Exercícios de Inferência Estatística

Paulo Justiniano Ribeiro Jr

Versão compilada em 13 de dezembro de 2018 às 17:07

1. Foram escolhidos ao acaso 500 animais (bovinos) de uma região para estimar a proporção de com propenção à uma certa doença. Destes, 120 testaram positivo.

   (a) Obtenha a estimativa pontual do percentual de susceptíveis na população.

   (b) Obtenha a estimativa intervalar (com confiança de 95%) do percentual de susceptíveis na população.

   (c) Idem anterior porém com confiança de 80%.

   (d) Deseja-se extender o levantamento para obter uma margem de erro de 1,5% para 95% de confiança. Quantos animais adicionais devem ser selecionados e testados?

   População:

$$\begin{aligned}
&\text{X : propenção à doença, 0 - não propenso, 1 - propenso}\\
&\text{X} \sim B(p)\\
&p : \text{parâmetro desconhecido}
\end{aligned}$$

   Amostra:

$$\begin{aligned}
\text{Dados:}&\\
&n = 500 \;\; ; \;\; \sum_{i=1}^{500} x_i = 120\\
\text{Estimador:}&\\
&\hat{p} = \frac{\sum_{i=1}^{500} x_i}{n}\\
\text{Estimativa:}&\\
&\hat{p} = \frac{120}{500} = 0,24\\
\text{Distribuição amostral:}&\\
&\hat{p} \sim \mathrm{N}(\mu_{\hat{p}} = p, \sigma^2_{\hat{p}} = \frac{p(1-p)}{n})
\end{aligned}$$

   (a) $\hat{p} = 0,24$

   (b) Intervalo assintótico (tomando $p = \hat{p}$)

$$\begin{aligned}
&\hat{p} \pm z_{0.95}\sqrt{\frac{p(1-p)}{n}}\\
&0,24 \pm 1,96 \cdot 0,019\\
&0,24 \pm 0,0374\\
&(0,203 \;\; ; \;\; 0,277)
\end{aligned}$$

(c)

$$\hat{p} \pm z_{0,80}\sqrt{\frac{p(1-p)}{n}}$$
$$0,24 \pm 1,28 \cdot 0,019$$
$$0,24 \pm 0,0245$$
$$(0,216 \;\; ; \;\; 0,264)$$

(d)

$$\text{ME} = z_{0.95}\sqrt{\frac{p(1-p)}{n}}$$
$$0,015 = 1,96^2\sqrt{\frac{0.24(1-0.24)}{n}}$$
$$n = \lceil 1,96^2 \frac{0.24(1-0.24)}{0,015^2}$$
$$n = 3115$$

Portanto 2615 novas amostras.

Soluções computacionais com o programa R:

```
> prop.test(120, 500, conf=0.95)$conf

[1] 0,2037 0,2804
attr(,"conf.level")
[1] 0,95

> prop.test(120, 500, conf=0.80)$conf

[1] 0,2154 0,2663
attr(,"conf.level")
[1] 0,8
```

---

2. Um pecuarista possui um rebanho e monitora o peso dos animais periodicamente. Como o rebanho é muito grande ele escolhe alguns animais ao acaso de cada vez. Um uma determinada pesagem ele seleciona 20 animais que mostram um peso médio de 430 kg com desvio padrão de 20 kg.

   (a) Obtenha uma estimativa intervalar (90% de confiança) para o peso do rebanho.

   (b) Se ele pesasse 10 animais, qual seria o intervalo de confiança?

   (c) Obtenha um intervalo de confiança (90% de confiança) para a variância do peso.

   População:

$$\text{X : peso de cada animal}$$
$$\text{X} \sim \text{N}(\mu, \sigma^2)$$
$$\mu \, , \, \sigma^2 \text{ : parâmetros desconhecidos}$$

Amostra:

Dados:

$$n = 20 \ ; \ \overline{x} = 430$$

Estimadores:

$$\hat{\mu} = \overline{x} = \frac{\sum_{i=1}^{500} x_i}{n} \ ; \ \hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{500} (x_i - \overline{x})^2}{n-1}$$

Estimativas:

$$\overline{x} = 430 \ ; \ S^2 = 20^2$$

Distribuições amostrais:

$$\overline{x} \sim \mathrm{N}(\mu, \sigma^2/n) \ ; \ S^2 \sim \chi^2(\nu = n-1)$$

(a)

$$\overline{x} \pm t_{0.90,19} \frac{S}{\sqrt{n}}$$
$$430 \pm 1,73 \cdot 4,47$$
$$430 \pm 7,73$$
$$(422,3 \ ; \ 437,7)$$

(b)

$$\overline{x} \pm t_{0.90,9} \frac{S}{\sqrt{n}}$$
$$430 \pm 1,83 \cdot 6,32$$
$$430 \pm 11,6$$
$$(418,4 \ ; \ 441,6)$$

(c)

$$\left(\frac{(n-1)S^2}{\chi^2_{sup}} \ ; \ \frac{(n-1)S^2}{\chi^2_{inf}} \right)$$
$$\left(\frac{(20-1)20^2}{30,1} \ ; \ \frac{(20-1)20^2}{10,1} \right)$$
$$(252 \ ; \ 751)$$

E para o desvio padrão teríamos:

$$\left(\sqrt{252} \ ; \ \sqrt{751} \right) \longrightarrow (15,9 \ ; \ 27,4)$$

---

3. Em uma avaliação de um novo algoritmo de classificação foi analisada uma amostra de 1200 cenários dentre os quais 780 foram classificados corretamente.

   (a) Identifique no contexto do problema: a população ("informal" e estatística), a amostra, o parâmetro, o estimador, a distribuição amostral e a estimativa (pontual).

   (b) Obtenha a estimativa pontual e intervalar (95% de confiança) para a proporção de classificações corretas.

   (c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse de 1,5% (com 95% de confiança)?

   (d) Um algoritmo atualmente utilizado possui um percentual de acerto de 62%. Há evidências baseadas no estudo de que o novo algoritmo tem desempenho diferente do utilizado atualmente? Justifique sua resposta.

   (e) Quais as suposições relevantes para os cálculos feitos nos item anteriores?

Solução:

$$X : \text{resultado da classificação (correto/incorreto)}$$
$$X \sim \mathrm{B}(p) \quad \mathrm{E}[X] = p \quad \mathrm{Var}[X] = p(1-p)$$
$$\hat{p} \sim \mathrm{N}(p, p(1-p)/n)$$

(a) **População "informal":** resultado (correto/incorreto) da classificação de imagens a serem analisadas pelo algorítmo

População "estatística":

$$X : \text{resultado da classificação (1: correto/0: incorreto)}$$
$$X \sim \mathrm{B}(p) \quad \mathrm{E}[X] = p \quad \mathrm{Var}[X] = p(1-p)$$

amostra: resultados das 1200 classificações feitas no teste (seriam representadas por um vetor de 0's e 1's)

parâmetro: a probabilidade (p) de acerto do algorítmo

estimador: o "processamento" que será feito na amostra para estimar o parâmetro p, no caso o cálculo da proporção de amostras classificadas corretamente

estimativa: o valor fornecido pelo estimador para a particular amostra que foi tomada, no caso $\hat{p} = 780/1200 = 0,65$

distribuição amostral: a distribuição das estimativas que seriam obtidas caso "diversas" amostras semelhantes à obtida fossem tomadas da população, no caso: $\hat{p} \sim \mathrm{N}(p, p(1-p)/n)$

(b)

$$\hat{p} = \frac{780}{1200} = 0,65$$

I.C. assintótico:

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow (0,623 \;\; ; \;\; 0,677)$$

I.C.conservador :

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{1}{4n}} \longrightarrow (0,622 \;\; ; \;\; 0,678)$$

(c) i. Utilizando $p = \hat{p}$

$$\begin{aligned}
\mathrm{ME} &= z_{95\%}\sqrt{\frac{p(1-p)}{n}} \\
0.015 &= 1,96\sqrt{\frac{0,65(1-0,65)}{n}} \\
n &= \lceil \frac{1,96^2}{0,015^2} 0,65(1-0,65) \\
n &= 3885
\end{aligned}$$

ii. Utilizando $p = 0,5$

$$\begin{aligned}
\mathrm{ME} &= z_{95\%}\sqrt{\frac{0,5(1-0,5)}{n}} = z_{95\%}\sqrt{\frac{1}{4n}} \\
0.015 &= 1,96\sqrt{\frac{0,5(1-0,5)}{n}} \\
n &= \lceil \frac{1,96^2}{0,015^2} \frac{1}{4} \\
n &= 4269
\end{aligned}$$

(d) Resposta e justificativa pode ser dada baseada no valor 0,62 estar ou não contido no I.C.. Alternativamente, pode-se construir um teste de hipótese como se segue.

Hipóteses: $H_0 : p = 0,62 \ vs \ H_a : p \neq 0,62$

Nível de significância: adotado $\alpha = 0,05$

Estatística de teste:

$$z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,65 - 0,62}{\sqrt{\frac{0,62(1-0,62)}{1200}}} = 2,14$$

Valor crítico: $z_t = \pm 1,96$

ou, alternativamente,

nível descritivo: p-valor $= P[Z > z_c] = 0,0323$

Conclusão: Como $|z_c| > z_t$ (ou, p-valor $< \alpha$), rejeita-se H_0 an nível de 5% de significância, ou seja, há evidências de que o novo algorítmo tem uma taxa de acerto diferente (no caso superior) à do utilizado atualmente.

(e)

---

4. Considere um exame no qual as notas possuem uma média de 100 pontos e um desvio padrão de 12 pontos. Obtenha a probabilidade de que

(a) a nota de um estudante selecionado ao acaso seja superior a 110,

(b) a nota média de um grupo de 25 estudantes tomados ao acaso seja superior a 105,

(c) a nota média de um grupo de 64 estudantes tomados ao acaso seja superior a 105,

(d) a nota média de um grupo de 16 estudantes tomados ao acaso esteja abaixo de 95 ou acima de 105,

(e) o número de estudantes que deveriam ser selecionados ao acaso para que a nota média estivesse entre 95 e 105 pontos com probabilidade de 0,98.

Solução:

$$\begin{aligned} X &: \text{nota de um estudante} \\ X &\sim N(\mu = 100, \sigma^2 = 12^2) \\ \overline{X}_n &: \text{nota média de um grupo de } n \text{ estudantes} \\ \overline{X}_n &\sim N(\mu_{\overline{X}} = 100, \sigma^2_{\overline{X}} = 12^2/n) \end{aligned}$$

(a) $P[X > 110] = P[Z > \frac{110-100}{12}] = P[Z > 0,8333] = 0,202$

(b) $P[\overline{X}_{25} > 105] = P[Z > \frac{105-100}{12/\sqrt{25}}] = P[Z > 2,083] = 0,0186$

(c) $P[\overline{X}_{64} > 105] = P[Z > \frac{105-100}{12/\sqrt{64}}] = P[Z > 3,333] = 0,000429$

(d) $P[\overline{X}_{16} < 95 \cup \overline{X}_{16} > 105] = P[Z < \frac{95-100}{12/\sqrt{16}} \cup Z > \frac{105-100}{12/\sqrt{16}}] = P[Z < -1,667 \cup Z > 1,667] = 0,0956$

(e)

$$\begin{aligned} P[95 < \overline{X}_n < 105] &= 0,98 \\ P[|\overline{X}_n - 100| < 5] &= 0,98 \\ z &= \frac{5}{12/\sqrt{n}} \\ n &= \lceil \left(\frac{z \cdot 12}{5}\right)^2 = 32 \end{aligned}$$

Soluções computacionais (linguagem R):

```
> pa <- pnorm(110, m=100, sd=12, lower=FALSE)
> pb <- pnorm(105, m=100, sd=12/sqrt(25), lower=FALSE)
> pc <- pnorm(105, m=100, sd=12/sqrt(64), lower=FALSE)
> pd <- 1-diff(pnorm(c(95,105), m=100, sd=12/sqrt(16)))
> n <- ceiling((qnorm(0.99)*12/(105-100))^2)
```

---

5. Explique os conceitos de ”estimador não viesado”e ”estimador mais eficiente”.

---

6. Uma amostra aleatória de uma população com distribuição de Poisson forneceu os seguintes valores: 1, 2, 0, 2, 2, 2, 5, 2, 3, 1, 2, 8. Obtenha o estimador de máxima verossimilhança para o parâmetro desta distribuição e o valor da estimativa para a amostra em questão.

Solução:

$$\begin{aligned}
X &\sim \mathrm{P}(\lambda) \\
P[X = x_i] &= \frac{\mathrm{e}^{-\lambda}\lambda^{x_i}}{x_i!} \\
l(\lambda) = \log\{L(\lambda)\} &= \log\{\prod_{i=1}^{n} P[X = x_i]\} \\
&= \sum_{i=1}^{n} \log\{P[X = x_i]\} = \sum_{i=1}^{n} \log\{\frac{\mathrm{e}^{-\lambda}\lambda^{x_i}}{x_i!}\} \\
&= \sum_{i=1}^{n} [-\lambda + x_i \log(\lambda) - \log(x_i!)] = -n\lambda + \log(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!)] \frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} \quad = -n + \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}} = 0 \\
\hat{\lambda} &= \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}
\end{aligned}$$

Para a amostra dada a estimativa é:

$$\hat{\lambda} = \frac{1 + 2 + \cdot + 2 + 8}{12} = 2,5$$

---

7. Seja X a variável tempo de serviço dos funcionários de determinada localidade. Se X tem distribuição normal com média 15 anos desvio padrão 10 anos:

   (a) Qual a probabilidade de um funcionário, aleatoriamente escolhido, ter pelo menos 10 anos de tempo de serviço?

   (b) Tomando-se uma amostra de 16 funcionários, qual a probabilidade do tempo médio de serviço estar entre 13 e 16 anos?

Solução:

$$\begin{aligned}
X &: \text{tempo de serviço de um funcionário} \\
X &\sim \mathrm{N}(\mu = 15, \sigma^2 = 10^2) \\
\overline{X} &: \text{tempo médio de serviço de um grupo de 16 funcionários} \\
\overline{X}_{16} &\sim \mathrm{N}(\mu = 15, \sigma^2 = 10^2/16)
\end{aligned}$$

(a) $P[X \geq 10] = P[Z \geq \frac{10-15}{10}] = P[Z \geq -0.5] = 0,6915$

(b) $P[13 \leq \overline{X}_{16} \leq 16] = P[\frac{13-15}{10/\sqrt{16}} \leq Z \leq \frac{16-15}{10/\sqrt{16}}] = P[-0,8 \leq Z \leq 0,4] = 0,4436$

Soluções computacionais (linguagem R):

```
> pa <- pnorm(10, m=15, sd=10, lower=FALSE)
> pb <- diff(pnorm(c(13,16), m=15, sd=10/4))
```

---

8. Uma companhia de cigarros anuncia que o índice médio de nicotina dos seus cigarros apresenta-se abaixo de 26 mg. Um laboratório realiza 6 análises, obtendo as seguintes quantidades de nicotina: 27, 24, 21, 25, 26, 22. Sabe-se que o índice de nicotina se distribui normalmente. Pode-se aceitar, ao nível de 10% de significância, a afirmação do fabricante? E ao nível de 5%?

Solução:

$$H_0 : \mu \geq 26 \text{ vs } H_a : \mu < 26$$
$$\alpha = 0,10 \text{ ou } \alpha = 0,05$$
$$t_c = \frac{\overline{x} - \mu_0}{S/\sqrt{n}} = \frac{24,17 - 26}{2,317/\sqrt{6}} = -1,938$$

(a) Para $\alpha = 0,10$

$$RC : \{t < -1,48\}$$
$$Rejeita - seH0...$$

(b) Para $\alpha = 0,05$

$$RC : \{t < -2,02\}$$
$$NaoRejeita - seH0...$$

Soluções computacionais (linguagem R):

```
> y <- c(27,  24,  21,  25,  26,  22)
> (m.y <- mean(y))

[1] 24,17

> (v.y <- var(y))

[1] 5,367

> (RC0.10 <- qt(0.10, df=length(y)-1))

[1] -1,476

> (RC0.05 <- qt(0.05, df=length(y)-1))

[1] -2,015

> (tc <- (mean(y)-26)/sqrt(var(y)/length(y)))

[1] -1,938

> ## solução "direta" usando função apropriada do R
> (T1 <- t.test(y, alternative="less", mu=26, conf=0.90))
```

```
One Sample t-test

data:  y
t = -1,9, df = 5, p-value = 0,06
alternative hypothesis: true mean is less than 26
90 percent confidence interval:
  -Inf 25,56
sample estimates:
mean of x
    24,17

> (T2 <- t.test(y, alternative="less", mu=26, conf=0.95))

One Sample t-test

data:  y
t = -1,9, df = 5, p-value = 0,06
alternative hypothesis: true mean is less than 26
95 percent confidence interval:
  -Inf 26,07
sample estimates:
mean of x
    24,17
```

---

9. Uma pesquisa mostrou que 30 dos 120 alunos entrevistados de uma universidade eram chefes de família.

   (a) Calcule um intervalo com 95% de confiança para a proporção de estudantes chefes de família.

   (b) Quantos alunos deveriam ser entrevistados para que o intervalo de confiança tivesse a metade da amplitude?

   (c) Uma segunda universidade fez a mesma pesquisa e encontrou 50 chefes de família entre os 150 entrevistados. Pode-se afirmar que há uma diferença significativa entre as proporções de chefes de família nas duas universidades?

Solução:

(a) IC (95%)

$$\hat{p} = \frac{30}{120} = 0,25 \text{Assintótico :}$$
$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$0,25 \pm 1,96\sqrt{\frac{0,25(1-0,25)}{120}}$$
$$0,25 \pm 0,0775$$
$$(0,173 \ ; \ 0,327)$$

Conservador :

$$\hat{p} \pm z\sqrt{\frac{1}{4n}}$$
$$0,25 \pm 1,96\sqrt{\frac{1}{4 \cdot 120}}$$
$$0,25 \pm 0,0895$$
$$(0,161 \ ; \ 0,339)$$

(b) n

$$
\begin{aligned}
&\text{baseado no IC assintótico :}\\
1,96\sqrt{\frac{0,25(1-0,25)}{n}} &= \frac{0,0775}{2}\\
n &= \left\lceil \frac{1,96^2 \cdot 0,25 \cdot (1-0,25)}{0,0387^2} \right.\\
n &= 480\\
&\text{baseado no IC conservador :}\\
1,96\sqrt{\frac{1}{4n}} &= \frac{0,0895}{2}\\
n &= \left\lceil \frac{1,96^2}{4 \cdot 0,0387^2} \right.\\
n &= 640
\end{aligned}
$$

(c) Comparação de duas proporções

$$
\begin{aligned}
H_0 :& p_1 = p_2 (p_1 - p_2 = 0) \text{ vs } H_a : p_1 \neq p_2 (p_1 - p_2 \neq 0)\\
\alpha &= 0,05\\
z_c &= \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{(0,25 - 0,33) - 0}{\sqrt{\frac{0,25(1-0,25)}{120} + \frac{0,33(1-0,33)}{150}}} = -1,43\\
RC :& \{z < 1,96 \text{ ou } z > 1,96\}\\
z_c &\notin RC
\end{aligned}
$$

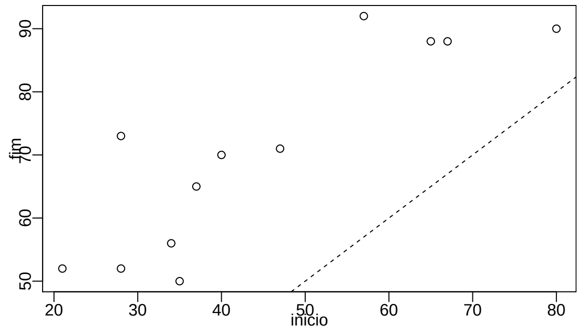
Não rejeita-se H_0, não há evidência suficiente para rejeitar a hipótese de que as proporções sejam iguais.

Soluções computacionais (linguagem R):

```
> ## resultado "direto" usando funções do R
> (a <- prop.test(30, 120, conf=0.95)$conf.int)

[1] 0,1775 0,3389
attr(,"conf.level")
[1] 0,95

> (TAB <- matrix(c(30,50,(120-30),(150-50)), nrow=2, dimnames=list(c("UniA","UniB"), c("Chefe","Na

     Chefe NaoChefe
UniA    30       90
UniB    50      100

> (b <- prop.test(TAB, conf=0.95))

2-sample test for equality of proportions with continuity correction

data:  TAB
X-squared = 1,8, df = 1, p-value = 0,2
alternative hypothesis: two.sided
95 percent confidence interval:
 -0,1990  0,0323
sample estimates:
prop 1 prop 2
0,2500 0,3333
```

10. Em um procedimento de avaliação um mesmo exame foi aplicado aos mesmos alunos quando estavam em períodos inicial e final de um curso. Os resultados das notas obtidas são apresentados a seguir. Interprete, faça análises estatísticas adequadas e discuta os resultados.

	media	desvio	CV	Amplitude	mediana	q1	q3	AIQ
inicio	44,92	18,39	40,95	59,00	38,50	31,00	61,00	30,00
fim	70,58	15,97	22,63	42,00	70,50	54,00	88,00	34,00



Solução:

(a) Será verificada a comparação e discussão das medidas descritivas da tabela e o gráfico.

(b) Teste de hipótese para duas amostras pareadas
Diferenças das observações:

```
[1] 28 35 22 30 31 45 15 10 23 24 24 21
```

$$H_0 : \mu_{fim} = \mu_{inicio} \ \ (d = 0) \text{ vs } \mu_{fim} = \mu_{inicio} \ \ (d > 0)$$
$$\alpha = 0,05$$
$$\overline{d} = 25,7 \text{ e } Var[\overline{d}] = 83,7$$
$$t_c = \frac{25,7 - 0}{\sqrt{83,7/12}} = 9,72$$
$$RC : \{t_c > 1,8\}$$
$$t_c \in RC$$

Rejeita-se H_0, ou seja, rejeita-se que a diferença das notas seja nula econsidera-se portanto que houve um aumento nas notas.

Soluções computacionais (linguagem R):

```
> t.test(fim, inicio, paired=TRUE, conf=0.95)

Paired t-test

data:  fim and inicio
t = 9,7, df = 11, p-value = 1e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 19,85 31,48
sample estimates:
mean of the differences
                 25,67
```

11. Considere um estudo na qual deseja-se estimar a proporção de solicitações atendidas e resolvidas de uma central do usuário através de uma amostra aleatória simples.

   (a) Qual a população, variável aleatória, parâmetro de interesse e o estimador?
   (b) Se a amostra for de 4000 solicitações, qual será a margem de erro (com confiança de 95%) para a estimação da proporção de resolvidas?
   (c) Para este mesmo tamanho de amostra, qual seria a confiança associada a uma margem de erro de $\pm 0,01$ (1 %)?
   (d) Qual deveria ser o tamanho da amostra para se obter a estimativa com uma margem de erro de 2,5% com 95% de confiança?
   (e) E para uma margem de erro de 3% com 99% de confiança?

Solução:

(a) População: solicitações atendidas, representada pela v.a.:

$$X : \text{resolução de solicitações atendidas}$$

que é binária e pode ser codificada assumindo valores 0, para não atendidas e, 1, para atendidas. O parâmetro de interesse é a proporção de solicitações atendidas na população, que representa a probabilidade p de uma solicitação ser atendida. O estimador é a proporção de atendidas na amostra. Assume-se então que:

$$X \sim B(p)$$

e os resultados das análises se baseiam no Teorema:

$$\text{Teo 2:} \quad \hat{p} = \overline{X} \approx N(p, \frac{p(1-p)}{n})$$

do qual se extraem os resultados:

$$\hat{p} \pm z_{(1-\alpha)} \sqrt{\frac{p(1-p)}{n}}$$

$$\hat{p} \pm \mathrm{ME}$$

$$\mathrm{ME} = z_{(1-\alpha)} \sqrt{\frac{p(1-p)}{n}}$$

$$n = \frac{(z_{(1-\alpha)})^2}{\mathrm{ME}^2} p(1-p)$$

$$p(1-p) \text{ é limitado superiormente para } p = 0,5$$

$$n = \frac{(z_{(1-\alpha)})^2}{\mathrm{ME}^2} 0,25$$

(b) $M.E. = z_{(1-\alpha)} \sqrt{\frac{p(1-p)}{n}} = 1,96 \sqrt{\frac{0,5(1-0,5)}{4000}} = 0,0155$

(c)

$$M.E. = z_{(1-\alpha)} \sqrt{\frac{p(1-p)}{n}}$$

$$z_{(1-\alpha)} = 0,01 / \sqrt{\frac{0,5(1-0,5)}{4000}} = 1,26$$

$$1 - \alpha = 79,4\%$$

(d) $n = \frac{(1,96)^2}{(0,025)^2} 0,25 = 1537$

(e) $n = \frac{(2,576)^2}{(0,03)^2} 0,25 = 1844$

12. Um conjunto de imagens foi submetido a dois algoritmos de tratamento (filtragem, correção e classificação) e foram registrados os tempos de processamento conforme a tabela a seguir.

Imagem	1	2	3	4	5	6	7	8	9	10
algorítmo A	23.7	27.9	35.3	17.7	20.9	32.2	50.9	45.4	76.8	31.1
algorítmo B	13.9	21.9	16.9	3.5	6.9	36.4	30.3	7.6	59.2	33.2

(a) Calcule a média, desvio padrão e coeficiente de variação de cada grupo

(b) Calcule a mediana, amplitude e amplitude interquartílica de cada grupo

(c) Faça um gráfico box-plot para comparar os algorítmos

(d) Faça um gráfico adequado e calcule alguma medida para verificar se existe associação entre os tempos de processamento dos dois algorítmos.

Considere agora apenas os dados do algorítmo B e suponha que possuam distribuição normal.

(e) Obtenha um intervalo de confiança (95%) para a variância dos tempos

(f) Obtenha um intervalo de confiança (95%) para o tempo médio

(g) qual deveria ser o tamanho da amostra para que a margem de erro no item anterior fosse metade da obtida?

Finalmente, considere que deseja-se adotar algum procedimento de inferência estatística para comparar os tempos médios de processamento dos dois algorítmos

(h) Descreva qual o procedimento que voce utilizaria e como interpretaria os possíveis resultados (não é preciso fazer os cálculos da comparação).

Solução:

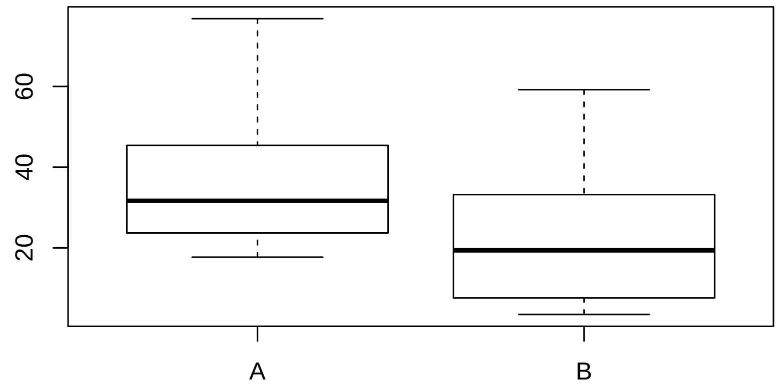
(a)

$$\overline{x}_A = 36,19 \qquad \overline{x}_B = 22,98$$
$$S_A = 17,62 \qquad S_B = 17,14$$
$$CV_A = 48,7\% \qquad CV_B = 74,6\%$$

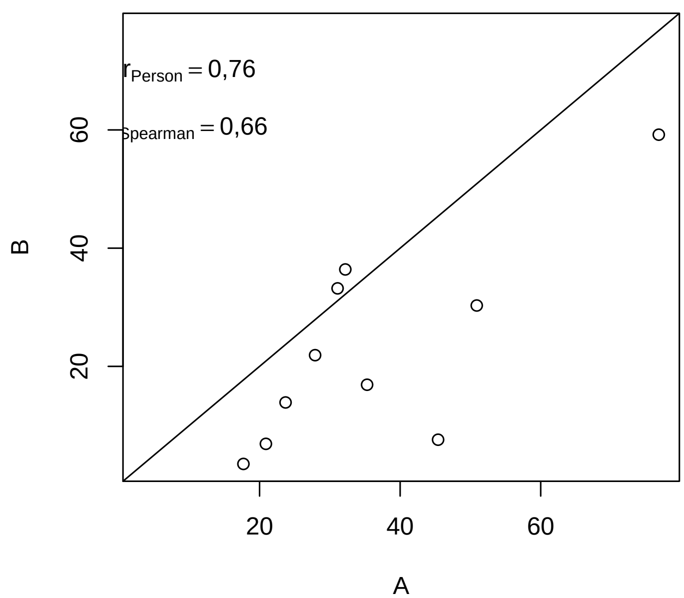
(b)

	md	min	max	A	Q1	Q3	AI
A	31,65	17,7	76,8	59,1	23,7	45,4	21,7
B	19,4	3,5	59,2	55,7	7,6	33,2	25,6

(c)



(d)



(e)

$$\left(\frac{(n-1)S_B^2}{\chi^2_{sup}} ; \frac{(n-1)S_B^2}{\chi^2_{inf}}\right)$$

$$\left(\frac{(10-1)\cdot 293,91}{19,02} ; \frac{(10-1)\cdot 293,91}{2,7}\right)$$

$$(139,05 \quad ; \quad 979,55)$$

(f)

$$\overline{x} \pm t_{0.95}\frac{S}{\sqrt{n}}$$
$$22,98 \pm 2,26\frac{17,14}{\sqrt{10}}$$
$$22,98 \pm 12,26$$
$$(10,72 \quad ; \quad 35,24)$$

(g)

$$t_{0.95}\frac{S}{\sqrt{n}} = 6,13$$
$$2,26\frac{17,14}{\sqrt{n}} = 6,13$$
$$n = \left\lceil 2,26^2\frac{17,14^2}{6,13^2}\right.$$
$$n = 41$$

(h)

---

13. Assume-se que em uma rede com um grande número de nós 20% deles podem estar inacessíveis a qualquer tempo. São feitas inspeções periódicas em 400 nós da rede escolhidos ao acaso e se 90 (22,5%) ou mais desses estão inacessíveis, é feito um rastreamento detalhado para verificação e detecção de problemas.

(a) Qual a probabilidade de encontrar mais que 85 nós inacessíveis em uma inspeção?

(b) Se a rede está normal, qual a probabilidade de encontrar 100 ou mais nós inacessíveis em uma inspeção?

(c) Qual os valores de proporções ao redor do valor médio (20%) de nós inacessíveis, dentre os quais devem-se estar 80% das inspeções.

(d) Qual a probabilidade de fazer um rastreamento desnecessário?

(e) Quantos nós deveriam ser inspecionados para que a chance de fazer um rastreamento desnecessário não ultrapasse 1%?

Solução:

$$X : \text{estado do nó (0 - acessível, 1 : inacessível)}$$
$$X \sim B(\mu = E[X] = p = 0,20, \sigma^2 = \text{Var}[X] = p(1-p) = 0,2 \cdot 0,8 = 0,16)$$
$$\hat{p} = \overline{X} : \text{proporção de acessíveis entre n (400) nós}$$
$$\hat{p} = \overline{X} \sim N(\mu_{\hat{p}} = p = 0,20, \sigma^2_{\hat{p}} = \frac{p(1-p)}{n} = \frac{0,2(1-0,2)}{400} = 0,0004 = 0,02^2)$$

(a) $P[\hat{p} > 85/400] = P[Z > \frac{85/400-0,20}{0,02}] = P[Z > 0,625] = 0,5 - P[0 < Z < 0,625] = 0,5 - 0,234 = 0,266$

(b) $P[\hat{p} > 100/400] = P[Z > \frac{100/400-0,20}{0,02}] = P[Z > 2,5] = 0,5 - P[0 < Z < 2,5] = 0,5 - 0,4938 = 0,00621$

(c)

$$P[p - \Delta_{\hat{p}} < X < p + \Delta_{\hat{p}}] = 0,80$$
$$z = \frac{p + \Delta_{\hat{p}} - p}{0,02} = 1,28$$
$$\Delta_{\hat{p}} = 0,0256$$
$$P[0,174 < X < 0,226] = 0,80$$

(d) $P[\hat{p} > 0,225] = P[Z > \frac{0,225-0,20}{0,02}] = P[Z > 1,25] = 0,5 - P[0 < Z < 1,25] = 0,5 - 0,3944 = 0,1056$

(e)

$$z_{0,99} = 2,33 = \frac{0,225 - 0,20}{0,16/n}$$
$$n = \frac{2,33^2}{(0,225 - 0,20)^2} 0,16 = 1386$$

Solução Computacional:

```
> sP <- sqrt(.2*.8/400)
> (p2.a <- pnorm(85/400, m=0.20, sd=sP, low=F))

[1] 0,266

> (p2.b <- pnorm(100/400, m=0.20, sd=sP, low=F))

[1] 0,00621

> (q2.c <- qnorm(c(0.10, 0.90), m=0.20, sd=sP))

[1] 0,1744 0,2256

> (p2.d <- pnorm(0.225, m=0.20, sd=sP, low=F))

[1] 0,1056

> (n.e <- ceiling(qnorm(0.99)^2 * (0.2*.8)/0.025^2))

[1] 1386
```

---

14. Uma locadora de veículos que possui uma grande frota decide fazer um estudo sobre vários aspectos relacionados ao desempenho. Para isto vai tomar uma amostra aleatória de 25 de seus veículos para inspeções detalhadas. Várias características serão medidas, mas vamos aqui nos ater apenas ao consumo de combustível, supondo que a variância do consumo de toda a frota é de $6,25 km/l$.

    (a) Qual a probabilidade do consumo médio aferido nos 25 veículos, diferir do consumo médio de toda a frota em mais que $0,5 km/l$? E em mais que $1 km/l$?
    (b) Qual a margem de erro na estimação do consumo médio da frota para uma confiança de 95%?
    (c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse a metade da calculada no item anterior?
    (d) Se uma amostra (com $n = 25$) fornecer uma estimativa intervalar de $(11,1\,;\,12,3) km/l$, qual a confiança desta estimativa?
    (e) identifique no problema: a população variável aleatória de interesse, o parâmetro de interesse, o estimador, a distribuição amostral, a estimativa pontual e a estimativa intervalar.

Solução:

$$
\begin{aligned}
X &: \text{consumo de veículo da frota} \\
\text{Distribuição da variável aleatória (população):} & \\
X &\sim \text{Dist.}(\mu_X = \text{E}[X], \sigma^2_X = \text{Var}[X]) \\
\text{distribuição amostral:} & \\
\overline{X} &\approx \text{N}(\mu_{\overline{X}} = \mu_X, \sigma_{\overline{X}} = \sigma^2_X/n) \\
\text{ou, equivalentemente,} & \\
\overline{X} - \mu &\approx \text{N}(0, \sigma^2/n) \\
\sigma^2_X &= 6,25 \ ; \ \sigma^2_{\overline{X}} = 0,25 \ ; \ \sigma_{\overline{X}} = 0,5
\end{aligned}
$$

(a)

$$
P[|\overline{X} - \mu| > 0,5] = P[\frac{|\overline{X} - \mu|}{\sigma} > 0,5/\sigma] = P[|Z| > 1] = 0,317
$$

$$
P[|\overline{X} - \mu| > 1] = P[\frac{|\overline{X} - \mu|}{\sigma} > 1/\sigma] = P[|Z| > 2] = 0,0455
$$

(b)

$$
\begin{aligned}
ME &= z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}} \\
ME &= 1,96\frac{2,5}{\sqrt{25}} \\
ME &= 0,98
\end{aligned}
$$

(c)

$$
\begin{aligned}
\frac{ME}{2} &= z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n^*}} \\
n^* &= \lceil(\frac{2}{ME})^2 z^2_{0,95}\sigma^2_X \\
n^* &= \lceil(\frac{2}{0,98})^2 1,96^2 6,25 \\
n^* &= 100
\end{aligned}
$$

(d)

$$
\begin{aligned}
(11,1\,;\,12,3) &\equiv 11,7 \pm 0,6 \\
ME &= z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}} \\
0,6 &= z_{1-\alpha}0,5 \\
z_{1-\alpha} &= \frac{0,6}{0,5} \\
1 - \alpha(\text{confiança}) &= 0,77(77\%)
\end{aligned}
$$

(e)

---

15. O (log-)tempo de processamento de *jobs* submetidos a um servidor de processamento possui um distribuição aproximandamente normal com média de 5,4 e desvio padrão de 1,2 (log-)minutos.

    (a) Qual a proporção esperada de *jobs* que devem ser concluídos em menos do que 5 (log-)minutos?

(b) Qual a probabilidade de que um lote de 10 *jobs* seja processado em menos que 50 (log-)minutos?

(c) Qual o intervalo de (log-)tempos ao redor da média de 5,4, que se espera que 95% de lotes de 16 *jobs* sejam processados?

(d) Quantos *jobs* deveriam ser submetidos em um lote para que o tempo médio de processamento fosse, com probabilidade de 0,90, de no máximo de 6 (log-)minutos.

Solução:

$$\begin{aligned}
X &: \text{(log-)tempo de processamento} \\
X &\sim \mathrm{N}(5,4;1,2^2) \\
\overline{X}_n &: \text{(log-)tempo médio de processamento de } n \text{ } jobs \\
\overline{X}_n &\sim \mathrm{N}(5,4;1,2^2)
\end{aligned}$$

(a) $P[X < 5] = 0,369$

(b) $P[\sum_{i=1}^{10} X_i < 50] = P[\overline{X}_{10} < 5] = 0,146$

(c) $P[t_1 < \overline{X}_{16} < t_2] = 0,95 \longrightarrow (t_1 = 4,81, t_2 = 5,99$

(d) $P[\overline{X}_n < 5,5] = 0,90 \longrightarrow n = 7$

Comandos em R para soluções:

```
> q1a <- pnorm(5, mean=5.4, sd=1.2)
> q1b <- pnorm(5, mean=5.4, sd=1.2/sqrt(10))
> q1c <- qnorm(c(0.025, 0.975), mean=5.4, sd=1.2/sqrt(16))
> q1d <- ceiling(qnorm(0.9)^2 * 1.2^2/(0.6^2))
```

---

16. Ainda no contexto do problema anterior acredita-se que 18% dos *jobs* encerram-se, por alguma razão) sem produzir o resultado correto.

(a) Qual a probabilidade de que em um lote de 200 *jobs*, ao menos 80% deles produzam o resultado correto?

(b) Decide-se fazer uma inspeção diária submetendo um lote de 150 *jobs* de controle. Se 30 ou mais não produzem o resultado correto o serviço é interrompido para uma inspeção na servidora. Qual a probabilidade de uma interrupção desnecessária?

(c) Qual deveria ser o número limite de *jobs* sem resultado correto para que tal probabilidade de interrupção desnecessária fosse de no máximo 0,15?

(d) Deseja-se manter o limite de 20% de *jobs* sem resultado correto como critério de interrupção com no máximo 0,15 de probabilidade de uma parada desnecesária. Neste caso, quantos *jobs* de controle deveriam ser testados?

Solução:

$$\begin{aligned}
&X : \text{número de } jobs \text{ sem resultado correto} \\
&X \sim \mathrm{B}(n; p = 0,18) \\
&p_n : \text{proporção de } jobs \text{ sem resultado correto em um lote de } n \text{ } jobs \text{ submetidos} \\
&p_n \sim \mathrm{N}(0,18; \frac{0,18(1-0,18)}{n})
\end{aligned}$$

(a) $P[p_{200} \leq 0,20] = 0,769$

(b) $P[p_{150} \geq \frac{30}{150}] = 0,262$

(c) $P[p_{150} \geq \frac{C}{150}] \leq 0,15 \longrightarrow C = 32$

(d) $P[p_n \geq 0,20] \leq 0,15 \longrightarrow n = 397$

Comandos em R para soluções:

```
> q2a <- pnorm(0.20, mean=0.18, sd=sqrt(0.18*(1-0.18)/200))
> q2b <- pnorm(30/150, mean=0.18, sd=sqrt(0.18*(1-0.18)/150), lower=FALSE)
> q2c <- ceiling(150 * qnorm(0.15, mean=0.18, sd=sqrt(0.18*(1-0.18)/150), lower=FALSE))
> q2d <- ceiling((qnorm(0.85)^2 * 0.18 * (1-0.18))/(0.20-0.18)^2)
```

---

17. Foi registrado o número de transações concretizadas em um sistema obtendo-se os seguintes valores:

$$3, 7, 2, 1, 5, 6, 3, 7, 8, 5, 6, 4, 3, 5, 2, 5, 4.$$

Deseja-se ajustar uma distribuição de Poisson ($P[X = x_i] = \mathrm{e}^{-\lambda}\lambda^{x_i}/x_i!$) aos dados. Assumindo-se independência entre os observações,

(a) encontre a expressão do estimador de máxima verossimilhança de λ,

(b) encontre o valor da estimativa de máxima verossimilhança obtida com os dados acima,

(c) usando a estimativa encontrada, encontre a probabilidade estimada de se obter uma ou menos transações em um dia qualquer.

Solução:

$$\begin{aligned} X &: \text{número de transações concretizadas} \\ X &\sim \mathrm{P}(\lambda) \\ P[X = x_i] &= \frac{\mathrm{e}^{-\lambda}\lambda_i^x}{x_i!} \end{aligned}$$

(a)

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^n P[X = x_i] = \prod_{i=1}^n \frac{\mathrm{e}^{-\lambda}\lambda_i^x}{x_i!} \\ l(\lambda) &= \log[L(\lambda)] = \sum_{i=1}^n \log\left[\frac{\mathrm{e}^{-\lambda}\lambda_i^x}{x_i!}\right] \\ &= \sum_{i=1}^n [-\lambda + x_i \log(\lambda) - \log(x_i!)] \\ &= -n\lambda + (\sum_{i=1}^n x_i)\log(\lambda) - \sum_{i=1}^n \log(x_i!) \\ \frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} &= -n + \frac{\sum_{i=1}^n x_i}{\lambda} \\ \frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} &= 0 \longrightarrow \hat{\lambda} = \frac{\sum_{i=1}^n X_i}{n} = \overline{X} \end{aligned}$$

(b) $\overline{x} = 4,47$

(c) $\hat{P}[X \leq 1] = \hat{P}[X = 1] + \hat{P}[X = 1] = \frac{\mathrm{e}^{-4,47}4,47^0}{0!} + \frac{\mathrm{e}^{-4,47}4,47^1}{1!} = 0,0626$

Comandos em R para soluções:

```
> dados <- c(3, 7, 2, 1, 5, 6, 3, 7, 8, 5, 6, 4, 3, 5, 2, 5, 4)
> (m <- mean(dados))
```

```
[1] 4,471
> (p <- ppois(1, lambda=m))
[1] 0,06259
```

---

18. Uma locadora de veículos que possui uma grande frota decide fazer um estudo sobre vários aspectos relacionados ao desempenho. Para isto vai tomar uma amostra aleatória de 25 de seus veículos para inspeções detalhadas. Várias características serão medidas, mas vamos aqui nos ater apenas ao rendimento de combustível, supondo que a variância do rendimento de toda a frota é de $6,25(km/l)^2$.

    (a) Qual a probabilidade do consumo médio aferido nos 25 veículos, diferir do consumo médio de toda a frota em mais que $0,5km/l$? E em mais que $1km/l$?

    (b) Qual a margem de erro na estimação do consumo médio da frota para uma confiança de 95%?

    (c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse a metade da calculada no item anterior?

    (d) Se uma amostra (com $n = 25$) fornecer uma estimativa intervalar de $(11,1\,;\,12,3)km/l$, qual a confiança desta estimativa?

    (e) identifique no problema: a população variável aleatória de interesse, o parâmetro de interesse, o estimador, a distribuição amostral, a estimativa pontual e a estimativa intervalar.

Solução:

$$
\begin{aligned}
& X : \text{consumo de veículo da frota} \\
\text{Distribuição da variável aleatória (população):} & \\
& X \sim \text{Dist.}(\mu_X = \text{E}[X], \sigma^2_X = \text{Var}[X]) \\
\text{distribuição amostral:} & \\
& \overline{X} \approx \text{N}(\mu_{\overline{X}} = \mu_X, \sigma_{\overline{X}} = \sigma^2_X/n) \\
\text{ou, equivalentemente,} & \\
& \overline{X} - \mu \approx \text{N}(0, \sigma^2/n) \\
& \sigma^2_X = 6,25\ ;\ \sigma^2_{\overline{X}} = 0,25\ ;\ \sigma_{\overline{X}} = 0,5
\end{aligned}
$$

(a)

$$P[|\overline{X} - \mu| > 0,5] = P[\frac{|\overline{X} - \mu|}{\sigma} > 0,5/\sigma] = P[|Z| > 1] = 0,317$$

$$P[|\overline{X} - \mu| > 1] = P[\frac{|\overline{X} - \mu|}{\sigma} > 1/\sigma] = P[|Z| > 2] = 0,0455$$

(b)

$$
\begin{aligned}
ME &= z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}} \\
ME &= 1,96\frac{2,5}{\sqrt{25}} \\
ME &= 0,98
\end{aligned}
$$

(c)

$$
\begin{aligned}
\frac{ME}{2} &= z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n^*}}\\
n^* &= \lceil(\frac{2}{ME})^2 z_{0,95}^2\sigma_X^2\\
n^* &= \lceil(\frac{2}{0,98})^2 1,96^2 6,25\\
n^* &= 100
\end{aligned}
$$

(d)

$$
\begin{aligned}
(11,1\,;\,12,3) &\equiv 11,7 \pm 0,6\\
ME &= z_{1-\alpha}\sigma_{\overline{X}} = z_{1-\alpha}\frac{\sigma_X}{\sqrt{n}}\\
0,6 &= z_{1-\alpha}\frac{2,5}{\sqrt{25}}\\
z_{1-\alpha} &= \frac{0,6}{0,5}\\
1-\alpha(\text{confiança}) &= 0,77(77\%)
\end{aligned}
$$

(e)

---

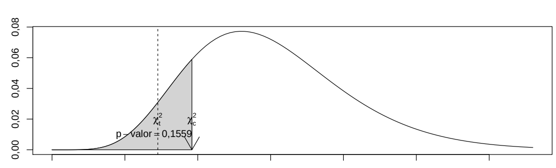
19. O tempo médio, por programador, para executar uma tarefa, tem sido 100 minutos, com um desvio padrão de 15 minutos. Introduziu-se uma modificação nos procedimentos do sistema para diminuir este tempo e também torná-lo mais homogêneo. Após certo período, sorteou-se uma amostra de 16 programadores, medindo-se o tempo para execução de cada um deles. O tempo médio da amostra foi de 85 minutos, e o desvio padrão de 12 minutos. Estes resultados trazem evidências estatísticas de melhoras tanto no tempo médio de execução quanto na maior homogeneidade dos tempos? Em cada critério (tempo médio e homogeneidade), caso afirmativo, forneça estimativas pontuais e intervalares dos parâmetros de interesse.

Solução:

i. Teste da variância

$$
\begin{aligned}
H_0: \sigma^2 \geq (15')^2 \text{ vs } H_a: \sigma^2 < (15')^2(\text{unilateral})\\
\alpha = 0,05 \longrightarrow \chi_t^2 = 7,26\\
\chi_c^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(16-1)12^2}{15^2} = 9,6\\
\text{p-valor} = 0,156
\end{aligned}
$$

Decisão: não rejeita-se H0



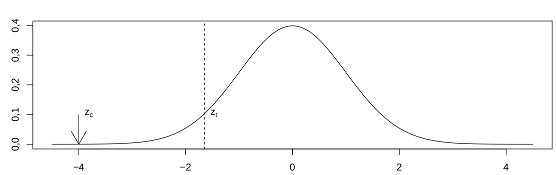
Comandos em R para soluções:

```
> (chi2t <- qchisq(0.05, df=16-1))
[1] 7,261
> (chi2c <- (16-1)*(12^2)/(15^2))
[1] 9,6
> (pvalor <- pchisq(chi2c, df=16-1))
[1] 0,1559
> curve(dchisq(x, df=16-1), from=0, to=33, xlab="", ylab="")
> polygon(cbind(c(chi2c, 0, seq(0, chi2c,l=100)),
+               c(0, 0, dchisq(seq(0, chi2c, l=100), df=16-1))),
+               col="lightgray")
> text(7, 0.01, substitute(p-valor==PV, list(PV=pvalor)))
> abline(v=chi2t, lty=2)
> arrows(chi2c, 0.02, chi2c, 0)
> text(chi2t, 0.02, expression(chi[t]^2))
> text(chi2c, 0.02, expression(chi[c]^2))
> D0 <- "não rejeita-se H0"
> Da <- "rejeita-se H0"
> ## resultado
> ifelse(chi2c > chi2t, D0, Da)
[1] "não rejeita-se H0"
> ## equivalentemente
> ifelse(pvalor > 0.05, D0, Da)
[1] "não rejeita-se H0"
```

ii. Teste da média

$$H_0 : \mu \geq 100 \text{ vs } H_a : \mu < 100 \text{(unilateral)}$$
$$\alpha = 0,05 \longrightarrow z_t = -1,64$$
$$z_c = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{85 - 100}{15/\sqrt{16}} = -4$$
$$\text{p-valor} = 0,00003$$

Decisão: rejeita-se H0
Estimativa pontual: $\hat{\mu} = 85$
Estimativa intervalar (95%): $(77,7,92,3)$



Comandos em R para soluções:

```
> (zt <- qnorm(0.05))
[1] -1,645
> (zc <- (85-100)/(15/sqrt(16)))
[1] -4
> (pvalor <- pnorm(zc))
```

```
[1] 3,167e-05
```

```
> curve(dnorm(x), from=-4.5, to=4.5, xlab="", ylab="")
> abline(v=zt, lty=2)
> arrows(zc, 0.1, zc, 0)
> text(zt, 0.1, expression(z[t]), pos=4)
> text(zc, 0.1, expression(z[c]), pos=4)
> D0 <- "não rejeita-se H0"
> Da <- "rejeita-se H0"
> ## resultado
> ifelse(zc > zt, D0, Da)
```

```
[1] "rejeita-se H0"
```

```
> ## equivalentemente
> ifelse(pvalor > 0.05, D0, Da)
```

```
[1] "rejeita-se H0"
```

```
> ## estimativa pontual
> (mu.est <- 85)
```

```
[1] 85
```

```
> ## estimativa intervalar (95%)
> (IC.mu <- mu.est + qnorm(c(0.025, 0.975)) * 15/sqrt(16))
```

```
[1] 77,65 92,35
```

---

20. Um fabricante garante que 90% dos equipamentos que fornece a uma fábrica estão de acordo com as especificações exigidas. O exame de uma amostra de 200 peças deste equipamento revelou 25 defeituosas. Teste (estatisticamente) a afirmativa do fabricante para os níveis de significância de 5 e 1%.

Solução:

$$X : \text{número de defeituosas}$$

$$X \sim \mathrm{B}(n = 200, p = 0,10)$$

Distribuição amostral utilizada:

$$\hat{p} \sim \mathrm{N}\left(\mu_p = p, \sigma_p^2 = \frac{p(1-p)}{n}\right)$$

Hipóteses:

$$H_0 : p \leq 0,10 \text{ vs } H_a : p > 0,10 (\text{unilateral})$$

Nível de significância:

$$\alpha = 0,05 \longrightarrow z_t =$$

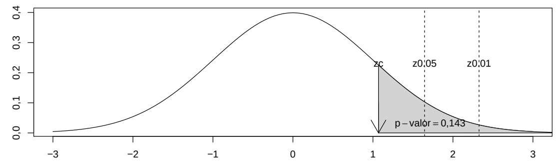
$$\alpha = 0,01 \longrightarrow z_t =$$

Estimativa pontual:

$$\hat{p} = \frac{25}{200} = 0,125$$

Estatística de teste:

$$z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,125 - 0,10}{\sqrt{\frac{0,10(1-0,10)}{200}}} = 1,07$$



Comandos em R para soluções:

```
> ## calculos passo a passo
> p.est <- 25/200
> sd.p <- sqrt(p.est*(1-p.est)/200)
> zc <- ((p.est)-0.10)/sd.p
> D0 <- "não rejeita-se H_0"
> Da <- "rejeita-se H_0"
> ## para 5%
> ifelse(zc < qnorm(0.95), D0, DA)

[1] "não rejeita-se H_0"

> ## para 1%
> ifelse(zc < qnorm(0.99), D0, DA)

[1] "não rejeita-se H_0"

> ## alternativamente, usando p p-valor
> (pv <- pnorm(zc, lower=FALSE))

[1] 0,1425

> ifelse(pv > 0.05, D0, DA)

[1] "não rejeita-se H_0"

> ifelse(pv > 0.01, D0, DA)

[1] "não rejeita-se H_0"

> ##
> ## Comandos para gráfico ilustranto resultado do teste
> curve(dnorm(x), from=-3, to=3, ylab="", xlab="")
> polygon(cbind(c(zc,seq(zc, 4, l=100),4), c(0, dnorm(seq(zc, 4, l=100)), dnorm(4))), col="lightg
> abline(v=qnorm(c(0.95, 0.99)), lty=2)
> arrows(zc, 0.1, zc, 0)
> text(c(zc, qnorm(c(0.95, 0.99))), 0.2, c("zc", "z0.05", "z0.01"), pos=3)
> text(1.2, 0.025, substitute(p-valor==PV, list(PV=round(pv, dig=3))), pos=4)
> ##
> ## teste já implementado no R
> prop.test(25, 200, 0.10, alternative="greater", conf.level=0.95)

1-sample proportions test with continuity correction

data:  25 out of 200, null probability 0.1
X-squared = 1,1, df = 1, p-value = 0,1
alternative hypothesis: true p is greater than 0,1
95 percent confidence interval:
```

```
 0,08933 1,00000
sample estimates:
    p
0,125

> prop.test(25, 200, 0.10, alternative="greater", conf.level=0.99)

1-sample proportions test with continuity correction

data:  25 out of 200, null probability 0.1
X-squared = 1,1, df = 1, p-value = 0,1
alternative hypothesis: true p is greater than 0,1
99 percent confidence interval:
 0,07831 1,00000
sample estimates:
    p
0,125
```

---

21. O tempo esperado (médio) de viagem de ônibus entre duas cidades é de 6 horas e 30 minutos com desvio padrão de 15 minutos.

    (a) Qual a probabilidade de que uma viagem tenha um atraso de mais que 20 minutos?

    (b) Se uma empresa faz 10 viagens por dia, qual a probabilidade de que tempo total das viagens seja superior a 67 horas?

Solução:
Notação:

$$\begin{aligned} X &: \text{tempo de viagem (em horas)} \\ X &\sim N(6,5; 0,25^2) \\ \overline{X} &: \text{tempo médio de 10 viagens (em horas)} \\ \overline{X}_{10} &\sim N(6,5; (0,25^2)/10) \end{aligned}$$

(a) $P[X > 6,5 + (20/60)] = P[Z > \frac{6,83-6,5}{0.25}] = P[Z > 27,33] = 0,0912$

(b) $P[\overline{X}_{10} > 6,7] = P[Z > \frac{6,7-6,5}{0.25/\sqrt{10}}] = P[Z > 84,75] = 0,00571$

$$\begin{aligned} X &: \text{tempo de viagem (em minutos)} \\ X &\sim N(390; 15^2) \\ \overline{X} &: \text{tempo médio de 10 viagens (em minutos)} \\ \overline{X} &\sim N(390; 15^2/10) \\ P[X > 410] &= 0,0912 \\ P[\overline{X} > 402] &= 0,00571 \end{aligned}$$

Solução computacional com o programa R:

```
> ita <- pnorm(6.5 + 20/60, mean=6.5, sd=0.25, lower=FALSE)
> itb <- pnorm(6.7, mean=6.5, sd=0.25/sqrt(10), lower=FALSE)
> ita1 <- pnorm(410, mean=390, sd=15, lower=FALSE)
> itb1 <- pnorm(402, mean=390, sd=15/sqrt(10), lower=FALSE)
```

22. Ainda no contexto da questão anterior, suspeitou-se que, com alterações nos trajetos e no trânsito, houve alteração nas características de tempo das viagens. Para investigar tomou-se uma amostra aleatória de 25 viagens. O tempo médio das viagens da amostra foi de 6 horas e 40 minutos e o desvio padrão de 24 minutos. Faça testes adequados para verificar se há evidência estatística de que houveram tais alterações. Caso tenham ocorrido forneça as estimativas intervalares (com confiança de 95%) para o(s) parametro(s) alterado(s).

Solução:
Solução computacional com o programa R:

```
> ## Teste para variância
> (chi2c <- (25-1)*(24^2)/(15^2))

[1] 61,44

> (chi2t <- qchisq(c(0.025, 0.975), df=25-1))

[1] 12,40 39,36

> (ICsigma <- sqrt((25-1)*(24^2)/qchisq(c(0.975, 0.025), df=25-1)))

[1] 18,74 33,39

> ## Teste para média
> (tc <- (400-390)/(24/sqrt(25)))

[1] 2,083

> (tt <- qt(c(0.025, 0.975), df=25-1))

[1] -2,064  2,064

> (pvalor <- 2*pt(tc, df=25-1, lower=FALSE))

[1] 0,04804

> (ICmu <- 400 + qt(c(0.025, 0.975), df=25-1) * 24/sqrt(25))

[1] 390,1 409,9
```

---

23. Dois medicamentos foram testados para comparar a sua eficácia no controle de infecções bucais. Para isto pacientes com infecção receberam o tratamento com um deles, selecionado aleatoriamente. O medicamento A foi utilizado em 80 pacientes tendo sido efetivo no controle para 63 deles. O medicamento B foi efetivo em 42 dentre os 55 que o receberam. Baseados nestes resultados, use um procedimento adequado para verificar se há evidência estatística de que haja diferença na eficiência dos medicamentos.

Solução:

---

24. Deseja-se fazer uma pesquisa para estimar a proporção de indivíduos de uma população que conhece um determinado aplicativo para telefones celulares. Qual deve ser o tamanho mínimo da amostra para que o erro da estimativa seja de no máximo 3% com 95% de confiança?
(Resolva sob a suposição de que será tomada uma amostra aleatória simples)

Solução:

$$\hat{p} \sim \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$
$$\text{M.E.} = z_{0,975}\sqrt{\frac{p(1-p)}{n}}$$
$$n = \left\lceil \frac{z_{0,975}^2 p(1-p)}{\text{M.E.}^2} \right.$$
$$\text{usando } p = 0,5$$
$$n = \left\lceil \frac{1,96^2 0,5(1-0,5)}{0,03^2} \right.$$
$$n = 1068$$

---

25. Em um levantamento geológico foram coletadas amostras de sedimentos de fundo de rios de uma bacia hidrográfica. Os teores obtidos de um certo elemento são mostrados a seguir.

```
35.1 5.7 5.8 4.7 1.5 2.8 70.2 3.0 8.3 6.3 17.8 16.3 7.0

1.8 1.9 5.4 4.5 4.5 19.2 8.4 9.9 6.8 2.8 21.1 1.3
```

Pretende-se usar os dados para estimar um valor que represente o teor característico da bacia. Neste contexto identifique na problema os seguintes elementos:

(a) A população e a variável aleatória em questão,
(b) O parâmetro, algum estimador e alguma estimativa
(c) Obtenha alguma estimativa intervalar com o nível de confiança de 90%
(d) Quais as suposições feitas para obtenção desta estimativa? Avalie se a estimativa é adequada para os dados obtidos.

---

26. Ainda no contexto anterior, considera-se que o teor está acima do nível de segurança se o nível na região estiver acima de 8,5 unidades e será feito um teste estatístico para determinar se a amostra indica que o nível de segurança foi ultrapassado com nível de significância de $\alpha = 0,10$.

(a) Identifique as hipóteses do estudo
(b) Descreva o que seriam os erros tipo I e II neste contexto
(c) Proceda o teste de hipótese descrevendo a conclusão no contexto do problema
(d) Suponha agora que o teor verdadeiro da região esteja em 11 unidades. e suponha ainda que o desvio padrão seja de 16 unuidades. Com estes valores, qual a probabilidade do erro tipo II?
(e) Discuta a adequação do teste considerando a natureza dos dados e as suposições feitas para realização do teste

---

27. Descreva uma situação (hipotética/inspirada em um estudo real) em que uma amostra é tomada de uma população para avaliar uma característica de interesse. Neste contexto especifique: a população (”informal”e no sentido estatístico), o(s) atributo(s) de interesse, o parâmetro de interesse, a forma como a amostra poderia ser tomada, se a forma de amostragem sugerida constitui uma amostra aleatória simples, algum estimador que poderia ser utilizado.

---

28. (B. & M.) Se uma amostra com 36 observações for tomada de uma população. qual deve ser o tamanho de uma outra amostra para que o desvio padrão da média desta amostra seja 2/3 do desvio padrão da média da primeira?

$$\begin{aligned}
\text{População: } & X \sim (E[X] = \mu, \mathrm{Var}[X] = \sigma^2) \\
\text{Amostra 1: } & \overline{X}_1 \approx \mathrm{N}(E[\overline{X}_1] = \mu, \mathrm{Var}[\overline{X}_1] = \sigma^2/36) \\
\text{Amostra 2: } & \overline{X}_2 \approx \mathrm{N}(E[\overline{X}_2] = \mu, \mathrm{Var}[\overline{X}_2] = \sigma^2/n_2) \\
\frac{\sigma}{\sqrt{n_2}} & = \frac{2}{3}\frac{\sigma}{\sqrt{36}} \\
n_2 & = 81
\end{aligned}$$

---

29. Em um experimento foram feitas 24 tentativas (independentes) de resolver um problema até que a terceira solução correta fosse encontrada. Encontre os estimadores de (i) métodos dos momentos; (ii) de máxima verossimilhança do parâmetro p que representa a probabilidade de obter a solução correta em uma tentativa qualquer.

$$\begin{aligned}
X & : \text{número de "falhas" até obter a terceira solução correta} \\
X & : \sim \mathrm{BN}(r = 3, p) \\
P[X = x] & \binom{x+r-1}{r-1} p^3 (1-p)^x \\
\text{amostra} & : \quad x = 21
\end{aligned}$$

método dos momentos :

$$\begin{aligned}
E[X] & = r\frac{(1-p)}{p} & = \overline{x} \\
3\frac{(1-\hat{p})}{\hat{p}} & = 21 \\
\hat{p} & = \frac{3}{24} = \frac{1}{8} = 0,125
\end{aligned}$$

método da máxima verossimilhança :

$$\begin{aligned}
L(p) & = \binom{31+3-1}{3-1} p^3 (1-p)^{21} \\
l(p) & = \log[L(p)] = \log\left(\binom{31+3-1}{3-1}\right) + 3\log(p) + 21\log(1-p) \\
\frac{\mathrm{d}l(p)}{\mathrm{d}p} & = \frac{3}{(\hat{p})} + \frac{21}{1-\hat{p}}(-1) = 0 \\
\hat{p} & = \frac{3}{24} = \frac{1}{8} = 0,125
\end{aligned}$$

---

30. O tempo para processamento e análise de determinado tipo de imagem médica possui distribuicão normal de média 5,2 segundo e variância de 0,64 segundos^2.

   (a) Se 1000 imagens serão analisadas, em quantas delas espera-se um tempo de processamento acima da 5,5 segundos?

   (b) Se as imagens forem agrupadas em lotes de 10, em quantos lotes espera-se um tempo de processamento acima da 55 segundos?

   (c) Considere agora que deseja-se determinar o número k de imagens para compor os lotes. Quantas imagens dever haver em um lote para que a probabilidade do tempo médio de processamento das k imagens de um lote ultrapassar 5,5 segundos seja de no máximo 0,02?

Solução:

$$X \sim \mathrm{N}(\mu = 5,2; \sigma^2 = 0,64)$$

(a) $1000 \cdot P[X > 5,5] = 1000 \cdot P[Z > \frac{5,5-5,2}{0,8}] = 1000 \cdot P[Z > 0,375] = 1000 \cdot [0,5 - 0,14617] = 354$

(b)

$$\overline{X} \sim \mathrm{N}(\mu_{\overline{X}} = 5,2; \sigma^2_{\overline{X}} = \frac{0,64}{10})$$

$$100 \cdot P[\sum_{i=1}^{10} > 55] = 100 \cdot P[\overline{X} > 5,5] = 100 \cdot P[Z > \frac{5,5-5,2}{0,8/\sqrt{10}}] = 100 \cdot P[Z > 1,19] = 100 \cdot [0,5 - 0,38216] = 12$$

(c)

$$P[\overline{X}_k > 5,5] = P[\sum_{i=1}^{k} > 55] \leq 0,02$$

$$P[Z > \frac{5,5-5,2}{0,8/\sqrt{k}}] \leq 0,02$$

$$z = \frac{5,5-5,2}{0,8/\sqrt{k}} \geq 2,05$$

$$k \geq \frac{0,8^2 2,05^2}{(5,5-5,2)^2}$$

$$k \leq 30$$

---

31. Um laboratório tomou uma amostra de 1500 materiais contaminados com uma determinada bactéria a fim de estimar a proporção θ de resistência na população a um determinado antibiótico;. Nos testes laboratoriais encontrou-se que 260 das amostras apresentavam resistência.

(a) Identifique no contexto deste problema a população (”informal” e estatística), a amostra, o parâmetro de interesse o estimador e a estimativa.

(b) Obtenha os intervalos de confiança (assintótico e conservador) para a proporção de resistência na população.

(c) Qual a suposição que foi feita a respeito da amostragem para o cálculo do intervalo de confiança. Discuta a adequação da suposição no contexto do problema.

Solução:

(a) População: materiais (contaminações) com a bactéria
População estatística: X : bactéria resistente nos materiais contaminados, $X \sim \mathrm{B}(\theta)$
Amostra: os 260 materiais colhidos e testados para resistência
Parâmetro: θ (proporção de materiais com bactérias resistêntes na população)
Estimador: $\hat{\theta} = \overline{x} = \frac{\sum X_i}{n}$ (proporção de materiais contaminados e com bactérias resistentes)
Estimativa : $\hat{\theta} = \frac{260}{1500} = 0,173$

(b) IC (95%)

$$\text{conservador: } \hat{\theta} \pm z\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$
$$0,173 \pm 1,96\sqrt{\frac{0,173(1-0,173)}{1500}}$$
$$0,173 \pm 0,0191$$
$$\text{assintótico: } \hat{\theta} \pm z\sqrt{\frac{0,5(1-0,5)}{n}}$$
$$0,173 \pm 1,96\sqrt{\frac{1}{4 \cdot 1500}}$$
$$0,173 \pm 0,0253$$

(c) Suposição de que foi tomada uma amostra aleatória simples, portanto as observações são independentes. (DISCUTIR NO CONTEXTO DO PROBLEMA)

---

32. Considere contexto da questão sobre o tempo de processamento e análise das imagens. Um novo algorítmo foi proposto para o processamento e análise das imagens e testado no processamento de 25 imagens. Os tempos de processamento são dados a seguir.

```
4,29  4,12  6,09  4,09  4,96  5,70  4,49  4,20  5,30  4,55  6,33  4,97  5,03
4,66  6,15  4,82  4,17  5,18  7,94  4,10  5,19  5,75  4,36  4,70  5,28
```

(a) Obtenha a média, desvio padrão e coeficiente de variação dos dados.

(b) Obtenha a mediana, quartis, amplitude e amplitude interquartílica.

(c) Obtenha o diagrama *box-plot* e comente sobre a distribuição dos dados.

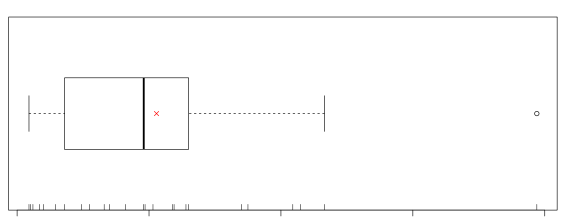
(d) Faça suposições adequadas e teste a hipótese de que o tempo médio com o novo algorítmo é inferior ao que vinha sendo utilizado.

Solução:

(a) $\overline{x} = 5,0568$; $S_x = 0,901178487685246$; $\text{CV}_x = 17,821121809944\%$

(b) $\text{md}(x) = Q_2 = 4,96$; $Q_1 = 4,36$; $Q_3 = 5,3$; $\text{A}_x = 3,85$; $\text{AI}_x = 0,94$

(c)



Comentar sobre: valor central, dispersão, assimetria e presença (ou não) de dado discrepante.

(d) Teste de hipótese sobre média com variância conhecida

i. $H_0 : \mu = 5,2$ vs $H_1 : \mu < 5,2$ (unilateral)

ii. $\alpha = 0,05$

Com todas as observações :

iii. $z_c = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{5{,}0568-5{,}2}{0{,}8/\sqrt{25}} = -0,895$

iv. $z_t = 1,64$, ou então $p - valor = 0,185$

v. Conclusão: $|z_c| < z_t$ ou, equivalentemente, $p - valor > \alpha(0,05)$, portanto não se rejeita H_0 ou seja não há evidência (com estes dados) de que o tempo tenha sido reduzido, para o nível de significância de 5%. Excluindo a observaçâo discrepante:

iii. $z_c = \frac{\overline{x}^*-\mu_0}{\sigma/\sqrt{n}} = \frac{4{,}937-5{,}2}{0{,}8/\sqrt{24}} = -1,61$

iv. $z_t = 1,64$, ou então $p - valor = 0,0534$

v. Conclusão: $|z_c| > z_t$ ou, equivalentemente, $p - valor < \alpha = 0,05$, portanto não rejeita-se H_0, ou seja, não há evidência (com estes dados, excluindo-se a observação atípica) de que o tempo tenha sido reduzido, para o nível de significância de 5%.

---

33. Sabe-se que o tempo de uso de um determinado serviço pode ser descrito pela distribuição exponencial e tem média de 30 segundos. Qual a probabilidade de um acesso:

(a) ter o tempo inferior a 12 segundos;

(b) durar entre 12 e 30 segundos;

(c) ultrapassar 42 segundos, sabendo-se que já chegou a 30 segundos?

Os acessos são divididos em três categorias: rápidos (com tempo inferior a 12 segundos), médios (com tempo entre 12 e 30 segundos), longos (com tempos entre 30 e 40 segundos) e demorados (com tempo superior a 42 segundos). Os custos e cada categoria são definidos como sendo, respectivamente 1,5; 2,8; 5,0 e 10,0 unidades de custo.

(d) Qual o custo médio dos acessos?

Se for tomada uma amostra de 100 tempos, qual a probabilidade de que o tempo médio desta amostra:

(e) ultrapasse 35 segundos;

(f) esteja entre 28 e 35 segundos?

(g) Qual deve ser o tamanho da amostra para que a probabilidade do tempo médio da amostra ultrapassar 35 segundos seja no máximo de 0,01?

Desconfia-se que houve alguma mudança no padrão dos tempos de acesso e para verificar isto tomou-se uma amostra de 100 tempos de acesso. A seguir estão alguns desses valores.

```
16,8   2,2   16,1   9,5   19,5   1,5   26   6,4   18,9   25,6   ...   ... 26,1   9,5   10,2
```

A média dos tempos amostrados é de 32,71.

(h) Obtenha um intervalo de confiança (95%) para o tempo médio.

(i) Voce diria que há evidências fortes de que houve uma mudança no comportamento dos tempos de acesso? Justifique.

Solução:

$$X : \text{tempo de acesso ao serviço}$$

$$X \sim \mathrm{E}(\lambda = 1/30) \quad f(x) = \frac{1}{30}\exp\{-x/30\} \quad F(x) = 1 - \exp\{-x/30\}$$

(a) $P[X < 12] = \int_{x=0}^{12} f(x)\mathrm{d}x = F(12) = 0,3297$

(b) $P[12 < X < 30] = \int_{x=12}^{30} f(x)\mathrm{d}x = F(30) - F(12) = 0,3024$

(c) Pela propriedade de falta de memória de exponencial,
$P[X > 42|X > 30] = P[X > 12] = 1 - P[X < 12] = 1 - \int_{x=0}^{12} f(x)\mathrm{d}x = 1 - F(12) = 0,6703$

(d) Y : custo de acesso

y	$1,5$	$2,8$	$5,0$	10
$P[Y=y]$	$P[X<12]=0,33$	$P[12<X<30]=0,302$	$P[30<X<42]=0,121$	$P[X>42]=0,247$

$\mathrm{E}(X)=\sum_i y_i P[Y=y_i]=1,5\cdot 0,33+2,8\cdot 0,302+5,0\cdot 0,121+10,0\cdot 0,247=4,4$ unidades de custo

$$X \sim \mathrm{Exp}(\lambda=1/30) \quad \mathrm{E}(X)=1/\lambda \quad \mathrm{Var}(X)=1/\lambda^2$$
$$\overline{X}_n \approx N(\mu=1/\lambda=30, \sigma^2=(1/\lambda^2)/n=30^2/n) \quad n=100$$

(e) $P[\overline{X}>35]=P[Z>\frac{35-30}{30/10}]=P[Z>1,667]=0,048$

(f) $P[28<\overline{X}<35]=P[\frac{28-30}{30/10}<Z<\frac{35-30}{30/10}]=P[-0,667<Z<1,667]=0,7$

(g)

$$\begin{aligned} P[\overline{X}>35] &\leq 0,01 \\ P[Z>\frac{35-30}{30/\sqrt{n}}] &\leq 0,01 \\ \frac{35-30}{30/\sqrt{n}} &\geq 2,326 \\ n &\geq \frac{2,326^2 30^2}{(35-30)^2}=195 \end{aligned}$$

$$\text{Assumindo que } \overline{X}_n \approx N(\mu=1/\hat{\lambda}=30, \sigma^2=(1/\hat{\lambda}^2)/n=30^2/n)$$

(h) $\hat{\lambda} \pm z_{0,95}\hat{\lambda}/\sqrt{n} \longrightarrow 32,71 \pm 1,96\frac{32,71}{10} \longrightarrow (26,3 \;;\; 39,12)$

(i) Justifique.

---

34. Foi feito um estudo na área de uma baía para verificar a salinidade (superficial) da água no local. No estudo foram tomadas amostras da água em 47 pontos da baía. Cada amostra foi analisada e o resultado de interesse é o nível de salinidade. O valor médio das 47 amostras foi de 28,35 $^o/^{oo}$. Neste contexto, identifique e descreva os seguintes elementos segundo os conceitos de estatística.

(a) a população,

(b) a variável aleatória de interesse,

(c) o parâmetro de interesse,

(d) a amostra,

(e) o estimador,

(f) a estimativa,

(g) a distribuição amostral.

Discuta ainda se, ou sob quais condições, a amostra pode ser considerada uma *amostra aleatória simples*.

---

35. Uma locadora de veículos que possui uma grande frota decide fazer um estudo sobre vários aspectos relacionados ao desempenho. Para isto vai tomar uma amostra aleatória de 25 de seus veículos para inspeções detalhadas. Várias características serão medidas, mas vamos aqui nos ater apenas ao consumo de combustível, supondo que a variância do consumo de toda a frota é de $6,25 km/l$.

(a) Qual a probabilidade do consumo médio aferido nos 25 veículos, diferir do consumo médio de toda a frota em mais que $0,5 km/l$? E em mais que $1 km/l$?

(b) Qual a margem de erro na estimação do consumo médio da frota para uma confiança de 95%?

(c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse a metade da calculada no item anterior?

(d) Se uma amostra (com $n = 25$) fornecer uma estimativa intervalar de $(11,1\,;\,12,3) km/l$, qual a confiança desta estimativa?

(e) identifique no problema: a população variável aleatória de interesse, o parâmetro de interesse, o estimador, a distribuição amostral, a estimativa pontual e a estimativa intervalar.

Solução:

$$X : \text{consumo de veículo da frota}$$

Distribuição da variável aleatória (população):

$$X \sim \text{Dist.}(\mu_X = \text{E}[X], \sigma_X^2 = \text{Var}[X])$$

distribuição amostral:

$$\overline{X} \approx \text{N}(\mu_{\overline{X}} = \mu_X, \sigma_{\overline{X}} = \sigma_X^2/n)$$

ou, equivalentemente,

$$\overline{X} - \mu \approx \text{N}(0, \sigma^2/n)$$

$$\sigma_X^2 = 6,25\ ;\ \sigma_{\overline{X}}^2 = 0,25\ ;\ \sigma_{\overline{X}} = 0,5$$

(a)

$$P[|\overline{X} - \mu| > 0,5] = P[\frac{|\overline{X} - \mu|}{\sigma} > 0,5/\sigma] = P[|Z| > 1] = 0,317$$

$$P[|\overline{X} - \mu| > 1] = P[\frac{|\overline{X} - \mu|}{\sigma} > 1/\sigma] = P[|Z| > 2] = 0,0455$$

(b)

$$ME = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}}$$

$$ME = 1,96\frac{2,5}{\sqrt{25}}$$

$$ME = 0,98$$

(c)

$$\frac{ME}{2} = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n^*}}$$

$$n^* = \lceil(\frac{2}{ME})^2 z_{0,95}^2 \sigma_X^2$$

$$n^* = \lceil(\frac{2}{0,98})^2 1,96^2 6,25$$

$$n^* = 100$$

(d)

$$(11,1\,;\,12,3) \equiv 11,7 \pm 0,6$$

$$ME = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}}$$

$$0,6 = z_{1-\alpha}0,5$$

$$z_{1-\alpha} = \frac{0,6}{0,5}$$

$$1 - \alpha(\text{confiança}) = 0,77(77\%)$$

(e)

---

36. Considere um estudo na qual deseja-se estimar a proporção de solicitações atendidas e resolvidas de uma central do usuário através de uma amostra aleatória simples.

   (a) Qual a população, variável aleatória e parâmetro de interesse?
   (b) Mostre como obter o estimador de máxima verossimilhança deste parâmetro.
   (c) Se a amostra for de 4000 solicitações, qual será a margem de erro (com confiança de 95%) para a estimação da proporção de resolvidas?
   (d) Para este mesmo tamanho de amostra, qual seria a confiança associada a uma margem de erro de $\pm 0,01$ (1 %)?
   (e) Qual deveria ser o tamanho da amostra para se obter a estimativa com uma margem de erro de 2,5% com 95% de confiança?
   (f) E para uma margem de erro de 3% com 99% de confiança?

Solução:

(a) População: solicitações atendidas, representada pela v.a.:

$$X : \text{resolução de solicitações atendidas}$$

que é binária e pode ser codificada assumindo valores 0, para não atendidas e, 1, para atendidas. O parâmetro de interesse é a proporção de solicitações atendidas, que representa a probabilidade p de uma solicitação ser atendida. Assume-se então que:

$$X \sim B(p)$$

e os resultados das análises se baseiam no Teorema:

$$\text{Teo 2:} \quad \hat{p} = \overline{X} \approx N(p, \frac{p(1-p)}{n})$$

do qual se extraem os resultados:

$$\hat{p} \pm z_{(1-\alpha)} \sqrt{\frac{p(1-p)}{n}}$$

$$\hat{p} \pm \text{ME}$$

$$\text{ME} = z_{(1-\alpha)} \sqrt{\frac{p(1-p)}{n}}$$

$$n = \frac{(z_{(1-\alpha)})^2}{\text{ME}^2} p(1-p)$$

$$p(1-p) \text{ é limitado superiormente para } p = 0,5$$

$$n = \frac{(z_{(1-\alpha)})^2}{\text{ME}^2} 0,25$$

(b) Seja a v.a. Y : número de solicitações resolvidas dentre n atendidas.

$$Y \sim \text{Bin}(n, p))$$

A função de verossimilhança é dada pela expressão de distribuição de Y, portanto:

$$L(p) = \binom{n}{y} p^y (1-p)^(n-y)$$

e a log-verossimilhança:

$$l(p) = \log \binom{n}{y} + y \log(p) + (n-y) \log(1-p).$$

O estimador de máxima verossimilhança de p é o ponto de máximo função (de p) acima e portanto encontrado derivando-se a expressão acima,

$$\frac{\mathrm{d}l(p)}{\mathrm{d}p} = \frac{y}{p} - \frac{n-y}{1-p}$$

que igualada a zero nos leva a:

$$\hat{p} = \frac{y}{n}.$$

Note-se que a derivada segunda fornece um valor negativo e valor $\hat{p}$ é então o estimador de máxima verossimilhança do parâmetro p.

(c) $M.E. = z_{(1-\alpha)}\sqrt{\frac{p(1-p)}{n}} = 1,96\sqrt{\frac{0,5(1-0,5)}{4000}} = 0,0155$

(d)

$$\begin{aligned}
&M.E. = z_{(1-\alpha)}\sqrt{\frac{p(1-p)}{n}}\\
&z_{(1-\alpha)} = 0,01/\sqrt{\frac{0,5(1-0,5)}{4000}} = 1,26\\
&1-\alpha = 79,4\%
\end{aligned}$$

(e) $n = \frac{(1{,}96)^2}{(0{,}025)^2}0,25 = 1537$

(f) $n = \frac{(2{,}576)^2}{(0{,}03)^2}0,25 = 1844$

---

37. O tempo que passageiros esperam no balcão de *check-in* de uma cia aérea é uma variável aleatória com média de 13,5 minutos e desvio padrão de 1,8 minutos. Vai ser tomada uma amostra aleatória do tempo de espera de 36 passageiros. Encontre a probabilidade:

(a) do tempo de espera de um passageiro escolhido ao acaso ser maior que 15 min;

(b) do tempo médio dos 36 passageiros estar acima de 15 min;

(c) do tempo médio de atendimento dos 36 passageiros estar entre 13 e 14 minutos.

(d) Se o tempo médio de espera fosse desconhecido e a ser estimado pela amostra, qual seria a amplitude do intervalo de confiança a 95% ?

(e) Suponha agora que a amostra dos tempos dos 36 foi tomada para verificar se houve alteração no tempo médio de espera. Use um procedimento estatístico adequado para verificar se é possível afirmar se houve alteração, para um tempo médio na amostra de 14,4 minutos?

Solução:

$$X \sim N(\mu = 13,5, \sigma^2 = 1,8^2)$$
$$\bar{X_{36}} \sim N(\mu = 13,5, \sigma^2 = 1,8^2/36)$$

(a) $P(X > 15) = P(Z > \frac{15-13{,}5}{1{,}8}) = P(Z > 0,833) = 0,202$

(b) $P(\bar{X} > 15) = P(Z > \frac{15-13{,}5}{1{,}8/\sqrt{36}}) = P(Z > 5) = 3e-07$

(c) $P(13 < \bar{X} < 14) = P(\frac{13-13{,}5}{1{,}8/\sqrt{36}} < Z < \frac{14-13{,}5}{1{,}8/\sqrt{36}}) = P(-1,667 < Z < 1,667) = 0,904$

(d)

$$\begin{aligned}
\text{M.E.} &= z_{0,95}\frac{\sigma}{\sqrt{n}}\\
\text{M.E.} &= 1,96\frac{1,8}{\sqrt{36}}\\
\text{M.E.} &= 0,588\\
\text{amplitude} &= 2\cdot\text{M.E.} = 1,18
\end{aligned}$$

(e) Duas possíveis soluções: pode-se obter um intervalo de confiança para estimativa e verificar se o intervalo inclui o valor 13,5 ou efetuar um teste de hipótese. Apresenta-se a seguir a segunda solução.

$$\begin{aligned} H_0 : \mu = 13,5 \text{ vs } H_a : \mu \neq 13,5 \\ \alpha = 0,05 \\ z_c = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{14,4 - 13,5}{1,8/\sqrt{36}} = 3 \\ \text{valor} - \text{p} = 0,0027 \\ \text{rejeita-se } H_0 \\ \text{para o nível de significância de 5\%,} \\ \text{afirma-se que a média se alterou} \end{aligned}$$

---

38. Uma locadora de veículos que possui uma grande frota decide fazer um estudo sobre vários aspectos relacionados ao desempenho. Para isto vai tomar uma amostra aleatória de 25 de seus veículos para inspeções detalhadas. Várias características serão medidas, mas vamos aqui nos ater apenas ao consumo de combustível, supondo que a variância do consumo de toda a frota é de $6,25 km/l$.

(a) Qual a probabilidade do consumo médio aferido nos 25 veículos, diferir do consumo médio de toda a frota em mais que $0,5km/l$? E em mais que $1km/l$?

(b) Qual a margem de erro na estimação do consumo médio da frota para uma confiança de 95%?

(c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse a metade da calculada no item anterior?

(d) Se uma amostra (com $n = 25$) fornecer uma estimativa intervalar de $(11,1\,;\,12,3)km/l$, qual a confiança desta estimativa?

(e) identifique no problema: a população variável aleatória de interesse, o parâmetro de interesse, o estimador, a distribuição amostral, a estimativa pontual e a estimativa intervalar.

Solução:

$$\begin{aligned} & X : \text{consumo de veículo da frota} \\ \text{Distribuição da variável aleatória (população):} & \\ & X \sim \text{Dist.}(\mu_X = \text{E}[X], \sigma^2_X = \text{Var}[X]) \\ \text{distribuição amostral:} & \\ & \overline{X} \approx \text{N}(\mu_{\overline{X}} = \mu_X, \sigma_{\overline{X}} = \sigma^2_X/n) \\ \text{ou, equivalentemente,} & \\ & \overline{X} - \mu \approx \text{N}(0, \sigma^2/n) \\ & \sigma^2_X = 6,25\ ;\ \sigma^2_{\overline{X}} = 0,25\ ;\ \sigma_{\overline{X}} = 0,5 \end{aligned}$$

(a)

$$P[|\overline{X} - \mu| > 0,5] = P[\frac{|\overline{X} - \mu|}{\sigma} > 0,5/\sigma] = P[|Z| > 1] = 0,317$$

$$P[|\overline{X} - \mu| > 1] = P[\frac{|\overline{X} - \mu|}{\sigma} > 1/\sigma] = P[|Z| > 2] = 0,0455$$

(b)

$$ME = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}}$$
$$ME = 1,96\frac{2,5}{\sqrt{25}}$$
$$ME = 0,98$$

(c)

$$\frac{ME}{2} = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n^*}}$$
$$n^* = \lceil (\frac{2}{ME})^2 z_{0,95}^2 \sigma_X^2$$
$$n^* = \lceil (\frac{2}{0,98})^2 1,96^2 6,25$$
$$n^* = 100$$

(d)

$$(11,1\,;\,12,3) \equiv 11,7 \pm 0,6$$
$$ME = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}}$$
$$0,6 = z_{1-\alpha}0,5$$
$$z_{1-\alpha} = \frac{0,6}{0,5}$$
$$1-\alpha(\text{confiança}) = 0,77(77\%)$$

(e)

---

39. Considere que vai ser feito um estudo sobre o tempo de persistência de um determinado serviço medindo-se, por um ano, o intervalo de tempo entre interrupções. Os resultados serão utilizados para conclusões sobre este e outros serviços similares. Será considerado que os tempos seguem uma distribuição Weibull de probabilidades que possui densidade:

$$X \sim Wei(\alpha,\beta) \qquad f(x) = \frac{\alpha}{\beta}(x/\beta)^{\alpha-1}\exp\{-(x/\beta)^{\alpha}\} \quad x > 0.$$

Neste contexto identifique a população e a variável aleatória, a amostra, discuta suas características e limitações, as suposições que serão feitas, identifique os parâmetros a serem estimados e descreva como seriam os procedimentos para ao menos duas formas de estimação destes parâmetros.

---

40. Foi feito um estudo na área de uma baía para verificar a salinidade (superficial) da água no local. No estudo foram tomadas amostras da água em 47 pontos da baía. Cada amostra foi analisada e o resultado de interesse é o nível de salinidade. O valor médio das 47 amostras foi de 28,35 $^o/^{oo}$ e o desvio padrão de 8,35 $^o/^{oo}$.

(a) Identifique e descreva, neste contexto, os seguintes elementos segundo os conceitos de estatística.

i. a população,

ii. a variável aleatória de interesse,

iii. o parâmetro de interesse,

iv. a amostra,

v. o estimador,

vi. a estimativa,

vii. a distribuição amostral.

(b) Discuta ainda se, ou sob quais condições, a amostra pode ser considerada uma *amostra aleatória simples*.

(c) Obtenha um intervalo de confiança (99%) para a salinidade média.

(d) O valor de 27 $^{o}/^{oo}$ é considerado um limite de normalidade para a salinidade nesta região. Efetue um teste estatístico de hipótese para verificar se pode-se afirmar, com um nível de significância de 5%, que a salinidade média ultrapassou tal valor de referência.

---

41. O volume de dados transmitido por dia em uma rede são independentes e possuem distribuição normal com média de 240 e variância de 900 unidades.

(a) Em quantos dias por ano (365 dias) espera-se que o volume de dados transmitido ultrapasse 300 unidades?

(b) Qual o volume deve ser transmitido em pelo menos 75% dos dias?

(c) Adota-se como dias *usuais* os que possuem valores ao redor da média em 80% dias dias. Quais volumes determinam esses limites?

(d) Qual a probabilidade do volume total transmitido em uma semana estar acima de 1800 unidades?

(e) Qual a probabilidade do volume de transmissão estar acima de 265 unidades em cinco dias consecutivos?

(f) Qual a probabilidade do volume diário não ultrapassar 250 unidades nenhuma vez em uma semana?

(g) Partindo de um dia qualquer, qual a probabilidade de ser necessário esperar mais que 3 dias para que o volume diário ultrapasse 280 unidades?

(h) Atribui-se um custo por uso da banda de 10 u.m. (unidades monetárias) para dias com volume abaixo de 200 unidades, 15 u.m. para dias com volume entre 200 e 250, 20 u.m. para dias com volume entre 250 e 280 u.m. e 30 u.m. para dias com volume acima de 280 u.m..

i. Monte a distribuição de probabilidades do custo diário.

ii. Qual o custo esperado em um mês (30 dias)?

Solução:

```
> # a)
> 365*pnorm(300, m=240, sd=30, lower=FALSE)

[1] 8,304

> # b)
> qnorm(0.25, m=240, sd=30)

[1] 219,8

> # c)
> qnorm(c(0.1, 0.9), m=240, sd=30)

[1] 201,6 278,4

> # d)
> c(pnorm(1800, m=240*7, sd=sqrt(7)*30, lower=FALSE), pnorm(1800/7, m=240, sd=30/sqrt(7), lower=F

[1] 0,06529 0,06529

> # e)
> (p <- pnorm(265, m=240, sd=30, lower=FALSE)); p^5

[1] 0,2023
```

```
[1] 0,0003391

> # f)
> (p <- pnorm(250, m=240, sd=30)); p^7

[1] 0,6306

[1] 0,03963

> # g)
> pgeom(4, prob=pnorm(280, m=240, sd=30, low=F))

[1] 0,3801

> Pr <- diff(pnorm(c(-Inf, 200, 250, 280, Inf), m=240, sd=30))
> names(Pr) <- c(10,15,20,30)
> # h1)
> Pr

     10      15      20      30
0,09121 0,53935 0,27823 0,09121

> # h2)
> 30*sum(Pr * c(10,15,20,30))

[1] 519,1
```

---

42. Considere o seguinte problema (adaptado de Magalhães & Lima, 2006):
Um fabricante afirma que sua vacina contra gripe imuniza em 80% dos casos. Uma amostra de 25 indivíduos entre os que tomaram a vacina foi sorteada e testes foram feitos para verificar a imunização ou não desses indivíduos.

   (a) No contexto do problema identifique:
      - a população,
      - o parâmetro de interesse,
      - o estimador,
      - a estimativa,
      - a distribuição amostral.

   (b) Se o fabricante estiver correto, qual é a probabilidade de proporção de imunizados ser inferior a 0,75? E superior a 0,85?

   (c) Quais seriam as estimativas pontuais e intervalares se fosse observados 18 imunizados dentre os 25 avaliados?

   (d) Qual deveria ser o tamanho da amostra em um novo estudo para que a margem de erro fosse de no máximo 0,03?

   (e) Suponha que se adote o critério de refutar a afirmativa do fabricante caso sejam observados 17 ou menos não imunizados. Qual a probabilidade de refutar a afirmativa mesmo quando ela é verdadeira (a vacina de fato imuniza 80%)?

   (f) Suponha agora que a imunização real seja de apenas 70%. Qual a probabilidade de mesmo assim não refutar a afirmativa do fabricante?

Solução:

$$X: \text{ imunizado (0: não, 1: sim)}$$
$$x \in \{0,1\}$$
$$X \sim \text{Ber}(p)$$

(a)

- Os indivíduos que receberam a vacina.
- A proporção (p) de indivíduos imunizados entre todos os que receberam a vacina (*na população*).
- O cálculo da proporção de indivíduos imunizados na amostra $\hat{p} = \sum_{i=1}^{n} X_i/n$.
- A proporção observada em uma determinada amostra (no caso na amostra de $n = 25$ indivíduos).
- A distribuição amostral do estimador, ou seja, a distribuição que seria obtida caso fossem obtidas estimativas de *diversas* amostras.

Distribuição exata:

$$Y = \sum_{i=1}^{n} x_i \sim \text{B}(n,p)$$

Distribuição exata:

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} \sim \text{N}(p, \frac{p(1-p)}{n})$$

(b)

$$\hat{p} \sim \text{N}\left(p = 0,80\ ,\ \frac{p(1-p)}{n} = \frac{0,80(1-80)}{25}\right)$$

$$P[\hat{p} < 0,75 | p = 0,80] = P[Z < \frac{0,75 - 0,80}{\sqrt{\frac{0,80(1-80)}{25}}}] = P[Z < -0,625] = 0,266$$

$$P[\hat{p} > 0,85 | p = 0,80] = P[Z < \frac{0,85 - 0,80}{\sqrt{\frac{0,80(1-80)}{25}}}] = P[Z < 0,625] = 0,266$$

(c)

- Usando $p = 0,80$

$$\hat{p} = \frac{18}{25} = 0,72$$
$$(\text{I.C.95\%})\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$$
$$0,72 \pm 1,96\sqrt{\frac{0,80(1-0,80)}{25}}$$
$$0,72 \pm 0,157$$
$$(0,563\ \ ;\ \ 0,877)$$

-

$$\hat{p} = \frac{18}{25} = 0,72$$
$$(\text{I.C.95\%}) : \hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$$
$$0,72 \pm 1,96\sqrt{\frac{0,72(1-0,72)}{25}}$$
$$0,72 \pm 0,176$$
$$(0,544\ \ ;\ \ 0,896)$$

(d)

$$z\sqrt{\frac{p(1-p)}{n}} = 0,03$$
$$1,96\sqrt{\frac{0,80(1-0,80)}{n}} = 0,03$$
$$n = \frac{1,96^2}{0,03^2}0,80(1-0,80) = 683$$

(e)

$$P[X \leq 17|p=0,80] \overset{\text{Binomial}}{=} \sum_{i=0}^{17}\binom{25}{i}(0,8)^i(1-0,8)^{25-i} = 0,109$$
$$\overset{\text{normal}}{\approx} P[\hat{p} \leq \frac{17,5}{25} = 0,7] = 0,106$$

(f)

$$P[X > 17|p=0,70] \overset{\text{Binomial}}{=} \sum_{i=18}^{25}\binom{25}{i}(0,7)^i(1-0,7)^{25-i} = 0,512$$
$$\overset{\text{normal}}{\approx} P[\hat{p} \leq \frac{17,5}{25} = 0,7] = 0,5$$

---

43. Imagens de satélite precisam ser (pós-)processadas para um certo objetivo. O tempo médio de processamento é de 96 min com desvio padrão de 18 minutos. Uma cena de uma determinada região é montada com o processamento de 12 imagens acrescido de 8 minutos para montagem da cena.

(a) Se as imagens forem processadas sequencialmente, qual a probabilidade do tempo de processamento ultrapassar 20 horas?

(b) Se as imagens forem processadas paralelamente, qual a probabilidade do tempo médio de processamento ficar entre 92 e 98 min?

(c) Quando o processamento é feito em paralelo a montagem só é feita após todos os processamentos estarem encerrados. Qual a probabilidade de ser necessário esperar mais que 100 minutos para iniciar a montagem?

Solução:
Notação:

$$X : \text{tempo de processamento}$$
$$X \ \text{Distribuição}(\mu = \mathrm{E}[X] = 96, \sigma^2 = \mathrm{Var}[X] = 18^2)$$
$$\overline{X} \approx \mathrm{N}(\mu = \mathrm{E}[X] = 96, \sigma^2/n = 18^2/12)$$

(a) $P[\sum_{i=1}^{n} X_i > 20 \cdot 60] = P[\overline{X} > 100] = P[Z > \frac{100-96}{\sqrt{18^2/12}}] = P[Z > 0,77 = 0,221$

(b) $P[92 < \overline{X} < 98] = P[\frac{92-96}{\sqrt{18^2/12}} < Z < \frac{98-96}{\sqrt{18^2/12}}] = P[-0,77 < Z < 0,385] = 0,429$

(c) $P[(X_1 > 100) \cup (X_2 > 100) \ldots (X_{12} > 100)] = 1 - P[(X_1 < 100) \cap (X_2 < 100) \ldots (X_{12} < 100)] = 1 - (P[(X < 100)])^{12} = 1 - (P[(Z < 0,77)])^{12} = 1 - (0,588)^{12} = 0,998$

---

44. No contexto anterior foi proposto um novo algorítmo para o processamento. O algorítmo foi avaliado processando-se 15 imagens. Os tempo obtidos foram:

`102,4 104,2 102,8  67,6 106,2  88,7  85,3  82,1 102,8 102,5  99,7  85,2 85,8 120,0 117,5.`

(a) Monte um IC (95%) para a nova média.

(b) Monte um IC (95%) para o novo desvio padrão.

(c) Teste a hipótese ($\alpha = 0,05$) que houve redução no tempo médio de processamento.

(Faça e declare as suposições necessárias)

Solução:
Notação:

$$Y : \text{novo tempo de processamento}$$
$$X \sim \text{Distribuição}(\mu_Y, \sigma_Y^2)$$
$$\text{Estimativas com a amostra} : \overline{y} = 96,9 \quad ; \quad S^2 = 198 \quad ; \quad S = 14,1 \quad ;$$

(a)

- Supondo $\sigma_Y^2 = \sigma^2 = 18^2$ (variância do novo algorítmo igual a do anterior)

$$\overline{x} \pm z\frac{\sigma}{\sqrt{n}}$$
$$96,9 \pm 1,96\frac{18}{\sqrt{15}}$$
$$96,9 \pm 9,1$$
$$(87,7 \ ; \ 106)$$

- Supondo σ_Y^2 estimado por S^2 (variância desconhecida do novo algorítmo, estimada pela variância amostral)

$$\overline{x} \pm t_{\nu=14}\frac{S}{\sqrt{n}}$$
$$96,9 \pm 2,14\frac{14}{\sqrt{15}}$$
$$96,9 \pm 7,8$$
$$(89,1 \ ; \ 105)$$

(b)

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\text{sup}}}} \ ; \ \sqrt{\frac{(n-1)S^2}{\chi^2_{\text{inf}}}}\right)$$
$$\left(\sqrt{\frac{(15-1)198}{26,1}} \ ; \ \sqrt{\frac{(15-1)198}{5,63}}\right)$$
$$(10,3 \ ; \ 22,2)$$

(c) Teste de hipótese:

$$H_0 : \mu = 96 \quad H_a : \mu < 96$$
$$\alpha = 0,05$$

- Supondo $\sigma_Y^2 = \sigma^2 = 18^2$ (variância do novo algorítmo igual a do anterior)

$$z_c = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{96,9 - 96}{18/\sqrt{15}} = 0,184$$
$$\text{Região Crítica (RC)} : Z < -1,64$$
$$\text{p} - \text{valor} = P[Z < z_c] = 0,573$$

x	1	2	3	4	5	6	7	8	9	10
y	0.9	2.2	4.1	3.7	4.5	5.7	6.3	7.7	9.1	10.6

- Supondo σ_Y^2 estimado por S^2 (variância desconhecida do novo algorítmo, estimada pela variância amostral)

$$\begin{aligned} t_c &= \frac{\overline{x} - \mu_0}{S/\sqrt{n}} \\ &= \frac{96,9 - 96}{14,1/\sqrt{15}} \\ &= 0,235 \\ \text{Região Crítica (RC)} &: Z < -1,76 \\ \text{p} - \text{valor} &= P[t < t_c] = 0,591 \end{aligned}$$

Conclusão:
z_c (ou t_c) $\notin$ RC,
não rejeitamos H_0, ou seja, não há evidências que o novo algoritmo reduza o tempo de processamento.

---

45. Considere que foram coletados dados de variáveis X e Y a assume-se que cada valor de Y pode ser explicado pelo respectivo valor de X segundo a equação:

$$Y_i = \theta x_i + \epsilon_i$$

em que θ é um parâmetro a ser estimado e ϵ_i é o desvio (aleatório) de cada observação. Considere os dados:

(a) Como o parâmetro θ pode ser estimado?
(b) Obtenha a expressão do estimador.
(c) Obtenha a estimativa para o conjunto de dados obtido.

---

46. Descreva as propriedades de um estimador e seu significado, usando como exemplo o exercício anterior.

---

47. (adaptado de Magalhães & Lima) O consumo de combustível é uma variável aleatória com parâmetros dependendo do tipo de veículo. Suponha que, para um certo automóvel, o desvio padrão do consumo seja conhecido e igual a 2 km/l Porém, precisamos informações sobre o consumo médio. Para tal, coletamos uma amostra de 40 automóveis desse modelo e observamos seu consumo.

(a) Identifique no contexto: a população, a amostra, o parâmetro de interesse, o estimador e a distribuição amostral.
(b) Se a amostra forneceu um consumo médio de 9,3 km/l, construa um intervalo de confiança (95%) para a média de consumo deste modelo de carro.
(c) Se a amplitude de um intervalo de confiança, construído a partir dessa amostra, é de 1,5 unidades, qual teria sido o nível de confiança?
(d) Com a amostra já obtida de média $9,3$, teste a hipótese de que o consumo de veículos do modelo está acima de 8,7 km/l
(e) Uma empresa possui 8 veículos do modelo considerado. Baseando-se nas informações disponíveis, qual a probabilidade de que o consumo médio dos 8 veículos fique acima de 9,5 km/l?
(f) Uma nova amostra vai ser tomada para outro modelo e supondo o mesmo desvio padrão. Qual deve ser o tamanho da amostra para que o intervalo de confiança (95%) tenha amplitude de 1,2 unidades.

Solução:

(a) A população são os veículos da marca considerada e representada pela v.a.:

$$X : \text{consumo de combustível dos veículos da marca considerada}$$

que possui distribuição com $\mathrm{E}[X] = \mu$ (deconhecida) e $\mathrm{Var}[X] = 2^2$ (conhecida). A amostra é dada pelos consumos $(x_1, x_2, \ldots, x_n)$ dos veículos observados. O parâmetro de interesse é a média populacional $\mathrm{E}[X] = \mu$, o estimador é a média amostral $\overline{X}$ e a distribuição amostral é

$$\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n = 2^2/40 = 0,1)$$

(b) I.C. (95%)

$$\overline{x} \pm z_(0,975)\frac{\sigma}{\sqrt{n}}$$
$$9,3 \pm 1,96\frac{2}{\sqrt{4}}$$
$$(8,68 \ ; \ \ 9,92)$$

(c)

$$1,5 = 2 \cdot (z_(1-\alpha/2)\frac{\sigma}{\sqrt{n}})$$
$$z_(1-\alpha/2 = \frac{1,5\sqrt{40}}{2 \cdot 2}$$
$$z_(1-\alpha/2 = 2,37$$
$$1-\alpha = 0,982$$

(d)

$$H_0 : \mu \leq 8,7 \quad vs \quad H_a : \mu > 8,7$$
$$\ldots$$

(e)

$$\overline{X}_8 \approx \mathrm{N}(\mu = 9,3, \sigma^2/n = 2^2/8 = 0,5)$$
$$P[\overline{X}_8 > 9,5] = 1,88e-41$$

(f)

$$1,2 = 2 \cdot (z_(0,975)\frac{2}{\sqrt{n}})$$
$$n = \lceil \frac{2^2 \cdot 1.96^2 \cdot 2^2}{1,2^2}$$
$$n = 43$$

---

48. Quinze homens com idades entre 35 e 50 anos participaram em um estudo para avaliar o efeito de uma dieta e exercícios no nível de colesterol. O colesterol total foi medido em cada indivíduo inicialmente e depois novamente medido após 3 meses após participação em um programa de exercícios aeróbicos combinado com uma dieta de baixa caloria. Os dados estão a seguir.

(a) Calcule a média e mediana para as medidas alteração do colesterol.

(b) Calcule desvio padrão e amplitude interquartílica para alteração do coleterol.

antes	265	240	258	295	251	245	287	314	260	279	283	240	238	225	247
depois	229	231	227	240	238	241	234	256	247	239	246	218	219	226	233

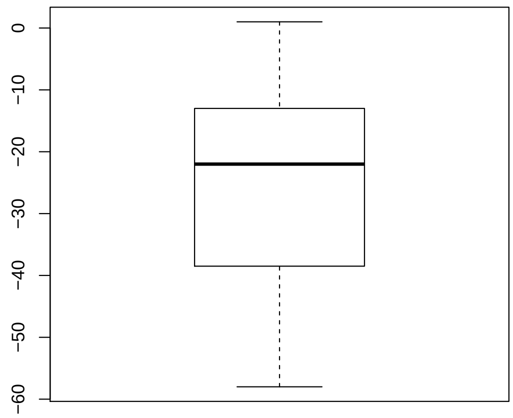
Tabela 1: Medidas de colesterol de 15 homens antes de depois de dieta combinada com exercícios.

(c) Construa um gráfico *boxplot* para as medidas de alteração do colesterol.

(d) Obtenha um intervalo de confiança (95%) para o nível médio de coleterol antes da dieta/exercícios.

(e) Use algum procedimento estatístico adequado para discutir se a dieta combinada com exercícios foi eficiente para reduzir o nível de colesterol total.

```
> antes <- c(265, 240, 258, 295, 251, 245, 287, 314, 260, 279, 283, 240, 238, 225, 247)
> depois <- c(229, 231, 227, 240, 238, 241, 234, 256, 247, 239, 246, 218, 219, 226, 233)
> (ad <- depois - antes)

 [1] -36  -9 -31 -55 -13  -4 -53 -58 -13 -40 -37 -22 -19   1 -14

> # a)
> c(media= mean(ad), mediana = median(ad))

  media mediana
 -26,87  -22,00

> # b)
> c(desvioP= sd(ad), AI = diff(fivenum(ad)[c(2,4)]))

desvioP      AI
  19,04   25,50

> # c)
> boxplot(ad)
> # d)
> mean(antes) + qt(c(0.025, 0.975), df=length(antes)-1) * sd(antes)/sqrt(length(antes))

[1] 248,0 275,6

> t.test(antes)$conf

[1] 248,0 275,6
attr(,"conf.level")
[1] 0,95

> # e)
```



---

49. Os números abaixo mostram as notas de um grupo de alunos em duas avaliações

Aluno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Prova 1	35	39	50	47	33	17	17	80	23	51	2	21	20	12	81	98	47	34
Prova 2	65	63	80	72	65	35	62	72	50	60	32	59	40	68	79	85	80	55

(a) Calcule média, variância e coeficiente de variação das notas em cada avaliação

(b) Calcule mediana, quantis, amplitude e amplitude interquartílica de cada avaliação

(c) Faça um diagrama *box-plot* para comparar as notas das duas avaliações

(d) Com as notas das duas provas juntas faça um único diagrama ramo-e-folhas sublinhando as notas da segunda prova.

(e) Usando as medidas e gráficos acima compara o rendimento dos alunos nas duas provas.

(f) Existe relação (associação) entre os resultados das duas provas? Faça um gráfico e calcule alguma(s) medida(s) estatística(s) para verificar se há associação.

(g) Suponha agora que as provas possam ser consideradas uma amostra aleatória. Faça um testes estatísticos adequado para verificar:
- Se a média da Prova 1 está significativamente abaixo de 50.
- Se houve aumento (significativo) de rendimento da primeira para segunda prova.

Solução:

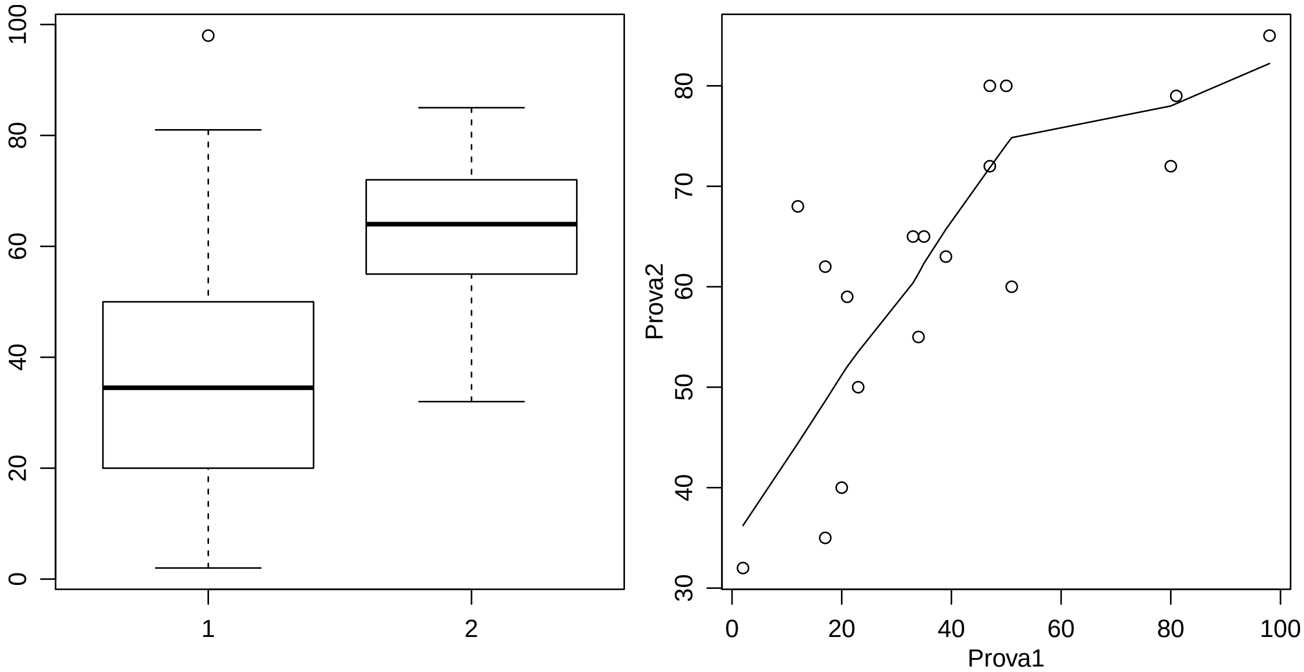
(a)

$$\bar{x}_1 = 39,28 \quad s_1^2 = 670,68 \quad CV_1 = 65,93\%$$
$$\bar{x}_2 = 62,33 \quad s_2^2 = 237,53 \quad CV_2 = 24,73\%$$

(b)

$$md_1 = 34,5 \quad Q1_1 = 20 \quad Q3_1 = 50 \quad A_1 = 96 \quad AI_1 = 30$$
$$md_2 = 64 \quad Q1_2 = 55 \quad Q3_2 = 72 \quad A_2 = 53 \quad AI_2 = 17$$

(c)



(d)
```
The decimal point is 1 digit(s) to the right of the |

0 | 2
1 | 277
2 | 013
3 | 234559
4 | 077
5 | 00159
6 | 023558
7 | 229
8 | 00015
9 | 8
```

(e) Comentários sobre: valores centrais, variabilidade, assimetria e dados discrepantes

(f) Coeficientes de correlação: Pearson $r_P = 0,75$ e Spearman $r_S = 0,732$
Comentários: ...

(g) • Teste de Hipótese para uma média

$$X_1 : \text{nota na Prova 1} \sim Dist(\mu_1, \sigma_1^2)$$
$$H_0 : \mu_1 \geq 50 \text{ vs } H_1 : \mu_1 < 50$$
$$\alpha = 0,05 \longrightarrow t_t = -1,74 \longrightarrow \hat{\mu_1} = \bar{x}_1 < 39,381$$
$$t_c = \frac{39,3 - 50}{25,9/\sqrt{18}} = -1,757$$
$$p - valor = 0,04849$$

Conclusão: Rejeita-se H_0 para $\alpha = 0,05$ e portanto há evidências que a média esteja abaixo de 50.

- Teste de Hipótese para comparação de médias pareadas

$$
\begin{aligned}
X_1 &: \text{nota na Prova 1 } \sim Dist(\mu_1, \sigma_1^2)\\
X_2 &: \text{nota na Prova 2 } \sim Dist(\mu_2, \sigma_2^2)\\
D = X_2 - X_1 &: \text{diferença de notas } \sim Dist(\mu_D = \mu_2 - \mu_1, \sigma_D^2)\\
H_0 &: \mu_2 \leq \mu_1 \longrightarrow \mu_D \leq 0 \text{ vs } H_1 : \mu_2 > \mu_1 \longrightarrow \mu_D > 0\\
\alpha &= 0,05 \longrightarrow t_t = 1,74 \longrightarrow \hat{\mu_D} = \bar{x}_D < 7,214\\
\text{Amostra} &:\\
\hat{\mu_D} &= \bar{x}_D = x_2 \bar{-} x_1 = 23,056 \quad \hat{s}_D = 17,595\\
t_c &= \frac{23,1 - 0}{17,6/\sqrt{18}} = 5,559\\
p - valor &= 0,99998
\end{aligned}
$$

Conclusão: Rejeita-se H_0 para $\alpha = 0,05$ e portanto há evidências que houve aumento de rendimento.

---

50. Um determinado elemento é medido regularmente em amostras água coletadas um reservatório. Assume-se que os valores individuais das coletas possuem distribuição normal de média 500 e desvio padrão de 20 unidades. Em um programa de monitoramento são feitas periodicamente como rotina a amostragem com análise de cinco coletas. Se algum anomalia é detectada é feita uma inspeção detalhada. Adota-se considerar uma *anomalia* se o teor médio de uma amostra ultrapassa valor crítico de 512 unidades.

    (a) Qual a proporção de coletas que devem apresentar teores individuais entre 490 e 515 unidades?

    (b) Qual a proporção de médias amostrais (das cinco coletas) que devem apresentar teores individuais entre 490 e 515 unidades?

    (c) Entre quais valores de teores das coletas individuais ao redor da média de 500 unidades espera-se obter 90% das medidas?

    (d) Entre quais valores ao redor da média de 500 unidades de teores médios das amostras de cinco coletas espera-se obter 90% das médias amostrais?

    (e) Verificou-se em uma amostra o valor médio de 510 unidades. Voce considera que há uma anomalia e recomendaria uma inspeção detalhada? Justifique.

    (f) Qual a probabilidade de encontrar um teor médio na amostra acima do valor crítico de 512 unidades, quando os valores do lago estão nos níveis usuais?

    (g) Qual deveria ser o *valor crítico* para que a probabilidade de encontrar teores médios acima dele não ultrapasse 0,02?

    (h) Se o *valor crítico* é mantido em 512 unidades qual deveria ser o número de coletas por amostra para que a probabilidade de encontrar teores médios acima de 512 unidades não ultrapasse 0,02?

    (i) Suponha que tenha havido uma alteração no reservatório elevando a concentração média de 500 para 515 unidades. Qual seria a probabilidade de obter uma média amostras abaixo de 500 unidades?

    (j) Ainda no contexto do item anterior, qual a probabilidade de não se recomendar uma inspeção detalhada?

Solução:

X : tempo de processamentos (segundos) $\quad X \sim N(\mu_X = 500, \sigma_X^2 = 20^2)$

$\overline{X}$: tempo de procesamentos (segundos) $\quad \overline{X} \sim N(\mu_{\overline{X}} = \mu_X, \sigma_{\overline{X}}^2 = \sigma_X^2) \sim N(\mu_{\overline{X}} = 500, \sigma_{\overline{X}}^2 = 20^2/5)$

(a) $P[490 < X < 515] = P[\frac{490-500}{20} < Z < \frac{515-500}{20}] = P[-0,5 < Z < 0,75] = P[0 < Z < 0,5] + P[0 < Z < 0,75] = 0,4648$

(b) $P[490 < \overline{X} < 515] = P[\frac{490-500}{20/\sqrt{5}} < Z < \frac{515-500}{20/\sqrt{5}}] = P[-1,118 < Z < 1,677] = P[0 < Z < 1,118] + P[0 < Z < 1,677] = 0,8215$

(c)

$$\begin{aligned}
P[500 - \Delta_X < X < 500 + \Delta_X] &= 0,90 \\
z = \frac{(500 + \Delta_X) - 500}{20} &= 1,64 \\
\Delta_X &= 32,9 \\
P[467,1 < X < 532,9] &= 0,90
\end{aligned}$$

(d)

$$\begin{aligned}
P[500 - \Delta_{\overline{X}} < \overline{X} < 500 + \Delta_{\overline{X}}] &= 0,90 \\
z = \frac{(500 + \Delta_{\overline{X}}) - 500}{20/\sqrt{5}} &= 1,64 \\
\Delta_{\overline{X}} &= 14,71 \\
P[485,3 < \overline{X} < 514,7] &= 0,90
\end{aligned}$$

(e) $P[\overline{X} > 510] = P[Z > \frac{510-500}{20/\sqrt{5}}] = P[Z > 1,118] = 0,5 - P[0 < Z < 1,118] = 0,5 - 0,368 = 0,132$

(f) $P[\overline{X} > 512] = P[Z > \frac{512-500}{20}] = P[Z > 1,342] = 0,5 - P[0 < Z < 1,342] = 0,5 - 0,41 = 0,0899$

(g)

$$\begin{aligned}
P[\overline{X} < \overline{x}_c] &= 0,02 \\
z = \frac{\overline{x}_c - 500}{20/\sqrt{5}} &= 2,05 \\
\overline{x}_c &= 518,4
\end{aligned}$$

(h)

$$\begin{aligned}
P[\overline{X}_n < 512] &= 0,02 \\
z = \frac{512 - 500}{20/\sqrt{n}} &= 2,05 \\
n = \left\lceil \frac{2,05^2 20^2}{(512 - 500)^2} \right. &= 12
\end{aligned}$$

(i) $P[\overline{X} < 500|\mu = 515] = P[Z < \frac{500-515}{20/\sqrt{5}}] = P[Z < -1,677] = 0,5 - P[0 < Z < 1,677] = 0,5 - 0,4532 = 0,04677$

(j) $P[\overline{X} < 512|\mu = 515] = P[Z < \frac{512-515}{20/\sqrt{5}}] = P[Z < -0,3354] = 0,5 - P[0 < Z < 0,3354] = 0,5 - 0,1313 = 0,3687$

Solução Computacional:

```
> (it.a <- diff(pnorm(c(490,515), m=500, sd=20)))

[1] 0,4648

> (it.b <- diff(pnorm(c(490,515), m=500, sd=20/sqrt(5))))

[1] 0,8215

> (it.c <- qnorm(c(0.05,0.95), m=500, sd=20))

[1] 467,1 532,9
```

```
> (it.d <- qnorm(c(0.05,0.95), m=500, sd=20/sqrt(5)))

[1] 485,3 514,7

> (it.e <- pnorm(510, m=500, sd=20/sqrt(5), low=F))

[1] 0,1318

> (it.f <- pnorm(512, m=500, sd=20/sqrt(5), low=F))

[1] 0,08986

> (it.g <- qnorm(0.98, m=500, sd=20/sqrt(5)))

[1] 518,4

> (it.h <- ceiling((qnorm(0.98)^2)*20^2/(512-500)^2))

[1] 12

> (it.i <- pnorm(500, m=515, sd=20/sqrt(5)))

[1] 0,04677

> (it.j <- pnorm(512, m=515, sd=20/sqrt(5)))

[1] 0,3687
```

---

51. Foi feita uma pesquisa junto a proprietários rurais para verificar o conhecimento sobre uma determinada legislação ambiental. Para isto foi selecionada uma amostra aleatória de 800 proprietários e vamos considerar aqui simplesmente que cada um deles era classificado como *0 - sem conhecimento* ou *1 - com conhecimento*. Feita a pesquisa, 504 foram classificados como *com conhecimento*.

    (a) Obtenha a estimativa pontual e intervalar (95% de confiança) para a proporção de proprietários com conhecimento da legislação.

    (b) Antes da pesquisa tabalhava-se com a hipótese de que o nível de conhecimento era da 60% dos proprietários. Há evidências estatísticas baseadas no estudo de que o o conhecimento seja diferente deste? Justifique sua resposta.

    (c) Após a pesquisa e uma campanha publicitária decidiu-se refazê-la porém deseja-se uma margem de erro de no máximo 1,5%. Qual deveria ser o tamanho da nova amostra?

Solução:

$$
\begin{aligned}
&X : \text{conhecimento sobre a legislação} \\
&X \sim \mathrm{B}(p) \;\; ; \;\; \mathrm{E}[X] = p \qquad \mathrm{Var}[X] = p(1-p) \\
&\hat{p} \sim \mathrm{N}(p, p(1-p)/n)
\end{aligned}
$$

(a)

$$\hat{p} = \frac{504}{800} = 0,63$$

I.C.assinttico :

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow (0,597 \;\;;\;\; 0,663)$$

I.C.conservador :

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{1}{4n}} \longrightarrow (0,595 \;\;;\;\; 0,665)$$

(b) Resposta e justificativa baseada no valor estar ou não contido no I.C..

(c)

$$\begin{aligned}
\text{ME} &= z_{95\%}\sqrt{\frac{p(1-p)}{n}} \\
0.015 &= 1,96\sqrt{\frac{0,63(1-0,63)}{n}} \\
n &= \lceil \frac{1,96^2}{0,015^2} 0,63(1-0,63) \\
n &= 3980
\end{aligned}$$

---

52. O acompanhamento de uma área de reflorestamento foi feito tomando uma amostra de 15 parcelas escolhidas ao acaso dentro da área. Em cada parcela são medidos diversos atributos. Abaixo estão as medidas de um deles, a soma de áreas basais das árvores da parcela.

`23,5  27,6  26,7  25,4  23,4  30,0  29,1  27,1  23,4  22,3  28,6  25,3  26,5  24,8  25,4`

(a) Obtenha estimativas pontuais da média e desvio padrão da área basal por parcela.

(b) Obtenha o intervalo de confiança (90%) do desvio padrão das áreas basais.

(c) Obtenha o intervalo de confiança (90%) da média da áreas basais.

Solução:

$$\begin{aligned}
X &: \text{área basal por parcela} \qquad X \sim N(\mu, \sigma^2) \\
\overline{X} &: \text{área basal média} \qquad \overline{X} \sim t_{n-1}(\mu, S^2/n) \\
\hat{\sigma}^2 = S^2 &: \text{variância da área basal por parcela} \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(\nu = n-1) \\
\hat{\sigma} &= S = \sqrt{S^2}
\end{aligned}$$

(a)

$$\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{n} x_i}{n} = 25,9 \\
\hat{\sigma}^2 = S^2 &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = 5,21 \\
\hat{\sigma} = S &= \sqrt{S^2} = 2,28
\end{aligned}$$

(b) I.C. (90%) para variância

$$\begin{gathered}
\left(\frac{(n-1)S^2}{\chi^2_{sup}} \quad ; \quad \frac{(n-1)S^2}{\chi^2_{inf}}\right) \\
\left(\frac{(15-1)5,21}{23,7} \quad ; \quad \frac{(15-1)5,21}{6,57}\right) \\
(3,079 \quad ; \quad 11,1)
\end{gathered}$$

I.C. (90%) desvio padrão:

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{sup}}} \quad ; \quad \sqrt{\frac{(n-1)S^2}{\chi^2_{inf}}}\right) = (1,755 \quad ; \quad 3,332)$$

(c) I.C. (90%) para média

$$IC_{90\%} : \overline{x} \pm t_{n-1,90\%} \frac{S}{\sqrt{n}}$$
$$IC_{90\%} : 25,9 \pm 1,76 \frac{2,28}{\sqrt{15}}$$
$$IC_{90\%} : (25,7 \;\; ; \;\; 26,2)$$

---

53. Assume-se que em uma rede com um grande número de nós 20% deles podem estar inacessíveis a qualquer tempo. São feitas inspeções periódicas em 400 nós da rede escolhidos ao acaso e se 90 (22,5%) ou mais desses estão inacessíveis, é feito um rastreamento detalhado para verificação e detecção de problemas.

- Qual a probabilidade de encontrar mais que 85 nós inacessíveis em uma inspeção?
- Se a rede está normal, qual a probabilidade de encontrar 100 ou mais nós inacessíveis em uma inspeção?
- Qual os valores de proporções ao redor do valor médio (20%) de nós inacessíveis, dentre os quais devem-se estar 80% das inspeções.
- Qual a probabilidade de fazer um rastreamento desnecessário?
- Quantos nós deveriam ser inspecionados para que a chance de fazer um rastreamento desnecessário não ultrapasse 1%?

Solução:

$$X : \text{estado do nó (0 - acessível, 1 : inacessív}$$
$$X \sim B(\mu = E[X] = p = 0,20, \sigma^2 = \text{Var}[X] = p(1-p) = 0,2 \cdot 0,8 = 0,16)$$
$$\hat{p} = \overline{X} : \text{proporção de acessíveis entre n (400) nós}$$
$$\hat{p} = \overline{X} \sim N(\mu_{\hat{p}} = p = 0,20, \sigma^2_{\hat{p}} = \frac{p(1-p)}{n} = \frac{0,2(1-0,2)}{400} = 0,02)$$

(a) $P[\hat{p} > 85/400] = P[Z > \frac{85/400-0,20}{0,02}] = P[Z > 0,625] = 0,5 - P[0 < Z < 0,625] = 0,5 - 0,234 = 0,266$

(b) $P[\hat{p} > 100/400] = P[Z > \frac{100/400-0,20}{0,02}] = P[Z > 2,5] = 0,5 - P[0 < Z < 2,5] = 0,5 - 0,4938 = 0,00621$

(c)

$$P[p - \Delta_{\hat{p}} < X < p + \Delta_{\hat{p}}] = 0,80$$
$$z = \frac{p + \Delta_{\hat{p}} - p}{0,02} = 1,28$$
$$\Delta_{\hat{p}} = 0,0256$$
$$P[0,174 < X < 0,226] = 0,80$$

(d) $P[\hat{p} > 0,225] = P[Z > \frac{0,225-0,20}{0,02}] = P[Z > 1,25] = 0,5 - P[0 < Z < 1,25] = 0,5 - 0,3944 = 0,1056$

(e)

$$z = 0,99 = \frac{0,225 - 0,20}{0,16/n}$$
$$n = \frac{0,99^2}{(0,225 - 0,20)^2} 0,16 = 1386$$

Solução Computacional com o programa R:

```
> sP <- sqrt(.2*.8/400)
> (p2.a <- pnorm(85/400, m=0.20, sd=sP, low=F))

[1] 0,266

> (p2.b <- pnorm(100/400, m=0.20, sd=sP, low=F))

[1] 0,00621

> (q2.c <- qnorm(c(0.10, 0.90), m=0.20, sd=sP))

[1] 0,1744 0,2256

> (p2.d <- pnorm(0.225, m=0.20, sd=sP, low=F))

[1] 0,1056

> (n.e <- ceiling(qnorm(0.99)^2 * (0.2*.8)/0.025^2))

[1] 1386
```

---

54. Assume-se que a geração diária de energia (em escala logarítmica) por uma turbina eólica possui distribuição normal de média 2,5 e desvio padrão 0,60 unidades.
    (a) Qual a probabilidade da geração ser inferior a 2 unidades em um particular dia?
    (b) Qual a probabilidade da geração diária ficar entre 2 e 3 unidades?
    (c) Define-se como o *o valor mínimo de referência* o valor gerado em pelo menos 80% dos dias? Qual é este valor?
    (d) Qual a probabilidade de gerar mais que 18 unidades em uma semana (7 dias)?
    (e) Qual a probabilidade da geração diária média em um mês (30 dias) ultrapassar 2,6 unidades?

Solução:

$$\begin{array}{rlr} X & : \text{geração diária (log)} & X \sim \mathrm{N}(2,5 \;\; ; \;\; 0,6^2) \\ \overline{X}_n & : \text{média da geração diária em } n \text{ dias} & \overline{X}_n \sim \mathrm{N}(2,5 \;\; ; \;\; 0,6^2/n) \end{array}$$

(a) $P[X < 2] = P[Z < \frac{2-2,5}{0,6}] = P[Z < -0,833] = 0,202$

(b) $P[2 < X < 3] = P[\frac{2-2,5}{0,6} < Z < \frac{3-2,5}{0,6}] = P[-0,833 < Z < 0,833] = 0,595$

(c)

$$\begin{aligned} P[X > x] = P[Z > z] &= 0,80 \\ z = \frac{x - 2,5}{0,6} &= -0,8416 \\ x = 2,5 + 0,6 \cdot (-0,8416) &= 2 \end{aligned}$$

(d) $P[\sum i = 1_7 X_i > 18] = P[\overline{X}_7 > 18/7] = P[Z < \frac{(18/7)-2,5}{0,6}] = P[Z < 0,119] = 0,376$

(e) $P[\overline{X}_{30} > 2,6] = P[Z < \frac{2,5-2,5}{0,6/\sqrt{30}}] = P[Z < 0,913] = 0,181$

Solução Computacional com o programa R:

```
> (it1 <- pnorm(2, m=2.5, sd=0.6))

[1] 0,2023

> (it2 <- diff(pnorm(c(2,3), m=2.5, sd=0.6)))

[1] 0,5953

> (it3 <- qnorm(0.2, m=2.5, sd=0.6))

[1] 1,995

> (it4 <- pnorm(18/7, m=2.5, sd=0.6/sqrt(7), low=F))

[1] 0,3764

> (it5 <- pnorm(2.6, m=2.5, sd=0.6/sqrt(30), low=F))

[1] 0,1807
```

---

55. Foi feita uma pesquisa com 1.500 pessoas que acessam um *site* para se estimar a proporção de mulheres que acessam o *site*. Encontrou-se que 1150 acessos eram de mulheres.

(a) Obtenha a margem de erro para pesquisa para confiança de 95%

(b) Qual seria esta margem e erro para uma amostra de 500 pessoas?

(c) Qual deveria ser o tamanho da amostra para que a margem de erro não ultrapasse 1%, com 95% de confiança.

(d) Teste, com base na pesquisa feita, a hipótese de que as mulheres representam mais que 75% dos acessos.

Solução:

$$X \text{sexo de quem acessa o site (0: homem / 1: mulher)}$$
$$X \sim Ber(p)$$
$$\hat{p} \sim \mathrm{N}(p, \frac{p(1-p}{n})$$
$$\hat{p} \sim \mathrm{N}(p, \frac{\hat{p}(1-\hat{p})}{n})(assinttica))$$
$$\hat{p} \sim \mathrm{N}(p, \frac{0,5(1-0,5}{n} = \frac{1}{4n})(conservadora)$$

(a)

$$\text{M.E.} = z\sqrt{\frac{p(1-p)}{n}}$$
$$\text{M.E. (assintótica)} = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96\sqrt{\frac{0,77(1-0,77)}{1500}} = 0,021$$
$$\text{M.E. (conservadora)} = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96\sqrt{\frac{1}{4 \cdot 1500}} = 0,025$$

(b)

$$\text{M.E.} = z\sqrt{\frac{p(1-p)}{n}}$$
$$\text{M.E. (assintótica)} = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96\sqrt{\frac{0,77(1-0,77)}{500}} = 0,037$$
$$\text{M.E. (conservadora)} = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96\sqrt{\frac{1}{4\cdot 500}} = 0,044$$

(c)

$$\text{M.E.} = z\sqrt{\frac{p(1-p)}{n}}$$
$$n = \frac{z^2}{\text{M.E.}^2}p(1-p)$$
$$\text{n (assintótico)} = \frac{1,96^2}{0,01^2}0,77(1-0,77) = 6872$$
$$\text{n (conservador)} = \frac{1,96^2}{0,01^2}0,5(1-0,5) = 9604$$

(d)

$$H_0 : p = 0,75 \quad vs \quad H_a : p > 0,75$$
$$\alpha = 0,95$$
$$z_c = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,77-0,75}{\sqrt{\frac{0,75(1-0,75)}{1500}}} = 1,526$$
$$z_{0,95} = 1,645$$
$$\text{p} - \text{valor} = 0,0635$$
$$\text{Conclusão : Não Rejeita H0}$$

Solução Computacional com o programa R:

```
> n <- 1500
> p.est <- 1150/n
> p0 <- 0.75
> # a)
> (MEa <- qnorm(0.975) * sqrt(p.est*(1-p.est)/n))

[1] 0,0214

> (MEc <- qnorm(0.975) * sqrt(1/(4*n)))

[1] 0,0253

> # b)
> (MEa0 <- qnorm(0.975) * sqrt(p.est*(1-p.est)/500))

[1] 0,03707

> (MEc0 <- qnorm(0.975) * sqrt(1/(4*500)))

[1] 0,04383
```

```
> # c)
> (n.a <- ceiling((qnorm(0.975)/0.01)^2 * p.est*(1-p.est)))

[1] 6872

> (n.c <- ceiling((qnorm(0.975)/0.01)^2 * (1/2)*(1/2)))

[1] 9604

> # d) H0: p > 0.75
> (zc <- (p.est - p0)/sqrt(p.est*(1-p.est)/n))

[1] 1,526

> (z.crit <- qnorm(0.95))

[1] 1,645

> (pvalor <- pnorm(p.est,m=0.75, sd=sqrt(p.est*(1-p.est)/n), low=F))

[1] 0,06348

> pnorm(zc, low=F)

[1] 0,06348
```

---

56. Foi tomada uma amostra para estimar a biomassa em uma área de regeneração a partir de uma amostra de um conjunto de parcelas, obtendo os valores a seguir. Faça as suposições necessárias e obtenha intervalos e confiança (90%) para a média e para a variância da biomassa. Teste a hipótese de que a biomassa está abaixo do valor esperado de 157 unidades.

```
156,4   155,7   157,1   155,9   156,9  160,1   154,9   156,1   157,4   158,5   155,4   157,1
```

Solução:

(a)

$$\begin{aligned}
&\text{I.C.}_{90\%}(\mu): \\
&\qquad \overline{x} \pm t_{(n-1),90\%}\frac{S}{\sqrt{n}} \\
&\qquad 156,8 \pm 0,819\frac{1,43}{\sqrt{12}} \\
&\qquad 156,8 \pm 0,743 \\
&\qquad (156 \;\; ; \;\; 157,5) \\
&\text{I.C.}_{90\%}(\sigma^2): \\
&\qquad \left(\frac{(n-1)S^2}{\chi^2_{sup}} \;\; ; \;\; \frac{(n-1)S^2}{\chi^2_{inf}}\right) \\
&\qquad \left(\frac{(12-1)2,05}{19,7} \;\; ; \;\; \frac{(12-1)2,05}{4,57}\right) \\
&\qquad (1,147 \;\; ; \;\; 4,933)
\end{aligned}$$

(b)

$$
\begin{aligned}
H_0 : \mu = 157 \quad vs \quad H_a : \mu < 157 \\
\alpha = 0,10 \\
t_c = \frac{\overline{x} - \mu_0}{S/\sqrt{n}} = \frac{156,8 - 157}{1,43/12} = -0,5038 \\
t_{0,90} = -1,363 \\
\text{p} - \text{valor} = 0,312 \\
\text{Conclusão : Não Rejeita H0}
\end{aligned}
$$

---

57. Em um site de compras, O valor de vendas diárias possui distribuição normal de média R$ 2.500,00 e desvio padrão de R$ 500,00.

(a) Qual a probabilidade de vender menos que R$ 2.000,00 em um particular dia?

(b) Qual a probabilidade da venda diária ficar entre R$ 2.000,00 e R$ 3.000,00 ?

(c) Qual o valor que tem 85% das vendas diárias acima dele?

(d) Qual a probabilidade de vender acima de R$18.000,00 em uma semana (7 dias)?

(e) Qual a probabilidade da venda diária média em um mês (30 dias) ficar acima de R$2.600,00?

Solução:

$$
\begin{array}{ll}
X : \text{geração diária (log)} & X \sim \text{N}(2.500 \ ; \ 500^2) \\
\overline{X}_n : \text{média da geração diária em } n \text{ dias} & \overline{X}_n \sim \text{N}(2.500 \ ; \ 500^2/n)
\end{array}
$$

(a) $P[X < 2.000] = P[Z < \frac{2.000-2.500}{500}] = P[Z < -1] = 0,159$

(b) $P[2.000 < X < 3.000] = P[\frac{2.000-2.500}{500} < Z < \frac{3.000-2.500}{500}] = P[-1 < Z < 1] = 0,683$

(c)

$$
\begin{aligned}
P[X > x] = P[Z > z] = 0,85 \\
z = \frac{x - 2.500}{500} = -1,036 \\
x = 2.500 + 500 \cdot (-1,036) = 1982
\end{aligned}
$$

(d) $P[\sum_{i=1}^{7} X_i > 18.000] = P[\overline{X}_7 > 18.000/7] = P[Z < \frac{(18.000/7)-2.500}{500/\sqrt{7}}] = P[Z < 1] = 0,353$

(e) $P[\overline{X}_{30} > 2.600] = P[Z < \frac{2.600-2.500}{500/\sqrt{30}}] = P[Z < 1,1] = 0,137$

---

58. Foi tomada uma amostra para estimar o rendimento (km/l) de ônibus de uma frota obtendo os valores a seguir.

```
6,4   5,7   7,1   5,9   6,9  10,1   4,9   6,1   7,4   8,5   5,4   7,1
```

(a) Calcule a média, desvio padrão e coeficiente de variação dos dados.

(b) Calcule a mediana, quartis e amplitude interquartílica.

(c) Faça as suposições necessárias e obtenha intervalos e confiança (90%) para o consumo médio e para a variância dos consumos.

(d) Teste a hipótese de que o consumo está abaixo de 6,8 km/l.

Solução:

(a)

$$\overline{x} = 6,8 \quad S = 1,4 \quad CV = 21$$

(b)

$$\text{md}(x) = 6,7 \quad q_1 = 5,8 \quad q_3 = 7,2 \quad AI = 1,4$$

(c)

$$\begin{aligned}
\text{I.C.}_{90\%}(\mu) :& \\
& \overline{x} \pm t_{(n-1),90\%}\frac{S}{\sqrt{n}} \\
& 6,79 \pm 1,8\frac{1,43}{\sqrt{12}} \\
& 6,79 \pm 0,743 \\
& (6,05 \;\; ; \;\; 7,53) \\
\text{I.C.}_{90\%}(\sigma^2) :& \\
& \left(\frac{(n-1)S^2}{\chi^2_{sup}} \;\; ; \;\; \frac{(n-1)S^2}{\chi^2_{inf}}\right) \\
& \left(\frac{(12-1)2,05}{19,7} \;\; ; \;\; \frac{(12-1)2,05}{4,57}\right) \\
& (1,15 \;\; ; \;\; 4,93)
\end{aligned}$$

(d)

$$\begin{aligned}
H_0 : \mu = 6,8 \quad vs \quad H_a : \mu < 6,8& \\
\alpha &= 0,10 \\
t_c &= \frac{\overline{x} - \mu_0}{S/\sqrt{n}} = \frac{6,79 - 6,8}{1,43/12} = -0,02015 \\
t_{0,90} &= -1,363 \\
\text{p} - \text{valor} &= 0,492 \\
\text{Conclusão} &: \text{Não Rejeita H0}
\end{aligned}$$

Soluções computacionais com o programa *R*

```
> x <- c(6.4, 5.7, 7.1, 5.9, 6.9, 10.1, 4.9, 6.1, 7.4, 8.5, 5.4, 7.1)
> n <- length(x)
> x.m <- mean(x)
> x.sd <- sd(x)
> x.var <- var(x)
> qs <- fivenum(x)
> ## a)
> c(media = x.m, desvio = x.sd, CV = 100*x.sd/x.m)

 media desvio     CV
 6,792  1,432 21,090

> ## b)
> c(mediana = median(x), q1 = qs[2], q3 = qs[4], AI=diff(qs[c(2,4)]))

mediana      q1      q3      AI
   6,65    5,80    7,25    1,45
```

```
> ## c)
> x.m + qt(c(0.05, 0.95), df=n-1)* x.sd/sqrt(n)

[1] 6,049 7,534

> (n-1)*x.var/qchisq(c(0.95, 0.05), df=n-1)

[1] 1,147 4,933

> ## d)
> tc <- (x.m-6.8)/(x.sd/sqrt(n))
> t.crit <- qt(0.10, df=n-1)
> pvalor <- pt(tc, df=n-1)
```

---

59. Quinze homens com idades entre 35 e 50 anos participaram em um estudo para avaliar o efeito de uma dieta e exercícios no nível de colesterol. O colesterol total foi medido em cada indivíduo inicialmente e depois novamente medido após 3 meses após participação em um programa de exercícios aeróbicos combinado com uma dieta de baixa caloria. Os dados estão a seguir.

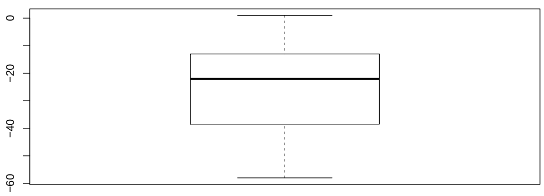
antes	265	240	258	295	251	245	287	314	260	279	283	240	238	225	247
depois	229	231	227	240	238	241	234	256	247	239	246	218	219	226	233

Tabela 2: Medidas de colesterol de 15 homens antes de depois de dieta combinada com exercícios.

(a) Calcule a média e mediana para as medidas alteração do colesterol.
(b) Calcule desvio padrão e amplitude interquartílica para alteração do coleterol.
(c) Construa um gráfico *boxplot* para as medidas de alteração do colesterol.
(d) Obtenha um intervalo de confiança (95%) para o nível médio de coleterol antes da dieta/exercícios.
(e) Use algum procedimento estatístico adequado para discutir se a dieta combinada com exercícios foi eficiente para reduzir o nível de colesterol total.

```
> antes <- c(265, 240, 258, 295, 251, 245, 287, 314, 260, 279, 283, 240, 238, 225, 247)
> depois <- c(229, 231, 227, 240, 238, 241, 234, 256, 247, 239, 246, 218, 219, 226, 233)
> (ad <- depois - antes)

 [1] -36  -9 -31 -55 -13  -4 -53 -58 -13 -40 -37 -22 -19   1 -14

> # a)
> c(media= mean(ad), mediana = median(ad))

  media mediana
 -26,87  -22,00

> # b)
> c(desvioP= sd(ad), AI = diff(fivenum(ad)[c(2,4)]))

desvioP      AI
  19,04   25,50

> # c)
> boxplot(ad)
> # d)
> mean(antes) + qt(c(0.025, 0.975), df=length(antes)-1) * sd(antes)/sqrt(length(antes))
```

```
[1] 248,0 275,6

> t.test(antes)$conf

[1] 248,0 275,6
attr(,"conf.level")
[1] 0,95

> # e)
```



60. Em levantamentos sobre a vegetação em uma determinada área foram feitas medidas em um conjunto de parcelas de 2×2m, e assume-se que as medidas são independentes entre os pontos de coleta. Em cada parcela anota-se as medidas de diversas variáveis e dentre elas as medidas consideradas aqui das variáveis *biomassa* e um *índice de fertilidade do solo*. Com base neste contexto, responda as questões a seguir.

    (a) Supondo que os valores de biomassa possuem distribuição normal de média 20 e desvio padrão de 3 unidades encontre:

        i. a probabilidade de encontrar um valor acima de 25 em uma parcela
        ii. a probabilidade de encontrar um valor entre 19 e 21 unidades em uma parcela
        iii. o valor abaixo do qual se espera encontrar apenas 10% das parcelas
        iv. a probabilidade de encontrar um valor médio de 5 parcelas entre 19 e 21 unidades

    (b) O *valor referência* das parcelas é definido como sendo de 100 unidades para parcelas com biomassa abaixo de 18, 120 unidades para biomassa entre 18 e 23, e 150 unidades para parcelas com biomassa acima de 23. Qual o valor de referência esperado para uma amostra de 20 parcelas?

    Em um dado levantamento foram obtidos os dados a seguir

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
biomassa	20,2	17,6	22,0	15,9	15,3	27,9	17,8	19,1	14,2	24,4	19,7	24,1	21,7	23,1	17,4	20,3	27,5	23,9	26,0	23,6
fertilidade	6,3	5,0	7,0	4,2	4,3	9,3	5,3	5,6	2,8	7,6	5,7	8,5	7,0	7,2	4,8	6,4	9,5	8,7	8,3	8,2

    (c) Obtenha a média, mediana e quartis para cada uma das variáveis.

    (d) Obtenha a amplitude, amplitude interquartílica e coeficiente de variação para cada uma das variáveis.

    (e) Qual variável apresenta maior variabilidade? Justifique.

    (f) Obtenha um gráfico *box-plot* para cada uma das variáveis

    (g) Investigue e relate baseando-se em um gráfico e alguma medida estatística adequada se a biomassa está relacionada com a fertilidade.

    (h) Obtenha intervalos de confiança (95%) para a variância e média da fertilidade

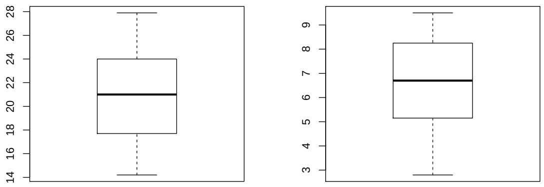
    (i) Verifique por um teste estatístico adequado e utilizando os dados obtidos a hipótese de que a biomassa média da área difere de 20 unidades.

    (j) No contexto do item anterior identifique: a população, a amostra, o parâmetro de interesse, o estimador, a estimativa, a distribuição amostral, a região crítica e o nível descritivo (valor-p).

Solução:

```
> ## 1a) P[X > 25]
> pnorm(25, m=20, s=3, lower=F)

[1] 0,04779

> ## 1b) P[19 < X < 21]
> diff(pnorm(c(19,21), m=20, s=3))

[1] 0,2611

> ## 1c) P[X < x] = 0,10
> qnorm(0.10, m=20, s=3)

[1] 16,16

> ## 1d) P[19 < \bar{X} < 21]
> diff(pnorm(c(19,21), m=20, s=3/sqrt(5)))

[1] 0,5439

> ##
> ## 2)
> (probs <- diff(pnorm(c(-Inf,18,23, Inf), m=20, s=3)))

[1] 0,2525 0,5889 0,1587

> VR <- c(100,120,150)
> sum(probs*VR)

[1] 119,7

> 20 * sum(probs*VR)

[1] 2394

> ##
> ## 3)
> t(summary(dat))

   biomassa  Min.   :14,2   1st Qu.:17,8   Median :21,0   Mean   :21,1
 fertilidade Min.   :2,80   1st Qu.:5,22   Median :6,70   Mean   :6,58

   biomassa  3rd Qu.:23,9   Max.   :27,9
 fertilidade 3rd Qu.:8,22   Max.   :9,50

> ##
> ## 4)
> t(apply(dat,2, function(x) c(A=diff(range(x)), AI=diff(fivenum(x)[c(2,4)]), CV=100*sd(x)/mean(x

               A  AI    CV
biomassa    13,7 6,3 18,94
fertilidade  6,7 3,1 28,25
```

```
> ##
> ## 5)
> ## Fertilidade: possui maior CV
> ##
> ## 6)
> par(mfrow=c(1,2))
> boxplot(dat[,1])
> boxplot(dat[,2])
> #boxplot(scale(dat))
> ##
> ## 7)
> par(mfrow=c(1,1))
> plot(dat)
> c(
+ rP=cor(dat[,"biomassa"],dat[,"fertilidade"], met="p"),
+ rS=cor(dat[,"biomassa"],dat[,"fertilidade"], met="s"),
+ rK=cor(dat[,"biomassa"],dat[,"fertilidade"], met="k")
+ )

    rP     rS     rK
0,9792 0,9786 0,9129

> ##
> ## 8)
> (length(dat[,2])-1)*var(dat[,2])/qchisq(c(0.975,0.025), df=length(dat[,2])-1)

[1] 2,002 7,384

> mean(dat[,2]) + qt(c(0.025, 0.975), df=length(dat[,2]-1)) * sd(dat[,2])/sqrt(length(dat[,2]))

[1] 5,717 7,453

> ##
> ## 9)
> t.test(dat[,1], mu=20)

One Sample t-test

data:  dat[, 1]
t = 1,2, df = 19, p-value = 0,2
alternative hypothesis: true mean is not equal to 20
95 percent confidence interval:
 19,22 22,95
sample estimates:
mean of x
    21,09

> ##
> ##10)
```



---

61. Considere uma pesquisa eleitoral na qual deseja-se estimar a intenção de voto de um candidato através de uma amostra aleatória simples.

    (a) Qual deveria ser o tamanho da amostra para se obter uma margem de erro de 1,5% com 95% de confiança?

    (b) E para uma margem de erro de 3% com 90% de confiança?

$$
\begin{aligned}
X &\sim B(p) \\
\text{Teo2}: \ \hat{p} = \overline{X} &\approx N(p, \frac{p(1-p)}{n}) \\
\hat{p} &\pm z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} \\
\hat{p} &\pm \text{ME} \\
\text{ME} &= z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} \\
n &= \frac{(z_{1-\alpha})^2}{\text{ME}^2}p(1-p) \\
p(1-p) &\text{ é limitado superiormente para } p = 0,5 \\
n &= \frac{(z_{1-\alpha})^2}{\text{ME}^2}0,25
\end{aligned}
$$

(a) $n = \frac{(1{,}96)^2}{(0{,}015)^2}0,25 = 4269$

(b) $n = \frac{(1{,}645)^2}{(0{,}03)^2}0,25 = 752$

---

62. Um veículo transporta até 12 passageiros e uma carga máxima de 1200 kg, incluindo pesos dos passageiros e bagagens. O peso dos passageiros possui distribuição normal de média 76 kg e desvio padrão de 15 kg. O peso das bagagens dos passageiros possui distribuição normal de média 22 kg e desvio padrão de 5 kg.

    (a) Qual a probabilidade de um passageiro (com sua bagagem) ultrapassar 100 kg?

    (b) Qual a probabilidade do peso total de 12 passageiros (com sua bagagem) ultrapassar a carga máxima?

    (c) Qual a probabilidade do peso total de 5 passageiros (com sua bagagem) ultrapassar 600 kg?

$$
\begin{aligned}
X_1 &: \text{ peso do passageiro}; \ \ X_1 \sim N(\mu_1 = 76, \sigma_1^2 = 15^2) \\
X_2 &: \text{ peso da bagagem}; \ \ X_2 \sim N(\mu_2 = 22, \sigma_2^2 = 5^2) \\
&\text{assumindo independência,} \\
X = X_1 + X_2 &: \text{ peso do passageiro com sua bagagem}; \ \ X \sim N(\mu = 98, \sigma^2 = 15^2 + 5^2 = 250) \\
\text{Teo1}: \ \hat{\mu} = \overline{X}_n &\sim N(\mu = 98, \sigma^2/n = 250/n)
\end{aligned}
$$

(a) $P[X > 100] = P[Z > \frac{100-98}{\sqrt{250}}] = 0,4497$

(b) $P[\sum_{i=1}^{12} X_i > 1200] = P[\overline{X}_{12} > 100] = P[Z > \frac{100-98}{\sqrt{250/12}}] = 0,3306$

(c) $P[\sum_{i=1}^{5} X_i > 600] = P[\overline{X}_5 > 120] = P[Z > \frac{120-98}{\sqrt{250/5}}] = 0,00093$

(d) $P[\sum_{i=1}^{6} X_i > 600] = P[\overline{X}_6 > 100] = P[Z > \frac{100-98}{\sqrt{250/6}}] = 0,3783$

---

63. Considere um estudo na qual deseja-se estimar a proporção de solicitações atendidas e resolvidas de uma central do usuário através de uma amostra aleatória simples.

(a) Qual deveria ser o tamanho da amostra para se obter a estimativa com uma margem de erro de 2,5% com 95% de confiança?

(b) E para uma margem de erro de 3% com 99% de confiança?

Solução:

$$
\begin{aligned}
& X \sim B(p) \\
\text{Teo 2:} \quad & \hat{p} = \overline{X} \approx N(p, \frac{p(1-p)}{n}) \\
& \hat{p} \pm z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} \\
& \hat{p} \pm \text{ME} \\
\text{ME} &= z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} \\
n &= \frac{(z_{1-\alpha})^2}{\text{ME}^2} p(1-p) \\
& p(1-p) \text{ é limitado superiormente para } p = 0,5 \\
n &= \frac{(z_{1-\alpha})^2}{\text{ME}^2} 0,25
\end{aligned}
$$

(a) $n = \frac{(1{,}96)^2}{(0{,}025)^2} 0,25 = 1537$

(b) $n = \frac{(2{,}576)^2}{(0{,}03)^2} 0,25 = 1844$

---

64. O consumo de combustível de uma frota de ônibus é uma v.a. medida pelo rendimento em km/l que tem distribuição normal com média de 8,2 e desvio padrão de 0,7. Selecionando-se um veículo ao acaso qual a probabilidade de

(a) ter rendimento inferior a 7,0 km/l,

(b) ter rendimento acima de 9,0 km/l,

(c) ter rendimento entre 8,0 e 8,5 km/l.

Selecionando-se uma amostra de 5 veículos, qual a probalidade de que

(d) nenhum deles tenha rendimento inferior a 7,0 km/l

(e) ao menos 4 tenham rendimento entre 8,0 e 8,5 km/l.

Considerando-se ainda a amostra de 5 veículos, que a probabilidade de que

(f) o rendimento médio esteja entre 8,0 e 8,5 km/l,

(g) o rendimento médio esteja abaixo de 7,5 km/l.

Solução:

$$X : \text{consumo de combustível de cada veículo}$$
$$X \sim N(\mu = 8,2; \sigma^2 = 0,7^2)$$

```
> (prA <- pnorm(7, m=8.2, sd=0.7))
```

```
[1] 0,04324
```

```
> (prB <- pnorm(9, m=8.2, sd=0.7, lower=FALSE))
```

```
[1] 0,1265
```

```
> (prC <- diff(pnorm(c(8, 8.5), m=8.2, sd=0.7)))
```

```
[1] 0,2783
```

$$Y_1 : \text{número de veículos com rendimento inferior a 7,0 } km/l$$
$$Y_1 \sim \mathrm{B}(n = 5, p = prA)$$

```
> (prD <- dbinom(0, size=5, prob=prA))
```

```
[1] 0,8017
```

$$Y_2 : \text{número de veículos com rendimento entre 8,0 e 8,5 } km/l$$
$$Y_2 \sim \mathrm{B}(n = 5, p = prC)$$

```
> (prE <- pbinom(3, size=5, prob=prC, lower=FALSE))
```

```
[1] 0,02333
```

$$\overline{X} : \text{rendimento médio de cinco veículos}$$
$$\mathrm{X} \sim N(\mu = 8,2; \sigma^2 = 0,7^2/5)$$

```
> (prF <- diff(pnorm(c(8, 8.5), m=8.2, sd=0.7/sqrt(5))))
```

```
[1] 0,5696
```

```
> (prG <- pnorm(7.5, m=8.2, sd=0.7/sqrt(5)))
```

```
[1] 0,01267
```

---

65. Com o objetivo de se dimensionar um serviço de atendimento ao cliente, foram tomados os tempos de atendimento em minutos de 50 consultas escolhidas ao acaso. Na análise dos dados adotou-se a distribuição exponencial e, o valor calculado do tempo médio de atendimento nas 50 consultas, foi de 6,5 minutos. Neste problema identifique e descreva: qual a população, qual a amostra, qual o parâmetro de interesse, qual o estimador e qual é a estimativa. Além disto discuta qual é a distribuição amostral relevante para inferências neste problema, justificando sua resposta.

Solução:

$$\begin{aligned} X &: \text{tempo de atendimento em consulta} \\ X &\sim Exp(\lambda) \\ f(x) &= \frac{1}{\lambda}\exp\{-\frac{x}{\lambda}\} \end{aligned}$$

- população: tempo de atendimento de cliente no SAC
- amostra: os 50 tempos selecionados
- parâmetro: o valor de λ na expressão de $f(x)$
- estimador: a regra para se estimar o parâmetro,no caso $\hat{\lambda} = \bar{X} = \frac{\sum_i X_i}{n}$
- estimativa: valor do estimador para uma particular amostra, $\hat{\lambda} = \bar{x} = 6,5$
- distribuição amostral: distribuição do estimador. No caso, como o estimador é uma média, pelo *teorema do limite central* tem-se que esta distribuição aproxima-se da normal na medida que cresce o tamanho da amostra.

---

66. Sabe-se que o tempo de uso de um determinado serviço pode ser descrito pela distribuição exponencial e tem média de 30 segundos. Qual a probabilidade de um acesso:

(a) ter o tempo inferior a 12 segundos;

(b) durar entre 12 e 30 segundos;

(c) ultrapassar 42 segundos, sabendo-se que já chegou a 30 segundos?

Os acessos são divididos em três categorias: rápidos (com tempo inferior a 12 segundos), médios (com tempo entre 12 e 30 segundos), longos (com tempos entre 30 e 40 segundos) e demorados (com tempo superior a 42 segundos). Os custos e cada categoria são definidos como sendo, respectivamente 1,5; 2,8; 5,0 e 10,0 unidades de custo.

(d) Qual o custo médio dos acessos?

Se for tomada uma amostra de 100 tempos, qual a probabilidade de que o tempo médio desta amostra:

(e) ultrapasse 35 segundos;

(f) esteja entre 28 e 35 segundos?

(g) Qual deve ser o tamanho da amostra para que a probabilidade do tempo médio da amostra ultrapassar 35 segundos seja no máximo de 0,01?

Desconfia-se que houve alguma mudança no padrão dos tempos de acesso e para verificar isto tomou-se uma amostra de 100 tempos de acesso. A seguir, apanas por ilustração, estão alguns dos valores obtidos.

```
16,8   2,2   16,1   9,5   19,5   1,5   26   6,4   18,9   25,6   ...
   ... 26,1   9,5   10,2
```

A média dos tempos amostrados é de 32,71.

(h) Obtenha um intervalo de confiança (95%) para o tempo médio.

(i) Voce diria que há evidências fortes de que houve uma mudança no comportamento dos tempos de acesso? Justifique.

Solução:

$$X : \text{tempo de acesso ao serviço}$$
$$X \sim \mathrm{E}(\lambda = 1/30) \quad f(x) = \frac{1}{30}\exp\{-x/30\} \quad F(x) = 1 - \exp\{-x/30\}$$

(a) $P[X < 12] = \int_0^{12} f(x)\mathrm{d}x = F(12) = 0,3297$

(b) $P[12 < X < 30] = \int_{12}^{30} f(x)\mathrm{d}x = F(30) - F(12) = 0,3024$

(c) Pela propriedade de falta de memória de exponencial,
$P[X > 42|X > 30] = P[X > 12] = 1 - P[X < 12] = 1 - \int_0^{12} f(x)\mathrm{d}x = 1 - F(12) = 0,6703$

(d) Y : custo de acesso

y	$1,5$	$2,8$	$5,0$	10
$P[Y = y]$	$P[X < 12] = 0,33$	$P[12 < X < 30] = 0,302$	$P[30 < X < 42] = 0,121$	$P[X > 42] = 0,247$

$\mathrm{E}(X) = \sum_i y_i P[Y = y_i] = 1,5 \cdot 0,33 + 2,8 \cdot 0,302 + 5,0 \cdot 0,121 + 10,0 \cdot 0,247 = 4,4$ unidades de custo

$$X \sim \mathrm{Exp}(\lambda = 1/30) \quad \mathrm{E}(X) = 1/\lambda \quad \mathrm{Var}(X) = 1/\lambda^2$$
$$\overline{X}_n \approx N(\mu = 1/\lambda = 30, \sigma^2 = (1/\lambda^2)/n = 30^2/n) \quad n = 100$$

(e) $P[\overline{X} > 35] = P[Z > \frac{35-30}{30/10}] = P[Z > 1,667] = 0,048$

(f) $P[28 < \overline{X} < 35] = P[\frac{28-30}{30/10} < Z < \frac{35-30}{30/10}] = P[-0,667 < Z < 1,667] = 0,7$

(g)

$$P[\overline{X} > 35] \leq 0,01$$
$$P[Z > \frac{35-30}{30/\sqrt{n}}] \leq 0,01$$
$$\frac{35-30}{30/\sqrt{n}} \geq 2,326$$
$$n \geq \frac{2,326^2 30^2}{(35-30)^2} = 195$$

$$\text{Assumindo que } \overline{X}_n \approx N(\mu = 1/\hat{\lambda} = 30, \sigma^2 = (1/\hat{\lambda}^2)/n = 30^2/n)$$

(h) $\hat{\lambda} \pm z_{0,95}\hat{\lambda}/\sqrt{n} \longrightarrow 32,71 \pm 1,96\frac{32,71}{10} \longrightarrow (26,3 \;;\; 39,12)$

(i) Justifique.

---

67. Para efetuar o monitoramento de poluentes em uma determinada área foram coletadas amostras. Os valores obtidos para um determinado elemento poluente são fornecidos a seguir.

```
[1] 39,7 33,9 43,3 36,3 37,6 27,2 34,6 36,6 35,2 37,8 34,7 37,9
```

(a) Caracterize o nível de poluição (deste elemento) na área através de um resumo estatístico adequado dos dados.

(b) Qual valor voce escolheria para estimar o nível de poluição na área?

(c) Qual a estimativa deste valor? Como voce representaria a incerteza sobre este estimativa?

(d) A legislação afirma que se o teor estiver acima de 35 unidades a área deve ser considerada contaminada e sujeita a intervenção de controle. Baseando-se nos dados, voce indicaria a intervenção na área?

Em um relatório foram reportadas análises dos dados que incluiam as informações a seguir.

O nível de poluição na área expresso pela média aritmética dos valores medidos nas amostras é de 36,23 $u.m.$[1]. A margem de erro é de 1,9 $u.m.$ obtida pela expressão $z \cdot 4/\sqrt{12}$ com $z = 1,645$ obtido da distribuição normal, e considerando-se que a variância dos teores é de 16 $u.m.^2$.

(e) Na análise estatística, qual a população, a variável aleatória e a amostra no contexto deste problema?

(f) Qual o estimador escolhido e as estimativa pontual obtida?

(g) Qual a estimativa intervalar e seu nível de confiança?

(h) Quais as suposições utilizadas nas análise?

(i) Como os resultados poderiam ser utilizados para determinar se deve ou não haver intervenção na área?

(j) Qual deveria ser o tamanho da amostra para que a margem de erro fosse de no máximo $1,2u.m.$?

---

68. Considere uma pesquisa para estimar a proporção de votos de um determinado candidato.

(a) Qual o tamanho necessário de uma amostra para uma margem de erro de 2,5% e confiança de 95% ?

(b) Comente sobre as suposições feitas para o cálculo anterior.

Solução:

$$\begin{aligned} X &\sim B(\theta) \\ \mu_x &= E[X] = \theta \\ \sigma^2_X &= Var[X] = \theta(1-\theta) \end{aligned}$$

(a)

$$\begin{aligned} \overline{\theta} &\approx N(\mu_{\overline{X}} = \mu_X = \theta; \sigma^2_{\overline{X}} = \sigma^2_X/n = \frac{\theta(1-\theta)}{n}) \\ M.E. &= z\sqrt{\frac{\theta(1-\theta)}{n}} = 0,025 \\ &\text{assumindo } \theta = 1/2 \\ M.E. &= 1.96\frac{1}{\sqrt{4 \cdot n}} = 0,025 \\ n &= \frac{1,96^2}{4 \cdot 0,025^2} = 1537 \end{aligned}$$

(b)

---

69. Assume-se que o tempo de votação em uma urna eletrônica possui distribuição uniforme com valores entre 5 e 30 segundos. Considere um grupo de 50 votantes.

(a) Qual a probabilidade do último votante gastar mais que 20 segundos?

(b) Qual a probabilidade do tempo de votação de todo o grupo ser superior a 15 minutos?

(c) Comente sobre as suposições feitas para o cálculo anterior.

[1] $u.m.$: unidade de medida

Solução:

$$
\begin{aligned}
X &\sim U(5, 30) \\
\mu_x = E[X] &= \frac{a+b}{2} = \frac{5+30}{2} = 17,5 \\
\sigma^2_X = Var[X] &= \frac{(b-a)^2}{12} = \frac{(30-5)^2}{12} = 52,08
\end{aligned}
$$

(a) $P[X > 20] = \frac{30-20}{30-5} = 0,4$

(b)

$$
\begin{aligned}
&\overline{X} \approx N(\mu_{\overline{X}} = \mu_X = 17,5 \;\; ; \;\; \sigma^2_{\overline{X}} = \sigma^2_X/n = 1,04) \\
&P[\sum_{i=1}^{50} X_i > 15' \cdot 60] = P[\frac{\sum_{i=1}^{50} X_i}{50} > \frac{15 \cdot 60''}{50}] = P[\overline{X} > 18] = 0,31
\end{aligned}
$$

(c)

---

70. Um máquina enche latas com um volume nominal de certo material, entretanto há variações de volume de uma lata para outra. A fim de monitorar e controlar o processo foi tomada uma amostra de latas e medido o volume individualmente (em ml). fornecendo os dados mostrados a seguir.

```
[1] 362 357 359 358 365 362 355 360 363 364 360 360
```

(a) Caracterize o volume de enchimento das latas através de um resumo estatístico adequado dos dados.

(b) Qual medida amostral voce escolheria para estimar o real volume de enchimento do equipamento?

(c) Qual a estimativa deste valor? Como voce representaria a incerteza sobre este estimativa?

(d) A especificação de volume é de 350 ml e caso o valor seja ultrapassado ou esteja abaixo do especificado, a processo deve ser parado para a máquina ser regulada. Baseando-se nos dados, voce indicaria a parada para regulagem?

(e) Idem anterior para volume de 360 ml.

Utilizando os conceitos e resultados discutidos nas aulas responda aos seguintes items. Faça a suposição de que os volumes das latas possuem uma variância de 9 unidades.

(f) Obtenha estimativas pontual e intervalar para o volume das latas.

(g) Quais as suposições que foram feitas na obtenção destas estimativas?

(h) Utilize os resultados para indicar se recomenda ou não uma parada para regulagem, justificando sua resposta.

(i) Qual deveria ser o tamanho mínimo da amostra para que a margem de erro da estimativa fosse de no máximo de 1 unidade (com nível de confiança de 95%).

(j) Para amostra do tamanho atual $n = 12$, qual deveria ser a variância dos volumes da lata para que a margem de erro não ultrapasse 0,5 unidades?

Solução:

$X \sim N(\mu \;\; ; \;\; \sigma^2 = 9 = 3^2)$

(a)

i.

Estimador Pontual: média amostral, $\overline{X} = \frac{\sum_{1=1}^{n} X_i}{n}$

Estimativa pontual: $\overline{x} = 360,42$

ii.

Estimador Intervalar: $\overline{X} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$

Estimativa Intervalar (para confiança de 90%): $360,42 \pm 1,645 \cdot \frac{3}{\sqrt{12}} = (359,94; 360,89)$

(b) amostra aleatória (independência), variância populacional conhecida (σ^2=9), confiança de 90%, média amostral como estimador adequado.

(c) Um possível critério é: verificar se o valor de referência (350) está ou não contido no intervalo; não estando rocomenda-se a parada. Respostas e justificativas serão analisadas.

(d) Vamos assumir nível de confiança de 90%

$$M.E. = z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}$$
$$1 = 1,645 \cdot \frac{3}{\sqrt{n_0}}$$
$$n_0 = \left(\frac{1,645 \cdot 3}{1}\right)^2 = 25$$

(e)

$$M.E. = z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}$$
$$1,5 = 1,645 \cdot \frac{\sigma}{\sqrt{12}}$$
$$\sigma = \frac{1.5^2 \cdot 12}{1,645} = 16,41$$

---

71. Considere uma pesquisa para estimar a proporção de votos de um determinado candidato.

(a) Qual o tamanho necessário de uma amostra para uma margem de erro de 2,5% e confiança de 95% ?

(b) Comente sobre as suposições feitas para o cálculo anterior.

Solução:

$$X \sim B(\theta)$$
$$\mu_x = E[X] = \theta$$
$$\sigma^2_X = Var[X] = \theta(1-\theta)$$

(a)

$$\overline{\theta} \approx N(\mu_{\overline{X}} = \mu_X = \theta; \sigma^2_{\overline{X}} = \sigma^2_X/n = \frac{\theta(1-\theta)}{n})$$
$$M.E. = z\sqrt{\frac{\theta(1-\theta)}{n}} = 0,025$$
$$\text{assumindo } \theta = 1/2$$
$$M.E. = 1.96\frac{1}{\sqrt{4 \cdot n}} = 0,025$$
$$n = \frac{1,96^2}{4 \cdot 0,025^2} = 1537$$

(b)

72. Assume-se que o tempo de votação em uma urna eletrônica possui distribuição uniforme com valores entre 5 e 30 segundos. Considere um grupo de 50 votantes.

(a) Qual a probabilidade do último votante gastar mais que 20 segundos?

(b) Qual a probabilidade do tempo de votação de todo o grupo ser superior a 15 minutos?

(c) Comente sobre as suposições feitas para o cálculo anterior.

Solução:

$$\begin{aligned} &X \sim U(5, 30) \\ &\mu_x = E[X] = \frac{a+b}{2} = \frac{5+30}{2} = 17,5 \\ &\sigma^2_X = Var[X] = \frac{(b-a)^2}{12} = \frac{(30-5)^2}{12} = 52,08 \end{aligned}$$

(a) $P[X > 20] = \frac{30-20}{30-5} = 0,4$

(b)

$$\begin{aligned} &\overline{X} \approx N(\mu_{\overline{X}} = \mu_X = 17,5 \ ; \ \sigma^2_{\overline{X}} = \sigma^2_X/n = 1,04) \\ &P[\sum_{i=1}^{50} X_i > 15' \cdot 60] = P[\frac{\sum_{i=1}^{50} X_i}{50} > \frac{15 \cdot 60''}{50}] = P[\overline{X} > 18] = 0,31 \end{aligned}$$

(c)

---

73. Um empresa é responsável pelo monitoramento de 12 reservatórios (A, B, C, . . . K, L) que são inspecionados periodicamente. Os custos de inspeção variam sendo de R$ 1.000,00 para A e B; R$ 2.000,00 para C e D, R$ 3.000,00 para E e F e R$ 4.000,00 para os demais. Em cada inspeção são selecionados aleatoriamente três reservatórios. Qual a probabilidade de que em uma inspeção:

(a) seja gasto R$ 6.000,00

(b) seja gasto mais que R$ 10.000,00

(c) seja gasto mais que R$ 8.000,00 sabendo-se que o custo é superior a R$ 5.000,00

(d) sejam selecionados dois ou mais reservatórios de mesmo custo

(e) sejam selecionados reservatórios de custos diferentes

Solução:

---

74. Em um levantamento geológico foram coletadas amostras de sedimentos de fundo de rios de uma bacia hidrográfica. Os teores obtidos de um certo elemento são mostrados a seguir.

```
35.1 5.7 5.8 4.7 1.5 2.8 70.2 3.0 8.3 6.3 17.8 16.3 7.0

1.8 1.9 5.4 4.5 4.5 19.2 8.4 9.9 6.8 2.8 21.1 1.3
```

(a) obtenha o teor médio e o desvio padrão,

(b) obtenha os quantis e a amplitude,

(c) obtenha o coeficiente de variação,

(d) obtenha um histograma,
(e) obtenha um box-plot,
(f) obtenha um diagrama de ramo-e-folhas.

Pretende-se usar os dados para estimar um valor que represente o teor característico da bacia. Neste contexto identifique na problema os seguintes elementos:

(g) A população e a variável aleatória em questão,
(h) O parâmetro, algum estimador e alguma estimativa
(i) Obtenha alguma estimativa intervalar com o nível de confiança de 90%
(j) Quais as suposições feitas para obtenção desta estimativa? Avalie se a estimativa é adequada para os dados obtidos.

Ainda no contexto do problema considera-se que o teor está acima do nível de segurança se o nível na região estiver acima de 8,5 unidades e será feito um teste estatístico para determinar se a amostra indica que o nível de segurança foi ultrapassado com nível de significância de $\alpha = 0,10$.

(k) Identifique as hipóteses do estudo
(l) Descreva o que seriam os erros tipo I e II neste contexto
(m) Proceda o teste de hipótese descrevendo a conclusão no contexto do problema
(n) Suponha agora que o teor verdadeiro da região esteja em 11 unidades. e suponha ainda que o desvio padrão seja de 16 unuidades. Com estes valores, qual a probabilidade do erro tipo II?
(o) Discuta a adequação do teste considerando a natureza dos dados e as suposições feitas para realização do teste

Solução:

---

75. Uma central de atendimento classifica as solicitações de atendimento em 12 categorias (A, B, C, ...K, L) de acordo com a complexidade. Assume-se que a probabilidade de classificação é a mesma para todas as categorias. Os custos são de R$ 100,00 para categorias A e B; R$ 200,00 para C e D, R$ 300,00 para E e F e R$ 400,00 para os demais. A cada dia são feitos três atendimentos. Qual a probabilidade de que em um dia o custo total:

    (a) seja de R$ 600,00 ,
    (b) seja de pelo menos R$ 1.000,00 ,
    (c) seja de pelo menos R$ 900,00 , sabendo-se que o custo é de pelo menos R$ 600,00 ,
    (d) sejam selecionados dois ou mais atendimentos de mesmo custo ,
    (e) sejam selecionados três atendimentos de custos diferentes.
    (f) Considerando uma semana de 5 dias úteis, qual o custo semanal (total) esperado?

76. O tempo de processamento de um novo algorítmo proposto para processamento de imagens foi anotado em um teste com 25 imagens selecionadas ao acaso e os valores mostrados a seguir.

```
35.1 5.7 5.8 4.7 1.5 2.8 70.2 3.0 8.3 6.3 17.8 16.3 7.0

1.8 1.9 5.4 4.5 4.5 19.2 8.4 9.9 6.8 2.8 21.1 1.3
```

    (a) obtenha o tempo médio e o desvio padrão,
    (b) obtenha os quantis e a amplitude,
    (c) obtenha o coeficiente de variação,
    (d) obtenha um histograma,

(e) obtenha um box-plot,

(f) obtenha um diagrama ramo-e-folhas.

Pretende-se usar os dados para estimar um valor que represente o tempo característico do algorítimo. Neste contexto identifique na problema os seguintes elementos:

(g) A população e a variável aleatória em questão.

(h) O parâmetro, algum estimador e alguma estimativa.

(i) Obtenha alguma estimativa intervalar com o nível de confiança de 90%.

(j) Quais as suposições feitas para obtenção desta estimativa? Avalie se a estimativa é adequada para os dados obtidos.

Ainda no contexto do problema, considera-se que o tempo está é aceitável (em comparação com o algorítmo atualmente utilizado) se o tempo característico médio estiver abaixo de 12,0 unidades. Será feito um teste estatístico para determinar se a amostra indica que o tempo desejado foi atingido com nível de significância de $\alpha = 0,10$.

(k) Identifique as hipóteses do estudo.

(l) Descreva o que seriam os erros tipo I e II neste contexto.

(m) Proceda o teste de hipótese descrevendo a conclusão no contexto do problema.

(n) Suponha agora que o teor verdadeiro da região esteja em 11 unidades. e suponha ainda que o desvio padrão seja de 16 unidades. Com estes valores, qual a probabilidade do erro tipo II?

(o) Discuta a adequação do teste considerando a natureza dos dados e as suposições feitas para realização do teste

Solução:

---

77. Uma amostra aleatória de 60 impressoras foi submetida a testes de impressão e em 12 delas algum defeito relevante foi observado.

(a) Encontre um intervalo de confiança (95%) para a proporção p de defeitos verdadeira (populacional).

(b) Uma impressora é considerada aceitável se a proporção p é menor que 0,25 (25%). Use um teste de hipótese adequado para decidir ($\alpha = 0,01$) se a impressora é aceitável.

(c) Usando a estimativa da proporção obtida na amostra como sendo o verdadeiro valor da proporção p, calcule um novo tamanho de amostra de impressoras a serem testadas para que a verdadeira proporção seja estimada com uma margem de erro de 0,025.

(d) Quanto deveria ser o novo tamanho da amostra no item anterior para que a margem de erro fosse de no máximo 0,025 independentemente do valor da verdadeira proporção?

Solução:

$$X : \text{defeito}$$
$$X \sim Ber(p)$$
$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$
$$n = 60 \qquad \hat{p} = 12/60 = 0,2$$

(a) IC assintótico:
$\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}} \longrightarrow 0,20 \pm 1,96\sqrt{\frac{0,2\cdot(1-0,2)}{60}} \longrightarrow 0,20 \pm 0,101 \longrightarrow (0,099; 0,301)$
IC conservador:
$\hat{p} \pm z\sqrt{\frac{1}{4n}} \longrightarrow 0,20 \pm 1,96\sqrt{\frac{1}{4\cdot 60}} \longrightarrow 0,20 \pm 0,127 \longrightarrow (0,073; 0,327)$

(b)

$$H_0: p \geq 0,25 \text{ (não aceitável)} \quad vs \;\; H_1: p < 0,25 \text{ (aceitável)}$$
$$\alpha = 0,01 \qquad RC: \{z > -2,326\}$$
$$z_c = \frac{0,20 - 0,25}{\sqrt{\frac{0,25(1-0,25)}{60}}} = -0,894$$
$$\text{valor-p} = 0,18555$$

Conclusão: $z_c \notin RC$ ou valor-p $> \alpha(0,01)$. Não rejeita-se H_0 para o nível de significância de 1%, ou seja, não há evidências suficientes de que a proporção seja inferior a 25% e a impressora é considerada não aceitável.

(c) $z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,025 \longrightarrow n = \lceil \frac{z^2\hat{p}(1-\hat{p})}{0,025^2} = \lceil \frac{1,96^2(0,2)(0,8)}{0,025^2} = 984$

(d) $z\sqrt{\frac{1}{4n}} = 0,025 \longrightarrow n = \lceil \frac{z^2}{4 \cdot 0,025^2} = \lceil \frac{1,96^2}{4 \cdot 0,025^2} = 1537$

---

78. Mostre que, assumindo-se que uma variável possa ser descrita por um valor médio μ mais um desvio aleatório tal que $y_i = \mu + \epsilon_i$, o estimador $\hat{\mu} = \bar{x} = \sum_{i=1}^n x_i$ é um estimador de mínimos quadrados de μ.

Solução:

Notando que:

$$y_i = \mu + \epsilon_i \longrightarrow \epsilon_i y_i - \mu Q(\mu) \qquad = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \mu)^2$$
$$\frac{\mathrm{d}Q(\mu)}{\mathrm{d}\mu} = 0 \longrightarrow \sum_{i=1}^n (-2)(y_i - \hat{\mu}) = 0 \longrightarrow \sum_{i=1}^n y_i - n\hat{\mu} = 0$$
$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n}$$

---

79. Encontre a expressão do estimador de máxima verossimilhança da parâmetro da distribuição exponencial. Usando este estimador obtenha a estimativa do parâmetro para uma amostra que forneceu os seguinte valores: `23,7; 12,8; 11,1; 5,0; 7,0; 10,1; 2,3; 5,4; 15,2; 8,5; 4,7; 1,5.`

Solução:

$$X \sim \mathrm{Exp}(\lambda)$$
$$f(x) = \lambda e^{-\lambda x} \mathrm{I}_{(0,\infty)}(x)$$
$$l(\lambda) = \log \prod_{i=1}^n f(x_i) = \sum_{i=1}^n \log\{f(x_i)\} =$$
$$= \sum_{i=1}^n (\log(\lambda) - \lambda x_i) = n\log(\lambda) - \lambda \sum_{i=1}^n x_i$$
$$\frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} = 0 \longrightarrow \frac{n}{\hat{\lambda}} - \sum_{i=1}^n x_i = 0 \longrightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$$

Para os dados da amostra a estimativa é:

$$\hat{\lambda} = 1/\bar{x} = 0,11$$

---

80. Um fabricante produz anéis de piston para automóveis. O processo produz anéis com diâmetro aproximadamente normal com desvio padrão de 0,001 mm. Uma amostra de 15 anéis mostrou um diâmetro médio de 74,036 mm.

   (a) Encontre intervalos de confiança a 90, 95 e 99% para o valor médio do diâmetro dos anéis.

   O processo deve produzir anéis com diâmetro médio de 74,03 mm. Determina-se que se a média das amostras ($n = 15$) estiver mais que 0,005 unidades afastada desta média, o processo é interrompido para inspeção e ajustes.

   (b) Com esta "regra", qual a probabilidade de interromper o processo desnecessariamente?

   (c) Com a amostra dada o processo é interrompido. Qual a probabilidade dessa decisão ter sido equivocada?

   (d) Qual deveria ser a "regra"para interrupção do processo para que a probabilidade de uma parada desnecessária não fosse superior a 0,01?

Solução:

$$X : \text{diâmetro dos anéis}$$
$$X \sim N(\mu, \sigma^2 = 0,001^2)$$
$$\text{amostra} : n = 15 \quad \bar{x} = 74,036$$

(a)

$$IC_{1-\alpha}(\mu) : \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$
$$IC_{90\%}(\mu) : 74,036 \pm 1,6449 \frac{0,001}{\sqrt{15}} \longrightarrow (74,0356; 74,0364)$$
$$IC_{95\%}(\mu) : 74,036 \pm 1,96 \frac{0,001}{\sqrt{15}} \longrightarrow (74,0355; 74,0365)$$
$$IC_{99\%}(\mu) : 74,036 \pm 2,5758 \frac{0,001}{\sqrt{15}} \longrightarrow (74,0353; 74,0367)$$

(b) $P[|\bar{X}_{15} - \mu| > 0,005 | \mu = 74,03] = 2 \cdot P[Z > \frac{0,005}{0,001/\sqrt{15}}] = 0$

(c) $P[\bar{X}_{15} \geq 74,036 | \mu = 74,03] = P[Z > \frac{74,036-74,03}{0,001/\sqrt{15}}] = P[Z > 23,238] = 0$

(d) $P[|\bar{X}_{15} - \mu| > k | \mu = 74,03] = 0,01 \longrightarrow z = 2,576 = \frac{k}{0,001/\sqrt{15}} \longrightarrow k = 0,00067$

---

81. Uma amostra aleatória de 50 capacetes foi submetida a testes de impacto e em 18 deles algum dano relevante foi observado.

   (a) Encontre um intervalo de confiança (95%) para a proporção p de capacetes verdadeira (populacional).

   (b) O capacete é considerado aceitável se a proporção p é menor que 0,40 (40%). Use um teste de hipótese adequado para decidir ($\alpha = 0,01$) se esse capacete é aceitável.

   (c) Usando a estimativa da proporção obtida na amostra como sendo o verdadeiro valor da proporção p, calcule um novo tamanho de amostra de capacetes a serem testados para que a verdadeira proporção seja estimada com uma margem de erro de 0,025.

   (d) Quanto deveria ser o novo tamanho da amostra no item anterior para que a margem de erro fosse de no máximo 0,025 independentemente do valor da verdadeira proporção?

Solução:

$$
\begin{aligned}
&X : \text{defeito}\\
&X \sim Ber(p)\\
&\hat{p} \sim N(p, \frac{p(1-p)}{n})\\
&n = 50 \quad \hat{p} = 18/50 = 0,36
\end{aligned}
$$

(a) IC assintótico:
$\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}} \longrightarrow 0,36 \pm 1,96\sqrt{\frac{0,36\cdot(1-0,36)}{50}} \longrightarrow 0,36 \pm 0,133 \longrightarrow (0,227, 0,493)$
ou para IC conservador:
$0,36 \pm 1,96\sqrt{\frac{1}{4\cdot 50}} \longrightarrow 0,36 \pm 0,139 \longrightarrow (0,221, 0,499)$

(b)

$$
\begin{aligned}
&H_0 : p \geq 0,40 \text{ (não aceitável)} \quad vs \quad H_0 : p < 0,40 \text{ (aceitável)}\\
&\alpha = 0,01 \qquad RC : \{z > -2,326\}\\
&z_c = \frac{0,36 - 0,40}{\sqrt{\frac{0,40(1-0,40)}{50}}} = -0,577\\
&\text{valor-p} = 0,28185
\end{aligned}
$$

Conclusão: $z_c \notin RC$ ou valor-p $> \alpha(0,01)$. Não rejeita-se H_0 para o nível de significância de 1%, ou seja, não há evidências suficientes de que a proporção seja inferior a 40% e o capacete é considerado não aceitável.

(c) $z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,025 \longrightarrow n = \lceil \frac{z^2\hat{p}(1-\hat{p})}{0,025^2} = \lceil \frac{1,96^2(0,36)(0,64)}{0,025^2} = 1417$

(d) $z\sqrt{\frac{1}{4n}} = 0,025 \longrightarrow n = \lceil \frac{z^2}{4\cdot 0,025^2} = \lceil \frac{1,96^2}{4\cdot 0,025^2} = 1537$

---

82. Estimação e estimadores:

(a) encontre a expressão do estimador de mínimos quadrados the β no modelo para uma v.a. Y em que $y_i = \beta x_i + \epsilon_i$.

(b) encontre a expressão do estimador de máxima verossimilhança para o parâmetro λ da distribuição de Poisson. Encontre a estimativa de λ para uma amostra que forneceu os seguintes valores:
`9 15 13 10 9 8 7 5 6 8 9`

Solução:

(a)

$$
\begin{aligned}
Q(\beta) &= \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n}(y_i - \beta x_i)^2\\
\frac{\mathrm{d}Q(\beta)}{\mathrm{d}\beta} &= 0 \longrightarrow \sum_{i=1}^{n}(-2x_i)(y_i - \hat{\beta}x_i) = 0 \longrightarrow \sum_{i=1}^{n} x_i y_i - \hat{\beta}\sum_{i=1}^{n} x_i = 0\\
\hat{\beta} &= \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}
\end{aligned}
$$

(b)

$$
\begin{aligned}
X &\sim P(\lambda) \\
P[X=x] &= \frac{e^{-\lambda}\lambda^x}{x!} \\
l(\lambda) &= \log \prod_{i=1}^{n} P[X=x_i] = \sum_{i=1}^{n} \log\left\{P[X=x_i]\right\} = \\
&= \sum_{i=1}^{n} \left(-\lambda + x_i \log(\lambda) - \log(x_i!)\right) = -n\lambda + \log(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!) \\
\frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} &= 0 \longrightarrow -n + \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}} = 0 \longrightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}
\end{aligned}
$$

Para os dados da amostra a estimativa é:

$\bar{x} = 9$

---

83. Um novo combustível foi desenvolvido e deseja-se verificar se ele aumenta o rendimento (medido em km/l) de veículos e se o aumento justifica o seu custo adicional. Para isto foi selecionada uma amostra de 12 veículos e medidas as diferenças em rendimentos (depois - antes) obtendo-se os seguintes valores:

Veículo	1	2	3	4	5	6	7	8	9	10	11	12
Diferença	2,1	0,7	-0,1	2,7	1,6	1,7	2,4	0,6	9,0	1,5	1,1	1,6

(a) Obtenha a média, desvio padrão e coeficiente de variação das diferenças de rendimento.

(b) Faça um gráfico adequado para resumir os dados.

(c) Com as suposições adequadas, faça um teste estatístico para verificar se há aumento (estatisticamente significativo) de rendimento usando $\alpha = 0,05$.

(d) Para justificar o custo adicional é necessário que o aumento de rendimento seja de ao menos $0,75km/l$. Utilize um teste estatístico ($\alpha = 0,10$) para decidir se o novo combustível deve ou não ser recomendado.

Solução:

---

84. Em um procedimento de monitoramento e controle de qualidade da água de uma cidade foram coletadas amostras em diferentes locais e analisados teores de vários elementos e substâncias. Vamos considerar aqui apenas o teor de um dos elementos para o qual foram obtidos os valores listados a seguir.

```
23,2  20,3  21,4  25,6  22,7  20,6  25,6  22,4  26,0  23,4  24,3  23,20
```

A análise destes dados forneceu um intervalo de confiança (95%) para o teor médio de $(22,015; 24.435)$. Neste contexto é importante que os teores sejam razoavelmente homogêneos e há um valor de referência para variabilidade dos teores expresso pela variância de 2,5 unidades. A variância calculada para a amostra é de $3,6275$. Nestas situações um teste estatístico de hipótese com nível de significância de 5% é feito para verificar se a variância está ultrapassando o limite adequado. Para este caso o teste forneceu um $valor - p$ de $0,1426$. Com base nesta informações responda:

(a) qual a população em questão e a variável aleatória (v.a.)?

(b) qual(ais) é(são) o(s) parâmetro(s) de interesse da distribuição de probabilidades da v.a.?

(c) qual(ais) é(são) o(s) estimador(es) e estimativa(s) deste(s) parâmetro(s) obtida(s) nas análises?

(d) qual(ais) é(são) a(s) distribuições amostrais utilizada(s) nas análises?

(e) qual seria o intervalo de confiança para o teor médio se fosse adotado o nível de 99% ?

(f) quais as hipóteses que são testadas no teste estatístico? (descreva as hipóteses no contexto do problema e formula as hipóteses estatísticas)

(g) qual a estatística de teste utilizada? Qual o seu valor para estes dados?

(h) qual a região crítica?

(i) qual a conclusão do teste de hipóteses e sua consequência para conclusões no contexto do estudo.

(j) quais seriam as conclusões se fosse adotado o nível de significância de 10% ?

(k) em uma nova amostragem deseja-se que a média seja estimada com margem de erro 0,6 unidades para intervalos de confiança de 95%. Qual tamanho de amostra deve ser utilizado neste novo estudo?

(l) cogita-se adotar uma norma segundo a qual se a variância dos teores na amostra estiver acima de $3,5$ unidades, é feita uma vistoria detalhada. Neste contexto, quais seriam as interpretações e consequências práticas do erro tipo I e II?

(m) calcule a probabilidade do erro tipo I caso a recomendação seja adotada.

Solução:

(a) *População:* elemento de interesse na área, v.a. X: teores do elemento de interesse

(b) $X \sim N(\mu, \sigma^2)$. No contexto os parâmetros de interesse para inferência são a média μ e a variância σ^2

(c) $\hat{\mu} = \bar{X} = \frac{\sum X_i}{n}$ e $\bar{x} = 23,23$
$\hat{\sigma}^2 = S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$ e $s^2 = 3,6275$

(d) Distribuição amostral da média $\bar{X} \sim N(\mu, \sigma^2/n)$ e da variância $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

(e)

$$IC : \bar{x} \pm t_{1-\alpha/2,\nu} S/\sqrt{n}$$
$$t_{0.005,11} = 3,1058$$
$$IC : [21,52, 24,93]$$

(f) A variância está em níveis aceitáveis *vs* a variância está acima do aceitável.
$H_0 : \sigma^2 \leq 2,5$ *vs* $H_1 : \sigma^2 > 2,5$

(g) $\chi^2_c = \frac{(n-1)S^2}{\sigma^2_0} = \frac{(12-1)3{,}6275}{2{,}5} = 17,412$

(h) $RC = \{\chi^2 : \chi^2_c > 21,0261\} = \{S^2 : S^2 > 4,7787\}$

(i) Não rejeita-se H_0 para $\alpha = 0,05$ pois $p - valor = 0,1426 > \alpha$, portanto não há evidência suficiente de que a variância tenha ultrapassado 2,5 (para nível de significância de $\alpha = 0,05$).

(j) A mesma da anterior pois ainda assim $p - valor = 0,1426 > \alpha$.

(k) Para $\sigma^2 = 2,5$:
$ME = z\frac{\sigma}{\sqrt{n}} \longrightarrow n = \frac{z^2\sigma^2}{0{,}6^2} = 14$
Se fosse adotado que $\sigma^2 = S^2$:
$ME = z\frac{\sigma}{\sqrt{n}} \longrightarrow n = \frac{z^2 3{,}6275}{0{,}6^2} = 20.$

(l) Erro I: fazer a vistoria quando não é necessário, o que implica em custos desnecessários.
Erro II: não fazer a vistoria quando é necessário, o que implica em não tomar medidas de controle quando estas são necessárias

(m) $P(ErroI) = P(S^2 > 3,5|\sigma^2 = 2,5) = P(\frac{(n-1)S^2}{2{,}5} > \frac{(12-1)3{,}5}{2{,}5}) = P(\chi^2 > 15,4) = 0,1649$

---

85. Considere uma variável aleatória com distribuição normal de média conhecida $\mu = a$ e variância σ^2 desconhecida a ser estimada a partir de uma amostra. Obtenha a expressão do estimador de σ^2 pelo método da máxima verossimilhança.

Solução:

---

86. Em uma pesquisa para verificar se uma população tem conhecimento sobre uma certa legislação são feitas pesquisas periodicamente. Há uma recomendação de que seja lançada uma campanha publicitária se a proporção de pessoas na população com conhecimento sobre a legislação estiver abaixo de 60%. Em uma destas pesquisas foram ser entrevistadas 1500 pessoas das quais 800 mostraram conhecer a legislação.

    (a) Quais o intervalo de confiança a 90% e 95% para a proporção da população que conhece a legislação?

    (b) Use um procedimento estatístico adequado para dizer se a campanha publicitária deve ou não ser lançada.

    (c) Deseja-se reduzir o tamanho da amostra em uma próxima pesquisa para 800 pessoas. Discuta mostrando qual o impacto esperado nos resultados das análises como as feitas nos itens anteriores.

Solução:

$$\begin{aligned} X &: \text{conhece a legislação} \\ X &\sim B(p) \\ \text{amostra} &: n = 1500; \sum x_i = 800 \end{aligned}$$

(a) $\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$

```
[1] 0,5077 0,5588
attr(,"conf.level")
[1] 0,95
```

```
[1] 0,4997 0,5666
attr(,"conf.level")
[1] 0,99
```

(b) Teste de Hipótese:

$$\begin{aligned} H_0 : p \geq 0.6(\text{campanha não é lançada}) \text{ vs } H_1 &: p < 0.6(\text{campanha é lançada}) \\ \alpha = 0,05 \longrightarrow \hat{p} < 0,579 \\ z_c = \frac{0,533 - 0,6}{\sqrt{0,6(1-0,6)/1500}} = -5,27 \\ p - valor = 1e - 07 \end{aligned}$$

Conclusão: Rejeita-se H_0 e a campanha deve ser lançada.

(c) Os intervalos de confiança seriam mais largos, as margens de erro seriam $0,035$ (95%) e $0,046$ (99%). No teste de hipótese a região crítica para $\alpha = 0,05$ passaria a ser $\hat{p} < 0,572$ ou seja, o limite na proporção na amostra deve ser menor para a campanha ser lançada.

---

87. O tempo que passageiros esperam no balcão de *check-in* de uma cia aérea é uma variável aleatória com média de 13,5 minutos e desvio padrão de 1,8 minutos. Vai ser tomada uma amostra aleatória do tempo de espera de 36 passageiros. Encontre a probabilidade:

    (a) do tempo de espera de um passageiro escolhido ao acaso ser maior que 15 min;

    (b) do tempo médio dos 36 passageiros estar acima de 15 min;

    (c) do tempo médio de atendimento dos 36 passageiros estar entre 13 e 14 minutos

Solução:

$$X \sim N(\mu = 13,5, \sigma^2 = 1,8^2)$$
$$\bar{X_{36}} \sim N(\mu = 13,5, \sigma^2 = 1,8^2/36)$$

(a) $P(X > 15) = P(Z > \frac{15-13,5}{1,8}) = P(Z > 0,833) = 0,2$

(b) $P(\bar{X} > 15) = P(Z > \frac{15-13,5}{1,8/\sqrt{36}}) = P(Z > 5) = 0$

(c) $P(13 < \bar{X} < 14) = P(\frac{13-13,5}{1,8/\sqrt{36}} < Z < \frac{14-13,5}{1,8/\sqrt{36}}) = P(-1,667 < Z < 1,667) = 0,9$

---

88. Considere ainda o contexto do problema anterior, porém considerando que a cia aérea desconhece o tempo médio de atendimento mas tem como meta que o tempo médio seja de 12 min. A cia aérea vai utilizar as informações da amostra (com $n = 36$) para decidir se disponibiliza mais um funcionário para atendimento de *check-in*. Decidiu-se que, se o tempo médio de atendimento calculado com os dados da amostra for superior a 13 min o funcionário é disponibilizado.

(a) Mostre como um teste de hipótese estatístico poderia ser utilizado para decidir sobre a contratação. Descreva as hipóteses no contexto do problema e em notação estatística. Descreva quais seriam os passos para efetuar o teste e como seriam tiradas as conclusões.

(b) Descreva, no contexto do problema, quais seriam os erros tipo I e II

(c) Calcule a probabilidade do erro tipo I

(d) Calcule a probabilidade do erro tipo II supondo que a verdadeira média é de fato 13,5 min

(e) Qual deve ser o tamanho da amostra para que a probabilidade do erro tipo I seja de 0.01 ($\alpha = 1\%$)

Solução:

(a)

(b) **Erro I:** contratar funcionário quando desnecessário
Erro II: não contratar funcionário quando necessário

(c) $P[\text{Erro I}] = P[\bar{X} > 13 | \mu = 12] = P[Z > \frac{13-12}{1,8/\sqrt{36}}] = 0,00043$

(d) $P[\text{Erro II}] = P[\bar{X} < 13 | \mu = 13,5] = P[Z < \frac{13-13,5}{1,8/\sqrt{36}}] = 0,048$

(e) $z_{0,99} = \frac{13-12}{1.8/\sqrt{n}} = 2,3263 \longrightarrow n = \frac{2,3263^2 1,8^2}{(13-12)^2} = 18$

---

89. Estuda-se a fração de circuitos integrados defeituosos produzidos em um processo de fotolitografia. Uma amostra aleatória de 300 circuitos é testada, revelando 33 defeitos.

(a) Obtenha intervalos de confiança a 90% e 95% para a proporção esperada de circuitos defeituosos na produção.

(b) A fábrica adota como prática uma parada para inspeção se houver evidência estatística que a proporção de defeituosos esteja acima de 10%. Use o procedimento estatístico adequado para dizer se a inspeção é recomendada com a amostra obtida.

Solução:

$$X : \text{apresenta defeito}$$
$$X \sim B(p)$$
$$\text{amostra} : n = 300; \sum x_i = 33$$

(a) $\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$

```
[1] 0,07799 0,15233
attr(,"conf.level")
[1] 0,95
```

```
[1] 0,07031 0,16715
attr(,"conf.level")
[1] 0,99
```

(b) Teste de Hipótese:

$$H_0 : p \leq 0.10(\text{não é recomendada inspeção}) \text{ vs } H_1 \quad : p > 0.10(\text{inspeção recomendada})$$
$$\alpha = 0,05 \longrightarrow \hat{p} < 0,128$$
$$z_c = \frac{0,11 - 0,10}{\sqrt{0,10(1-0,10)/300}} = 0,577$$
$$p - valor = 0,2818514$$

Conclusão: Não rejeita-se H_0 e a inspeção não é recomendada.

---

90. Em um procedimento de monitoramento e controle de tempos de atendimento de requisições foram coletados dados compondo uma amostra aleatória que forneceu os seguintes valores.

```
23,2  20,3  21,4  25,6  22,7  20,6  25,6  22,4  26,0  23,4  24,3  23,20
```

A análise destes dados forneceu um intervalo de confiança (95%) para o tempo médio de $(22,015; 24.435)$. Neste contexto é importante que os tempos sejam razoavelmente homogêneos e há um valor de referência para variabilidade dos tempos expresso pela variância de 2,5 unidades. A variância calculada para a amostra é de $3,6275$. Nestas situações um teste estatístico de hipótese com nível de significância de 5% é feito para verificar se a variância está ultrapassando o limite adequado. Para este caso o teste forneceu um $valor - p$ de $0,1426$. Com base nesta informações responda:

(a) qual a população em questão e a variável aleatória (v.a.)?

(b) qual(ais) é(são) o(s) parâmetro(s) de interesse da distribuição de probabilidades da v.a.?

(c) qual(ais) é(são) o(s) estimador(es) e estimativa(s) deste(s) parâmetro(s) obtida(s) nas análises?

(d) qual(ais) é(são) a(s) distribuições amostrais utilizada(s) nas análises?

(e) qual seria o intervalo de confiança para o tempo médio se fosse adotado o nível de 99% ?

(f) quais as hipóteses que são testadas no teste estatístico? (descreva as hipóteses no contexto do problema e formula as hipóteses estatísticas)

(g) qual a estatística de teste utilizada? Qual o seu valor para estes dados?

(h) qual a região crítica?

(i) qual a conclusão do teste de hipóteses e sua consequência para conclusões no contexto do estudo.

(j) quais seriam as conclusões se fosse adotado o nível de significância de 10% ?

(k) em uma nova amostragem deseja-se que a média seja estimada com margem de erro 0,6 unidades para intervalos de confiança de 95%. Qual tamanho de amostra deve ser utilizado neste novo estudo?

(l) cogita-se adotar uma norma segundo a qual se a variância dos tempos na amostra estiver acima de $3,5$ unidades, é feita uma vistoria detalhada no processo. Neste contexto, quais seriam as interpretações e consequências práticas do erro tipo I e II?

(m) calcule a probabilidade do erro tipo I caso a recomendação seja adotada.

Solução:

(a) *População:* tempos de atendimento de requisições, v.a. X: tempos

(b) Assume-se $X \sim N(\mu, \sigma^2)$. No contexto os parâmetros de interesse para inferência são a média μ e a variância σ^2

(c) $\hat{\mu} = \bar{X} = \frac{\sum X_i}{n}$ e $\bar{x} = 23,23$
$\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$ e $s^2 = 3,6275$

(d) Distribuição amostral da média $\bar{X} \sim N(\mu, \sigma^2/n)$ e da variância $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

(e)

$$IC : \bar{x} \pm t_{1-\alpha/2,\nu} S/\sqrt{n}$$
$$t_{0.005,11} = 3,1058$$
$$IC : [21,52, 24,93]$$

(f) A variância está em níveis aceitáveis *vs* a variância está acima do aceitável.
$H_0 : \sigma^2 \leq 2,5$ *vs* $H_1 : \sigma^2 > 2,5$

(g) $\chi^2_c = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(12-1)3,6275}{2,5} = 17,412$

(h) $RC = \{\chi^2 : \chi^2_c > 21,0261\} = \{S^2 : S^2 > 4,7787\}$

(i) Não rejeita-se H_0 para $\alpha = 0,05$ pois $p - valor = 0,1426 > \alpha$, portanto não há evidência suficiente de que a variância tenha ultrapassado 2,5 (para nível de significância de $\alpha = 0,05$).

(j) A mesma da anterior pois ainda assim $p - valor = 0,1426 > \alpha$.

(k) Para $\sigma^2 = 2,5$:
$ME = z\frac{\sigma}{\sqrt{n}} \longrightarrow n = \frac{z^2\sigma^2}{0,6^2} = 14$
Se fosse adotado que $\sigma^2 = S^2$:
$ME = z\frac{\sigma}{\sqrt{n}} \longrightarrow n = \frac{z^2 3,6275}{0,6^2} = 20.$

(l) Erro I: fazer a vistoria quando não é necessário, o que implica em custos desnecessários.
Erro II: não fazer a vistoria quando é necessário, o que implica em não tomar medidas de controle quando estas são necessárias

(m) $P(ErroI) = P(S^2 > 3,5|\sigma^2 = 2,5) = P(\frac{(n-1)S^2}{2,5} > \frac{(12-1)3,5}{2,5}) = P(\chi^2 > 15,4) = 0,1649$

---

91. Considere uma variável aleatória com distribuição de Rayleigh $f(x) = (x/\theta)e^{\{-x^2/2\theta\}} I_{x>0}(x)$ com $\theta > 0$. Obtenha o estimador de θ pelo método da máxima verossimilhança.

Solução:

---

92. O tempo que passageiros esperam no balcão de *check-in* de uma cia aérea é uma variável aleatória com média de 13,5 minutos e desvio padrão de 1,8 minutos. Vai ser tomada uma amostra aleatória do tempo de espera de 36 passageiros. Encontre a probabilidade:

(a) do tempo de espera de um passageiro escolhido ao acaso ser maior que 15 min;

(b) do tempo médio dos 36 passageiros estar acima de 15 min;

(c) do tempo médio de atendimento dos 36 passageiros estar entre 13 e 14 minutos

Solução:

$$X \sim N(\mu = 13,5, \sigma^2 = 1,8^2)$$
$$\bar{X_{36}} \sim N(\mu = 13,5, \sigma^2 = 1,8^2/36)$$

(a) $P(X > 15) = P(Z > \frac{15-13,5}{1,8}) = P(Z > 0,833) = 0,202$

(b) $P(\bar{X} > 15) = P(Z > \frac{15-13,5}{1,8/\sqrt{36}}) = P(Z > 5) = 3e - 07$

(c) $P(13 < \bar{X} < 14) = P(\frac{13-13,5}{1,8/\sqrt{36}} < Z < \frac{14-13,5}{1,8/\sqrt{36}}) = P(-1,667 < Z < 1,667) = 0,904$

---

93. Os números abaixo mostram as notas de um grupo de alunos em duas avaliações

Aluno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Prova 1	35	39	50	47	33	17	17	80	23	51	2	21	20	12	81	98	47	34
Prova 2	65	63	80	72	65	35	62	72	50	60	32	59	40	68	79	85	80	55

(a) Calcule média, variância e coeficiente de variação das notas em cada avaliação

(b) Calcule mediana, quartis, amplitude e amplitude interquartílica de cada avaliação

(c) Faça um diagrama *box-plot* para comparar as notas das duas avaliações

(d) Com as notas das duas provas juntas faça um único diagrama ramo-e-folhas sublinhando as notas da segunda prova.

(e) Usando as medidas e gráficos acima compare o rendimento dos alunos nas duas provas.

(f) Existe relação (associação) entre os resultados das duas provas? Faça um gráfico e calcule alguma(s) medida(s) estatística(s) para verificar se há associação.

(g) Suponha agora que as provas possam ser consideradas uma amostra aleatória. Faça um testes estatísticos adequado para verificar:

- Se a média da Prova 1 está significativamente abaixo de 50.
- Se houve aumento (significativo) de rendimento da primeira para segunda prova.

Solução:

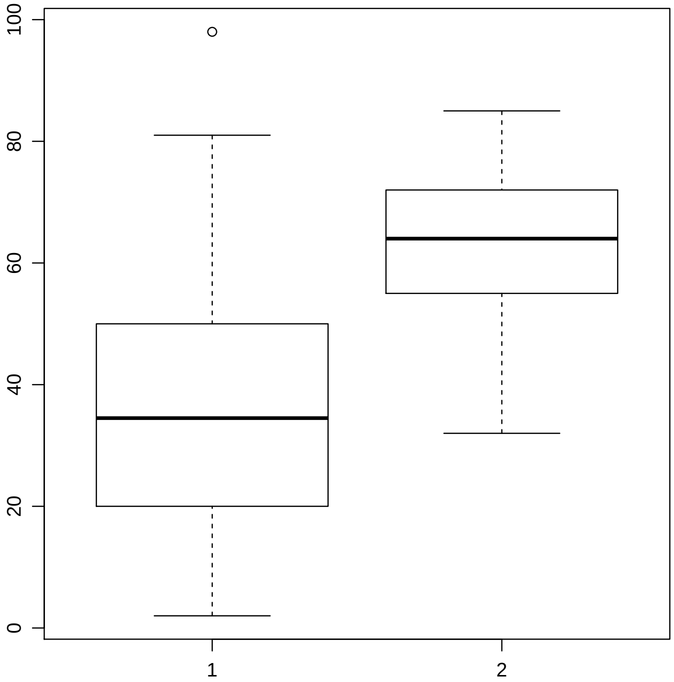
(a)

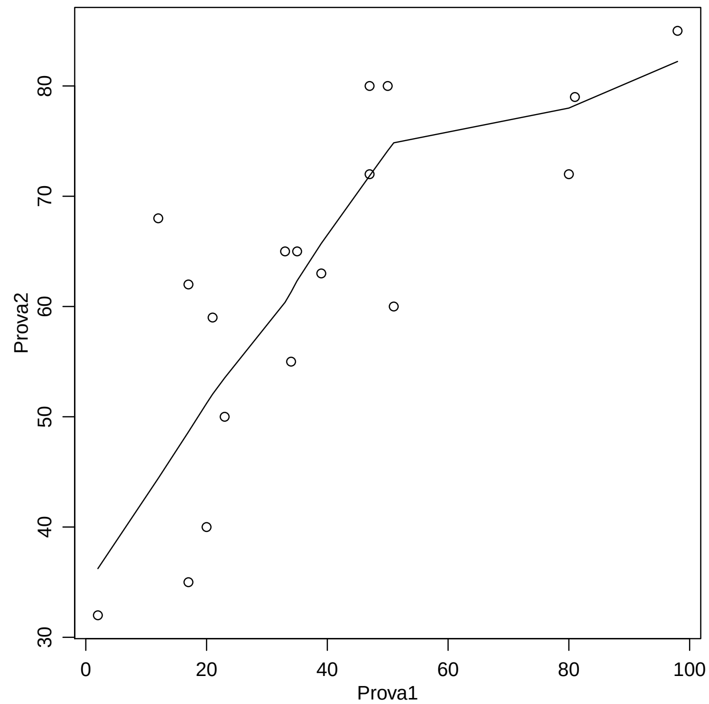
$$\bar{x}_1 = 39,28 \quad s_1^2 = 670,68 \quad CV_1 = 65,93\%$$
$$\bar{x}_2 = 62,33 \quad s_2^2 = 237,53 \quad CV_2 = 24,73\%$$

(b)

$$md_1 = 34,5 \quad Q1_1 = 20 \quad Q3_1 = 50 \quad A_1 = 96 \quad AI_1 = 30$$
$$md_2 = 64 \quad Q1_2 = 55 \quad Q3_2 = 72 \quad A_2 = 53 \quad AI_2 = 17$$

(c)





(d)
```
The decimal point is 1 digit(s) to the right of the |

0 | 2
1 | 277
2 | 013
3 | 234559
4 | 077
5 | 00159
6 | 023558
7 | 229
8 | 00015
9 | 8
```

(e) Comentários sobre: valores centrais, variabilidade, assimetria e dados discrepantes

(f) Coeficientes de correlação: Pearson $r_P = 0,75$ e Spearman $r_S = 0,732$
Comentários: ...

(g) • Teste de Hipótese para uma média

$$X_1 : \text{nota na Prova 1 } \sim Dist(\mu_1, \sigma_1^2)$$
$$H_0 : \mu_1 \geq 50 \text{ vs } H_1 : \mu_1 < 50$$
$$\alpha = 0,05 \longrightarrow t_t = -1,74 \longrightarrow \hat{\mu_1} = \bar{x}_1 < 39,381$$
$$t_c = \frac{39,3 - 50}{25,9/\sqrt{18}} = -1,757$$
$$p - valor = 0,04849$$

Conclusão: Rejeita-se H_0 para $\alpha = 0,05$ e portanto há evidências que a média esteja abaixo de 50.

- Teste de Hipótese para comparação de médias pareadas

$$X_1 : \text{nota na Prova 1} \;\sim Dist(\mu_1, \sigma_1^2)$$
$$X_2 : \text{nota na Prova 2} \;\sim Dist(\mu_2, \sigma_2^2)$$
$$D = X_2 - X_1 : \text{diferença de notas} \;\sim Dist(\mu_D = \mu_2 - \mu_1, \sigma_D^2)$$
$$H_0 : \mu_2 \leq \mu_1 \longrightarrow \mu_D \leq 0 \text{ vs } H_1 : \mu_2 > \mu_1 \longrightarrow \mu_D > 0$$
$$\alpha = 0,05 \longrightarrow t_t = 1,74 \longrightarrow \hat{\mu_D} = \bar{x}_D < 7,214$$
$$\text{Amostra :}$$
$$\hat{\mu_D} = \bar{x}_D = x_2 \bar{-} x_1 = 23,056 \quad \hat{s}_D = 17,595$$
$$t_c = \frac{23,1 - 0}{17,6/\sqrt{18}} = 5,559$$
$$p - valor = 0,99998$$

Conclusão: Rejeita-se H_0 para $\alpha = 0,05$ e portanto há evidências que houve aumento de rendimento.

---

94. Uma lavoura é monitorada inspecionando periodicamente 100 plantas para verificar a necessidade de aplicação de inseticida. A aplicação se dá quando a proporção ultrapassa 0,25 (25%). Suponha que uma lavoura esteja em um certo momento com 20% das plantas infectadas.

    (a) qual a probabilidade de que seja decidido aplicar o inseticida nesta lavoura?

    (b) qual seria o valor desta probabilidade caso a inspeção fosse feita em 200 plantas?

    (c) quantas plantas deveriam ser inspecionadas para que esta probabilidade não ultrapassasse 0,02?

    (d) suponha agora que a lavoura está com 32% das plantas infectadas. Qual a probabilidade de se deixar de aplicar o inseticida?

Solução:

$$X : \text{número de plantas infectadas}$$
$$X \sim B(n, p)$$
$$\hat{p} \approx \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$

(a) $n = 100, p = 0,25, \hat{p} \approx N(0,25; 0,0019)$
$P[\hat{p} > 0,25] = 0.50$

(b) $n = 200, p = 0,25, \hat{p} \approx N(0,25; 9e-04)$
$P[\hat{p} > 0,25] = 0.50$

(c) será sempre 0.5 para qualquer n

(d) $n = 100, p = 0,32, \hat{p} \approx N(0,32; 0,0022)$
$P[\hat{p} < 0,25|p = 0,32] = 0,0667$

---

95. O consumo de combustível é uma variável aleatória com parâmetros dependendo do tipo de veículo. Suponha que, para um certo automóvel, o desvio padrão do consumo seja conhecido e igual a 2 km/l. Porém, precisamos de informações sobre o consumo médio. Para tal, coletamos uma amostra de 40 automóveis desse modelo e observamos seu consumo.

    (a) Quem seria o estimador para o consumo médio de todos automóveis desse tipo?

    (b) Se a amostra forneceu um consumo médio de 9,3 km/l, construa um intervalo de confiança (95%) para média do consumo desses carros.

    (c) Qual é a confiança se a amplitude de um intervalo de confiança, construído a partir dessa amostra é de 1,5?

Solução:

(a) A média amostral $\bar{X} \sim N(\mu; 2^2/40)$

(b)
```
> 9.3 + qnorm(c(0.025, 0.975)) * 2/sqrt(40)
[1] 8,68 9,92
```

(c)
```
> 1 - 2*pnorm(sqrt(40)*(1.5/2)/2, lower=FALSE)
[1] 0,9823
```

---

96. Uma fibra sintética possui resistência à tensão com distribuição normal de média 75,5 psi e desvio padrão 3,5 psi.

(a) encontre a probabilidade de que uma amostra aleatória de tamanho $n = 6$ fibras contenha uma média que exceda 75,75 psi.

(b) qual seria o erro padrão da média amostral caso o tamanho da amostra passasse de $n = 6$ para $n = 49$?

Solução:

$$X : \text{tensão da fibra; } X \sim N(75,5\ ;\ 3,5^2)$$
$$\overline{X} \sim N(75,5\ ;\ 3,5^2/n)$$

(a) $P[\bar{X}_6 > 75,75] = P[Z > \frac{75,75-75,5}{3,5/\sqrt{6}}] = 0,431$

(b) $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 3.5/\sqrt{49} = 0,5$

---

97. Para se ajustar uma máquina, a correia deve ter entre 60 e 62 cm de comprimento. Tendo em vista o processo de fabricação, o comprimento destas correia pode ser considerado como uma variável aleatória com distribuição normal, de média 60,7 e desvio padrão 0,8 cm. Pergunta-se:

(a) qual a probabilidade de uma correia, escolhida ao acaso, poder ser usada na máquina?

(b) um grande revendedor dessas correias estabelece um controle de qualidade nos lotes que compra da fábrica: ele sorteia 4 correias do lote e só aceita o lote se o comprimento médio estiver dentro do tamanho aceito pela máquina. Calcule a probabilidade de aceitação do lote.

Solução:

$$X : \text{comprimento da correia}$$
$$X \sim \mathrm{N}(60,7; 0.8^2)$$
$$\bar{X}_4 \sim \mathrm{N}(60,7; 0.8^2/4)$$

(a) $P[\text{correia poder ser usada}] = P[60 < X < 62] = P[\frac{60-60,7}{0.8} < Z < \frac{62-60.7}{0.8}] = 0,757$

(b) $P[\text{lote ser aceito}] = P[60 < \bar{X}_4 < 62] = P[\frac{60-60,7}{0.8/\sqrt{4}} < Z < \frac{62-60.7}{0.8/\sqrt{4}}] = 0,959$

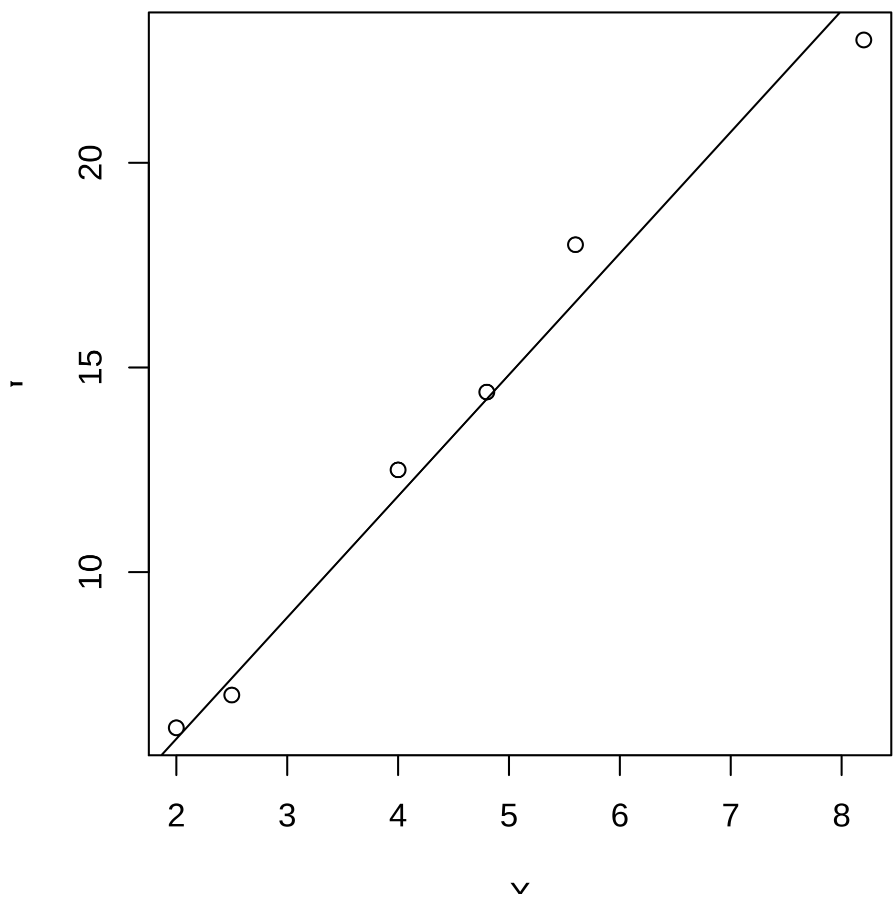
---

98. Suponha que os dados a seguir foram obtidos em um estudo e deseja-se explicar o comportamento da variável Y pelos valores de uma outra variavel X pelo modelo $Y = \theta X + \epsilon$. Encontre a expressão do estimador de mínimos quadrados de θ e obtenha a estimativa para este conjunto de dados.

```
X    2,0      2,5      4,0      4,8     5,6      8,2
----------------------------------------------------
Y    6,2      7,0     12,5     14,4    18,0     23,0
```

Solução:

$$\begin{aligned}
y_i &= \theta x_i + \epsilon \\
Q(\theta) &= \sum (y - \theta x_i)^2 \\
\frac{\delta Q(\theta)}{\delta \theta} &= -2 \sum x_i (y_i - \theta x_i) \\
\frac{\delta Q(\theta)}{\delta \theta} &= 0 \longrightarrow \hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{438,42}{147,89} = 2,96
\end{aligned}$$



---

99. **Pagar por serviços da Web?**
Uma pesquisa foi feita com 13.000 usuários de certo serviço da internet que foram perguntados sobre a sua disposição em pagar para acessar estes serviços. Destes, 2938 se declararam *completamente contrários* a pagar pelos serviços. Assume-se que os 13.000 usuários foram selecionados aleatoriamente.

   - Construa um intervalo de confiança de 95% para a proporção de usuários dispostos a pagar pelo acesso
   - Verifique a largura (amplitude) do intervalo calculado anteriormente. Para muitas aplicações o intervalo é desnecessáriamente estreito. Neste caso o que voce sugere para o tamanho da amostra?
   - qual seria a confiança com o tamanho de amostra original de um intervalo com o dobro da largura?
   - Qual deveria ser o tamanho da amostra necessário para estimar a proporção de dispostos a pagar com 2% de margem de erro com 95% de confiança?

100. No contexto do problema anterior suponha que os prestadores deste serviço decidem cobrar por eles se tiverem evidência suficiente de que ao menos 75% da população de usuários não são completamente aversos ao pagamento. Estruture o problema na forma de um teste de hipótese de hipóteses indicando:

   (a) quais seriam as interpretação do o ERRO I e o ERRO II neste contexto?

(b) formule as hipóteses de interesse a as transcreve para notação adequada no problema.

(c) defina as condições e suposições necessária e proceda o teste de hipótese, indicando qual deveria ser a decisão dos prestadores do serviço.

Solução:

101. Uma certa requisição computacional é processada passando em diferentes nós de uma rede. Suponha que em cada nó o tempo de processamento/operação é de 35 unidades com desvio padrão de 5 unidades.

(a) qual a probabilidade do tempo de processamento ser superior a 42 unidades em um determinado nó?

(b) qual a probabilidade do tempo total de processamento ultrapassar 175 unidades quando a requisição passa por 4 nós da rede?

(c) quantos nós deveria ter a rede para que o tempo médio de processamento estivesse, com probabilidade de 0.90, no intervalo $[34, 36]$ unidades?

(d) qual a probabilidade de que, passando por seis nós, o tempo de processamento seja inferior a 208 unidades?

Solução:

---

102. O consumo de combustível é uma variável aleatória com parâmetros dependendo do tipo de veículo. Suponha que, para um certo automóvel, o desvio padrão do consumo seja conhecido e igual a 2,5 km/l. Porém, precisamos de informações sobre o consumo médio. Para tal, coletamos uma amostra de 40 automóveis desse modelo e observamos seu consumo.

(a) Quem seria o estimador para o consumo médio de todos automóveis desse tipo? Qual(ais) a rezão para usar este estimador?

(b) Se a amostra forneceu um consumo médio de 8,5 km/l, construa um intervalo de confiança (95%) para média do consumo desses carros.

(c) Qual é a confiança de um intervalo de confiança de *amplitude* 1,2?

Solução:

---

103. Estuda-se a resistência de vigas de madeira utilizadas na construção. O fornecedor atesta que, em média, cada viga resiste a 3 toneladas, com desvio padrão de aproximadamente 2 toneladas. Vinte destas vigas serão tomadas ao acaso para serem utilizadas em uma obra. Considerando que é verdadeira a afirmação do fornecedor e assumindo-se o modelo normal para o peso das vigas, pergunta-se:

(a) qual a probabilidade de uma destas vigas suportar menos que 1 tonelada?

(b) qual a probabilidade de as 20 vigas suportarem, em média, pelo menos 2,5 toneladas?

(c) qual a probabilidade no item anterior considerando 40 vigas?

(d) quantas vigas seriam necessárias para que a probabilidade do item anterior fosse de 0,99 ?

Solução:

---

104. Uma população normal tem média 100 e variância 25. Qual deve ser o tamanho de uma amostra aleatória para que o erro padrão da média amostral seja 1,5?

Solução:

$$\begin{aligned}
X &\sim N(\mu_x = 100, \sigma_x^2 = 25) \\
\bar{X} &\sim N(\mu_{\bar{x}} = 100, \sigma_{\bar{x}}^2 = 25/n) \\
\sigma_{\bar{x}}^2 &= 25/n = 1,5 \\
n &= 25/1,5 = 17
\end{aligned}$$

---

105. Em um estudo foram selecionados ao acaso 1000 casos de câncer de pulmão dos quais 740 resultaram em morte dentro de um período pré-especificado.

   (a) Obtenha um intervalo de confiança (95%) para taxa de óbitos.

   (b) Qual deveria ser o tamanho da amostra para que a margem de erro no intervalo anterior fosse de, no máximo, 0,03 ?

Solução:

$$\begin{aligned}
X &: \text{óbito} \\
X &\sim B(p) \\
\hat{p} &\sim N(p, p(1-p)/n)
\end{aligned}$$

```
> p.est <- 740/1000
> n <- 1000
> ## a)
> round(prop.test(740, 1000, conf=0.95)$conf.int, dig=4)

[1] 0,7114 0,7667
attr(,"conf.level")
[1] 0,95

> round(prop.test(740, 1000, conf=0.95, corr=F)$conf.int, dig=4)

[1] 0,7119 0,7662
attr(,"conf.level")
[1] 0,95

> round(p.est + qnorm(c(0.025, 0.975)) * sqrt(p.est * (1 - p.est)/n), dig=4)

[1] 0,7128 0,7672

> round(p.est + qnorm(c(0.025, 0.975)) * sqrt(0.5 * (1 - 0.5)/n), dig=4)

[1] 0,709 0,771

> ## b)
> ceiling(qnorm(0.975)^2 * 0.74 * (1 - 0.74)/(0.03)^2)

[1] 822

> ceiling(qnorm(0.975)^2 * 0.5 * (1 - 0.5)/(0.03)^2)

[1] 1068
```

---

106. Certo material deve ter ao menos 12 unidades de um certo elemento em sua composição, e com um desvio padrão de 0,5 unidades. Um laboratório deseja testar a hipótese de que um lote está dentro da especificação contra a de que está fora (ou seja com média inferior a 12) usando uma amostra aleatória de quatro unidades.

    (a) qual a probabilidade do erro tipo I se a região crítica é definida como sendo $\bar{x} < 11,5$?
    (b) Qual é a probabilidade do erro tipo II se o valor verdadeiro do lote for 11.25 ?
    (c) qual deveria ser o tamanho da amostra para que com a mesma região crítica, a probabilidade do erro tipo I não superasse 0,05 ?
    (d) e qual seria o erro tipo II conforme definido no item (b) para este tamanho de amostra?

Solução:

Supondo distribuição normal:

$$X \sim N(\mu = 12, \sigma^2 = 0,5^2)$$
$$H_0 = \mu \geq 12$$
$$H_1 = \mu < 12$$
$$n = 4$$
$$\bar{X} \sim N(\mu = 12, \sigma^2 = 0,5^2/4)$$

(a)
$$P[\text{Erro I}] = P[\bar{X} < 11,5|\mu = 12] = 0,023$$

(b)
$$P[\text{Erro II}] = P[\bar{X} > 11,5|\mu = 11,25] = 0,159$$

(c)
$$z = \frac{11,5 - 12}{0,5/\sqrt{n}} = -1,645 \longrightarrow n = 3$$

(d)
$$\bar{X} \sim N(12; 0,5/3) \longrightarrow P[\text{Erro II}] = P[\bar{X} > 11,5|\mu = 11.25] = 0,193$$

---

107. O índice de produtividade diário de um processo está sendo estudado. De experiências anteriores sabe-se que este índice varia com desvio padrão de 3 unidades. Nos últimos cinco dias os índices foram:
    `91,6   88,75    90,95   89,95   91,3`
    Usando $\alpha = 0,05$, use procedimentos estatísticos adequados para responder:

    (a) há evidência de que o índice de produtividade difere de 90,0 ?
    (b) qual o valor-p para o teste do item anterior?
    (c) qual é a probabilidade do erro tipo II se a verdadeira média é de 92,0 ?

Solução:

```
> dat <- c(91.6, 88.75, 90.95, 89.95, 91.3)
> (zc <- (mean(dat) - 90)/(3/sqrt(5)))

[1] 0,3801

> ## a)
> (ifelse(abs(zc) > qnorm(0.975), "Rejeita H_0, difere de 90",
+                          "Nao rejeita H_0, nao difere significativamente de 90"))
```

```
[1] "Nao rejeita H_0, nao difere significativamente de 90"

> ## b)
> round(2 * pnorm(abs(zc), low=F), dig=4)

[1] 0,7038

> ## c)
> (crit <- 90 + qnorm(c(0.025, 0.975)) * 3/sqrt(5))

[1] 87,37 92,63

> diff(pnorm(crit, mean=92, sd=3/sqrt(5)))

[1] 0,6803
```

---

108. Uma amostra de uma população com distribuição de Poisson de parâmetro λ forneceu os seguinte valores:
`15 18 12 10 11 13 11 10 12 9 10 13 12 11`
Encontre a expressão do estimador de máxima verossimilhança do parâmetro e obtenha o valor da estimativa do parâmetro.

Solução:

$$\begin{aligned}
X &\sim \mathrm{P}(\lambda)\\
P[X=x_i] &= \frac{\mathrm{e}^{-\lambda}\lambda^{x_i}}{x_i!}\\
l(\lambda) = \log\{L(\lambda)\} &= \log\{\prod_{i=1}^{n} P[X=x_i]\}\\
&= \sum_{i=1}^{n}\log\{P[X=x_i]\} = \sum_{i=1}^{n}\log\{\frac{\mathrm{e}^{-\lambda}\lambda^{x_i}}{x_i!}\}\\
&= \sum_{i=1}^{n}[-\lambda + x_i\log(\lambda) - \log(x_i!)] = -n\lambda + \log(\lambda)\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\log(x_i!)]\frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} \quad = -n + \frac{\sum_{i=1}^{n}x_i}{\hat{\lambda}} = 0\\
\hat{\lambda} &= \frac{\sum_{i=1}^{n}x_i}{n} = \overline{x}
\end{aligned}$$

Para a amostra dada a estimativa é:

$$\hat{\lambda} = \bar{X} \longrightarrow \hat{\lambda} = 11,93$$

---

109. O tempo de processamento de certo tipo de requisição tem distribuição normal com média de 10 segundos e variância de 4 segundos. Sob as suposições convenientes, qual a probabilidade de que uma sequência de 9 requisições tenha duração superior a 91 segundos?

Solução:

$$\begin{aligned}
X &\sim N(\mu_x = 10, \sigma_x^2 = 4)\\
n &= 9\\
\bar{X} &\sim N(\mu_{\bar{x}} = 10, \sigma_{\bar{x}}^2 = 4/9)\\
P[\bar{X} > 91/9] = P[Z &> \frac{(91/9) - 10}{\sqrt{4/9}}] = 0,4338
\end{aligned}$$

---

110. Uma amostra de uma população forneceu os seguinte valores:
`216 237 249 204 225 301 281 263 318 255 275 295.`
Faça as suposições necessárias e obtenha:

(a) intervalos de confiança para o valor médio de 90 e 95%
(b) intervalos de confiança para a variância 90 e 95%

Solução:

```
> am <- c(216, 237, 249, 204, 225, 301, 281, 263, 318, 255, 275, 295)
> n <- length(am)
> t.test(am, conf=0.90)$conf.int

[1] 241,5 278,4
attr(,"conf.level")
[1] 0,9

> t.test(am, conf=0.95)$conf.int

[1] 237,3 282,5
attr(,"conf.level")
[1] 0,95

> ((n-1) * var(am))/qchisq(c(0.95, 0.05), df=n-1)

[1]  707,3 3042,1

> ((n-1) * var(am))/qchisq(c(0.975, 0.025), df=n-1)

[1]  634,9 3647,2
```

---

111. Em um estudo foram examinados 1000 processos dos quais 740 resultaram em algum tipo de erro.

(a) Obtenha um intervalo de confiança (95%) para taxa de erros.
(b) Qual deveria ser o tamanho da amostra para que a margem de erro no intervalo anterior fosse de, no máximo, 0,03 ?

Solução:

$$X : \text{erro}$$
$$X \sim B(p)$$
$$\hat{p} \sim N(p, p(1-p)/n)$$

```
> p.est <- 740/1000
> n <- 1000
> ## a)
> round(prop.test(740, 1000, conf=0.95)$conf.int, dig=4)

[1] 0,7114 0,7667
attr(,"conf.level")
[1] 0,95
```

```
> round(prop.test(740, 1000, conf=0.95, corr=F)$conf.int, dig=4)

[1] 0,7119 0,7662
attr(,"conf.level")
[1] 0,95

> round(p.est + qnorm(c(0.025, 0.975)) * sqrt(p.est * (1 - p.est)/n), dig=4)

[1] 0,7128 0,7672

> round(p.est + qnorm(c(0.025, 0.975)) * sqrt(0.5 * (1 - 0.5)/n), dig=4)

[1] 0,709 0,771

> ## b)
> ceiling(qnorm(0.975)^2 * 0.5 * (1 - 0.5)/(0.03)^2)

[1] 1068
```

---

112. Certo material deve ter ao menos 12 unidades de um certo elemento em sua composição, e com um desvio padrão de 0,5 unidades. Um laboratório deseja testar a hipótese de que um lote está dentro da especificação contra a de que está fora (ou seja com média inferior a 12) usando uma amostra aleatória de quatro unidades.

   (a) qual a probabilidade do erro tipo I se a região crítica é definida como sendo $\bar{x} < 11,5$?
   (b) Qual é a probabilidade do erro tipo II se o valor verdadeiro do lote for 11.25 ?
   (c) qual deveria ser o tamanho da amostra para que com a mesma região crítica, a probabilidade do erro tipo I não superasse 0,05 ?
   (d) e qual seria o erro tipo II conforme definido no item (b) para este tamanho de amostra?

Solução:

Supondo distribuição normal:

$$X \sim N(\mu = 12, \sigma^2 = 0,5^2)$$
$$H_0 = \mu \geq 12$$
$$H_1 = \mu < 12$$
$$n = 4$$
$$\bar{X} \sim N(\mu = 12, \sigma^2 = 0,5^2/4)$$

(a)

$$P[\text{Erro I}] = P[\bar{X} < 11,5|\mu = 12] = 0,023$$

(b)

$$P[\text{Erro II}] = P[\bar{X} > 11,5|\mu = 11,25] = 0,159$$

(c)

$$z = \frac{11,5 - 12}{0,5/\sqrt{n}} = -1,645 \longrightarrow n = 3$$

(d)

$$\bar{X} \sim N(12; 0,5/3) \longrightarrow P[\text{Erro II}] = P[\bar{X} > 11,5|\mu = 11.25] = 0,193$$

---

113. O índice de produtividade diário de um processo está sendo estudado. De experiências anteriores sabe-se que este índice varia com desvio padrão de 3 unidades. Nos últimos cinco dias os índices foram:
`91,6   88,75    90,95   89,95   91,3`
Usando $\alpha = 0,05$, use procedimentos estatísticos adequados para responder:

(a) há evidência de que o índice de produtividade difere de 90,0 ?
(b) qual o valor-p para o teste do item anterior?
(c) qual é a probabilidade do erro tipo II se a verdadeira média é de 92,0 ?

Solução:

```
> dat <- c(91.6, 88.75, 90.95, 89.95, 91.3)
> (zc <- (mean(dat) - 90)/(3/sqrt(5)))

[1] 0,3801

> ## a)
> (ifelse(abs(zc) > qnorm(0.975), "Rejeita H_0, difere de 90",
+                          "Nao rejeita H_0, nao difere significativamente de 90"))

[1] "Nao rejeita H_0, nao difere significativamente de 90"

> ## b)
> round(2 * pnorm(abs(zc), low=F), dig=4)

[1] 0,7038

> ## c)
> (crit <- 90 + qnorm(c(0.025, 0.975)) * 3/sqrt(5))

[1] 87,37 92,63

> diff(pnorm(crit, mean=92, sd=3/sqrt(5)))

[1] 0,6803
```

---

114. (B & M, 2002) A precipitação pluviométrica anual numa numa certa região tem desvio padrão $\sigma = 3,1$ e média desconhecida. Para os últimos nove anos, foram obtidos os seguintes resultados:
`30,5   34,1   27,9   35,0   26,9   30,2   28,3   31,7   25,8.`

(a) Construa um teste de hipótese para saber se a média da precipitação pluviométrica anual é maior que 30,0 unidades.
(b) Discuta o mesmo problema porém considerando σ desconhecido.
(c) Supondo que, na realidade, $\mu = 33,0$ qual a probabilidade de tirarmos uma conclusão errada?

Solução:

```
> prec <- c(30.5, 34.1, 27.9, 35.0, 26.9, 30.2, 28.3, 31.7, 25.8)
> n <- length(prec)
> (m.prec <- mean(prec))

[1] 30,04

> (dp.prec <- sd(prec))

[1] 3,153
```

(a)
```
> (p.valor <- 1 - pnorm(m.prec, mean=30, sd=3.1/sqrt(n)))
[1] 0,4828
```

(b)
```
> t.test(prec, conf=0.95, alt="greater", mu=30)
One Sample t-test

data:  prec
t = 0,042, df = 8, p-value = 0,5
alternative hypothesis: true mean is greater than 30
95 percent confidence interval:
 28,09   Inf
sample estimates:
mean of x
    30,04
```

(c) $P[\bar{X} \notin RC | \mu = 33] = P[\bar{X} < \bar{X}_c | \mu = 33\ ]$:
```
> (tc <- qt(0.95, df=n-1))
[1] 1,86
> (xbc <- 30 + tc * (dp.prec/sqrt(n)))
[1] 31,95
> pt((xbc - 33)/(dp.prec/sqrt(n)), df=n-1)
[1] 0,1744
```

---

115. Sabe-se que o peso de um certo tipo de peça tem desvio padrão de 5 cm. Em uma amostra de 36 peças observou-se que a média era de 150 cm.

(a) Monte um intervalo de confiança (95%) para a média populacional.
(b) Que tamanho deveria ter a amostra para que o intervalo (149,02 ; 150,98) tenha 95% de confiança?

Solução:

$$X : \text{peso da peça}$$
$$X \sim N(\mu, 5^5)$$
$$\text{amostra} : n = 36 \ \ ; \ \ \bar{x} = 150$$
$$\bar{X} \sim N(\mu, 5^2/36)$$

(a) (148,37 ; 151,63)
(b) $z_{\alpha/2} \frac{5}{\sqrt{n}} = \frac{150{,}98 - 149{,}02}{2} \longrightarrow n = 100$

116. Foi feita uma entrevista com uma amostra aleatória de 500 pessoas que foram perguntadas se são a favor do impedimento (*impeachment*) de um ocupante de cargo público acusado de conduta indevida. Entre os entrevistados 265 manifestaram-se a favor do empedimento.

(a) Monte intervalos de confiança a 90, 95 e 99% para o percentual de favoráveis.

(b) Qual deveria ser o tamanho da amostra para que a margem de erro fosse de 1,5% com 95% de confiança?

(c) Com os dados da pesquisa, poderia-se afirmar, com um nível de significância de 5%, que o impedimento é apoiado pela maioria da população?

Solução:

$$\begin{aligned} X &: \text{número de favoráveis} \\ X &\sim B(n = 500, p) \\ \text{amostra} : n = 500 \;\; ; \;\; \hat{p} &= \frac{265}{500} = 0,53 \\ \hat{p} &\sim N(p, p(1-p)/n) \end{aligned}$$

(a)
```
> p.est <- 265/500
> p.est + qnorm(c(0.05, 0.95)) * sqrt(p.est * (1-p.est)/500)
[1] 0,4933 0,5667
> p.est + qnorm(c(0.025, 0.975)) * sqrt(p.est * (1-p.est)/500)
[1] 0,4863 0,5737
> p.est + qnorm(c(0.005, 0.9995)) * sqrt(p.est * (1-p.est)/500)
[1] 0,4725 0,6034
```

(b) $z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.015 \longrightarrow n = 4253$

(c)

$$\begin{aligned} \mathrm{H}_0 : p \leq 0,5 \quad &versus \quad \mathrm{H}_1 : p > 0,5 \\ \alpha &= 0.05 \\ z_c &= \frac{\hat{p} - 0,5}{\sqrt{0,5(1-0,5)/n}} = 1,3416 \\ p_{valor} &= 0,0899 \end{aligned}$$

Conclusão: não rejeita-se H_0 ao nível de significância de 5%, portanto não há evidências suficientes para se afirmar que a maioria seja favorável ao impedimento.

---

117. O nível de confiança dos consumidores em um produto era de 65%. Houve um problema envolvendo o produto e decidiu-se fazer uma pesquisa para investigar se tal problema reduziu este nível de confiança. Para isto foi feita uma pesquisa com uma amostra de 1200 consumidores dos quais 710 afirmaram ainda confiar no produto. Teste a hipótese de que houve uma redução no nível de confiança do produto.

Solução:

$$\begin{aligned} \mathrm{H}_0 : p \geq 0,65 \quad &versus \quad \mathrm{H}_1 : p \leq 0,65 \\ \alpha &= 0.05 \\ z_c &= \frac{\hat{p} - 0,65}{\sqrt{0,65(1-0,65)/n}} = -4,1111 \\ p_{valor} &= 2e - 05 \end{aligned}$$

Conclusão: rejeita-se H_0 ao nível de significância de 5%, portanto não há evidências suficientes para se afirmar que diminuiu o nível de confiança no produto entre os consumidores.

---

118. Defina e ilustre com um exemplo/situação hipotética os seguintes conceitos:

   (a) parâmetro
   (b) estimador
   (c) estimativa
   (d) distribuição amostral
   (e) intervalo de confiança
   (f) nível de significância

   Solução:

---

119. A distribuição dos comprimentos dos elos de uma corrente de bicicleta é normal com média $2cm$ e variância igual a $0.01cm^2$. Para que uma corrente se ajusta a bicicleta deve ter comprimento total entre 58 e $61cm$.

   (a) Qual a probabilidade de um elo escolhido ao acaso ser maior que $2,15cm$?
   (b) Qual a probabilidade de uma corrente com 30 elos não se ajustar à bicicleta?
   (c) E uma corrente de 29 elos?

   Solução:

$$\begin{aligned}
&X : \text{comprimento dos elos} \\
&X \sim N(\mu = 2, \sigma^2 = 0.01) \\
&\bar{X} \sim N(\mu_{\bar{x}} = 2, \sigma^2_{\bar{x}} = 0.01/n)
\end{aligned}$$

   (a) $P[X > 2.15] = 0,0668$
   (b) $P[\sum_{i=1}^{n=30} < 58 \text{ ou } \sum_{i=1}^{n=30} > 61] = P[\bar{X} < \frac{58}{30} \text{ ou } \bar{X} > \frac{61}{30}] = 0,0341$
   (c) $P[\sum_{i=1}^{n=29} < 58 \text{ ou } \sum_{i=1}^{n=29} > 61] = P[\bar{X} < \frac{58}{29} \text{ ou } \bar{X} > \frac{61}{29}] = 0,5$

---

120. A duração do "tonner"da uma impressora pode ser modelada pela distribuição normal com média 10.000 cópias e desvio padrão de 1.200 cópias. Em uma amostra de 12 impressoras a duração do "tonner"será anotada e pergunta-se a probabilidade da média destes ser:

   (a) inferior a 9.000 cópias;
   (b) superior a 10.500 cópias;
   (c) entre 9.200 e 10.000 cópias.

   Solução:

$$\begin{aligned}
X &: \text{duração do tonner (em número de cópias)} \\
X &\sim N(\mu = 10.000, \sigma^2 = 1.200^2) \\
\text{amostra} &: n = 12 \\
\bar{X} &\sim N(\mu = 10.000, \sigma^2 = 1.200^2/12)
\end{aligned}$$

(a) $P[\bar{X} < 9.000] = P[Z < \frac{9.000-10.000}{1.200/\sqrt{12}}] = 0,0019$

(b) $P[\bar{X} > 10.500] = P[Z < \frac{10.500-10.000}{1200/\sqrt{12}}] = 0,0745$

(c) $P[9.200 < \bar{X} < 10.000] = P[\frac{9.200-10.000}{1.200/\sqrt{12}} < Z < \frac{10.000-10.000}{1.200/\sqrt{12}}] = 0,4895$

---

121. Os produtores de um programa de televisão pretendem modificá-lo se for assistido regularmente por menos de um quarto dos possuidores de televisão. Uma pesquisa encomendada a uma empresa especializada mostrou que, de 400 famílias entrevistadas, 80 assistem ao programa regularmente. Com base nos dados, qual deve ser a decisão dos produtores.

Solução:

$$\begin{aligned}
\mathrm{H}_0 : p \geq 0,25 \quad &versus \quad \mathrm{H}_1 : p \leq 0,25 \\
\alpha &= 0.05 \\
z_c &= \frac{\hat{p} - 0,25}{\sqrt{0,25(1-0,25)/n}} = -2,5 \\
p_{valor} &= 0,00621
\end{aligned}$$

Conclusão: rejeita-se H_0 ao nível de significância de 5%, portanto não há evidências suficientes para se afirmar que a audiência está abaixo de um quarto e portanto a modificação é recomendada.

---

122. Defina e identifique na questão anterior: (a) a população, (b) o parâmetro de interesse; (c) o estimador; (d) a estimativa; (e) a distribuição amostral.

---

123. Defina e discuta a importância e uso dos conceitos relacionados a estimadores: (a) não tendenciosidade; (b) eficiência.

---

124. O tempo de processamento (em segundos) de certo tipo de requisição a um servidor tem distribuição aproximadamente normal. Foram anotados os seguintes tempos em uma amostra de oito requisições: 15,2; 16,0; 14,5; 13,1; 15,4; 17,0; 16,2 e 14,5. Obtenha intervalos de confiança (90%) para a média e para a variância dos tempos de processamento.

Solução:
IC para média: $\bar{x} \pm t_{\alpha/2} s/\sqrt{n}$
IC para variância: $\left(\frac{(n-1)s^2}{\chi^2_{0,95}}; \frac{(n-1)s^2}{\chi^2_{0,05}}\right)$

```
> tp <- c(15.2, 16.0, 14.5, 13.1, 15.4, 17.0, 16.2, 14.5)
> xbar <- mean(tp)
> xbar

[1] 15,24

> s2 <- var(tp)
> s2

[1] 1,471

> n <- length(tp)
> t.test(tp, conf.level=0.90)$conf.int
```

```
[1] 14,43 16,05
attr(,"conf.level")
[1] 0,9

> c((n-1)*s2/qchisq(0.95, df=n-1), (n-1)*s2/qchisq(0.05, df=n-1))

[1] 0,7321 4,7518
```

---

125. Suponha agora que o tempo de processamento das requisições tenha distribuição normal com média de 15 seg e variância de 1,5 segundos.

   (a) Qual a probabilidade de que o tempo total de processamento de 10 requisições (que podem ser consideradas uma amostra aleatória) ultrapasse 160 segundos?
   (b) E de fossem 11 requisições?
   (c) Para uma amostra de 10 requisições qual a confiança de um intervalo de amplitude de 0,8 segundos?
   (d) Quantas requisições seriam necessárias para que a confiança do intervalo de amplitude de 0,8 segundos fosse de 99%?

Solução:

$$X : \text{tempo de processamento (em seguundos)}$$
$$X \sim N(\mu = 15, \sigma^2 = 1,5^2)$$
$$\bar{X} \sim N(\mu = 15, \sigma^2 = 1,5^2/n)$$

(a) $P[\sum X_{i=1}^{10} > 160] = P[\bar{X}_{10} > 16] = 0,0175$

(b) $P[\sum X_{i=1}^{11} > 160] = P[\bar{X}_{11} > 160/11] = 0,8426$

(c)

$$0,4 = z\frac{1,5}{\sqrt{10}}$$
$$z = \frac{(0,4)(\sqrt{10})}{1,5} = 0,84$$
$$1 - \alpha = 0,8005$$

(d) Quantas requisições seriam necessárias para que a confiança do intervalo de amplitude de 0,8 segundos fosse de 99%?

$$0,4 = z\frac{1,5}{\sqrt{n}} \quad ; \quad z = 2,58$$
$$n = \frac{(2,58^2)(1,5^2)}{0.4^2} = 94$$

---

126. Os tempos de montagem de um módulo, sem segundos, em uma linha de produção forma anotados em uma amostra de 20 execuções.

200.2	196.7	177.6	184.9	197.6	193.0	158.6	223.3	194.6	198.4
178.5	191.4	174.4	208.4	187.2	184.7	177.1	212.5	198.2	182.7

   (a) Monte um diagrama ramo-e-folhas dos dados.

(b) Monte um gráfico ”box-plot”com os dados obtidos.

(c) Calcule o tempo médio, o desvio padrão e o coeficiente de variação

(d) Monte intervalos de confianças a 95% e 99% para a média populacional.

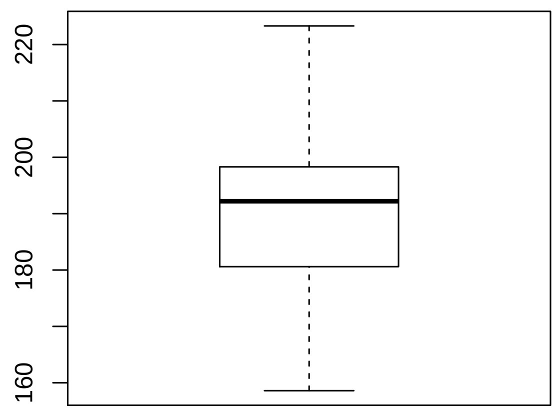
(e) Que tamanho deveria ter a amostra para que o intervalo (186 ; 196) tenha 95% de confiança?

(f) há interesse particular em saber se o tempo médio de montagem está abaixo de 195 segundos em conformidade com especificações de processo. Teste esta hipótese para o nivel de significância de 5%

Solução:

(a)
```
The decimal point is 1 digit(s) to the right of the |

14 | 9
16 | 4789
18 | 35571357888
20 | 083
22 | 3
```

(b)



(c) $\bar{x} = 191$; $S_x = 14,7$; $\text{CV} = 7,7\%$

(d)

$$IC_{90\%} : \bar{x} \pm t_{0,05} \cdot \frac{S}{\sqrt{n}} \quad : \quad 191 \pm 1,729 \cdot \frac{14,7}{\sqrt{20}} \quad : \quad (185,32; 196,68)$$

$$IC_{95\%} : \bar{x} \pm t_{0,025} \cdot \frac{S}{\sqrt{n}} \quad : \quad 191 \pm 2,093 \cdot \frac{14,7}{\sqrt{20}} \quad : \quad (184,12; 197,88)$$

(e)

$$X : \text{tempo de reação}$$
$$X \sim N(\mu, \sigma^2)$$
$$\bar{X} \sim t_{n-1}$$

Aproximando a distribuição t pela normal:

$$z_{\alpha/2}\frac{S}{\sqrt{n}} \approx \frac{196-186}{2}$$
$$z_{\alpha/2}\frac{14,7}{\sqrt{20}} = 5$$
$$n = 34$$

(f)

$$\mathrm{H}_0 : \mu \leq 195 \quad versus \quad \mathrm{H}_1 : \mu > 195$$
$$\alpha = 0.05$$
$$t_c = \frac{\bar{x}-195}{S/\sqrt{n}} = -1,217$$
$$p_{valor} = 0,1193$$

Conclusão: não rejeita-se H_0 ao nível de significância de 5%, portanto não há evidências suficientes para se afirmar que o tempo esteja abaixo de 195.

---

127. Em uma amostra aleatória simples com 200 edifícios com cinco anos, em certa cidade de grande porte, 55% apresentaram problemas estéticos relevantes após a entrega da obra.

(a) construir um intervalo de confiança (95%) para proporção de edifícios da cidade que apresentam problems estéticos nos cinco primeiros anos;

(b) qual a confiança do intervalo $(52\% - -58\%)$?

(c) qual deveria ser o tamanho da amostra para que a confiança do intervalo do item anterior fosse de 90%?

Solução: considerando ambos, IC assintótico e conservador

(a)
```
> pc <- 0.55 ; n = 200
> pc + qnorm(c(0.025, 0.975)) * sqrt(pc*(1-pc)/n)
[1] 0,4811 0,6189
> pc + qnorm(c(0.025, 0.975)) * sqrt(1/(4*n))
[1] 0,4807 0,6193
> prop.test(110, 200)$conf
[1] 0,4783 0,6198
attr(,"conf.level")
[1] 0,95
```

(b)
```
> m.err <- 0.03
> z <- m.err/(sqrt(pc*(1-pc)/n)); 1 - 2*pnorm(z, lower=FALSE)
[1] 0,6062
> z <- m.err/sqrt(1/(4*n)); 1 - 2*pnorm(z, lower=FALSE)
[1] 0,6039
```

(c) qual deveria ser o tamanho da amostra para que a confiança do intervalo do item anterior fosse de 90%?
```
> ceiling(((qnorm(0.95)/0.03)^2) * (pc*(1-pc)))
[1] 745
> ceiling(((qnorm(0.95)/0.03)^2)/4)
[1] 752
```

128. O consumo de combustível é uma variável aleatória com parâmetros dependendo do tipo de veículo. Suponha que, para um certo automóvel, o desvio padrão do consumo seja conhecido e igual a 2 km/l. Porém, precisamos de informações sobre o consumo médio. Para tal, coletamos uma amostra de 40 automóveis desse modelo e observamos seu consumo.

(a) Quem seria o estimador para o consumo médio de todos automóveis desse tipo?

(b) Se a amostra forneceu um consumo médio de 9,3 km/l, construa um intervalo de confiança (95%) para média do consumo desses carros.

(c) Qual é a confiança se a amplitude de um intervalo de confiança, construído a partir dessa amostra é de 1,5?

Solução:

(a) A média amostral $\bar{X} \sim N(\mu; 2^2/40)$

(b)
```
> 9.3 + qnorm(c(0.025, 0.975)) * 2/sqrt(40)
[1] 8,68 9,92
```

(c)
```
> 1 - 2*pnorm(sqrt(40)*(1.5/2)/2, lower=FALSE)
[1] 0,9823
```

---

129. O tempo de permanência de engenheiros recém formados no 1^o emprego, em anos, foi estudado considerando o modelo Normal com média e variâncias desconhecidas. Por analogia com outras categorias profisionais, deseja-se testar se a média é 2 anos contra a alternativa de ser 3 anos. Para uma amostra de 15 engenheiros, a média obtida foi de 2,7 anos e o desvio padrão amostral de 1,4 anos. Qual a conclusão ao nível de 1% ?

Solução:

```
> tc0 <- (2.7 - 2)/(1.4/sqrt(15))
> err1 <- pt(tc0, df=14, lower=F)
> err1

[1] 0,03663

> tca <- (2.7 - 3)/(1.4/sqrt(15))
> err2 <- pt(tca, df=14)
> err2

[1] 0,2103
```

---

130. A experiência mostra que a taxa de complicações, associada a um determinado procedimento cirúrgico, é de 0,20. Com o objetivo de reduzir essa taxa, um pesquisador desenvolveu um novo procedimento e o aplicou a uma amostra de pacientes.

(a) Se ele usar a nova técnica em 100 pacientes, qual deveria ser a taxa limite para que conclua que a nova técnica é melhor que a anterior? (use $\alpha = 0,05$)

(b) Se a verdadeira taxa de complicações associada à nova técnica for 0,08; qual é a probabilidade de que, em uma amostra de tamanho 100, ele não consiga rejeitar a hipótese nula?

Solução:

(a) `> p.crit <- optimize(function(p){(((p-0.2)/sqrt((p*(1-p))/100))-qnorm(0.05))^2}, c(0,1))$min`
`> p.crit`

```
[1] 0,1425
```

(b) `> pnorm(p.crit, m=0.08, sd=sqrt((0.08*0.92)/100), lower=FALSE)`

```
[1] 0,0106
```

---

131. Com o objetivo de avaliar a confiabilidade de um novo sistema de transmissão de dados, torna-se necessário verificar a proporção de *bits* transmitidos com erro em cada lote de 100 Mb. Considere que seja tolerável um erro amostral de no máximo 2% e que em sistemas similares a taxa de erro na transmissão é de 10%. Qual deve ser o tamanho da amostra:
(a) para confiança de 90% ? ; (b) para confiança de 99% ?

Solução:

- `> ceiling(0.1 * 0.9 * qnorm(0.95)^2/(0.02)^2)`

  ```
  [1] 609
  ```

  `> ceiling(0.1 * 0.9 * qnorm(0.995)^2/(0.02)^2)`

  ```
  [1] 1493
  ```

---

132. Sob condições normais, realizaram-se dez observações sobre o tempo de resposta de uma consulta a um certo banco de dados. Os resultados, em segundos, foram:

```
28   35   43   23   62   38   34   27   32   37
```

(a) Construa o intervalo de confiança (95%) para o tempo médio de uma consulta.

(b) Voce afirmaria que o tempo médio é *estatisticamente (significativamente)* inferior a 50 segundos? Fundamente a sua resposta com um procedimento e cálculos adequados.

Solução:

`> dados <- c(28,   35,   43,   23,   62,   38,   34,   27,   32,   37)`

(a) `> t.test(dados)$conf`

```
[1] 28,12 43,68
attr(,"conf.level")
[1] 0,95
```

(b) `> t.test(dados, alt="less", mu=50)`

```
One Sample t-test

data:  dados
t = -4,1, df = 9, p-value = 0,001
alternative hypothesis: true mean is less than 50
95 percent confidence interval:
  -Inf 42,21
sample estimates:
mean of x
     35,9
```

---

133. Para testar se um sistema computacional de classificação binária "inteligente" adquiriu algum conhecimento sobre certo assunto são submetidas 60 informações para classificação das quais 40 são classificadas corretamente. Efetue um procedimento adequado e diga qual a conclusão para o nível de significância de 5%?

Solução:

```
> prop.test(40, 60, alt="greater", p=0.5, conf=0.95)

1-sample proportions test with continuity correction

data:  40 out of 60, null probability 0.5
X-squared = 6, df = 1, p-value = 0,007
alternative hypothesis: true p is greater than 0,5
95 percent confidence interval:
 0,5528 1,0000
sample estimates:
     p
0,6667
```

---

134. Em um certo banco de dados, o tempo para realização de buscas á aproximadamente normal, com média de 53 segundos e desvio padrão de 14 segundos. Depois de atualizações e modificações no sistema, observou-se que, em 30 consultas, o tempo médio foi de 45 segundos e o desvio padrão foi de 10 segundos. Teste as hipóteses ($\alpha = 0.05$) (i) de que a variância dimunuiu e (ii) que a média diminuiu.

Solução:

- ```
  > chisq.c <- (30-1)*10^2/(14^2)
  > p.valor <- pchisq(chisq.c, df=(30-1))
  > p.valor
  [1] 0,01346
  ```
- ```
  > t.c <- (45 - 53)/(10/sqrt(30))
  > p.valor <- pt(t.c, df=(30-1))
  > p.valor
  [1] 7,035e-05
  ```

---

135. A fim de comparar o desempenho de dois grupos de estudantes, foi tomada uma amostra de 10 indivíduos de cada grupo que, submetidos a um teste, obtiveram os escores indicados a seguir.

```
Grupo  I: 73  35  35  40  81  45  48  35  93  50
Grupo II: 75   5  55  50  50  30  75  65  90  50
```

(a) Faça uma análise descritiva utilizando gráficos para resumir os dados e discutir os resultados.

(b) Ainda na análise descritiva, obtenha medidas descritivas (medidas resumo) para os dados e discuta os resultados.

(c) Teste a hipótese de que os grupos possuem a mesma variabilidade.

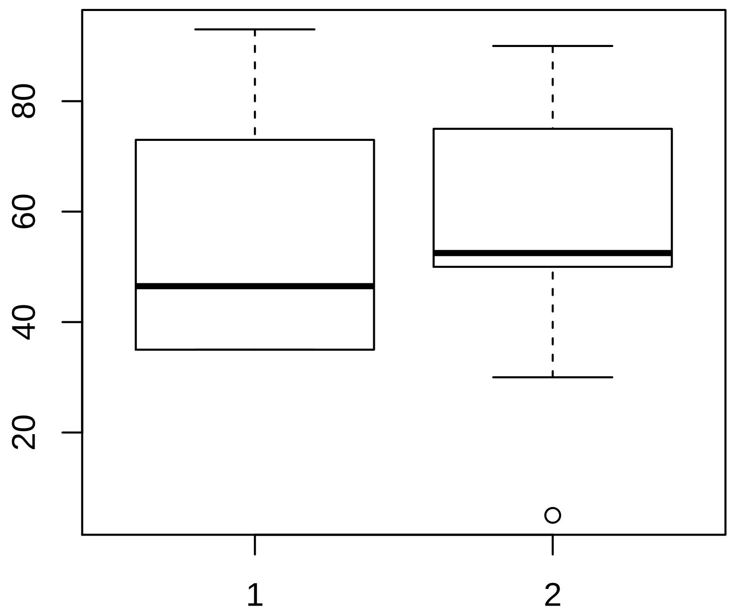
(d) Teste a hipótese da igualdade da média dos grupos.

Solução:

```
> g1 <- c(73,  35,  35,  40,  81,  45,  48,  35,  93,  50)
> g2 <- c(75,   5,  55,  50,  50,  30,  75,  65,  90,  50)
```

(a) Faça uma análise descritiva utilizando gráficos para resumir os dados e discutir os resultados.

```
> boxplot(g1, g2)
```



(b)
```
> res.f <- function(x){
+   res <- c(fivenum(x), mean(x), var(x), sd(x), 100*sd(x)/mean(x))
+   names(res) <- c("Min","Q1","Md","Q3","Max","Média","Variância","Desv.Pad","CV")
+   return(round(res, dig=1))
+ }
> res.f(g1)
```

```
     Min        Q1        Md        Q3       Max     Média Variância  Desv.Pad
    35,0      35,0      46,5      73,0      93,0      53,5     446,7      21,1
      CV
    39,5
```

```
> res.f(g2)
```

```
     Min        Q1        Md        Q3       Max     Média Variância  Desv.Pad
     5,0      50,0      52,5      75,0      90,0      54,5     591,4      24,3
      CV
    44,6
```

(c)
```
> var.test(g1, g2)
```

```
F test to compare two variances

data:  g1 and g2
F = 0,76, num df = 9, denom df = 9, p-value = 0,7
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
 0,1876 3,0411
sample estimates:
ratio of variances
            0,7554
```

(d) `> t.test(g1, g2, var.equal=TRUE)`

```
Two Sample t-test

data:  g1 and g2
t = -0,098, df = 18, p-value = 0,9
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22,41  20,41
sample estimates:
mean of x mean of y
     53,5      54,5
```

---

136. Defina dizendo como devem ser usados e interpretados os seguintes conceitos e/ou métodos estatísticos:

(a) coeficiente de correlação de Spearman
(b) coeficiente de contingência
(c) erro Tipo II
(d) nível de significância
(e) distribuição amostral
(f) valor-p

Solução:

---

137. Um atributo de interesse tem distribuição normal com média 50 e variância desconhecida. Uma amostra aleatória de 16 valores deste atributo foi selecionada da população e apresentou média $\bar{x} = 52$ e desvio padrão $s = 1,5$. A probabilidade da média amostral ser maior ou igual ao resultado obtido reflete o quão atípico é o resultado da amostra, ou seja, quanto menor a probabilidade, mais atípico é considerado o resultado. Desta forma, calcule:

(a) quão atípico é o resultado da amostra obtida?
(b) considerando os mesmos valores de $\bar{x}$ e s, qual deveria ser o tamanho da amostra para que a probabilidade calculada no item anterior fosse de no máximo 0,005 ?

Solução:

$$\begin{aligned} X &: \text{atributo de interesse} \\ X &\sim \text{N}(\mu = 50, \sigma^2) \\ \bar{X} &\sim \text{t}(\nu = n - 1) \\ n = 16 &\quad ; \quad \bar{x} = 52 \; ; \; s = 1,5 \\ \frac{\bar{x - \mu}}{s/\sqrt{n}} &\sim t(v = n - 1) \end{aligned}$$

(a) $P[\bar{X} > 52] = P[T_{\nu=15} > \frac{52-50}{1,5/\sqrt{16}}] = 4e-05$

(b) $t_{15}(0.995) = 2.95 \Rightarrow \frac{\bar{x}-\mu}{s/\sqrt{n}} = 2.95 \Rightarrow n = 10$

---

138. O tempo de vida de um semicondutor laser trabalhando em condições constantes é distribuído segundo uma normal de média 7.000 *horas* e variância de 360.000 *horas*^{2}.

(a) qual a probabilidade de que uma peça dure menos que 5.000 horas?

(b) qual o tempo de duração para o qual 90% das peças superam este valor?

(c) se três lasers são usados em um produto e assumindo-se que a probabilidade falha entre eles é independente, qual a probabilidade de que os três ainda estejam funcionando após 7.300 horas?

(d) um produto contém apenas um dispositivo laser. Ao adquirir o produto, o comprador compra também dois dispositivos extras para troca quando necessário. Qual a probabilidade de que os três dispositivos garantam o funcionamento do produto por pelo menos 23.000 horas?

(e) nas condições do item anterior, qual o número necessário de dispositivos para que a probabilidade do produto funcionar por pelo menos 50.000 horas seja de pelo menos 0.90?

Solução:

$$\begin{aligned} X &: tempodevida \\ X &\sim N(\mu = 7000, \sigma^2 = 360.000) \\ \bar{X} &\sim N(\mu = 7000, \sigma^2 = 360.000/n) \end{aligned}$$

(a) $P[X < 5000] = 0,00043$

(b) $P[P > k] = 0.90 \Rightarrow k = 6231,07$

(c) $P[X_1 > 7.300 \cap X_2 > 7.300 \cap X_3 > 7.300] = P[X_1 > 7.300].P[X_2 > 7.300].P[X_3 > 7.300] = 0,0294$

(d) $P[\sum_{i=1}^{3} X_i > 23.000] = P[\bar{X}_3 > 23.000/3] = 0,0271$

(e) $P[\bar{X} > 50.000/n] = 0.90 \Rightarrow n = 8$

---

139. Para verificar a taxa de resposta a um determinado sinal foram enviados 1.500 sinais para os quais foram obtidas 350 respostas.

(a) obtenha um intervalo de confiança (99%) para a proporção de respostas

(b) quantos sinais deveriam ser enviados para que a margem de erro na estimativa de proporção de resposta seja de 0.5% com 99% de confiança?

Solução:

$$\begin{aligned} X &: respostaaosinal \\ X &\sim Bernoulli(p) \\ \hat{p} &= 350/1500 = 0.233 \end{aligned}$$

Resultado baseado em IC *conservador*

(a) $\hat{p} \pm z_{\alpha/2}\sqrt{1/4n} \Rightarrow (0.2; 0.267)$

(b) $\hat{p} \pm z_{\alpha/2}\sqrt{1/4n} \Rightarrow n = 66349$

Resultado baseado em IC *assintótico*

(a) $\hat{p} \pm z_{\alpha/2}\sqrt{1/4n} \Rightarrow (0.205; 0.261)$

(b) $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \Rightarrow n = 47477$

---

140. Uma fibra sintética possui resistência à tensão com distribuição normal de média 75,5 psi e desvio padrão 3,5 psi.

(a) encontre a probabilidade de que uma amostra aleatória de tamanho $n = 6$ fibras contenha uma média que exceda 75,75 psi.

(b) qual seria o erro padrão da média amostral caso o tamanho da amostra passasse de $n = 6$ para $n = 49$?

Solução:

$$X : \text{tensão da fibra}$$
$$X \sim \mathrm{N}(75,5\ ;\ 3,5^2)$$
$$\overline{X} \sim \mathrm{N}(75,5\ ;\ 3,5^2/n)$$

(a) $P[\bar{X}_6 > 75,75] = P[Z > \frac{75,75-75,5}{3,5/\sqrt{6}}] = 0,431$

(b) $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 3.5/\sqrt{49} = 0,5$

---

141. Tintas para pavimentos de rodovias são fornecidas em duas cores: amarela e branca. Deseja-se estudar o tempo de secagem e em particular suspeita-se que a tinta amarela seca mais rápido do que a branca. Para isto foram tomadas amostras de pinturas em cada cor e mediu-se os tempos de secagem. Teste a hipótese de interesse ao nível de 5% de significância.

Branca:	120, 132, 123, 122, 140, 110, 120, 107
Amarela:	126, 124, 116, 125, 109, 130, 125, 117, 129, 120

Solução: Primeiro vamos testar a igualdade das variâncias.

```
> branca <- c(120, 132, 123, 122, 140, 110, 120, 107)
> amarela <- c(126, 124, 116, 125, 109, 130, 125, 117, 129, 120)
> var.test(branca, amarela)

F test to compare two variances

data:  branca and amarela
F = 2,7, num df = 7, denom df = 9, p-value = 0,2
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0,6379 12,9133
sample estimates:
ratio of variances
             2,677
```

A seguir as médias.

$$H_0 : \mu_B = \mu_A \ \ versus \ \ H_a : \mu_B > \mu_A$$

```
> t.test(branca, amarela, alternative="greater", var.eq=TRUE, conf=0.95)
```

```
Two Sample t-test

data:  branca and amarela
t = -0,086, df = 16, p-value = 0,5
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -7,481    Inf
sample estimates:
mean of x mean of y
    121,8     122,1
```

A conclusão é que não se rejeita H_0 para o nível de significância de 5%, ou seja, não há evidência estatística para afirmar que a tinta amarela seque mais rapidamente do que a branca.

---

142. Um procedimento de controle de qualidade foi planejado para garantir um máximo de 5% de itens defeituosos na produção. A cada 15 minutos sorteia-se uma amostra de 20 peças e, havendo mais que 12% de defeituosas, pára-se a produção para verificações. Qual a probabilidade de uma parada desnecessária?

Solução:

$$\begin{aligned} P[\text{Parada desnecessária}] &= P[\hat{p} > 0,12 | p = 0,05] = \\ &= P[\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} > \frac{0,12 - 0,05}{\sqrt{(0,05)(1-0,05)/20}}] = P[Z > 1,44] = 0,075 \end{aligned}$$

---

143. Defina e ou discuta de forma comparativa (se quiser use exemplos) os conceitos:

(a) parâmetro, estimador e estimativa

(b) estimador consistente, estimador não viciado e estimador eficiente

(c) erro tipo I e erro tipo II em testes de hipótese

---

144. Antes de uma eleição, um determinado partido está interessado em estimar a proporção p de eleitores favoráveis a seu candidato. Uma amostra piloto de tamanho 200 revelou que 58% dos eleitores eram favoráveis ao candidato em questão.

(a) determine o tamanho necessário da amostra a ser tomada para que o erro cometido na estimação seja de, no máximo, 0,01 com probabilidade de 0,80.

(b) Se na amostra final com tamanho n obtido no item anterior, observou-se que 54% dos eleitores eram favoráveis ao candidato em questão, construa um intervalo de confiança (99%) para a proporção p.

Solução:

(a)

$$\begin{aligned} z_\alpha \sqrt{\frac{p(1-p)}{n}} &= 0.01 \\ n &= \lceil \frac{z_\alpha^2 p(1-p)}{0,01^2} \\ n &= \lceil \frac{1,28^2 \cdot (0.58)(1-0.58)}{0,01^2} \\ n &= 4001 \end{aligned}$$

(b) > *n <- ceiling((qnorm(0.90)^2)*(0.58)*(1-0.58)/(0.01^2))*
> *prop.test(0.54*n, n, conf=0.99)$conf*

```
[1] 0,5195 0,5603
attr(,"conf.level")
[1] 0,99
```

---

145. Um produtor de certo componente eletrônico está interessado em estimar a fração de itens defeituosos produzidos. Uma amostra aleatória de 800 unidades apresentou 40 defeituosos. Obtenha um intervalo de confiança (99%) para a proporção de defeituosos.

Solução:

> *prop.test(40, 800, conf.level = 0.99, corr=F)$conf*

```
[1] 0,03359 0,07381
attr(,"conf.level")
[1] 0,99
```

---

146. Dois tipos de plástico, I e II são adequados para uso na produção de certo componente eletrônico. A tensão de ruptura do material é uma característica importante. Sabe-se que $\sigma_I = \sigma_{II} = 1$ psi. Uma amostra aleatória com $n_I = 10$ e $n_{II} = 12$ fornece $\bar{x}_I = 162,5$ e $\bar{x}_{II} = 155,0$. Por razões de custo, a companhia não irá adotar o plástico I a menos que este tenha uma resistência média que exceda a do plástico II em pelo menos 10 psi. Proceda um teste de hipótese adequado para saber se, baseando-se nas informações da amostra, a companhia deve adotar o plástico I (use $\alpha = 0,05$).

Solução:

- $H_0 : \mu_I - \mu_{II} = 10$ versus$H_0 : \mu_I - \mu_{II} > 10$
- $\alpha = 0,05$
- $z_c = \frac{(\bar{x_I} - \bar{x_{II}}) - (\mu_I - \mu_{II})}{\sqrt{(\sigma_I^2/n_I) + (\sigma_{II}^2/n_{II})}} = \frac{(162{,}5 - 155) - (10)}{\sqrt{(1/10) + (1/12)}} = -5,84$
- p-valor $= 0,999999997370179$
- Não rejeita-se H_0, não há evidências ($\alpha = 0,05$) de que a média do tipo I supere a do tipo II em pelo menos 10 unidades.

---

147. Para se ajustar uma máquina, a correia deve ter entre 60 e 62 cm de comprimento. Tendo em vista o processo de fabricação, o comprimento destas correia pode ser considerado como uma variável aleatória com distribuição normal, de média 60,7 e desvio padrão 0,8 cm. Pergunta-se:

(a) qual a probabilidade de uma correia, escolhida ao acaso, poder ser usada na máquina?

(b) um grande revendedor dessas correias estabelece um controle de qualidade nos lotes que compra da fábrica: ele sorteia 4 correias do lote e só aceita o lote se o comprimento médio estiver dentro do tamanho aceito pela máquina. Calcule a probabilidade de aceitação do lote.

Solução:

$$X \; : \; \text{comprimento da correia}$$
$$X \sim N(60,7; 0.8^2)$$

(a) $P[\text{correia poder ser usada}] = P[60 < X < 62] = P[\frac{60-60{,}7}{0.8} < Z < \frac{62-60.7}{0.8}] = 0,757$

(b) $P[\text{lote ser aceito}] = P[60 < \bar{X} < 62] = P[\frac{60-60,7}{0.8/\sqrt{4}} < Z < \frac{62-60.7}{0.8/\sqrt{4}}] = 0,959$

---

148. Para uma variável aleatória com densidade normal e desvio padrão 5, o teste da média $\mu = 10$ contra $\mu = 14$, teve a região crítica dada por $\{\bar{x} \in \mathbb{R} : \bar{x} > 12\}$ para uma amostra de tamanho 25. Determine as probabilidades dos erros tipo I e II.

Solução:

$$\begin{aligned} X &\sim N(\mu; 5^2) \\ H_0 &: \mu = 10 \;\; versus \;\; \mu = 14 \end{aligned}$$

(a) $P[\text{erro tipo I}] = P[\text{rejeitar} H_0 | H_0 \text{é verdadeira}] = P[\bar{x} > 12 | \mu = 10] = P[Z > \frac{12-10}{5/\sqrt{25}}] = 0,023$

(b) $P[\text{erro tipo II}] = P[\text{não rejeitar} H_0 | H_0 \text{é falsa}] = P[\bar{x} < 12 | \mu = 14] = P[Z > \frac{12-14}{5/\sqrt{25}}] = 0,023$

---

149. O volume de vendas, no ramo de vestuário, tem se mantido estável de ano para ano, mas acedita-se que sofra mudança de um quadrimestre para outro, dentro de um mesmo ano. Através de uma metodologia adequada, foi criado um índice que reflete a quantidade vendida. Para o primeiro e terceiro quadrimestres do ano, foram escolhidas aleatóriamente algumas empresas de mesmo porte e seus índices de venda foram calculados conforme tabela abaixo.

Quad. 1	114,7	144,7	119,1	113,7	108,9	96,7	87,6	132,4	
Quad. 3	153,1	192,5	145,5	168,8	141,5	141,2	189,6	178,4	208,6

Faça testes de hipótese adequados, com nível de significância $\alpha = 0.10$ para verificar se:

(a) a variabilidade das vendas difere entre os dois períodos

(b) pode-se afirmar que os índices de venda são maiores no terceiro trimestre

Solução: Na solução computacional apresentada a seguir primeiro criamos objetos que contém os dados:

```
> q1 <- c(114.7, 144.7, 119.1, 113.7, 108.9, 96.7, 87.6, 132.4)
> ##q2 <- c(144.7, 173.4, 154.2, 154.7, 125.9, 119.5, 155.7, 213.9, 156.2, 159.0)
> q3 <- c(153.1, 192.5, 145.5, 168.8, 141.5, 141.2, 189.6, 178.4, 208.6)
```

(a) Neste item testamos a igualdade das variâncias, portanto a hipótese testada é:

$$H_0 : \sigma^2_{q1} = \sigma^2_{q3} \text{ versus } H_a : \sigma^2_{q1} \neq \sigma^2_{q3}$$

O resultado do teste mostra que a hipótese H_0 não deve ser rejeitada, o que pode ser visto de duas formas no resultado mostrado a seguir : (i) o intervalo de confiança contém o valor 1, que representa a igualdade no quociente de variâncias, ou, (ii) o valor-P (*p-value*) é superior a 0,10, o nível de significância adotado.

```
> var.test(q1,q3, conf=0.9)

F test to compare two variances

data:  q1 and q3
F = 0,54, num df = 7, denom df = 8, p-value = 0,4
alternative hypothesis: true ratio of variances is not equal to 1
90 percent confidence interval:
 0,1529 1,9937
sample estimates:
ratio of variances
            0,5351
```

(b) Neste item testamos a igualdade das variâncias, portanto a hipótese testada é:

$$H_0 : \mu_{q3} = \mu_{q1} \text{ versus } H_a : \mu_{q3} > \mu_{q1}$$

O resultado do teste-t unilateral para duas amostras, com variâncias desconhecidas, porém consideradas iguais, mostram que a hipótese H_a deve ser rejeitada, o que pode ser visto de duas formas no resultado mostrado a seguir : (i) o intervalo de confiança não contém o valor 0, que representa a igualdade des médias, ou, (ii) o valor-P (*p-value*) é inferior a 0,10, o nível de significância adotado.

```
> t.test(q3,q1, var.equal=T, conf=0.9, alt="greater")
Two Sample t-test

data:  q3 and q1
t = 5, df = 15, p-value = 7e-05
alternative hypothesis: true difference in means is greater than 0
90 percent confidence interval:
 39,71   Inf
sample estimates:
mean of x mean of y
    168,8     114,7
```

---

150. Suponha que uma indústria farmacêutica deseja saber a quantos voluntários se deva aplicar uma vacina, de modo que a proporção de indivíduos imunizados na amostra difira de menos de 2% da proporção verdadeira de imunizados na população, com probabilidade de 0.90. Qual o tamanho da amostra a escolher?

Solução:

$$\begin{aligned}
\hat{p} &\sim N(p; \frac{p(1-p)}{n}) \\
\hat{p} - p &\sim N(0; \frac{p(1-p)}{n}); \text{e adotamos } p = 0,5 \\
P[|(\hat{p} - p)| < 0,02] &\geq 0,90 \\
P[-0,02 < (\hat{p} - p) < 0,02] &\geq 0,90 \\
P[\frac{-0,02 - 0}{\sqrt{1/4n}} < (\hat{p} - p) < \frac{0,02 - 0}{\sqrt{1/4n}}] &\geq 0,90 \\
z &= 1.645 \\
\frac{1}{\sqrt{1/4n}} &= 1.645 \\
n &= 1691
\end{aligned}$$

---

151. Se uma amostra com 36 observações for tomada de uma população, qual deverá ser o tamanho de uma outra amostra para que o desvio padrão dessa amostra seja 2/3 do desvio padrão da média da primeira?

Solução:

$$\begin{aligned}
X &\sim N(\mu; \sigma^2) \\
\bar{X}_{36} &\sim N(\mu; \sigma^2/36) \\
\bar{X}_n &\sim N(\mu; \sigma^2/n) \\
\frac{\sigma}{\sqrt{n}} &= \frac{2}{3}\frac{\sigma}{\sqrt{36}} \\
n &= 81
\end{aligned}$$

---

152. (30 pontos) Um artigo no *Journal of Structural Engineering* (vol 115, 1989) descreve um experimento para testar a resistência de tubos circulares com tampas soldadas nas extremidades. Os primeiros dados (em kN) são: 96, 128, 102, 102, 104, 160, 96, 108, 126, 104, 128, 140, 156, 102 e 160.

   (a) obtenha um intervalo de confiança para a média populacional, indicando quais as suposições feitas na obtenção deste intervalo

   (b) teste a hipótese ($\alpha = 0,10$) de que a resistência média é superior a 110 kN.

Solução:

   (a) Supõe-se a normalidade dos dados. A seguir mostramos resultados para niveis de confiança de 90, 95 e 99%, respectivamente

```
> t.test(dados, conf=0.90)$conf
[1] 29,59 42,21
attr(,"conf.level")
[1] 0,9
> t.test(dados, conf=0.95)$conf
[1] 28,12 43,68
attr(,"conf.level")
[1] 0,95
> t.test(dados, conf=0.99)$conf
[1] 24,72 47,08
attr(,"conf.level")
[1] 0,99
```

   (b) O resultado abaixo mostra que a hipótese nula pode ser rejeitada para $\alpha = 0.10$ e portanto pode-se afirmar que a média é superior a 110 kN

```
> t.test(dados, conf=0.90, alt="greater", mu=110)
One Sample t-test

data:  dados
t = -22, df = 9, p-value = 1
alternative hypothesis: true mean is greater than 110
90 percent confidence interval:
 31,14   Inf
sample estimates:
mean of x
     35,9
```

---

153. Um tipo de tubo de PVC é manufaturado com diâmetro médio de 1,01 polegadas e um desvio padrão de 0,003 polegadas. Encontre a probabilidade de que uma amostra de tamanha $n = 9$ tenha uma média amostral do diâmetro maior que 1,009 e menor que 1,012 polegadas.

Solução:

$$\begin{aligned}
X &: \text{ diâmetro do tubo} \\
X &\sim N(1,01; (0,003)^2) \\
\bar{X}_9 &\sim N(1,01; (0,003)^2/9) \\
P[1,009 < \bar{X} < 1,012] &= P[\frac{1,009 - 1,01}{0,003/\sqrt{9}} < Z < \frac{1,012 - 1,01}{0,003/\sqrt{9}}] = P[-1 < Z < -1] = 0,819
\end{aligned}$$

---

154. De 1000 casos selecionados ao acaso de câncer de pulmão, 823 resultaram em morte.

(a) Obtenha um intervalo de confiança (95%) para a taxa de óbitos de câncer de pulmão,

(b) qual deveria ser o tamanho amostral para que com confiança de ao menos 95% o erro na estimação da taxa de óbitos seja inferior a 0,03?

Solução:

(a) Há duas possíveis soluções vistas no curso: o intervalo assintótico, tomando $p = \hat{p}$ ou conservador onde $p = 0,5$

```
> p.est <-  823/1000
> p.est + qnorm(c(0.025, 0.975)) * sqrt(p.est*(1-p.est)/1000)
[1] 0,7993 0,8467
> p.est + qnorm(c(0.025, 0.975)) * sqrt(0.5*(1-0.5)/1000)
[1] 0,792 0,854
```

(b) Aqui também há duas possíveis soluções considerando $p = \hat{p}$ ou $p = 0,5$

$$\begin{aligned}
P[|\hat{p} - p| < 0,03] &\geq 0.95 \\
z &= 1,96 = \frac{0,03 - 0}{p(1-p)/n} \\
n &= \left(\frac{1,96}{0,03}\right)^2 p(1-p) \\
n &= 622 \text{ para } p = \hat{p} \\
n &= 1068 \text{ para } p = 0.5
\end{aligned}$$

---

155. Suponha que amostras de tamanho $n = 25$ são selecionadas ao acaso de uma distribuição normal com média $\mu = 100$ e desvio padrão $\sigma = 10$.

(a) Qual a probabilidade de que a média amostral esteja no intervalo $\mu_{\bar{x}} - 1.8\sigma_{\bar{x}}$ a $\mu_{\bar{x}} + 1.0\sigma_{\bar{x}}$?

(b) Qual deveria ser o tamanho n da amostra para que a probabilidade de que a média amostral esteja no intervalo $(99, 101)$ seja de 0.90?

Solução:

$$X \sim N(100, 10^2) \quad \bar{X}_{25} \sim N(100, 10^2/25)$$

(a) $P[\mu_{\bar{x}} - 1.8\sigma_{\bar{x}} < \bar{X} < \mu_{\bar{x}} + 1.0\sigma_{\bar{x}}] = P[-1.8 < Z < 1.0] =$

```
> pnorm(1) - pnorm(-1.8)
[1] 0,8054
```

(b) $X \sim N(100, 10^2) \qquad \bar{X}_n \sim N(100, 10^2/n)$

$$\begin{aligned}
P[99 < \bar{X} < 101] &= 0.9 \\
P[\frac{99 - 100}{\sqrt{100/n}} < Z < \frac{101 - 100}{\sqrt{100/n}}] &= 0.9 \\
z &= 1,645 \\
z &= \frac{101 - 100}{\sqrt{100/n}} \\
n &= \frac{1,645^2 * 100}{(101 - 100)^2} \\
n &= 271
\end{aligned}$$

156. Um cientista da computação solicitou a 12 programadores para programar uma funcionalidade padrão em certa linguagem (A), e o tempo (em minutos) foi anotado obtendo-se:

programador	1	2	3	4	5	6	7	8	9	10	11	12
tempo	17	16	21	14	18	24	16	14	21	23	13	18

Assumindo que os tempos tem distribuição normal,

(a) construa intervalos de confiança a 90% e 99% para o tempo médio de programação;

(b) construa intervalos de confiança a 90% e 99% para a variância do tempo de programação.

Solução:

```
> tA <- c(17, 16, 21, 14, 18, 24, 16, 14, 21, 23, 13, 18)
> n <- length(tA)
```

(a)
```
> t.test(tA, conf=0.90)$conf
[1] 16,03 19,80
attr(,"conf.level")
[1] 0,9
> t.test(tA, conf=0.99)$conf
[1] 14,66 21,17
attr(,"conf.level")
[1] 0,99
```

(b)
```
> c((n-1)* var(tA)/ qchisq(0.95, df=n-1), (n-1)* var(tA)/ qchisq(0.05, df=n-1))
[1]  7,365 31,677
> c((n-1)* var(tA)/ qchisq(0.995, df=n-1), (n-1)* var(tA)/ qchisq(0.005, df=n-1))
[1]  5,416 55,668
```

157. Considere agora solicitou-se aos mesmos programadores do problema acima para implementarem a mesma funcionalidade em uma segunda linguagem (B). O tempo (em minutos) foi anotado obtendo-se:

programador	1	2	3	4	5	6	7	8	9	10	11	12
tempo	18	14	19	11	23	21	10	13	19	24	15	20

(a) Use um procedimento adequado para testar a hipótese de que as linguagens não diferem quanto ao tempo médio de programação

(b) Suponha que decide-se usar a seguinte regra: se a diferença entre os tempos médios de programação for superior a 2 unidades, decide-se que as linguagens possuem tempos médios diferentes para implementação da funcionalidade desejada. Qual a probabilidade do erro tipo I neste caso?

Solução:

```
> tB <- c(18, 14, 19, 11, 23, 21, 10, 13, 19, 24, 15, 20)
```

(a) `> t.test(tA, tB, paired = TRUE, conf=0.95)`

```
Paired t-test

data:  tA and tB
t = 0,78, df = 11, p-value = 0,5
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1,217  2,550
sample estimates:
mean of the differences
                 0,6667
```

(b) $P[|\bar{D}| > 2 \mid \mu_d = 0]$

```
> 2 * pnorm(mean(abs(tB - tA)), m=0, sd=sqrt(var(tB-tA)/n), low=F)
[1] 0,003485
```

---

158. Uma indústria submete seus novos operários a um teste de aptidão e três meses depois mede a produtividade destes operários. Para uma amostra de 6 operários os dados obtidos foram:

operário	1	2	3	4	5	6
teste de aptidão	22	25	15	19	22	18
produtividade	45	37	25	40	33	30

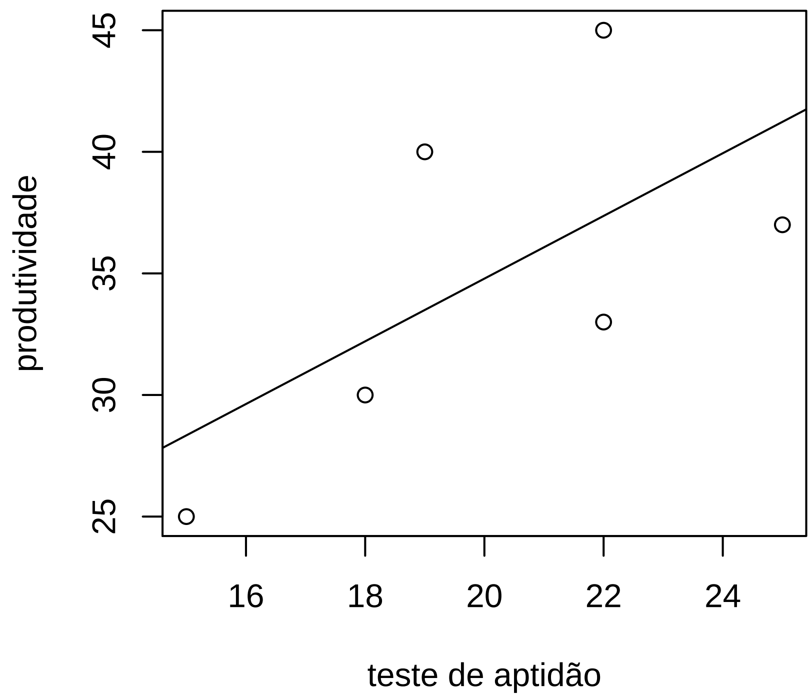
Use análises estatísticas adequadas para determinar se a produtividade está associada ao resultado no teste de aptidão e caso positivo, descreva a forma desta associação e obtendo ainda a produtividade esperada para um operário que obtém resultado 20 no teste de aptidão.

Solução:

```
> x <- c(22, 25, 15, 19, 22, 18)
> y <- c(45, 37, 25, 40, 33, 30)
> cor(x,y)

[1] 0,6362

> reg <- lm(y ~ x)
> reg

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
      9,00         1,29

> predict(reg, data.frame(x=20))
```

```
    1
34,79
```

```
> plot(x,y, xlab="teste de aptidão", ylab="produtividade")
> abline(reg)
```



---

159. Foi tomada uma amostra de 18 representantes de venda dos quais se obteve os gastos (em centenas de reais) em ligações telefônicas de longa distância durante 6 meses como a seguir:

6.2 5.4 6.7 5.8 4.5 4.7 7.1 6.3 6.5 3.4 5.6 3.1 6.4 5.0 6.1 6.1 5.6 5.1

(a) assumindo que os gastos tem distribuição normal, obtenha um intervalo de confiança (a 95%) para o gasto médio

(b) obtenha também um intervalo de confiança (95%) para variância dos gastos

Solução:

```
> gasto <- c(6.2 , 5.4 , 6.7 , 5.8 , 4.5 , 4.7 , 7.1 , 6.3 , 6.5,
+             3.4 , 5.6 , 3.1 , 6.4 , 5.0 , 6.1 , 6.1 , 5.6 , 5.1)
```

(a) `> t.test(gasto, conf=0.95)$conf`

```
[1] 4,994 6,072
attr(,"conf.level")
[1] 0,95
```

(b) `> (length(gasto)-1) * var(gasto)/qchisq(c(0.975, 0.025), df=length(gasto)-1)`

```
[1] 0,6618 2,6414
```

160. (40 pontos) Um experimento foi feito para comparar os custos de reparo de dois tipos de bombas (A e B). Para isto registrou-se o custo para 16 bombas de cada tipo, ao longo de um ano de operação. Os custos anuais de reparo (em milhares de reais) foram os seguintes:

Tipo A:	0.4	1.8	4.0	1.8	2.3	1.9	0.8	1.4	3.4	1.0	2.0	0.5	1.6	3.9	1.6	1.8
Tipo B:	1.2	0.8	2.7	0.8	0.5	1.1	0.7	1.3	0.9	3.9	1.4	2.6	0.6	0.6	1.5	2.0

(a) faça um teste de hipótese adequado com nível de significância $\alpha = 0,05$ para comparar os custos de reparo das duas bombas.

(b) qual seria a probabilidade do *erro tipo I* caso no teste acima decida-se adotar que, sendo a diferença das médias dos dois grupos superior a 0.4 unidades, declara-se que os custos de reparo diferem entre os dois tipos de bombas?

Solução:

```
> tipoA <-c(0.4, 1.8, 4.0, 1.8, 2.3, 1.9, 0.8, 1.4, 3.4, 1.0, 2.0, 0.5, 1.6, 3.9, 1.6, 1.8)
> tipoB <-c(1.2, 0.8, 2.7, 0.8, 0.5, 1.1, 0.7, 1.3, 0.9, 3.9, 1.4, 2.6, 0.6, 0.6, 1.5, 2.0)
> nA <- length(tipoA)
> nB <- length(tipoB)
```

(a) faça um teste de hipótese adequado com nível de significância

```
> var.test(tipoA, tipoB)

F test to compare two variances

data:  tipoA and tipoB
F = 1,3, num df = 15, denom df = 15, p-value = 0,6
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0,4555 3,7311
sample estimates:
ratio of variances
             1,304

> t.test(tipoA, tipoB, var.equal=T)

Two Sample t-test

data:  tipoA and tipoB
t = 1,3, df = 30, p-value = 0,2
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0,2585  1,2085
sample estimates:
mean of x mean of y
    1,887     1,413
```

(b) qual seria a probabilidade do *erro tipo I* caso no teste acima decida-se adotar que, sendo a diferença das médias dos dois grupos superior a 0,4 unidades, declara-se que os custos de reparo diferem entre os dois tipos de bombas?

```
> s2p <- ((nA-1) * var(tipoA)  + (nB-1) * var(tipoB))/(nA+nB-2)
> 2*pnorm(0.4, mean=0, sd=sqrt(s2p * (1/nA)+(1/nB)), low=F)

[1] 0,2617
```

161. Um escore de aptidão escolar para alunos de um certo ano tem média $\mu = 475$ e desvio padrão $\sigma = 80$.

(a) para uma amostra de 225 alunos, qual a probabilidade de que a média amostral esteja entre 465 e 485?

(b) para esta mesma amostra, quais os valores simétricos ao redor da média populacional entre os quais espera-se encontrar a média amostral com 0,80 de probabilidade?

(c) qual deveria ser o tamanho da amostra para que a probabilidade da média amostral estar entre 460 e 490 seja de 0,90?

Solução:

$$X \sim N(\mu, \sigma^2) \qquad \bar{X} \sim N(\mu, \sigma^2/n)$$

(a) `> pnorm(485, m=475, sd=80/sqrt(225)) - pnorm(465, m=475, sd=80/sqrt(225))`

`[1] 0,9392`

(b) `> qnorm(c(0.10, 0.90), m=475, sd=80/sqrt(225))`

`[1] 468,2 481,8`

(c) `> ceiling((80 * qnorm(0.95)/15)^2)`

`[1] 77`

---

162. Assume-se que a durabilidade de um certo tipo de estrutura tem distribuição normal de média 50 anos e desvio padrão de 10 anos.

(a) Toma-se uma amostra de 16 destas estruturas. Qual a probabilidade da durabilidade média destas amostras estar entre 45 e 53 anos?

(b) Suponha que não se conhece o tempo médio de duração e que a média desta amostra foi de 51,4 anos. Obtenha um intervalo de confiança para a durabilidade média.
Usando $1 - \alpha = 0.95$:

(c) Qual deveria ser o tamanho da amostra para que a amplitude do intervalo de confiança para média fosse de 4 anos?

(d) Suspeita-se que a estrutura está fora da especificação, apresentando tempo de duração inferior aos 50 anos. Para testar esta hipótese decide-se que, se a média de uma amostra de 20 estruturas for inferior a 47 anos, a estrutura é considerada fora da especificação. Qual a probabilidade do erro tipo I neste teste?

(e) Qual a probabilidade do erro tipo II no teste mencionado no ítem anterior caso o tempo médio de duração estivesse na verdade em 45 anos?

Solução:

$$\begin{aligned} X &: \text{durabilidade da estrutura (em anos)} \\ X &\sim N(50, 10^2) \end{aligned}$$

(a) `> diff(pnorm(c(45,53), mean=50, sd=10/sqrt(16)))`

`[1] 0,8622`

(b) `> 51.4 + qnorm(c(0.025, 0.975)) * 10/sqrt(16)`

`[1] 46,5 56,3`

(c) `> ceiling((qnorm(0.975) * 10/(4/2))^2)`

`[1] 97`

(d) `> pnorm(47, m=50, sd=10/sqrt(20))`

`[1] 0,08986`

(e) `> pnorm(47, m=45, sd=10/sqrt(20), lower=F)`

```
[1] 0,1855
```

---

163. Uma pesquisa ouviu uma amostra de 1500 pessoas para saber se apoiavam uma certa mudança na constituição. Destas, 730 declararam-se favoráveis.

   (a) Obtenha um intervalo de confiança (90%) para proporção de favoráveis.
   (b) Qual deveria ter sido o tamanho da amostra para que a margem de erro fosse de 1% ?
   (c) Baseado nesta amostra pode-se afirmar que a proporção de favoráveis é inferior a 50% ? Teste esta hipótese para $\alpha = 0,05$.

Solução:

(a)
```
> p.est <- 730/1500
> IC.assintotico <- p.est + qnorm(c(0.05, 0.95)) * sqrt((p.est)*(1-p.est)/1500)
> IC.assintotico
```
```
[1] 0,4654 0,5079
```
```
> IC.conservador <- p.est + qnorm(c(0.05, 0.95)) * sqrt(1/(4*1500))
> IC.conservador
```
```
[1] 0,4654 0,5079
```

(b) *Usando o IC assintótico*:
```
> ceiling(((qnorm(0.95)^2) * (p.est)*(1-p.est))/(0.01)^2)
```
```
[1] 6760
```
Usando o IC conservador:
```
> ceiling((qnorm(0.95)^2)/(4*(0.01)^2))
```
```
[1] 6764
```

(c)
```
> zc <- (p.est - 0.5)/sqrt((0.5)*(1-0.5)/1500)
> (pvalor <- pnorm(zc))
```
```
[1] 0,1508
```

---

164. A sequência de operações executadas por operários para executar certa tarefa está sendo estudada. Uma nova sequência está sendo proposta e deseja-se verificar se é adequado substituir a anterior. Para tanto, 9 operários foram sorteados e mediu-se o tempo necessário, em minutos, para que cada um realizasse a tarefa, com dois tipos de sequências. Os dados colotados foram:

Operário	1	2	3	4	5	6	7	8	9
Atual	24	25	27	22	23	28	26	28	29
Nova	21	23	28	27	24	26	25	22	23

Suponha que o modelo normal é adequado e teste a hipótese de que houve redução no tempo de execução ($\alpha = 0,05$). Apresente os passo a passo os procedimentos do teste explicando a interpretação das hipóteses e justificando seu formato. Calcule também o *valor-p* (nível descritivo) do teste.

Solução:

```
> atual <- c(24, 25, 27, 22, 23, 28, 26, 28, 29)
> nova <-  c(21, 23, 28, 27, 24, 26, 25, 22, 23)
> t.test(nova, atual, paired=T, alt="less")

Paired t-test

data:  nova and atual
t = -1,2, df = 8, p-value = 0,1
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
   -Inf 0,7275
sample estimates:
mean of the differences
                 -1,444
```

---

165. De 50.000 válvulas fabricadas por uma companhia retira-se uma amostra de 400 válvulas e obtém-se a vida média de 800 horas e desvio padrão de 100 horas.

(a) Qual o intervalo de confiança de 99% para vida média da população de válvulas?

(b) Com que confiança dir-se-ia que a vida média é de $800 \pm 0,98$?

(c) Que tamanho deve ter uma amostra para que seja de 95% a estimativa intervalar de 800 ± 7.84 ?

Solução: O tamanho da amostra $n = 400$ é bem menor que o tamanho $N = 50.000$ da população (n/N = 0,008 < 0,05) e portanto vamos considerar a população como "infinita".

$$
\begin{aligned}
&\text{População:}\\
&X : \text{tempo de vida da válvula (horas)}\\
&\text{Amostra:}\\
&n = 30 \; ; \overline{x} = 800 \text{ e } S = 100\\
&\text{Distribuição Amostral:}\\
&\overline{X} \sim \text{t}(\nu = n-1)\\
&IC(1-\alpha) : \overline{X} \pm t_{1-\alpha/2} S/\sqrt{n}\\
&\text{Mas, como a amostra é "grande"}(n > 30)\text{pode-se usar:}\\
&\text{Distribuição Amostral (aproximada):}\\
&\overline{X} \sim \text{t}(\nu = n-1) \approx \text{N}(\mu, S^2/n) (\text{usualmente para n} > 30)\\
&IC(1-\alpha) : \overline{X} \pm z_{1-\alpha/2} S/\sqrt{n}
\end{aligned}
$$

(a)

$$
\begin{aligned}
IC(0,99) &: \overline{X} \pm z_{0,995} \cdot S/\sqrt{n}\\
IC(0,99) &: 800 \pm 2,576 \cdot 100/\sqrt{400}\\
IC(0,99) &: (787,1 \; ; \; 812,9)
\end{aligned}
$$

(b)

$$
\begin{aligned}
IC(1-\alpha/2) &: 800 \pm 0,98\\
z_{1-\alpha/2} \cdot S/\sqrt{n} &= 0,98\\
z_{1-\alpha/2} \cdot 100/\sqrt{400} &= 0,98\\
z_{1-\alpha/2} &= 0,196\\
1-\alpha/2 &= 15,5\%
\end{aligned}
$$

(c)

Soluções computacionais com o programa R:

```
> ## a)
> 800 + qt(c(0.005, 0.995), df=399) * (100/sqrt(400))

[1] 787,1 812,9

> (ICa <- 800 + qnorm(c(0.005, 0.995)) * (100/sqrt(400)))

[1] 787,1 812,9

> ## b)
> 1 - 2*pt(0.98 * sqrt(400)/100, df=399,lower=F)

[1] 0,1553

> (CONFa <- 1 - 2*pnorm(0.98 * sqrt(400)/100,lower=F))

[1] 0,1554

> ## c)
> ceiling((qt(0.975, df=399) * 100/7.84)^2)

[1] 629

> (na <- ceiling((qnorm(0.975) * 100/7.84)^2))

[1] 625
```

---

166. Suponha que estamos interessados em estimar a porcentagem de consumidores de um certo produto. Se a amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:

(a) O intervalo de confiança de p, a proporção na população, a 95%. Interprete o resultado deste intervalo, ou seja, explique com palavras co significado do intervalo obtido.

(b) O tamanho da amostra para que o erro da estimativa não exceda a 0,02 unidades com probabilidade de 90%.

(c) Teste a hipótese com $\alpha = 0.05$ de que a proporção é superior a 30%.

Solução:

(a)
```
> p.est <-100/300
> p.est + qnorm(c(0.025, 0.975)) * sqrt(p.est*(1-p.est)/300)

[1] 0,2800 0,3867

> p.est + qnorm(c(0.025, 0.975)) * sqrt(1/(4*300))

[1] 0,2768 0,3899
```

(b)
```
> ceiling((qnorm(0.95)^2)*(p.est)*(1-p.est)/(0.02^2))

[1] 1504

> ceiling((qnorm(0.95)^2)*(0.5)*(0.5)/(0.02^2))

[1] 1691
```

(c)
```
> pnorm((p.est - 0.30)/sqrt(p.est*(1-p.est)/300))
```

```
[1] 0,8897
```

---

167. Explique os seguinte conceitos:

    (a) estimador não viciado.
    (b) estimador eficiente.
    (c) erro tipo I.
    (d) erro tipo II.
    (e) hipótese unilateral.

---

168. Assume-se que o tempo de processamento de uma certa requisição tem distribuição normal de média 500 segundos e desvio padrão de 20 segundos. Para checar se a processo está sob controle foi tomada uma amostra de 15 requisições para as quais mediu-se o tempo de processamento. Responda:

    (a) qual a probabilidade do tempo de processamento de uma requisição ser superior a 530 segundos?
    (b) se o processo está dentro da especificação, qual a probabilidade da média da amostra estar acima de 505 segundos?
    (c) decide-se adotar o seguinte critério: se o tempo médio verificado na amostra for superior a 508 segundos o processo é considerado fora de controle e é interrompido. Formule as hipóteses adequadas para testar a necessidade de interromper o processo.
    (d) considerando o critério citado qual a probabilidade do erro tipo I ?
    (e) se a média do processo alterou-se para 510 segundos, qual a probabilidade do erro tipo II ?
    (f) qual deveria ser o tamanho da amostra para detectar com probabilidade de 90% uma diferença de 5 segundos no tempo médio de processamento?
    (g) suponha agora que não se conhece o tempo médio de processamento. A média da amostra forneceu o valor de $\bar{x} = 497$. Obtenha um intervalo de confiança (99%) para o tempo médio de processamento.
    (h) qual deveria ser o tamanho da amostra para que o intervalo calculado no item anterior tivesse metade da amplitude?
    (i) baseado nos dados da amostra, teste ao nível $\alpha = 0.05$ a hipótese de que o tempo de processamento é inferior a 502 segundos.
    (j) está se estudando a aquisição de um outro servidor que tem a mesma varância no tempo de processamento, porém espera-se que com um tempo de processamento inferior ao servidor atual. Para comparar os servidores foi tomada também uma amostra do tempo de processamento deste segundo servidor. A amostra de tamanho $n = 18$ forneceu um tempo de processamento médio de $\bar{x} = 495$. Utilize as amostras dos 2 servidores para testar a hipótese de que o segundo servidor é mais rápido que o primeiro.

Solução:

(a) *> pnorm(530, mean=500, sd=20, lower=F)*

```
[1] 0,06681
```

(b) *> pnorm(505, mean=500, sd=20/sqrt(15), lower=F)*

```
[1] 0,1665
```

(c)

$$\begin{cases} H_0: & \mu \leq 500 \\ H_a: & \mu > 500 \end{cases}$$

(d) *> pnorm(508, mean=500, sd=20/sqrt(15), lower=F)*

```
[1] 0,06067
```

(e) *Adotando-se* $\alpha = 0,05$:

```
> xc <- qnorm(0.05, mean=500, sd=20/sqrt(15), lower=F)
> pnorm(xc, mean=510, sd=20/sqrt(15))
[1] 0,3853
```

(f) *Adotando-se* $\alpha = 0,05$:

```
> ceiling(((qnorm(0.975) + qnorm(0.90))^2)*(20^2)/(5^2))
[1] 169
```

(g)
```
> IC <- 497 + qnorm(c(0.005, 0.995)) * 20/sqrt(15)
> IC
[1] 483,7 510,3
```

(h)
```
> ceiling((qnorm(0.995) * 20/((diff(IC)/2)/2))^2)
[1] 60
```

(i)

$$\begin{cases} H_0: & \mu \geq 502 \\ H_a: & \mu < 502 \end{cases}$$
$$\alpha = 0,05$$
$$z_c = \frac{497 - 502}{20/\sqrt{15}}$$
$$\text{P-valor} = 0,166$$

(j)

$$\begin{cases} H_0: & \mu_2 \geq \mu_1 \\ H_a: & \mu_2 < \mu_1 \end{cases}$$
$$\alpha = 0.05$$
$$z_c = \frac{(497 - 495) - 0}{20\sqrt{(1/15) + (1/18)}}$$
$$\text{P-valor} = 0,613$$

---

169. Antes de uma eleição para um cargo majoritário, um determinado partido está interessado em estimar a probabilidade p de eleitores favoráveis ao seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores eram favoráveis ao candidato.

(a) Utilizando a informação da amostra piloto, determine o tamanho da amostra para que, com 0.80 de probabilidade, o erro cometido na estimação seja no máximo de 5 pontos percentuais.

(b) Se na amostra final, com tamanho obtido no ítem anterior, observou-se que 51% dos eleitores eram favoráveis ao candidato. Construa um intervalo de confiança para p, com confiança 95%.

Solução:

(a)
```
> (n <- ceiling((qnorm(0.90)^2)*0.6*0.4/(0.05^2)))
[1] 158
```

(b)
```
> 0.51 + qnorm(c(0.025, 0.975)) * sqrt(0.51*0.49/n)
[1] 0,4321 0,5879
```

---

170. Um experimento foi conduzido para comparar a capacidade de enchimento de equipamentos em 2 vinícolas, **A** e **B**. Em cada uma delas, 20 garrafas de vinho *Pinot Noir* foram selecionadas aleatoriamente e medidas e os dados estão na tabela a seguir.

Vinícola A	755	753	754	752	755	751	753	752	753	753
	752	753	751	753	750	753	754	753	752	753
Vinícola B	756	755	754	754	756	754	756	755	756	756
	757	756	755	755	756	756	753	754	756	756

```
> vinA <- c(755, 753, 754, 752, 755, 751, 753, 752, 753, 753, 752, 753, 751, 753, 750, 753, 754,
> vinB <- c(756, 755, 754, 754, 756, 754, 756, 755, 756, 756, 757, 756, 755, 755, 756, 756, 753,
```

(a) Os dados apoiam a conjectura de que nas duas vinícolas as garrafas são cheias com o mesmo volume médio? (Use $\alpha = 0,05$ e assuma que os desvios padrões são iguais)

(b) Qual o **valor-P** do teste?

(c) Construa um gráfico *box-plot* para as duas amostras e compare os dois grupos. Baseando-se neste gráfico discuta se o pressuposto feito no ítem anterior de que as variâncias são iguais parece adequado.

Solução:

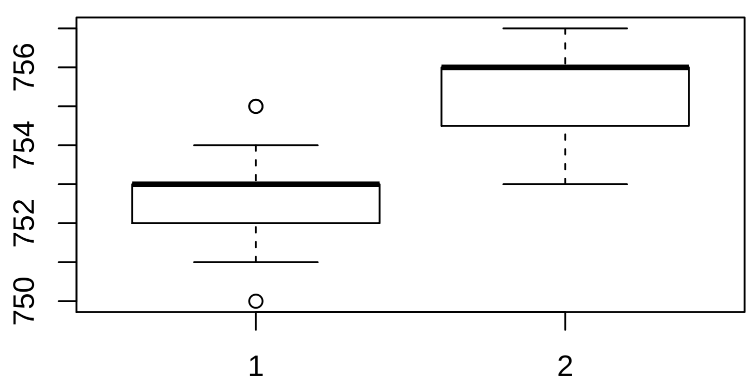
(a) `> t.test(vinA, vinB, var.equal=T, conf=0.95)`

```
Two Sample t-test

data:  vinA and vinB
t = -7, df = 38, p-value = 2e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3,284 -1,816
sample estimates:
mean of x mean of y
    752,8     755,3
```

(b) `> t.test(vinA, vinB, var.equal=T, conf=0.95)$p.value`

```
[1] 2,217e-08
```

(c) `> boxplot(vinA, vinB)`



171. Numa pesquisa de possuidores de carros numa universidade, entre homens e mulheres, foram obtidos: 24 de 100 homens possuem automóveis e 13 de 100 mulheres possuem automóveis. Baseado nestes resultados voce afirmaria que a proporção de homens e mulheres que possuem carros é a mesma na população? (adote um procedimento estatístico para verificar com confiança de 95%). Assumindo as mesmas proporções observadas na amostra original, qual deveria ser o tamanho da amostra, com o mesmo número de homens e mulheres, para que a hipótese da igualdade fosse rejeitada com $\alpha = 0,01$.

Solução:

```
> prop.test(x=c(24, 13), n=c(100, 100), alt="two.sided", correct=F)

2-sample test for equality of proportions without continuity correction

data:  c(24, 13) out of c(100, 100)
X-squared = 4, df = 1, p-value = 0,05
alternative hypothesis: two.sided
95 percent confidence interval:
 0,003457 0,216543
sample estimates:
prop 1 prop 2
  0,24   0,13

> p1e <- 24/100
> p2e <- 13/100
> ceiling(((qnorm(0.995)^2) * ((p1e*(1-p1e))+(p2e*(1-p2e))))/((p1e-p2e)^2))

[1] 163
```

---

172. Uma companhia de cigarros anuncia que o índice médio de nicotina dos cigarros que fabrica apresenta-se abaixo de 23 mg por cigarro. Um laboratório realiza 6 análises desse índice, obtendo: 27, 24, 21, 25, 26, 22.

(a) Pode-se aceitar ao nível de 10% a afirmação do fabricante?

(b) No teste acima, qual a probabilidade do erro tipo II caso o teor seja de fato 24,7 mg e assumindo que o índice de nicotina tem distribuição normal com variância de 4,86 mg^2.

(c) Assumindo que o índice de nicotina tem distribuição normal com variância de 4,86 mg^2, qual a probabilidade do erro tipo I no teste feito no ítem (a) se adotar-se a região crítica $\bar{x} > 25$?

Solucão:

```
> nic <- c(27, 24, 21, 25, 26, 22)
```

(a)
```
> t.test(nic, alt="less", mu=23, conf=0.90)

One Sample t-test

data:  nic
t = 1,2, df = 5, p-value = 0,9
alternative hypothesis: true mean is less than 23
90 percent confidence interval:
  -Inf 25,56
sample estimates:
mean of x
    24,17
```

(b)
```
> xc <- 23 + qnorm(0.90) * sqrt(4.68)
> pnorm(xc, m=24.7, sd=sqrt(4.68), low=F)
[1] 0,31
```

(c)
```
> pnorm(25, m=23, sd=sqrt(4.68), low=F)
[1] 0,1776
```

---

173. Os dados a seguir correspondem a teores de um elemento indicador da qualidade de um certo produto vegetal. Foram coletadas 2 amostras referentes a 2 métodos de produção.

Método 1	0.9	2.5	9.2	3.2	3.7	1.3	1.2	2.4	3.6	8.3
Método 2	5.3	6.3	5.5	3.6	4.1	2.7	2.0	1.5	5.1	3.5

(a) Encontre um intervalo de confiança (99%) para a média do método 1
(b) Encontre um intervalo de confiança (95%) para a variância do método 1
(c) Teste a hipótese de igualdade das variâncias entre os dois métodos ao nível de 10%
(d) Teste a hipótese de igualdade das médias entre os dois métodos ao nível de 5%
(e) Qual o nível de confiança para o intervalo (3.16 ; 4.76)?
(f) Qual deveria ser o tamanho de uma amostra para obter um intervalo com amplitude de 1 unidade com 90% de confiança?

Solução:

(a)
(b)
(c)
(d)
(e)
(f)

---

174. Seja X uma variável aleatória com função de densidade $f(x) = \lambda \exp(-\lambda x)$ para $x \geq 0$. Encontre a expressão do estimador de máxima verossimilhança de X baseado em uma amostra de tamanho n

---

175. O processo de produção de uma determinada peça é controlado de maneira que a percentagem de ítens defeituosos é de 10%. Se os itens são vendidos em caixas de 100 unidades qual a probabilidade de que em uma caixa:

(a) haja mais que 12% de defeituosos?
(b) não seja encontrado nenhum defeituoso?
(c) se um cliente encontrar mais de 15 defeituosos ele recebe uma caixa grátis. Qual a proporção esperada de clientes bonificados?

---

176. Foi feita uma pesquisa para eleição ouvindo-se uma amostra aleatória de 1100 eleitores.

(a) Teste a hipótese ao nível de 1% de que um candidato que tenha obtido 52.4% dos votos poderia se considerar eleito, se a eleição fosse no dia da pesquisa.
(b) Qual a margem de erro da pesquisa, a 95% de confiança?
(c) Qual deveria ser o tamanho da amostra para obter uma margem de erro de 1% com 95% de confiança?

Solução:

(a)

(b)

(c)

---

177. Uma enchedora automática de garrafas de refrigerantes está regulada para que o volume médio de líquido seja de $1.000cm^3$ e o desvio padrão de $10cm^3$. Pode-se admitir que a distribuição da variável seja normal.

(a) Qual a porcentagem esperada de garrafas em que o volume é menor de $990cm^3$?

(b) Qual a porcentagem esperada de garrafas em que o volume de líquido não se desvia da média em mais que 2 desvios padrão?

(c) O que acontecerá com a porcentagem do ítem anterior se a máquina for regulada de forma que a média seja de $1.200cm^3$ e o desvio padrão de $20cm^3$

(d) Mantendo o desvio padrão de $10cm^3$, em quanto deveria ser regulada a média para garantir que 90% ou mais da garrafas tenham volume superior a $1.000cm^3$?

(e) Mantendo a média de $1.000cm^3$ quanto deveria ser o desvio padrão para garantir que 95% das garrafas tenham volume entre 990 e $1010cm^3$?

(f) Suspeita-se que o processo está produzindo garrafas com volume médio inferior a $1.000cm^3$. Foi então tomada uma amostra de garrafas e medido o seu volume para ver se o processo está sob controle. Os valores obtidos foram:

```
986  998  1002  990  995  984  993  1000  999  1004  985
```

Teste a hipótese de interesse ao nível de 5% de significância.

(g) Encontre um intervalo de confiança (99%) para média usando os dados desta amostra.

(h) Com os dados acima teste ao nível de 10% de significância a hipótese de que o desvio padrão difere de $10cm^3$.

(i) Qual o nível de confiança para o intervalo (991.18 ; 997.18)?

(j) Assumindo-se que a variância da população é igual à variância da amostra, qual deveria ser o tamanho de uma amostra para obter um intervalo com amplitude de 2 unidades com 90% de confiança?

(k) Encontre um intervalo de confiança (95%) para a variância da amostra.

(l) Foi obtida também uma amostra em uma segunda fábrica engarrafadora a fim de determinar se as duas fábricas estão produzindo garrafas com o mesmo volume médio:

```
1002  988  998  1001  1005  1007  997  996  999
```

Teste ao nível de significância de 5% a hipóteses de igualdade das médias entre as duas fábricas.

Solução:

(a)
```
> pnorm(990, m=1000, sd=10)
[1] 0,1587
```

(b)
```
> pnorm(1020, m=1000, sd=10)- pnorm(980, m=1000, sd=10)
[1] 0,9545
```

(c)
```
> pnorm(1240, m=1200, sd=20) - pnorm(1160, m=1200, sd=20)
[1] 0,9545
```

(d)
```
> 1000 - qnorm(0.10) * 10
[1] 1013
```

(e)
```
> (1010-1000)/qnorm(0.975)
```

```
[1] 5,102
```

(f)
```
> m1 <- c(986, 998, 1002, 990, 995, 984, 993, 1000, 999, 1004, 985)
> t.test(m1, mu=1000, alt="less")
```

```
One Sample t-test

data:  m1
t = -2,7, df = 10, p-value = 0,01
alternative hypothesis: true mean is less than 1000
95 percent confidence interval:
  -Inf 998,1
sample estimates:
mean of x
    994,2
```

(g)
```
> pchisq((length(m1)-1)* var(m1)/100, df=10)
```

```
[1] 0,1113
```

(h)
```
> t.test(m1,conf=0.99)$conf.int
```

```
[1]  987,4 1001,0
attr(,"conf.level")
[1] 0,99
```

(i)
```
> 1 - 2*pt(3 * sqrt(10)/sd(m1), df=10, low=F)
```

```
[1] 0,7891
```

(j)
```
> (qnorm(0.95)*sd(m1))^2
```

```
[1] 136,3
```

(k)
```
> c((length(m1)-1)*var(m1)/qchisq(0.975, df=10),(length(m1)-1)*var(m1)/qchisq(0.025, df=10))
```

```
[1]  24,59 155,11
```

```
> var(m1)*10/100
```

```
[1] 5,036
```

(l)
```
> m2 <- c(1002,  988, 998, 1001, 1005, 1007, 997, 996, 999)
> var.test(m1,m2)
```

```
F test to compare two variances

data:  m1 and m2
F = 1,6, num df = 10, denom df = 8, p-value = 0,5
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0,3789 6,2740
sample estimates:
ratio of variances
             1,628
```

```
> t.test(m1,m2, var.eq=T)
```

```
Two Sample t-test

data:  m1 and m2
t = -1,7, df = 18, p-value = 0,1
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11,14   1,06
sample estimates:
mean of x mean of y
    994,2     999,2
```

178. Foi feita uma pesquisa para eleição ouvindo-se uma amostra aleatória de 1050 eleitores.

(a) Teste a hipótese ao nível de significância de 1% de que um candidato que tenha obtido 52.7% dos votos poderia se considerar eleito, se a eleição fosse no dia da pesquisa.

(b) Qual a margem de erro da pesquisa, a 95% de confiança?

(c) Qual deveria ser o tamanho da amostra para obter uma margem de erro de 1% com 95% de confiança?

Solução:

(a)
```
> prop.test(0.527*1050, 1050, p=0.5, alt="gr", conf=0.99)
```
```
1-sample proportions test with continuity correction

data:  0.527 * 1050 out of 1050, null probability 0.5
X-squared = 3, df = 1, p-value = 0,04
alternative hypothesis: true p is greater than 0,5
99 percent confidence interval:
 0,4906 1,0000
sample estimates:
    p
0,527
```

(b)
```
> teste <- prop.test(0.527*1050, 1050, p=0.5)
> teste$conf.int
```
```
[1] 0,4963 0,5575
attr(,"conf.level")
[1] 0,95
```
```
> diff(teste$conf.int)/2
```
```
[1] 0,03062
```

(c)
```
> ((qnorm(0.975)/0.01)^2)/4
```
```
[1] 9604
```

---

179. Vinte observações do tempo de falha (em horas) de um material de isolamento elétrico são mostradas a seguir.

204	228	252	300	324	444	624	720	816	912
1176	1296	1392	1488	1512	2520	2856	3192	3528	3710

(a) Baseado nos dados e nas medidas acima o que voce pode dizer sobre a distribuição dos tempos de vida?

(b) Suponha que os tempos de vida tenham distribuição exponencial $f(x) = \lambda \exp\{-\lambda x\}$; $x \geq 0$. Encontre o valor da estimativa do parâmetro λ da distribuição.

(c) Usando a estimativa acima como o verdadeiro valor do parâmetro, calcule a probabilidade de que o tempo de falha de um componente seja superior a 2000 horas.

(d) Calcule novamente a probabilidade acima agora assumindo que o valor do parâmetro é 0,001.

Solução:

```
> tp <- c(204, 228, 252, 300, 324, 444, 624, 720, 816, 912, 1176, 1296, 1392, 1488, 1512, 2520, 28
```

(a) *variabilidade alta com $CV = 84\%$, com média superior à mediana, diferenças de quartis à mediana indicam uma distribuição assimétrica*

(b) *O estimador de MV de λ é o recíproco média amostral:* $\hat{\lambda} = 1/\bar{x}$

```
> lh <- 1/mean(tp)
> lh
```

```
[1] 0,0007274
```

(c) $P[X > 2000] = \int_{2000}^{\infty} \lambda \exp\{-\lambda x\} = 1 - \int_{2000}^{\infty} \lambda \exp\{-\lambda x\}$ para $\lambda = 7e - 04$.

```
> 1 - integrate(dexp, 0, 2000, rate=lh)$value
```

```
[1] 0,2334
```

(d)
```
> 1 - integrate(dexp, 0, 2000, rate=0.001)$value
```

```
[1] 0,1353
```

---

180. Os dados abaixo são da temperatura dos efluentes na saída de uma estação de tratamento de esgotos em dias consecutivos.

43	47	45	48	44	50	46	49	45	44	46	46
52	49	51	51	49	51	52	50	48	50	49	50

(a) construa um *box-plot* para estes dados

(b) construa um diagrama ramos-e-folhas para estes dados

(c) assumindo que as temperaturas tem distribuição normal, obtenha um intervalo da confiança (95%) para temperatura média

(d) construa um intervalo de confiança (95%) para variância das temperaturas

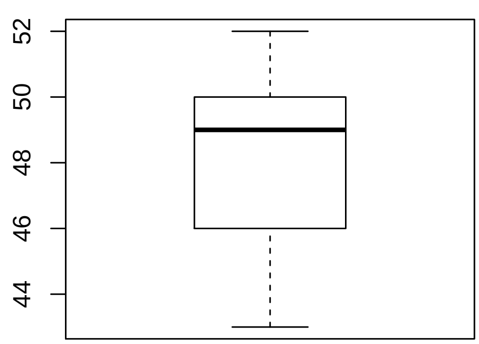
(e) considere agora que os 12 dados da primeira linha foram obtidos antes de uma mudança no processo de tratamento, e os demais depois da mudança. Faça um teste adequado para verificar se a mudança alterou a variância das temperaturas

(f) nas condições do ítem anterior faça um teste para verificar se as mudanças alteraram a média das temperaturas.

Solução:

```
> temp <- c(43,47,45,48,44,50,46,49,45,44,46,46,52,49,51,51,49,51,52,50,48,50,49,50)
```

(a)
```
> boxplot(temp)
```



(b)
```
> stem(temp)
```

```
  The decimal point is at the |

  42 | 0
  44 | 0000
  46 | 0000
  48 | 000000
  50 | 0000000
  52 | 00
```

(c) `> t.test(temp)$conf`

```
[1] 46,99 49,26
attr(,"conf.level")
[1] 0,95
```

(d)
```
> n <- length(temp)
> t.v <- var(temp)
> c((n-1)*t.v/qchisq(0.975, df=n-1), (n-1)*t.v/qchisq(0.025, df=n-1))
```

```
[1]  4,376 14,255
```

(e)
```
> temp1 <- temp[1:12]
> temp2 <- temp[-(1:12)]
> var.test(temp1, temp2)
```

```
F test to compare two variances

data:  temp1 and temp2
F = 2,8, num df = 11, denom df = 11, p-value = 0,1
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0,7971 9,6182
sample estimates:
ratio of variances
             2,769
```

(f) `> t.test(temp1, temp2, var.eq=TRUE)`

```
Two Sample t-test

data:  temp1 and temp2
t = -5,7, df = 22, p-value = 9e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5,556 -2,610
sample estimates:
mean of x mean of y
    46,08     50,17
```

---

181. A distribuição dos comprimentos dos elos de uma corrente de bicicleta é normal com média $2cm$ e variância igual a $0.01cm^2$. Para que uma corrente se ajusta a bicicleta deve ter comprimento total entre 58 e $61cm$.

(a) Qual a probabilidade de um elo escolhido ao acaso ser maior que $2,15cm$?

(b) Qual a probabilidade de uma corrente com 30 elos não se ajustar à bicicleta?

(c) E uma corrente de 29 elos?

Solução:

(a) `> 1-pnorm(2.15, mean=2, sd=sqrt(0.01))`

```
[1] 0,06681
```

(b) `> pnorm(58/30, mean=2, sd=sqrt(0.01/30)) + (1-pnorm(61/30, mean=2, sd=sqrt(0.01/30)))`

```
[1] 0,03407
```

(c) `> pnorm(58/29, mean=2, sd=sqrt(0.01/29)) + (1-pnorm(61/29, mean=2, sd=sqrt(0.01/29)))`

```
[1] 0,5
```

---

182. Sendo X uma variável seguindo o modelo Normal com média $\mu = 130$ e variância $\sigma^2 = 64$, pergunta-se:

  (a) Qual a probabilidade de obter uma média amostral maior do que 132 em uma amostra de tamanho 20?
  (b) Qual o tamanho de amostra necessário para que a probabilidade acima seja de 0.05?
  (c) Pretende-se adotar procedimentos para reduzir a variância do processo. Para quanto σ^2 deve ser reduzido para que a probabilidade acima seja de 0.10 para uma amostra de tamanho 15?
  (d) Qual a amplitude do intervalo de confiança (95%) para média caso seja retirada uma amostra de tamanho 10?
  (e) Qual deve ser o tamanho da amostra para reduzir à metade a amplitude do IC calculada no item anterior?

Solução:

---

183. Em uma instituição pública com um número muito grande de funcionários deseja-se estimar a proporção de favoráveis a uma proposta de reestruturação de cargos.

  (a) Qual a margem de erro estimada caso seja tomada uma amostra de tamanho 1000, com 95% de confiança?
  (b) Qual o tamanho de amostra para que a estimativa tenha uma margem de erro de 3% com 99% de confiança?
  (c) Qual a confiança de um intervalo com margem de erro 2% obtido com uma amostra de tamanho 500?

Solução:

---

184. Foram coletados dados de produtividade de 2 linhas de montagem A e B de uma fábrica.

  - Linha A: `54 47 53 51 55 49 54 53 48 45`
  - Linha B: `40 52 53 32 50 30 43 68 45 44`

  Teste a hipótese de igualdade das médias ($\alpha = 5\%$), supondo que as variâncias são iguais e desconhecidas.

Solução:

---

185. Seja X uma v.a. com distribuição binomial, com $n = 15$. Considere testar $H_0 : p \geq 0,5$ contra $H_1 : p < 0,5$, com região crítica (região de rejeição de H_0) $= \{0, 1, 2\}$

  (a) Calcule a probabilidade do erro tipo I.
  (b) Calcule a probabilidade do erro tipo II quando $p = 0,3$.

Solução:

---

186. Os novos operários de uma empresa são treinados a operarem uma máquina, cujo o tempo X (em horas) de aprendizado é anotado. Observou-se que X segue uma distribuição $N(25, 100)$. Uma nova técnica de ensino, que deve melhorar o aprendizado, é testada em 16 novos empregados, os quais apresentaram 20,5 horas como tempo médio de aprendizado. Usando o nível descritivo (P-valor), você diria que a nova técnica é superior à anterior?

Solução:

---

187. Os dados abaixo referem-se a meses de experiência de 10 datilógrafos e o número de erros cometidos na datilografia de um determinado texto.

Meses	1	2	3	4	6	7	8	9	10	10
Erros	30	28	24	20	18	14	13	10	7	6

Dados: $\sum_i x_i = 60$ $\sum_i x_i^2 = 460$ $\sum_i y_i = 170$ $\sum_i x_i y_i = 768$

(a) Ajuste uma reta de regressão aos dados.

(b) Faça um esboço do gráfico dos dados com a reta ajustada e comente o ajuste

(c) Obtenha o coeficiente de determinação R^2 e explique o seu significado.

Solução:

---

188. Diga se as afirmativas a seguir são verdadeiras ou falsas. No caso de *FALSA* justifique a sua resposta:

(a) quanto maior o tamanho de uma amostra maior a amplitude dos intervalos de confiança dos parâmetros.

(b) quanto maior a variância de um estimador menor a confiança na estimativa fornecida por ele.

(c) o estimador não-viciado será sempre o de menor variância.

(d) o erro tipo II é dado pela probabilidade rejeitar a hipótese H_0 quando H_1 é verdadeira.

(e) um teste mais poderoso é um teste que tem menor probabilidade de erro tipo I.

Solução:

---

189. O consumidor de um certo produto acusou o fabricante, dizendo que mais de 20% da unidades fabricadas apresentam defeito. Para embasar sua acusação, ele usou uma amostra de tamanho $n = 50$ na qual 27% eram defeituosas. Mostre se fabricante poderia refutar a acusação com um teste estatístico de hipótese. Use um nível de significância de 10%.

Solução:

$$\begin{aligned}
\text{População:}& \\
&X : \text{presença de defeito (0 - não, 1 - sim)} \\
&X : \mathrm{B}(p) \\
\text{Amostra:}& \\
&n = 50 \\
&y = \sum_{i=1}^{50} x_i = 27 \\
\text{Teste de hipótese:}& \\
H_0 : p \leq 0,20 \; vs \; H_1 &: p \geq 0,20 \\
\alpha = 0,10 &\longrightarrow z_t = 1,28 \\
&z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,27 - 0,20}{\sqrt{\frac{0,20(1-0,20)}{50}}} = 1,24
\end{aligned}$$

Conclusão: Como $z_c < z_t$, ou, equivalentemente como o *p-valor* $= 0,108 \geq \alpha$, não rejeita-se H_0 ao nível de 10% de significância, ou seja, não há evidência suficiente na amostra para acusar o fabricante.

---

190. Uma variável aleatória tem distribuição normal e desvio padrão igual a 12. Estamos testando se sua média é igual ou diferente a 20 e coletamos uma amostra de 100 valores desta variável, obtendo uma média amostral de 17,4.

(a) Formule as hipóteses para um teste estatístico de hipótese de interesse no contexto.

(b) Obtenha a região crítica (RRH_0) e dê a conclusão do teste para os níveis de 1%, 5% e 10% de significância.

Solução:

$$\begin{aligned}
&\text{População:}\\
&\quad X : \mathrm{N}(\mu, \sigma^2 = 12^2)\\
&\text{Amostra:}\\
&\quad n = 100\\
&\quad \overline{x} = \frac{\sum_{i=1}^{100} x_i}{n} = 17,4\\
&\text{Teste de hipótese:}\\
&\quad H_0 : \mu = 20 \; vs \; H_1 : \mu \neq 20\\
&\quad \alpha = 0,01; 0,05 \text{ e } 0,10 \longrightarrow z_t = 2,58, z_t = 1,96 \text{ e } z_t = 1,64\\
&\quad z_c = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{17.4 - 20}{12/\sqrt{10}} = -2,17
\end{aligned}$$

Conclusão (1%): Como $|z_c| < z_t$, ou, equivalentemente como o $p\text{-}valor = 0,0303 \geq \alpha = 0,01$, não rejeita-se H_0 ou seja, não há evidência suficiente na amostra para afirmar que a média difere de 20 ao nível de 1% de significância,

Conclusão (5%): Como $|z_c| \geq z_t$, ou, equivalentemente como o $p\text{-}valor = 0,0303 < \alpha 0,05$, rejeita-se H_0 ou seja, há evidência suficiente na amostra para afirmar que a média difere de 20 ao nível de 5% de significância,

Conclusão (10%): Como $|z_c| \geq z_t$, ou, equivalentemente como o $p\text{-}valor = 0,0303 < \alpha = 0,10$, rejeita-se H_0 ou seja, há evidência suficiente na amostra para afirmar que a média difere de 20 ao nível de 10% de significância,

---

191. A duração do "tonner"da uma impressora pode ser modelada pela distribuição normal com média 10.000 cópias e desvio padrão de 1.200 cópias. Em uma amostra de 12 impressoras a duração do "tonner"será anotada e pergunta-se a probabilidade da média destes ser:

(a) inferior a 9.000 cópias;

(b) superior a 10.500 cópias;

(c) entre 9.200 e 10.000 cópias.

Solução:

$$\begin{aligned}
&\quad X : \text{duração do tonner (em número de cópias)}\\
&\quad X \sim N(\mu = 10.000, \sigma^2 = 1.200^2)\\
&\text{amostra} : n = 12\\
&\quad \bar{X} \sim N(\mu = 10.000, \sigma^2 = 1.200^2/12)
\end{aligned}$$

(a) $P[\bar{X} < 9.000] = P[Z < \frac{9.000-10.000}{1.200/\sqrt{12}}] = 0,0019$

(b) $P[\bar{X} > 10.500] = P[Z > \frac{10.500-10.000}{1200/\sqrt{12}}] = 0,0745$

(c) $P[9.200 < \bar{X} < 10.000] = P[\frac{9.200-10.000}{1.200/\sqrt{12}} < Z < \frac{10.000-10.000}{1.200/\sqrt{12}}] = 0,4895$

Solução computacional com o sistema **R**:

```
> (qa <- pnorm(9000, m=10000, sd=1200/sqrt(12)))

[1] 0,001946

> (qb <- pnorm(10500, m=10000, sd=1200/sqrt(12), low=F))

[1] 0,07446

> (qc <- diff(pnorm(c(9200, 10000), m=10000, sd=1200/sqrt(12))))

[1] 0,4895
```

---

192. Para se ajustar uma máquina, a correia deve ter entre 60 e 62 cm de comprimento. Tendo em vista o processo de fabricação, o comprimento destas correia pode ser considerado como uma variável aleatória com distribuição normal, de média 60,7 e desvio padrão 0,8 cm. Pergunta-se:

    (a) qual a probabilidade de uma correia, escolhida ao acaso, poder ser usada na máquina?

    (b) um grande revendedor dessas correias estabelece um controle de qualidade nos lotes que compra da fábrica: ele sorteia 4 correias do lote e só aceita o lote se o comprimento médio estiver dentro do tamanho aceito pela máquina. Calcule a probabilidade de aceitação do lote.

Solução:

$$\begin{aligned} X &: \text{comprimento da correia} \\ X &\sim \mathrm{N}(60,7; 0.8^2) \\ \bar{X}_4 &\sim \mathrm{N}(60,7; 0.8^2/4) \end{aligned}$$

(a) $P[\text{correia poder ser usada}] = P[60 < X < 62] = P[\frac{60-60{,}7}{0.8} < Z < \frac{62-60.7}{0.8}] = 0,757$

(b) $P[\text{lote ser aceito}] = P[60 < \bar{X}_4 < 62] = P[\frac{60-60{,}7}{0.8/\sqrt{4}} < Z < \frac{62-60.7}{0.8/\sqrt{4}}] = 0,959$
