ACH0021 – Tratamento e Análise de Dados/Informações

Lista de Exercícios 1

Observação 1: Os exercícios desta lista devem ser resolvidos SEM o uso de ferramentas computacionais
Observação 2: Alguns dos exercícios foram adaptados ou retirados do livro de M. N. Magalhães & A. C. P. de Lima, *Noções de Probabilidade e Estatística*, Edusp (2008).

Notação: Os símbolos $d_{a\vdash b}$ e $f_{a\vdash b}$ indicam, respectivamente, a densidade e frequência relativa no intervalo $a \vdash b$. Ademais, n_i e f_i indicam, respectivamente, a i-ésima frequências absoluta e relativa.

Formulário (para um conjunto $\{x_i\}$ de n dados)

Média (amostral): $\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$ Variância (amostral): $\sigma^2 :\approx \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ Desvio padrão (amostral): σ

---

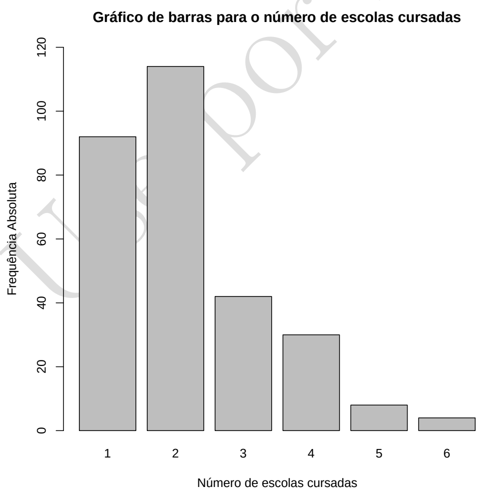
1) Um grupo de pedagogos estuda a influência da troca de escolas no desempenho de alunos do ensino fundamental. Como parte do levantamento realizado, foi anotado o número de escolas cursadas pelos alunos participantes do estudo.

Escolas cursadas	Frequência
1	92
2	114
3	42
4	30
5	8
6	4

a) Determinar a porcentagem dos alunos que cursaram mais de duas escolas.
b) Construir o gráfico de barras.
c) Classificar os alunos em dois grupos segundo a rotatividade: *alta* para alunos com mais de 2 escolas e *baixa* para os demais. Obter a tabela de frequência dessa variável.

1a) Porcentagem dos alunos que cursaram mais de duas escolas: $\frac{42+30+8+4}{92+114+42+30+8+4} = \frac{42}{145} = 0,2896\cdots \approx 28,97\%$.

1b) Gráfico de barras:



1c) Tabela de rotatividade:

Rotatividade	Frequência absoluta	Frequência relativa
Baixa	206	206/290 ≈ 0,71
Alta	84	84/290 ≈ 0,29
TOTAL	290	1,00

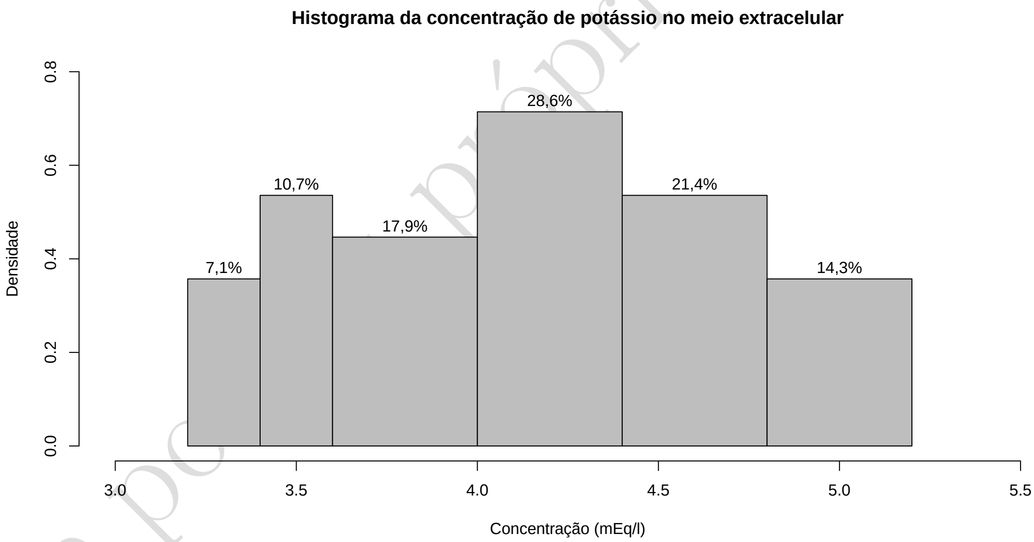
2) Cinquenta e seis pacientes de uma clínica médica tiveram a sua concentração de potássio no meio extracelular medida (em mEq/l – miliequivalente por litro[1]). Os resultados foram os seguintes:

Concentração (mEq/l)	Frequência
3,20 $\vdash$ 3,40	04
3,40 $\vdash$ 3,60	06
3,60 $\vdash$ 4,00	10
4,00 $\vdash$ 4,40	16
4,40 $\vdash$ 4,80	12
4,80 $\vdash$ 5,20	08

a) Construir o histograma.
b) Estimar a média e o desvio padrão.
c) Considerando "típicas" as concentrações de potássio até a uma distância de um desvio padrão em relação à média, estimar a fração de pacientes que apresenta a concentração "típica".
d) Considerando "pacientes normais" aqueles que têm a concentração de potássio (no meio extracelular) entre $3,5$mEq/l a $5,0$mEq/l, estimar a fração dos pacientes fora desta faixa.

2a) Tabela de frequências, amplitudes e densidades para as concentrações de potássio no meio extracelular:

Concentração de potássio (mEq/l)	g_i	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
3,20 $\vdash$ 3,40	3,30	04	$04/56 \approx 0{,}071$	0,20	$5/14 \approx 0{,}357$
3,40 $\vdash$ 3,60	3,50	06	$06/56 \approx 0{,}107$	0,20	$15/28 \approx 0{,}536$
3,60 $\vdash$ 4,00	3,80	10	$10/56 \approx 0{,}179$	0,40	$25/56 \approx 0{,}446$
4,00 $\vdash$ 4,40	4,20	16	$16/56 \approx 0{,}286$	0,40	$5/7 \approx 0{,}714$
4,40 $\vdash$ 4,80	4,60	12	$12/56 \approx 0{,}214$	0,40	$15/28 \approx 0{,}536$
4,80 $\vdash$ 5,20	5,00	08	$08/56 \approx 0{,}143$	0,40	$5/14 \approx 0{,}357$
TOTAL	-	56	1,000	-	-



2b) Admitir-se-á a hipótese da distribuição uniforme dos dados nas barras do histograma. Consequentemente, pode-se eleger o ponto médio g_i do i-ésimo intervalo como sendo o representante do mesmo para as estimativas que seguem.

Estimativa da média $\overline{g}$ do total de $n = 56$ dados:

$$\begin{aligned} \overline{g} &= \frac{1}{n}\sum_i n_i g_i = \frac{1}{56}\Big(4 \cdot 3,30 + 6 \cdot 3,50 + 10 \cdot 3,80 + 16 \cdot 4,20 + 12 \cdot 4,60 + 8 \cdot 5,00\Big) \\ &= \frac{1173}{280} \approx 4,19 (mEq/l)\,. \end{aligned}$$

Estimativa da variância σ^2:

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_i n_i\left(g_i-\overline{g}\right)^2 = \frac{1}{56}\left[4\left(3,30-\frac{1173}{280}\right)^2+6\left(3,50-\frac{1173}{280}\right)^2+10\left(3,80-\frac{1173}{280}\right)^2+\right. \\
&\left. +16\left(4,20-\frac{1173}{280}\right)^2+12\left(4,60-\frac{1173}{280}\right)^2+8\left(5,00-\frac{1173}{280}\right)^2\right]=\frac{20739}{78400}\,,
\end{aligned}$$

que implica um desvio padrão de $\sigma=\sqrt{\frac{20739}{78400}}\approx 0,51(mEq/l)$.

2c) A fração dos pacientes situados no intervalo $(\overline{g}-\sigma,\overline{g}+\sigma)\approx(3,67;4,70)$ é estimada assumindo uma distribuição uniforme dos dados nos intervalos $3,60\vdash 4,00$ e $4,40\vdash 4,80$. É imediato que

$$d_{3,60\vdash 4,00}=d_{\overline{g}-\sigma\vdash 4,00}\Rightarrow\frac{25}{56}=\frac{f_{\overline{g}-\sigma\vdash 4,00}}{4,00-(\overline{g}-\sigma)}\Rightarrow f_{\overline{g}-\sigma\vdash 4,00}=\frac{25}{56}\left(\sqrt{\frac{20739}{78400}}-\frac{53}{280}\right)$$

e

$$d_{4,40\vdash 4,80}=d_{4,40\vdash\overline{g}+\sigma}\Rightarrow\frac{15}{28}=\frac{f_{4,40\vdash\overline{g}+\sigma}}{(\overline{g}+\sigma)-4,40}\Rightarrow f_{4,40\vdash\overline{g}+\sigma}=\frac{15}{28}\left(\sqrt{\frac{20739}{78400}}-\frac{59}{280}\right).$$

Logo, chega-se a

$$\begin{aligned}
f_{\overline{g}-\sigma\vdash 4,00}+f_{4,00\vdash 4,40}+f_{4,40\vdash\overline{g}+\sigma} &= \frac{25}{56}\left(\sqrt{\frac{20739}{78400}}-\frac{53}{280}\right)+\frac{16}{56}+\frac{15}{28}\left(\sqrt{\frac{20739}{78400}}-\frac{59}{280}\right) \\
&= \frac{277}{3136}+\frac{55}{56}\sqrt{\frac{20739}{78400}}\approx 0,593\,,
\end{aligned}$$

e a fração de pacientes que apresenta a concentração “típica” é cerca de 59,3%.

2d) Assumindo uma distribuição uniforme dos dados nos intervalos $3,40\vdash 3,60$ e $4,80\vdash 5,20$, tem-se

$$d_{3,40\vdash 3,60}=d_{3,40\vdash 3,50}\Rightarrow\frac{15}{28}=\frac{f_{3,40\vdash 3,50}}{3,50-3,40}\Rightarrow f_{3,40\vdash 3,50}=\frac{3}{56}$$

e

$$d_{4,80\vdash 5,20}=d_{5,00\vdash 5,20}\Rightarrow\frac{5}{14}=\frac{f_{5,00\vdash 5,20}}{5,20-5,00}\Rightarrow f_{5,00\vdash 5,20}=\frac{1}{14}\,.$$

Desta forma, a fração de pessoas fora da faixa $(3,50;5,00)$ é estimada por

$$f_{3,20\vdash 3,40}+f_{3,40\vdash 3,50}+f_{5,00\vdash 5,20}=\frac{4}{56}+\frac{3}{56}+\frac{1}{14}=\frac{11}{56}\approx 0,196\,,$$

que é cerca de 19,6%.

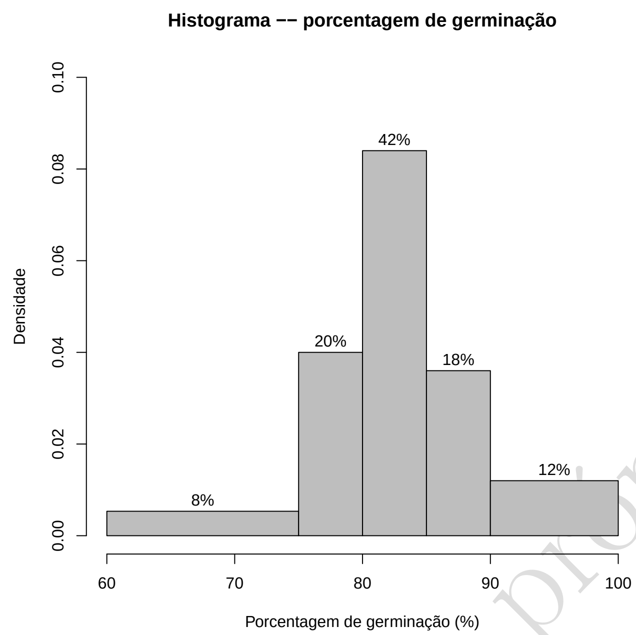
3) O índice de germinação é um dos principais fatores para definir a qualidade das sementes. Ele é determinado em experimento científico conduzido pelo fabricante e regulamentado pelos órgãos fiscalizadores. Um fabricante afirma que o índice de germinação de suas sementes de milho é de 85%. Para verificar tal afirmação, uma cooperativa de agricultores sorteou 150 amostras com 100 sementes em cada uma e anotou a porcentagem de germinação em cada amostra.

Germinação (%)	Frequência
$60\vdash 75$	12
$75\vdash 80$	30
$80\vdash 85$	63
$85\vdash 90$	27
$90\vdash 100$	18

a) Fazer uma representação gráfica da tabela ao lado.
b) Estimar a média e o desvio padrão.
c) Comentar a afirmação do fabricante.

3a) Tabela de frequências, amplitudes e densidades para os índices de germinação:

Germinação (%)	g_i	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
$60 \vdash 75$	67,5	12	$12/150 = 0{,}08$	15	$2/375 \approx 0{,}0053$
$75 \vdash 80$	77,5	30	$30/150 = 0{,}20$	5	$1/25 = 0{,}0400$
$80 \vdash 85$	82,5	63	$63/150 = 0{,}42$	5	$21/250 = 0{,}0840$
$85 \vdash 90$	87,5	27	$27/150 = 0{,}18$	5	$9/250 = 0{,}0360$
$90 \vdash 100$	95,0	18	$18/150 = 0{,}12$	10	$3/250 = 0{,}0120$
TOTAL	-	150	1,00	-	-



3b) Admitir-se-á a hipótese de distribuição uniforme dos dados nas barras do histograma. Consequentemente, pode-se eleger o ponto médio g_i do i-ésimo intervalo como sendo representante do mesmo para as estimativas que seguem.
Estimativa da média $\overline{g}$ dos $n = 100$ dados:

$$\begin{aligned}\overline{g} &= \frac{1}{n}\sum_i n_i g_i = \sum_i f_i g_i = 0,08 \cdot 67,5 + \\ &+0,20 \cdot 77,5 + 0,42 \cdot 82,5 + \\ &+0,18 \cdot 87,5 + 0,12 \cdot 95,0 = 82,7(\%)\,.\end{aligned}$$

Estimativa da variância σ^2:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_i n_i\,(g_i - \overline{g})^2 = \frac{1}{150}\Big[\\ &12\,(67,5 - 82,7)^2 + 30\,(77,5 - 82,7)^2 + \\ &+63\,(82,5 - 82,7)^2 + \\ &+27\,(87,5 - 82,7)^2 + \\ &+18\,(95,0 - 82,7)^2\Big] = \frac{4621}{100}\,,\end{aligned}$$

que implica um desvio padrão de $\sigma = \sqrt{\frac{4621}{100}} \approx 6,8(\%)$.

3c) A porcentagem de germinação de 85% (declarado pelo fabricante) encontra-se dentro de uma distância de um desvio padrão da média, que compreende a região $(\overline{g} - \sigma, \overline{g} + \sigma) \approx (75,9; 89,5)$ – em porcentagem. Admitindo que este último represente o intervalo que contenha os valores típicos que flutuam ao redor de um valor "central" (representado pela média), o valor 85% pode ser considerado "típico" do conjunto de dados. **Segundo este critério**, a afirmação do fabricante é dotada de plausibilidade.

4) Um exame vestibular para uma faculdade tem 80 questões, sendo 40 de português e 40 de matemática. Para os 10 melhores classificados, apresentamos o número de acertos em cada disciplina, em ordem decrescente do total de pontos.

Aluno	1	2	3	4	5	6	7	8	9	10
Português	35	35	34	32	31	30	26	26	24	23
Matemática	31	29	27	28	28	26	27	23	24	24

a) Organizar uma tabela de frequência para cada variável.
b) Fazer uma representação gráfica das tabelas obtidas em (a)
c) Construir a tabela de frequência da variável "*total de pontos*".
d) Comentar sobre a afirmação: "os aprovados são melhores em português do que em matemática". (Nota: Supor que a aprovação está condicionada somente a uma pontuação total igual ou superior a 50)

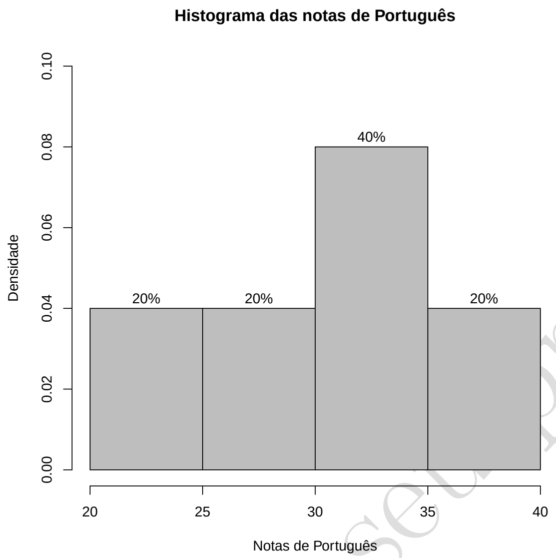
4a) Tabela de frequências, amplitudes e densidades para as notas de Português:

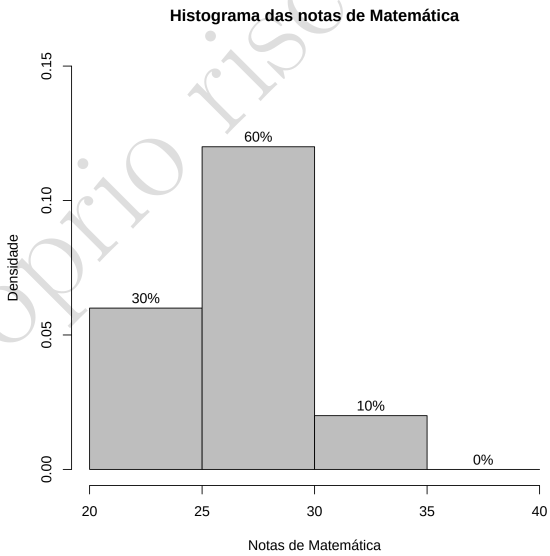
Nota	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
20 $\vdash$ 25	02	$2/10 = 0{,}2$	5	$1/25 = 0{,}04$
25 $\vdash$ 30	02	$2/10 = 0{,}2$	5	$1/25 = 0{,}04$
30 $\vdash$ 35	04	$4/10 = 0{,}4$	5	$2/25 = 0{,}08$
35 $\vdash$ 40	02	$2/10 = 0{,}2$	5	$1/25 = 0{,}04$
TOTAL	10	1,0	-	-

Tabela de frequências, amplitudes e densidades para as notas de Matemática:

Nota	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
20 $\vdash$ 25	03	$3/10 = 0{,}3$	5	$3/50 = 0{,}06$
25 $\vdash$ 30	06	$6/10 = 0{,}6$	5	$3/25 = 0{,}12$
30 $\vdash$ 35	01	$1/10 = 0{,}1$	5	$1/50 = 0{,}02$
35 $\vdash$ 40	00	$0/10 = 0{,}0$	5	0,00
TOTAL	10	1,0	-	-

4b) Histogramas das tabelas obtidas anteriormente.





4c) Tabela com a variável "total de pontos":

Aluno	1	2	3	4	5	6	7	8	9	10
Total de pontos	66	64	61	60	59	56	53	49	48	47

Tabela de frequências, amplitudes e densidades para o total de pontos:

Total de pontos	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
45 $\vdash$ 50	03	$3/10 = 0{,}3$	5	$3/50 = 0{,}06$
50 $\vdash$ 55	01	$1/10 = 0{,}1$	5	$1/50 = 0{,}02$
55 $\vdash$ 60	02	$2/10 = 0{,}2$	5	$1/25 = 0{,}04$
60 $\vdash$ 65	03	$3/10 = 0{,}3$	5	$3/50 = 0{,}06$
65 $\vdash$ 70	01	$1/10 = 0{,}1$	5	$1/50 = 0{,}02$
TOTAL	10	1,0	-	-

4d) Dados para os alunos aprovados (acima de 50 pontos no total):

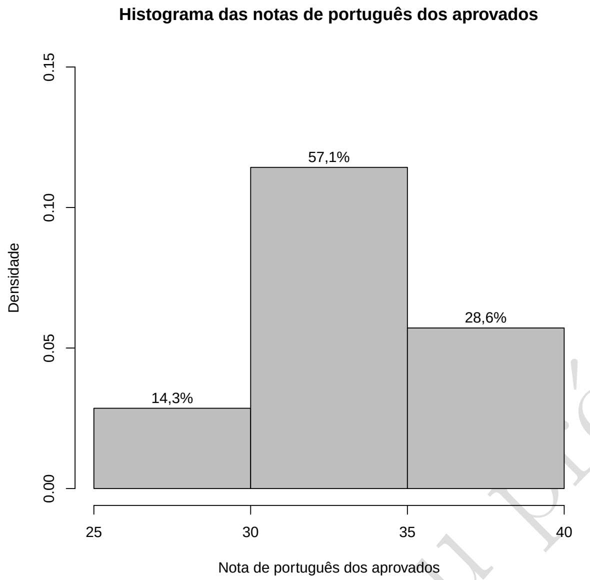
Aluno	1	2	3	4	5	6	7
Português	35	35	34	32	31	30	26
Matemática	31	29	27	28	28	26	27

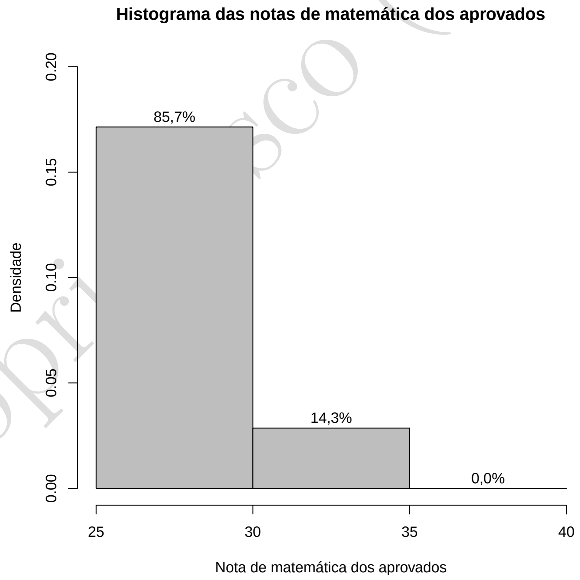
Tabela de frequências, amplitudes e densidades para as notas de Português (alunos aprovados):

Nota	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
$25 \vdash 30$	01	$1/7 \approx 0{,}143$	5	$1/35 \approx 0{,}029$
$30 \vdash 35$	04	$4/7 \approx 0{,}571$	5	$4/35 \approx 0{,}114$
$35 \vdash 40$	02	$2/7 \approx 0{,}286$	5	$2/35 \approx 0{,}057$
TOTAL	07	1,000	-	-

Tabela de frequências, amplitudes e densidades para as notas de Matemática (alunos aprovados):

Nota	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
$25 \vdash 30$	06	$6/7 \approx 0{,}857$	5	$6/35 \approx 0{,}171$
$30 \vdash 35$	01	$1/7 \approx 0{,}143$	5	$1/35 \approx 0{,}029$
$35 \vdash 40$	00	$0/7 = 0{,}000$	5	0,000
TOTAL	07	1,000	-	-





Além de ser visualmente sugestivo nos histogramas, é possível, através dos dados, notar que $\frac{6}{7} \approx 86\%$ dos aprovados (alunos 1 a 7) foram melhores em português que em matemática; ademais, nota-se que as notas da primeira disciplina destes alunos são todas superiores que a segunda maior nota de matemática. Estas informações sugerem que os aprovados foram melhores em português.

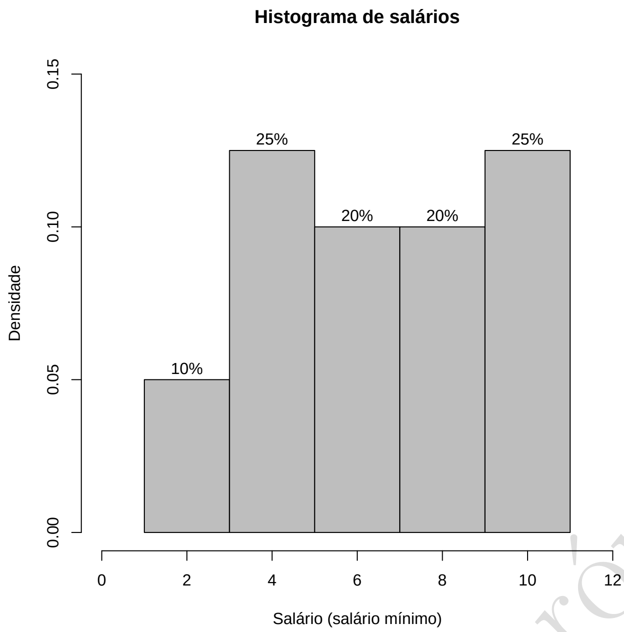
5) Os dados abaixo referem-se ao salário (em salários mínimos) de 20 funcionários administrativos em uma indústria.

10,1	7,3	8,5	5,0	4,2
3,1	2,2	9,0	9,4	6,1
3,3	10,7	1,5	8,2	10,0
4,7	3,5	6,5	8,9	6,1

a) Construir uma tabela de frequência agrupando os dados em faixas a partir de 1 e com amplitude de 2 salários mínimos. Construir, também, o histograma.
b) Analisando o histograma construído, estimar quantos funcionários poderiam financiar uma compra à prestação de R$703,80 mensais de sorte que este valor não seja superior a 30% de seus salários. (Nota: 1 salário mínimo: R$510,00)

5a) Tabela de frequência dos salários dos funcionários:

Salário (salários mínimos)	Frequência absoluta	Frequência relativa	Amplitude	Densidade
1⊢3	2	0,10	2	0,050
3⊢5	5	0,25	2	0,125
5⊢7	4	0,20	2	0,100
7⊢9	4	0,20	2	0,100
9⊢11	5	0,25	2	0,125
TOTAL	20	1,00	-	-



5b) Para que R\$703,80 não ultrapasse 30% do salário, o funcionário deve ter uma renda de, no mínimo, $R\$703,80/0,30 = R\$2346,00$ – ou 4,6 salários mínimos. Dos dados, nota-se que 14 funcionários satisfazem esta condição. Por outro lado, admitindo a hipótese de uma distribuição uniforme dos dados nas barras do histograma quando necessário, pode-se estimar a fração $f_{4,6\vdash 5}$ de pessoas que têm salário entre 4,6 e 5 salários mínimos. Esta estimativa pode ser calculada mediante a equação

$$d_{3\vdash 5} = d_{4,6\vdash 5} \Rightarrow 0,125 = \frac{f_{4,6\vdash 5}}{5-4,6},$$

donde $f_{4,6\vdash 5} = 0,05 = 5\%$. Logo,

$$\underbrace{f_{4,6\vdash 5}}_{5\%} + \underbrace{f_{5\vdash 7}}_{20\%} + \underbrace{f_{7\vdash 9}}_{20\%} + \underbrace{f_{9\vdash 11}}_{25\%} = 70\%$$

é a fração dos funcionários que podem aderir ao supracitado financiamento, e esta quantidade corresponde a $20 \cdot 70\% = 14$ pessoas.

6) Um estudo pretende verificar se o problema da desnutrição em adultos medida pelo peso, em quilogramas, em uma região agrícola (denotada por Região A), é maior do que em uma região industrial (Região B). Para tanto, uma amostra foi tomada em cada região, fornecendo a tabela de frequências a seguir:

Região A

Massa (kg)	n_i
25 ⊢ 40	08
40 ⊢ 50	25
50 ⊢ 60	28
60 ⊢ 70	12
70 ⊢ 100	09

Região B

Massa (kg)	n_i
25 ⊢ 60	10
60 ⊢ 70	34
70 ⊢ 80	109
80 ⊢ 90	111
90 ⊢ 100	55

a) Construir um histograma para cada região.
b) Estimar a média e o desvio padrão para cada região e discutir se o grau de desnutrição em ambas é diferente.

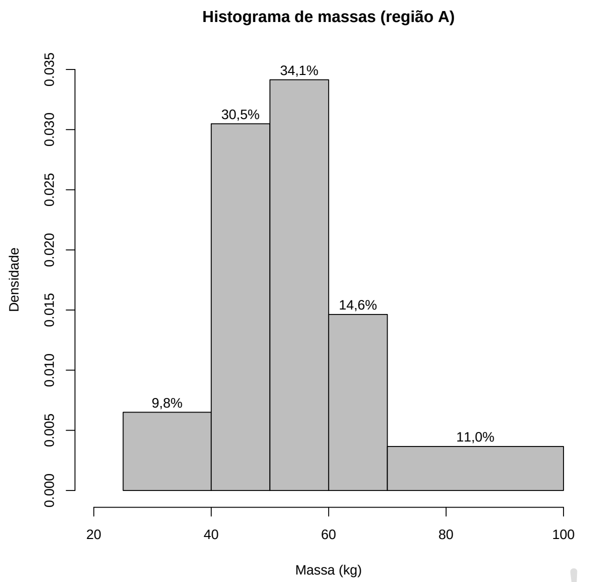
6a) Tabela de frequências, amplitudes e densidades para a região A:

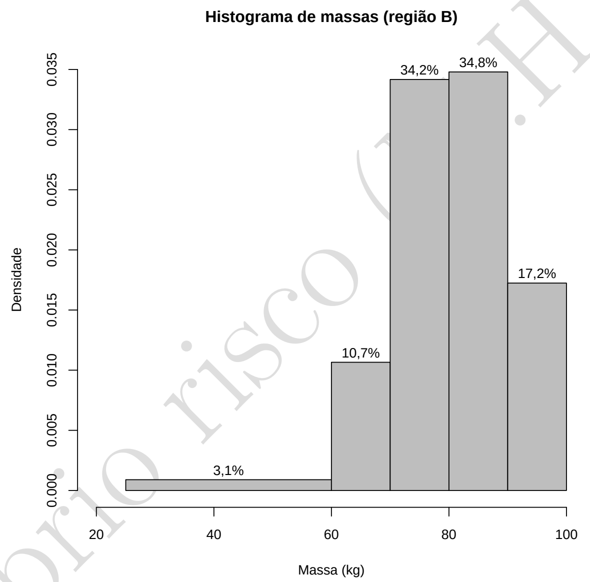
Massa (kg)	a_i	Frequência absoluta	Frequência relativa	Amplitude (kg)	Densidade
25 ⊢ 40	32,5	08	08/82 ≈ 0,098	15	4/615 ≈ 0,0065
40 ⊢ 50	45,0	25	25/82 ≈ 0,305	10	5/164 ≈ 0,0305
50 ⊢ 60	55,0	28	28/82 ≈ 0,341	10	7/205 ≈ 0,0341
60 ⊢ 70	65,0	12	12/82 ≈ 0,146	10	3/205 ≈ 0,0146
70 ⊢ 100	85,0	09	09/82 ≈ 0,110	30	3/820 ≈ 0,0037
TOTAL	-	82	1,000	-	-

Tabela de frequências, amplitudes e densidades para a região B:

Massa (kg)	b_i	Frequência absoluta	Frequência relativa	Amplitude (kg)	Densidade
$25 \vdash 60$	42,5	10	$10/319 \approx 0{,}031$	35	$2/2233 \approx 0{,}00090$
$60 \vdash 70$	65,0	34	$34/319 \approx 0{,}107$	10	$17/1595 \approx 0{,}01066$
$70 \vdash 80$	75,0	109	$109/319 \approx 0{,}342$	10	$109/3190 \approx 0{,}03417$
$80 \vdash 90$	85,0	111	$111/319 \approx 0{,}348$	10	$111/3190 \approx 0{,}03480$
$90 \vdash 100$	95,0	55	$55/319 \approx 0{,}172$	10	$11/638 \approx 0{,}01724$
TOTAL	-	319	1,000	-	-

Histograma das massas dos adultos das duas regiões.





6b) Assumir-se-á uma distribuição uniforme dos dados em cada barra do histograma. Como consequência, admite-se o ponto médio a_i (b_i) do i-ésimo intervalo para a região A (B) como sendo representante do intervalo correspondente para as estimativas.

Estimativa da média $\overline{a}$ (região A) dos $n_A = 82$ dados:

$$\overline{a} = \frac{1}{n_A}\sum_i n_i a_i = \frac{1}{82}\Big(8 \cdot 32,5 + 25 \cdot 45,0 + 28 \cdot 55,0 + 12 \cdot 65,0 + 9 \cdot 85,0\Big) = \frac{2235}{41} \approx 54,5(kg)$$

Estimativa da variância σ_A^2 (região A)

$$\begin{aligned} \sigma_A^2 &= \frac{1}{n_A}\sum_i n_i\left(a_i - \overline{a}\right)^2 = \frac{1}{82}\left[8\left(32,5 - \frac{2235}{41}\right)^2 + 25\left(45,0 - \frac{2235}{41}\right)^2 + \right. \\ & \left. +28\left(55,0 - \frac{2235}{41}\right)^2 + 12\left(65,0 - \frac{2235}{41}\right)^2 + 9\left(85,0 - \frac{2235}{41}\right)^2\right] = \frac{324525}{1681}, \end{aligned}$$

que implica um desvio padrão de $\sigma_A = \sqrt{\frac{324525}{1681}} \approx 13,9(kg)$.

Estimativa da média $\overline{b}$ (região B) dos $n_B = 319$ dados:

$$\overline{a} = \frac{1}{n_B}\sum_i n_i b_i = \frac{1}{319}\Big(10 \cdot 42,5 + 34 \cdot 65,0 + 109 \cdot 75,0 + 111 \cdot 85,0 + 55 \cdot 95,0\Big) = \frac{25470}{319} \approx 79,8(kg)$$

Estimativa da variância σ_B^2 (região B)

$$\begin{aligned} \sigma_B^2 &= \frac{1}{n_B}\sum_i n_i\left(b_i - \overline{b}\right)^2 = \frac{1}{319}\left[10\left(42,5 - \frac{25470}{319}\right)^2 + 34\left(65,0 - \frac{25470}{319}\right)^2 + \right. \\ & \left. +109\left(75,0 - \frac{25470}{319}\right)^2 + 111\left(85,0 - \frac{25470}{319}\right)^2 + 55\left(95,0 - \frac{25470}{319}\right)^2\right] = \frac{25251825}{203522}, \end{aligned}$$

que implica um desvio padrão de $\sigma_B = \sqrt{\frac{25251825}{203522}} \approx 11,1(kg)$.

Nota-se que os intervalos $(\bar{a} - \sigma_A, \bar{a} + \sigma_A) \approx (40,6; 68,4)$ (em kg) e $(\bar{b} - \sigma_B, \bar{b} + \sigma_B) \approx (68,7; 91,0)$ (em kg) não têm intersecção. **Assumindo** que $(\bar{a} - \sigma_A, \bar{a} + \sigma_A)$ ($(\bar{b} - \sigma_B, \bar{b} + \sigma_B)$) compreende uma parcela significativa dos dados da região A(B), a disjunção dos intervalos mencionada sugere que o grau de desnutrição em ambas regiões agrícolas sejam distintas, sendo maior na região A, onde os dados são, também, mais dispersos em relação à média.

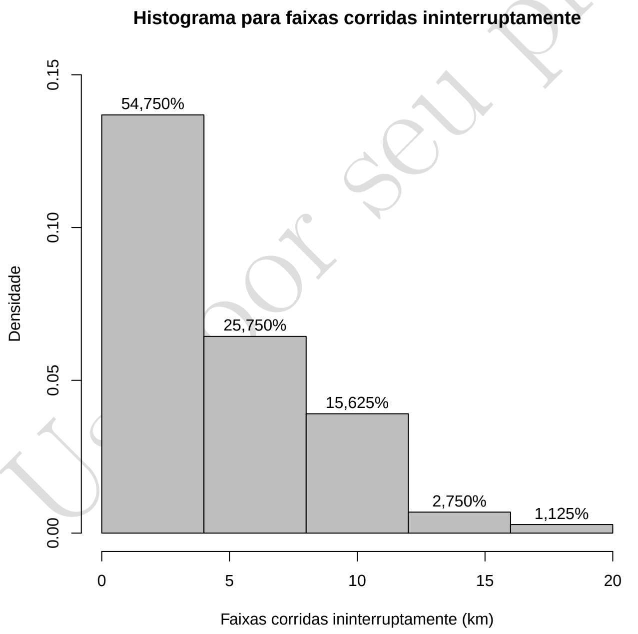
7) Alunos da Escola de Educação Física foram submetidos a um teste de resistência quanto ao número de quilômetros que conseguiriam correr sem parar. Os dados estão apresentados a seguir.

Faixas (km)	Frequência
$0 \vdash 4$	438
$4 \vdash 8$	206
$8 \vdash 12$	125
$12 \vdash 16$	22
$16 \vdash 20$	9

a) Construir o histograma.
b) A tradicional corrida de São Silvestre tem um trajeto de cerca de 15km. Estimar quantos alunos, a princípio, estariam aptos a participar deste evento sem necessitar parar durante a corrida (ignorando as peculiaridades do trajeto).

7a) Tabela de frequências, amplitudes e densidades para faixas corridas ininterruptamente:

Faixas (km)	Frequência absoluta	Frequência relativa	Amplitude (km)	Densidade
$0 \vdash 4$	438	438/800 = 0,54750	4	219/1600 = 0,1368750
$4 \vdash 8$	206	206/800 = 0,25750	4	103/1600 = 0,0643750
$8 \vdash 12$	125	125/800 = 0,15625	4	5/128 = 0,0390625
$12 \vdash 16$	22	22/800 = 0,02750	4	11/1600 = 0,0068750
$16 \vdash 20$	09	09/800 = 0,01125	4	9/3200 = 0,0028125
TOTAL	800	1,00000	-	-



7b) Deve-se estimar o número de alunos que conseguem correr, ininterruptamente, mais de 15km. A fração $f_{15\vdash 16}$ destes alunos, situados na faixa $12 \vdash 16$ (km), pode ser computada pela equação

$$d_{12\vdash 16} = d_{15\vdash 16} \Rightarrow \frac{11}{1600} = \frac{f_{15\vdash 16}}{16-15},$$

onde assumiu-se uma distribuição uniforme dos dados nesta barra do histograma. Esta equação implica $f_{15\vdash 16} = \frac{11}{1600} = 0,006875$ $(0,68750\%)$. Logo, $f_{15\vdash 16} + f_{16\vdash 20} = 1,8125\%$ dos alunos estariam, em princípio, aptos a participar do evento; tal fração corresponde a $800 \cdot 1,8125\% = 14,5$ (aproximadamente 15 alunos).

8) Mostrar que, para um conjunto $\{x_i\}$ de n dados, tem-se as seguintes fórmulas para a média ($\overline{x}$) e desvio padrão (σ):

$$\overline{x} = \frac{1}{n} \sum_{\substack{i\text{: tipo de} \\ \text{variável}}} n_i x_i = \sum_{\substack{i\text{: tipo de} \\ \text{variável}}} f_i x_i$$

e

$$\sigma = \sqrt{\frac{1}{n} \sum_{\substack{i\text{: tipo de} \\ \text{variável}}} n_i \left(x_i - \overline{x}\right)^2} = \sqrt{\sum_{\substack{i\text{: tipo de} \\ \text{variável}}} f_i \left(x_i - \overline{x}\right)^2}$$

onde n_i e f_i são, respectivamente, a frequência absoluta e relativa do i-ésimo tipo de variável.

8) Organizando os n dados de sorte a juntar aqueles que assumem o mesmo valor, obtém-se m subconjuntos ($m \leq n$) com n_α elementos no α-ésimo ($1 \leq \alpha \leq m$) conjunto. Denotando por x_α o elemento do α-ésimo subconjunto (que contém n_α elementos), é imediato que

$$\sum_{i=1}^{n} x_i = \sum_{\alpha=1}^{m} n_\alpha x_\alpha = \sum_{\substack{i\text{: tipo de} \\ \text{variável}}} n_i x_i \, ,$$

donde segue o resultado desejado. Naturalmente, um elemento x_i do primeiro membro **não necessariamente coincide** com o elemento x_i do terceiro membro.

O argumento supracitado aplica-se, *mutatis mutandis*, para o caso da fórmula para o desvio padrão.

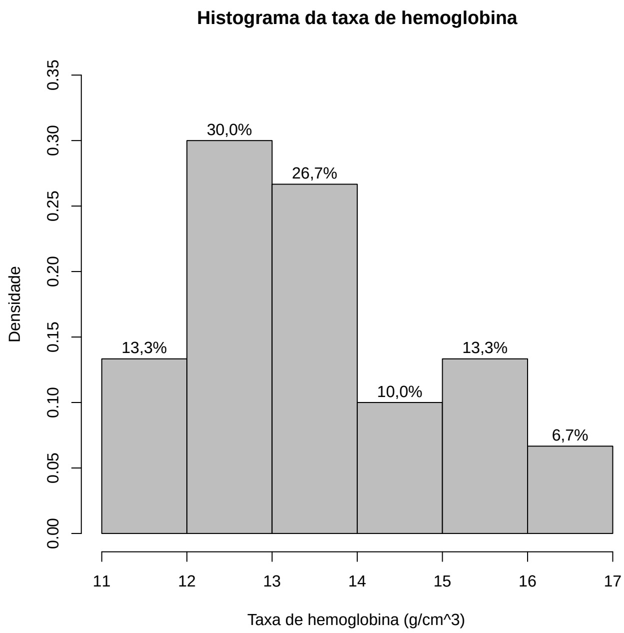
9) Foram feitas medidas da taxa de hemoglobina no sangue (em g/cm^3) em um grupo de voluntários com os seguintes resultados:

11,1	12,2	11,7	12,5	13,9	12,3	14,4	13,6	12,7	12,6
11,3	11,7	12,6	13,4	15,2	13,2	13,0	16,9	15,8	14,7
13,5	12,7	12,3	13,5	15,4	16,3	15,2	12,3	13,7	14,1

a) Organizar os dados em faixas de tamanho 1 (g/cm^3) a partir de $11g/cm^3$ e construir o histograma.
b) Calcular, pelos dados, a média e o desvio padrão.
c) Estimar, pelo histograma, a média e o desvio padrão. Comparar os valores obtidos com os resultados do exercício (8b) e comentar as possíveis diferenças.
d) Taxas abaixo de $12g/cm^3$ ou acima de $16g/cm^3$ são consideradas alteradas e requerem acompanhamento médico. Estimar a porcentagem dessas pessoas que se encontram nestas condições segundo o histograma.

9a) Tabela de frequências, amplitudes e densidades para as taxas de hemoglobina:

Taxa (g/cm^3)	x_i	Frequência absoluta	Frequência relativa	Amplitude (g/cm^3)	Densidade
$11 \vdash 12$	11,5	4	$4/30 \approx 0{,}133$	1	$2/15 \approx 0{,}133$
$12 \vdash 13$	12,5	9	$9/30 = 0{,}300$	1	$3/10 = 0{,}300$
$13 \vdash 14$	13,5	8	$8/30 \approx 0{,}267$	1	$4/15 \approx 0{,}267$
$14 \vdash 15$	14,5	3	$3/30 = 0{,}100$	1	$1/10 = 0{,}100$
$15 \vdash 16$	15,5	4	$4/30 \approx 0{,}133$	1	$2/15 \approx 0{,}133$
$16 \vdash 17$	16,5	2	$2/30 \approx 0{,}067$	1	$1/15 \approx 0{,}067$
TOTAL	-	30	1,000	-	-



9b) Denotando os $n = 30$ dados por $\{y_i\}$, a média $\overline{y}$ é dada por

$$\begin{aligned}\overline{y} &= \frac{1}{n}\sum_{i=1}^{30} y_i = \frac{1}{30}\Big(11,1 + 12,2 + 11,7 + \\ &+12,5 + 13,9 + 12,3 + 14,4 + 13,6 + \\ &+12,7 + 12,6 + 11,3 + 11,7 + 12,6 + \\ &+13,4 + 15,2 + 13,2 + 13,0 + 16,9 + \\ &+15,8 + 14,7 + 13,5 + 12,7 + 12,3 + \\ &+13,5 + 15,4 + 16,3 + 15,2 + 12,3 + \\ &+13,7 + 14,1\Big) = 13,46(g/cm^3)\,.\end{aligned}$$

Por outro lado, a variância σ_y^2 dos dados é

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{30}\left(y_i - \overline{y}\right)^2 \ ,$$

o que leva a

$$\begin{aligned}\sigma_y^2 &= \frac{1}{30}\Big[(11,1-13,46)^2 + (12,2-13,46)^2 + (11,7-13,46)^2 + (12,5-13,46)^2 + \\ &+(13,9-13,46)^2 + (12,3-13,46)^2 + (14,4-13,46)^2 + (13,6-13,46)^2 + (12,7-13,46)^2 + \\ &+(12,6-13,46)^2 + (11,3-13,46)^2 + (11,7-13,46)^2 + (12,6-13,46)^2 + (13,4-13,46)^2 + \\ &+(15,2-13,46)^2 + (13,2-13,46)^2 + (13,0-13,46)^2 + (16,9-13,46)^2 + (15,8-13,46)^2 + \\ &+(14,7-13,46)^2 + (13,5-13,46)^2 + (12,7-13,46)^2 + (12,3-13,46)^2 + (13,5-13,46)^2 + \\ &+(15,4-13,46)^2 + (16,3-13,46)^2 + (15,2-13,46)^2 + (12,3-13,46)^2 + (13,7-13,46)^2 + \\ &+(14,1-13,46)^2\Big] = \frac{8024}{3750} \approx 2,14\,,\end{aligned}$$

donde se tem o desvio padrão $\sigma = \sqrt{\frac{8024}{3750}} \approx 1,46(g/cm^3)$.

9c) Assumir-se-á uma distribuição uniforme dos dados em cada barra do histograma, o que possibilita eleger o ponto médio x_i do i-ésimo intervalo (da variável na tabela de frequência) como seu representante para as estimativas abaixo. Desta forma, a média $\overline{x}$ e a variância σ^2, determinados a partir do histograma, são dados, respectivamente, por

$$\overline{x} = \frac{1}{n}\sum_i n_i x_i = \frac{1}{30}\Big(4\cdot 11,5 + 9\cdot 12,5 + 8\cdot 13,5 + 3\cdot 14,5 + 4\cdot 15,5 + 2\cdot 16,5\Big) = 13,5(g/cm^3)$$

e

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_i n_i\left(x_i - \overline{x}\right)^2 = \frac{1}{30}\Big[4\,(11,5-13,5)^2 + 9\,(12,5-13,5)^2 + 8\,(13,5-13,5)^2 + \\ &+3\,(14,5-13,5)^2 + 4\,(15,5-13,5)^2 + 2\,(16,5-13,5)^2\Big] = \frac{31}{15} \approx 2,07\,,\end{aligned}$$

donde se tem o desvio padrão $\sigma = \sqrt{\frac{31}{15}} \approx 1,44(g/cm^3)$.

As diferenças nos valores das médias ($\overline{y}$ e $\overline{x}$) e desvios padrão (σ_y e σ) são consequências da hipótese adotada para as estimativas: os dados não se distribuem de forma uniforme em cada intervalo do histograma.

9d) Segundo os dados, a porcentagem das pessoas com taxas alteradas é dada por

$$f_{11\vdash 12} + f_{16\vdash 17} = \frac{2}{15} + \frac{2}{30} = \frac{1}{5} = 20\%\,.$$

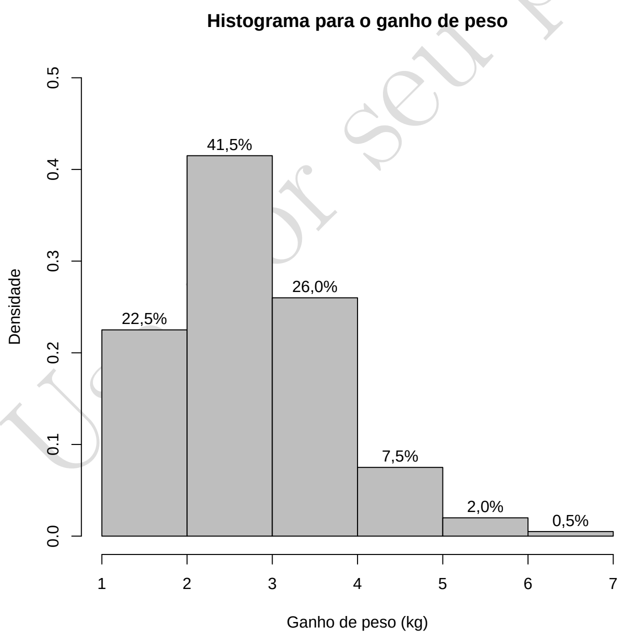
10) Uma nova ração foi fornecida a suínos recém desmamados e deseja-se avaliar sua eficiência. A ração tradicional dava um ganho de peso ao redor de 3,5kg em um mês. A seguir, apresenta-se os dados referentes ao ganho, em quilogramas, para essa nova ração, aplicada em um mês em 200 animais nas condições acima.

Ganho (kg)	Frequência
1,0 $\vdash$ 2,0	45
2,0 $\vdash$ 3,0	83
3,0 $\vdash$ 4,0	52
4,0 $\vdash$ 5,0	15
5,0 $\vdash$ 6,0	4
6,0 $\vdash$ 7,0	1

a) Construir o histograma.
b) Estimar a média e o desvio padrão.
c) Analisar se a nova ração é mais eficiente que a tradicional.

10a) Tabela de frequências, amplitudes e densidades para o ganho de peso:

Ganho (kg)	x_i	Frequência absoluta	Frequência relativa	Amplitude (kg)	Densidade
1,0 $\vdash$ 2,0	1,5	45	45/200 = 0,225	1,0	9/40 = 0,225
2,0 $\vdash$ 3,0	2,5	83	83/200 = 0,415	1,0	83/200 = 0,415
3,0 $\vdash$ 4,0	3,5	52	52/200 = 0,260	1,0	13/50 = 0,260
4,0 $\vdash$ 5,0	4,5	15	15/200 = 0,075	1,0	3/40 = 0,075
5,0 $\vdash$ 6,0	5,5	04	4/200 = 0,020	1,0	1/50 = 0,020
6,0 $\vdash$ 7,0	6,5	01	1/200 = 0,005	1,0	1/200 = 0,005
TOTAL	-	200	1,000	-	-



10b) Admitindo uma distribuição uniforme dos dados nas barras do histograma, pode-se tomar o ponto médio x_i do i-ésimo intervalo como sendo seu representante para as estimativas da média $\overline{x}$ e variância σ^2.
Estimativa da média:

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_i n_i x_i = \sum_i f_i x_i = 0,225\cdot 1,5 + \\ &\quad +0,415\cdot 2,5 + 0,260\cdot 3,5 + \\ &\quad +0,075\cdot 4,5 + 0,020\cdot 5,5 + 0,005\cdot 6,5 \\ &= 2,765(kg) \end{aligned}$$

Estimativa da variância:

$$\begin{aligned} \sigma^2 &= \frac{1}{n}\sum_i n_i\,(x_i-\overline{x})^2 = \frac{1}{200}\Big[\\ &\quad 45\,(1,5-2,765)^2 + 83\,(2,5-2,765)^2 + \\ &\quad +52\,(3,5-2,765)^2 + 15\,(4,5-2,765)^2 + \\ &\quad +4\,(5,5-2,765)^2 + 1\,(6,5-2,765)^2\Big] \\ &= \frac{38991}{40000}\,, \end{aligned}$$

que implica um desvio padrão de $\sigma = \sqrt{\frac{38991}{40000}} \approx 0,987$ (kg).

10c) Admitindo uma distribuição uniforme dos dados no intervalo $3,0 \vdash 4,0$ do histograma, tem-se

$$d_{3,0\vdash 4,0} = d_{3,0\vdash 3,5} \Leftrightarrow \frac{13}{50} = \frac{f_{3,0\vdash 3,5}}{3,5-3,0} \Rightarrow f_{3,0\vdash 3,5} = \frac{13}{100} = 0,13\,,$$

donde se estima que cerca de $f_{1,0\vdash 2,0} + f_{2,0\vdash 3,0} + f_{3,0\vdash 3,5} = \frac{77}{100} = 0,77 = 77\%$ das rações novas apresentam um ganho inferior em relação à ração tradicional, o que sugere esta última ser mais eficiente **segundo este ponto de vista**. Por outro lado, nota-se que o ganho de $3,5kg$ por mês está previsto no intervalo $(\overline{x}-\sigma, \overline{x}+\sigma) \approx (1,778; 3,752)$ (em kg); admitindo como critério de tipicidade os valores situados nesta região, a ração atual contempla a eficiência da ração tradicional.

11) Num estudo sobre rotatividade de mão-de-obra na indústria, anotou-se o número de empregos nos últimos 3 anos para operários especializados e não especializados.

Não especializados

Empregos	n_i
1	106
2	222
3	338
4	292
5	164
Total	1122

Especializados

Empregos	n_i
1	210
2	342
3	109
4	91
5	35
Total	787

a) Construir um histograma para cada grupo de operários (especializados e não especializados).
b) Analisar se os trabalhadores especializados trocam menos de emprego.
c) Juntar as informações das duas tabelas em uma só e obter um histograma da rotatividade de mão-de-obra na indústria (sem diferenciar a especialização).

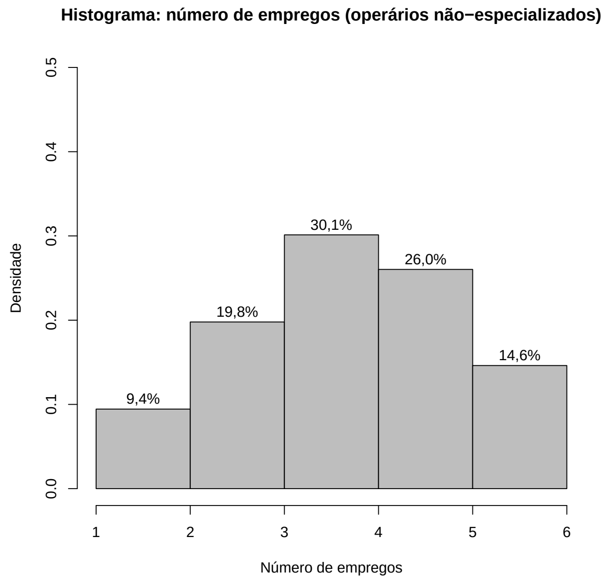
d) Com base no histograma do item (c), estimar por quantos empregos passam, no mínimo, os 50% dos operários que mais trocam de emprego.

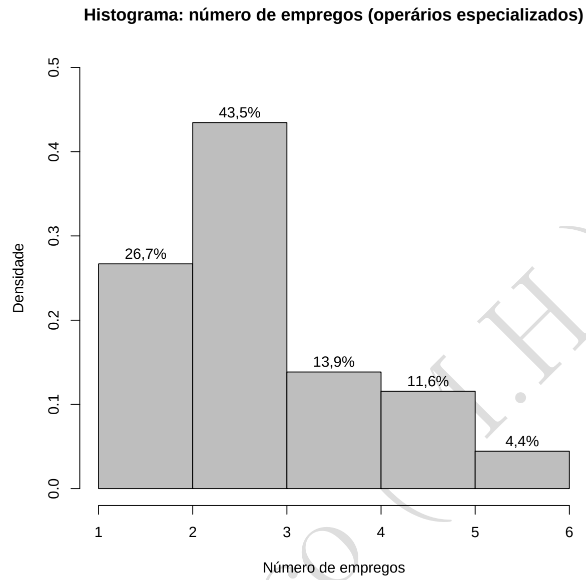
11a) Tabela de frequências, amplitudes e densidades para número de empregos – operários não especializados:

Empregos	Frequência absoluta	Frequência relativa	Amplitude	Densidade
1	106	$106/1122 \approx 0{,}094$	1	$53/561 \approx 0{,}094$
2	222	$222/1122 \approx 0{,}198$	1	$37/187 \approx 0{,}198$
3	338	$338/1122 \approx 0{,}301$	1	$169/561 \approx 0{,}301$
4	292	$292/1122 \approx 0{,}260$	1	$146/561 \approx 0{,}260$
5	164	$164/1122 \approx 0{,}146$	1	$82/561 \approx 0{,}146$
TOTAL	1122	$\approx 1{,}000$	-	-

Tabela de frequências, amplitudes e densidades para número de empregos – operários especializados:

Empregos	Frequência absoluta	Frequência relativa	Amplitude	Densidade
1	210	$210/787 \approx 0{,}267$	1	$210/787 \approx 0{,}267$
2	342	$342/787 \approx 0{,}435$	1	$342/787 \approx 0{,}435$
3	109	$109/787 \approx 0{,}139$	1	$109/787 \approx 0{,}139$
4	091	$91/787 \approx 0{,}116$	1	$91/787 \approx 0{,}116$
5	035	$35/787 \approx 0{,}044$	1	$35/787 \approx 0{,}044$
TOTAL	787	$\approx 1{,}000$	-	-

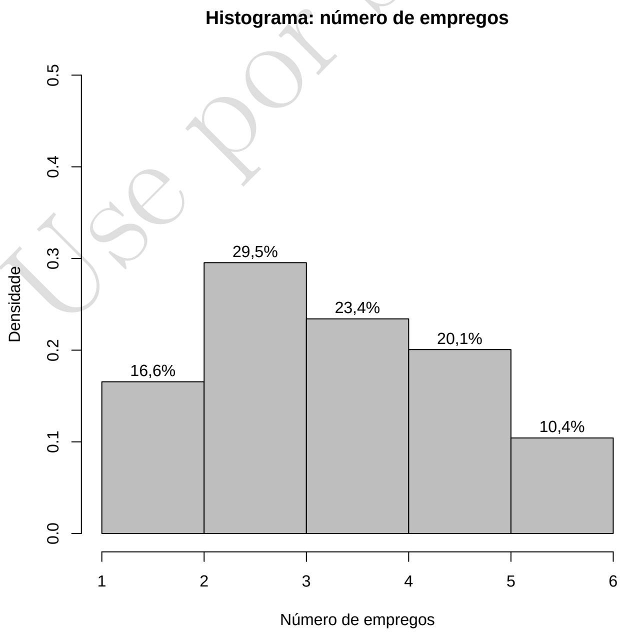




11b) Notando, pela tabela de frequência, que $\frac{338}{1122} + \frac{292}{1122} + \frac{164}{1122} = \frac{397}{561} \approx 70,77\%$ dos operários não-especializados trocam de emprego três ou mais vezes, ao passo que $\frac{210}{787} + \frac{342}{787} = \frac{552}{787} \approx 70,14\%$ dos especializados trocam de emprego menos de três vezes, há indícios apontando para o fato dos primeiros trocarem mais de emprego que os últimos.

11c) Tabela de frequências, amplitudes e densidades para número de empregos.

Empregos	Frequência absoluta	Frequência relativa	Amplitude	Densidade
1 ($1 \vdash 2$)	316	$316/1909 \approx 0{,}166$	1	$316/1909 \approx 0{,}166$
2 ($2 \vdash 3$)	564	$564/1909 \approx 0{,}295$	1	$564/1909 \approx 0{,}295$
3 ($3 \vdash 4$)	447	$447/1909 \approx 0{,}234$	1	$447/1909 \approx 0{,}234$
4 ($4 \vdash 5$)	383	$383/1909 \approx 0{,}201$	1	$383/1909 \approx 0{,}201$
5 ($5 \vdash 6$)	199	$199/1909 \approx 0{,}104$	1	$199/1909 \approx 0{,}104$
TOTAL	1909	1,000	-	-



11d) O número mínimo de empregos pelos quais passam os 50% dos operários que mais trocam de emprego é a mediana Md dos dados. Assumindo a uniformidade da distribuição de dados na barra $3 \vdash 4$ do histograma (onde se situa a mediana), tem-se

$$d_{3\vdash 4} = d_{3\vdash \mathrm{Md}} \Rightarrow \frac{447}{1909} = \frac{\frac{1}{2} - \left(\frac{316}{1909} + \frac{564}{1909}\right)}{\mathrm{Md} - 3},$$

donde se tem $\mathrm{Md} = \frac{19}{6} \approx 3,17$.

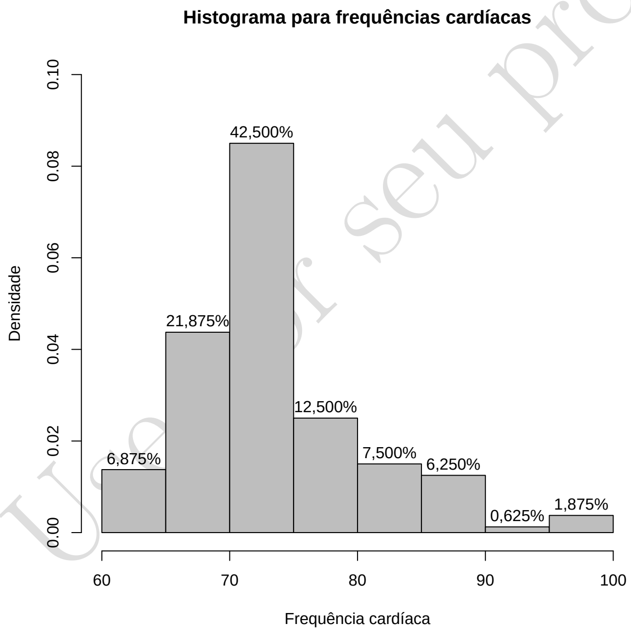
12) Como parte de uma avaliação médica, foi medida a frequência cardíaca de um grupo de pessoas. Os dados (frequência cardíaca em batidas por minuto) são apresentados em seguida.

Frequência cardíaca	Frequência
$60 \vdash 65$	11
$65 \vdash 70$	35
$70 \vdash 75$	68
$75 \vdash 80$	20
$80 \vdash 85$	12
$85 \vdash 90$	10
$90 \vdash 95$	01
$95 \vdash 100$	3

a) Obter o histograma.
b) Para a faixa etária dessas pessoas, frequências cardíacas que estejam abaixo de 62 e acima de 92 requerem acompanhamento médico. Estimar a porcentagem de pessoas nessas condições.
c) Para a faixa etária dessas pessoas, uma frequência cardíaca ao redor de 72 batidas por minuto é considerada padrão. Analisar se, de modo geral, essas pessoas encaixam nesse caso.

12a) Tabela de frequências, amplitudes e densidades para frequências cardíacas (em batidas por minuto):

Frequência cardíaca	x_i	Frequência absoluta	Frequência relativa	Amplitude	Densidade
$60 \vdash 65$	62,5	11	$11/160 = 0{,}06875$	5	$11/800 = 0{,}01375$
$65 \vdash 70$	67,5	35	$35/160 = 0{,}21875$	5	$7/160 = 0{,}04375$
$70 \vdash 75$	72,5	68	$68/160 = 0{,}42500$	5	$17/200 = 0{,}08500$
$75 \vdash 80$	77,5	20	$20/160 = 0{,}12500$	5	$1/40 = 0{,}02500$
$80 \vdash 85$	82,5	12	$12/160 = 0{,}07500$	5	$3/200 = 0{,}01500$
$85 \vdash 90$	87,5	10	$10/160 = 0{,}06250$	5	$1/80 = 0{,}01250$
$90 \vdash 95$	92,5	01	$1/160 = 0{,}00625$	5	$1/800 = 0{,}00125$
$95 \vdash 100$	97,5	03	$3/160 = 0{,}01875$	5	$3/800 = 0{,}00375$
TOTAL	-	160	1,00000	-	-



12b) Assumindo uma distribuição uniforme dos dados nas barras do histograma nos intervalos $60 \vdash 65$ e $90 \vdash 95$, deve-se determinar a porcentagem $f_{60\vdash 62}$ de pessoas com frequência cardíaca inferior a 62 (e superior a 60, que é a cota inferior entre os dados) e a fração de pessoas $f_{92\vdash 95}$ com frequência cardíaca entre 92 e 95. A primeira é calculada através de

$$d_{60\vdash 65} = d_{60\vdash 62} \Rightarrow \frac{11}{800} = \frac{f_{60\vdash 62}}{62-60},$$

donde se tem $f_{60\vdash 62} = \frac{11}{400} = 2,75\%$. Por outro lado, a porcentagem $f_{92\vdash 95}$ pode ser computada mediante a equação

$$d_{90\vdash 95} = d_{92\vdash 95} \Rightarrow \frac{1}{800} = \frac{f_{92\vdash 95}}{95-92},$$

donde segue $f_{92\vdash 95} = \frac{3}{800} = 0,375\%$. Desta forma, a porcentagem das pessoas que requerem acompanhamento médico é $f_{60\vdash 62} + f_{92\vdash 95} + f_{95\vdash 100} = 2,75\% + 0,375\% + 1,875\% = 5,00\%$.

12c) A fim de estudar a questão, analisar-se-á, inicialmente, a dispersão dos dados baseando-se no desvio padrão. Admitir-se-á como sendo padrão as frequências cardíacas situadas até a uma distância de um desvio padrão em relação à média. Assumindo uniformidade na distribuição dos dados em cada barra do histograma, pode-se tomar o ponto médio x_i de cada intervalo como sendo seu respectivo representante. Esta hipótese permite estimar a média e a variância dos $n = 160$ dados apresentados.

Estimativa da média $\overline{x}$:

$$\begin{aligned}\overline{x} &= \frac{1}{n}\sum_i n_i x_i = \sum_i f_i x_i = 0,06875 \cdot 62,5 + 0,21875 \cdot 67,5 + 0,42500 \cdot 72,5 + 0,12500 \cdot 77,5 + \\ &\quad +0,07500 \cdot 82,5 + 0,06250 \cdot 87,5 + 0,00625 \cdot 92,5 + 0,01875 \cdot 97,5 = 73,625 \text{ (batimentos/minuto)}. .\end{aligned}$$

Estimativa da variância σ^2:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_i n_i\,(x_i - \overline{x})^2 = \frac{1}{160}\Big[11\,(62,5 - 73,625)^2 + 35\,(67,5 - 73,625)^2 + \\ &\quad +68\,(72,5 - 73,625)^2 + 20\,(77,5 - 73,625)^2 + 12\,(82,5 - 73,625)^2 + 10\,(87,5 - 73,625)^2 + \\ &\quad +1\,(92,5 - 73,625)^2 + 3\,(97,5 - 73,625)^2\Big] = \frac{3199}{64},\end{aligned}$$

que implica um desvio padrão de $\sigma = \sqrt{\frac{3199}{64}} \approx 7,070$.

Nota-se que o valor de 72 batidas por minuto está previsto no intervalo $(\overline{x} - \sigma, \overline{x} + \sigma) \approx (66,555; 80,695)$, que foi admitido como a região dos dados "típicos". **Deste ponto de vista, e por este critério**, há indícios de que o conjunto de pessoas analisado encaixa-se no padrão.

13) Dado um conjunto $\{x_i\}$ de n (> 1) dados cuja média e variância são, respectivamente, $\overline{x}$ e σ^2, mostrar que

a) $\displaystyle\sum_{i=1}^{n} (x_i - \overline{x}) = 0$ b) $\sigma^2 = \overline{x^2} - (\overline{x})^2$ $\left(\text{por definição, } \overline{x^2} := \frac{1}{n}\sum_{i=1}^{n} x_i^2\right)$

13a) Levando em consideração a definição de média, tem-se

$$\sum_{i=1}^{n} (x_i - \overline{x}) = \underbrace{(x_1 - \overline{x}) + \cdots + (x_n - \overline{x})}_{n \text{ termos}} = x_1 + \cdots + x_n - (\underbrace{\overline{x} + \cdots + \overline{x}}_{n \text{ termos}}) = \underbrace{\sum_{i=1}^{n} x_i}_{n\overline{x}} - n\overline{x} = n\overline{x} - n\overline{x} = 0\,.$$

13b) Invocando a definição de $\overline{x^2}$ e de média, tem-se

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n}\Big[\,(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2\,\Big] \\ &= \frac{1}{n}\Big[\,\left(x_1^2 - 2x_1\overline{x} + \overline{x}^2\right) + \cdots + \left(x_n^2 - 2x_n\overline{x} + \overline{x}^2\right)\Big] \\ &= \frac{1}{n}\Big[x_1^2 + \cdots + x_n^2 - (2\overline{x}x_1 + \cdots + 2\overline{x}x_n) + \underbrace{\overline{x}^2 + \cdots + \overline{x}^2}_{n \text{ termos}}\Big] \\ &= \frac{1}{n}\Big[\underbrace{\sum_{i=1}^{n} x_i^2}_{n\overline{x^2}} - 2\overline{x}(x_1 + \cdots + x_n) + n\overline{x}^2\Big] \\ &= \frac{1}{n}\Big[n\overline{x^2} - 2\overline{x}\underbrace{\sum_{i=1}^{n} x_i}_{n\overline{x}} + n\overline{x}^2\Big] \\ &= \frac{1}{n}\Big[n\overline{x^2} - 2n\overline{x}^2 + n\overline{x}^2\Big] = \frac{1}{n}\Big[n\overline{x^2} - n\overline{x}^2\Big] = \overline{x^2} - \overline{x}^2\,.\end{aligned}$$

14) Em um estudo sobre a renda de professores do ensino fundamental, coletou-se os dados referentes ao salário desses. Os resultados foram dispostos na tabela abaixo.

Salário (salários mínimos)	Frequência absoluta
$1 \vdash 3$	15
$3 \vdash 5$	25
$5 \vdash 7$	18
$7 \vdash 9$	9
$9 \vdash 10$	4
TOTAL	71

a) Estimar a média $\overline{s}$ e desvio padrão σ. Explicitar a hipótese assumida para os cálculos.
b) Estimar a porcentagem dos professores situados no intervalo $(\overline{s}-\sigma, \overline{s}+\sigma)$.

14a) Assumindo uma distribuição uniforme dos dados em cada intervalo, é razoável escolher o representante s_i de cada faixa como sendo o ponto médio (do respectivo intervalo). Desta forma, chega-se à seguinte tabela de frequências:

Salário (salários mínimos)	s_i	Frequência	Frequência relativa	Amplitude	Densidade
$1 \vdash 3$	2	15	$15/71 \approx 0{,}2113$	2	$15/142 \approx 0{,}1056$
$3 \vdash 5$	4	25	$25/71 \approx 0{,}3521$	2	$25/142 \approx 0{,}1761$
$5 \vdash 7$	6	18	$18/71 \approx 0{,}2535$	2	$9/71 \approx 0{,}1268$
$7 \vdash 9$	8	9	$9/71 \approx 0{,}1268$	2	$9/142 \approx 0{,}0634$
$9 \vdash 10$	9,5	4	$4/71 \approx 0{,}0563$	1	$4/71 \approx 0{,}0563$
TOTAL	-	71	1	-	-

Estimativa da média:

$$\begin{aligned}
\overline{s} &= \frac{1}{n}\sum_i n_i s_i = \sum_i f_i s_i \\
&= \frac{15}{71}\cdot 2 + \frac{25}{71}\cdot 4 + \frac{18}{71}\cdot 6 + \frac{9}{71}\cdot 8 + \frac{4}{71}\cdot 9,5 = \frac{348}{71} \approx 4,9014 \text{ (salários mínimos)}\,.
\end{aligned}$$

Estimativa da variância:

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_i n_i\left(s_i-\overline{s}\right)^2 \\
&= \frac{1}{71}\left[15\left(2-\frac{348}{71}\right)^2 + 25\left(4-\frac{348}{71}\right)^2 + 18\left(6-\frac{348}{71}\right)^2 + 9\left(8-\frac{348}{71}\right)^2 + 4\left(9,5-\frac{348}{71}\right)^2\right] \\
&= \frac{24091}{5041}\,.
\end{aligned}$$

O desvio padrão é, pois, $\sigma = \sqrt{\frac{24091}{5041}} \approx 2,1861$ (salários mínimos).

14b) Deve-se estimar a área do histograma compreendida entre $(\overline{s}-\sigma, \overline{s}+\sigma) \approx (2,7153; 7,0875)$, que é o valor $f_{\overline{s}-\sigma \vdash 3} + f_{3\vdash 5} + f_{5\vdash 7} + f_{7\vdash \overline{s}+\sigma}$. Como a hipótese da distribuição uniforme dos dados em cada intervalo do histograma fora assumida, sabe-se que

$$d_{1\vdash 3} = d_{\overline{s}-\sigma\vdash 3} = \frac{f_{\overline{s}-\sigma\vdash 3}}{3-(\overline{s}-\sigma)} \Rightarrow f_{\overline{s}-\sigma\vdash 3} = \frac{15}{142}\left(\sqrt{\frac{24091}{5041}} - \frac{135}{71}\right) \approx 0,0301\,.$$

e

$$d_{7\vdash 9} = d_{7\vdash\overline{s}+\sigma} = \frac{f_{7\vdash\overline{s}+\sigma}}{(\overline{s}+\sigma)-7} \Rightarrow f_{7\vdash\overline{s}+\sigma} = \frac{9}{142}\left(\sqrt{\frac{24091}{5041}} - \frac{149}{71}\right) \approx 0,0055\,.$$

A fração dos professores com salário situado no intervalo $(\overline{s}-\sigma, \overline{s}+\sigma)$ é, então,

$$\begin{aligned}
f_{\overline{s}-\sigma\vdash 3} + f_{3\vdash 5} + f_{5\vdash 7} + f_{7\vdash\overline{s}+\sigma} &= \frac{15}{142}\left(\sqrt{\frac{24091}{5041}} - \frac{135}{71}\right) + \frac{25}{71} + \frac{18}{71} + \frac{9}{142}\left(\sqrt{\frac{24091}{5041}} - \frac{149}{71}\right) \\
&= \frac{12}{71}\sqrt{\frac{24091}{5041}} + \frac{1370}{71^2} = 0,6412\cdots \approx 64,13\%\,.
\end{aligned}$$

15) Num estudo sobre consumo de combustível, 200 automóveis do mesmo ano e modelo tiveram seu consumo observado durante 1000 quilômetros. A informação obtida é apresentada na tabela abaixo em km/litro.

Faixas	Frequência
$7 \vdash 8$	27
$8 \vdash 9$	29
$9 \vdash 10$	46
$10 \vdash 11$	43
$11 \vdash 12$	55

a) Estimar o desvio padrão de consumo (por litro).
b) Estimar o número de carros cuja taxa de consumo situa-se no intervalo de até um desvio padrão em torno da média.
c) Estimar a menor taxa de consumo observada dentre os 90% dos carros que mais consomem combustível (por litro).

15a) Assumindo uma distribuição uniforme dos dados em cada faixa de valores, pode-se escolher o ponto médio como o ponto representante x_i em cada intervalo. Desta forma, tem-se

Faixa (km/l)	x_i (km/l)	Frequência absoluta	Frequência relativa	Amplitude	Densidade
$7 \vdash 8$	7,5	27	$27/200 = 0{,}135$	1	$27/200 = 0{,}135$
$8 \vdash 9$	8,5	29	$29/200 = 0{,}145$	1	$29/200 = 0{,}145$
$9 \vdash 10$	9,5	46	$46/200 = 0{,}230$	1	$23/100 = 0{,}230$
$10 \vdash 11$	10,5	43	$43/200 = 0{,}215$	1	$43/200 = 0{,}215$
$11 \vdash 12$	11,5	55	$55/200 = 0{,}275$	1	$11/40 = 0{,}275$
TOTAL	-	200	1	-	-

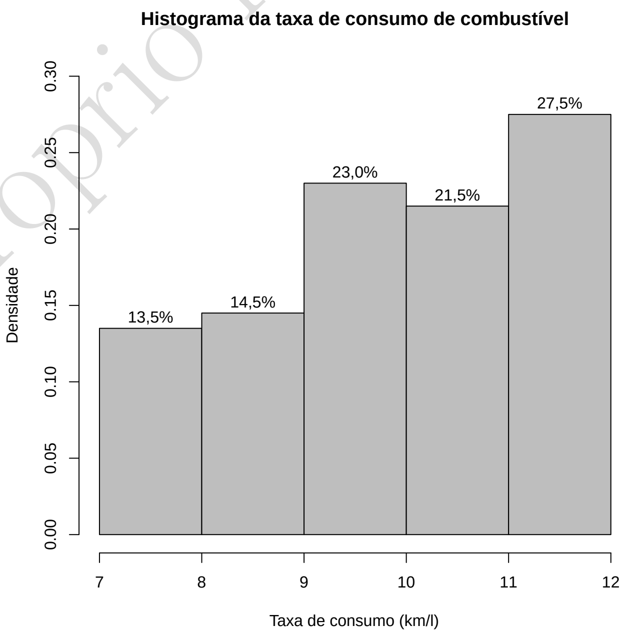
Estimativa da média $\overline{x}$:

$$\begin{aligned}\overline{x} &= \sum_i f_i x_i = 0,135 \cdot 7,5 + 0,145 \cdot 8,5 + \\ &\quad +0,230 \cdot 9,5 + 0,215 \cdot 10,5 + \\ &\quad +0,275 \cdot 11,5 = 9,85 \ (km/l)\,.\end{aligned}$$

Estimativa da variância σ^2 dos $n = 200$ dados:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_i n_i\,(x_i - \overline{x})^2 \\ &= \frac{1}{200}\Big[27\,(7,5-9,85)^2 + \\ &\quad +29\,(8,5-9,85)^2 + 46\,(9,5-9,85)^2 + \\ &\quad +43\,(10,5-9,85)^2 + 55\,(11,5-9,85)^2\Big] \\ &= \frac{751}{400}\,,\end{aligned}$$

o que implica um desvio padrão de $\sigma = \sqrt{\frac{751}{400}} \approx 1,3702\ (km/l)$.



15b) Deve-se estimar o número de carros cuja taxa de consumo encontra-se no intervalo $(\overline{x}-\sigma, \overline{x}+\sigma) \approx (8,48; 11,22)$, lembrando que a hipótese de distribuição uniforme dos dados em cada barra do histograma fora assumida. Como o ponto $\overline{x}-\sigma \approx 8,48$ localiza-se no intervalo $8 \vdash 9$ e $\overline{x}+\sigma \approx 11,22$ no intervalo $11 \vdash 12$, a fração de carros que atende as condições especificadas é dada por $f_{\overline{x}-\sigma \vdash 9} + f_{9 \vdash 10} + f_{10 \vdash 11} + f_{11 \vdash \overline{x}+\sigma}$. Como

$$d_{8\vdash 9} = d_{\overline{x}-\sigma\vdash 9} \Rightarrow \frac{29}{200} = \frac{f_{\overline{x}-\sigma\vdash 9}}{9-(\overline{x}-\sigma)} \Rightarrow f_{\overline{x}-\sigma\vdash 9} = \frac{29}{200}\left(\sqrt{\frac{751}{400}} - \frac{17}{20}\right) \approx 0,0754$$

e

$$d_{11\vdash 12} = d_{11\vdash\overline{x}+\sigma} \Rightarrow \frac{11}{40} = \frac{f_{11\vdash\overline{x}+\sigma}}{(\overline{x}+\sigma)-11} \Rightarrow f_{11\vdash\overline{x}+\sigma} = \frac{11}{40}\left(\sqrt{\frac{751}{400}} - \frac{23}{20}\right) \approx 0,0606\,,$$

tem-se

$$f_{\overline{x}-\sigma\vdash 9}+f_{9\vdash 10}+f_{10\vdash 11}+f_{11\vdash \overline{x}+\sigma} = \frac{29}{200}\left(\sqrt{\frac{751}{400}}-\frac{17}{20}\right)+\frac{46}{200}+\frac{43}{200}+\frac{11}{40}\left(\sqrt{\frac{751}{400}}-\frac{23}{20}\right)$$
$$= \frac{11}{2000}+\frac{21}{50}\sqrt{\frac{751}{400}} \approx 0,5810 = 58,10\% \text{ (de um total de 200 carros)},$$

o que implica cerca de 116 automóveis.

15c) Os 90% dos carros que mais consomem combustível são aqueles que têm menos distância percorrida por litro. Pela tabela de frequência, nota-se que os 90% dos automóveis com pior desempenho percorrem até y quilômetros por litro, onde y pertence ao intervalo $11 \vdash 12$. Logo, lembrando que a hipótese de distribuição uniforme dos dados em cada barra do histograma fora assumida, chega-se a

$$d_{11\vdash 12} = d_{y\vdash 12} \Rightarrow \frac{11}{40} = \frac{0,10}{12-y} \Rightarrow y = \frac{128}{11} \approx 11,6364\ (km/l).$$

A menor taxa de consumo estimada dentre os 90% dos carros que mais consomem combustível é de $\frac{128}{11} \approx 11,64 (km/l)$.

16) O tempo, em horas, necessário para um certo medicamento fazer efeito é apresentado abaixo:

0,21	2,71	2,12	2,81	3,30	0,15	0,54	3,12	0,80	1,76
1,14	0,16	0,31	0,91	0,18	0,04	1,16	2,16	1,48	0,63

a) Calcular a média e o desvio padrão para o conjunto de dados.
b) Construir uma tabela de frequência para classes com amplitude de 0,5 hora, começando do zero.
c) Suponha que o conjunto original de dados foi perdido e só se dispõe da tabela construída em (b). Construir o histograma a partir desta tabela e, utilizando alguma hipótese conveniente (e razoável), estimar a média e a variância. Comentar as possíveis diferenças encontradas.

16a) Cálculo da média $\overline{t}$ para os $n = 20$ dados do conjunto $\{t_i\}$:

$$\overline{t} = \frac{1}{n}\sum_i t_i = \frac{1}{20}\Big[0,21+2,71+2,12+2,81+3,30+0,15+0,54+3,12+0,80+1,76+1,14+$$
$$+0,16+0,31+0,91+0,18+0,04+1,16+2,16+1,48+0,63\Big] = \frac{2569}{2000} = 1,2845 \text{ horas}.$$

Cálculo da variância σ^2:

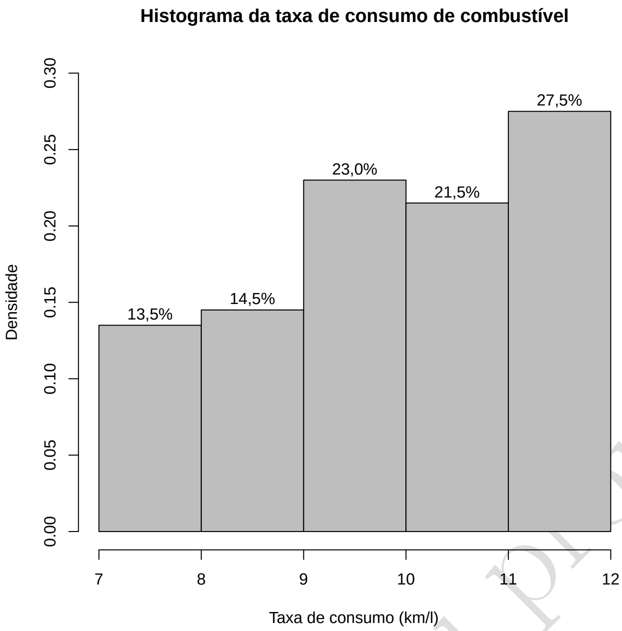
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(t_i-\overline{t})^2 = \frac{1}{20}\left[\left(0,21-\frac{2569}{2000}\right)^2+\left(2,71-\frac{2569}{2000}\right)^2+\left(2,12-\frac{2569}{2000}\right)^2+\right.$$
$$+\left(2,81-\frac{2569}{2000}\right)^2+\left(3,30-\frac{2569}{2000}\right)^2+\left(0,15-\frac{2569}{2000}\right)^2+\left(0,54-\frac{2569}{2000}\right)^2+$$
$$+\left(3,12-\frac{2569}{2000}\right)^2+\left(0,80-\frac{2569}{2000}\right)^2+\left(1,76-\frac{2569}{2000}\right)^2+\left(1,14-\frac{2569}{2000}\right)^2+$$
$$+\left(0,16-\frac{2569}{2000}\right)^2+\left(0,31-\frac{2569}{2000}\right)^2+\left(0,91-\frac{2569}{2000}\right)^2+\left(0,18-\frac{2569}{2000}\right)^2+$$
$$+\left(0,04-\frac{2569}{2000}\right)^2+\left(1,16-\frac{2569}{2000}\right)^2+\left(2,16-\frac{2569}{2000}\right)^2+\left(1,48-\frac{2569}{2000}\right)^2+$$
$$\left.+\left(0,63-\frac{2569}{2000}\right)^2\right] = \frac{4467579}{4000000},$$

que fornece um desvio padrão de $\sigma = \sqrt{\frac{4467579}{4000000}} \approx 1,0568$ horas.

16b) Tabela de frequência (o significado da coluna "h_i" é explicitado no exercício 2c):

Tempo (horas)	h_i	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
0,0 ⊢ 0,5	0,25	6	$6/20 = 0{,}30$	0,5	$3/5 = 0{,}60$
0,5 ⊢ 1,0	0,75	4	$4/20 = 0{,}20$	0,5	$2/5 = 0{,}40$
1,0 ⊢ 1,5	1,25	3	$3/20 = 0{,}15$	0,5	$3/10 = 0{,}30$
1,5 ⊢ 2,0	1,75	1	$1/20 = 0{,}05$	0,5	$1/10 = 0{,}10$
2,0 ⊢ 2,5	2,25	2	$2/20 = 0{,}10$	0,5	$1/5 = 0{,}20$
2,5 ⊢ 3,0	2,75	2	$2/20 = 0{,}10$	0,5	$1/5 = 0{,}20$
3,0 ⊢ 3,5	3,25	2	$2/20 = 0{,}10$	0,5	$1/5 = 0{,}20$
TOTAL	-	20	1	-	-

16c) Assumindo uma distribuição uniforme dos dados em cada intervalo de tempo, os dados podem ser analisados assumindo a média h_i da i-ésima faixa como a representante da mesma.



Estimativa da média $\overline{h}$ a partir dos dados da tabela de frequência:

$$\begin{aligned}\overline{h} &= \sum_i f_i h_i = 0,30 \cdot 0,25 + 0,20 \cdot 0,75 + \\ &+0,15 \cdot 1,25 + 0,05 \cdot 1,75 + \\ &+0,10 \cdot 2,25 + 0,10 \cdot 2,75 + \\ &+0,10 \cdot 3,25 = 1,325 \text{ (horas)}\end{aligned}$$

Estimativa da variância σ_h^2 a partir dos dados da tabela de frequência:

$$\begin{aligned}\sigma_h^2 &= \frac{1}{n} \sum_i n_i \left(h_i - \overline{h}\right)^2 = \frac{1}{20} \Big[\\ & 6\,(0,25 - 1,325)^2 + 4\,(0,75 - 1,325)^2 + \\ &+3\,(1,25 - 1,325)^2 + 1\,(1,75 - 1,325)^2 + \\ &+2\,(2,25 - 1,325)^2 + 2\,(2,75 - 1,325)^2 + \\ &+2\,(3,25 - 1,325)^2 \Big] = \frac{1731}{1600},\end{aligned}$$

donde segue que o desvio padrão é $\sigma_h = \sqrt{\frac{1731}{1600}} \approx 1,040$ (horas). A discrepância entre os valores obtidos nos exercícios (2a) e (2c) indicam os efeitos da hipótese adotada para a aritmética, que não foi completamente satisfatória.

17) Um consumidor está indeciso na compra de uma televisão e decide avaliar algumas informações estatísticas, fornecidas pelo fabricante, sobre a duração (em horas) do tubo de imagem. Discutir qual seria a escolha mais "conservadora" e a escolha mais "ousada".

Marca da TV	A	B	C
Média	8000	8100	8000
Desvio padrão	3000	3000	300

17) Deve-se notar, inicialmente, que as três médias são (visualmente) próximas[2], sendo descartadas como critério de diferenciação entre as três marcas de televisão. O aparelho da marca C, contudo, apresenta o menor desvio padrão, constituindo a escolha mais "conservadora". Por outro lado, as marcas A e B possuem as maiores dispersões no tempo de duração de seus tubos de imagem, sendo ambas as opções mais "ousadas" em relação à escolha da marca C; entretanto, ao apresentar a maior média, a televisão B seria a opção preferível dentre as escolhas menos conservadoras.

[2] A "proximidade" é entendida no sentido da diferença entre as médias ser inferior ao menor dos desvios padrão.

18) A distribuição de nota dos alunos de um curso de física é dada na tabela abaixo.

Nota	Frequência
0,0 ⊢ 2,0	04
2,0 ⊢ 4,0	02
4,0 ⊢ 6,0	36
6,0 ⊢ 8,0	28
8,0 ⊢ 10,0	01

a) Estimar a média $\overline{n}$ e o desvio padrão σ.
b) Estimar a fração dos alunos com as notas situadas no intervalo $\overline{n} \pm 2,0\sigma$.
c) Estimar a fração dos alunos com as notas situadas no intervalo $(\overline{n} - \sigma, \overline{n} + 1,5\sigma)$.

18a) Tabela de frequência:

Nota	x_i	Frequência Absoluta	Frequência Relativa	Amplitude	Densidade
0,0 ⊢ 2,0	1,0	04	$4/71 \approx 0{,}056$	2,0	$2/71 \approx 0{,}028$
2,0 ⊢ 4,0	3,0	02	$2/71 \approx 0{,}028$	2,0	$1/71 \approx 0{,}014$
4,0 ⊢ 6,0	5,0	36	$36/71 \approx 0{,}507$	2,0	$18/71 \approx 0{,}254$
6,0 ⊢ 8,0	7,0	28	$28/71 \approx 0{,}394$	2,0	$14/71 \approx 0{,}197$
8,0 ⊢ 10,0	9,0	01	$1/71 \approx 0{,}014$	2,0	$1/142 \approx 0{,}007$
TOTAL	-	71	1	-	-

Assumindo uma distribuição uniforme dos dados em cada intervalo da tabela acima, as análises estatísticas podem ser realizadas assumindo a média x_i do i-ésimo intervalo como o representante do mesmo.

Estimativa da média $\overline{n}$ para os $n = 71$ dados do conjunto $\{x_i\}$:

$$\overline{n} = \frac{1}{n}\sum_i n_i x_i = \frac{1}{71}\Big(4 \cdot 1,0 + 2 \cdot 3,0 + 36 \cdot 5,0 + 28 \cdot 7,0 + 1 \cdot 9,0\Big) = \frac{395}{71} \approx 5,56\,.$$

Estimativa da variância σ^2:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_i n_i\,(x_i - \overline{n})^2 = \frac{1}{71}\left[4\left(1,0 - \frac{395}{71}\right)^2 + 2\left(3,0 - \frac{395}{71}\right)^2 + 36\left(5,0 - \frac{395}{71}\right)^2 + \right.\\ &\left. +28\left(7,0 - \frac{395}{71}\right)^2 + 1\left(9,0 - \frac{395}{71}\right)^2\right] = \frac{18}{7}\,,\end{aligned}$$

que fornece um desvio padrão de $\sigma = \sqrt{\frac{18}{7}} \approx 1,60$.

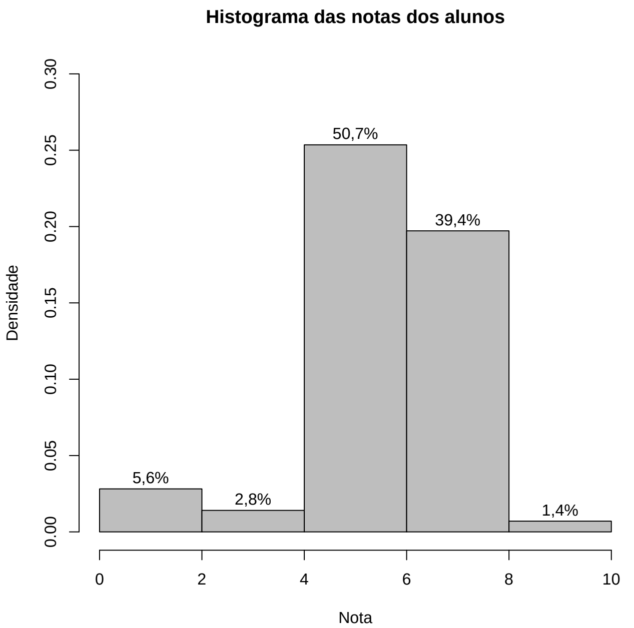
18b) Deve-se estimar a fração dos alunos com nota no intervalo $(\overline{n} - 2,0\sigma; \overline{n} + 2,0\sigma) \approx (2,36; 8,77)$, e os pontos $\overline{n} - 2,0\sigma$ e $\overline{n} + 2,0\sigma$ encontram-se, respectivamente, nos intervalos $2,0 \vdash 4,0$ e $8,0 \vdash 10,0$. Lembrando que a hipótese da distribuição uniforme dos dados em cada barra do histograma havia sido invocada, tem-se

$$\begin{aligned}d_{2,0\vdash 4,0} &= d_{(\overline{n}-2,0\sigma)\vdash 4,0}\\ \frac{1}{71} &= \frac{f_{(\overline{n}-2,0\sigma)\vdash 4,0}}{4,0 - (\overline{n} - 2,0\sigma)}\,,\end{aligned}$$

donde se tem $f_{(\overline{n}-2,0\sigma)\vdash 4,0} = \frac{1}{71}\left(2\sqrt{\frac{18}{7}} - \frac{111}{71}\right) \approx 0,0232 = 2,32\%$. Analogamente, tem-se

$$\begin{aligned}d_{8,0\vdash 10,0} &= d_{8,0\vdash(\overline{n}+2,0\sigma)}\\ \frac{1}{142} &= \frac{f_{8,0\vdash(\overline{n}+2,0\sigma)}}{(\overline{n} + 2,0\sigma) - 8,0}\,,\end{aligned}$$

donde se tem $f_{8,0\vdash(\overline{n}+2,0\sigma)} = \frac{1}{142}\left(2\sqrt{\frac{18}{7}} - \frac{173}{71}\right) \approx 0,0054 = 0,54\%$.



Desta forma, a fração dos alunos com nota no intervalo $\overline{n} \pm 2,0\sigma$ é

$$f_{(\overline{n}-2,0\sigma)\vdash 4,0} + f_{4,0\vdash 6,0} + f_{6,0\vdash 8,0} + f_{8,0\vdash(\overline{n}+2,0\sigma)} =$$
$$\frac{1}{71}\left(2\sqrt{\frac{18}{7}} - \frac{111}{71}\right) + \frac{36}{71} + \frac{28}{71} + \frac{1}{142}\left(2\sqrt{\frac{18}{7}} - \frac{173}{71}\right) = \frac{3}{71}\sqrt{\frac{18}{7}} + \frac{8693}{10082} \approx 0,9300\,,$$

ou cerca de $93,00\%$.

18c) Deve-se estimar a fração dos alunos com nota no intervalo $(\overline{n}-\sigma;\overline{n}+1,5\sigma) \approx (3,96;7,97)$, sendo que os pontos $\overline{n}-\sigma$ e $\overline{n}+1,5\sigma$ encontram-se, respectivamente, nos intervalos $2,0 \vdash 4,0$ e $6,0 \vdash 8,0$. Lembrando que a hipótese da distribuição uniforme dos dados em cada barra do histograma havia sido invocada, tem-se

$$\begin{aligned} d_{2,0\vdash 4,0} &= d_{(\overline{n}-\sigma)\vdash 4,0} \\ \frac{1}{71} &= \frac{f_{(\overline{n}-\sigma)\vdash 4,0}}{4,0-(\overline{n}-\sigma)}\,, \end{aligned}$$

donde se tem $f_{(\overline{n}-\sigma)\vdash 4,0} = \frac{1}{71}\sqrt{\frac{18}{7}} - \frac{111}{71^2} \approx 0,0006 = 0,06\%$. Analogamente, tem-se

$$\begin{aligned} d_{6,0\vdash 8,0} &= d_{6,0\vdash(\overline{n}+1,5\sigma)} \\ \frac{14}{71} &= \frac{f_{6,0\vdash(\overline{n}+1,5\sigma)}}{(\overline{n}+1,5\sigma)-6,0}\,, \end{aligned}$$

donde se tem $f_{6,0\vdash(\overline{n}+1,5\sigma)} = \frac{21}{71}\sqrt{\frac{18}{7}} - \frac{434}{71^2} \approx 0,3882 = 38,82\%$.

Desta forma, a fração dos alunos com nota no intervalo $(\overline{n}-\sigma,\overline{n}+1,5\sigma)$ é

$$f_{(\overline{n}-\sigma)\vdash 4,0} + f_{4,0\vdash 6,0} + f_{6,0\vdash(\overline{n}+1,5\sigma)} =$$
$$\left(\frac{1}{71}\sqrt{\frac{18}{7}} - \frac{111}{71^2}\right) + \frac{36}{71} + \left(\frac{21}{71}\sqrt{\frac{18}{7}} - \frac{434}{71^2}\right) = \frac{22}{71}\sqrt{\frac{18}{7}} + \frac{2011}{5041} \approx 0,8958\,,$$

ou cerca de $89,58\%$.

19) Uma amostra de vinte empresas, de porte médio, foi escolhida para um estudo sobre o nível educacional dos funcionários do setor de vendas. Os dados coletados, quanto ao número de empregados com curso superior completo, são apresentados abaixo.

Empresa	01	02	03	04	05	06	07	08	09	10
Funcionários	02	00	00	03	00	01	03	02	02	01

Empresa	11	12	13	14	15	16	17	18	19	20
Funcionários	01	01	00	00	04	03	01	01	02	02

a) Organizar uma tabela de frequência e calcular a média.
b) Determinar o desvio padrão.
c) As empresas pretendem incentivar o estudo de seus funcionários oferecendo um adicional de 3 salários mínimos para cada funcionário com curso superior. Calcular a despesa média adicional nessas empresas.

19a) Tabela de frequência:

Funcionários com curso superior completo	Frequência absoluta	Frequência relativa
0	5	5/20 = 0,25
1	6	6/20 = 0,30
2	5	5/20 = 0,25
3	3	3/20 = 0,15
4	1	1/20 = 0,05
TOTAL	20	1

Cálculo da média $\overline{x}$:

$$\overline{x} = \sum_i f_i x_i = 0,25 \cdot 0 + 0,30 \cdot 1 + 0,25 \cdot 2 + 0,15 \cdot 3 + 0,05 \cdot 4 = 1,45\,.$$

19b) Cálculo da variância σ^2 para os $n = 20$ dados:

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_i n_i\,(x_i - \overline{x})^2 \\
&= \frac{1}{20}\Big[5\,(0-1,45)^2 + 6\,(1-1,45)^2 + 5\,(2-1,45)^2 + 3\,(3-1,45)^2 + 1\,(4-1,45)^2\Big] \\
&= \frac{539}{400}\,,
\end{aligned}$$

o que implica num desvio padrão de $\sigma = \sqrt{\frac{539}{400}} \approx 1,1608$.

19c) Tabela de frequência com despesa adicional:

Funcionários com curso superior completo	Frequência absoluta	Frequência relativa	Despesa adicional (salários mínimos)
0	5	$5/20 = 0{,}25$	0
1	6	$6/20 = 0{,}30$	3
2	5	$5/20 = 0{,}25$	6
3	3	$3/20 = 0{,}15$	9
4	1	$1/20 = 0{,}05$	12
TOTAL	20	1	-

Despesa média adicional $\overline{d}$:

$$\overline{d} = \sum_i f_i \cdot (\text{salário})_i = 0,25 \cdot 0 + 0,30 \cdot 3 + 0,25 \cdot 6 + 0,15 \cdot 9 + 0,05 \cdot 12 = 4,35\ (\text{salários mínimos})\,.$$

Nota: Este exercício admite uma solução alternativa, onde o salário pago seria dado pelo produto (3 salários mínimos por funcionário) × (número médio de funcionários) = 3 × 1,45 = 4,35 salários mínimos. Deve-se apontar, contudo, que esta forma de resolução é válida devido à uniformidade na distribuição de salários por funcionários com curso superior completo.

20) Estudando-se o número de acertos em 100 lances-livres de bola ao cesto, uma amostra com 20 jogadores forneceu os seguintes resultados: 68, 73, 61, 66, 96, 79, 65, 86, 84, 79, 65, 78, 78, 62, 80, 67, 75, 88, 75 e 82. Agrupar as observações em intervalos de comprimento 5 a partir de 60 e, usando alguma suposição adicional, estimar a média e a variância do número de acertos em 100 arremessos.

20) Tabela de frequência:

Número de acertos	a_i	Frequência Absoluta	Frequência Relativa
$60 \vdash 65$	62,5	2	$2/20 = 0,10$
$65 \vdash 70$	67,5	5	$5/20 = 0,25$
$70 \vdash 75$	72,5	1	$1/20 = 0,05$
$75 \vdash 80$	77,5	6	$6/20 = 0,30$
$80 \vdash 85$	82,5	3	$3/20 = 0,15$
$85 \vdash 90$	87,5	2	$2/20 = 0,10$
$90 \vdash 95$	92,5	0	$0/20 = 0,00$
$95 \vdash 100$	97,5	1	$1/20 = 0,05$
TOTAL	-	20	1

As estimativas da média $\overline{a}$ e desvio padrão σ assumirão uma distribuição uniforme dos dados em cada intervalo, podendo ser possível, então, escolher o ponto médio de cada faixa como seu representante para a análise.
Estimativa da média dos $n = 20$ dados $\{a_i\}$:

$$\begin{aligned}
\overline{a} &= \sum_i f_i a_i = 0,10 \cdot 62,5 + 0,25 \cdot 67,5 + 0,05 \cdot 72,5 + \\
&\quad +0,30 \cdot 77,5 + 0,15 \cdot 82,5 + 0,10 \cdot 87,5 + \\
&\quad +0,00 \cdot 92,5 + 0,05 \cdot 97,5\Big) = 76,0 \text{ acertos}\,.
\end{aligned}$$

Estimativa da variância σ^2:

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_i n_i\,(a_i-\overline{a})^2 = \frac{1}{20}\Big[2\,(62,5-76,0)^2+5\,(67,5-76,0)^2+1\,(72,5-76,0)^2+ \\
&\quad +6\,(77,5-76,0)^2+3\,(82,5-76,0)^2+2\,(87,5-76,0)^2+0\,(92,5-76,0)^2+1\,(97,5-76,0)^2\Big] \\
&= \frac{321}{4}\,,
\end{aligned}$$

que leva a um desvio padrão de $\sigma=\sqrt{\frac{321}{4}}\approx 9,0$ acertos.

21) Estudando uma nova técnica de sutura, foram contados os dias necessários para a completa cicatrização de determinada cirrugia. Os resultados de 25 pacientes foram os seguintes: 6, 8, 9, 7, 8, 6, 6, 7, 8, 9, 10, 7, 8, 10, 9, 9, 9, 7, 6, 5, 7, 7, 8, 10 e 11. Organizar os dados numa tabela de frequência e calcular a média e a variância.

21) Tabela de frequência:

Tempo para cicatrização (dias)	Frequência Absoluta	Frequência Relativa
05	1	1/25 = 0,04
06	4	4/25 = 0,16
07	6	6/25 = 0,24
08	5	5/25 = 0,20
09	5	5/25 = 0,20
10	3	3/25 = 0,12
11	1	1/25 = 0,04
TOTAL	25	1

Média $\overline{t}$ dos $n=25$ dados $\{t_i\}$:

$$\begin{aligned}
\overline{t} &= \sum_i f_i t_i = 0,04\cdot 5+0,16\cdot 6+0,24\cdot 7+ \\
&\quad +0,20\cdot 8+0,20\cdot 9+0,12\cdot 10+0,04\cdot 11 \\
&= 7,88\ (\text{dias})\,.
\end{aligned}$$

variância σ^2:

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_i n_i\left(t_i-\overline{t}\right)^2 = \frac{1}{25}\Big[1\,(5-7,88)^2+4\,(6-7,88)^2+6\,(7-7,88)^2+5\,(8-7,88)^2+ \\
&\quad +5\,(9-7,88)^2+3\,(10-7,88)^2+1\,(11-7,88)^2\Big] = \frac{1416}{625}\,,
\end{aligned}$$

que leva a um desvio padrão de $\sigma=\sqrt{\frac{1416}{625}}\approx 1,51$ dias.

22) O departamento de atendimento ao consumidor de uma concessionária de veículos recebe, via telefone, as reclamações dos clientes. O número de chamadas dos últimos 30 dias foram anotados e os resultados foram: 3, 4, 5, 4, 4, 5, 6, 9, 4, 4, 5, 6, 4, 3, 6, 7, 4, 5, 4, 5, 7, 8, 8, 5, 7, 5, 4, 5, 7 e 6.
a) Construir uma tabela de frequência.
b) Calcular a média e o desvio padrão.
c) Admitindo que cada telefonema acarreta serviços sob a garantia avaliados em R$ 0,50 por chamada, calcular a média e o desvio padrão das despesas oriundas do atendimento ao consumidor.

22a) Tabela de frequência:

Número de chamadas	Frequência Absoluta	Frequência Relativa
3	2	2/30 ≈ 0,067
4	9	9/30 = 0,300
5	8	8/30 ≈ 0,267
6	4	4/30 ≈ 0,133
7	4	4/30 ≈ 0,133
8	2	2/30 ≈ 0,067
9	1	1/30 ≈ 0,033
TOTAL	30	1

22b) Cálculo da média $\overline{c}$ dos $n = 30$ dados $\{c_i\}$:

$$\overline{c} = \frac{1}{n}\sum_i n_i c_i = \frac{1}{30}\big(2\cdot 3 + 9\cdot 4 + 8\cdot 5 + 4\cdot 6 + 4\cdot 7 + 2\cdot 8 + 1\cdot 9\big) = \frac{53}{10} = 5,3 \text{ (chamadas/dia)}\,.$$

Cálculo da variância σ^2:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_i n_i\,(c_i - \overline{c})^2 = \frac{1}{30}\big[2\,(3-5,3)^2 + 9\,(4-5,3)^2 + 8\,(5-5,3)^2 + 4\,(6-5,3)^2 + \\ &\quad +4\,(7-5,3)^2 + 2\,(8-5,3)^2 + 1\,(9-5,3)^2\,\big] = \frac{683}{300}\,,\end{aligned}$$

que implica num desvio padrão de $\sigma = \sqrt{\frac{683}{300}} \approx 1,5$.

22c) Tabela de frequência das despesas provenientes do atendimento ao consumidor:

Número de chamadas	Despesas (R$)	Frequência Absoluta	Frequência Relativa
3	1,50	2	$2/30 \approx 0{,}067$
4	2,00	9	$9/30 = 0{,}300$
5	2,50	8	$8/30 \approx 0{,}267$
6	3,00	4	$4/30 \approx 0{,}133$
7	3,50	4	$4/30 \approx 0{,}133$
8	4,00	2	$2/30 \approx 0{,}067$
9	4,50	1	$1/30 \approx 0{,}033$
TOTAL	-	30	1

Cálculo da média das despesas $\overline{d}$ dos $n = 30$ dados $\{d_i\}$:

$$\begin{aligned}\overline{d} &= \frac{1}{n}\sum_i n_i d_i = \frac{1}{30}\big(2\cdot 1,50 + 9\cdot 2,00 + 8\cdot 2,50 + 4\cdot 3,00 + 4\cdot 3,50 + 2\cdot 4,00 + 1\cdot 4,50\big) \\ &= \frac{53}{20} = 2,65 \text{ (reais)}\,.\end{aligned}$$

Cálculo da variância σ_d^2:

$$\begin{aligned}\sigma_d^2 &= \frac{1}{n}\sum_i n_i\,\big(d_i - \overline{d}\big)^2 = \frac{1}{30}\big[2\,(1,50-2,65)^2 + 9\,(2,00-2,65)^2 + 8\,(2,50-2,65)^2 + \\ &\quad +4\,(3,00-2,65)^2 + 4\,(3,50-2,65)^2 + 2\,(4,00-2,65)^2 + 1\,(4,50-2,65)^2\,\big] = \frac{683}{1200}\,,\end{aligned}$$

que implica num desvio padrão de $\sigma_d = \sqrt{\frac{683}{1200}} \approx 0,75$ (reais).

Nota: Este exercício admite uma solução alternativa, onde a despesa seria dado pelo produto (Despesa por chamada de R$ 0,50) $\times$ (número médio de chamadas) = R$0,50 $\times$ 5,3 = R$ 2,65. Deve-se apontar, contudo, que esta forma de resolução é válida devido à uniformidade na distribuição de despesa por chamada. O mesmo comentário aplica-se no cálculo da variância (e, consequentemente, na determinação do desvio padrão), em que σ_d^2 = (R$ 0,50)^2 $\times$ σ^2 (e σ_d = R$ 0,50$\times\sigma$).

23) Considere um conjunto de n dados $\{x_i\}$ onde o desvio padrão coincide com a média $\overline{x}$.
a) Determinar a média se $x_i = a$ (a constante) para $i = 1, \cdots, n$. Determinar, ainda, o valor de a.
b) Determinar a média se metade dos n dados (assuma n par) for da forma $\overline{x} + b$ e, a outra metade, da forma $\overline{x} - b/2$ (b constante).

23a) se $x_i = a$ para $i = 1, \cdots, n$, então $\overline{x} = a$, donde se tem

$$\overline{x} = \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\big(\underbrace{x_i}_{=a} - \underbrace{\overline{x}}_{=a}\big)^2} = 0\,;$$

logo, $\overline{x} = a = 0$.

23b) Se metade dos n dados for da forma $\overline{x} + b$, e a outra metade, da forma $\overline{x} - b/2$, chega-se a

$$\overline{x} = \sigma = \left[\frac{1}{n} \underbrace{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}_{\frac{n}{2}\left[(\overline{x}+b)-\overline{x}\right]^2 + \frac{n}{2}\left[\left(\overline{x}-\frac{b}{2}\right)-\overline{x}\right]^2} \right]^{\frac{1}{2}} = \sqrt{\frac{5}{8}b^2} = \frac{\sqrt{10}}{4}|b|\,;$$

logo, $\overline{x} = \frac{\sqrt{10}}{4}|b|$.

Ademais, como

$$\overline{x} = \frac{1}{n}\left[\frac{n}{2}\left(\overline{x} + b\right) + \frac{n}{2}\left(\overline{x} - \frac{b}{2}\right)\right] = \overline{x} + \frac{b}{4}\,,$$

donde $b = 0$ (vide primeiro e terceiro membros), tem-se $\overline{x} = \frac{\sqrt{10}}{4}|b| = 0$.

24) Os batimentos cardíacos de dez pacientes foram medidos, chegando-se a uma média de 80 batidas por minuto. Após o cálculo desta média, os dados de um dos pacientes foram perdidos, restando as medidas dos outros nove, que são 75, 83, 77, 88, 82, 76, 79, 80 e 83 (em batidas por minuto). Determinar o desvio padrão dos dez dados.

24) Denotando por b o batimento cardíaco do paciente cujos dados foram perdidos, e por $\overline{x} = 80$ a média amostral, tem-se

$$\overline{x} = 80 = \frac{75 + 83 + 77 + 88 + 82 + 76 + 79 + 80 + 83 + b}{10}\,,$$

donde $b = 77$ (batidas por minuto). A variância σ^2 é dada, pois, por

$$\begin{aligned}\sigma^2 \quad &= \quad \frac{1}{10}\Big[(75-80)^2 + (83-80)^2 + (77-80)^2 + (88-80)^2 + (82-80)^2 + (76-80)^2 + \\ &\quad + (79-80)^2 + (80-80)^2 + (83-80)^2 + (77-80)^2\Big] = \frac{73}{5}\,,\end{aligned}$$

que implica um desvio padrão de $\sqrt{\frac{73}{5}} \approx 4$ (batimentos por minuto).

25) Em uma experiência em um laboratório didático, um aluno foi requisitado para medir a massa de um material três vezes com um instrumento precário. Após duas medidas ($1g$ e $2g$), o estudante (desonestamente) inventou a terceira medida de sorte que a variância fosse $1,00g^2$ para sua futura conveniência. Determinar o(s) possível(is) valor(es) para a terceira medida forjada.

25) Denotando por m a terceira medida (forjada), por $\overline{x}$ a média dos três dados (em g) e sua variância por σ^2, tem-se

$$\begin{cases} \overline{x} &= \dfrac{1 + 2 + m}{3} \\[2ex] \sigma^2 &= 1,00 = \dfrac{1}{3}\Big[(1-\overline{x})^2 + (2-\overline{x})^2 + (m-\overline{x})^2\Big] \end{cases}\,,$$

donde segue, imediatamente, que $2m^2 - 6m - 3 = 0$. Logo, tem-se $m = \frac{3-\sqrt{15}}{2}(g)$ ou $m = \frac{3+\sqrt{15}}{2}(g)$, implicando, respectivamente, cerca de $-0,44g$ e $3,44g$ como possíveis candidatos para a terceira medida (forjada). Como a primeira opção não é fisicamente viável, a medida forjada deve ser de $\frac{3+\sqrt{15}}{2} \approx 3,44g$.