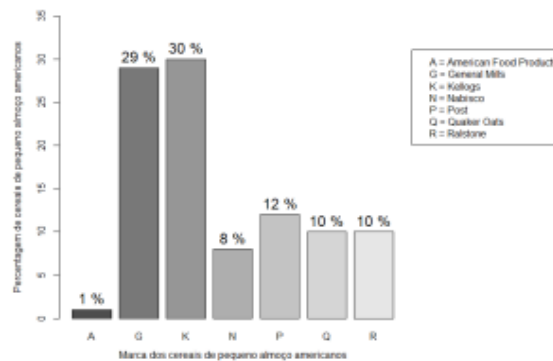


1. No gráfico seguinte são apresentados os dados referentes a todos os cereais de pequeno almoço americanos comercializados por marca:



A marca que mais cereais de pequeno almoço comercializa tem 60 tipos de cereais americanos.

Um grande hipermercado pretende estudar a preferência dos consumidores relativamente aos cereais de pequeno almoço americanos que comercializa, tendo em conta as componentes nutricionais de cada um e a localização nas prateleiras. No ficheiro `cereall.txt`, disponível no Moodle, são apresentados os dados recolhidos no estudo e no qual se encontram os seguintes campos:

- name = nome do cereal de pequeno almoço
- calories = calorias por dose do cereal (em quilocalorias)
- carbo = hidratos de carbono por dose do cereal (em gramas)
- shelf = prateleira do hipermercado onde se encontra o cereal
(1 = prateleira mais baixa, 2 = prateleira do meio, 3 = prateleira mais alta)
- rating = pontuação atribuída ao cereal

- [1.0] (a) Em relação ao estudo, identifique a População, a unidade estatística, as variáveis estatísticas e os dados estatísticos, classificando-os.
- [1.0] (b) Qual é a dimensão da População? Qual é a dimensão da Amostra?

1. (a) População: todos os cereais de pequeno almoço americanos

Unidade estatística: cereal de pequeno almoço americano

Variável estatística: calories

Dados estatísticos: 70, 120,...

Classificação: Quantitativa contínua (*)

Variável estatística: shelf

Dados estatísticos: 1, 2, 3 (códigos)

Classificação: Qualitativa ordinal

Variável estatística: carbo

Dados estatísticos: 5, 8,...

Classificação: Quantitativa contínua (*)

Variável estatística: rating

Dados estatísticos: 68.4, 34,...

Classificação: Quantitativa contínua

(*) foi recolhida como Quantitativa discreta.

- (b) Como o gráfico refere-se a todos os cereais de pequeno almoço americanos, então a informação é sobre a População. A marca que mais cereais de pequeno almoço comercializa tem 60 tipos de cereais americanos, logo como a

$$\text{Moda} = K$$

pois apresenta a maior percentagem (30%), então

$$\text{Dimensão da População} = \frac{60 \times 100}{30} = 200 \text{ cereais}$$

Como o ficheiro só tem os cereais de pequeno almoço americanos utilizados no estudo (comercializados pelo hipermercado), então a informação é sobre a amostra recolhida, logo

$$\text{Dimensão da Amostra} = n = 77 \text{ cereais}$$

- [1.5] (c) Construa a tabela de frequências completa relativamente à prateleira em que os cereais se encontram e calcule as medidas de localização e dispersão adequadas. Justifique a sua resposta.

(c) Tabela de frequências:

i	Prateleira x_i	Freq. Absoluta n_i	Freq. Relativa f_i	Freq. Absoluta Acum. N_i	Freq. Relativa Acum. F_i
1	mais baixa	20	0.25974	20	0.25974
2	do meio	21	0.27273	41	0.53247
3	mais alta	36	0.46753	77	1
		$n = 77$	1		

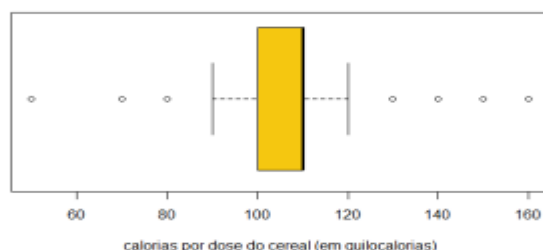
Como os dados são qualitativos, a única medida adequada é a Moda:

$$\text{moda} = \text{prateleira mais alta}$$

(pois $n_3 = 36$ é a maior frequência absoluta).

- [1.5] (d) Construa o diagrama de extremos e quartis das calorias por dose de cereal e avalie a existência de *outliers* moderados e severos. Comente os resultados.

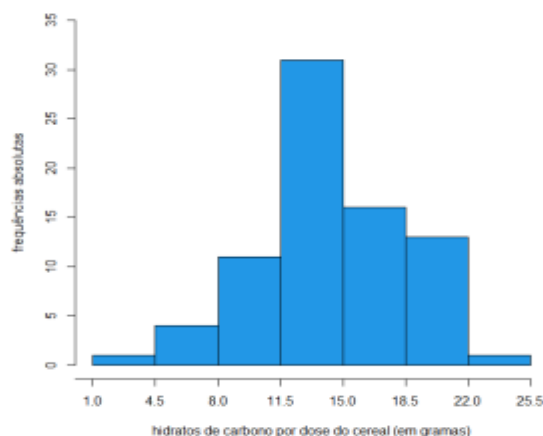
- (d) Diagrama de extremos e quartis das calorias por dose de cereal com indicação dos *outliers* (moderados e severos)



No total há 14 dados considerados *outliers*, 6 pelo limite inferior e 8 pelo limite superior. Dos 14 dados considerados *outliers*, 6 podem ser classificados como *outliers* severos (3 pelo limite inferior e 3 pelo limite superior) e 8 podem ser classificados como *outliers* moderados (3 pelo limite inferior e 5 pelo limite superior).

- [1.5] (e) Com base num histograma, classifique os dados referentes aos hidratos de carbono quanto à simetria. Na construção do histograma considere 7 classes, a primeira classe a começar no mínimo dos dados, amplitude por classe de 3.5 gramas e classes fechadas à direita e abertas à esquerda.

(e) Histograma



O histograma aparenta que os dados se distribuem de forma simétrica, com uma ligeira assimetria negativa (ou enviesada para a esquerda).

[1.5] (f) Compare quanto à dispersão os dados referentes às calorias e aos hidratos de carbono por dose de cereal. Qual apresenta menor dispersão? Justifique a sua resposta.

(f) Como os dados têm unidades de medidas diferentes, a única medida de dispersão adequada é o coeficiente de variação:

$$CV_{\text{calories}} = 18.2294\% \quad \text{e} \quad CV_{\text{carbo}} = 28.6722\%$$

Como $CV_{\text{calories}} < CV_{\text{carbo}}$, os dados das calorias por dose de cereal apresentam menor dispersão do que os dados dos hidratos de carbono por dose de cereal.

2. O responsável da loja I-TEL de uma grande superfície comercial está a analisar a estratégia de mercado que está a ser aplicada. Seja X a variável aleatória discreta, que representa a procura diária de um dado modelo de *smartphone* na loja I-TEL, com a seguinte função distribuição:

$$F(x) = \begin{cases} 0 & , x < 0 \\ \frac{1}{4} & , 0 \leq x < 1 \\ \frac{3}{8} & , 1 \leq x < 2 \\ \frac{7}{8} & , 2 \leq x < 3 \\ 1 & , x \geq 3 \end{cases}$$

[1.5] (a) Num dia em que já houve procura de pelo menos um daqueles *smartphones*, qual a probabilidade da procura ser inferior a 3?

2. Seja X a variável aleatória discreta, que representa a procura diária de um dado modelo de *smartphone* na loja I-TEL.

(a)

$$\begin{aligned} P(X < 3 | X \geq 1) &= \frac{P(X < 3 \wedge X \geq 1)}{P(X \geq 1)} = \frac{P(1 \leq X < 3)}{1 - P(X < 1)} = \frac{P(0 < X \leq 2)}{1 - P(X \leq 0)} \\ &= \frac{F(2) - F(0)}{1 - F(0)} = \frac{\frac{7}{8} - \frac{1}{4}}{1 - \frac{1}{4}} = \frac{5}{6} \end{aligned}$$

[2.0] (b) Determine a variância do número de *smartphones* daquele modelo que são procurados diariamente.

(b) Para determinar a variância é necessário calcular a função probabilidade. Temos que:

$$f(0) = F(0) = \frac{1}{4} = 0.25$$

$$f(1) = F(1) - F(0) = \frac{3}{8} - \frac{1}{4} = \frac{1}{8} = 0.125$$

$$f(2) = F(2) - F(1) = \frac{7}{8} - \frac{3}{8} = \frac{1}{2} = 0.5$$

$$f(3) = F(3) - F(2) = 1 - \frac{7}{8} = \frac{1}{8} = 0.125$$

Assim:

x	0	1	2	3
$f(x)$	0.25	0.125	0.5	0.125

Pretende-se calcular $V[X] = E[X^2] - E^2[X]$

onde

$$E[X] = 0 \times 0.25 + 1 \times 0.25 + 2 \times 0.5 + 3 \times 0.125 = 1.5$$

e

$$E[X^2] = 0^2 \times 0.25 + 1^2 \times 0.25 + 2^2 \times 0.5 + 3^2 \times 0.125 = 3.25$$

Logo

$$V[X] = 3.25 - (1.5)^2 = 1$$

- [1.5] (c) A remuneração mensal (em unidades monetárias) dos vendedores da I-TEL é uma variável aleatória Y com a seguinte função densidade de probabilidade:

$$f(y) = \begin{cases} \frac{y^2}{7000} & , \quad 10 < y \leq 20 \\ k & , \quad 20 < y \leq 30 \\ 0 & , \quad \text{caso contrário} \end{cases}.$$

Determine, justificando, o valor de k .

- (c) Seja Y = remuneração mensal (em unidades monetárias) dos vendedores da I-TEL, uma variável aleatória contínua. Como f é uma função densidade de probabilidade, então $f(y) \geq 0, \forall y$ e $\int_{-\infty}^{+\infty} f(y) dy = 1$.

Portanto

$$\begin{aligned} \int_{-\infty}^{+\infty} f(y) dy = 1 &\Leftrightarrow \int_{10}^{20} \frac{y^2}{7000} dy + \int_{20}^{30} k dy = 1 \Leftrightarrow \frac{1}{7000} \left[\frac{y^3}{3} \right]_{10}^{20} + k [y]_{20}^{30} = 1 \Leftrightarrow \\ &\Leftrightarrow \frac{1}{7000} \left(\frac{20^3}{3} - \frac{10^3}{3} \right) + k(30 - 20) = 1 \Leftrightarrow \frac{1}{3} + 10k = 1 \Leftrightarrow k = \frac{1}{15} \end{aligned}$$

e para $f(y) \geq 0, \forall y$ é necessário que $k \geq 0$, e tem-se $k = \frac{1}{15} \geq 0$.

Logo $k = \frac{1}{15}$ é o valor pretendido.

- (d) O tempo entre entradas consecutivas de clientes na loja I-TEL é uma variável aleatória Exponencial com valor esperado 15 minutos.

- [1.0] i. Sabe-se que a última entrada nas loja ocorreu há, pelo menos 10 minutos. Qual a probabilidade de se esperar, pelo menos mais 40 minutos pelo próximo cliente?

- [1.5] ii. Determine a probabilidade de, em 3 horas, entrarem pelo menos 10 clientes na loja.

- (d) Seja B – tempo entre entradas consecutivas de clientes na loja, em minutos, então, $B \sim \text{Exp}(15)$ porque $\theta = E[B] = 15$

i.

$$P(B \geq 40 + 10 | B \geq 10) \underset{(*)}{=} P(B \geq 40) = 1 - P(B < 40) \underset{\text{v.a. cont.}}{=} 0.0695$$

(*) falta de memória da distribuição exponencial

- ii. Considere a variável aleatória discreta C – número de clientes que entram na loja, em 3 horas. Pela relação entre a distribuição Exponencial e a distribuição de Poisson tem-se

$$C \sim P\left(\lambda = \frac{t}{\theta} = \frac{3 \times 60}{15} = 12\right)$$

$$P(C \geq 10) = 1 - P(C < 10) \underset{\text{v.a. discreta}}{=} 1 - P(C \leq 9) = 0.7576$$

- (e) O tempo de produção de um lote de *smartphones* é uma variável aleatória normal de média 50 dias e desvio padrão 5 dias.

- [1.0] i. A empresa tem de estabelecer um prazo de entrega para um lote cuja produção vai ser iniciada agora. Que prazo deve ser indicado ao cliente de forma a ser cumprido em 95% das vezes?
- [2.0] ii. Foram encomendados 10 lotes de *smartphones*. Qual a probabilidade de apenas 3 lotes demorarem mais de 60 dias em produção?

(e) Seja L — tempo de produção de um lote, em dias, onde

$$L \sim N(50, 5)$$

i. Pretende-se determinar k (prazo de entrega), tal que:

$$P(L \leq k) = 0.95 \Leftrightarrow F(k) = 0.95 \Leftrightarrow k = F^{-1}(0.95) \Leftrightarrow k = 58.224$$

Pelo que, para garantir que em 95% das vezes o prazo é cumprido, este terá de ser cerca de 58 dias.

ii. Seja V - número de lotes em 10, que demoram mais de 60 dias em produção.

Por outro lado,

$$p = P(\text{lote demorar mais de 60 dias}) = P(L > 60) = 1 - P(L \leq 60) = 0.0227 \text{ e } n = 10.$$

Assim $V \sim B(10, 0.0227)$ e

$$P(V = 3) = 0.0012$$

[1.5] (f) Seja W a variável aleatória que representa a distância (em quilómetros) que a empresa transportadora percorre, numa hora, para entrega das encomendas de *smartphones*. Considerando ainda que W tem distribuição uniforme definida no intervalo $[30; 40]$, calcule o desvio padrão da variável $T = -3W + 2$.

(f) Seja W = distância (em quilómetros) que a empresa transportadora percorre, numa hora, para entrega das encomendas de *smartphones*. Como $W \sim U_{(30,40)}$ então

$$V[W] = \frac{(40 - 30)^2}{12} = \frac{25}{3}$$

$$V[T] = V[-3W + 2] = (-3)^2 \times V[W] = 9 \times \frac{25}{3} = 75$$

e

$$\sigma_T = \sqrt{V[T]} = \sqrt{75} = 8.6603$$