35.

a) A variável em estudo, que representa o grau de satisfação de um aluno escolhido ao acaso de entre o universo de estudantes do curso em questão, é uma variável qualitativa ou categórica, uma vez que representa uma qualidade (neste caso, uma opinião) e não uma quantidade. Além disso, é possível estabelecer uma "ordem" entre as categorias no que diz respeito ao grau de satisfação, já que "NS" é um grau de satisfação menor do que "SP", que por sua vez é um grau de satisfação menor do que "S" e assim por diante. Utilizando a notação "<<" para representar um grau de satisfação menor, pode então dizer-se que: "NS" << "SP" << "S" << "B" << "MB". Portanto, existe uma ordem entre as categorias que definem a variável e diz-se, então, que é **qualitativa ordinal**.

b) Designe-se por:

n a dimensão da amostra – n.º de indivíduos na amostra (n = 80, no caso presente);

n_i a frequência absoluta simples da categoria *i* – n.º de indivíduos na amostra com grau de satisfação na categoria *i*;

N_i a frequência absoluta acumulada da categoria *i* – n.º de indivíduos na amostra com grau de satisfação na categoria *i* ou em categorias que representam graus de satisfação menores;

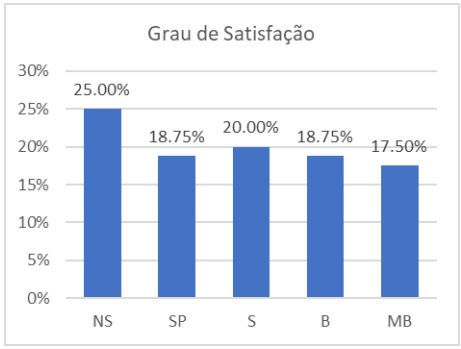
f_i a frequência relativa simples da categoria *i* – percentagem de indivíduos na amostra com grau de satisfação na categoria *i*: $fi = n_i/n$

F_i a frequência relativa acumulada da categoria *i* – percentagem de indivíduos na amostra com grau de satisfação na categoria *i* ou em categorias que representam graus de satisfação menores: $Fi = N_i/n$.

Pode então construir-se a seguinte tabela de frequências (que é exaustiva):

Grau Satisf.	n_i	N_i	f_i	F_i
NS	20	20	0.25	0.25
SP	15	35	0.1875	0.4375
S	16	51	0.2	0.6375
B	15	66	0.1875	0.825
MB	14	80	0.175	1

Uma representação adequada para dados qualitativos (observações de uma variável qualitativa) é um diagrama de barras, em que na linha horizontal se colocam as categorias, igualmente espaçadas, e a cada categoria se associa uma barra, não adjacente a qualquer outra, de altura igual a uma das frequências simples. No caso de dados que não são apenas qualitativos mas também ordinais, a colocação das categorias na linha horizontal deve respeitar a ordem entre elas.
O diagrama de barras que se apresenta em baixo é baseado nas frequências relativas simples da variável em estudo.

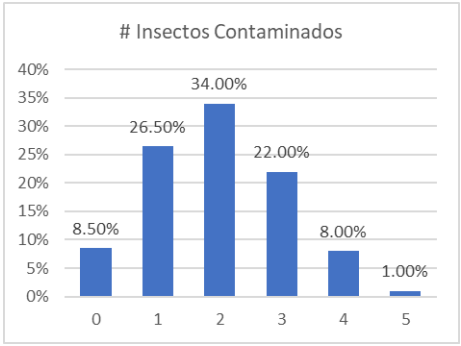


A única característica amostral que se pode calcular para qualquer conjunto de dados qualitativos é a Moda – categoria mais frequente na amostra. Neste caso, a **moda é "NS"**.
Para dados qualitativos ordinais, podem ainda ser calculados quantis, utilizando a definição:
o quantil de ordem *p*, com 0< p< 1, corresponde à "menor" categoria cuja frequência relativa acumulada é de pelo menos p. Por exemplo, para os dados em questão:
a **mediana** (p = 0.5) **é "S"**, pois F_{SP} = .4375< 0.5 e F_S = .6375> 0.5;
o **primeiro quartil** (p = 0.25) **é "NS"**, pois F_{SP} = .25 e
o **terceiro quartil** (p = 0.75) **é "B"**, pois F_S = .6375< 0.75 e F_B = .825> 0.75.

36. Seja X a variável em estudo (população); X é a v.a. que representa o n.º de insectos contaminados, numa amostra de 5 insectos seleccionados ao acaso. X só pode tomar um n.º finito de valores distintos: 0, 1, 2, 3, 4 ou 5. Deste modo, é uma variável discreta e os **dados** recolhidos, sendo observações de uma variável discreta, dizem-se **discretos**.
    a) Os dados recolhidos já são apresentados agrupados numa tabela com as frequências absolutas simples. Com base nestas é, ainda, possível construir a seguinte tabela de frequências exaustiva:

# Ins. Cont.	n_i	N_i	f_i	F_i
0	17	17	0.085	0.085
1	53	70	0.265	0.35
2	68	138	0.34	0.69
3	44	182	0.22	0.91
4	16	198	0.08	0.99
5	2	200	0.01	1

    A representação gráfica adequada para dados discretos é um gráfico de barras, representado num sistema de eixos coordenados, onde as abcissas são os valores possíveis para a v.a. em estudo e as ordenadas uma das correspondentes frequências simples (absoluta ou relativa). O gráfico de barras que se apresenta em baixo é baseado nas frequências relativas simples dos dados recolhidos.



    b) Moda: 2 (valor mais frequente na amostra)

    Média: $\bar{x} = \frac{1}{200}\sum_{i=1}^{200} x_i = (0\times17 + 1\times53 + \ldots + 5\times2)/200 = 1.975$

    Variância: $s^2 = \frac{1}{199}\left(\sum_{i=1}^{200} x_i^2 - 200\,\bar{x}^2\right) = (0^2\times17 + 1^2\times53 + \ldots + 5^2\times2 - 200\times1.975^2)/199 \approx 1.24$

    Desvio-padrão: $s = \sqrt{s^2} = \sqrt{1.24} \approx 1.1$

    Seja $x_{(i)}$ o elemento na posição *i* da amostra ordenada (por ordem crescente), *i* = 1, ..., n

    Mediana: n = 200 é par $\Rightarrow Q_{1/2} = (x_{(200/2)} + x_{(200/2+1)})/2 = (x_{(100)} + x_{(101)})/2 = (2 + 2)/2 = 2$

    Quantil de ordem 1/3: $n\times1/3 = 200/3 = 66.(6)$ é não inteiro $\Rightarrow Q_{1/3} = x_{(67)} = 1$

37. Seja X a variável em estudo (população); X – v.a. que representa o tempo (em segundos) entre duas reclamações consecutivas, escolhidas ao acaso, que chegam à central telefónica. X pode tomar um número infinito, não numerável (contável) de valores distintos, já que está associada a um fenómeno de carácter contínuo – tempo – e, formalmente, X pode tomar qualquer valor no intervalo $]0, +\infty[$. Assim, X é uma variável contínua, pelo que, os **dados** recolhidos são **contínuos**.
    a) A representação adequada de dados contínuos é o gráfico do **histograma**, onde os dados são distribuídos por intervalos disjuntos e adjacentes – "classes".

    É usual construir as classes por forma a que todas tenham a mesma amplitude h (a não ser que exista alguma justificação específica para que se proceda de modo distinto). Segundo este princípio, se os dados forem distribuídos por k classes, C_1, C_2, ..., C_k, então para garantir que as classes em conjunto contém todos os dados, basta que o limite inferior de C_1 seja menor ou igual do que o mínimo da amostra, $x_{(1)}$, e que o limite superior de C_k seja maior ou igual do que o máximo da amostra, $x_{(n)}$. Ou seja, pretende distribuir-se a amplitude amostral, $r = x_{(n)} - x_{(1)}$, por k classes com amplitude comum h; para tal, basta tomar h tão pequeno quanto possível e tal que

$$h \geq (x_{(n)} - x_{(1)})/k\ .$$

Resta decidir qual o n.º de classes, k, a considerar. Não existe um valor que seja o "melhor" para atribuir a k; mas uma regra comum e aceite como bastante boa numa série de situações distintas é devida a Sturges, segundo a qual se deve tomar

$$k = [\log_2 n]+1 ,$$

onde [x] representa a parte inteira de x. Excepto nos casos em que n é exactamente igual a uma potência de 2, a regra de Sturges pode ser enunciada do seguinte modo: tome-se k como o menor inteiro tal que

$$2^k \geq n .$$

Sempre que n é exactamente igual a uma potência de 2, condição anterior conduz a exactamente uma classe a menos do que a regra de Sturges. No entanto, a regra é meramente indicadora de um valor em torno do qual é adequado tomar o valor de k, pelo que, frequentemente se deve experimentar uma classe a mais ou uma classe a menos para se perceber qual a representação que parece descrever melhor a distribuição subjacente. Assim, é costume simplificar e utilizar a condição $2^k \geq n$ para a escolha inicial de k.

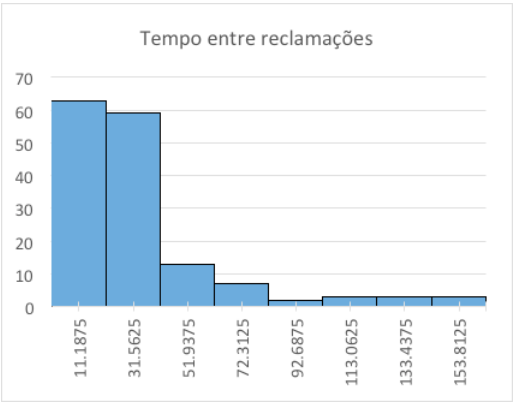
Vejamos como se faz, segundo o exposto, a classificação dos dados do problema presente.
n = 153: menor k tal que $2^k \geq 153 \Rightarrow \underline{k = 8}$ ($2^7 = 128$ (< 153) e $2^8 = 256$ (> 153))
$x_{(1)} = 1$, $x_{(153)} = 164 \Rightarrow r = x_{(153)} - x_{(1)} = 163$
$r/k = 163/8 = 20.375 \Rightarrow \underline{h = 20.375}$
As classes C_j vão ser consideradas da forma $C_j = [\, a_j, a_j+h[$, fechadas à esquerda e abertas à direita, j = 1, ..., k–1, tomando $a_1 = x_{(1)}$; a última classe será da forma $C_k = [\, a_k, a_k+h]$, fechada também à direita para incluir $x_{(n)}$ que coincide com a_k+h (o que é resultado de não se ter feito qualquer arredondamento a r/k). Deste modo, os dados vão ser distribuídos pelas classes:

$C_1 = [1, 21.375[$; $C_2 = [21.375, 41.75[$; ...; $C_7 = [123.25, 143.625[$ e $C_8 = [143.625, 164]$.

A cada classe C_j associa-se uma frequência absoluta n_j que não é mais do que o n.º de observações que pertencem a C_j, obtendo-se a seguinte tabela de frequências simples para os dados classificados:

Classe C_j	m_j	n_j
[1.000, 21.375[	11.1875	63
[21.375, 41.750[	31.5625	59
[41.750, 62.125[	51.9375	13
[62.125, 82.500[	72.3125	7
[82.500, 102.875[	92.6875	2
[102.875, 123.250[	113.0625	3
[123.250, 143.625[	133.4375	3
[143.625, 164.000]	153.8125	3
		153

Note-se que a tabela anterior inclui uma coluna onde estão os pontos médios das classes, m_j, que frequentemente são utilizados como os representantes das classes na representação gráfica do histograma.
Um histograma para os dados observados obtém-se associando a cada classe uma barra com altura igual à correspondente frequência absoluta e representando classes *versus* frequências num referencial, obtendo-se:



b) Média: $\bar{x} = \frac{1}{153}\sum_{i=1}^{153} x_i = 5065/153 \approx \underline{33.1}$

Variância: $s^2 = \frac{1}{152}\left(\sum_{i=1}^{153} x_i^2 - 153\bar{x}^2\right) = (326169 - 5065^2/153)/152 \approx \underline{1042.73}$

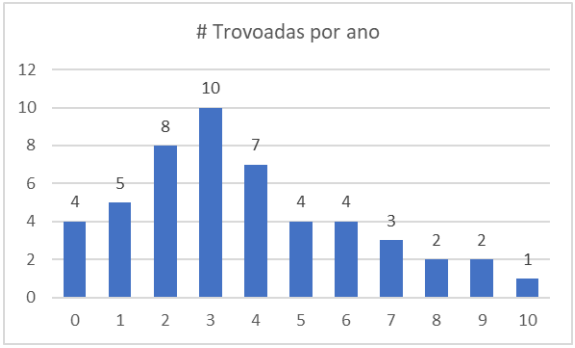
Mediana: n = 153 é ímpar $\Rightarrow Q_{1/2} = x_{((153+1)/2)} = x_{(77)} = \underline{24}$

1º Quartil: n/4 = 38.25 é não inteiro $\Rightarrow Q_{1/4} = x_{(39)} = \underline{14}$

3º Quartil: 3n/4 = 114.75 é não inteiro $\Rightarrow Q_{3/4} = x_{(115)} = \underline{38}$

38. Seja X a variável em estudo (população); X – v.a. que representa o n.º de trovoadas ocorridas, num ano escolhido ao acaso. X representa uma contagem e é, portanto, uma variável discreta. Deste modo, os dados recolhidos são discretos.

a) A representação adequada é um gráfico de barras. O gráfico de barras que se apresenta em baixo é baseado nas frequências absolutas simples dos dados recolhidos.



b) Média: $\bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = (0\times4 + 1\times5 + \ldots + 10\times1)/50 = \underline{3.76}$

Variância: $s^2 = \frac{1}{49}\left(\sum_{i=1}^{50} x_i^2 - 50\bar{x}^2\right) = (0^2\times4 + 1^2\times5 + \ldots + 10^2\times1 - 50\times3.76^2)/49 \approx \underline{6.39}$

Mediana: n = 50 é par $\Rightarrow Q_{1/2} = (x_{(25)} + x_{(26)})/2 = (3 + 3)/2 = \underline{3}$

1º Quartil: n/4 = 12.5 é não inteiro $\Rightarrow Q_{1/4} = x_{(13)} = \underline{2}$

3º Quartil: 3n/4 = 37.5 é não inteiro $\Rightarrow Q_{3/4} = x_{(38)} = \underline{5}$