\Longrightarrow} \sqrt{n}\,\frac{\overline{X}-\mu}{S'} \dot{\sim} N(0,1).$$

Teorema do Limite Central

Reformulação de [Murteira et al. (2015), Teorema 5.12]

Distribuição assintótica da média da amostra

$$\mu, \sigma^2 \longrightarrow n\overline{X} \dot{\sim} N(n\mu, n\sigma^2) \xrightarrow[\frac{\sigma^2}{n}]{\mu} \overline{X} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Obs:

- $E(\overline{X}) = \mu$ & $V(\overline{X}) = \dfrac{\sigma^2}{n}$;
- $\dfrac{\overline{X} - E(\overline{X})}{\sqrt{V(\overline{X})}} = \sqrt{n}\dfrac{\overline{X} - \mu}{\sigma}$.

Aulas 16 & 17

Como irão funcionar

Aula 16

Vamos aplicar estimação por intervalos a **populações normais** $N(\mu, \sigma^2)$.

Aula 17

Vamos recorrer ao **Teorema do Limite Central** para determinar assintoticamente intervalos de confiança.

Figura: Fita de Möbius por M. C. Escher (1964)

Conteúdos

Exercícios Resolvidos

Exercício 28., página 454

Enunciado do Exercício 28.

- Para a v.a. X, temos $X \sim N(\mu, (0.06)^2)$ – μ desconhecido;
- $n = 16$ é a dimensão da amostra de $X \sim N(\mu, (0.06)^2)$;
- $\overline{x} = 0.95$ é a média de litros da amostra.

Resolução Exercício 28. a)

Pretendemos determinar $P(X > 1)$ ($X \sim N(\mu, (0.06)^2)$). Não conhecendo o valor de μ – **média da população** – vamos recorrer à estimativa de máxima verossimilhança para μ **(estimativa $\overline{x}$)** para calcular uma aproximação para $P(X > 1)$, sendo

$$P(X > 1) = P\left(\frac{X-\mu}{0.06} > \frac{1-\mu}{0.06}\right) = 1 - \Phi\left(\frac{1-\mu}{0.06}\right).$$

Resolução Exercício 28. a) em *R*

```
> x<-0.95
> 1-pnorm((1-x)/0.06)
[1] 0.2023284
```

Resolução Exercício 28. b)

Pretende-se calcular uma estimativa de $P(0.911 < \overline{X} < 0.989)$ para avaliar a **percentagem de confiança**, $(1-\alpha) \times 100\%$, sendo

$$P(0.911 < \overline{X} < 0.989) = \Phi\left(\frac{0.989-\mu}{0.015}\right) - \Phi\left(\frac{0.911-\mu}{0.015}\right).$$

Resolução Exercício 28. b) em *R*

```
> x<-0.95
> pnorm((.989-x)/.015)-pnorm((.911-x)/.015)
[1] 0.9906776
```

Resolução Exercício 28. c) – Ideia

1. Tomamos $\overline{x} = 0.95$, $t_1 = 0.911$ e $t_2 = 0.989$.
2. Calcular a amplitude de $]t_1, t_2[$ $(t_2 - t_1 = 0.078)$.
3. Escolher $\varepsilon = \frac{t_2 - t_1}{2}$ (**metade da amplitude** de $]t_1, t_2[$) e considerar $]\overline{x} - \frac{\varepsilon}{2}, \overline{x} + \frac{\varepsilon}{2}[$ como novo intervalo de confiança de amplitude ε.

Solução: $]\overline{x} - \frac{\varepsilon}{2}, \overline{x} + \frac{\varepsilon}{2}[=]0.9305, 0.9695[$.

Formulação 31.

1. Apenas conheço a dimensão da amostra de $N(\mu, \sigma^2)$ $(n = 25)$.
2. Uso a variável fulcral $Z = \sqrt{n-1}\frac{\overline{X}-\mu}{S} \sim t_{n-1}$ para determinar um intervalo de confiança a 95% para μ em **31. a)**.
3. Uso a variável fulcral $Z = \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$ determinar um intervalo de confiança a 95% para σ em **31. b)**.

Cálculos Auxiliares 31.

- $1-\alpha = 0.95 \Longrightarrow \frac{\alpha}{2} = 0.025$ e $1-\frac{\alpha}{2} = 0.975$;
- $\overline{x} = \frac{\sum_{i=1}^{25} x_i}{25} = \frac{75}{25} = 3$;
- $s^2 = \frac{\sum_{i=1}^{25} x_i^2}{25} - \overline{x}^2 = \frac{321}{25} - 3^2 = 3.84$.

Resolução 31. a) – Simetria da t-Student $(1 - \frac{\alpha}{2} = 0.975)$

$$\overbrace{q_{\frac{\alpha}{2}}}^{-q_{1-\frac{\alpha}{2}}} < \sqrt{24}\frac{\overline{x}-\mu}{s} < q_{1-\frac{\alpha}{2}}$$

$$\Longleftrightarrow \overline{x} - \frac{s}{\sqrt{24}}q_{1-\frac{\alpha}{2}} < \mu < \overline{x} + \frac{s}{\sqrt{24}}q_{1-\frac{\alpha}{2}}$$

Resolução Exercício 31. a) em *R*

```
> x<-3
> s<-sqrt(3.84)
> x-s/sqrt(24)*qt(0.975,24)
[1] 2.174441
> x+s/sqrt(24)*qt(0.975,24)
[1] 3.825559
```

Exercícios Resolvidos

Exercício 31., página 455

Resolução 31. b) – Não tenho simetria para χ^2_{24}

$$q_{\frac{\alpha}{2}} < \frac{25s^2}{\sigma^2} < q_{1-\frac{\alpha}{2}} \iff \frac{1}{q_{1-\frac{\alpha}{2}}} < \frac{\sigma^2}{25s^2} < \frac{1}{q_{\frac{\alpha}{2}}}$$

$$\iff \frac{5s}{\sqrt{q_{1-\frac{\alpha}{2}}}} < \sigma < \frac{5s}{\sqrt{q_{\frac{\alpha}{2}}}}$$

Resolução Exercício 31. a) em *R*

```
> x<-3
> s<-sqrt(3.84)
> 5*s/sqrt(qchisq(0.975,24))
[1] 1.561657
> 5*s/sqrt(qchisq(0.025,24))
[1] 2.782304
```

Exercícios Resolvidos

Exercício 34., página 456

Formulação 34. (análogo ao Exercício 31.)

1. Vou usar a variável fulcral $Z = \sqrt{15}\frac{\overline{X}-\mu}{S'} \sim t_{14}$ para determinar um intervalo de confiança a 95% para μ em **31. a)**.
2. Uso o intervalo de confiança associado variável fulcral $Z = \frac{15S^2}{\sigma^2} \sim \chi^2_{14}$, $0.3082 < \sigma < 0.5851$, para determinar $1-\alpha$ em **31. b)**.

Cálculos Auxiliares 34.

- $\overline{x} = \dfrac{\sum_{i=1}^{15} x_i}{15} = \dfrac{30}{15} = 2;$
- $s^2 = \dfrac{\sum_{i=1}^{15} x_i^2}{15} - \overline{x}^2 = \dfrac{62.25}{15} - 2^2 = 0.15.$
- $(s')^2 = \dfrac{n}{n-1}s^2 = \dfrac{15}{14} \times 0.15 \approx 0.1607143$

Exercícios Resolvidos

Exercício 34., página 456

Resolução 34. a) – Simetria da t-Student $(1-\frac{\alpha}{2}=0.975)$

$$\overbrace{q_{\frac{\alpha}{2}}}^{-q_{1-\frac{\alpha}{2}}} < \sqrt{15}\frac{\overline{x}-\mu}{s'} < q_{1-\frac{\alpha}{2}}$$

$$\Longleftrightarrow \overline{x}-\frac{s'}{\sqrt{15}}q_{1-\frac{\alpha}{2}} < \mu < \overline{x}+\frac{s'}{\sqrt{15}}q_{1-\frac{\alpha}{2}}$$

Resolução Exercício 34. a) em *R*

```
> x<-2
> s1<-sqrt(15*0.15/14)
> x-s1/sqrt(15)*qt(0.975,14)
[1] 1.777993
> x+s1/sqrt(15)*qt(0.975,14)
[1] 2.222007
```

Exercícios Resolvidos

Exercício 34., página 456

Resolução 34. b) – vou usar $Z = \frac{15S^2}{\sigma^2} \sim \chi^2_{14}$

$$0.3082 < \sigma < 0.5851 \implies \frac{15s^2}{(0.5851)^2} < \overbrace{\frac{15S^2}{\sigma^2}}^{=Z} < \frac{15s^2}{(0.3082)^2}$$

$$\implies P\left(\frac{15s^2}{(0.5851)^2} < Z < \frac{15s^2}{(0.3082)^2}\right) = ?$$

Resolução Exercício 34. b) em *R*

```
> p2<-pchisq(15*0.15/(0.3082)^2,14)
> p1<-pchisq(15*0.15/(0.5851)^2,14)
> p2-p1
[1] 0.8999788
```

Teorema do Limite Central

Amostras casuais envolvendo distribuições discretas

$\overline{X} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$

- **Bernoulli** $X \sim B(1,\theta)$
  - $n\overline{X} \sim B(n,\theta)$
  - $\mu = \theta$
  - $\sigma^2 = \theta(1-\theta)$
- **Geométrica** $X \sim BN(1,\theta)$
  - $n\overline{X} \sim BN(n,\theta)$
  - $\mu = \frac{1}{\theta}$
  - $\sigma^2 = \frac{1-\theta}{\theta^2}$
- **Poisson** $X \sim Po(\lambda)$
  - $n\overline{X} \sim Po(n\lambda)$
  - $\mu = \lambda$
  - $\sigma^2 = \lambda^2$

Formulação 41. (envolve provas de Bernoulli)

População $X \sim B(1,\theta)$, sendo:

- $X = \begin{cases} 1, & \text{utiliza metropolitano} \\ 0, & \text{não utiliza metropolitano} \end{cases}$
- $\overline{x} = \frac{50}{200} = 0.25$ dá-nos a estimativa para o **número médio de pessoas** que usa o metro, numa amostra de dimensão $n = 200$.

Aplicação do Teorema do Limite Central

$$\overline{X} \dot{\sim} N\left(\theta, \frac{\theta(1-\theta)}{n}\right) \Longrightarrow q_{\frac{\alpha}{2}} < \frac{\overline{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} < q_{1-\frac{\alpha}{2}}$$

Problema: Determinação do intervalo de confiança a partir da variável fulcral $Z = \frac{\overline{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$ envolve cálculos demasiado tediosos **(resolução de inequações de grau 2)**.

Exercícios Resolvidos

Exercício 41., página 458

Resolução 41. a) – Nova variável fulcral

$$\frac{\overline{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \dot{\sim} N(0,1) \overset{\text{estimador } V(X)}{\Longrightarrow} \frac{\overline{X}-\theta}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}} \dot{\sim} N(0,1)$$

Algumas Observações:

1. Como $\widehat{\theta}=\overline{X}$ é um **estimador de máxima verossimilhança (m.v.)** para a média de $X \sim B(1,\theta)$ $(E(X)=\theta)$, a propriedade de invariância permite-nos concluir que $\overline{X}(1-\overline{X})$ é um **estimador de m.v.** para $V(X)=\theta(1-\theta)$.

2. Caso no enunciado nos fosse indicado, para além de $\overline{x}$, **dados adicionais** que nos permitissem calcular s^2 ou $(s')^2$, poderíamos utilizar uma das seguintes variáveis fulcrais:

$$\sqrt{n-1}\,\frac{\overline{X}-\mu}{S} \dot{\sim} N(0,1) \text{ ou } \sqrt{n}\,\frac{\overline{X}-\mu}{S'} \dot{\sim} N(0,1).$$

Exercícios Resolvidos

Exercício 41., página 458

Resolução 41. a) – Simetria de $N(0,1)$ $(1-\frac{\alpha}{2}=0.975)$

$$\overbrace{q_{0.025}}^{-q_{0.975}} < \frac{\overline{x}-\theta}{\sqrt{\frac{\overline{x}(1-\overline{x})}{200}}} < q_{0.975}$$

$$\Longleftrightarrow \overline{x}-\sqrt{\frac{\overline{x}(1-\overline{x})}{200}}\times q_{0.975} < \theta < \overline{x}+\sqrt{\frac{\overline{x}(1-\overline{x})}{200}}\times q_{0.975}$$

Resolução Exercício 41. a) em *R*

```
> x<-50/200
> sd<-sqrt(x*(1-x)/200)
> x-sd*qnorm(0.975)
[1] 0.1899886
> x+sd*qnorm(0.975)
[1] 0.31001140
```

Resolução Exercício 41. b) – Como resolver

1. De **41.a)** conhecemos $\overline{x} = 50/200 = 0.25$ e o intervalo de confiança para $\theta = E(X)$ ($t_1 < \theta < t_2$, com $t_1 \approx 0.19$ e $t_2 \approx 0.31$).
2. Nas condições do exercício, **pretendemos calcular a dimensão [mínima] da amostra**, n, de modo a que $|\overline{x} - \theta| < \frac{t_2 - t_1}{4}$ tenha o mesmo grau de confiança de **41. a)** $(1 - \alpha = 0.95)$.
3. Vou recorrer à **variável fulcral** $\dfrac{\overline{X} - \theta}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}} \dot{\sim} N(0,1)$

   (n desconhecido).

Resolução 41. b) – Variável Fulcral

$$|\overline{x}-\theta| < \overbrace{\frac{0.31-0.19}{4}}^{=0.03} \implies \left|\frac{\overline{X}-\theta}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}}\right| < \overbrace{\frac{0.03}{\sqrt{\frac{0.25\times(1-0.25)}{n}}}}^{=\frac{0.03\sqrt{n}}{\sqrt{0.25\times 0.75}}} = q_{0.975}$$

Resolução 41. b) – Cálculo da dimensão da amostra

$$\frac{0.03\sqrt{n}}{\sqrt{0.25\times 0.75}} = q_{0.975} \iff n = \left(\frac{q_{0.975}}{0.03}\right)^2 \times 0.25 \times 0.75 \approx 800$$

Instrução R: `> (qnorm(0.975)/0.03)^2*0.25*0.75`

Formulação 45. (envolve um processo de Poisson)

População $X \sim Po(\lambda)$, sendo λ o **tempo médio de espera**. Temos ainda para uma **amostra casual** de dimensão $n = 100$:

- $\overline{x} = 17$ (segundos) como o **tempo médio de espera**;
- $s = 6$ (segundos) como o **desvio padrão** em relação ao tempo de espera.

Aplicação do Teorema do Limite Central

$$\sqrt{n-1}\frac{\overline{X}-\lambda}{S} \dot{\sim} t_{n-1} \quad \Longrightarrow \sqrt{n-1}\frac{\overline{X}-\lambda}{S} \dot{\sim} N(0,1)$$

Observação: A variável fulcral $Z = \sqrt{n-1}\frac{\overline{X}-\lambda}{S}$ dá-nos uma **aproximação assintótica** da **distribuição normal** $N(0,1)$, dado a **dimensão da amostra** ser **elevada** ($n = 100$).

Resolução 45 – Simetria de $N(0,1)$ $(1-\frac{\alpha}{2}=0.975)$

$$\overbrace{q_{0.025}}^{-q_{0.975}} < \sqrt{99}\frac{\overline{x}-\lambda}{s} < q_{0.975}$$

$$\Longleftrightarrow \overline{x} - \frac{s}{\sqrt{99}} \times q_{0.975} < \lambda < \overline{x} + \frac{s}{\sqrt{99}} \times q_{0.975}$$

Resolução Exercício 45 em *R*

```
> x<-17
> s<-6
> x-s/sqrt(99)*qnorm(0.975)
[1] 15.8181
> x+s/sqrt(99)*qnorm(0.975)
[1] 18.1819
```

Exercícios Resolvidos

Exercício 45., página 458

Resolução 45 – Variável fulcral alternativa

$$\frac{\overline{X}-\lambda}{\sqrt{\frac{\lambda}{n}}} \dot{\sim} N(0,1) \overset{\text{estimador } V(X)}{\Longrightarrow} \frac{\overline{X}-\lambda}{\sqrt{\frac{\overline{X}}{n}}} \dot{\sim} N(0,1)$$

Observação:

$\widehat{\lambda}=\overline{X}$ é um **estimador de máxima verossimilhança (m.v.)** para a média e a variância de $X \sim Po(\lambda)$, uma vez que $E(X)=V(X)=\lambda$.

Resolução Exercício 45 em *R*

```
> x<-17
> s<-sqrt(x)
> x-s/sqrt(100)*qnorm(0.975)
[1] 16.19189
> x+s/sqrt(100)*qnorm(0.975)
[1] 17.80811
```

Resolução Exercício 47 – Ideia

1. Considerar a **população** $X \sim Po(\lambda)$.
2. Verificar que $\overline{x} = 4/3$ corresponde à média de peças produzidas por dia para uma **amostra casual de dimensão** $n = 30$:

   '(…) Nos 30 dias de teste, registou-se um total de 40 componentes defeituosas (…)'.

3. Considerar a variável fulcral $Z = \dfrac{\overline{X} - \lambda}{\sqrt{\frac{\overline{X}}{n}}} \dot{\sim} N(0, 1)$.
4. Determinar o **intervalo de confiança** para λ com grau de confiança $1 - \alpha = 0.95$: $-q_{0.975} < \dfrac{\overline{x} - \lambda}{\sqrt{\frac{\overline{x}}{n}}} < q_{0.975}$.
5. Verificar se $\lambda = 2$ pertence ao **intervalo de confiança**.

Conteúdos

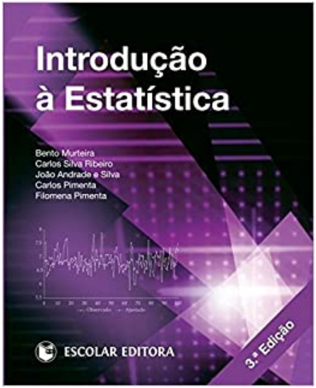
Exercícios Propostos

Complemento das Aulas 16 & 17

Estimação por Intervalos

- **págs. 454–458:** 29., 30., 32., 33., 39., 40.

Murteira, B., C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta, F. Pimenta (2015), **Introdução à Estatística, 3ª Edição**', Escolar Editora.



[Murteira et al. (2015), Exercício 24., p. 453]

Alguns comentários:

a) Compare o valor de $V(T) = V(0.5X_1 + 0.4X_2 + 0.1X_3)$ com o valor de $V(\overline{X})$, sendo $\overline{X} = \frac{1}{3}(X_1 + X_2 + X_3)$.

b) Para discutir e justificar a veracidade desta afirmação (**(V)** ou **(F)**) vide **[Murteira et al. (2015), Exercício 18.a), p. 451]** que se encontra resolvido em `Semana7.pdf`. Outra pergunta complementar a esta alínea é a seguinte:

'Será que o estimador centrado e mais eficiente é o estimador de menor erro quadrático médio?'

c) Tome como referência o estimador S^2 – estimador T_2 de **[Murteira et al. (2015), Exercício 18.a), p. 451]**.