Instituto Politécnico de Setúbal
Escola Superior de Tecnologia de Setúbal

DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º Semestre - 2022/2023
1.º Teste

Data: 22 de abril de 2023 — Duração: 2 horas

Resolução

O teste foi resolvido recorrendo ao software R: ver script_1Teste_ME_22_23.

1. Variável de interesse: X = preferência em relação ao sabor do café

   Níveis da variável de interesse: arábica, excelsa, liberia, robusta

   Amostra de dimensão $n = 1000$ consumidores

   (a) Tabela de frequências:

i	preferência em relação ao sabor do café x_i	Frequência Absoluta n_i	Frequência Relativa f_i
1	arábica	$0.405 \times 1000 = 405$	$\frac{40.5}{100} = 0.405$
2	excelsa	$0.180 \times 1000 = 180$	$\frac{64.8}{360} = 0.180$
3	liberia	$0.205 \times 1000 = 205$	$\frac{20.5}{100} = 0.205$
4	robusta	$1000 - (405 + 180 + 205) = 210$	$\frac{210}{1000} = 0.210$
		$n = 1000$	1

   Como os dados são qualitativos nominais, a única medida adequada é a Moda:

$$\text{moda} = \text{arábica}$$

   (pois $n_1 = 405$ é a maior frequência absoluta).

   (b) i. População: todos os países com plantações de café em 2019
   Dimensão da População: Não é indicado, no máximo serão todos os países do mundo

   Amostra: os países que produziram café em 2019 e que se encontram no ficheiro cafe.txt
   Dimensão da Amostra: $n = 55$ países

   Unidade estatística: países

   Variável estatística: bags
   Dados estatísticos: 51841, 917, 81265,...
   Classificação: Quantitativa discreta (*)

   Variável estatística: price
   Dados estatísticos: 120.52160, 24.00670,...
   Classificação: Quantitativa contínua

   Variável estatística: month
   Dados estatísticos: 4 (abril), 6 (junho), 10 (outubro)
   Classificação: Qualitativa ordinal

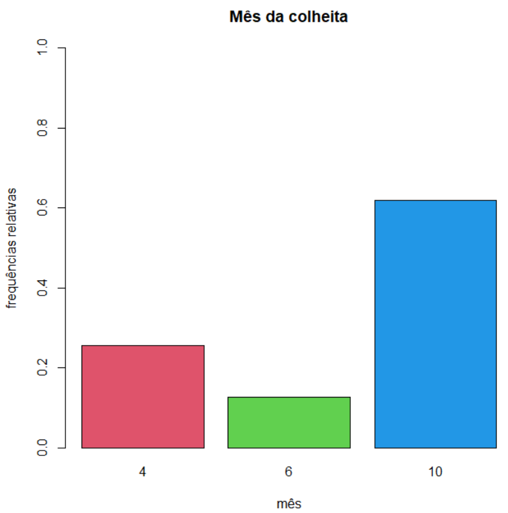
   (*) Como "bags" está representada em milhares, pode ser classificada como Quantitativa contínua.

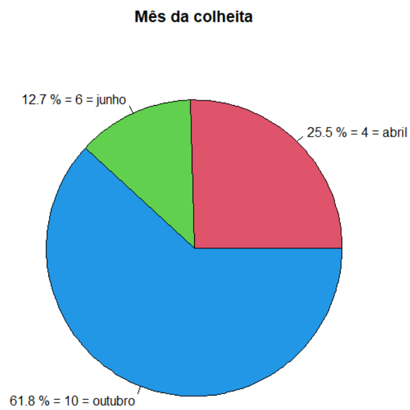
ii. O Brasil foi o país que produziu mais café em 2019 (58210712 sacos de café de 60 kg). A Bolívia foi o país que foi melhor pago em 2019 (158.2941 cêntimos de dólar por tonelada).

iii. Tabela de frequências:

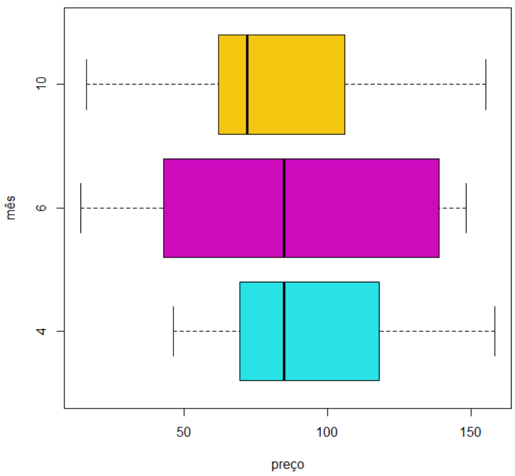
i	mês da colheita x_i	Freq. Absoluta n_i	Freq. Relativa f_i	Freq. Abs. Acumulada N_i	Freq. Rel. Acumulada F_i
1	4 = abril	14	0.255	14	0.255
2	6 = junho	7	0.127	21	0.382
3	10 = outubro	34	0.618	55	1
		$n = 55$	1		

Representação gráfica: gráfico de barras ou diagrama circular





iv. Nenhum dos meses apresenta valores considerados "outliers". O preço pago aos produtores apresenta maior dispersão quando a colheita é feita no mês de outubro (10), mas é no mês de junho (6) que a dispersão do preço é maior no que se refere aos 50% dos preços centrais (entre o 1º quartil e o 3º quartil). O preço máximo é atingido no mês de abril (4) e o mínimo no mês de junho (6), o preço mediano mais baixo ocorre no mês de outubro (10).



v. Como os dados têm unidades de medidas diferentes, a única medida de dispersão adequada é o coeficiente de variação:

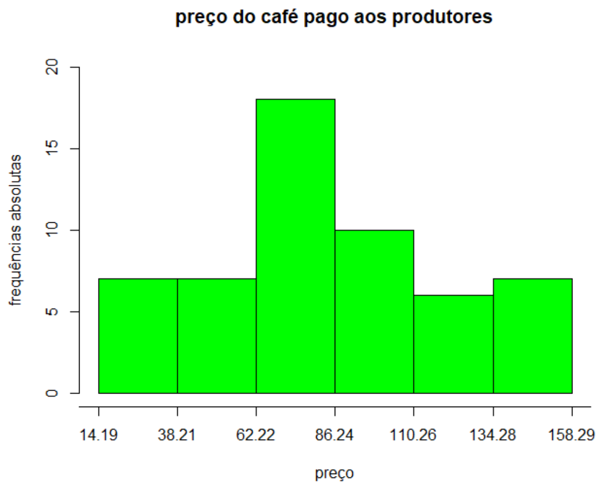
$$CV_{\text{sacos}} = \frac{8984.848}{3000.98} \times 100\% = 299.3971\% \qquad CV_{\text{preço}} = \frac{37.1775}{83.5274} \times 100\% = 44.5094\%$$

Como $CV_{\text{sacos}} > CV_{\text{preço}}$, o número de sacos de café de 60kg apresenta maior dispersão do que o preço pago aos produtores.

vi. Tabela de frequências:

i	Preço pago aos produtores Classe - c_i	Freq. Absoluta n_i	Freq. Relativa f_i	Freq. Abs. Acumulada N_i	Freq. Rel. Acumulada F_i
1	$[14.19, 38.21[$	7	0.127	7	0.127
2	$[38.21, 62.22[$	7	0.127	14	0.255
3	$[62.22, 86.24[$	18	0.327	32	0.582
4	$[86.24, 110.3[$	10	0.182	42	0.764
5	$[110.3, 134.3[$	6	0.109	48	0.873
6	$[134.3, 158.3]$	7	0.127	55	1
		$n = 55$	1		

Representação gráfica: histograma



2. $X =$ diâmetro, em milímetros, dos parafusos produzidos numa determinada linha de montagem

(a) Como X é uma variável aleatória contínua, a função de distribuição é $F(x) = \int_{-\infty}^{x} f(t)\,dt$.

$$\text{Se } x \leq 0, \quad F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0dt = 0$$

$$\text{Se } 0 < x < 1, \quad F(x) = \int_{-\infty}^{x} f(t)\,dt = 0 + \int_{0}^{x} tdt = \left[\frac{t^2}{2}\right]_0^x = \frac{x^2}{2}$$

$$\text{Se } 1 \leq x < 3, \quad F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1^2}{2} + \int_{1}^{x} \frac{t}{8}dt = \frac{1}{2} + \left[\frac{t^2}{16}\right]_1^x = \frac{1}{2} + \frac{x^2}{16} - \frac{1^2}{16} = \frac{7}{16} + \frac{x^2}{16}$$

$$\text{Se } x \geq 3, \quad F(x) = \int_{-\infty}^{x} f(t)\,dt = \frac{7}{16} + \frac{3^2}{16} + \int_{3}^{x} 0dt = 1 + 0 = 0$$

logo

$$F(x) = P\left(X \leq x\right) = \begin{cases} 0 & , x \leq 0 \\ \frac{x^2}{2} & , 0 < x < 1 \\ \frac{7}{16} + \frac{x^2}{16} & , 1 \leq x < 3 \\ 1 & , x \geq 3 \end{cases} .$$

(b)

$$P\left(X \leq \frac{5}{3} \,|\, \frac{1}{4} < X < \frac{8}{3}\right) = \frac{P\left(X \leq \frac{5}{3} \wedge \frac{1}{4} < X < \frac{8}{3}\right)}{P\left(\frac{1}{4} < X < \frac{8}{3}\right)} = \frac{P\left(\frac{1}{4} < X \leq \frac{5}{3}\right)}{P\left(\frac{1}{4} < X < \frac{8}{3}\right)} \underset{\text{v.a. cont.}}{=}$$

$$= \frac{F\left(\frac{5}{3}\right) - F\left(\frac{1}{4}\right)}{F\left(\frac{8}{3}\right) - F\left(\frac{1}{4}\right)} = \frac{\frac{7}{16} + \frac{\left(\frac{5}{3}\right)^2}{16} - \frac{\left(\frac{1}{4}\right)^2}{2}}{\frac{7}{16} + \frac{\left(\frac{8}{3}\right)^2}{16} - \frac{\left(\frac{1}{4}\right)^2}{2}} = 0.6816$$

<u>OU</u>

$$P\left(X \leq \frac{5}{3} \,|\, \frac{1}{4} < X < \frac{8}{3}\right) = \frac{P\left(X \leq \frac{5}{3} \wedge \frac{1}{4} < X < \frac{8}{3}\right)}{P\left(\frac{1}{4} < X < \frac{8}{3}\right)} = \frac{P\left(\frac{1}{4} < X \leq \frac{5}{3}\right)}{P\left(\frac{1}{4} < X < \frac{8}{3}\right)} \underset{\text{v.a. cont.}}{=}$$

$$= \frac{\int_{\frac{1}{4}}^{1} x dx + \int_{1}^{\frac{5}{3}} \frac{x}{8} dx}{\int_{\frac{1}{4}}^{1} x dx + \int_{1}^{\frac{8}{3}} \frac{x}{8} dx} = 0.6816$$

(c) Pretende-se determinar k tal que:

$$P\left(X > k\right) = 0.05 \Leftrightarrow 1 - P\left(X \leq k\right) = 0.05 \Leftrightarrow P\left(X \leq k\right) = 0.95 \Leftrightarrow F\left(k\right) = 0.95$$

Como

$$F(1) = \frac{1^2}{2} = 0.5 < 0.95 \text{ então } k > 1$$

$$F(3) = 1 > 0.95 \text{ então } k < 3$$

Como $1 < k < 3$ tem-se

$$F\left(k\right) = 0.95 \Leftrightarrow \frac{7}{16} + \frac{k^2}{16} = 0.95 \Leftrightarrow k^2 = 8.2 \Leftrightarrow k = \sqrt{8.2} \Leftrightarrow k = \pm 2.8636 \underset{1<k<3}{\Rightarrow} k = 2.8636$$

<u>OU</u>

$$P\left(X > k\right) = 0.05 \Leftrightarrow 1 - P\left(X \leq k\right) = 0.05 \Leftrightarrow P\left(X \leq k\right) = 0.95 \Leftrightarrow \int_{-\infty}^{k} f\left(x\right) dx = 0.95$$

portanto

$$\text{Se } k \leq 0, \quad \int_{-\infty}^{k} f(x)\,dx = 0.95 \Leftrightarrow \int_{-\infty}^{k} 0dx = 0.95 \Leftrightarrow 0 = 0.95 \text{ impossível}$$

$$\text{Se } 0 < k < 1, \quad \int_{-\infty}^{k} f(x)\,dx = 0.95 \Leftrightarrow \int_{-\infty}^{0} 0dx + \int_{0}^{k} xdx = 0.95 \Leftrightarrow 0 + \left[\frac{x^2}{2}\right]_0^k = 0.95 \Leftrightarrow \frac{k^2}{2} = 0.95 \Leftrightarrow$$

$$\Leftrightarrow k = \pm 1.3784 \text{ impossível pois } 0 < k < 1$$

$$\text{Se } 1 \leq k < 3, \quad \int_{-\infty}^{k} f(x)\,dx = 0.95 \Leftrightarrow \int_{-\infty}^{0} 0dx + \int_{0}^{1} xdx + \int_{1}^{k} \frac{x}{8}dx = 0.95 \Leftrightarrow$$

$$\Leftrightarrow 0 + \left[\frac{x^2}{2}\right]_0^1 + \left[\frac{x^2}{16}\right]_1^k = 0.95 \Leftrightarrow \frac{7}{16} + \frac{k^2}{16} = 0.95 \Leftrightarrow k = \pm 2.8636 \underset{1<k<3}{\Rightarrow} k = 2.8636$$

(d) Seja $Y =$ número de parafusos defeituosos por dia de produção, $Y \sim P(1.2)$ pois $E(Y) = \lambda = 1.2$ parafusos defeituosos por dia de produção.

i. Seja $W =$ número de parafusos defeituosos encontrados em 10 dias de produção,

$$W \sim P(10 \times 1.2) \Leftrightarrow W \sim P(12)$$

Pretende-se

$$P(W = 20) = f_W(20) = 0.0097$$

ii. Seja $T =$ tempo de produção (em horas) entre dois parafusos defeituosos.
Como 1 dia produção = 8 horas, então recorrendo à relação entre a distribuição de Poisson e a distribuição Exponencial, tem-se

$$T \sim Exp\left(\frac{8}{1.2}\right)$$

Pretende-se

$$P(T \geq 6) = 1 - P(T < 6) \underset{\text{v.a. cont.}}{=} 1 - F_T(6) = 0.4066$$

3. Seja $X =$ quantidade de areia vendida diariamente, em toneladas, $X \sim N(2, \sigma)$ pois $E[X] = \mu = 2$.

(a) Tem-se $X \sim N(2, \sigma) \Leftrightarrow Z = \frac{X-2}{\sigma} \sim N(0,1)$
1500 quilos = 1.5 toneladas
Pretende-se determinar σ sabendo

$$P(X > 1.5) = 0.95 \Leftrightarrow 1 - P(X \leq 1.5) = 0.95 \Leftrightarrow P(X \leq 1.5) = 0.05 \Leftrightarrow P\left(Z \leq \frac{1.5-2}{\sigma}\right) = 0.05 \Leftrightarrow$$

$$\Leftrightarrow \Phi\left(\frac{1.5-2}{\sigma}\right) = 0.05 \Leftrightarrow \frac{1.5-2}{\sigma} = z_{0.05} \Leftrightarrow \frac{1.5-2}{\sigma} = -1.645 \Leftrightarrow \sigma = 0.304$$

(b) Sabe-se que $\sigma^2 = 0.04$, então $\sigma = \sqrt{0.04} = 0.2$, logo $X \sim N(2, 0.2)$.

i.

$$P(X > 2.5 | X \geq 1) = \frac{P(X > 2.5 \wedge X \geq 1)}{P(X \geq 1)} = \frac{P(X > 2.5)}{P(X \geq 1)} = \frac{1 - P(X \leq 2.5)}{1 - P(X < 1)} \underset{\text{v.a. cont.}}{=}$$

$$= \frac{1 - F(2.5)}{1 - F(1)} = 0.0062$$

ii. Seja V = número de dias em 20, onde as vendas não atingem a média, $V \sim B\,(20, 0.5)$ pois

$$n = 20$$

$$p = P(\text{vendas não atingirem a média}) = P(X < 2) \underset{\text{v.a. cont.}}{=} F_X(2) = 0.5$$

Pretende-se

$$P\,(V \geq 5) = 1 - P(V < 5) \underset{\text{v.a. disc}}{=} 1 - P(V \leq 4) = 1 - F_V(4) = 1 - 0.0059 = 0.9941$$

(c) Pretende-se determinar $E[W]$. Como

$$E[W] = E[3Y^2 - 1] = 3E[Y^2] - 1$$

e sabe-se que Y é dado por:

y	0	2	3
$f(y)$	0.1	0.3	0.6

vem que

$$E[Y^2] = 0^2 \times 0.1 + 2^2 \times 0.3 + 3^2 \times 0.6 = 6.6$$

logo

$$E[W] = 3 \times 6.6 - 1 = 18.8$$