

## Parte C - Introdução à Inferência Estatística

### Notas:

- Os exercícios marcados com ✓ são de resolução fortemente recomendada.
- Os exercícios marcados com ⊗ não poderão ser resolvidos com a matéria lecionada em 2023/24.

### Exercícios de Distribuições de Amostragem e de Estimadores

- 3.1. De uma população normal de média 8 e desvio padrão 4, extraíu-se uma amostra aleatória de dimensão 100.
- a) Qual a probabilidade de a diferença, em valor absoluto, entre a média da amostra e a média da população ser superior a 0.5?
  - b) Se a população não fosse normal, qual seria essa probabilidade? Justifique.
- ✓ 3.2. (*Exame 12.01.2009*) O peso médio dos indivíduos duma certa espécie de bivalves é 31 g e o respectivo desvio padrão é 2.4 g. Recolhe-se uma amostra aleatória de 100 indivíduos desta espécie.
- a) Qual a probabilidade, aproximada, de a média da amostra ser inferior a 30 g?
  - b) Qual a probabilidade, aproximada, de o peso total da amostra ser superior a 3150 g?
- ✓ 3.3. Mediu-se o comprimento (em cm) de 25 coelhos adultos de uma dada raça escolhidos aleatoriamente. A média da amostra foi 56.2 cm e o desvio padrão 4 cm. Admita que a população dos comprimentos dos coelhos adultos daquela raça é bem modelada por uma distribuição normal.
- a) Indique estimativas para a média e o desvio padrão da população.
  - b) Usando as estimativas da alínea anterior como os parâmetros da população, determine a probabilidade de o comprimento de um coelho adulto daquela raça, escolhido ao acaso, ser superior a 57 cm.
  - c) Chama-se erro padrão da média amostral,  $\bar{X}$ , a uma estimativa do desvio padrão da v.a. média amostral. Com base nos resultados obtidos, determine o erro padrão da média amostral.
- 3.4. Seja  $X$  uma variável aleatória com distribuição de Bernoulli de parâmetro  $p$  desconhecido e  $(X_1, X_2, X_3)$  uma amostra aleatória associada a esta variável.
- a) Quais as propriedades da amostra aleatória  $(X_1, X_2, X_3)$ ?
  - b) Sugerem-se os dois estimadores para  $p$ ,  $\hat{P}_1 = \bar{X}$  e  $\hat{P}_2 = (X_1 + 2X_2 + X_3)/4$ . Determine o valor esperado de cada um dos estimadores. Interprete.

- c) Analise os estimadores quanto à sua variância. Qual dos dois estimadores considera que deve ser escolhido para estimar  $p$ ? Justifique.
- d) Procedeu-se à recolha de uma amostra, nas condições do enunciado, tendo-se obtido os valores  $(0, 1, 0)$ . Obtenha duas estimativas para  $p$ .

- ✓ 3.5. Depois de fabricado e embalado, a actividade de um certo adubo segue aproximadamente uma distribuição normal com  $\mu = 120$  dias e  $\sigma = 40$  dias.

Pretende-se enviar um lote de embalagens do referido adubo de modo que a actividade média amostral ( $\bar{X}$ ) seja superior a 118 dias com probabilidade de 0.95. Qual o tamanho do lote a enviar?

- ✓ 3.6. Seja  $(X_1, \dots, X_n)$  uma amostra aleatória de tamanho  $n$  proveniente de uma população  $X$  com distribuição *Uniforme*(0, 1). Exprima, em função de  $n$ , a probabilidade, aproximada, de  $\bar{X}$  ser superior 0.9.

- ✓ 3.7. Uma amostra aleatória de dimensão 50,  $(X_1, \dots, X_{50})$ , é extraída de uma população de Poisson com  $\lambda = 10$ . Recorra à distribuição normal para calcular um valor aproximado de  $P(\bar{X} > 11)$ .

- 3.8. Seja  $X_1, X_2, \dots, X_n$  uma amostra de variáveis aleatórias independentes e com a mesma distribuição, de valor médio  $\mu$  e variância  $\sigma^2$ . Considere os seguintes estimadores de  $\mu$  e  $\sigma^2$ , respectivamente,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- a) Prove que  $E[\bar{X}] = \mu$  e  $Var[\bar{X}] = \sigma^2/n$ .
- b) Prove que

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

- c) Usando os resultados das alíneas anteriores, mostre que  $E[S^2] = \sigma^2$ .

- ✓ 3.9. (*Exame 31.01.2011*) Considere a amostra aleatória  $X_1, X_2, \dots, X_{100}$ , retirada de uma população  $X$  para a qual se tem  $E[X] = 0.5$  e  $Var[X] = 0.0625$ . Considere a variável aleatória  $S = \sum_{i=1}^{100} X_i$ .

- a) Defina “amostra aleatória”.
- b) Determine o valor esperado e a variância de  $S$ . Justifique convenientemente.
- c) Determine, aproximadamente,  $P[S < 45]$ .

- 3.10. (*Exame 17.01.2011*) Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de dimensão  $n$ , proveniente de uma população  $X$  com valor médio  $\mu$  e variância  $\sigma^2$ . Seja  $\bar{X}$  a média da amostra aleatória.

Responda, **justificando convenientemente**, se são verdadeiras ou falsas as afirmações nas seguintes alíneas.

- a)  $\bar{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$ .

- b)  $E[\mu] = \bar{X}$ .  
 c) Se  $X \sim \text{Bernoulli}(p)$  então  $n\bar{X} \sim \text{Binomial}(n, p)$ .

**3.11.** Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de dimensão  $n$ , proveniente de uma população com distribuição de Bernoulli de parâmetro  $p$ , desconhecido. Considere a v.a.  $Y = \sum_{i=1}^n X_i$ .

- a) Caracterize, justificando, a distribuição da v.a.  $Y$ .  
 b) Seja  $\hat{P} = Y/n$ . Mostre que  $\hat{P}$  é um estimador centrado de  $p$ .

**3.12.** (Adaptado do *Exame 12.01.2015*) Seja  $(X_1, \dots, X_n)$  uma amostra aleatória de dimensão  $n$  proveniente de uma população  $X$  de valor médio  $\mu$  e variância  $\sigma^2$ . Seja  $\bar{X}$  a média associada aquela amostra aleatória.

- a) i) Defina “amostra aleatória de dimensão  $n$ ”.  
 ii) Determine  $\text{Cov}(X_1, \bar{X})$  em função da variância de  $X$  e da dimensão da amostra.

b) Considere os seguintes estimadores de  $\mu$ :  $\bar{X}$  e  $T = \frac{2\sum_{i=1}^n i X_i}{n(n+1)}$ .

- i) Mostre que ambos são estimadores centrados de  $\mu$ .  
 ii) Determine a variância de cada um dos estimadores.

**Nota:** Tenha em conta que  $\sum_{i=1}^n i = \frac{n(n+1)}{2}$  e  $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$ .

- iii) Qual dos estimadores de  $\mu$  considerados nas alíneas anteriores é o mais eficiente? Justifique.  
 iv) Supondo que foi observada a amostra  $(1, 3, 2, 5, 4, 0)$ , obtenha o valor de 2 estimativas.

**3.13.** (*Exame 26.01.2018*) Recolheram-se de forma independente duas amostras aleatórias  $(X_1, \dots, X_n)$  e  $(Y_1, \dots, Y_n)$  de uma população exponencial de valor médio  $\beta$ . Sendo  $\bar{X}$  e  $\bar{Y}$  as médias respectivas, considere os seguintes estimadores para o valor médio da população:

$$T_1 = \bar{X} \quad T_2 = \frac{\bar{X} + \bar{Y}}{2}$$

- a) Determine o valor médio e a variância de cada um dos estimadores em função de  $\beta$ .  
 b) Os estimadores  $T_1$  e  $T_2$  são estimadores centrados de  $\beta$ ? Porquê?  
 c) Qual destes estimadores considera ser o melhor estimador para  $\beta$ ?  
 d) Escreva a função geradora de momentos do estimador  $T_1 = \bar{X}$ . Será que  $T_1$  tem distribuição exponencial? Justifique.

**3.14.** (*Exame 9.01.2019*) Seja  $(X_1, X_2)$  um par de variáveis aleatórias independentes tais que  $E[X_i] = \mu_i$  e  $\text{Var}[X_i] = \sigma_i^2$ ,  $i = 1, 2$ .

- a) i) Quando se poderá dizer que  $(X_1, X_2)$  é uma amostra aleatória de dimensão 2 retirada de uma população  $X$  com valor médio  $\mu$  e variância  $\sigma^2$ ?

ii) No caso de  $(X_1, X_2)$  ser uma amostra aleatória retirada de  $X$ , considere os seguintes estimadores de  $\mu$ :

$$T_1 = \frac{3X_1 - X_2}{2} \qquad T_2 = \frac{X_1 + X_2}{2}$$

Mostre que  $T_1$  e  $T_2$  são estimadores centrados de  $\mu$  e diga, justificando convenientemente, qual dos dois é “o melhor” estimador.

b) Se  $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ ,  $i = 1, 2$ , demonstre que  $aX_1 + bX_2$ , com  $a, b \in \mathbb{R}$  é também normal e indique os respectivos parâmetros.

## Exercícios de Intervalos de Confiança e Testes de Hipóteses

- ✓ **3.15.** Seja  $X$  uma população com distribuição normal, de média  $\mu$  e desvio padrão  $\sigma=2$ . Uma amostra aleatória de dimensão 25 foi extraída desta população, tendo-se obtido  $\bar{x} = 78.3$ .
- Calcule o intervalo de confiança a 99% para  $\mu$ .
  - Qual o erro máximo cometido (a 99% de confiança) ao estimar  $\mu$  por  $\bar{x}=78.3$ ?
  - Qual deverá ser a dimensão da amostra para que o erro máximo cometido, a 99% de confiança, ao estimar  $\mu$  por  $\bar{x}$ , não exceda 0.1?
- ✓ **3.16.** Para avaliar a tensão máxima suportada por uma barra de aço testaram-se  $n$  barras tendo-se obtido, nas unidades adequadas,  $\bar{x}_n = 20$  e para extremo superior do intervalo de confiança a 95% para o verdadeiro valor médio da tensão obteve-se 21.7. Sabendo que se admite que a tensão suportada por uma barra de aço é uma v.a. normal com desvio padrão  $\sigma = 3$ , determine o extremo inferior do intervalo de confiança e a dimensão  $n$  da amostra.
- ✓ **3.17.** Uma empresa de peixe congelado está a ser investigada com o objectivo de se verificar se cada embalagem pesa de facto 1 kg em média. Numa amostra aleatória de 100 embalagens de peixe registou-se o peso (kg) de cada embalagem,  $x_i$  ( $i = 1, \dots, 100$ ), tendo-se obtido os seguintes valores:
- $$\sum_{i=1}^{100} x_i = 95.9 \text{ kg} \quad \sum_{i=1}^{100} x_i^2 = 93.12 \text{ kg}^2$$
- Indique uma estimativa para a média e para a variância do peso de uma embalagem de peixe.
  - O que pode dizer sobre o resultado da investigação? Justifique convenientemente, especificando as hipóteses que necessitou de considerar.
  - Determine, explicitando as hipóteses necessárias, um intervalo a 95% de confiança para a variância do peso de uma embalagem de peixe.
- ✓ **3.18.** O grau de acidez do azeite produzido em certa região supõe-se ter distribuição normal. Uma amostra da produção de dimensão 25 conduziu a uma acidez média de 1 grau e a um desvio padrão de 0.33 graus.
- Face a estes valores, alguém sugeriu o intervalo  $]0.815; 1.185[$  para o valor esperado da acidez. Qual o nível de confiança associado a este intervalo?
  - Para o nível de confiança determinado na alínea anterior, qual o erro máximo cometido ao estimar o verdadeiro valor médio da acidez por 1 grau?
- ✓ **3.19.** Numa amostra de 16 elementos, que se supõe ter sido retirada de uma população com distribuição normal, o desvio padrão obtido foi de 5.2.

- a) Foi calculado um intervalo de confiança para a média populacional tendo-se obtido  $]24.2297; 29.7703[$ .  
Indique justificando:
- Qual a média da amostra;
  - Qual o grau de confiança do intervalo calculado.
- b) Qual o intervalo de confiança a 95% para a variância populacional?

- ✓ **3.20.** (*Exame 31.01.2012*) Um silvicultor sabe que a altura das árvores para madeira é importante para os compradores. Afirmou a um comprador que a altura média das suas árvores era 28 metros. O comprador fechou negócio, mas aquando do abate escolheu ao acaso 65 árvores e mediu as alturas. Obteve  $\sum_{i=1}^{65} x_i^2 = 43959.32 m^2$ , onde  $x_i$  designa a altura de cada árvore seleccionada. O comprador obteve o seguinte intervalo de confiança para a altura média, em metros:

$]25.1538; 26.5462[$ .

- Calcule uma estimativa para a altura média das árvores daquela floresta.
- Determine a confiança do intervalo dado acima.
- Considera que o comprador tem razão para reclamar? Justifique.
- Pretende-se construir um intervalo de confiança para a variância da altura das árvores.
  - Indique o estimador da variância usado na construção deste intervalo, apresentando a sua expressão.
  - Com base nos resultados fornecidos construa o intervalo de confiança a 95% para a variância das alturas, indicando o(s) pressuposto(s) necessário(s).

- ✓ **3.21.** Suponha que o rendimento de um pé de tomateiro expresso em kg é uma variável aleatória com distribuição normal de valor médio 1 kg. Numa parte da produção foi utilizado um novo fertilizante. Observada uma amostra de 10 pés de tomateiro da parte da produção em que foi utilizado o novo fertilizante obtiveram-se os seguintes resultados:

1.375   1.223   1.773   1.752   0.779   1.407   1.068   1.633   1.201   1.042

expressos em kg. Que decisão se deverá tomar perante estes resultados, face ao novo fertilizante?

- ✓ **3.22.** (*Exame 26.01.2018*) Para analisar a qualidade de um determinado tipo de azeite virgem, proveniente de uma região, R1, foram seleccionadas aleatoriamente 20 garrafas e determinada a acidez do azeite de cada uma delas, expressa em concentração de ácido oleico (g/100g). Os dados obtidos foram introduzidos no  $\mathbb{R}$ , apresentando-se de seguida alguns dos comandos e respectivos resultados.

```
> acidez<-c(0.635, 0.822, 0.833, ... , 0.862, 0.694, 0.665)
```

```
> mean(acidez)
[1] 0.76925
```

```
> shapiro.test(acidez)
Shapiro-Wilk normality test
data:  acidez
```

```
> var(acidez)
[1] 0.01504357
```

W = 0.95549, p-value = 0.4581

- a) De acordo com a legislação em vigor, o azeite virgem só pode ser classificado como Extra se tiver uma acidez média inferior a 0.8. Será que o azeite virgem analisado pode ter essa classificação? Justifique convenientemente a sua resposta.
- b) Numa outra região, R2, analisou-se uma amostra de 20 garrafas de azeite virgem do mesmo tipo. Com os valores observados contruiu-se o seguinte intervalo de confiança a 95% para a diferença da acidez média do azeite virgem daquele tipo entre as regiões R1 e R2.

$$] - 0.0357; 0.0667 [$$

Justificando convenientemente, responda às seguintes questões:


- i) Que pressupostos foi necessário verificar para construir o intervalo de confiança dado?
- ii) Determine o valor da acidez média observada na região R2.
- iii) Compare a acidez média deste tipo de azeite nas duas regiões R1 e R2.

**3.23.** (Exame 28.01.2013) Pretende-se comparar dois antibióticos A e B usados para o tratamento de um certo tipo de infecção em bovinos. Escolhem-se ao acaso dois grupos de bovinos doentes: a um grupo é administrado o antibiótico A e ao outro o antibiótico B, registando-se o tempo (h) até os sintomas desaparecerem.

- ⊗ a) Considere a v.a.  $X$ , de valor médio  $\mu$  e variância  $\sigma^2$  desconhecidos, que designa o tempo até os sintomas desaparecerem no grupo de bovinos a que se aplica o antibiótico A. Seja  $(X_1, \dots, X_{10})$  uma amostra aleatória retirada de  $X$ . Sugere-se dois estimadores para  $\mu$ ,

$$T_1 = \bar{X} \quad \text{e} \quad T_2 = \frac{X_1 + 3X_{10}}{4},$$

com  $\bar{X}$  média da amostra aleatória.

- i) Defina “amostra aleatória de dimensão 10”.
  - ii) Mostre que  $T_1$  e  $T_2$  são estimadores centrados de  $\mu$  e obtenha as respectivas variâncias.
  - iii) Qual dos dois estimadores deve ser escolhido para estimar  $\mu$ ? Justifique.
- b) Os resultados obtidos pela aplicação dos dois antibióticos foram introduzidos no *software* . Utilize, sempre que possível, os resultados apresentados no Anexo para responder às seguintes questões.
- i) Com base nos dados recolhidos indique estimativas para o tempo médio no caso de aplicação do antibiótico A, associadas aos estimadores dados em a).
  - ii) Face aos resultados obtidos poder-se-á concluir que o antibiótico A é mais lento do que o antibiótico B no desaparecimento dos sintomas da infecção? Explícite e valide os pressupostos necessários à resolução do problema.

## Anexo

```
> A<-c(49.5,48,52.5,46.5,52,48.5,40,50.5,46,46.5)
> B<-c(47.5,48.5,40.5,43.5,49,43,42,45,48.5,40)

> var(A)          > var(B)          > var(A-B)
[1] 13.05556      [1] 11.84722      [1] 19.29167

> shapiro.test(A)
      Shapiro-Wilk normality test
data:  A
W = 0.9189, p-value = 0.3481

> shapiro.test(B)
      Shapiro-Wilk normality test
data:  B
W = 0.9009, p-value = 0.2244

> shapiro.test(A-B)
      Shapiro-Wilk normality test
data:  A - B
W = 0.9482, p-value = 0.6474

> var.test(A,B)
      F test to compare two variances
data:  A and B
F = 1.102, num df = 9, denom df = 9, p-value = 0.8873
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2737195 4.4366172
sample estimates:
ratio of variances
      1.101993

> t.test(A,B,paired=T,alternative="two.sided")
      Paired t-test
data:  A and B
t = 2.3399, df = 9, p-value = 0.04403
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1079894 6.3920106
sample estimates:
mean of the differences
      3.25

> t.test(A,B,paired=T,alternative="less")
      Paired t-test
data:  A and B
t = 2.3399, df = 9, p-value = 0.978
```



```

alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
  -Inf 5.796092
sample estimates:
mean of the differences
      3.25


> t.test(A,B,paired=T,alternative="greater")
      Paired t-test
data:  A and B
t = 2.3399, df = 9, p-value = 0.02201
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
  0.703908      Inf
sample estimates:
mean of the differences
      3.25

> t.test(A,B,paired=F,var.equal=T,alternative="two.sided")
      Two Sample t-test
data:  A and B
t = 2.0595, df = 18, p-value = 0.05421
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.06538397  6.56538397
sample estimates:
mean of x mean of y
  48.00    44.75

> t.test(A,B,paired=F,var.equal=T,alternative="less")
      Two Sample t-test
data:  A and B
t = 2.0595, df = 18, p-value = 0.9729
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
  -Inf 5.986459
sample estimates:
mean of x mean of y
  48.00    44.75

> t.test(A,B,paired=F,var.equal=T,alternative="greater")
      Two Sample t-test
data:  A and B
t = 2.0595, df = 18, p-value = 0.02711
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
  0.5135412      Inf
sample estimates:
mean of x mean of y
  48.00    44.75

```

✓ 3.24. (Exame 25.01.2016) Um investigador está interessado em avaliar, numa dada região, a produção por macieira de duas variedades de maçã: Gala e Golden. Para essa finalidade, foram registadas as produções, em kg, de 10 macieiras de cada uma das variedades de maçã em estudo. Os dados foram introduzidos no *software* . No ANEXO apresentam-se resultados de comandos, alguns inadequados. Responda, de forma completa, às seguintes questões utilizando os resultados do Anexo.

- Indique uma estimativa para o valor médio da produção da variedade Gala. Qual é o erro máximo, a 95% de confiança, que se comete ao estimar  $\mu$  pela estimativa indicada?
- Conjectura-se que as variabilidades das produções das duas variedades são iguais. Será esta conjectura compatível com os dados obtidos?
- Com base nos valores obtidos poder-se-á concluir que a produção média da variedade Golden é significativamente superior à produção média da variedade Gala?

## ANEXO

```
> gala<-c(84,82,90,86,80,91,85,79,81,82)
> golden<-c(95,102,85,93,104,89,98,99,107,106)

> var(gala); var(golden); var(gala-golden)
[1] 16.44444
[1] 53.51111
[1] 121.2889

> shapiro.test(gala); shapiro.test(golden); shapiro.test(gala-golden)

Shapiro-Wilk normality test
data: gala
W = 0.9241, p-value = 0.3922

Shapiro-Wilk normality test
data: golden
W = 0.9573, p-value = 0.7546

Shapiro-Wilk normality test
data: gala - golden
W = 0.9014, p-value = 0.2269

> var.test(gala, golden)

      F test to compare two variances

data: gala and golden
F = 0.3073, num df = 9, denom df = 9, p-value = 0.09365
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
```

```
0.0763312 1.2372241
sample estimates:
ratio of variances
      0.307309
```

```
> t.test(gala, golden, paired=T)
```

```
      Paired t-test
data:  gala and golden
t = -3.9625, df = 9, p-value = 0.003292
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -21.678314  -5.921686
sample estimates:
mean of the differences
      -13.8
```

```
> t.test(gala, golden, paired=F, var.equal=T)
```

```
      Two Sample t-test
data:  gala and golden
t = -5.2176, df = 18, p-value = 5.811e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -19.356752  -8.243248
sample estimates:
mean of x mean of y
      84.0      97.8
```

```
> t.test(gala, golden, paired=F, var.equal=T, alternative="less")
```

```
      Two Sample t-test
data:  gala and golden
t = -5.2176, df = 18, p-value = 2.906e-05
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -9.213556
sample estimates:
mean of x mean of y
      84.0      97.8
```

```
> t.test(gala, golden, paired=F, var.equal=T, alternative="greater")
```

```
      Two Sample t-test
data:  gala and golden
t = -5.2176, df = 18, p-value = 1
alternative hypothesis: true difference in means is greater than 0
```

```

95 percent confidence interval:
-18.38644      Inf
sample estimates:
mean of x mean of y
      84.0      97.8

```

- ✓ **3.25.** (*Exame 11.01.2016*) Os rolamentos produzidos por uma empresa devem ter diâmetro entre 150 e 160 mm (inclusive). Uma amostra de 35 rolamentos foi seleccionada aleatoriamente e os valores observados foram introduzidos no *software R*. Utilize os resultados apresentados no **Anexo** e responda, de forma completa, às seguintes questões. (Note que alguns resultados podem não ser necessários ou adequados).

- Indique uma estimativa para o valor médio e para a variância do diâmetro dos rolamentos. Explícite os estimadores associados.
- Determine um intervalo de confiança a 95% para o valor médio dos diâmetros das peças produzidas?
- Poder-se-á admitir que mais de 60% dos rolamentos produzidos pela empresa satisfazem o exigido? Justifique.

#### ANEXO

```

> diametro<-c(135, 154, 159, 155, 167, 159, 158, 159, 154, 155,
+             152, 169, 154, 158, 140, 149, 145, 157, 151,154,168,153,151,
+             160, 155, 155, 143,157, 139, 159, 139, 159, 162, 151, 150)
> sort(diametro)
 [1] 135 139 139 140 143 145 149 150 151 151 151 152 153 154 154 154 154 155
[19] 155 155 155 157 157 158 158 159 159 159 159 159 160 162 167 168 169

> mean(diametro)           >var(diametro)
 [1] 153.8571              [1] 63.2437

> shapiro.test(diametro)

      Shapiro-Wilk normality test

data:  diametro
W = 0.9476, p-value = 0.09577

> t.test(diametro)

      One Sample t-test

data:  diametro
t = 114.4572, df = 34, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:

```

```

151.1253 156.5890
sample estimates:
mean of x
153.8571

> t.test(diametro,mu=160,alternative="less")


One Sample t-test

data: diametro
t = -4.5698, df = 34, p-value = 3.080e-05
alternative hypothesis: true mean is less than 160
95 percent confidence interval:
 -Inf 156.1301
sample estimates:
mean of x
153.8571

```

- ✓ **3.26.** Num estudo sobre o número de folhas por planta de tabaco, obtiveram-se os seguintes dados:

n <sup>o</sup> de folhas	17	18	19	20	21	22	23	24
n <sup>o</sup> de plantas	3	22	44	42	22	10	6	1

- Indique a unidade estatística e a variável em estudo.
  - Determine a média e a mediana da amostra. Compare-as e comente.
  - Com base nesta amostra poder-se-á afirmar que 90% das plantas de tabaco têm 21 ou menos folhas? Justifique.
- ✓ **3.27.** Duas marcas de chocolate (A e B) produzem chocolate rotulado “70% de cacau”. Seleccionaram-se aleatoriamente 15 chocolates de cada uma das marcas, que foram analisados quanto ao teor de cacau. Os resultados foram introduzidos no *software* . Utilize, sempre que possível, os resultados apresentados abaixo para responder às seguintes questões (alguns cálculos apresentados são desnecessários ou inadequados).
- Conjectura-se que as variabilidades do teor de cacau dos chocolates das duas marcas são diferentes. Será esta conjectura compatível com os dados obtidos? Explícite e valide os pressupostos necessários à resolução do problema.
  - Indique estimativas para o valor esperado do teor de cacau dos chocolates de cada uma das marcas.
  - Os dados recolhidos evidenciam que o teor médio de cacau do chocolate da marca A é significativamente superior ao da marca B? Explícite e valide os pressupostos necessários à resolução do problema.
  - O produtor de chocolate da marca A afirma que o teor médio de cacau do seu chocolate é superior a 70%. O que pode dizer sobre a afirmação do produtor? Justifique convenientemente. Especifique as hipóteses nula e alternativa de um teste de hipóteses que permita averiguar a validade desta afirmação.

```

> A
[1] 72.8 69.6 70.0 73.3 69.8 71.1 71.0 71.0 71.0 76.3 71.3
[12] 75.4 72.1 70.4 68.6
> B
[1] 69.8 71.3 67.3 69.4 67.7 67.9 70.7 70.5 70.2 71.4 71.6
[12] 73.1 70.6 68.5 70.2

> var(A)                > var(B)                > var(A-B)
[1] 4.483143            [1] 2.616952            [1] 3.643810

> shapiro.test(A)      > shapiro.test(B)
  Shapiro-Wilk normality test      Shapiro-Wilk normality test
data: A                          data: B
W = 0.9066, p-value = 0.1199      W = 0.9613, p-value = 0.7144

> shapiro.test(A-B)
  Shapiro-Wilk normality test
data: A - B
W = 0.9754, p-value = 0.9283

> var.test(A,B)
  F test to compare two variances
data: A and B
F = 1.7131, num df = 14, denom df = 14, p-value = 0.3254
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5751437 5.1026658
sample estimates:
ratio of variances
 1.713116

> t.test(A,B,paired=T,alternative="less")
  Paired t-test
data: A and B
t = 3.1787, df = 14, p-value = 0.9967
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 2.434763
sample estimates:
mean of the differences
 1.566667

> t.test(A,B,paired=T,alternative="greater")
  Paired t-test
data: A and B
t = 3.1787, df = 14, p-value = 0.003349
alternative hypothesis: true difference in means is greater than 0


```

```
95 percent confidence interval:
 0.6985701      Inf
sample estimates:
mean of the differences
      1.566667
```

```
> t.test(A,B,paired=F,var.equal=T,alternative="less")
      Two Sample t-test
data:  A and B
t = 2.2771, df = 28, p-value = 0.9847
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 2.737039
sample estimates:
mean of x mean of y
 71.58000  70.01333
```

```
> t.test(A,B,paired=F,var.equal=T,alternative="greater")
      Two Sample t-test
data:  A and B
t = 2.2771, df = 28, p-value = 0.01531
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.3962939      Inf
sample estimates:
mean of x mean of y
 71.58000  70.01333
```

- ✓ **3.28.** (*Exame 10.01.2018*) Com o objectivo de estudar o teor de gordura no leite de duas raças de vacas leiteiras (*Ayrshire*, raça A e *Holstein-Fresian*, raça B), foram escolhidas aleatoriamente 20 vacas de cada raça e foi determinada a percentagem de gordura no respectivo leite.

Os dados obtidos foram introduzidos no  e foram efectuados cálculos que se encontram no **Anexo**. Utilize os resultados que considere adequados para responder às seguintes alíneas.

- Indique uma estimativa pontual para a variância da percentagem de gordura no leite da raça A. Indique, explicitando-o, o estimador associado.
- Comente a veracidade da seguinte afirmação: “Não existe diferença significativa entre as variâncias das percentagens de gordura no leite das raças A e B”. Justifique convenientemente.
- Será que a percentagem média de gordura no leite da raça A é superior à da raça B? Justifique convenientemente a sua resposta.

## Anexo

```
> A<-c(3.53, 3.71, ..., 4.37, 4.41, 4.44)
> B <-c(3.30,3.38 ,..., 3.94 ,3.95 ,4.25)
> mean(A)          > mean(B)          > var(A)          > var(B)
[1] 4.06           [1] 3.6605           [1] 0.0680           [1] 0.0542

> shapiro.test(A)

      Shapiro-Wilk normality test

data:  A
W = 0.9569, p-value = 0.4836

> shapiro.test(B)

      Shapiro-Wilk normality test

data:  B
W = 0.945, p-value = 0.2974

> shapiro.test(A-B)

      Shapiro-Wilk normality test

data:  A - B
W = 0.9241, p-value = 0.1191

> var.test(A,B)

      F test to compare two variances

data:  A and B
F = 1.2552, num df = 19, denom df = 19, p-value = 0.6253
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4968133 3.1711368
sample estimates:
ratio of variances
      1.255174

> t.test(A,B,paired=TRUE,alternative="greater")
```



Paired t-test

```
data: A and B
t = 18.4124, df = 19, p-value = 7.122e-14
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.3619823      Inf
sample estimates:
mean of the differences
                0.3995
```

```
> t.test(A,B,paired=TRUE,alternative="less")
```

Paired t-test

```
data: A and B
t = 18.4124, df = 19, p-value = 1
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 0.4370177
sample estimates:
mean of the differences
                0.3995
```

```
> t.test(A,B,paired=FALSE,var.equal=TRUE,alternative="greater")
```

Two Sample t-test

```
data: A and B
t = 5.109, df = 38, p-value = 4.72e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.2676668      Inf
sample estimates:
mean of x mean of y
 4.0600      3.6605
```

```
> t.test(A,B,paired=FALSE,var.equal=TRUE,alternative="less")
```

Two Sample t-test

```
data: A and B
```

$t = 5.109$ ,  $df = 38$ ,  $p\text{-value} = 1$   
 alternative hypothesis: true difference in means is less than 0  
 95 percent confidence interval:  
     -Inf 0.5313332  
 sample estimates:  
 mean of x mean of y  
     4.0600    3.6605

- ✓ **3.29.** A fim de investigar os efeitos de ambientes nitrosos e de ambientes fosfatados no desenvolvimento de colónias de bactérias, contaminam-se 10 plaquetas envolvidas em cada um daqueles ambientes com as bactérias em estudo, e deixa-se incubar durante 24 horas. Após esse tempo, procede-se à contagem do número de colónias de bactérias em cada plaqueta, tendo-se obtido os seguintes resultados:

Ambiente nitroso	60	47	12	29	51	46	49	74	63	101
Ambiente fosfatado	8	46	21	13	58	33	20	46	31	38

Investigue a hipótese de o tipo de ambiente não influir no desenvolvimento das colónias de bactérias. Responda de forma completa à questão.

- ✓ **3.30.** Numa Estação Florestal estudam-se problemas de intercepção de precipitação. Nesse sentido em 12 dias de chuva (suponha que se trata de uma amostra aleatória de dias com precipitação) são colocados dois udómetros para medir a quantidade de precipitação: um numa zona desarborizada e outro sob as copas das árvores. As leituras da quantidade de água em cada udómetro (medidas em cm de altura) deram os seguintes valores (cada coluna corresponde a um dia):

zona descoberta	5.87	1.30	2.34	2.82	5.89	9.09	1.93	9.27	4.65	4.35	5.00	8.43
sob coberto	4.96	1.14	2.24	2.26	4.75	7.83	1.86	8.85	4.17	3.65	4.08	7.99

- a) Estime a precipitação média na zona desarborizada e na zona sob coberto.
- b) Será admissível supor que a 95% de confiança as precipitações médias nos dois casos são iguais? Explícite as hipóteses necessárias à resolução desta questão.
- c) Tendo em conta que, pela própria natureza do problema, o nível de precipitação sob coberto não deverá ser superior ao nível da correspondente precipitação a descoberto, indique um procedimento estatístico mais adequado para avaliar se existem diferenças significativas nos dois casos.
- ✓ **3.31.** Pretende-se testar se a proporção de ulmeiros afectados pela grafiose é idêntica em duas zonas A e B. Na zona A foi recolhida uma amostra aleatória de 30 ulmeiros e verificou-se que 20 estavam afectados pela grafiose. Na zona B recolheu-se uma amostra de 35 ulmeiros e verificou-se que 27 estavam afectados pela grafiose. Que conclusão se pode tirar ao nível de significância de 0.05?
- ✓ **3.32.** Diga, justificando, se cada uma das afirmações seguintes é verdadeira ou falsa:


- a) Num teste de hipóteses  $H_0 : \mu = 0$  vs.  $H_1 : \mu \neq 0$  em que se obteve um  $p\text{-value} = 0.034$ , não se rejeita a hipótese de a média populacional ser nula com um nível de significância de 0.01.
- b) Se  $]2.15; 3.24[$  é um intervalo de confiança a 95% para  $\sigma^2$  então a probabilidade de  $\sigma^2$  pertencer a este intervalo é igual a 0.95.

- ✓ 3.33. É desencadeado um programa de controlo da poluição de um rio em que são efectuadas medições, antes de lançar a campanha antipoluição e um ano após. As medições são combinações de vários índices; quanto maior for o valor resultante maior é a poluição. Obtiveram-se os seguintes resultados:

Ponto de controlo	1	2	3	4	5	6	7	8	9	10
Antes da campanha	68	88	101	82	96	74	65	74	52	99
Um ano após	67	87	90	76	98	69	68	65	59	70

Será que a campanha antipoluição reduziu de facto a poluição? Explícite e verifique todas as hipóteses necessárias à resolução do problema, justificando convenientemente.

- ✓ 3.34. (Exame 31.01.2011) Para comparar dois tipos de máquina de ceifar (segadeiras) quanto à sua eficiência, foram seleccionadas ao acaso 9 searas tendo sido cada uma dividida em dois lotes. Em cada seara uma das segadeiras foi atribuída ao acaso a um dos lotes, ficando a outra para o outro lote.

A eficiência é avaliada num intervalo de valores de 0 (eficiência mínima) a 10 (eficiência máxima). Os valores registados relativos à eficiência de cada segadeira foram introduzidos no *software* . Responda às seguintes questões utilizando os resultados de alguns dos comandos apresentados abaixo.

- a) Indique uma estimativa da eficiência média de cada uma das segadeiras.
- b) Calculou-se um intervalo a 95% de confiança para a variância da eficiência da segadeira 1, tendo-se obtido  $]0.5144; 4.1381[$ . Qual o valor observado para a variância da segadeira 1?
- c) Será que a segadeira 1 é mais eficiente do que a segadeira 2? Justifique convenientemente indicando e validando as condições necessárias à resolução.

```
> # seg1 designa os valores observados de eficiência da segadeira 1
> # seg2 designa os valores observados de eficiência da segadeira 2
```

```
> mean(seg1)      > mean(seg2)      > var(seg2)      > var(seg1-seg2)
[1] 7.4           [1] 6.488889      [1] 0.5861111    [1] 0.6761111
```

```
> shapiro.test(seg1)                               > shapiro.test(seg2)
  Shapiro-Wilk normality test                        Shapiro-Wilk normality test
data: seg1                                           data: seg2
W = 0.8848, p-value = 0.1762                         W = 0.9105, p-value = 0.3193
```

```
> shapiro.test(seg1-seg2)
```

```

        Shapiro-Wilk normality test
data:  seg1 - seg2
W = 0.9648, p-value = 0.8472

> t.test(seg1,seg2,paired=T,alternative="two.sided")
      Paired t-test
data:  seg1 and seg2
t = 3.3242, df = 8, p-value = 0.01047
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2790663 1.5431559
sample estimates:
mean of the differences
      0.9111111

> t.test(seg1,seg2,paired=T,alternative="greater")
      Paired t-test
data:  seg1 and seg2
t = 3.3242, df = 8, p-value = 0.005237
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.4014339      Inf
sample estimates:
mean of the differences
      0.9111111


> t.test(seg1,seg2,paired=T,alternative="less")
      Paired t-test
data:  seg1 and seg2
t = 3.3242, df = 8, p-value = 0.9948
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 1.420788
sample estimates:
mean of the differences
      0.9111111

> t.test(seg1,seg2,paired=F,alternative="two.sided")
      Welch Two Sample t-test
data:  seg1 and seg2
t = 2.088, df = 14.548, p-value = 0.05482
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02147036 1.84369258
sample estimates:
mean of x mean of y
 7.400000  6.488889

```

```
> t.test(seg1,seg2,paired=F,alternative="greater")
      Welch Two Sample t-test
data:  seg1 and seg2
t = 2.088, df = 14.548, p-value = 0.02741
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1446037      Inf
sample estimates:
mean of x mean of y
 7.400000  6.488889
```

```
> t.test(seg1,seg2,paired=F,alternative="less")
      Welch Two Sample t-test
data:  seg1 and seg2
t = 2.088, df = 14.548, p-value = 0.9726
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 1.677619
sample estimates:
mean of x mean of y
 7.400000  6.488889
```

✓ 3.35. (Exame 17.01.2011) Uma máquina de ensacar está regulada para encher sacos com 16 kg de açúcar. Para controlar o seu funcionamento escolheram-se, aleatoriamente, 15 sacos, que foram pesados. Os valores obtidos foram introduzidos no *software* . Responda às seguintes questões utilizando, sempre que possível, os resultados dos comandos apresentados.

- Indique uma estimativa para o peso médio e para a variância do peso de cada saco.
- Foi calculado um intervalo de confiança para a variância do peso de cada saco,  $\sigma^2$ , tendo-se obtido ]0.01368 ; 0.04931[.
  - Qual o estimador que foi utilizado na construção do intervalo?
  - Determine o nível de confiança do intervalo dado.
- Que conclusão pode tirar quanto à regulação da máquina? Justifique convenientemente, indicando e validando os pressupostos necessários à sua resposta.

```
> peso
[1] 16.1 15.8 15.9 16.1 15.8 16.2 16.0 15.9 16.0 15.7 15.8 15.7 16.0 16.0 15.8

> mean(peso)
[1] 15.92

> var(peso)
[1] 0.02314286

> shapiro.test(peso)
      Shapiro-Wilk normality test
data:  peso
W = 0.9377, p-value = 0.3544
```

- ✓ 3.36. Um enólogo pretende avaliar a acidez total de um vinho. Para isso selecciona aleatoriamente 20 garrafas de vinho na adega e mede a acidez através do método clássico e de um dispositivo de titulação automática. Alguns resultados das análises, em g/l, foram:

garrafa	1	2	3	...	18	19	20
método clássico	4.8	3.4	2.5	...	2.9	5.4	2.1
titulação automática	6.1	5.1	2.1	...	4.5	3.9	1.5

Os dados foram introduzidos no *software* . Abaixo apresentam-se resultados de comandos, alguns inadequados. Responda às seguintes questões utilizando os resultados apresentados abaixo.

- De acordo com a legislação em vigor um vinho de mesa deverá ter uma acidez total superior a 3.5 g/l. Com base nos resultados das análises efectuadas pelo método clássico, o enólogo poderá concluir que o seu vinho cumpre os requisitos de acidez impostos pela legislação? Explícite e valide os pressupostos necessários à resolução do problema.
- Com base nos valores obtidos poder-se-á concluir que os dois métodos de análise da acidez total do vinho têm resultados significativamente diferentes? Explícite e valide os pressupostos necessários à resolução do problema.

```
> classico
[1] 4.8 3.4 2.5 3.8 4.3 3.6 3.5 3.5 4.0 3.6 6.3 3.0 3.1 3.7 2.8
[16] 5.1 4.0 2.9 5.4 2.1
> automatico
[1] 6.1 5.1 2.1 4.9 6.6 4.0 4.5 1.0 4.7 3.5 8.2 3.9 3.6 4.5 3.5
[16] 6.3 4.7 4.5 3.9 1.5

> var(classico)          > var automatico)          > var(classico-automatico)
[1] 1.040105              [1] 2.887868              [1] 1.326605

> shapiro.test(classico)                > shapiro.test automatico)
Shapiro-Wilk normality test            Shapiro-Wilk normality test
data: classico                          data: automatico
W = 0.951, p-value = 0.3827             W = 0.9625, p-value = 0.5959

> shapiro.test(classico-automatico)
Shapiro-Wilk normality test
data: classico - automatico
W = 0.9163, p-value = 0.08413

> t.test(classico,mu=3.5)
One Sample t-test
data: classico
t = 1.184, df = 19, p-value = 0.2510
alternative hypothesis: true mean is not equal to 3.5
95 percent confidence interval:
```

```

3.292693 4.247307
sample estimates:
mean of x
  3.77

> t.test(classico,mu=3.5,alternative="greater")
      One Sample t-test
data:  classico
t = 1.184, df = 19, p-value = 0.1255
alternative hypothesis: true mean is greater than 3.5
95 percent confidence interval:
 3.375677      Inf
sample estimates:
mean of x
  3.77

> t.test(classico,mu=3.5,alternative="less")
      One Sample t-test
data:  classico
t = 1.184, df = 19, p-value = 0.8745
alternative hypothesis: true mean is less than 3.5
95 percent confidence interval:
 -Inf 4.164323
sample estimates:
mean of x
  3.77

> t.test(classico,automatico,paired=FALSE, var.equal=TRUE)
      Two Sample t-test
data:  classico and automatico
t = -1.32, df = 38, p-value = 0.1947
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.4821486  0.3121486
sample estimates:
mean of x mean of y
  3.770    4.355

> t.test(classico,automatico,paired=TRUE)
      Paired t-test
data:  classico and automatico
t = -2.2714, df = 19, p-value = 0.03493
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.12405128 -0.04594872
sample estimates:
mean of the differences
      -0.585

```

- ✓ 3.37. 10 pessoas asmáticas participaram numa experiência destinada a estudar o efeito de um novo tratamento da função pulmonar. Durante 1 segundo registou-se o volume de ar expirado (em litros) antes e depois da aplicação do tratamento. Os resultados obtidos foram os seguintes:

	1	2	3	4	5	6	7	8	9	10
Antes	1.69	2.77	1.00	1.66	0.85	1.42	2.82	2.58	1.98	2.02
Depois	2.69	2.22	3.07	3.35	2.74	3.61	5.14	2.44	2.25	3.41

**Nota: Para responder a esta pergunta utilize, sempre que possível, resultados que lhe são apresentados no Anexo.**

- Determine um intervalo a 99% de confiança para o volume médio de ar expirado por uma pessoa antes da aplicação do tratamento. Indique e valide os pressupostos necessários à sua resolução.
- Qual é o erro máximo cometido ao usar a média da amostra para estimar o volume médio de ar expirado por pessoa, com o nível de confiança da alínea anterior?
- Os dados recolhidos dão indicação de que o volume de ar expirado é superior após o tratamento? Justifique, indicando e validando os pressupostos necessários à resolução desta questão.

#### ANEXO

```
>antes<-c(1.69,2.77,1.00,1.66,.85,1.42,2.82,2.58,1.98,2.02)
>depois<-c(2.69,2.22,3.07,3.35,2.74,3.61,5.14,2.44,2.25,3.41)
```

```
> var(antes);var(depois)
[1] 0.4807433
[1] 0.7567511
```

```
> var(antes-depois)
[1] 1.056112
```

```
> shapiro.test(antes)
```

Shapiro-Wilk normality test

```
data: antes
W = 0.9402, p-value = 0.5558
```

```
> shapiro.test(depois)
```

Shapiro-Wilk normality test



```
data: depois
W = 0.8633, p-value = 0.08348
```

```
> shapiro.test(antes-depois)
```

```
Shapiro-Wilk normality test
```

```
data: antes - depois
W = 0.8972, p-value = 0.2039
```

```
> t.test(antes,depois,paired=F,var.equal=T)
```

```
Two Sample t-test
```

```
data: antes and depois
t = -3.4482, df = 18, p-value = 0.002868
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.9520632 -0.4739368
sample estimates:
mean of x mean of y
1.879      3.092
```

```
> t.test(antes,depois,paired=F,var.equal=T,alternative="less")
```

```
Two Sample t-test
```

```
data: antes and depois
t = -3.4482, df = 18, p-value = 0.001434
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf -0.6029904
sample estimates:
mean of x mean of y
1.879      3.092
```

```
> t.test(antes,depois,paired=F,var.equal=T,alternative="greater")
```

```
Two Sample t-test
```

```
data: antes and depois
t = -3.4482, df = 18, p-value = 0.9986
alternative hypothesis: true difference in means is greater than 0
```

95 percent confidence interval:

-1.823010            Inf

sample estimates:

mean of x mean of y

1.879            3.092

> t.test(antes,depois,paired=T)

Paired t-test

data: antes and depois

t = -3.7326, df = 9, p-value = 0.004679

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.9481531 -0.4778469

sample estimates:

mean of the differences

-1.213

> t.test(antes,depois,paired=T,alternative="less")

Paired t-test

data: antes and depois

t = -3.7326, df = 9, p-value = 0.002339

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -0.6172772

sample estimates:

mean of the differences

-1.213

> t.test(antes,depois,paired=T,alternative="greater")

Paired t-test

data: antes and depois

t = -3.7326, df = 9, p-value = 0.9977

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-1.808723            Inf

sample estimates:

mean of the differences

- ✓ **3.38.** Uma associação de defesa de consumidores pretende publicar um estudo comparativo dos preços de dois supermercados. Para isso seleccionou 20 produtos e registou os preços de cada um nos supermercados A e B (em cêntimos):

Produto	1	2	3	4	5	6	...	15	16	17	18	19	20
Super. A	202	201	560	253	384	332	...	624	851	742	501	476	765
Super. B	185	187	516	239	349	295	...	613	851	731	546	490	762

**Utilize os resultados apresentados abaixo para responder às seguintes questões.**

- Determine uma estimativa para a diferença dos preços médios nos dois supermercados.
- Com base nos dados recolhidos, o que pode afirmar relativamente aos preços médios praticados pelos dois supermercados? Explícite e valide as hipóteses necessárias à resolução do problema.


```

> A
[1] 202 201 560 253 384 332 549 722 153 676 804 535 472 335 624 851
[17] 742 501 476 765
> B
[1] 185 187 516 239 349 295 552 667 132 676 745 529 460 316 613 851
[17] 731 546 490 762

> mean(A)      > mean(B)      > var(A)      > var(B)      > var(A-B)
[1] 506.85      [1] 492.05      [1] 46116.77   [1] 46454.68   [1] 568.379

> shapiro.test(A)                > shapiro.test(B)
  Shapiro-Wilk normality test      Shapiro-Wilk normality test
data: A                             data: B
W = 0.957, p-value = 0.4864        W = 0.9552, p-value = 0.4521

> shapiro.test(A-B)
  Shapiro-Wilk normality test
data: A - B
W = 0.9514, p-value = 0.3889
    
```

- ✓ **3.39.** (Exame 27.01.2014) Uma empresa agrícola pretende adquirir um gerador. Face a dois tipos de geradores (que designamos por ger1 e ger2), de preços semelhantes, quer decidir comparando a produção de energia eléctrica, em kWh, de cada um dos geradores. Recolheu uma amostra de 25 observações da quantidade de energia eléctrica produzida por cada gerador, em diferentes situações. Os dados foram introduzidos no *software* . Considere os seguintes resultados referentes ao gerador tipo 1:

```

>var(ger1); mean(ger1)
[1] 18.90528
[1] 9.6056
>shapiro.test(ger1)
Shapiro-Wilk normality test
data: ger1
W = 0.9514, p-value = 0.2698

```

- Indique uma estimativa da variância da produção de energia eléctrica do gerador 1. Explícite o estimador que lhe está associado.
- Poder-se-á considerar que a variância da produção de energia eléctrica no gerador 1 é superior a  $15(\text{kWh})^2$ ? Justifique convenientemente a sua resposta.
- Para obter um intervalo de confiança a 95% para a diferença entre as produções médias de energia eléctrica dos geradores 1 e 2, usou-se o comando do R que se apresenta de seguida e respectivo resultado (no qual três valores foram substituídos por ????):

```
>t.test(ger1,ger2,paired=???,var.equal=T)
```


Two Sample t-test

```

data: ger1 and ger2
t = -0.6632, df = ????, p-value = 0.5104
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.009280  ???
sample estimates:
mean of x mean of y
 9.6056  10.3520

```

- Qual a diferença entre as médias observadas para a produção de energia eléctrica dos dois geradores? Qual o erro máximo que se cometeria, a 95% de confiança, ao identificar esta diferença com a verdadeira diferença entre as produções médias de energia eléctrica dos geradores?
- Complete o *output* indicando a informação em falta nos três locais assinalados por ????
- Que pressupostos deverão ser verificados para construir o intervalo de confiança dado no *output*?
  - O que pode afirmar, a 95% de confiança, sobre a diferença entre as produções médias de energia eléctrica dos geradores?
  - A conclusão da alínea anterior seria alterada se considerasse uma confiança de 99%? Justifique.

- ✓ **3.40.** (Exame 26.01.2015) Para estudar o peso de uma certa espécie de peixe, recolheram-se dados do peso (g) de 15 peixes fêmea e de 15 peixes macho. Os resultados obtidos foram introduzidos no software . Utilize, sempre que possível, os resultados apresentados no Anexo II para responder às seguintes questões.

- a) Pretende-se estudar a variabilidade do peso dos peixes fêmea.
- i) Face aos dados recolhidos indique uma estimativa da variância do peso dos peixes fêmea.
  - ii) Qual o estimador associado à estimativa obtida em i)? Explícite-o e mostre que é um estimador centrado da variância populacional.
- b) Justificando convenientemente, diga se os dados recolhidos são compatíveis com as seguintes conjecturas:
- i) O peso de um peixe fêmea daquela espécie segue uma distribuição normal.
  - ii) O peso médio de um peixe fêmea daquela espécie é 1000 g.
- c) Comente, justificando convenientemente, a seguinte afirmação “O peso médio das fêmeas é superior ao dos machos”.

## ANEXO

```

> femea
[1] 1007 965 ... 978 1020
> macho
[1] 934 1014 ... 1017 1023

> var(femea)          > var(macho)          > var(femea-macho)
[1] 679.5714          [1] 765.6952          [1] 1629.695

> shapiro.test(femea)
      Shapiro-Wilk normality test
data:  femea
W = 0.9545, p-value = 0.5977

> shapiro.test(macho)
      Shapiro-Wilk normality test
data:  macho
W = 0.9371, p-value = 0.3477

> shapiro.test(femea-macho)
      Shapiro-Wilk normality test
data:  femea - macho
W = 0.9638, p-value = 0.7581

> var.test(femea, macho)
      F test to compare two variances
data:  femea and macho
F = 0.8875, num df = 14, denom df = 14, p-value = 0.8265
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2979674 2.6435622
sample estimates:
ratio of variances
 0.8875221

```

```

> t.test(femea,mu=1000,alternative="greater")
      One Sample t-test
data:  femea
t = 1.6343, df = 14, p-value = 0.06224
alternative hypothesis: true mean is greater than 1000
95 percent confidence interval:
 999.1448      Inf
sample estimates:
mean of x
      1011

> t.test(femea,mu=1000)
      One Sample t-test
data:  femea
t = 1.6343, df = 14, p-value = 0.1245
alternative hypothesis: true mean is not equal to 1000
95 percent confidence interval:
 996.5637 1025.4363
sample estimates:
mean of x
      1011

> t.test(femea, macho, paired=F, var.equal=T, alternative="greater")
      Two Sample t-test
data:  femea and macho
t = 2.8389, df = 28, p-value = 0.004167
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 11.1686      Inf
sample estimates:
mean of x mean of y
1011.0000  983.1333


> t.test(femea, macho, paired=T, alternative="greater")
      Paired t-test
data:  femea and macho
t = 2.6735, df = 14, p-value = 0.009088
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
  9.507904      Inf
sample estimates:
mean of the differences
      27.86667

```

- ✓ 3.41. (Exame 25.01.2019) Realizou-se um estudo de investigação em modelos animais (ratinhos), onde se pretende conhecer a importância de algumas proteínas do sistema imune na infecção com malária na gravidez. Quantificou-se a expressão relativa dos genes associados a uma proteína, A, em placentas de 6 fêmeas não infectadas (NI) e 6 fêmeas infectadas (INF), tendo-se obtido os seguintes indicadores:

	Média	Variância
NI	1.520	0.612
INF	0.347	0.131

O cientista suspeita que a expressão relativa dos genes da proteína A pode sofrer alterações, em média, com a infecção.

Os resultados da experiência foram analisados com recurso ao *software*  e estão apresentados no Anexo Utilize-os, sempre que possível, para responder de forma completa às questões apresentadas a seguir.

- Utilizando um intervalo de confiança adequado (consultar o Anexo) esclareça a dúvida do cientista. Responda de forma completa à questão, justificando convenientemente a sua resposta.
- Um parâmetro de interesse neste estudo é a variabilidade da expressão relativa dos genes quando há infecção.
  - Indique o estimador associado ao estudo desse parâmetro?
  - Com base nos dados recolhidos poder-se-á admitir que o valor daquele parâmetro é inferior a 0.3? Justifique convenientemente.

#### ANEXO

```
> var.test(A_NI, A_INF)
      F test to compare two variances
data:  A_NI and A_INF
F = 4.6731, num df = 5, denom df = 5, p-value = 0.1159
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6539058 33.3954743
sample estimates:
ratio of variances
 4.67306

> shapiro.test(A_NI)
Shapiro-Wilk normality test

> shapiro.test(A_INF)
Shapiro-Wilk normality test
```

```

data: A_NI                      data: A_INF
W = 0.92543, p-value = 0.5453   W = 0.79852, p-value = 0.05699

```

```
> shapiro.test(placenta$A_NI-placenta$A_INF)
```

Shapiro-Wilk normality test

```

data: placenta$A_NI - placenta$A_INF
W = 0.9378, p-value = 0.6414

```

```
> t.test(A_NI,A_INF, paired=F, var.equal=T,alternative="two.sided")
```

Two Sample t-test

```

data: A_NI and A_INF
t = 3.334, df = 10, p-value = 0.007566
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.3891588      ????
sample estimates:
mean of x      mean of y
1.5199600      0.3467328

```


```
> t.test(A_NI, A_INF, paired=T, var.equal=T,alternative="two.sided")
```

Paired t-test

```


data: A_NI and A_INF      t = 2.66, df = 5, p-value = 0.04488
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.03945195      ??????
sample estimates:
mean of the differences
1.173227

```

- ✓ **3.42.** (*Exame 09.01.2020*) Num estudo sobre o efeito da congelação no valor nutricional de legumes e frutos, determinou-se o valor de um índice nutricional por caloria ingerida para cada um de 13 alimentos (legumes ou frutos) frescos e congelados. Os dados obtidos (ilustrados abaixo) foram introduzidos nos objectos `fresco` e `congel` do *software* .

Legume/Fruto	amora	espinafre	...	pêssego
Fresco	0.731	3.903	...	0.657
Congelado	0.490	3.490	...	0.865



Responda às seguintes questões utilizando, sempre que possível, o *output* do  apresentado em Anexo.

- a) Face à amostra observada obteve-se o intervalo de confiança ]1.226, 2.223[ para o valor esperado do índice nutricional,  $\mu$ , de um alimento fresco. Responda às seguintes questões apresentando os cálculos efectuados:
- Determine o índice nutricional médio na amostra dos alimentos frescos.
  - Qual o erro máximo cometido ao usar a média da amostra para estimar  $\mu$ .
  - Determine a confiança deste intervalo, justificando convenientemente.
- b) Poder-se-á afirmar que a congelação dos legumes e frutos provoca uma redução do índice nutricional por caloria ingerida? Explícite e valide os pressupostos necessários.

## ANEXO

```
> fresco <- c(0.731, 3.903, ... 0.657)
> congel <- c(0.490, 3.490, ... 0.865)

> var(fresco)           > var(congel)           > var(fresco-congel)
[1] 1.017156             [1] 0.8771506           [1] 0.09266823

> shapiro.test(fresco)      > shapiro.test(congel)
  Shapiro-Wilk normality test      Shapiro-Wilk normality test
data: fresco                      data: congel
W = 0.89088, p-value = 0.1003      W = 0.92141, p-value = 0.2621

> shapiro.test(fresco-congel)
  Shapiro-Wilk normality test
data: fresco - congel
W = 0.89526, p-value = 0.1151

> t.test(fresco, congel, paired=TRUE)
  Paired t-test
data: fresco and congel
t = 3.21, df = 12, p-value = 0.007493
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.0870592 0.4549710
sample estimates:
mean of the differences
 0.2710151

> t.test(fresco, congel, paired=TRUE, alternative="greater")
  Paired t-test
data: fresco and congel
t = 3.21, df = 12, p-value = 0.003746
```

```
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1205376      Inf
```

```
> t.test(fresco, congel, paired=TRUE, alternative="less")
Paired t-test
data: fresco and congel
t = 3.21, df = 12, p-value = 0.9963
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf 0.4214926
```

```
> t.test(fresco, congel, paired=FALSE, var.equal=TRUE)
Two Sample t-test
data: fresco and congel
t = 0.70997, df = 24, p-value = 0.4846
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.5168317  1.0588619
```

```
> t.test(fresco, congel, paired=FALSE, var.equal=TRUE, alternative="greater")
Two Sample t-test
data: fresco and congel
t = 0.70997, df = 24, p-value = 0.2423
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-0.3820756      Inf
```

```
> t.test(fresco, congel, paired=FALSE, var.equal=TRUE, alternative="less")
Two Sample t-test
data: fresco and congel
t = 0.70997, df = 24, p-value = 0.7577
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf 0.9241058
```

## Exercícios de Revisão de Inferência Estatística

**R3.1.** (Exame 9.01.2020) Considere a v.a.  $X$  com distribuição uniforme contínua no intervalo  $]0, \theta[$ ,  $\theta > 0$ . Seja  $\bar{X}$  a média da amostra aleatória  $(X_1, \dots, X_n)$  e  $T = 2\bar{X}$  um estimador de  $\theta$ .

- Defina “amostra aleatória de dimensão  $n$ ”.
- Mostre que  $T$  é um estimador centrado de  $\theta$ .
- Determine  $Var[T]$ .

**R3.2.** Recolheu-se uma amostra aleatória de dimensão 5 de uma população normal. Determine a probabilidade de o desvio padrão da amostra ser inferior ao desvio padrão da população.

**R3.3.** Seja  $X$  uma variável aleatória com distribuição binomial,  $X \cap B(n, p)$ . Considere os estimadores para  $p$  assim definidos  $\hat{P}_1 = \frac{X+1}{n+1}$  e  $\hat{P}_2 = \frac{X}{n+1}$ .

- ⊗ a) Determine, em função dos parâmetros da distribuição de  $X$ , o valor médio e a variância dos estimadores  $\hat{P}_1$  e  $\hat{P}_2$ . Qual deles preferiria? Justifique.
- b) Suponha que numa amostra de dimensão 20, se observou a ocorrência de 5 sucessos. Determine uma estimativa para  $p$ .

**R3.4.** (Exame 12.01.2009) A quantidade (em ppm) de um poluente no solo de uma certa região é uma v.a.  $X$  que se admite ter distribuição normal com valor médio  $\mu$  e variância  $\sigma^2$ , desconhecidos.

Recolheu-se uma amostra de dimensão 10 tendo-se obtido o desvio padrão de 1.3943 ppm. Com base na amostra construiu-se o seguinte intervalo de confiança para  $\mu$

$$]22.447 ; 25.313[$$

- Indique estimativas do valor médio e da variância da quantidade daquele poluente no solo.
- Qual o grau de confiança daquele intervalo?
- Construa um intervalo a 95% de confiança para  $\sigma^2$ .

**R3.5.** Num certo processo químico é muito importante que uma dada solução tenha um pH de exactamente 8.20. O método utilizado na determinação do pH fornece medições que se admite terem distribuição normal de valor médio igual ao verdadeiro valor do pH da solução e desvio padrão de 0.02.

Para avaliar o pH de uma solução, efectuaram-se 10 medições independentes tendo-se obtido os seguintes valores:

8.18 8.16 8.17 8.22 8.19 8.17 8.15 8.21 8.16 8.18

- Indique uma estimativa do valor médio do pH da solução.
- Com base nestas 10 medições, o que pode concluir relativamente à utilização desta solução no referido processo químico?

- c) Pretende-se efectuar um novo conjunto de medições para diminuir o erro máximo cometido na estimativa do verdadeiro valor do pH da solução. Mantendo-se todas as condições referidas acima, qual deverá ser o tamanho da amostra para que aquele erro máximo não exceda 0.01, a 95% de confiança?

**R3.6.** Faz-se uma experiência para saber se dois regimes alimentares A e B produzem o mesmo aumento de peso nos animais, durante um período de tempo fixado. Tomam-se 20 animais e de entre eles 10 ao acaso aos quais é dado o alimento A. Aos outros 10 é dado o alimento B. Os aumentos de peso (expressos em kg) no mesmo intervalo de tempo são os seguintes:

Regime A	-2.0	0.0	4.2	6.3	9.6	4.3	10.2	11.0	12.4	13.1
Regime B	4.0	6.0	8.0	11.3	12.3	14.4	14.5	14.7	14.7	16.0

Diga se existe diferença significativa entre os dois regimes alimentares, justificando convenientemente todas as hipóteses necessárias à resolução do problema.

**R3.7.** Um estudo pretende comparar um tipo de semente melhorada com o tipo de semente usado anteriormente. A nova semente passará a ser utilizada se, em média, o crescimento das plantas após 20 dias for superior ao das obtidas das 'velhas' sementes. São criadas 15 diferentes situações laboratoriais, variando temperatura e humidade. Em cada situação semeia-se uma semente de cada tipo e obtêm-se os seguintes resultados para o crescimento (em cm) das plantas após 20 dias :

Situação	1	2	3	4	5	6	7	8
Sementes melhoradas	3.46	3.48	2.74	2.83	4.00	4.95	2.24	6.92
sementes tradicionais	3.18	3.67	2.92	3.10	4.10	4.86	2.21	6.91
Situação	9	10	11	12	13	14	15	
sementes melhoradas	6.57	6.18	8.30	3.44	4.47	7.59	3.87	
sementes tradicionais	6.83	6.19	8.05	3.46	4.18	7.43	3.85	

Deverá passar a usar-se as sementes melhoradas? Responda justificando e explicitando quaisquer hipóteses adicionais que seja necessário impôr.


**R3.8.** Num estudo sobre a incidência de certa doença numa população de insectos, um grupo de biólogos registou ao longo de um ano o número de insectos contaminados em cada amostra de 5 insectos, tendo para tal recolhido 100 amostras. Os resultados obtidos foram:

Num. de insectos contaminados	0	1	2	3	4	5
Num. de amostras	9	26	34	22	8	1

- Construa uma tabela de frequências absolutas, relativas e relativas acumuladas para o número de insectos contaminados por amostra.
- Determine a média e a mediana do número de insectos contaminados por amostra.
- Considerando agora a totalidade de insectos observados como uma amostra aleatória da população de insectos:
  - Determine uma estimativa da verdadeira proporção de insectos contaminados.

ii) Qual é o erro máximo associado à estimativa obtida na alínea anterior, com uma confiança de 95% ?

- R3.9.** O dono de uma ervanária produz um chá relativamente ao qual afirma que é 90% eficaz para curar dores de cabeça. Num inquérito feito a 250 pessoas, 198 concordaram que o chá cura as dores de cabeça. Acha que o resultado do inquérito é compatível com a pretensão do produtor ?
- R3.10.** Um investigador pretende estudar a incidência a nível nacional, de uma doença que ataca os pinheiros. Observações efectuadas através do país resultaram em 1233 casos de pinheiros afectados (a nível nacional) num total de 4250 observações.
- a) Estime a percentagem de pinheiros afectados a nível nacional.
  - b) Determine um intervalo a 95% de confiança para a verdadeira proporção de pinheiros afectados.

**R3.11.** (*Exame 16.01.2012*) Em várias espécies de animais, altos níveis de testosterona tornam os machos mais atractivos para as fêmeas. No entanto, conjectura-se que altos níveis de testosterona podem enfraquecer o sistema imunitário, i.e., reduzem o número de anticorpos. Para responder a esta questão consideraram-se 13 melros de asa-vermelha, em cada um dos quais foi implantado um tubo contendo testosterona. Em cada melro avaliou-se o número de anticorpos no sangue antes e após o implante. Os resultados obtidos, em unidades convenientes, foram introduzidos no *software* .

Utilize, sempre que possível, os resultados apresentados no **Anexo** para responder às seguintes questões (alguns dos procedimentos apresentados são inadequados).

- a) Indique estimativas para o valor médio do número de anticorpos antes e depois do implante.
- b) Determine um intervalo de confiança a 95% para a variabilidade do número de anticorpos antes da introdução do implante.
- c) A conjectura de que o aumento do nível de testosterona enfraquece o sistema imunitário será compatível com os resultados obtidos? Explícite e valide os pressupostos necessários à resolução do problema.

## ANEXO

```
antes <-  
c(4.65, 4.21, 4.91, 4.50, 4.80, 5.01, 4.88, 4.78, 4.98, 4.41, 4.75, 4.70, 4.93)  
depois <-  
c(4.54, 3.80, 4.98, 4.45, 5.00, 5.00, 4.35, 4.96, 5.02, 4.73, 4.77, 4.60, 5.01)  
  
> var(antes)           > var(depois)  
[1] 0.05638077        [1] 0.1299974  
  
> shapiro.test(antes)
```

```

                Shapiro-Wilk normality test
data:  antes
W = 0.9196, p-value = 0.2477

> shapiro.test(depois)
                Shapiro-Wilk normality test
data:  depois
W = 0.831, p-value = 0.01631

> shapiro.test(antes-depois)
                Shapiro-Wilk normality test
data:  antes - depois
W = 0.9169, p-value = 0.2275

> var.test(antes,depois)
                F test to compare two variances
data:  antes and depois
F = 0.4337, num df = 12, denom df = 12, p-value = 0.1622
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1323375 1.4213773
sample estimates:
ratio of variances
 0.4337068

> t.test(antes,depois,paired=F,var.equal=T,alternative="two.sided")
                Two Sample t-test
data:  antes and depois
t = 0.1927, df = 24, p-value = 0.8488
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2240467  0.2702005
sample estimates:
mean of x mean of y
 4.731538  4.708462

> t.test(antes,depois,paired=F,var.equal=T,alternative="less")
                Two Sample t-test
data:  antes and depois
t = 0.1927, df = 24, p-value = 0.5756
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 0.2279316

```

```

sample estimates:
mean of x mean of y
  4.731538  4.708462

> t.test(antes,depois,paired=F,var.equal=T,alternative="greater")
      Two Sample t-test
data:  antes and depois
t = 0.1927, df = 24, p-value = 0.4244
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.1817778      Inf
sample estimates:
mean of x mean of y
  4.731538  4.708462

> t.test(antes,depois,paired=T,alternative="two.sided")
      Paired t-test
data:  antes and depois
t = 0.3562, df = 12, p-value = 0.7279
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1180915  0.1642453
sample estimates:
mean of the differences
      0.02307692

> t.test(antes,depois,paired=T,alternative="less")
      Paired t-test
data:  antes and depois
t = 0.3562, df = 12, p-value = 0.6361
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 0.1385539
sample estimates:
mean of the differences
      0.02307692

> t.test(antes,depois,paired=T,alternative="greater")
      Paired t-test
data:  antes and depois
t = 0.3562, df = 12, p-value = 0.3639
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:

```


```

-0.09240003          Inf
sample estimates:
mean of the differences
          0.02307692

```

**R3.12.** Um agricultor pretende experimentar dois tipos (I e II) de semeador de milho para comparar a produtividade das duas máquinas.

Para isso, num campo agrícola marcaram-se aleatoriamente 10 parcelas de igual área sendo cada uma delas dividida em duas secções iguais. Em cada parcela sorteou-se a atribuição de um dos tipos de semeador a uma das secções.

A produtividade (em unidades adequadas) de cada um dos semeadores foi registada, e os dados foram introduzidos no *software* . Utilizando os resultados apresentados abaixo (alguns desnecessários), poderá o agricultor admitir que a produtividade esperada das duas máquinas é igual? Justifique convenientemente a sua resposta.

```

>x<-c(5.6, 8.4, 8.0, 8.6, 6.4, 7.8, 6.2, 7.7, 8.0, 7.7)
observações para o semeador tipo I.
>y<-c(6.0, 7.4, 7.3, 7.5, 6.4, 6.0, 5.5, 6.6, 5.6, 6.1)
observações para o semeador tipo II.

```

```

>mean(x)          >mean(y)          >mean(x-y)
[1] 7.44          [1] 6.44          [1] 1

```

```

>var(x)           >var(y)           >var(x-y)
[1] 1.018222      [1] 0.5448889     [1] 0.68

```

```

> shapiro.test(x)

```

```

      Shapiro-Wilk normality test

```

```

data:  x
W = 0.8731, p-value = 0.1087

```

```

> shapiro.test(y)

```

```

      Shapiro-Wilk normality test

```

```

data:  y
W = 0.9026, p-value = 0.2341

```

```

> shapiro.test(x-y)

```

```

      Shapiro-Wilk normality test

```



data: x - y  
W = 0.9765, p-value = 0.9435

**R3.13.** (Exame 31.01.2012) Um estudante de agronomia pretende estudar a produção de uma certa variedade de milho numa dada região. Para esse efeito, feita a colheita, pesou 50 espigas e organizou os resultados na tabela abaixo:

Peso(g)	$x'_i$	$n_i$	$N_i$	$f_i$	$F_i$
] 180, 190]	185	3			
] 190, 200]			12		
] 200, 210]		9		0.18	
] 210, 220]			30		
] 220, 230]		12			0.84
] 230, 240]		8	50		

- Identifique e classifique, justificando, a variável estatística em estudo.
- Diga o que representa o símbolo no cabeçalho de cada coluna e complete a tabela. Esboce o histograma associado.
- Calcule um valor aproximado da média e da mediana do peso das espigas.
- Indique uma estimativa da percentagem de espigas com peso superior a 220 g.
- O estudante tirou a seguinte conclusão: “Esta variedade de milho tem uma boa produção porque mais de 30% das espigas têm peso superior a 220 g”. Está de acordo com o seu colega? Justifique convenientemente.
- Suponha que outro estudante quer estudar a produção dessa variedade de milho noutra região. Para isso decidiu colher e pesar 50 espigas verificando que 22 tinham peso superior a 220 g. Poder-se-á dizer que a percentagem de espigas com peso superior a 220 g é diferente nas duas regiões? Justifique.

**R3.14.** Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória retirada de uma população com distribuição normal com valor médio  $\mu$  e variância  $\sigma^2$  conhecida.

- Qual é a probabilidade de o intervalo aleatório

$$\left] \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right[$$

conter o valor  $\mu$ ?

- Qual é o erro máximo cometido ao usar  $\bar{x}$  para estimar  $\mu$ ?
- Qual é a redução do erro máximo que se obtém quando a dimensão da amostra duplica?
- Mostre que  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  é um estimador centrado de  $\sigma^2$ .

**R3.15.** (Exame 14.01.2013) Seja  $X$  uma variável aleatória com valor esperado  $\mu$  e variância  $\sigma^2$  desconhecidos. Recolheram-se aleatoriamente 15 dados da população  $X$ , que se introduziram no *software*  $\mathbb{R}$  no objecto  $x$ . Responda às seguintes questões utilizando, sempre que possível, o *output* apresentado em Anexo.

- a) Pretende-se estimar o valor médio,  $\mu$ , da população  $X$ .
  - i) Mostre que  $\bar{X}$ , média de uma amostra aleatória de dimensão  $n$  retirada de  $X$ , é um estimador centrado de  $\mu$ . Exprima a variância de  $\bar{X}$  como função dos parâmetros da população.
  - ii) Com base na amostra recolhida indique estimativas do valor médio de  $X$  e da variância de  $\bar{X}$ .
- b) Pretende-se averiguar se a variância da população é superior a 1, pelo que se decide realizar um teste de hipóteses.
  - i) Formule as hipóteses do teste que permite responder a esta questão.
  - ii) Indique a estatística de teste e a sua distribuição. Valide os pressupostos necessários à existência dessa distribuição.
  - iii) Conclua o teste de hipóteses.
- c) Recolheu-se uma amostra de dimensão 15 de outra população  $Y$ . Os dados foram introduzidos no objecto  $y$  do  $\mathbb{R}$ .
  - i) Indique um intervalo de confiança a 95% para a diferença das médias das populações  $X$  e  $Y$ . Explícite e valide os pressupostos necessários.
  - ii) Os dados recolhidos evidenciam que as médias das duas populações são iguais? Justifique convenientemente.

#### ANEXO

```

> mean(x)           > mean(y)
[1] 5.686667        [1] 4.646667

> var(x)           > var(y)
[1] 1.084095        [1] 1.855524

> shapiro.test(x)           > shapiro.test(y)
  Shapiro-Wilk normality test      Shapiro-Wilk normality test
data:  x                          data:  y
W = 0.9747, p-value = 0.6733      W = 0.9577, p-value = 0.2704

> shapiro.test(x-y)
  Shapiro-Wilk normality test
data:  x - y
W = 0.924, p-value = 0.2214

> var.test(x, y)

```

F test to compare two variances

```
data: x and y
F = 0.5843, num df = 14, denom df = 14, p-value = 0.3261
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.196151 1.740249
sample estimates:
ratio of variances
 0.5842529

> t.test(x,y,paired=T)
    Paired t-test
data: x and y
t = 2.1864, df = 14, p-value = 0.04627
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.01977868 2.06022132
sample estimates:
mean of the differences
      1.04

> t.test(x,y,paired=F,var.equal=T)
    Two Sample t-test
data: x and y
t = 2.3493, df = 28, p-value = 0.0261
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1331903 1.9468097
sample estimates:
mean of x mean of y
 5.686667  4.646667
```

**R3.16.** (Exame 23.01.2020) Num inventário florestal de montado de sobre pretende-se comparar o diâmetro das árvores em duas parcelas, A e B. Para isso em cada uma das parcelas registou-se o diâmetro de 25 árvores, em cm. Os dados obtidos foram introduzidos nos objectos `parcelaA` e `parcelaB` do software  $\mathbb{R}$ .

Responda às seguintes questões utilizando, sempre que possível, o *output* do  $\mathbb{R}$  apresentado em Anexo.

- a) Considere a v.a.  $X$ , de valor médio  $\mu$  e variância  $\sigma^2$  desconhecidos, que designa o diâmetro dos sobreiros na parcela A. Seja  $(X_1, \dots, X_{25})$  a amostra aleatória retirada de  $X$ . Sugere-se dois estimadores para  $\mu$ ,

$$T_1 = \frac{\sum_{i=1}^{25} X_i}{25} \quad \text{e} \quad T_2 = \frac{2X_1 + 8X_{25}}{10},$$

- i) Mostre que  $T_1$  e  $T_2$  são estimadores centrados de  $\mu$  e obtenha as respectivas variâncias.
- ⊗ ii) Qual dos dois estimadores deve ser escolhido para estimar  $\mu$ ? Justifique.
- iii) Com base nos dados recolhidos indique estimativas para o diâmetro médio dos sobreiros na parcela A, associadas aos estimadores  $T_1$  e  $T_2$ .
- iv) Determine um intervalo de confiança a 95% para a variância do diâmetro dos sobreiros na parcela A. Responda de forma completa.
- b) Face aos resultados obtidos poder-se-á concluir que as árvores da parcela B, apresentam em média um diâmetro superior ao dos sobreiros da parcela A? Explícite e valide os pressupostos necessários à resolução do problema.

## ANEXO

```

> parcelaA
[1] 45.00 36.61 41.70 ... 32.47 38.52 38.20
[13] 36.92 21.96 19.42 ... 19.10 52.20 28.65
[25] 39.79
> parcelaB
[1] 33.74 58.89 32.15 ... 45.84 49.02 29.60
[13] 58.57 47.75 62.39 ... 36.29 37.56 41.06
[25] 29.92

> mean(parcelaA)      > mean(parcelaB)      > var(parcelaA)      >var(parcelaB)
[1] 35.152             [1] 43.4812             [1] 82.57123         [1] 93.6749

> shapiro.test(parcelaA)                               > shapiro.test(parcelaB)

      Shapiro-Wilk normality test                        Shapiro-Wilk normality test
data:  parcelaA                                         data:  parcelaB
W = 0.9352, p-value = 0.1147                            W = 0.9757, p-value = 0.7895

> shapiro.test(parcelaA-parcelaB)

      Shapiro-Wilk normality test
data:  parcelaA - parcelaB
W = 0.9528, p-value = 0.2890

> var.test(parcelaA,parcelaB)

      F test to compare two variances

data:  parcelaA and parcelaB
F = 0.8815, num df = 24, denom df = 24, p-value = 0.7598

```

```
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3884346 2.0002904
sample estimates:
ratio of variances
      0.8814658
```

```
> t.test(parcelaA,parcelaB,var.equal=T,paired=F,alternative="greater")
```

Two Sample t-test

```
data: parcelaA and parcelaB
t = -3.137, df = 48, p-value = 0.9985
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-12.78249      Inf
sample estimates:
mean of x mean of y
 35.1520  43.4812
```

```
> t.test(parcelaA,parcelaB,var.equal=T,paired=F,alternative="less")
```

Two Sample t-test

```
data: parcelaA and parcelaB
t = -3.137, df = 48, p-value = 0.001457
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -3.875911
sample estimates:
mean of x mean of y
 35.1520  43.4812
```

```
> t.test(parcelaA,parcelaB,paired=T,alternative="less")
```

Paired t-test

```
data: parcelaA and parcelaB
t = -2.6753, df = 24, p-value = 0.006617
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -3.002535
sample estimates:
mean of the differences
      -8.3292
```

```
> t.test(parcelaA,parcelaB,paired=T,alternative="greater")
```

Paired t-test

```
data: parcelaA and parcelaB
t = -2.6753, df = 24, p-value = 0.9934
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-13.65587      Inf
sample estimates:
mean of the differences
-8.3292
```

## Soluções de alguns Exercícios – Parte C

- 3.1. a) 0.2112  
b)  $\approx 0.2112$

3.2. a) Como  $n = 100$  é “grande”, pelo Teorema Limite Central, verifica-se para a média amostral,  $\bar{X} = \sum_{i=1}^{100} X_i/n$ ,  $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \implies \frac{\bar{X}-31}{0.24} \sim N(0,1)$ . Logo,  $P[\bar{X} < 30] \approx \Phi\left(\frac{30-31}{0.24}\right) = \Phi(-4.17) = 1 - \Phi(4.17) = 1 - 1 = 0$

b) Seja  $S_{100} = \sum_{i=1}^{100} X_i$  o peso total da amostra aleatória. Pelo Teorema Limite Central tem-se

$$\frac{S_{100} - 100\mu}{\sigma\sqrt{100}} \sim N(0,1) \implies \frac{S_{100} - 3100}{24} \sim N(0,1).$$

Então,  $P[S_{100} > 3150] = 1 - P[S_{100} \leq 3150] \approx 1 - \Phi\left(\frac{3150-3100}{24}\right) = 1 - \Phi(2.08) = 1 - 0.9812 = 0.0188$ .

- 3.3. a)  $\bar{x} = 56.2$  cm e  $s = 4$  cm  
b) 0.42074  
c)  $4/5=0.8$  cm.

- 3.4. b)  $p$   
d)  $\hat{p}_1 = 1/3$  e  $\hat{p}_2 = 1/2$

3.5. 1082

3.6.  $\approx 1 - \Phi(0.4\sqrt{12n})$ , para  $n$  elevado

3.7.  $\approx 0.01267$

**3.10. Resolução detalhada na colectânea de exames**

**3.12. Resolução detalhada na colectânea de exames**

**3.14. Resolução detalhada na colectânea de exames**

- 3.15. a) ]77.27; 79.33[  
b) 1.03  
c) 2654

3.16. 18.3 e  $n = 12$

- 3.17. a)  $\bar{x} = 0.959$  kg e  $s^2 = 0.0116$  kg<sup>2</sup>  
c) ]0.00886, 0.0155[

3.18. a) 99%

b) 0.185

**3.19.** a) i) 27    ii) 0.95

b) ]14.76; 64.77[

**3.20.** a) 25.85

b) 95%

c) O IC não inclui o valor 28 m. Logo, com 95% de confiança, pode-se afirmar que o silvicultor não tem razão na afirmação.

d) i)  $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$     ii) IC para variância a 95% de confiança =]5.9; 11.7[

**3.25. Resolução detalhada na colectânea de exames**

**3.26.** a) Unidade estatística: planta de tabaco; variável:  $n^{\circ}$  de folhas por planta

b)  $\bar{x} = 2967/150 = 19.78$  e  $\tilde{x} = 20$

**3.28. Resolução detalhada na colectânea de exames**

**3.32.** a) Verdadeira

b) Falsa

**3.35. Resolução detalhada na colectânea de exames**

**3.39. Resolução detalhada na colectânea de exames**

**3.40. Resolução detalhada na colectânea de exames**

**3.42. Soluções com alguns detalhes na colectânea de exames**

**R3.2.** 0.7127

**R3.3.** a)  $E[\hat{P}_1] = \frac{np+1}{n+1}$ ;  $E[\hat{P}_2] = \frac{np}{n+1}$ ;  
 $Var[\hat{P}_1] = Var[\hat{P}_2] = \frac{npq}{(n+1)^2}$ .

$\hat{P}_2$  é o que apresenta menor EQM portanto é o melhor.

b) Usando o melhor estimador,  $\hat{P}_2$ , temos uma estimativa para  $p$ ,  $\hat{p}_2 = 5/21$ .

**R3.4.** a) Estimativa do valor médio  $\mu$  é  $\bar{x} = 23.88$  ppm e de  $\sigma^2$  é  $s^2 = 1.9441$  ppm<sup>2</sup>

b) O grau de confiança é 99%.

**R3.5.** a)  $\bar{x} = 8.179$

c)  $n \geq 16$



**R3.8.** a)

$x_i$	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Freq. Rel Acum. $F_i$
0	9	0.09	0.09
1	26	0.26	0.35
2	34	0.34	0.69
3	22	0.22	0.91
4	8	0.08	0.99
5	1	0.01	1.00

b)  $\bar{x} = 197/100$  e  $\tilde{x} = 2$

c) i)  $\hat{p} = 197/500 = 0.394$     ii) 0.0428

**R3.9.** IC para  $p$  ]0.7417; 0.8423[

Como o IC não contém 0.9, a afirmação do vendedor não é correcta.

**R3.10.** a) 29.01%

b) ]0.2765; 0.3037[

**R3.11.** Resolução detalhada na colectânea de exames