RESOLUÇÃO ASSISTIDA DE PROBLEMAS

CAPÍTULO 8 – AMOSTRAGEM ALEATÓRIA. DISTRIBUIÇÕES POR AMOSTRAGEM

PROBLEMA 8.1

APOIOS À RESOLUÇÃO

Alínea (i)

Apoio 1 (apenas o resultado)

Valor esperado: 3.9

Desvio padrão: 1.302. ■

Apoio 2 (sugestão)

A variável aleatória "vendas quinzenais" resulta da soma de duas variáveis aleatórias independentes. ■

Apoio 3 (resolução completa)

A variável aleatória Y denota as vendas semanais de computadores. O valor esperado e a variância de Y são, respectivamente:

$$\mathrm{E}(Y) = \mu_Y = \sum_{y=0}^{3} y \cdot p(y) = 0\bullet 0.10 + 1\bullet 0.15 + 2\bullet 0.45 + 3\bullet 0.30 = 1.95$$

e

$$\mathrm{Var}(Y) = \sigma_Y^2 = \sum_{y=0}^{3} (y-\mu_Y)^2 \cdot p(y) =$$

$$= (0 - 1.95)^2\bullet 0.010 + (1 - 1.95)^2\bullet 0.15 + \ldots + (2 - 1.95)^2\bullet 0.45 + (3 - 1.95)^2\bullet 0.30 =$$
$$= 0.8475.$$

Seja Q a variável aleatória que denota as vendas quinzenais de computadores. Ora, considerando que Y_1 e Y_2 que correspondem às vendas que se verificam em duas semanas consecutivas, então:

$Q = Y_1 + Y_2$,

$\mathrm{E}(Q) = \mathrm{E}(Y_1) + \mathrm{E}(Y_2) = 2\bullet \mathrm{E}(Y) = 2\bullet 1.95 = 3.9$

e

$\mathrm{Var}(Q) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) = 2\bullet \mathrm{Var}(Y) = 2\bullet 0.8475 = 1.695$,

ou

$\sigma(Q) = \sqrt{1.695} = 1.302$. ■

Alínea (ii)

Apoio 2 (sugestão)

Procure representar numa tabela todas as combinações possíveis das variáveis Y_1 e Y_2 (que correspondem às vendas que se verificam em duas semanas consecutivas) juntamente com as probabilidades de tais combinações ocorrerem. Note que por cada combinação de Y_1 e Y_2 se obtém um valor particular de Q. ■

Apoio 3 (resolução completa)

Na tabela seguinte apresentam-se todas as combinações possíveis das variáveis Y_1 e Y_2, juntamente com as probabilidades de tais combinações ocorrerem.

y_1	y_2	q	$p(q)$
0	0	0	0.10 · 0.10 = 0.01
0	1	1	0.10 · 0.15 = 0.015
0	2	2	0.10 · 0.45 = 0.045
0	3	3	0.10 · 0.30 = 0.03
1	0	1	0.15 · 0.10 = 0.015
1	1	2	0.15 · 0.15 = 0.0225
1	2	3	0.15 · 0.45 = 0.0675
1	3	4	0.15 · 0.30 = 0.045
2	0	2	0.45 · 0.10 = 0.045
2	1	3	0.45 · 0.15 = 0.0675
2	2	4	0.45 . 0.45 = 0.2025
2	3	5	0.45 . 0.30 = 0.135
3	0	3	0.30 . 0.10 = 0.03
3	1	4	0.30 . 0.15 = 0.045
3	2	5	0.30 . 0.45 = 0.135
3	3	6	0.30 . 0.30 = 0.09

confirmação: $\sum p(q) = 1$

A função de probabilidade da variável Q é então:

q		$p(q)$
0		= 0.01
1	0.015 + 0.015	= 0.03
2	0.045 + 0.0225 + 0.045	= 0.1125
3	0.03 + 0.0675 + 0.0675 + 0.03	= 0.195
4	0.045 + 0.2025 + 0.045	= 0.2925
5	0.135 + 0.135	= 0.27
6		= 0.09

Verifiquem-se agora os resultados obtidos na alínea anterior.

$$\mathrm{E}(Q) = \mu_Q = \sum_{z=0}^{6} q \cdot p(q) =$$

$$= 0 \bullet 0.01 + 1 \bullet 0.03 + 2 \bullet 0.1125 + 3 \bullet 0.195 + 4 \bullet 0.2925 + 5 \bullet 0.27 + 6 \bullet 0.09 = 3.9.$$

$$\text{Var}(Q) = \sigma_Q^2 = \sum_{z=0}^{6} (q-\mu_Q)^2 \cdot p(q) =$$

$$= (0 – 3.9)^2 \bullet 0.01 + (1 – 3.9)^2 \bullet 0.03 + (2 – 3.9)^2 \bullet 0.1125 + (3 – 3.9)^2 \bullet 0.195 +$$

$$+ (4 – 3.9)^2 \bullet 0.2925 + (5 – 3.9)^2 \bullet 0.27 + (6 – 3.9)^2 \bullet 0.09 = 1.695. \blacksquare$$

Alínea (iii)

Apoio 2 (sugestão)

Procure gerar duas amostras aleatórias (de dimensão 500) de vendas semanais (ou seja, das variáveis Y_1 e Y_2) e, a partir destas, gere uma outra de vendas quinzenais ($Q = Y_1 + Y_2$). Para gerar as amostras aleatórias pode recorrer a qualquer calculadora (ou aplicação informática) que produza números aleatórios pertencentes a uma distribuição Uniforme $U(0, 1)$. No caso do "Microsoft Excel", basta recorrer à função RAND para se obter um número aleatório maior ou igual a 0 e menor do que 1. ■

Apoio 3 (resolução completa)

Começa-se por produzir duas amostras aleatórias (de dimensão 500) de vendas semanais (ou seja, das variáveis Y_1 e Y_2). Para tal, geram-se números aleatórios u_n pertencentes à distribuição Uniforme $U(0, 1)$. Divide-se o intervalo [0, 1[em tantos intervalos quantos os valores que a variável discreta Y pode tomar, associando a cada valor desta um intervalo com uma amplitude proporcional à probabilidade com que tal valor pode ocorrer. Assim:

$y_n = 0$, se $0.0000 \leq u_n < 0.1000$ (10%)

$y_n = 1$, se $0.1000 \leq u_n < 0.2500$ (15%)

$y_n = 2$, se $0.2500 \leq u_n < 0.7000$ (45%)

$y_n = 3$, se $0.7000 \leq u_n < 1.0000$ (30%)

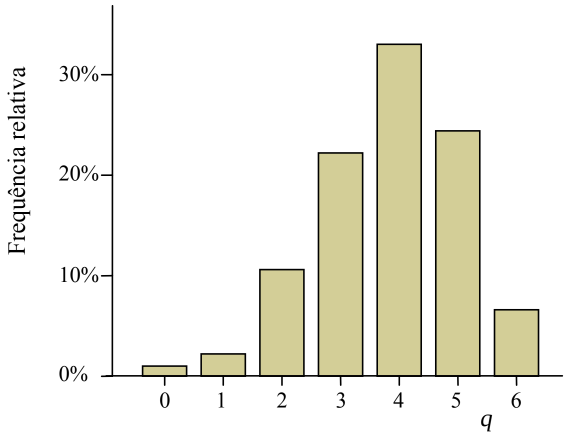
Apresentam-se de seguida duas amostras aleatórias de vendas semanais (y_1 e y_2) e uma de vendas quinzenais ($q = y_1 + y_2$) criada a partir das duas anteriores. Os números aleatórios u_n foram obtidos recorrendo ao "Microsoft Excel".

N	$u_1 = U(0, 1)$	$u_2 = U(0, 1)$	y_1	y_2	$q = y_1+y_2$
1	0.3277	0.2197	2	1	3
2	0.7264	0.3982	2	2	4
3	0.8929	0.5029	3	2	5
4	0.4834	0.7599	2	3	5
		(…)			
498	0.7640	0.2814	3	2	5
499	0.8103	0.8252	3	3	6
500	0.5623	0.9127	2	3	5

No quadro seguinte apresentam-se algumas estatísticas da amostra de vendas quinzenais.

Estatística	$Q = Y_1+Y_2$
Dimensão da amostra	500
Média	3,8360
Desvio padrão	1,2199
Coeficiente de assimetria	-0,4417

Na figura seguinte mostra-se o histograma de frequências relativas construído com as 500 observações da variável *Q*.



■

PROBLEMA 8.2

APOIOS À RESOLUÇÃO

Alínea (i)

Apoio 1 (apenas o resultado)

10.03%. ■

Apoio 2 (sugestão)

Repare que a probabilidade de uma variável tomar um valor superior a uma outra é equivalente a considerar a probabilidade de a variável que resulta da diferença entre as duas anteriores ser superior a zero. Recorde que a combinação linear de duas variáveis Normais independentes segue ainda uma distribuição Normal. ■

Apoio 3 (resolução completa)

Denote-se por X o número de jogos vendidos por dia na loja do Porto e por Y o número de jogos vendidos por dia na loja de Lisboa. O que se pretende determinar é $P(X > Y)$ ou, o que é equivalente, $P[(X - Y) > 0]$. Como X e Y são Normais, a variável $(X – Y)$ também o é, com valor esperado

$E(X – Y) = E(X) – E(Y) = 150 – 200 = -50$

e variância

$$\mathrm{Var}\,(X - Y) = (1)^2 \cdot \mathrm{Var}\,(X) + 2 \cdot (1) \cdot (-1) \cdot \mathrm{Cov}\,(X, Y) + (-1)^2 \cdot \mathrm{Var}\,(Y)$$

$$= \mathrm{Var}\,(X) + \mathrm{Var}\,(Y) = 25^2 + 30^2 = 1525$$ (supondo que X e Y são independentes).

Assim

$$P[(X - Y) > 0] = P\left[Z = \frac{(X - Y) - (-50)}{\sqrt{1525}} > \frac{0 - (-50)}{\sqrt{1525}} \right]$$

$$= P(Z > 1.28) = 0.1003 .$$ ■

Alínea (ii)

Apoio 1 (apenas o resultado)

96.49%. ■

Apoio 2 (sugestão)

Em primeiro lugar, procure caracterizar a variável “número de jogos vendidos por dia no conjunto das duas lojas”. Seguidamente, caracterize a variável “número total de jogos vendidos em 200 dias no conjunto das duas lojas” (que corresponde à soma de 200 variáveis aleatórias independentes com a mesma distribuição). Para tal recorra ao Teorema do Limite Central. ■

Apoio 3 (resolução completa)

Denote-se por $V = X + Y$ a variável Normal que corresponde à soma do número de jogos vendidos por dia em cada uma das duas lojas (Porto e Lisboa). O valor esperado e variância de V são, respectivamente:

$E(V) = E(X) + E(Y) = 150 + 200 = 350$

$$\mathrm{Var}\,(V) = (1)^2 \cdot \mathrm{Var}\,(X) + 2 \cdot (1) \cdot (1) \cdot \mathrm{Cov}\,(X, Y) + (1)^2 \cdot \mathrm{Var}\,(Y)$$

$$= \mathrm{Var}\,(X) + \mathrm{Var}\,(Y) = 25^2 + 30^2 = 1525$$ (admitindo que X e Y são independentes).

Considere agora a variável T que representa o número de jogos vendidos em 200 dias nas duas lojas. Ou seja, $T = \sum_{i=1}^{200} V_i = V_1 + V_2 + \cdots + V_{200} = 200 \cdot \overline{V}$. Esta variável, que é também Normal (soma de variáveis Normais independentes), tem valor esperado e variância, respectivamente,

$$E(T) = E(V_1) + E(V_2) + E(V_3) + ... + E(V_{200}) = 200 \cdot 350 = 700000$$

$$\text{Var}(T) = \text{Var}(V_1) + \text{Var}(V_2) + \text{Var}(V_3) + \ldots + \text{Var}(V_{200}) = 200 \cdot 1525 = 305000 .$$

Assim,

$$P(T > 69000) = P\left(Z = \frac{T - (70000)}{\sqrt{305000}} > \frac{69000 - 70000}{\sqrt{305000}} \right)$$

$$= P(Z > -1.81) = 1 - P(Z < -1.81) = 1 - P(Z > 1.81) = 1 - 0.0351 = 0.9649 . \blacksquare$$

PROBLEMA 8.3

APOIOS À RESOLUÇÃO

Alínea (i)

Apoio 1 (apenas o resultado)

17.91%. ■

Apoio 2 (sugestão)

Na secção 8.2 veja qual a definição do Teorema do Limite Central. Repare que o peso dos 50 mineiros, que se denota por *S*, consiste numa variável que resulta da soma de 50 variáveis aleatórias independentes com a mesma distribuição. ■

Apoio 3 (resolução completa)

Sejam:

X: o peso de cada mineiro [em kg] (com E(X) = 75 [kg] e Var(X) = 64 [kg^2])

C: a capacidade nominal do elevador (com C = 3800 kg)

M: a dimensão da população (com M = 650)

N: a dimensão da amostra (com N = 50)

S: o peso de 50 mineiros seleccionados aleatoriamente, e

$\overline{X}$: a média amostral de X em amostras de dimensão N = 50.

Nestas condições,

$$\text{E}(\overline{X}) = \text{E}(X) = 75 \text{ [kg]}$$

e

$$\text{Var}(\overline{X}) = \left(\frac{M - N}{M - 1} \right) \cdot \frac{1}{N} \cdot \text{Var}(X) = \left(\frac{650 - 50}{650 - 1} \right) \cdot \frac{1}{50} \cdot 64 .$$

Ora,

$$S = \sum_{i=1}^{50} X_i = 50 \cdot \frac{\sum_{i=1}^{50} x_i}{50} = 50 \cdot \overline{X} \, ,$$

vindo

$$\mathrm{E}(S) = 50 \cdot \mathrm{E}(\overline{X}) = 50 \cdot 75 = 3750 \text{ [kg]}$$

e

$$\mathrm{Var}(S) = 50^2 \cdot \mathrm{Var}(\overline{X}) = 50^2 \cdot \left(\frac{650-50}{650-1} \right) \cdot \frac{1}{50} \cdot 64 = 2956.8 \text{ [kg}^2\text{]},$$

ou

$$\sigma_S = \sqrt{2956.8} = 54.38 \text{ [kg]}.$$

De acordo com o Teorema do Limite Central, a variável *S* (soma de 50 variáveis aleatórias independentes com a mesma distribuição) seguirá aproximadamente uma distribuição Normal.

Assim,

$$S \frown N\left(\mu_S = 3750\, , \sigma_S^2 = 54.38^2\right).$$

Transformando a variável *S* na variável normal padronizada *N*(0, 1) vem

$Z = \dfrac{S - \mu_S}{\sigma_S}$. Assim,

$$P(S \geq 3800) = P\left(\frac{S - 3750}{54.38} = Z \geq \frac{3800 - 3750}{54.38} \right)$$
$$= P(Z \geq 0.919).$$

Ora,

$$\left. \begin{array}{l} P(Z > 0.91) = 0.1814 \\ P(Z > 0.92) = 0.1788 \end{array} \right\} \text{(ver Tabela 3, do Anexo «Tabelas»).}$$

Efectuando uma interpolação linear, obtém-se finalmente

$$P(Z > 0.919) = 0.1814 - \frac{0.1814 - 0.1788}{0.92 - 0.91} \cdot (0.919 - 0.91) \approx 0.1791 \; \blacksquare$$

PROBLEMA 8.4

APOIOS À RESOLUÇÃO

Alínea (i)

Apoio 1 (apenas o resultado)

69.57%. ■

Apoio 2 (sugestão)

Independentemente da forma da distribuição original do salário horário de um engenheiro recém-formado na Alemanha, pode considerar-se que a dimensão da amostra é suficientemente grande para que a variável média amostral siga uma distribuição aproximadamente Normal (Teorema do Limite Central). Na secção 8.2 veja quais as expressões do valor esperado e da variância da média amostral. Assuma que os parâmetros das populações coincidem com as correspondentes estatísticas amostrais. ■

Apoio 3 (resolução completa)

Denote-se por A a variável aleatória "salário horário de um engenheiro recém-formado na Alemanha" Admita-se que os parâmetros das populações coincidem com as correspondentes estatísticas amostrais. A variável A terá então valor esperado μ_A = 15.4 e desvio padrão σ_A = 3.0. Atendendo ao Teorema do Limite Central, admite-se que a média dos salários de uma amostra de 40 engenheiros a trabalhar na Alemanha, que se denota por $\overline{A}$, segue uma distribuição Normal com valor esperado

$$\mathrm{E}\left(\overline{A}\right) = \mathrm{E}(A) = 15.4 \ \text{[Euros]}$$

e variância

$$\mathrm{Var}\left(\overline{A}\right) = \frac{1}{N} \cdot \mathrm{Var}(A) = \frac{9}{40} = 0.225 \ [\text{Euros}^2].$$

Padronizando esta variável vem:

$$P\left(15 \le \overline{A} \le 16\right) = P\left(\frac{15-15.4}{\sqrt{0.225}} \le Z = \frac{\overline{A}-15.4}{\sqrt{0.225}} \le \frac{16-15.4}{\sqrt{0.225}}\right) = P(-0.84 \le Z \le 1.26) =$$

$$= 1 - P(Z \ge 1.26) - P(Z \ge 0.84) = 1 - 0.1038 - 0.2005 = 0.6957 \,. \ \blacksquare$$

Alínea (ii)

Apoio 1 (apenas o resultado)

31.92%. ■

Apoio 2 (sugestão)

Procure caracterizar a variável aleatória que resulta da diferença entre as médias amostrais dos salários horários na Alemanha e em Portugal. Recorde que a combinação linear de duas variáveis Normais independentes segue ainda uma distribuição Normal. ■

Apoio 3 (resolução completa)

De acordo com o a resolução da alínea (i), a variável $\overline{A}$ segue uma distribuição $N\left(\mu_{\overline{A}} = 15.4\,, \sigma^2_{\overline{A}} = 0.225\right)$. Denote-se por P a variável aleatória que representa o salário horário de um engenheiro recém-formado em Portugal. Do mesmo modo, atendendo ao Teorema do Limite Central, a variável $\overline{P}$, que representa média dos salários de uma amostra de 50 engenheiros a trabalhar em Portugal, segue uma distribuição Normal com valor esperado

$$\mathrm{E}\left(\overline{P}\right) = \mathrm{E}(P) = 11.2 \text{ [Euros]}$$

e variância

$$\mathrm{V}ar\left(\overline{P}\right) = \frac{1}{N} \cdot \mathrm{Var}(P) = \frac{9}{50} = 0.18 \text{ [Euros}^2\text{]}.$$

A variável "diferença entre as médias amostrais", ($\overline{A} - \overline{P}$**),** segue também uma distribuição Normal com valor esperado e variância, respectivamente,

$$\mathrm{E}\left(\overline{A} - \overline{P}\right) = \mathrm{E}\left(\overline{A}\right) - \mathrm{E}\left(\overline{P}\right) = 15.4 - 11.2 = 4.2 \text{ [Euros]}$$

$$\mathrm{Var}\,(\overline{A} - \overline{P}) = (1)^2 \cdot \mathrm{Var}\,(\overline{A}) + 2 \cdot (1) \cdot (-1) \cdot \mathrm{Cov}\,(\overline{A}, \overline{P}) + (-1)^2 \cdot \mathrm{Var}\,(\overline{P})$$

$$= \mathrm{Var}\,(\overline{A}) + \mathrm{Var}\,(\overline{P}) = 0.225 + 0.18 = 0.405 \text{ [Euros}^2\text{]}$$

(admite-se que A e P são independentes e que, portanto, $\mathrm{Cov}\,(\overline{A}, \overline{P}) = 0$).

Assim

$$P\left[\left(\overline{A} - \overline{P}\right) > 4.5\right] = P\left[Z = \frac{\left(\overline{A} - \overline{P}\right) - 4.2}{\sqrt{0.405}} > \frac{4.5 - 4.2}{\sqrt{0.405}}\right]$$

$$= P(Z > 0.47) = 0.3192\,. \blacksquare$$

PROBLEMA 8.5

APOIOS À RESOLUÇÃO

Alínea (i)

Apoio 1 (apenas o resultado)

$\mu_V = 27 \cdot 10^3$

$\sigma_V \approx 7.79 \cdot 10^3$. ■

Apoio 2 (sugestão)

Embora V (variável que denota o volume das peças) resulte de uma transformação não linear de variáveis Normais, considere que pode ser aproximada por uma relação linear em torno de μ_{d_1}, μ_{d_2} e μ_{d_3}. Na Secção 5.7 do livro, veja quais as expressões do valor esperado e da variância neste tipo de transformações. ■

Apoio 3 (resolução completa)

Seja $V = d_1 \cdot d_2 \cdot d_3$ a variável aleatória que denota o volume das peças. Embora V resulte de uma transformação não linear de variáveis Normais e independentes, admita-se que pode ser aproximada por uma relação linear em torno de μ_{d_1}, μ_{d_2} e μ_{d_3}. Assim,

$$\mu_V \approx E(d_1) \cdot E(d_2) \cdot E(d_3) = \mu_d^3 = 27 \cdot 10^3$$

e

$$\begin{aligned}
\operatorname{Var}(V) \approx & \left[\left. \frac{\partial V}{\partial d_1} \right|_{\mu_{d_1}, \mu_{d_2}, \mu_{d_3}} \right]^2 \cdot \operatorname{Var}(d_1) + \\
& + \left[\left. \frac{\partial V}{\partial d_2} \right|_{\mu_{d_1}, \mu_{d_2}, \mu_{d_3}} \right]^2 \cdot \operatorname{Var}(d_2) + \\
& + \left[\left. \frac{\partial V}{\partial d_3} \right|_{\mu_{d_1}, \mu_{d_2}, \mu_{d_3}} \right]^2 \cdot \operatorname{Var}(d_3) \\
& = 3 \cdot \left[\mu_d^2\right]^2 \cdot \sigma_d^2 \\
& = 3 \cdot 900^2 \cdot 25 = 60.75 \cdot 10^6,
\end{aligned}$$

ou

$\sigma_V \approx 7.79 \bullet 10^3$. ■

Alínea (ii)

Apoio 1 (apenas o resultado)

16.0%. ■

Apoio 2 (sugestão)

Verifique através de um histograma que a hipótese de que a variável V pode ser aproximada por uma relação linear em torno de μ_{d_1}, μ_{d_2} e μ_{d_3} não é rigorosa. A partir de amostras de d_1, d_2 e d_3 gere amostras de $V = d_1 \cdot d_2 \cdot d_3$, e recorra à definição frequencista de probabilidade para calcular $P(V > 35)$. Gere as amostras pelo método de Monte Carlo recorrendo ao "Microsoft Excel". ■

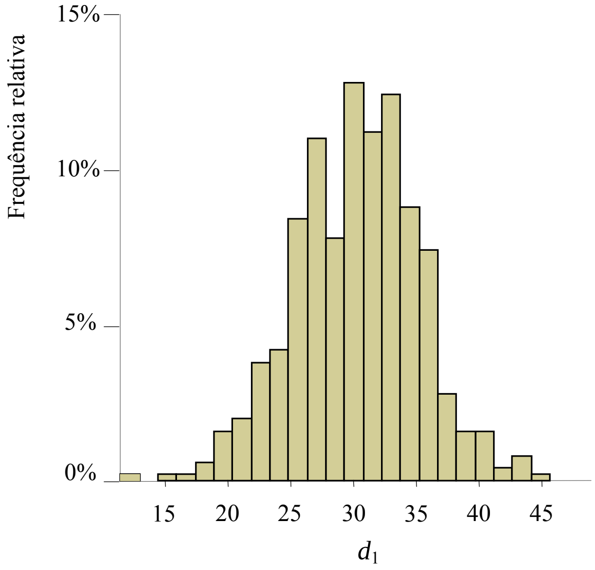
Apoio 3 (resolução completa)

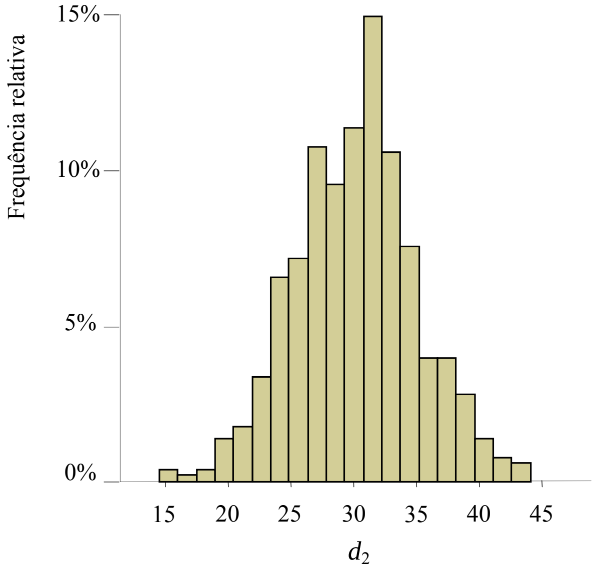
Se a hipótese de que a variável V pode ser aproximada por uma relação linear em torno de μ_{d_1}, μ_{d_2} e μ_{d_3} for válida, então V resultará de uma combinação linear de variáveis Normais e seguirá também uma distribuição Normal $N(\mu_V, \sigma_V)$. Nessas condições seria

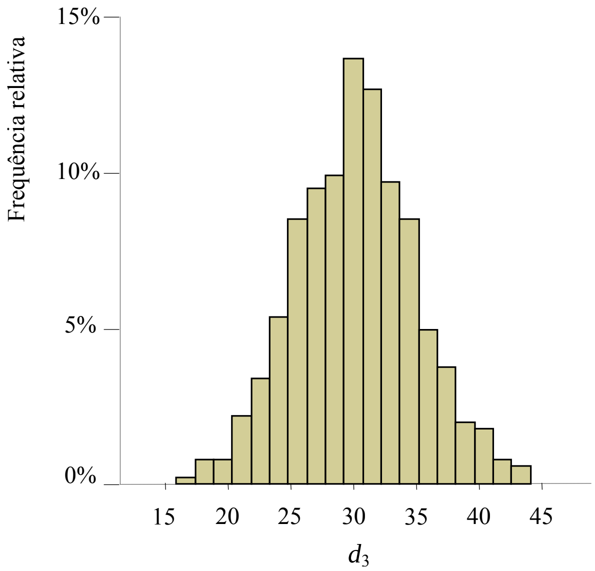
$$P\left(V > 35 \cdot 10^3\right) = P\left(Z = \frac{V - 27 \cdot 10^3}{7.79 \cdot 10^3} > \frac{35 - 27}{7.79}\right) = P(Z > 1.027) \approx 0.1522\,.$$

No entanto, não parece provável que a aproximação linear seja muito rigorosa, pois é de esperar que V tenha uma distribuição assimétrica à direita (isto é, com uma cauda maior à direita). Verifique-se se assim é gerando uma amostra de 500 observações de V, construída a partir de outras tantas observações independentes de d_1, d_2 *e* d_3 (com d_i sendo $IN(\mu_d = 30, \sigma_d = 5)$. As amostras de d_1, d_2 *e* d_3 podem ser produzidas recorrendo ao "Microsoft Excel", da seguinte forma. Para cada uma das variáveis d_i, produzem-se 500 observações pertencentes à distribuição $U(0, 1)$ recorrendo à função RAND. Em seguida, utilizando a função NORMINV, que produz valores da inversa da função distribuição da Normal (μ, σ), geram-se as observações de d_i.

Nas figuras seguintes apresentam-se os histogramas de frequências das três variáveis d_1, d_2 *e* d_3.



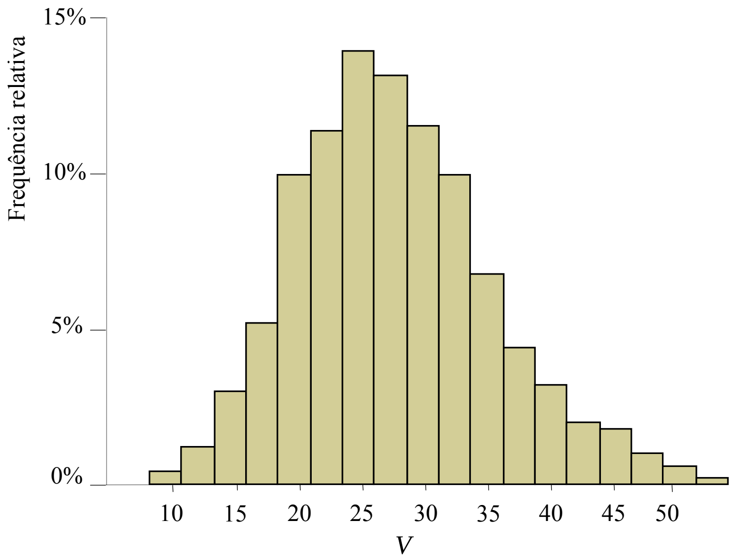




Na tabela seguinte apresentam-se algumas estatísticas amostrais de d_1, d_2 *e* d_3.

Estatística amostral	d_1	d_2	d_3
Dimensão da amostra	500	500	500
Média	30.208	30.132	30.209
Desvio padrão	4.943	4.856	4.763
Coeficiente de kurtose	1.046	0.269	-0.302
Coeficiente de assimetria	-1.096	-0.249	0.614

As 500 observações de *V* foram obtidas multiplicando os valores referentes à primeira observação das variáveis d_1, d_2 *e* d_3, à segunda observação, e assim sucessivamente até se produzir uma amostra de dimensão 500.



Na tabela seguinte apresentam-se algumas medidas amostrais obtidas a partir da amostra de *V*.

Estatística amostral	*V*
Dimensão da amostra	500
Média	27.545
Desvio padrão	7.924
Coeficiente de kurtose	0.917
Coeficiente de assimetria	4.633

Tal como se suspeitava, o histograma permite observar uma moderada assimetria à direita, que é confirmada pelo valor positivo do coeficiente de assimetria amostral.

Para estimar $P(V > 35000)$ a partir da amostra de dimensão 500 da variável *V*, ordena-se o vector de valores de *V* (por exemplo, por ordem decrescente) e verifica-se qual a proporção de observações com valor superior a 35000. Na tabela seguinte pode verificar-se que a partir da octogésima primeira observação os valores de *V* são inferiores a 35000.

Observação no.	V	
74	35.431	
75	35.205	
76	35.181	
77	35.173	
78	35.045	
79	35.014	
80	35.003	> 35000 ↑
81	34.991	< 35000 ↓
82	34.950	
83	34.873	

Assim

$$\hat{P}(V > 35.000) \approx \frac{80}{500} = 0.160 . \blacksquare$$

Alínea (iii)

Apoio 1 (apenas o resultado)

$\mu_V \approx 28324$

$\sigma_V \approx 10207$

$P(V > 35000) = 23.0\%$. ■

Apoio 2 (sugestão)

Note que a nova distribuição de V terá uma dispersão maior do que a calculada na alínea anterior. Assim, torna-se ainda mais recomendável estimar o valor esperado e a variância de V, bem como determinar $P(V > 35000)$, através de um processo de geração artificial de amostras. ■

Apoio 3 (resolução completa)

Note-se que, agora, o volume é dado por $V = d^3$, em que d é uma variável $N(\mu_d = 30,\ \sigma_d = 5)$. Admita-se, em primeiro lugar, que V é aproximável por uma relação linear em d (em torno de μ_d). Nessas condições viria:

$$\mu_V \approx 3 \cdot \mathrm{E}(d) = \mu_d^3 = 27 \cdot 10^3$$

e

$$\mathrm{Var}(V) \approx \left[\left. \frac{\partial V}{\partial d} \right|_{\mu_d} \right]^2 \cdot \mathrm{Var}(d) = \left(3 \cdot \left[\mu_d^2\right]\right)^2 \cdot \sigma =$$

$$= 9 \cdot 900^2 \cdot 25 = 1.8225 \cdot 10^8 ,$$

ou

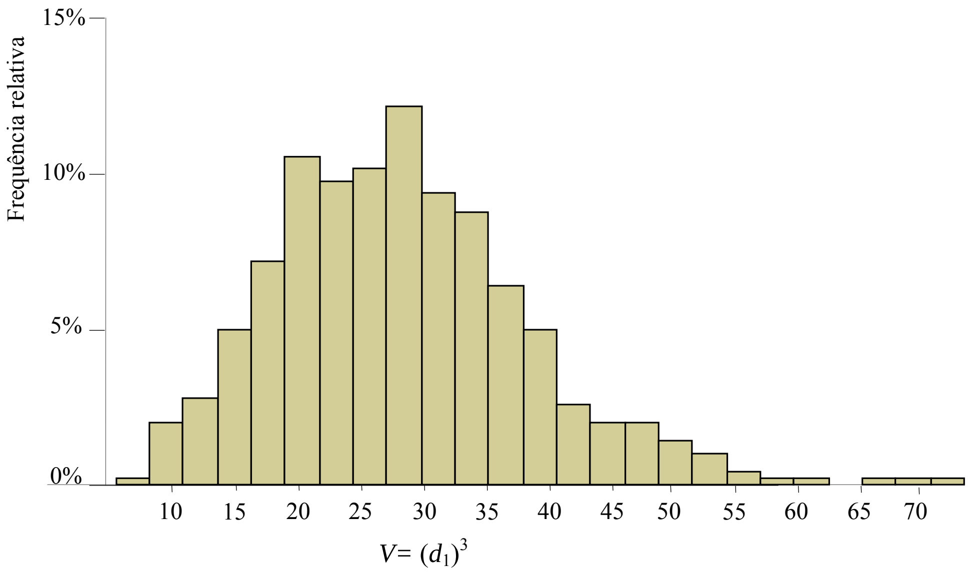
$\sigma_V \approx 1.350 \cdot 10^4$.

Por outro lado, se a hipótese formulada for válida, então V também seria Normal (μ_V, σ_V). Então:

$$P\left(V > 35 \cdot 10^3\right) = P\left(V > 35 \cdot 10^3\right) = P\left(Z = \frac{V - 27 \cdot 10^3}{1.350 \cdot 10^4} > \frac{35 - 27}{13.5}\right) = P(Z > 0.593) \approx 0.2767 .$$

No entanto, se anteriormente a aproximação linear era questionável agora mais o será (visto que a nova distribuição de V terá uma dispersão maior do que a anterior). Nestas condições, será ainda mais recomendável estimar o valor esperado e a variância de V, bem como $P(V > 35000)$, através de um processo de geração artificial de amostras.

Seguidamente apresentam-se os resultados obtidos a partir da amostra de dimensão 500 da variável V, agora com $V = (d_1)^3$.



Na tabela seguinte apresentam-se algumas medidas amostrais obtidas a partir da amostra de $V = (d_1)^3$.

Estatística amostral	$V = (d_1)^3$
Dimensão da amostra	500
Média	28324
Desvio padrão	10207
Coeficiente de kurtose	6.837
Coeficiente de assimetria	5.392

Procedendo de forma idêntica à da alínea (ii) estima-se $P(V > 35000)$ Verifica-se (ver tabela seguinte) que a partir observação colocada em centésimo décimo sexto lugar os valores de V são inferiores a 35000.

Observação no.	$V = (d_1)^3$	
109	35.580	
110	35.177	
111	35.159	
112	35.110	
113	35.085	
114	35.053	
115	35.014	> 35000 ↑
116	34.964	< 35000 ↓
117	34.943	
118	34.933	

Assim

$$\hat{P}(V > 35.000) \approx \frac{115}{500} = 0.230\,,$$

valor substancialmente diferente de 0.2767 que havia sido determinado anteriormente. ■

PROBLEMA 8.6

APOIOS À RESOLUÇÃO

Alínea (i)

Apoio 1 (apenas o resultado)

Não aplicável. ■

Apoio 2 (sugestão)

Na Secção 8.6 - Geração de Amostras Recorrendo à Técnica de Monte Carlo - veja como se podem obter amostras aleatórias provenientes de uma distribuição Exponencial Negativa. Gere uma amostra de dimensão 500 e compare as estatísticas amostrais obtidas a partir desta via experimental com os valores teoricamente esperados. ■

Apoio 3 (resolução completa)

Sejam as variáveis $X_1, \ldots, X_{15}$ provenientes de uma distribuição Exponencial Negativa $EN(\lambda = 1/3)$. Sejam ainda

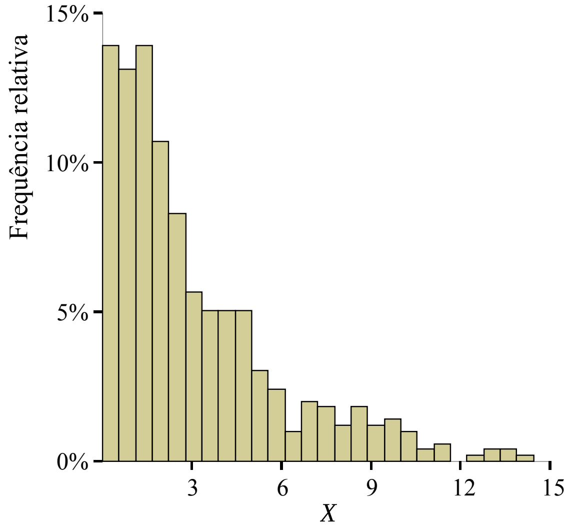
*X*barra2: $\overline{X}_2 = \frac{1}{2} \cdot \sum_{i=1}^{2} X_i$

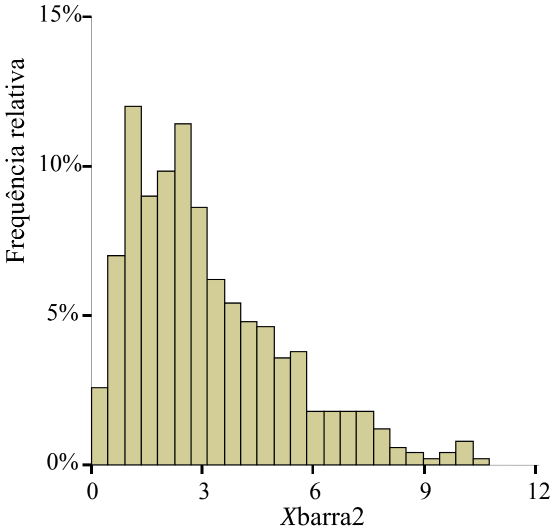
*X*barra5: $\overline{X}_5 = \frac{1}{5} \cdot \sum_{i=1}^{5} X_i$

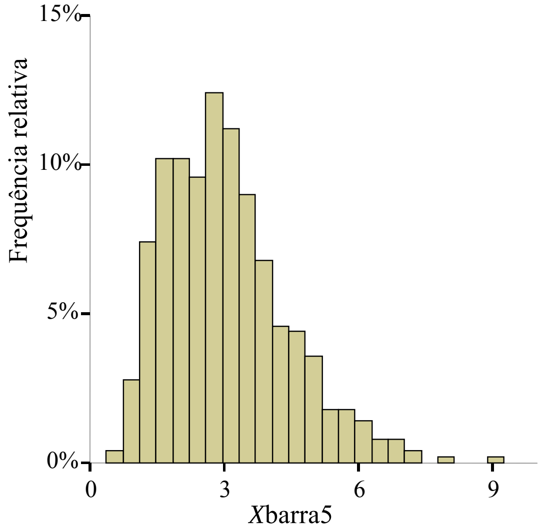
*X*barra10: $\overline{X}_{10} = \frac{1}{10} \cdot \sum_{i=1}^{10} X_i$

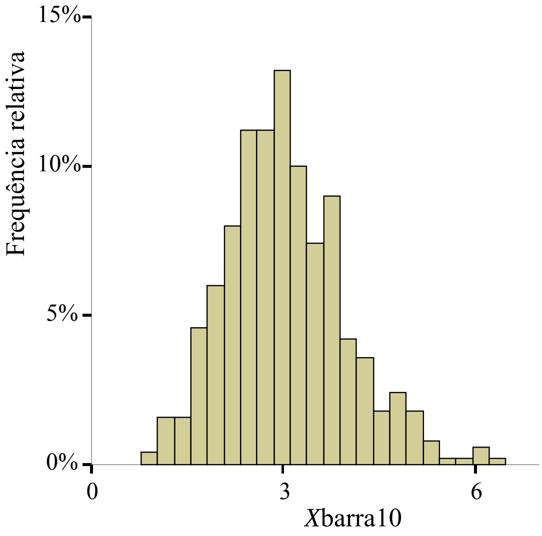
*X*barra15: $\overline{X}_{15} = \frac{1}{15} \cdot \sum_{i=1}^{15} X_i$.

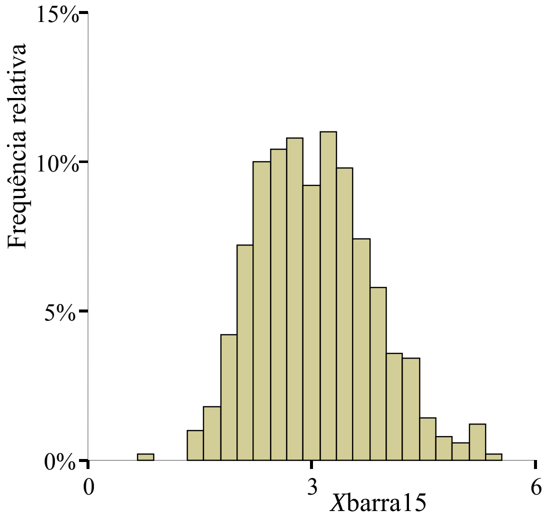
Recorde-se que a função de distribuição de uma variável Exponencial Negativa (λ) é $F(X) = 1 - e^{-\lambda x}$. Seja U uma variável aleatória pertencente a uma distribuição $U(0,\ 1)$. Fazendo $U = F(X)$, resulta que a função inversa de $F(X)$, $F^{-1}(U)$, é $X = 2 \cdot \ln\left[1/(1-U)\right]$. Recorrendo a esta expressão, podem gerar-se observações de uma qualquer distribuição Exponencial Negativa a partir de observações $U(0,\ 1)$, obtidas, por exemplo, recorrendo à função RAND no "Microsoft Excel". Apresentam-se de seguida os histogramas das distribuições de amostras de dimensão 500 de X, $\overline{X}_2$, $\overline{X}_5$, $\overline{X}_{10}$ e $\overline{X}_{15}$ produzidas seguindo esta metodologia.











Na tabela seguinte apresentam-se algumas medidas amostrais obtidas a partir das amostras das diferentes variáveis.

Estatística amostral	X	$\overline{X}_2$	$\overline{X}_5$	$\overline{X}_{10}$	$\overline{X}_{15}$
Dimensão da amostra	500	500	500	500	500
Média	3.216	3.144	3.044	3.034	3.043
Variância	9.968	4.573	1.841	0.878	0.602
Desvio padrão	3.157	2.139	1.357	0.937	0.776
Coeficiente de assimetria	16.521	11.318	7.836	5.079	3.315
Coeficiente de kurtose	18.204	8.995	4.096	2.391	-0.332

Para interpretar estes resultados recorda-se que a dispersão da média amostral é inferior à dispersão da variável original, e é tanto menor quanto maior for a dimensão da amostra. De facto, os resultados da tabela anterior verificam a expressão

$$\sigma_{\overline{X}}^2 = \frac{1}{N} \cdot \sigma_X^2 ,$$

válida para amostras aleatórias simples. Recorda-se ainda que o valor esperado da média amostral, independentemente da dimensão da amostra, é idêntico ao da variável original (ou seja, $\mu_{\overline{X}} = \mu_X$).

Por outro lado, a distribuição da média amostral, obtida a partir de uma amostra aleatória simples, tende para uma distribuição Normal à medida que a dimensão da amostra cresce (Teorema do Limite Central). No entanto, dado que se está perante uma distribuição original X muito assimétrica, a média amostral (mesmo quando estimada com N = 15) apresenta ainda uma ligeira assimetria. Como regra prática, é frequente indicar-se que a aproximação Normal é adequada quando $N \geq 50$ se a distribuição original for muito assimétrica. ■

PROBLEMA 8.7

APOIOS À RESOLUÇÃO

<u>Apoio 1</u> (apenas o resultado)

Não aplicável. ■

<u>Apoio 2</u> (sugestão)

Comece por gerar duas variáveis $U_1(0,1)$ e $U_2(0,1)$, obtidas a partir de duas séries distintas produzidas recorrendo à função RAND do "Microsoft Excel". Recorra às expressões do Apêndice 8.3 do livro para converter directamente pares de observações $U(0,1)$ em pares de observações $N(0,1)$. Verifique se as variáveis transformadas são de facto independentes. ■

<u>Apoio 3</u> (resolução completa)

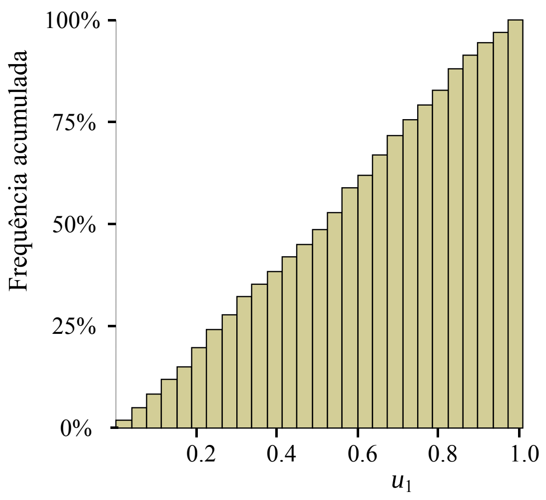
Utilizando a função RAND no "Microsoft Excel" geram-se 500 observações de duas amostras independentes U_1 e U_2, ambas pertencentes a uma distribuição $U(0, 1)$. A partir destes valores, geram-se 500 observações X_1 e X_2 de uma distribuição $N(0, 1)$ recorrendo às expressões de Box-Muller:

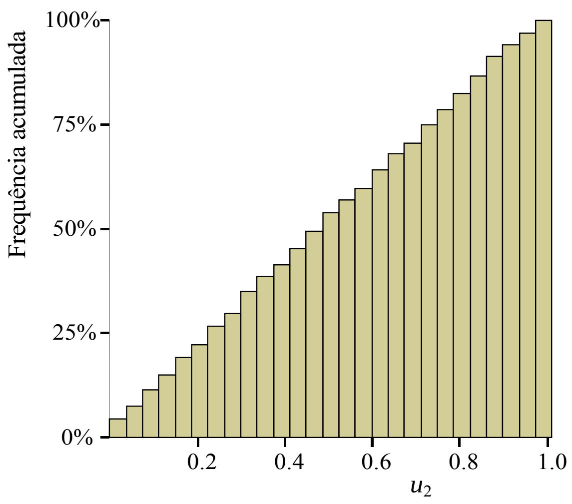
$$X_1 = \sqrt{-2 \cdot \ln U_1} \cdot \cos(2\pi \cdot U_2)$$

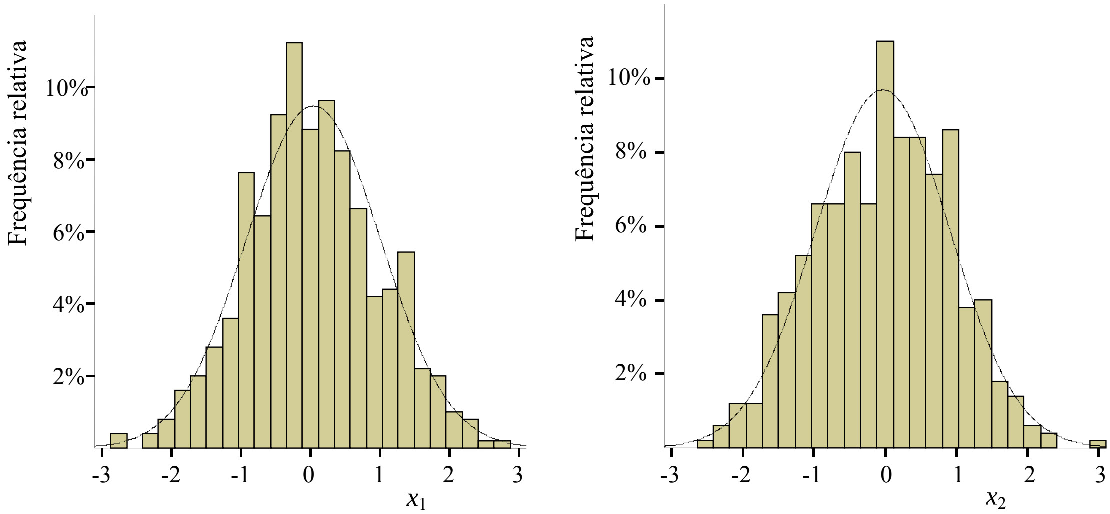
e

$$X_1 = \sqrt{-2 \cdot \ln U_1} \cdot \operatorname{sen}(2\pi \cdot U_2).$$

Nas figuras seguintes apresentam-se os histogramas de frequências acumuladas das observações de U_1, U_2, e os histogramas de frequências relativas das observações de X_1 e X_2. Nos dois últimos casos, a cada um dos histogramas sobrepôs-se a funções densidade de probabilidade da variável Normal padronizada.







Na tabela seguinte apresentam-se algumas medidas amostrais das variáveis X_1 e X_2.

Estatística amostral	X_1	X_2
Dimensão da amostra	500	500
Média	0.044	-0.033
Desvio padrão	0.970	0.945
Coeficiente de kurtose	0.459	-1.842
Coeficiente de assimetria	1.533	-0.637

Na base do método proposto por Box-Muller está o pressuposto de as variáveis transformadas definidas de acordo com as expressões apresentadas anteriormente serem independentes. De facto, verifica-se que o coeficiente de correlação amostral entre as observações das variáveis X_1 e X_2 é $r_{X_1,X_2} = -0.002$. ■

PROBLEMA 8.8

APOIOS À RESOLUÇÃO

Alínea (i)

Apoio 1 (apenas o resultado)

Não aplicável. ■

Apoio 2 (sugestão)

Comece por especificar a função de probabilidade e a função de distribuição da variável em causa. Seguidamente, divida o intervalo [0, 1] em tantos intervalos quantos os valores que a variável discreta pode tomar, associando a cada um deles um intervalo com uma amplitude proporcional à probabilidade com que tal valor pode ocorrer. Por fim, recorra à técnica de Monte Carlo para gerar amostras. ■

Apoio 3 (resolução completa)

Denote-se por Y a variável de Poisson ($\lambda = 4$) que corresponde ao número de navios que chegam ao cais por hora. Na tabela seguinte incluem-se os valores da função de probabilidade e da função de distribuição de Y (tais valores foram obtidos recorrendo à Tabela 2 do Anexo "Tabelas").

Y	$P(y)$	$F(y)$
0	0.0183	0.0183
1	0.0733	0.0916
2	0.1465	0.2381
3	0.1954	0.4335
4	0.1953	0.6288
5	0.1563	0.7851
6	0.1042	0.8893
7	0.0596	0.9489
8	0.0297	0.9786
≥ 9	0.0214	1.0000

Divida-se o intervalo [0, 1[em tantos intervalos quantos os valores que a variável discreta Y pode tomar. Associe-se a cada um deles um intervalo com uma amplitude proporcional à probabilidade com que tal valor pode ocorrer. Ou seja, por exemplo, como $P(0) = 0.0183$, Y tomará o valor 0 se $F(y) < 0.0188$, como $P(1) = 0.0733$, tomará o valor 1 se $F(y) \geq 0.0188$ e $F(y) < 0.0916$, e assim sucessivamente. Sejam u_n observações de uma variável $U(0, 1)$. Estas observações podem ser transformadas em número de navios, y_n, do seguinte modo:

$y_n = 0$, se $0.0000 \leq u_n < 0.0183$ (1.83%)

$y_n = 1$, se $0.0183 \leq u_n < 0.0916$ (7.33%)

$y_n = 2$, se $0.0916 \leq u_n < 0.2381$ (14.65%)

$y_n = 3$, se $0.2381 \leq u_n < 0.4335$ (19.54%)

$y_n = 4$, se $0.4335 \leq u_n < 0.6288$ (19.53%)

$y_n = 5$, se $0.6288 \leq u_n < 0.7851$ (15.63%)

$y_n = 6$, se $0.7851 \leq u_n < 0.8893$ (10.42%)

$y_n = 7$, se $0.8893 \leq u_n < 0.9489$ (5.96%)

$y_n = 8$, se $0.9489 \leq u_n < 0.9786$ (2.97%)

$y_n = 9$, se $0.9786 \leq u_n < 1.0000$ (2.14%)

Seja T a variável que denota o número de navios que chegam ao cais num período de 3 horas. Recorrendo à série de números aleatórios equiprováveis independentes da Tabela 7 do Anexo «Tabelas», geraram-se as seguintes três amostras:

	1ª amostra		2ª amostra		3ª amostra	
1ª hora	$u_1 = 0.2683$	$y_1 = 3$	$u_1 = 0.0993$	$y_1 = 2$	$u_1 = 0.4366$	$y_1 = 4$
2ª hora	$u_2 = 0.6621$	$y_2 = 5$	$u_2 = 0.0315$	$y_2 = 1$	$u_1 = 0.2821$	$y_2 = 3$
3ª hora	$u_3 = 0.2429$	$y_3 = 3$	$u_3 = 0.7919$	$y_3 = 6$	$u_1 = 0.9765$	$y_3 = 8$
T	$T = 11$		$T = 9$		$T = 15$	

■

Alínea (ii)

Apoio 1 (apenas o resultado)

3.89. ■

Apoio 2 (sugestão)

A partir dos valores registados para o número de navios que chegam ao cais num período de 3 horas (T_1, T_2 e T_3), calculados para as 3 amostras anteriormente geradas, procure estimar o número de navios que chegam ao cais por hora. ■

Apoio 3 (resolução completa)

Tomando os resultados da simulação efectuada na alínea (i), o número médio de navios que chegam ao cais por hora será:

$$\bar{y} = \left(\frac{11}{3} + \frac{9}{3} + \frac{15}{3}\right) \Big/ 3 = 3.89 .$$ ■