

# **RESOLUÇÃO ASSISTIDA DE PROBLEMAS**

## **CAPÍTULO 2 – ESTATÍSTICA DESCRITIVA**

### **PROBLEMA 2.1**

#### *APOIOS À RESOLUÇÃO*

Apoio 1 (apenas o resultado)

Não aplicável. ■

Apoio 2 (sugestão)

Para efectuar comparações entre modelos podem ser utilizados todos os tipos de escala, excepto a nominal. ■

Apoio 3 (resolução completa)

(i) Consumo de combustível

escala ordinal: consumo elevado, médio ou baixo

escala absoluta: consumo expresso em litros aos 100 km. ■

(ii) Conforto

escala ordinal: ruído elevado, médio ou baixo

escala absoluta: ruído expresso em decibéis. ■

(iii) Segurança na travagem

escala ordinal: segurança reduzida, considerável ou elevada

escala absoluta: distância requerida para parar o veículo a 100 km/h.

Em princípio, os dados expressos numa escala absoluta fornecem mais informação do que aquela que é contida em dados expressos noutra escala (de facto, os dados expressos numa escala absoluta podem ser expressos numa escala de intervalo ou numa escala nominal). No entanto, é necessário que quem utiliza os dados saiba interpretar a informação que encerram. Por exemplo, o ruído dentro da cabine, se expresso em decibéis, pode ser convertido em ruído elevado, médio ou baixo. No entanto o cidadão comum percebe mais facilmente o significado de “ruído elevado dentro do habitáculo de um automóvel” do que “ruído de 15 decibéis”. ■

### **PROBLEMA 2.2**

#### *APOIOS À RESOLUÇÃO*

Apoio 1 (apenas o resultado)

Não aplicável. ■

### Apoio 2 (sugestão)

A forma mais comum de descrever amostras univariadas com dados expressos numa escala nominal é recorrer a tabelas de frequências, a diagrama de barras ou a diagrama circulares.

■

### Apoio 3 (resolução completa)

Os dados apresentados na tabela do enunciado do problema estão expressos numa escala nominal. A forma mais comum de descrever amostras univariadas com este tipo de dados é recorrer a tabelas de frequências, a diagrama de barras ou a diagrama circulares.

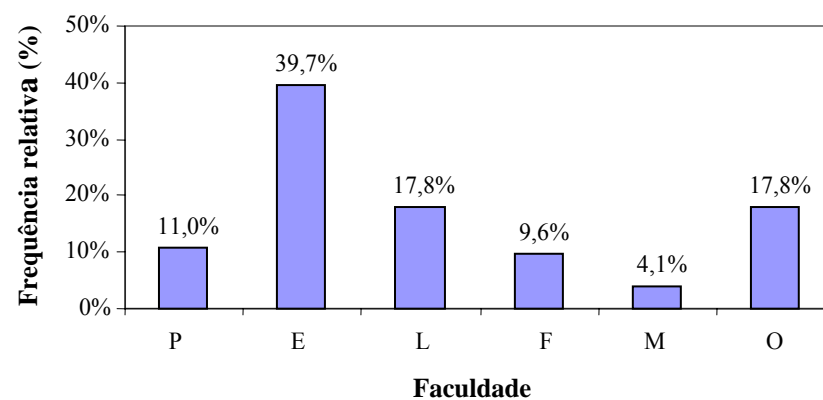
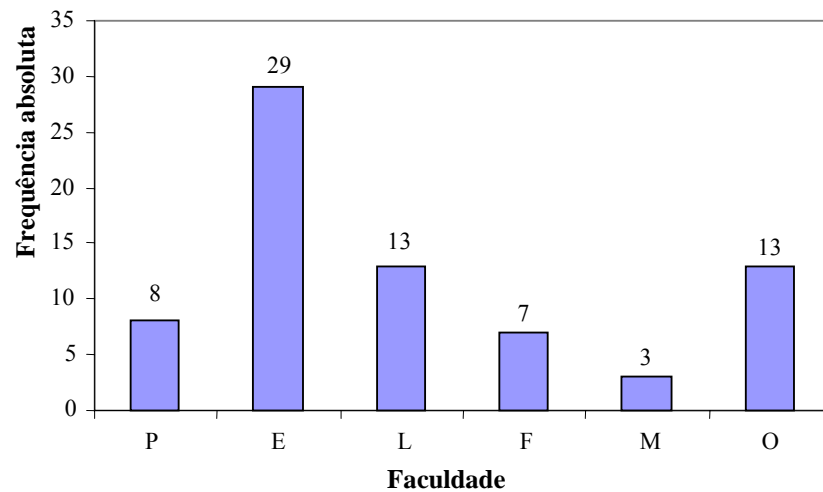
Apresenta-se seguidamente uma tabela de frequências, um diagrama de barras e um diagrama circular, que permitem pôr em evidência a importância relativa das diferentes categorias de alunos.

#### (i) Tabela de frequências

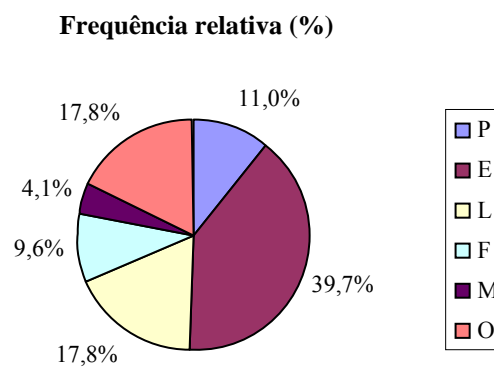
<i>Faculdade (k)</i>	<i>Frequência absoluta (<math>N_k</math>)</i>	<i>Frequência relativa (<math>f_k</math>)</i>
Psicologia (P)	8	0.110
Engenharia (E)	29	0.390
Letras (L)	13	0.178
Farmácia (F)	7	0.096
Medicina (M)	3	0.041
Outras (O)	13	0.178
Total	$N = 73$	1.000

Note-se que  $N = \sum_{k=1}^K N_k$  e que  $f_k = \frac{N_k}{N}$ .

#### (ii) Diagramas de barras



(iii) Diagrama circular



Quer a tabela de frequências quer estes diagramas podem ser facilmente obtidos a partir do programa “Microsoft Excel”. Em primeiro lugar, devem preencher-se as células da folha de cálculo como se indica na figura:

	A	B	C	D	E	F	G	H	I	J	K
1	E	L	F	E	L	E	E	P	F	L	
2	M	E	E	O	E	P	O	E	F	O	
3	O	L	O	E	L	F	L	E	L	E	
4	O	E	O	L	P	E	E	L	E	M	
5	E	O	E	O	P	P	E	F	E	O	
6	P	E	L	E	F	O	L	F	P	E	
7	E	O	E	E	E	M	E	L	E	P	
8	O	E	L								

Seguidamente, constitui-se a lista das faculdades (note-se que a lista é nominal) preenchendo-se as células necessárias (ver figura seguinte). Para calcular as frequências absolutas utiliza-se a função “COUNTIF” aplicada ao conjunto de células que contém os dados, para cada uma das abreviaturas associadas às faculdades - “P”, “E”, ..., “O” - (ver figura seguinte).

	L	M	N	O
			=COUNTIF(\$A\$1:\$J\$8, "P")	
				=M10/\$M\$16
9	Faculdade	Frequência absoluta	Frequência relativa (%)	
10	Psicologia (P)	8	11,0%	
11	Engenharia (E)	29	39,7%	
12	Letras (L)	13	17,8%	
13	Farmácia (F)	7	9,6%	
14	Medicina (M)	3	4,1%	
15	Outras (O)	13	17,8%	
16	Total	73	100,0%	
17				

## **PROBLEMA 2.3**

### *APOIOS À RESOLUÇÃO*

Apoio 1 (apenas o resultado)

Não aplicável. ■

Apoio 2 (sugestão)

Construa um diagrama de *box-and-whisker*, medidas amostrais (ou estatísticas), tabela de frequências, histograma e polígono de frequências acumuladas. Note que para os três últimos é necessário agrupar os dados em classes (considere classes de igual amplitude). ■

Apoio 3 (resolução completa)

Apresentam-se seguidamente as medidas amostrais (ou estatísticas).

Dimensão da amostra	64
Média	1787.88
Mediana	1021
Moda	444*
Desvio padrão ( $s$ )	1914.95
Variância ( $s^2$ )	3.66705 E6
Kurtose ( $g_2$ )	2.12
Assimetria ( $g_1$ )	1.74
Amplitude	7398
Mínimo	97
Máximo	7495
1º Quartil	536
3º Quartil	2258.75
Amplitude interquartil	1722.75

\* Existem múltiplas modas. Apresenta-se o valor mínimo.

Recorda-se que

$$s = \sqrt{\frac{1}{N-1} \cdot \sum_{n=1}^N (x_n - \bar{x})^2}.$$

Por outro lado, seja  $m_i = \frac{1}{N} \cdot \sum_{n=1}^N (x_n - \bar{x})^i$  o momento centrado de ordem  $i$ . A variância amostral pode ser calculada por

$$s^2 = \frac{N}{N-1} \cdot m_2.$$

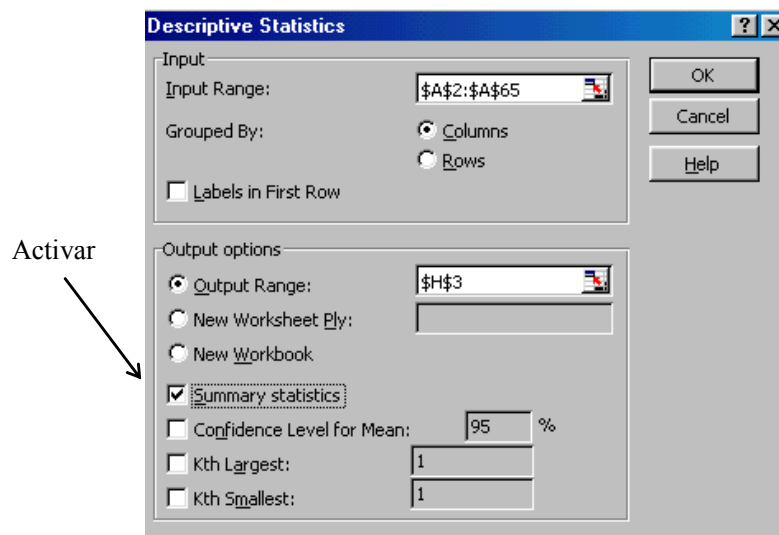
Adicionalmente,

$$g_1 = \frac{1}{s^3} \cdot \frac{N^2}{(N-1) \cdot (N-2)} \cdot m_3$$

e

$$g_2 = \frac{1}{s^4} \cdot \left[ \frac{N^2 \cdot (N+1)}{(N-1) \cdot (N-2) \cdot (N-3)} \cdot m_4 - \frac{3 \cdot (N-1)^2}{(N-2) \cdot (N-3)} \cdot s^4 \right].$$

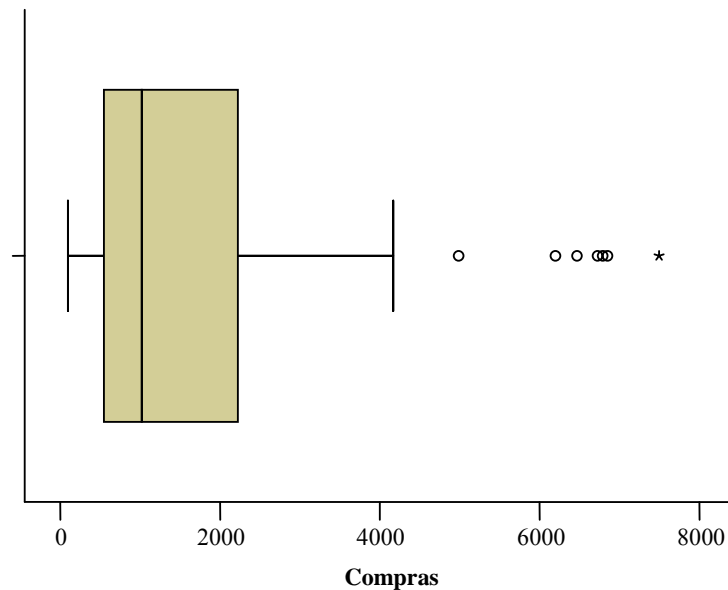
As medidas amostrais podem ser facilmente obtidas a partir do programa “Microsoft Excel”. Em primeiro lugar, introduzem-se os dados na folha de cálculo. Em seguida, acciona-se o menu “Tools” e, dentro das opções disponíveis, escolhe-se “Data Analysis” (se esta opção não estiver visível, selecciona-se “Add-Ins”, seguidamente “Analysis ToolPack” e prime-se o botão “OK”). Dentro da opção “Data Analysis” escolhe-se “Descriptive Statics”. Deverá surgir a janela seguinte, na qual se deverá activar “Summary statistics” (indicando em “Input Range” as células que contêm os dados e em “Output Range” as células que deverão ser ocupadas pelos resultados).



Na sequência desta operação surgirá a seguinte listagem:

<i>Column1</i>	
Mean	1787,875
Standard Error	239,3693014
Median	1021
Mode	444
Standard Deviation	1914,954411
Sample Variance	3667050,397
Kurtosis	2,121334946
Skewness	1,73715751
Range	7398
Minimum	97
Maximum	7495
Sum	114424
Count	64

No diagrama do tipo caixa (*box-and-whisker plot*) representam-se graficamente a mediana, o primeiro quartil, o terceiro quartil e os valores máximo e mínimo dos dados.



(i) Tabela de frequências

Para construir a tabela de frequências (e também para o histograma e polígono de frequências) deve, sempre que possível, definir-se células de igual amplitude e em número aproximadamente igual à raiz quadrada do número de dados. Neste caso, as células foram definidas da seguinte forma:

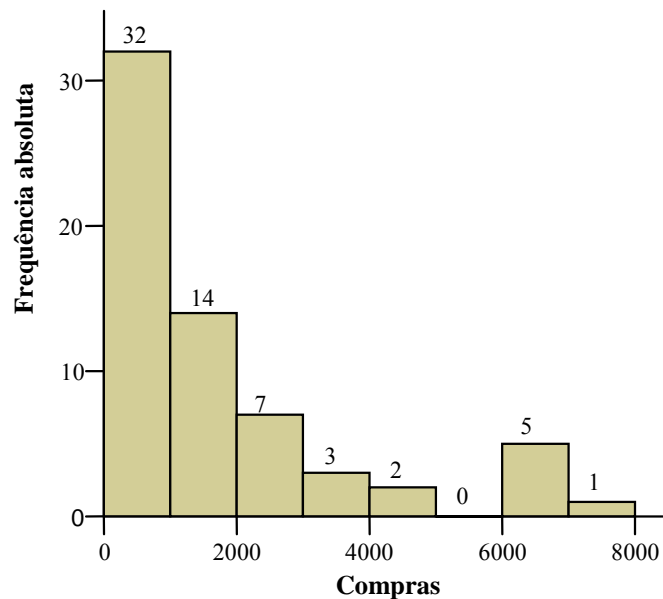
1. amplitude da amostra:  $A = (x_n)_{\max} - (x_n)_{\min} = 7496 - 97 = 7398$
2. amplitude das células: para amplitudes de 1000, 2000 e 3000 obter-se-iam os valores indicativos  $\frac{A}{1000} = 7.40$ ,  $\frac{A}{2000} = 3.70$  e  $\frac{A}{5000} = 1.48$
3. número de células ( $K$ ): para células com igual amplitude recorre-se, tipicamente, à seguinte regra:

$$K \approx \sqrt{N.^\circ \text{ de dados}(N)}.$$

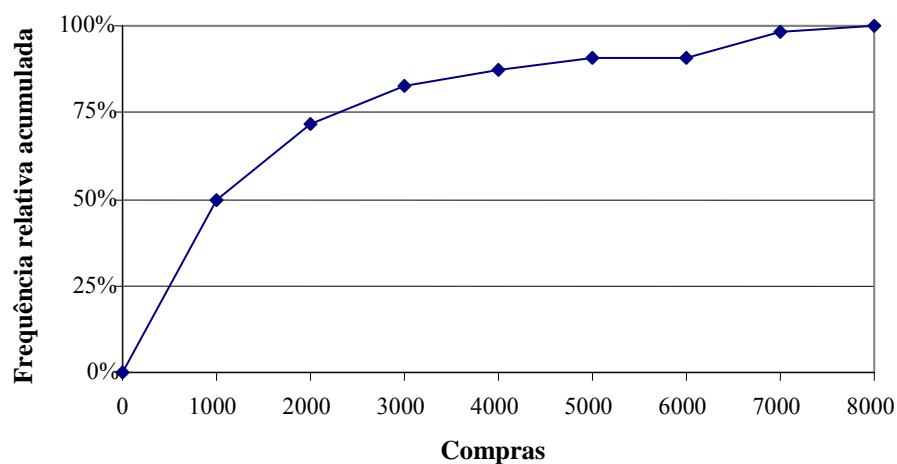
Finalmente, tudo ponderado, decidiu-se adoptar o número de células  $K = \sqrt{64} = 8$ , com amplitude de 1000 (iniciando a tabela em 0 e terminando em 8000, garantindo a cobertura de todos os valores observados). Note-se que há vantagens óbvias em associar às células amplitudes expressas por números “redondos”.

Compras [euros]	Frequência absoluta	Frequência absoluta acumulada	Frequência relativa	Frequência relativa acumulada
]0; 1000]	32	32	0.5000	0.5000
]1000; 2000]	14	46	0.2188	0.7188
]2000; 3000]	7	53	0.1094	0.8281
]3000; 4000]	3	56	0.0469	0.8750
]4000; 5000]	2	58	0.0313	0.9063
]5000; 6000]	0	58	0.0000	0.9063
]6000; 7000]	5	63	0.0781	0.9844
]7000; 8000]	1	64	0.0156	1.0000

(ii) Histograma

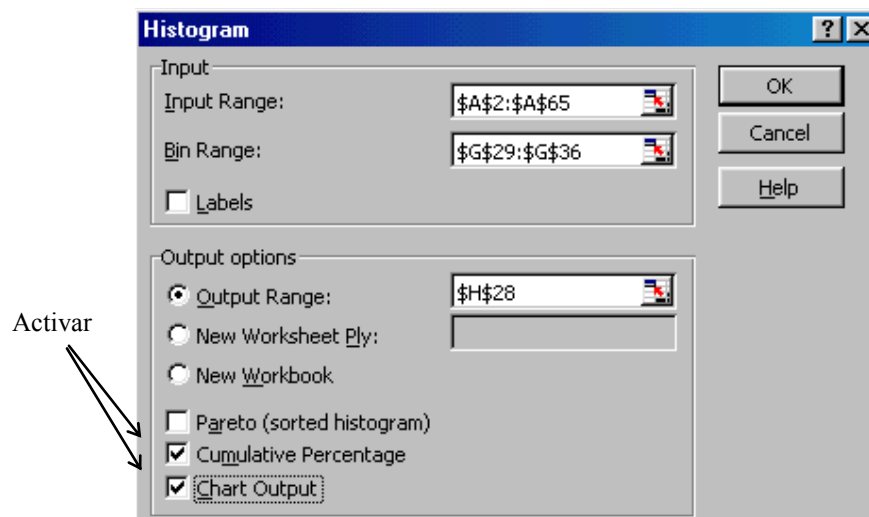


(iii) Polígono de frequências acumuladas



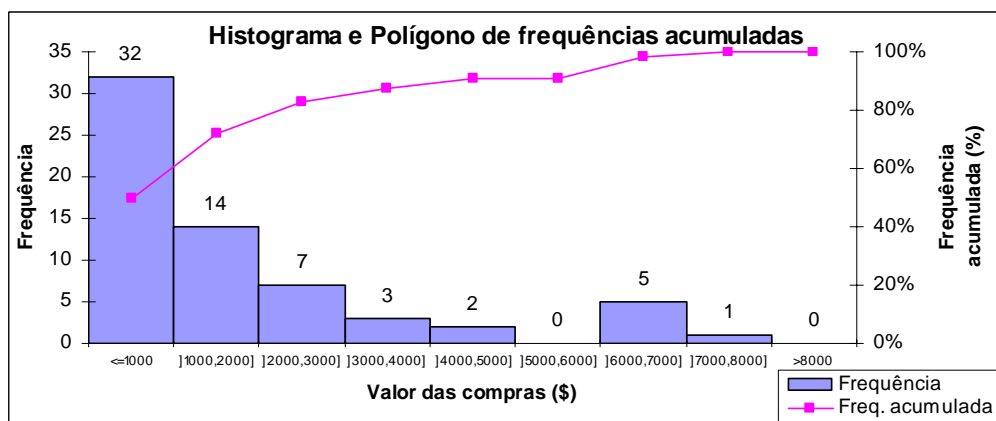
O histograma de frequências e o respectivo polígono de frequências acumuladas, podem igualmente ser obtidos com auxílio do programa “Microsoft Excel”. Em primeiro lugar, introduzem-se os dados na folha de cálculo e cria-se uma coluna com os valores referentes aos limites superiores de cada classe. Em seguida, acciona-se o menu “Tools” e, dentro das opções disponíveis, escolhe-se “Data Analysis” (se esta opção não estiver visível, selecciona-se sucessivamente “Add-Ins”, “Analysis ToolPack” e prime-se “OK”). Dentro da opção “Data Analysis” escolhe-se “Histogram”. Deverá surgir a janela que se apresenta na figura seguinte, na qual se selecciona “Summary statistics” (indicando em “Input Range” as células que contêm os dados, em “Bin Range” as células em que se inscreveram os limites superiores de cada classe e em “Output Range” as células que serão ocupadas pelos resultados). Para além da opção “Chart Output”, que se utiliza para construir o histograma, se se pretender sobrepor-lhe o polígono de frequências acumuladas deverá activar-se a opção “Cumulative Percentage” (ver figura seguinte).





Na listagem e a figura seguintes ilustra-se o resultado destas operações.

<i>Bin</i>	<i>Frequency</i>	<i>Cumulative %</i>
1000	32	50,00%
2000	14	71,88%
3000	7	82,81%
4000	3	87,50%
5000	2	90,63%
6000	0	90,63%
7000	5	98,44%
8000	1	100,00%
More	0	100,00%



## **PROBLEMA 2.4**

### *APOIOS À RESOLUÇÃO*

Apoio 1 (apenas o resultado)

Não aplicável. ■

### Apoio 2 (sugestão)

Procure representar os dados numa tabela de frequências, num histograma e num polígono de frequências acumuladas. ■

### Apoio 3 (resolução completa)

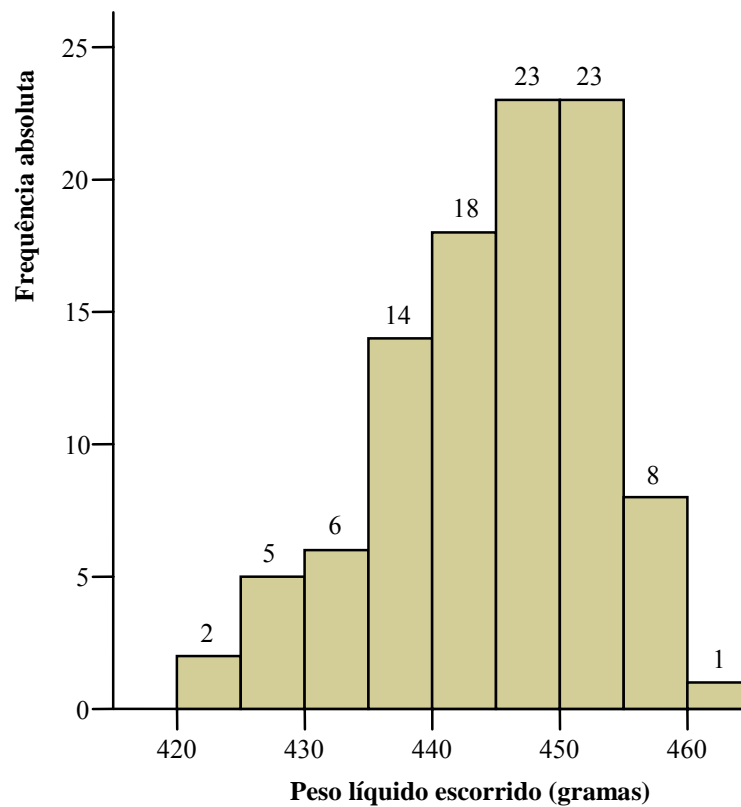
Apresentam-se seguidamente uma tabela de frequências, um histograma e um polígono de frequências acumuladas. As células foram definidas de tal modo que ficasse clara a distinção entre valores do peso líquido escorrido inferiores a 450 gramas (células 1 a 6) e valores iguais ou superiores a 450 gramas (células 7 a 9).

Registe-se que, na amostra, 68% das latas têm um peso líquido escorrido inferior ao indicado como médio (450 g). A média amostral de 444.3 gramas (ver resolução do Problema 2.5) é, também, inferior àquele valor.

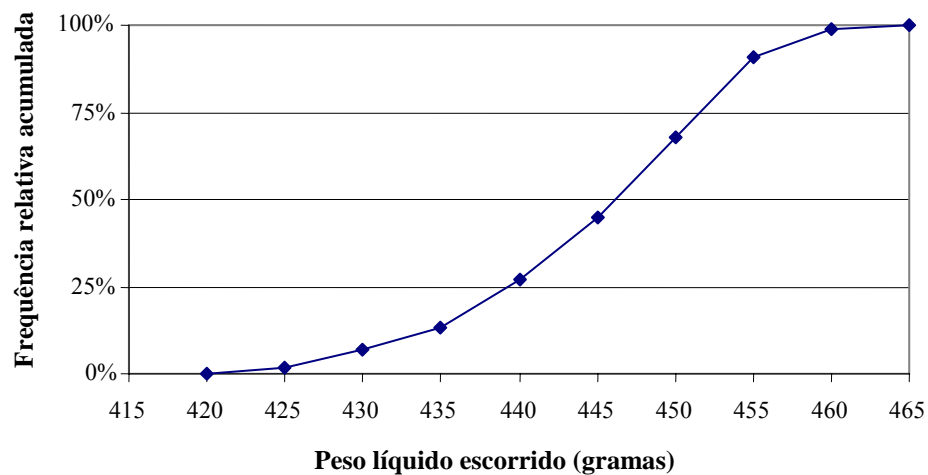
#### (i) Tabela de frequências

Classe	Frequência absoluta	Frequência absoluta acumulada	Frequência relativa	Frequência relativa acumulada
[420,425[	2	2	0.02	0.02
[425,430[	5	7	0.05	0.07
[430,435[	6	13	0.06	0.13
[435,440[	14	27	0.14	0.27
[440,445[	18	45	0.18	0.45
[445,450[	23	68	0.23	0.68
[450,455[	23	91	0.23	0.91
[455,460[	8	99	0.08	0.99
[460,465[	1	100	0.01	1.00

(ii) Histograma



(iii) Polígono de frequências relativas acumuladas



Para obter a tabela de frequências, o histograma de frequências e o respectivo polígono de frequências acumuladas com auxílio do programa “Microsoft Excel” proceder-se-ia de modo idêntico ao que se descreve no Problema 2.3. As estatísticas amostrais são apresentadas no Problema 2.5. ■

## **PROBLEMA 2.5**

### *APOIOS À RESOLUÇÃO*

#### Apoio 1 (apenas o resultado)

Não aplicável. ■

#### Apoio 2 (sugestão)

Para calcular as estatísticas a partir dos dados agrupados recorra ao ponto central de cada célula  $M_k$ . Admita que os dados incluídos em cada célula tomam o valor correspondente a  $M_k$ . ■

#### Apoio 3 (resolução completa)

Na tabela seguinte apresenta-se uma comparação entre as estatísticas calculadas com base nos dados do problema anterior (variável PLE – peso líquido escorrido) e as estatísticas calculadas com base nos dados agrupados (variável PLEGRUP). Notar que, neste caso, se considerou que os dados se concentravam no centro de cada uma das células (o que corresponde aos valores 422, 427, 432, 437, 442, 447, 452, 457 e 462). Registe-se que se verificam diferenças em todas as estatísticas (embora de pequena monta). Naturalmente que a aproximação será tanto melhor quanto menor for a amplitude das células.

Variável	PLE	PLEGRUP
Dimensão da amostra	100	100
Média	444.29	444.2
Mediana	445	447
Moda	450	447
Desvio padrão ( $s$ )	8.787	8.508
Variância ( $s^2$ )	77.211	72.384
Kurtose ( $g_2$ )	-0.207	-0.059
Assimetria ( $g_1$ )	-0.484	-0.529
Amplitude	43	40
Mínimo	421	422
Máximo	464	462
1º Quartil	438.5	437
3º Quartil	450.5	452
Amplitude interquartil	12	15

■

## **PROBLEMA 2.6**

### *APOIOS À RESOLUÇÃO*

#### Apoio 1 (apenas o resultado)

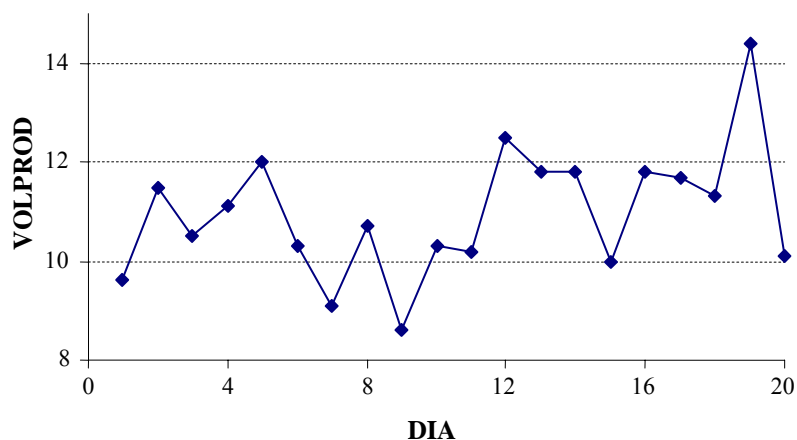
Não aplicável. ■

#### Apoio 2 (sugestão)

Note que se está perante uma amostra bivariada (constituída por pares ordenados de dados). Procure representar num diagrama  $(x, y)$  os dados do enunciado. Verifique se os volumes de produção são estáveis ao longo do tempo. Para caracterizar a relação entre os volumes de produção e os dias sucessivos, ajuste uma relação linear entre as variáveis. ■

#### Apoio 3 (resolução completa)

Ao contrário do que sucedia nos os problemas anteriores, os dados correspondem agora a uma série temporal (volumes de produção em 20 dias consecutivos). Faz então sentido começar por verificar se os volumes de produção são estáveis ao longo do tempo. O gráfico VOLPROD (volume de produção) *versus* DIA, que se apresenta em seguida, sugere que a primeira variável aumenta ligeiramente com o tempo.



Nestas condições, recorrendo ao método dos mínimos quadrados, ajustou-se uma relação linear entre as variáveis, do tipo seguinte:

$$y_n = a' + b \cdot x_n.$$

Os parâmetros  $a'$  e  $b$  foram estimados a partir das expressões

$$a' = \bar{y} - b \cdot \bar{x}$$

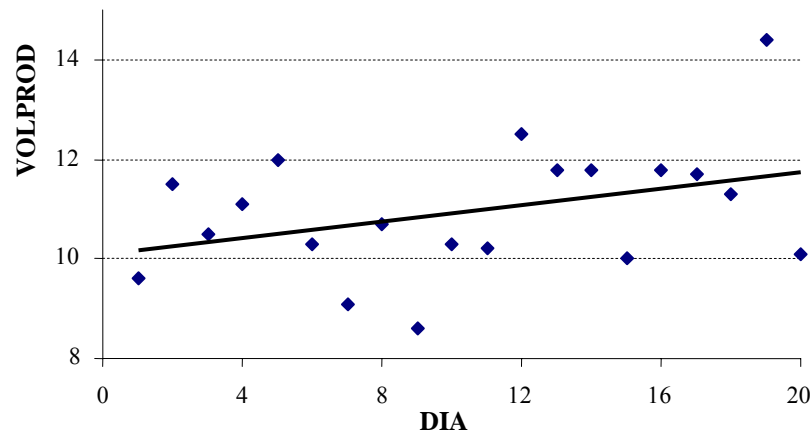
e

$$b = \frac{s_{XY}}{s_{XX}}.$$

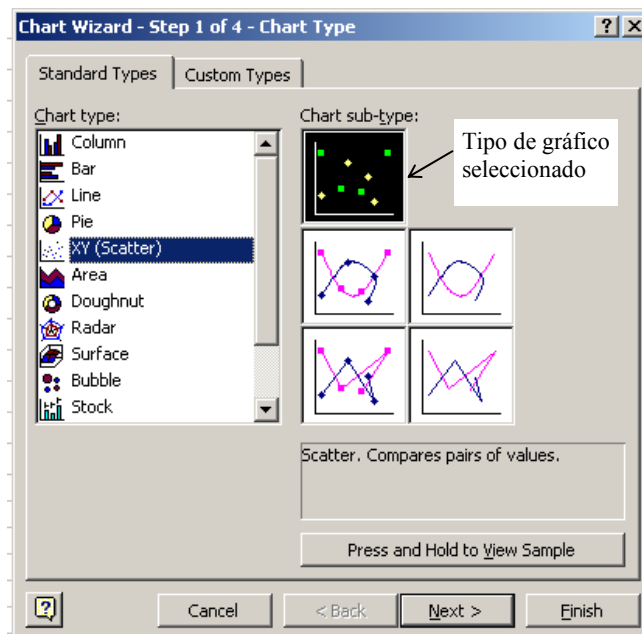
Os resultados conduzem à seguinte expressão:

$$VOLPROD = 10.1 + 0.0825 \cdot DIA.$$

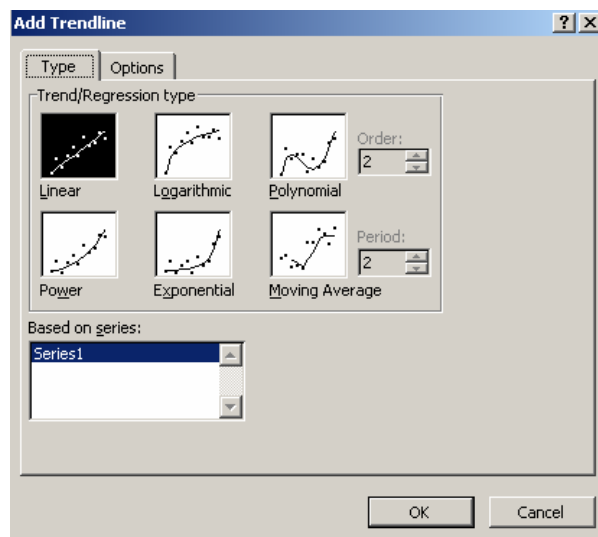
O coeficiente de determinação,  $R^2 = 13.9\%$ , representa a proporção da variação dos dados de *VOLPROD* que é explicada pela regressão linear, a partir da variação dos dias – variável *DIA*. Registe-se que, atendendo as grandes oscilações do volume de produção de dia para dia, o valor de  $R^2 = 13.9\%$  é baixo. No diagrama (*VOLPROD*, *DIA*) seguinte apresenta-se a relação linear ajustada.



Para obter esta figura a partir do programa “Microsoft Excel” procede-se do seguinte modo. Após ter introduzido os dados na folha de cálculo, selecciona-se o ícone “Chart Wizard” e a opção “XY (Scatter)”. Por defeito, é seleccionado o gráfico tipo “Scatter” tal como se indica na figura.

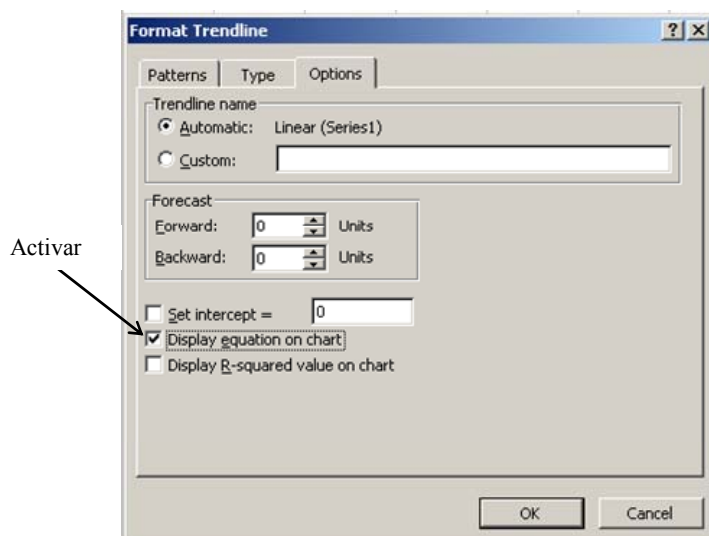


Seguidamente, acciona-se o botão “Next >” e, na janela que fica disponível, define-se o “Data range” seleccionando-se as células da folha de cálculo que contêm os dados e os respectivos descritivos (é necessário verificar se os dados estão dispostos em linha ou em coluna e seleccionar a opção correspondente: “Rows” ou “Columns”). Desta forma o gráfico ( $x$ ,  $y$ ) fica definido. Para acrescentar a recta que, pelo método dos mínimos quadrados, melhor se ajusta aos dados, selecciona-se um dos pontos do gráfico e, com o auxílio do botão do lado direito do “rato”, escolhe-se a opção “Add Trendline”. Automaticamente, surge a janela seguinte



Uma vez que se pretende ajustar uma relação linear, de entre as várias opções que ficam disponíveis ao seleccionar o separador “Type” escolhe-se “Linear” (ver figura anterior). Fazendo “OK” o processo termina com a sobreposição da recta pretendida no gráfico (x, y).

A equação da recta pode ser também obtida. Para tal, basta seleccionar a recta, accionar o botão do lado direito do “rato” e escolher “Format Trendline”, fazendo surgir a seguinte janela:



Escolhendo “Options”, activa-se a opção relativa a “Display equation on chart” (ver figura anterior) e carregando em “OK” obtém-se a equação pretendida. ■

## **PROBLEMA 2.7**

### *APOIOS À RESOLUÇÃO*

Apoio 1 (apenas o resultado)

Não aplicável. ■

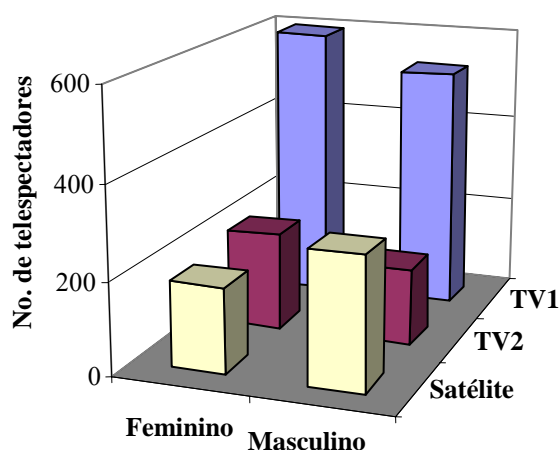
Apoio 2 (sugestão)

Procure representar os dados através de uma tabela de informação cruzada e de um diagrama de barras tridimensional (ou diagrama de barras sobrepostas). ■

Apoio 3 (resolução completa)

Está-se perante uma amostra bivariada com dados expressos em escalas nominais. Seguidamente apresenta-se a respectiva tabela de informação cruzada e o correspondente diagrama de barras tridimensional.

		Estação			
		TV1	TV2	Satélite	Total
Género	Masculino	528 26,8%	164 8,3%	285 14,4%	977 49,5%
	Feminino	598 30,3%	212 10,7%	186 9,4%	996 50,5%
	Total	1126 57,1%	376 19,1%	471 23,9%	1973 100,0%



Pode agora, facilmente, ordenar-se as estações por ordem decrescente de importância, de acordo com o género:

Sexo masculino: TV1, SATÉLITE, TV2

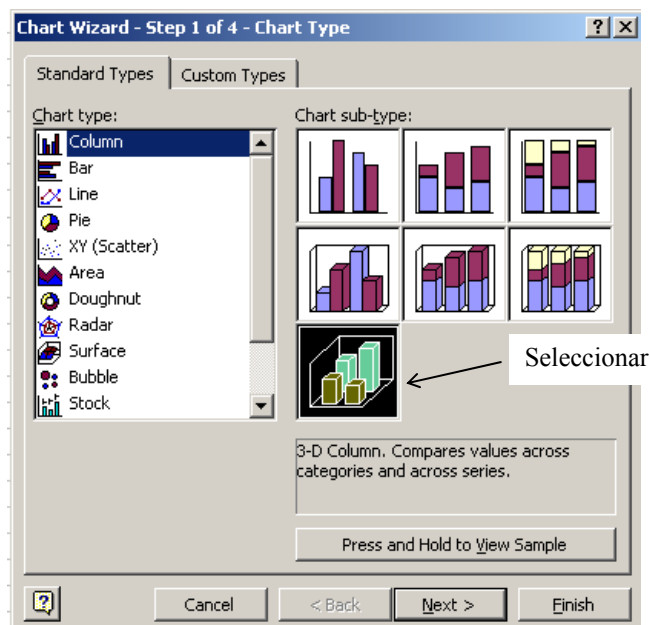
Sexo feminino: TV1, TV2, SATÉLITE.



O diagrama de barras tridimensional é facilmente obtido com auxílio do programa “Microsoft Excel”. Em primeiro lugar, introduzem-se os dados na folha de cálculo, criando-se uma tabela tal como se representa na figura seguinte.

	A	B	C	D	E	F
1						
2			TV1	TV2	Satélite	
3		Masculino	528	164	285	
4		Feminino	598	212	186	
5						

Seguidamente, selecciona-se esta tabela (ver figura anterior), acciona-se o ícone “Chart Wizard” e escolhe-se a opção “Column”. Finalmente, seleccionado o tipo gráfico que se indica na figura seguinte, acciona-se o botão “Next” e seguem-se as instruções do programa até se obter a figura pretendida.



## **PROBLEMA 2.8**

### *APOIOS À RESOLUÇÃO*

Apoio 1 (apenas o resultado)

Não aplicável. ■

Apoio 2 (sugestão)

Repare que se está perante uma amostra bivariada (constituída por pares ordenados de dados). Procure representar num diagrama  $(x, y)$  os dados do enunciado. Ajuste uma relação linear entre as variáveis “custo directo de produção” e “dimensão do lote”. ■

### Apoio 3 (resolução completa)

No diagrama seguinte apresenta-se uma relação linear (do tipo  $y_n = a' + b \cdot x_n$ ) ajustada ao conjunto de dados (pelo método dos mínimos quadrados). A variável  $y$  (*CUSTDIRECT*) representa o custo directo de produção e a variável  $x$  (*DIMLOTES*) a dimensão do lote.

Os parâmetros  $a'$  e  $b$ , estimados a partir das expressões

$$a' = \bar{y} - b \cdot \bar{x}$$

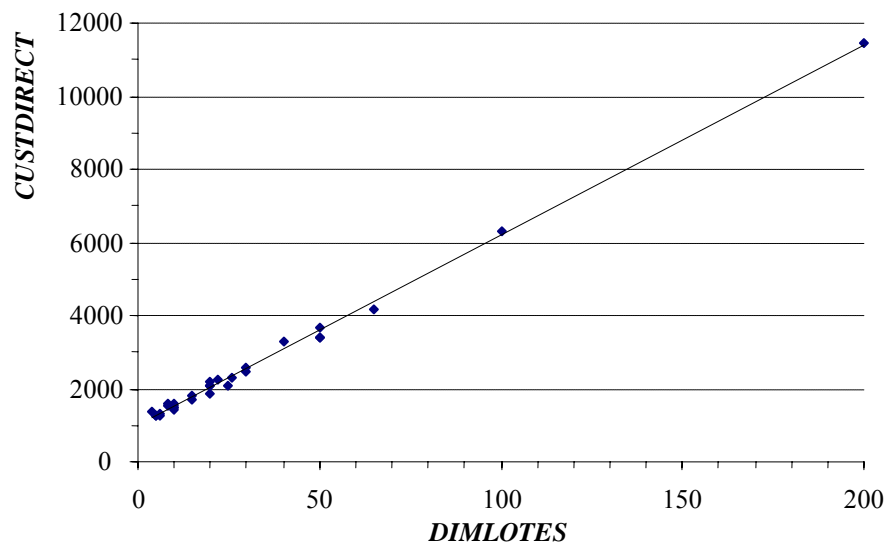
e

$$b = \frac{s_{XY}}{s_{XX}},$$

conduzem à seguinte expressão da recta:

$$CUSTDIRECT = 997.8 + 51.918 \cdot DIMLOTES.$$

O coeficiente de determinação é muito elevado ( $R^2 = 99.6\%$ ), traduzindo o bom ajuste da recta aos dados.



Para obter esta figura a partir do programa “Microsoft Excel” (bem como a equação da recta e também o coeficiente de determinação) proceder-se-ia de forma idêntica à que foi indicada no problema 2.6. ■