

Gabarito da 2ª Lista de Exercícios - MAE1512

Professora: SILVIA NAGIB ELIAN

Monitores: DANIELLE VELLOSO E RODRIGO PASSOS MARTINS

Exercício 1

A população dos pesos dos passageiros de uma companhia de aviação tem média de 150 libras e desvio padrão de 25 libras. Denotando como X a variável aleatória que descreve o comportamento dos pesos dessa população, temos que:

$$E(X) = 150 \text{ e } DP(X) = 50$$

Utilizando o **Teorema do Limite Central**, podemos dizer que a variável aleatória Y que descreve o comportamento do peso da população de 50 passageiros é tal que:

$$\sum_{i=1}^{50} X_i = Y \sim N(50.E(X), 50.VAR(X)) \Leftrightarrow Y \sim N(50.150, 50.25^2) \Leftrightarrow Y \sim N(7500, 31250)$$

Assim, a probabilidade de 50 passageiros ultrapassarem o limite do avião de 7800 libras é dada por:

$$P(Y > 7800) = P\left(Z_Y > \frac{7800 - 7500}{\sqrt{31250}}\right) \cong P(Z_Y > 1,70) = 0,5 - P(0 < Z_Y < 1,70) = 0,5 - 0,45543 = 0,04457 = 4,457\%$$

Logo, a probabilidade de uma lotação de 50 passageiros ultrapassar o limite de 7800 libras é de 4,457%.

Agora, se quisermos reduzir para 0,01 a chance de sobrecarga, a capacidade deve ser k , tal que:

$$P(Y > k) = 0,01 \Leftrightarrow P\left(Z_Y > \frac{k - 7500}{\sqrt{31250}}\right) = 0,01 \Leftrightarrow 0,5 - P\left(0 < Z_Y < \frac{k - 7500}{\sqrt{31250}}\right) = 0,01 \Leftrightarrow P\left(0 < Z_Y < \frac{k - 7500}{\sqrt{31250}}\right) = 0,49$$

Da Tabela Normal Padrão, temos que:

$$\frac{k - 7500}{\sqrt{31250}} = 2,33 \Leftrightarrow k = 2,33 \cdot \sqrt{31250} + 7500 = 7911,89$$

Portanto, a capacidade do avião deve aumentar em $7911,89 - 7800 = 111,89$ libras. ■

Exercício 2

Vamos supor que um agrimensor está procurando determinar a área de um campo retangular, em que as medidas do comprimento, X , e da largura, Y , são variáveis aleatórias independentes que flutuam em redor dos verdadeiros valores, de acordo com as seguintes distribuições de probabilidade:

- $X = \{8; 10; 11\}$
- $P(X = x) = \{0,25; 0,25; 0,5\}$

- $Y = \{4; 6\}$
- $P(Y = y) = \{0,5; 0,5\}$

A área calculada $A = XY$ é, naturalmente, uma variável aleatória, e é usada para estimar a área verdadeira.

Vamos considerar que o verdadeiro comprimento e a verdadeira largura são 10 e 5, respectivamente.

a)

Vamos verificar se X é um estimador não viesado do verdadeiro comprimento:

$$E(X) = \sum_{i=1}^3 x_i \cdot P(X = x_i) = 8 \cdot 0,25 + 10 \cdot 0,25 + 11 \cdot 0,5 = 2 + 2,5 + 5,5 = 10$$

Como $E(X)$ é igual ao verdadeiro comprimento, X é um estimador não viesado.

b)

Vamos verificar se Y é um estimador não viesado da verdadeira largura:

$$E(Y) = \sum_{i=1}^2 y_i \cdot P(Y = y_i) = 4 \cdot 0,5 + 6 \cdot 0,5 = 2 + 3 = 5$$

Como $E(Y)$ é igual à verdadeira largura, Y é um estimador não viesado.

c)

Vamos verificar se A é um estimador não viesado da verdadeira área:

$$E(A) = E(XY)$$

Como X e Y são independentes, $E(XY) = E(X).E(Y)$, logo:

$$E(A) = E(XY) = E(X).E(Y) = 10.5 = 50$$

Como $E(A)$ é igual a verdadeira área ($10.5 = 50$), A é um estimador não viesado.

d)

Os possíveis valores de A e as probabilidades deles são:

- $A = \{32; 40; 44; 48; 60; 66\}$
- $P(A = a) = \{0,125; 0,125; 0,25; 0,125; 0,125; 0,25\}$

Com isso, calculamos a variância de A por:

$$VAR(A) = \sum_{i=1}^6 (a_i - E(A))^2 . p_i = (32 - 50)^2 . 0,125 + (40 - 50)^2 . 0,125 + \dots + (66 - 50)^2 . 0,25 = 139 \blacksquare$$

Exercício 3

Vamos supor que dois economistas estimem μ (despesa familiar média com alimentação) com dois estimadores não viesados $\hat{\mu}_1$ e $\hat{\mu}_2$ (estatisticamente independentes). Assim,

$$E(\mu_1) = E(\mu_2) = \mu$$

O segundo economista é menos cuidadoso que o primeiro, de maneira que o desvio padrão de $\hat{\mu}_2$ é cinco vezes maior do que o de $\hat{\mu}_1$. Logo,

$$DP(\mu_2) = 5.DP(\mu_1) \Leftrightarrow VAR(\mu_2) = 25.VAR(\mu_1)$$

Solicitado a informar sobre como combinar $\hat{\mu}_1$ e $\hat{\mu}_2$ para obter uma estimativa global publicável, um grupo de estatísticos apresentou as seguintes propostas:

- **1)** $\hat{W}_1 = \frac{1}{2} \cdot (\hat{\mu}_1 + \hat{\mu}_2)$
- **2)** $\hat{W}_2 = \left(\frac{4}{5} \hat{\mu}_1 + \frac{1}{5} \hat{\mu}_2 \right)$
- **3)** $\hat{W}_3 = \left(\frac{5}{6} \hat{\mu}_1 + \frac{1}{6} \hat{\mu}_2 \right)$
- **4)** $\hat{W}_4 = \hat{\mu}_1$

a)

Vamos calcular quais dos estimadores são não viesados:

$$1) E(\hat{W}_1) = E\left(\frac{1}{2} \cdot (\hat{\mu}_1 + \hat{\mu}_2)\right) = \frac{1}{2} \cdot E(\hat{\mu}_1 + \hat{\mu}_2) = \frac{1}{2} \cdot [E(\hat{\mu}_1) + E(\hat{\mu}_2)] = \frac{1}{2} \cdot (\mu + \mu) = \frac{2\mu}{2} = \mu$$

$$2) E(\hat{W}_2) = E\left(\frac{4}{5} \hat{\mu}_1 + \frac{1}{5} \hat{\mu}_2\right) = \frac{4}{5} \cdot E(\hat{\mu}_1) + \frac{1}{5} \cdot E(\hat{\mu}_2) = \frac{4\mu}{5} + \frac{1\mu}{5} = \mu$$

$$3) E(\hat{W}_3) = E\left(\frac{5}{6} \hat{\mu}_1 + \frac{1}{6} \hat{\mu}_2\right) = \frac{5}{6} \cdot E(\hat{\mu}_1) + \frac{1}{6} \cdot E(\hat{\mu}_2) = \frac{5\mu}{6} + \frac{1\mu}{6} = \mu$$

$$4) E(\hat{W}_4) = E(\hat{\mu}_1) = \mu$$

Portanto, todos os estimadores são não viesados.

b)

Vamos calcular a variância de cada um dos estimadores. Para tal, vamos assumir que:

$$VAR(\hat{\mu}_1) = \sigma^2$$

Assim,

$$1) VAR(\hat{W}_1) = VAR\left(\frac{1}{2} \cdot (\hat{\mu}_1 + \hat{\mu}_2)\right) = \frac{1}{4} \cdot VAR(\hat{\mu}_1 + \hat{\mu}_2) = \frac{1}{4} \cdot [VAR(\hat{\mu}_1) + VAR(\hat{\mu}_2)] = \frac{1}{4} \cdot (\sigma^2 + 25\sigma^2) = \frac{26}{4}\sigma^2 = 6,5\sigma^2$$

$$2) VAR(\hat{W}_2) = VAR\left(\frac{4}{5}\hat{\mu}_1 + \frac{1}{5}\hat{\mu}_2\right) = \frac{16}{25} \cdot VAR(\hat{\mu}_1) + \frac{1}{25} \cdot VAR(\hat{\mu}_2) = \frac{16\sigma^2}{25} + \frac{25\sigma^2}{25} = \frac{41\sigma^2}{25} = 1,64\sigma^2$$

$$3) VAR(\hat{W}_3) = VAR\left(\frac{5}{6}\hat{\mu}_1 + \frac{1}{6}\hat{\mu}_2\right) = \frac{25}{36} \cdot VAR(\hat{\mu}_1) + \frac{1}{36} \cdot VAR(\hat{\mu}_2) = \frac{25\sigma^2}{36} + \frac{25\sigma^2}{36} = \frac{50\sigma^2}{36} = 1,36\sigma^2$$

$$4) VAR(\hat{W}_4) = VAR(\hat{\mu}_1) = \sigma^2$$

Assim, em ordem decrescente de eficiência, temos do estimador com menor variância ao de maior variância:

- 1) $\hat{W}_4 = \hat{\mu}_1$
- 2) $\hat{W}_3 = \left(\frac{5}{6}\hat{\mu}_1 + \frac{1}{6}\hat{\mu}_2\right)$
- 3) $\hat{W}_2 = \left(\frac{4}{5}\hat{\mu}_1 + \frac{1}{5}\hat{\mu}_2\right)$
- 4) $\hat{W}_1 = \frac{1}{2} \cdot (\hat{\mu}_1 + \hat{\mu}_2)$

Podemos perceber que quanto maior a composição de $\hat{\mu}_1$ no estimador, mais eficiente ele é.

Exercício 4

Um banco está monitorando a duração do tempo que seus clientes ficam nas filas de caixa até serem atendidos. Especificamente, deseja estimar a proporção p de clientes que demoram 5 minutos ou mais para serem atendidos. Uma amostra aleatória de clientes desse banco será coletada, e o tempo de espera será registrado.

a)

Vamos calcular o tamanho da amostra, para que o erro da estimativa seja no máximo 0,04 com nível de confiança de 0,92:

$$n = \left(\frac{Z}{\varepsilon}\right)^2 \cdot p \cdot (1 - p) = \left(\frac{1,75}{0,04}\right)^2 \cdot 0,5 \cdot (1 - 0,5)$$

Assumimos $p = 0,5$ por ser o valor que fornece maior tamanho amostral (conservador):

$$n = \left(\frac{1,75}{0,04}\right)^2 \cdot 0,5 \cdot (1 - 0,5) \cong 478,52 \cong 479$$

Portanto, o tamanho amostral para que a estimação corresponda às condições anteriores é $n = 479$.

b)

A direção do banco sabe que essa proporção p não ultrapassa 26%. Com essa informação, vamos recalculer o tamanho da amostra:

$$n = \left(\frac{Z}{\varepsilon}\right)^2 \cdot p \cdot (1 - p) = \left(\frac{1,75}{0,04}\right)^2 \cdot 0,26 \cdot (1 - 0,26) \cong 368,26 \cong 369$$

Assim, por conta da nova informação, diminuimos a amostra de $n = 479$ para $n = 369$

c)

Uma amostra de 50 clientes forneceu as seguintes medidas desse tempo (em minutos):

25, 28, 4, 5, 15, 12, 34, 6, 8, 42, 16, 7, 2, 6, 5, 8, 15, 4, 8, 9, 14, 8, 3, 9, 4, 13, 8, 9, 17, 8, 10, 15, 4, 6, 8, 7, 7, 12, 9, 6, 13, 14, 8, 4, 7, 9, 16, 20, 11, 21

A estimativa pontual de p é dada por:

$$\hat{p} = \frac{\text{n}^\circ \text{ de clientes que demoraram 15 minutos ou mais}}{\text{total de clientes da amostra}} = \frac{12}{50} = 0,24$$

Agora, vamos construir um intervalo de confiança otimista de 92% para p :

$$IC(p; 92\%) = \left[\hat{p} \pm Z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right] = \left[0,24 \pm 1,75 \cdot \sqrt{\frac{0,24 \cdot (1 - 0,24)}{50}} \right] \cong [0,1343; 0,3457]$$

O erro amostra dessa estimativa é dado por:

$$\varepsilon = 1,75 \cdot \sqrt{\frac{0,24 \cdot (1 - 0,24)}{50}} \cong 0,1057 = 10,57\%$$

Exercício 7

Numa pesquisa de mercado deseja-se analisar a preferência da população de uma cidade em relação a um determinado produto, com base em uma amostra aleatória simples com reposição de 300 indivíduos.

a)

Vamos estimar a probabilidade de que a estimativa pontual dessa proporção não difira do verdadeiro valor em mais de 0,001:

$$P(|\hat{p} - p| < 0,001) = P(-0,001 < \hat{p} - p < 0,001)$$

Como a \hat{p} já está substituído pela sua média, basta dividirmos pelo desvio padrão para usarmos a Normal Padrão, descrita por Z :

$$P(-0,001 < \hat{p} - p < 0,001) = P\left(\frac{-0,001}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < Z_{\hat{p}} < \frac{0,001}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\right) = P\left(\frac{-0,001\sqrt{300}}{\sqrt{0,5(1-0,5)}} < Z_{\hat{p}} < \frac{0,001\sqrt{300}}{\sqrt{0,5(1-0,5)}}\right) \cong$$

$$\cong P(-0,03 < Z_{\hat{p}} < 0,03) = 2.P(0 < Z_{\hat{p}} < 0,03) = 2.0,01197 = 0,02394 = 2,394\%$$

Portanto, a probabilidade de que a estimativa pontual dessa proporção não difira do verdadeiro valor em mais de 0,001 é de 0,02394.

b)

Vamos verificar o que deve ser feito para obter uma estimativa pontual dessa proporção que não difira do valor verdadeiro em mais de 0,0005 com probabilidade 0,95. A ideia é calcular o tamanho amostral que assegure esses parâmetros:

$$n = \left(\frac{Z}{\varepsilon}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) = \left(\frac{1,96}{0,0005}\right)^2 \cdot 0,5 \cdot (1 - 0,5) = 3841600$$

Portanto, para se obter uma estimativa pontual dessa proporção que não difira do valor verdadeiro em mais de 0,0005 com probabilidade 0,95, precisamos de uma amostra de 3841600, ou seja, um aumento de 3841300 pessoas da população analisada.

c)

Vamos considerar que sabemos que dos 300 indivíduos, 180 preferem o referido produto. Logo, é fácil determinar que

$$\hat{p} = \frac{180}{300} = 0,6$$

Vamos determinar intervalos com 95% de confiança para a proporção da população que prefere o produto em estudo, sendo um otimista e outro conservador:

Otimista:

$$IC(p; 95\%) = \left[\hat{p} \pm Z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right] = \left[0,6 \pm 1,96 \cdot \sqrt{\frac{0,6 \cdot (1 - 0,6)}{300}} \right] \cong [0,5446; 0,6554]$$

Conservador:

$$IC(p; 95\%) = \left[\hat{p} \pm Z \cdot \sqrt{\frac{0,5 \cdot (1 - 0,5)}{n}} \right] = \left[0,6 \pm 1,96 \cdot \sqrt{\frac{0,25}{300}} \right] \cong [0,5434; 0,6566]$$

Exercício 8

Utilizando o programa **R**, vamos gerar 100 réplicas de uma distribuição binomial, com parâmetros $n = 90$ e $p = 0,75$ (probabilidade de sucesso):

```
# Para gerar 100 réplicas da binomial n = 90, p = 0,75
n_replicas <- 100
replicas <- rbinom(n_replicas, 90, 0.75)
```

Para cada réplica estimamos a probabilidade de sucesso:

```
# Para estimar a probabilidade de sucesso
p_sucesso <- replicas/90
```

```
# para visualizar
p_sucesso
```

```
## [1] 0.7222222 0.7111111 0.7555556 0.7444444 0.7444444 0.7666667 0.7444444
## [8] 0.7555556 0.7777778 0.8000000 0.7111111 0.8111111 0.6555556 0.6666667
## [15] 0.7777778 0.8111111 0.7555556 0.7888889 0.7333333 0.7111111 0.7888889
## [22] 0.7222222 0.7111111 0.6444444 0.7888889 0.8222222 0.8222222 0.7666667
## [29] 0.8000000 0.8000000 0.8000000 0.7222222 0.7444444 0.7333333 0.7666667
## [36] 0.7444444 0.7777778 0.7666667 0.6888889 0.8000000 0.8333333 0.7555556
## [43] 0.8111111 0.7111111 0.7666667 0.6888889 0.6888889 0.7555556 0.7666667
## [50] 0.7111111 0.7888889 0.8111111 0.7111111 0.7666667 0.7666667 0.7222222
## [57] 0.7555556 0.7666667 0.7555556 0.7111111 0.8111111 0.8222222 0.7666667
## [64] 0.7666667 0.7111111 0.7000000 0.7666667 0.7000000 0.7222222 0.7888889
## [71] 0.7888889 0.8000000 0.8444444 0.7333333 0.7777778 0.6888889 0.7444444
## [78] 0.7777778 0.7666667 0.8333333 0.7888889 0.7111111 0.6777778 0.7222222
## [85] 0.7222222 0.7555556 0.6777778 0.7111111 0.7666667 0.7555556 0.6666667
## [92] 0.7888889 0.7555556 0.7777778 0.7000000 0.6888889 0.7777778 0.7666667
## [99] 0.7222222 0.7888889
```

Após, construímos o intervalo de confiança (Limite inferior, Limite superior), com $\gamma = 0,95$ e visualizamos os resultados:

```
# Para construir Intervalo de confiança de 95% para cada réplica de tamanho 90
erro <- 1.96*sqrt(p_sucesso*(1-p_sucesso)/90)
Lim_inf <- p_sucesso - erro
Lim_sup <- p_sucesso + erro

# Para visualizar os resultados
cbind(p_sucesso, Lim_inf, Lim_sup)
```

##		p_sucesso	Lim_inf	Lim_sup
##	[1,]	0.7222222	0.6296845	0.8147600
##	[2,]	0.7111111	0.6174695	0.8047527
##	[3,]	0.7555556	0.6667668	0.8443443
##	[4,]	0.7444444	0.6543302	0.8345587
##	[5,]	0.7444444	0.6543302	0.8345587
##	[6,]	0.7666667	0.6792838	0.8540496
##	[7,]	0.7444444	0.6543302	0.8345587
##	[8,]	0.7555556	0.6667668	0.8443443
##	[9,]	0.7777778	0.6918851	0.8636705
##	[10,]	0.8000000	0.7173591	0.8826409
##	[11,]	0.7111111	0.6174695	0.8047527
##	[12,]	0.8111111	0.7302428	0.8919794
##	[13,]	0.6555556	0.5573810	0.7537302
##	[14,]	0.6666667	0.5692735	0.7640598
##	[15,]	0.7777778	0.6918851	0.8636705
##	[16,]	0.8111111	0.7302428	0.8919794
##	[17,]	0.7555556	0.6667668	0.8443443
##	[18,]	0.7888889	0.7045752	0.8732026
##	[19,]	0.7333333	0.6419704	0.8246962
##	[20,]	0.7111111	0.6174695	0.8047527
##	[21,]	0.7888889	0.7045752	0.8732026
##	[22,]	0.7222222	0.6296845	0.8147600
##	[23,]	0.7111111	0.6174695	0.8047527
##	[24,]	0.6444444	0.5455479	0.7433410
##	[25,]	0.7888889	0.7045752	0.8732026
##	[26,]	0.8222222	0.7432329	0.9012115
##	[27,]	0.8222222	0.7432329	0.9012115
##	[28,]	0.7666667	0.6792838	0.8540496
##	[29,]	0.8000000	0.7173591	0.8826409
##	[30,]	0.8000000	0.7173591	0.8826409
##	[31,]	0.8000000	0.7173591	0.8826409
##	[32,]	0.7222222	0.6296845	0.8147600
##	[33,]	0.7444444	0.6543302	0.8345587
##	[34,]	0.7333333	0.6419704	0.8246962
##	[35,]	0.7666667	0.6792838	0.8540496
##	[36,]	0.7444444	0.6543302	0.8345587
##	[37,]	0.7777778	0.6918851	0.8636705
##	[38,]	0.7666667	0.6792838	0.8540496
##	[39,]	0.6888889	0.5932428	0.7845350
##	[40,]	0.8000000	0.7173591	0.8826409
##	[41,]	0.8333333	0.7563373	0.9103294
##	[42,]	0.7555556	0.6667668	0.8443443
##	[43,]	0.8111111	0.7302428	0.8919794
##	[44,]	0.7111111	0.6174695	0.8047527
##	[45,]	0.7666667	0.6792838	0.8540496
##	[46,]	0.6888889	0.5932428	0.7845350
##	[47,]	0.6888889	0.5932428	0.7845350
##	[48,]	0.7555556	0.6667668	0.8443443
##	[49,]	0.7666667	0.6792838	0.8540496
##	[50,]	0.7111111	0.6174695	0.8047527
##	[51,]	0.7888889	0.7045752	0.8732026
##	[52,]	0.8111111	0.7302428	0.8919794
##	[53,]	0.7111111	0.6174695	0.8047527
##	[54,]	0.7666667	0.6792838	0.8540496
##	[55,]	0.7666667	0.6792838	0.8540496
##	[56,]	0.7222222	0.6296845	0.8147600
##	[57,]	0.7555556	0.6667668	0.8443443

```
## [58,] 0.7666667 0.6792838 0.8540496
## [59,] 0.7555556 0.6667668 0.8443443
## [60,] 0.7111111 0.6174695 0.8047527
## [61,] 0.8111111 0.7302428 0.8919794
## [62,] 0.8222222 0.7432329 0.9012115
## [63,] 0.7666667 0.6792838 0.8540496
## [64,] 0.7666667 0.6792838 0.8540496
## [65,] 0.7111111 0.6174695 0.8047527
## [66,] 0.7000000 0.6053230 0.7946770
## [67,] 0.7666667 0.6792838 0.8540496
## [68,] 0.7000000 0.6053230 0.7946770
## [69,] 0.7222222 0.6296845 0.8147600
## [70,] 0.7888889 0.7045752 0.8732026
## [71,] 0.7888889 0.7045752 0.8732026
## [72,] 0.8000000 0.7173591 0.8826409
## [73,] 0.8444444 0.7695649 0.9193240
## [74,] 0.7333333 0.6419704 0.8246962
## [75,] 0.7777778 0.6918851 0.8636705
## [76,] 0.6888889 0.5932428 0.7845350
## [77,] 0.7444444 0.6543302 0.8345587
## [78,] 0.7777778 0.6918851 0.8636705
## [79,] 0.7666667 0.6792838 0.8540496
## [80,] 0.8333333 0.7563373 0.9103294
## [81,] 0.7888889 0.7045752 0.8732026
## [82,] 0.7111111 0.6174695 0.8047527
## [83,] 0.6777778 0.5812269 0.7743287
## [84,] 0.7222222 0.6296845 0.8147600
## [85,] 0.7222222 0.6296845 0.8147600
## [86,] 0.7555556 0.6667668 0.8443443
## [87,] 0.6777778 0.5812269 0.7743287
## [88,] 0.7111111 0.6174695 0.8047527
## [89,] 0.7666667 0.6792838 0.8540496
## [90,] 0.7555556 0.6667668 0.8443443
## [91,] 0.6666667 0.5692735 0.7640598
## [92,] 0.7888889 0.7045752 0.8732026
## [93,] 0.7555556 0.6667668 0.8443443
## [94,] 0.7777778 0.6918851 0.8636705
## [95,] 0.7000000 0.6053230 0.7946770
## [96,] 0.6888889 0.5932428 0.7845350
## [97,] 0.7777778 0.6918851 0.8636705
## [98,] 0.7666667 0.6792838 0.8540496
## [99,] 0.7222222 0.6296845 0.8147600
## [100,] 0.7888889 0.7045752 0.8732026
```

a)

Vamos destacar os intervalos que não contêm o verdadeiro valor de p :

```
# para contar quantos IC não contêm o verdadeiro valor 0,75
aux1 <- ifelse(Lim_inf < 0.75, 1, 0)
aux2 <- ifelse(Lim_sup < 0.75, 1, 0)
```

```
indicadora <- ifelse(aux1 == aux2, 1, 0)
sum(indicadora)
```

```
## [1] 4
```

São 4 intervalos que não contém o verdadeiro valor de p .

b)

Vamos calcular a proporção de intervalos que contém o verdadeiro valor de p , dentre os 100 gerados:

$$\text{Proporção: } \frac{4}{100} = 0,04 = 4\%$$

O esperado é que $1 - \gamma = 5\%$ dos intervalos não contesse o verdadeiro valor de p . O resultado observado por nós de 4% é condizente com o esperado, já que o γ da simulação foi de 95%, que significa que se pudéssemos reproduzir o experimento muitas vezes, 95% delas conteria o verdadeiro valor de p .

c)

Agora, vamos repetir o que foi feito em b) e c), agora com 1000 intervalos gerados:

```
# Para gerar 1000 réplicas da binomial n = 90, p = 0,75
n_replicas <- 1000
replicas <- rbinom(n_replicas, 90, 0.75)

# Para estimar a probabilidade de sucesso
p_sucesso <- replicas/90

# Para construir Intervalo de confiança de 95% para cada réplica de tamanho 90
erro <- 1.96*sqrt(p_sucesso*(1-p_sucesso)/90)
Lim_inf <- p_sucesso - erro
Lim_sup <- p_sucesso + erro

# para contar quantos IC não contêm o verdadeiro valor 0,75
aux1 <- ifelse(Lim_inf < 0.75, 1, 0)
aux2 <- ifelse(Lim_sup < 0.75, 1, 0)
indicadora <- ifelse(aux1 == aux2, 1, 0)
sum(indicadora)
```

```
## [1] 57
```

São 57 intervalos que não contém o verdadeiro valor de p .

Vamos calcular a proporção de intervalos que contêm o verdadeiro valor de p , dentre os 1000 gerados:

$$\text{Proporção: } \frac{57}{1000} = 0,057 = 5,7\%$$

O esperado é que $1 - \gamma = 5\%$ dos intervalos não contesse o verdadeiro valor de p . O resultado observado por nós de 5,7% é condizente com o esperado (apesar de acima dos 5%), já que o γ da simulação foi de 95%, que significa que se pudéssemos reproduzir o experimento muitas vezes, 95% delas conteria o verdadeiro valor de p .

Pela teoria, conforme vai se aumentando o número de réplicas do experimento, mais próximo de 5% será a quantidade de intervalos que não contém o valor verdadeiro de p .

Exercício 9

Em uma população de homens com mais de 55 anos de determinado município, constatou-se, a partir de uma amostra aleatória de 50 homens que 30% apresentaram pressão arterial fora do intervalo apropriado para essa faixa etária. Um ano após a implementação de um programa de exercícios físicos, observou-se em uma nova amostra de 50 homens que 18% apresentavam pressão fora do referido intervalo.

a)

Deseja-se saber entre que limites estava a verdadeira proporção p de homens com pressão fora do intervalo apropriado, antes e após a implementação do programa de exercícios. Vamos utilizar um coeficiente de confiança de 95%, logo:

Antes da implementação do programa:

$$IC(p; 95\%) = \left[\hat{p} \pm Z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right] = \left[0,3 \pm 1,96 \cdot \sqrt{\frac{0,3 \cdot (1 - 0,3)}{50}} \right] \cong [0,1730; 0,4270]$$

Após a implementação do programa:

$$IC(p; 95\%) = \left[\hat{p} \pm Z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right] = \left[0,18 \pm 1,96 \cdot \sqrt{\frac{0,18 \cdot (1 - 0,18)}{50}} \right] \cong [0,0735; 0,2865]$$

b)

O erro máximo associado a cada uma das estimativas é dado por:

Antes da implementação do programa:

$$\varepsilon = Z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} = 1,96 \cdot \sqrt{\frac{0,3 \cdot (1 - 0,3)}{50}} \cong 0,1270$$

Após a implementação do programa:

$$\varepsilon = Z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} = 1,96 \cdot \sqrt{\frac{0,18 \cdot (1 - 0,18)}{50}} \cong 0,1065$$

c)

Num trabalho futuro, o tamanho amostral necessário para obter um intervalo de 95% de confiança para a proporção p de homens com pressão arterial fora do intervalo apropriado, com um erro amostral máximo de 0,05, considerando a última estimativa obtida para p , ou seja, $p = 18\%$, é dado por:

$$n = \left(\frac{Z}{\varepsilon}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) = \left(\frac{1,96}{0,05}\right)^2 \cdot 0,18 \cdot (1 - 0,18) \cong 226,8 = 227$$

Exercício 36 (Seção 7.5)

Vamos estudar a simulação da distribuição amostral de \bar{X} .

a)

Vamos gerar 100 amostras de tamanho 30 de uma $N(200, 5^2)$:

```
# Gerando 100 amostras de tamanho 30 de N(200, 25) no R
n_amostras <- 100
amostras <- matrix(nrow = n_amostras, ncol = 30)

for(i in 1:n_amostras){
  amostras[i, 1:30] <- rnorm(n = 30, mean = 200, sd = 5)
}
```

Agora, vamos calcular a média de cada amostra:

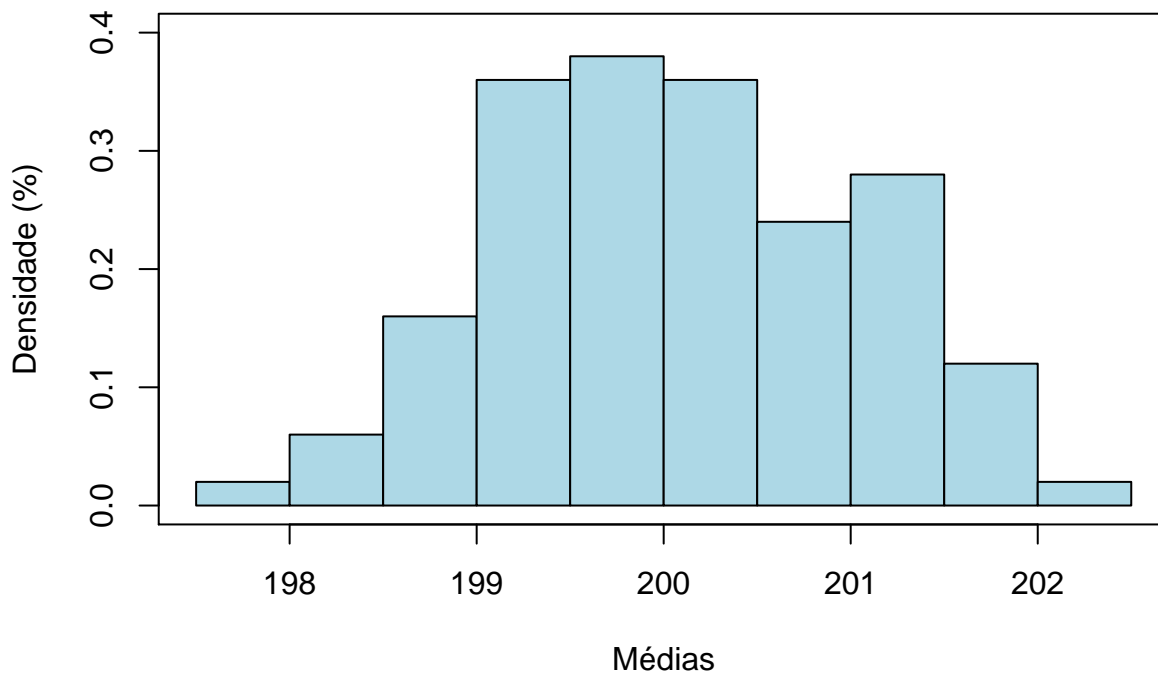
```
# Calculando a média de cada amostra no R
media_amostras <- c()

for(i in 1:n_amostras){
  media_amostras[i] <- mean(amostras[i, ])
}
```

Agora, vamos montar um histograma para a média de cada amostra:

```
# Construindo o histograma no R
hist(media_amostras,
     main = "Histograma das 100 amostras",
     xlab = "Médias", ylab = "Densidade (%)",
     xlim = c(197.5, 202.5), ylim = c(0, 0.4),
     freq = F, col = "lightblue")
box()
```

Histograma das 100 amostras



O histograma apresenta um comportamento de uma curva gaussiana ao redor da média original das amostras, que era de 200. Assim, o gráfico se assemelha a uma distribuição Normal com média 200.

Vamos verificar as medidas descritivas dessas 100 médias:

```
# Medidas descritivas das 100 médias amostrais no R  
summary(media_amostras) # Mínimo, Quartis, Média e Máximo
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
## 197.8 199.4   200.1   200.1  200.7   202.3
```

```
sd(media_amostras) # Desvio padrão amostral
```

```
## [1] 0.9521953
```

Observando as medidas descritivas, temos 100 valores centrados no 200 de maneira simétrica, com uma amplitude de valores baixa. Além disso, podemos observar um desvio-padrão amostral menor que 1, o que nos indica uma pouca variabilidade em torno do 200.

b)

Vamos gerar 100 amostras de uma $Bin(n = 50; p = 0,45)$:

```
# Gerando 100 amostras de Bin(50; 0,45) no R
n_amostras <- 100
amostras <- matrix(nrow = n_amostras, ncol = 50)

for(i in 1:n_amostras){
  amostras[i, 1:50] <- rbinom(n = 50, size = 50, p = 0.45)
}
```

Agora, vamos calcular a média de cada amostra:

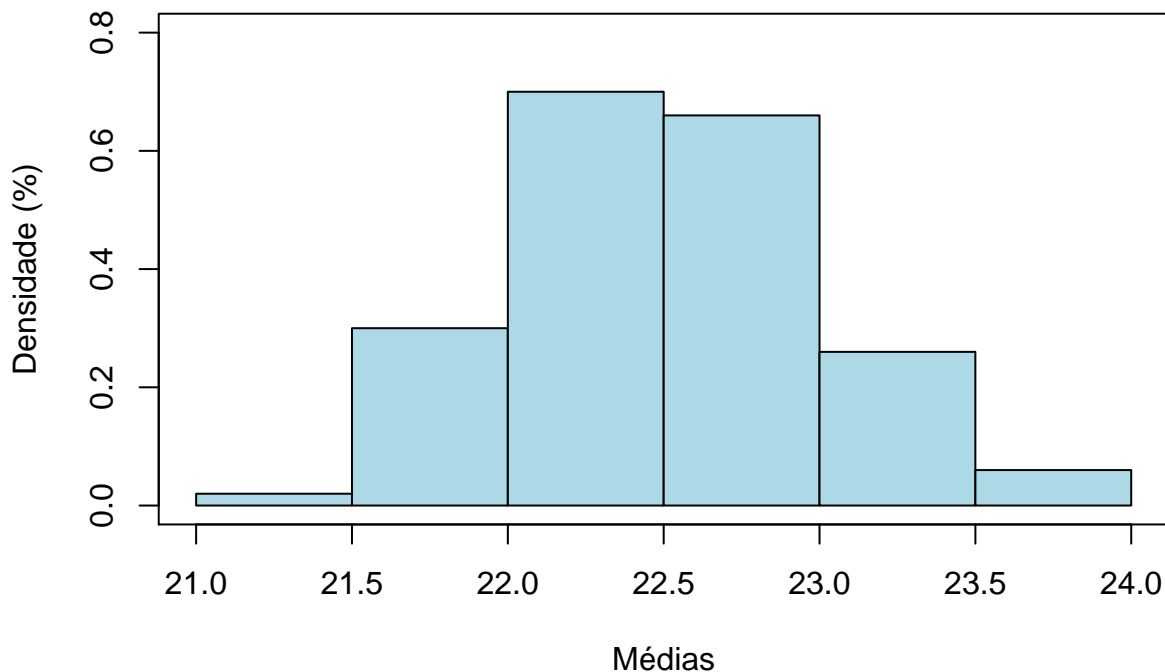
```
# Calculando a média de cada amostra no R
media_amostras <- c()

for(i in 1:n_amostras){
  media_amostras[i] <- mean(amostras[i, ])
}
```

Agora, vamos montar um histograma para a média de cada amostra:

```
# Construindo o histograma no R
hist(media_amostras,
      main = "Histograma das 100 amostras",
      xlab = "Médias", ylab = "Densidade (%)",
      ylim = c(0, 0.8),
      freq = F, col = "lightblue")
box()
```

Histograma das 100 amostras



O histograma apresenta um comportamento de uma curva gaussiana ao redor da média original das amostras, que era de $np = 50,0, 45 = 22,5$. Assim, o gráfico se assemelha a uma distribuição Normal com média 22,5.

Vamos verificar as medidas descritivas dessas 100 médias:

```
# Medidas descritivas das 100 médias amostrais no R  
summary(media_amostras) # Mínimo, Quartis, Média e Máximo
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
## 21.30  22.16   22.50   22.50  22.82   23.66
```

```
sd(media_amostras) # Desvio padrão amostral
```

```
## [1] 0.4923717
```

Observando as medidas descritivas, temos 100 valores centrados no 22,5 de maneira simétrica, com uma amplitude de valores baixa. Além disso, podemos observar um desvio-padrão amostral menor que 0,5, o que nos indica uma pouca variabilidade em torno do 22,5.

Isso ratifica o que era esperado de uma média de Binomiais: é razoável considerar uma aproximação para a distribuição Normal quando se trata da média amostral vinda de uma Binomial. Então, pelo Teorema do Limite Central, visualizamos que:

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Observamos que \bar{X} é um estimador de para p , a proporção populacional, e, desse modo, para n suficientemente grande podemos considerar a distribuição amostral de $\hat{p} = \bar{X}$ como aproximadamente Normal:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

O que podemos representar como

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

Exercício 39 (Seção 7.5)

Vamos simular a coleta de 80 observações de uma variável $N(10, 25)$ 100 vezes:

Vamos gerar 100 amostras de tamanho 30 de uma $N(200, 5^2)$:

```
# Gerando 100 amostras de tamanho 80 de N(10, 25) no R
n_amostras <- 100
amostras <- matrix(nrow = n_amostras, ncol = 80)

for(i in 1:n_amostras){
  amostras[i, 1:80] <- rnorm(n = 80, mean = 10, sd = 5)
}
```

Agora, vamos calcular a média de cada amostra:

```
# Calculando a média de cada amostra no R
media_amostras <- c()

for(i in 1:n_amostras){
  media_amostras[i] <- mean(amostras[i, ])
}
```

Com a média de cada amostra, vamos construir um intervalo de confiança de 92% para a média e verificar quantos deles contêm a verdadeira média de 10. Lembrando que,

$$IC(\mu; 92\%) = \left[\hat{\mu} \pm Z \cdot \sqrt{\frac{\hat{\sigma}^2}{n}} \right]$$

```
# Para construir Intervalo de confiança de 92% para cada amostra de tamanho 80
erro <- 1.75*sd(media_amostras)
Lim_inf <- media_amostras - erro
Lim_sup <- media_amostras + erro

# para contar quantos IC não contêm o verdadeiro valor 10
aux1 <- ifelse(Lim_inf < 10, 1, 0)
aux2 <- ifelse(Lim_sup < 10, 1, 0)
indicadora <- ifelse(aux1 == aux2, 1, 0)
sum(indicadora)
```

```
## [1] 6
```

Temos que apenas 6 intervalos dos 100 não contêm o verdadeiro valor da média, que é 10. Esse é resultado esperado da simulação, uma vez que um intervalo de confiança de 92% nessa situação pode levar até, teoricamente, 8 intervalos que não contenham a verdadeira média.