

2. Representação e redução de dados. Tabelas e gráficos

2.1- Introdução

Num módulo anterior de Estatística, já foram apresentados vários processos de organizar os dados, de forma a realçar as características principais e a estrutura subjacente da população de onde esses dados foram retirados.

Quer estejamos perante uma variável de tipo discreto ou contínuo, o processo de organizar a informação consiste em, de um modo geral, começar por construir tabelas de frequência e proceder a representações gráficas adequadas. Vamos seguidamente utilizar o Excel na construção dessas tabelas de frequência.

2.2 – Utilização do Excel na obtenção de tabelas de frequência

Vamos exemplificar a utilização do Excel na construção de tabelas de frequência a partir do ficheiro *Deputados.xls*, apresentado no capítulo anterior.

2.2.1 – Tabela de dados qualitativos ou quantitativos discretos

O procedimento para a construção das tabelas de frequência é idêntico, quer tenhamos um conjunto de dados qualitativos ou quantitativos discretos, já que as classes que se consideram são as diferentes categorias ou valores que surgem, respectivamente, no conjunto de dados. A seguir apresentamos a construção destas tabelas utilizando a função *COUNTIF*. Numa secção posterior veremos a sua construção utilizando a metodologia das *PivotTables*.

Exemplo 2.2.1 – Considere o ficheiro *Deputados.xls*. Obtenha uma tabela de frequência para a variável Grupo Parlamentar.

Começámos por copiar a coluna correspondente ao Grupo parlamentar para um novo ficheiro. Ordenámos os elementos por ordem crescente e inserimos na coluna **Classes** os diferentes elementos do conjunto de dados. Utilizámos de seguida a função *COUNTIF* (*CONTAR.SE*) para obter as frequências absolutas de deputados de cada um dos grupos parlamentares:

	A	B	C	D	E
1	Grupo				
2	Parlamentar			Tabela de frequências	
3	BE		Classes	Freq. Abs.	Freq. Rel.
4	BE		BE	=COUNTIF(\$A\$3:\$A\$232;C4)	=D4/\$D\$10
5	BE		CDS-PP	=COUNTIF(\$A\$3:\$A\$232;C5)	=D5/\$D\$10
6	BE		PCP	=COUNTIF(\$A\$3:\$A\$232;C6)	=D6/\$D\$10
7	BE		PEV	=COUNTIF(\$A\$3:\$A\$232;C7)	=D7/\$D\$10
8	BE		PS	=COUNTIF(\$A\$3:\$A\$232;C8)	=D8/\$D\$10
9	BE		PSD	=COUNTIF(\$A\$3:\$A\$232;C9)	=D9/\$D\$10
10	BE			=SUM(D4:D9)	=SUM(E4:E9)

ou

	A	B	C	D	E
1	Grupo				
2	Parlamentar			Tabela de frequências	
3	BE		Classes	Freq. Abs.	Freq. Rel.
4	BE		BE	=COUNTIF(\$A\$3:\$A\$232;"BE")	=D4/\$D\$10
5	BE		CDS-PP	=COUNTIF(\$A\$3:\$A\$232;"CDS-PP")	=D5/\$D\$10
6	BE		PCP	=COUNTIF(\$A\$3:\$A\$232;"PCP")	=D6/\$D\$10
7	BE		PEV	=COUNTIF(\$A\$3:\$A\$232;"PEV")	=D7/\$D\$10
8	BE		PS	=COUNTIF(\$A\$3:\$A\$232;"PS")	=D8/\$D\$10
9	BE		PSD	=COUNTIF(\$A\$3:\$A\$232;"PSD")	=D9/\$D\$10
10	BE			=SUM(D4:D9)	=SUM(E4:E9)

As fórmulas apresentadas anteriormente, deram origem à seguinte tabela:

	A	B	C	D	E	F
1	Grupo					
2	Parlamentar			Tabela de frequências		
3	BE		Classes	Freq. Abs.	Freq.	
4	BE		BE	8	0,035	
5	BE		CDS-PP	12	0,052	
6	BE		PCP	12	0,052	
7	BE		PEV	2	0,009	
8	BE		PS	121	0,526	
9	BE		PSD	75	0,326	
10	BE			230	1	
11	CDS-PP					

2.2.2 – Tabela de dados quantitativos contínuos

Como se viu no módulo anterior de Estatística, no caso de dados contínuos o processo de construção das tabelas é um pouco mais elaborado, já que a definição das classes não é tão imediata. De um modo geral as classes são intervalos com a mesma amplitude, fechados à esquerda e abertos à direita ou abertos à esquerda e fechados à direita. Em certos casos não é conveniente que as classes tenham a mesma amplitude, o que em si não é um problema para a construção da tabela de frequências, mas que implica alguma complicação na construção do histograma associado, quando pretendemos utilizar Excel.

Vamos utilizar ainda o ficheiro *Deputados.xls* para estudar a variável Idade, que é uma variável quantitativa contínua.

Exemplo 2.2.2 – Utilizando a informação contida no ficheiro *Deputados.xls*, construa uma tabela de frequências para a variável Idade.

Vamos dividir esta tarefa em duas partes: uma primeira parte consistirá na definição das classes e uma segunda parte no cálculo das frequências.

Copie a coluna “Data de nascimento” para um ficheiro novo com 230 elementos que ocupam as células A2:A231. Para obter a idade em 31/12/2007, podemos utilizar a seguinte metodologia:

- Passo 1 – Inserir na célula B1 a data 31/12/2007;
- Passo 2 – Colocar o cursor na célula B2 e introduzir a expressão: =B\$1-A2;
- Passo 3 – Replicar esta função através das células B3 a B231;
- Passo 4 - Se no passo anterior se obteve uma coluna de datas, formatar essa coluna com o *Format Geral*, por exemplo. Obtém-se a idade em dias;
- Passo 5 – Para obter a idade em anos, colocar o cursor na célula C2 e introduzir a seguinte função: =INT(B2/365), a qual devolve o maior inteiro contido no quociente (n.º de dias do deputado)/(n.º de dias do ano).
- Replicar esta função através das células C3 a C231.

Definição das classes:

- a) Determinar a amplitude da amostra, subtraindo o mínimo do máximo;
- b) Dividir essa amplitude pelo número K de classes pretendido. Existe uma regra empírica que nos dá um valor aproximado para o número K de classes e que consiste no seguinte: para uma amostra de dimensão n, considerar para K o menor inteiro tal que $2^K \geq n$. Uma expressão equivalente para obter K, consiste em considerar $K = \text{INT}(\text{LOG}(n;2)) + 1$ ou $K = \text{ROUNDUP}(\text{LOG}(n;2);0)$, em que a função *ROUNDUP*(x;m), devolve um valor de x, arredondado por excesso, com m casas decimais;
- c) Calcular a amplitude de classe h, dividindo a amplitude da amostra por K e tomando para h um valor aproximado por excesso do quociente anteriormente obtido;

- d) Construir as classes C_1, C_2, \dots, C_k . Vamos considerar como classes os intervalos $[\text{mínimo}, \text{mínimo} + h[, [\text{mínimo} + h, \text{mínimo} + 2h[, \dots, [\text{mínimo} + (k-1)h, \text{mínimo} + kh[$. Uma alternativa a este procedimento seria considerar as classes abertas à esquerda e fechadas à direita, da seguinte forma: $]\text{max} - Kh, \text{max} - (K-1)h[,]\text{max} - (K-1)h, \text{max} - (K-2)h[,]\text{max} - h, \text{max}[$.

Estes passos são representados na figura seguinte:

	C	D	E	F	G	H	I
1	Idade						
2	=INT(B2/365)	Mínimo	=MIN(C2:C231)			Classes	
3	=INT(B3/365)	Máximo	=MAX(C2:C231)			Limite inferior	Limite superior
4	=INT(B4/365)	Amplitude	=F3-F2		=F2		=H4+\$F\$8
5	=INT(B5/365)	n	=COUNT(C2:C231)		=4		=H5+\$F\$8
6	=INT(B6/365)	k	=INT(LOG(F5;2))+1		=5		=H6+\$F\$8
7	=INT(B7/365)	amplitude/k	=F4/F6		=6		=H7+\$F\$8
8	=INT(B8/365)	h	5,7		=7		=H8+\$F\$8
9	=INT(B9/365)				=8		=H9+\$F\$8
10	=INT(B10/365)				=9		=H10+\$F\$8
11	=INT(B11/365)				=10		=H11+\$F\$8

com os seguintes resultados:

	C	D	E	F	G	H	I
1	Idade						
2	53	Mínimo	28			Classes	
3	32	Máximo	73			Limite inferior	Limite superior
4	61	Amplitude	45		28		33,7
5	51	n	230		33,7		39,4
6	48	k	8		39,4		45,1
7	56	amplitude/k	5,625		45,1		50,8
8	50	h	5,7		50,8		56,5
9	53				56,5		62,2
10	44				62,2		67,9
11	39				67,9		73,6

Cálculo das frequências:

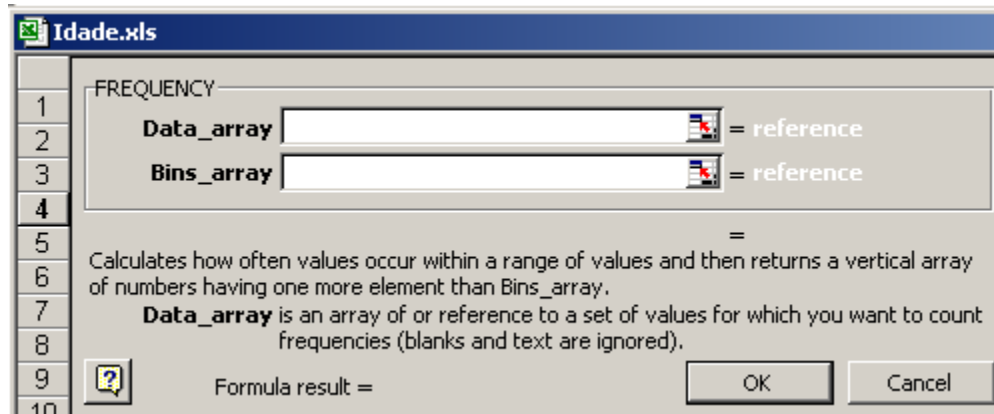
Para obter as frequências absolutas, vamos utilizar a função *COUNTIF* do seguinte modo:

L5		fx		=COUNTIF(\$C\$2:\$C\$231;"<39,4")-COUNTIF(\$C\$2:\$C\$231;"<33,7")							
	C	D	E	F	Classes			Tabela de frequências			
1	Idade				Limite inferior	Limite superior	Classes	Freq. Abs.	Freq. Rel.		
2	53	Mínimo		28							
3	32	Máximo		73							
4	61	Amplitude		45	28	33,7	c1 [28,0; 33,7[19	0,083		
5	51	n		230	33,7	39,4	c2 [33,7; 39,4[29	0,126		
6	48	k		8	39,4	45,1	c3 [39,4; 45,1[39	0,170		
7	56	amplitude/k		5,625	45,1	50,8	c4 [45,1; 50,8[38	0,165		
8	50	h		5,7	50,8	56,5	c5 [50,8; 56,5[50	0,217		
9	53				56,5	62,2	c6 [56,5; 62,2[38	0,165		
10	44				62,2	67,9	c7 [62,2; 67,9[10	0,043		
11	39				67,9	73,6	c8 [67,9; 73,6[7	0,030		
12	37							230	1		

As frequências das classes c1, c3..., c8, são obtidas de forma idêntica à de c2, mudando os limites das classes.

2.2.3 - Construção de uma tabela de frequências utilizando a função *Frequency* do Excel

O Excel tem uma função, que é a função *Frequency(Data_array;Bins_array)*, que calcula o número de elementos da variável - cujos valores se encontram na **Data_array**, existentes nas



classes - cujos limites se encontram em **Bins_array**. Este vector **Bins_array** é constituído por um conjunto de k valores b_1, b_2, \dots, b_k , formando (k+1) classes, tais que:

- A 1ª classe é dada por $(-\infty, b_1]$, isto é, conterà todos os elementos $\leq b_1$;
- A 2ª classe é dada por $]b_1, b_2]$;
- A 3ª classe é dada por $]b_2, b_3]$;
- A késima classe é dada por $]b_{k-1}, b_k]$;
- A (k+1)ésima classe é dada por $]b_k, +\infty)$;

Vamos exemplificar construindo uma tabela de frequências para a variável idade.

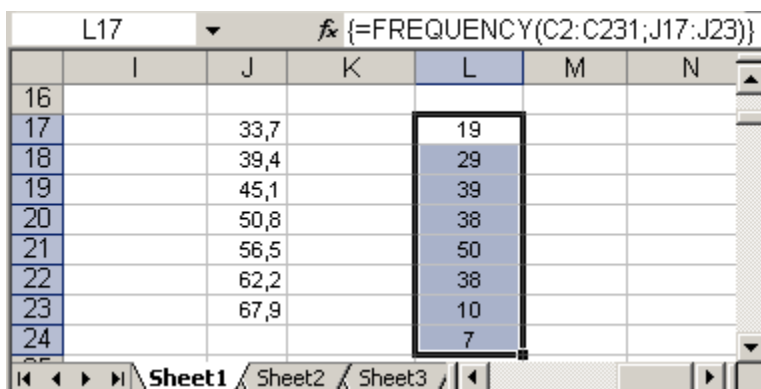
Definição das classes:

Considerando as classes definidas em 2.2 e tendo em atenção o que dissemos anteriormente sobre as classes para a utilização da função *Frequency*, o nosso conjunto de valores para o **Bins_array**, será constituído por {33,7; 39,4; 45,1; 50,8; 56,5; 62,2; 67,9};

Para utilizar a função *Frequency(Data_array;Bin_array)*, procede-se do seguinte modo:

- Definir a coluna de separadores ou limites das classes, que constituirá o **Bins_array**;
- Seleccionar tantas células em coluna, quantas as classes consideradas para a tabela de frequências (não esquecer que o número de classes é superior em uma unidade ao número de separadores, pelo que o número de células seleccionadas deverá ser, neste caso, de 8);
- Introduzir a função *Frequency*, considerando como primeiro argumento o conjunto de células onde se encontram os dados a agrupar, chamado de **Data_array**, e como segundo argumento as células que constituem o **Bins_array**;
- Carregar **CTRL+SHIFT+ENTER**

Na figura seguinte apresentamos o resultado deste procedimento:



	I	J	K	L	M	N
16						
17		33,7		19		
18		39,4		29		
19		45,1		39		
20		50,8		38		
21		56,5		50		
22		62,2		38		
23		67,9		10		
24				7		

Verifique que os valores devolvidos pela função *Frequency*, nas células L17: L24, são iguais às frequências obtidas anteriormente e apresentadas na tabela de frequências já construída. Esta situação nem sempre se verifica, nomeadamente se os limites das classes fossem números inteiros, já que agora as classes são consideradas fechadas à direita e abertas à esquerda. Assim, alguns valores da amostra que anteriormente não pertenciam a determinadas classes, poderiam agora pertencer.

2.3 – Utilização do Excel na representação gráfica de dados

De forma idêntica à que fizemos para a construção das tabelas de frequências, vamos também considerar separadamente o caso da variável em estudo ser de natureza qualitativa ou quantitativa discreta, ou de natureza quantitativa contínua.

2.3.1 – Variáveis qualitativas ou quantitativas discretas. Diagrama de barras

Neste caso vimos que a construção da tabela de frequências se resume, de um modo geral, a considerar como classes as diferentes categorias ou valores que surgem na amostra. Uma representação gráfica adequada para estes dados, é o diagrama de barras, que já foi introduzido no módulo de Estatística.


Diagrama de barras – Representação gráfica que consiste em marcar num sistema de eixos coordenados, no eixo dos xx, pontos representando as categorias ou os valores considerados para as classes na tabela de frequências, e nesses pontos barras verticais de altura igual à frequência absoluta ou à frequência relativa.

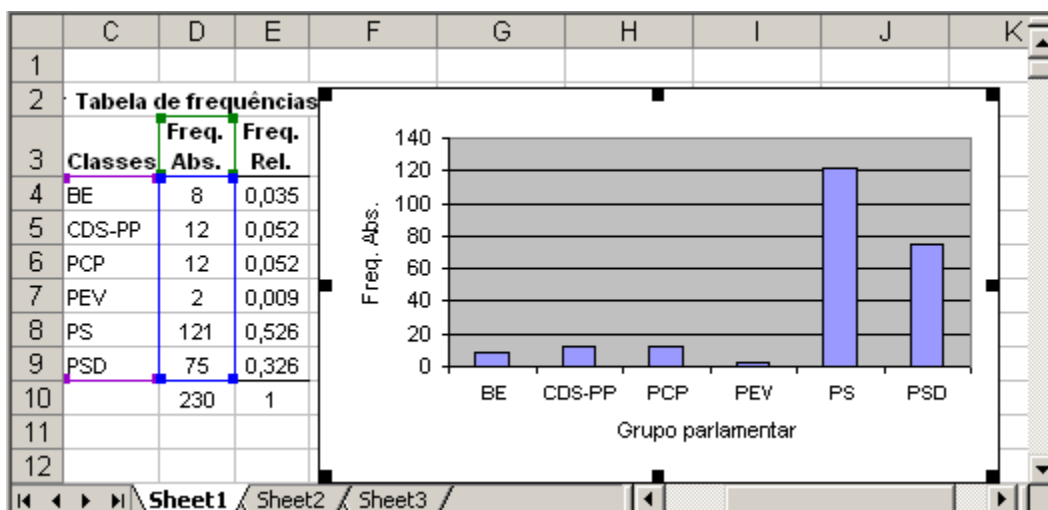
2.3.1.1 - Variável de tipo qualitativo

Exemplo 2.3.1 - Vamos exemplificar a construção de um diagrama de barras de uma variável qualitativa, considerando a tabela de frequências construída em 2.2.1, para estudar a variável Grupo Parlamentar, do ficheiro *Deputados.xls*:

Classes	Freq. Abs.	Freq. Rel.
BE	8	0,035
CDS-PP	12	0,052
PCP	12	0,052
PEV	2	0,009
PS	121	0,526
PSD	75	0,326
	230	1

A metodologia seguida para construir o diagrama de barras, consiste em, na folha Excel, que contém a tabela:

- Seleccionar as células que contêm as classes e as frequências absolutas (por exemplo);
- Seleccionar, no menu, o ícone Chart ;
- Na caixa de diálogo que aparece, seleccionar a opção *Column*;
- Clicar no botão *Next*, duas vezes, para passar dois passos, até aparecer uma caixa de diálogo, que apresenta várias opções: Em *Legend*, desactivar a legenda e em *Titles*, acrescentar o título no eixo dos Y's e no eixo dos X's, como se apresenta a seguir, e carregar em *Finish*:



2.3.1.2 - Variável de tipo quantitativo discreto

2.3.1.2.1 – Diagrama de barras

No caso de dados discretos, para construir a tabela de frequência consideram-se como classes os diferentes valores que surgem na amostra. Estes valores devem ser apresentados, na tabela de frequência, ordenados.

Exemplo 2.3.2 – Suponhamos que para uma amostra de 30 deputados da actual legislatura, se tinha recolhido a informação sobre o número de filhos, tendo-se obtido os seguintes valores:

2, 1, 2, 3, 0, 0, 1, 1, 4, 1, 2, 1, 0, 0, 0, 2, 3, 1, 1, 6, 3, 1, 3, 2, 0, 1, 2, 0, 2, 3

Resuma os dados numa tabela de frequências e construa o diagrama de barras associado.

Introduzimos os dados numa folha de Excel, a que chamámos *Filhos.xls* e a seguir procedemos do seguinte modo:

1ª parte – Procedimento para a construção da tabela de frequências:



- Seleccionar as células que contêm os dados e ordená-los utilizando o botão  da barra de Excel;
- Considerar para classes os diferentes valores que surgem na amostra. Se faltar algum valor entre o máximo e o mínimo, considerá-lo também na tabela de frequências, se a seguir se pretende construir um diagrama de barras;
- Utilizando a função *COUNTIF*, determinar as frequências absolutas das classes consideradas no ponto anterior; calcular a partir destas, as frequências relativas:

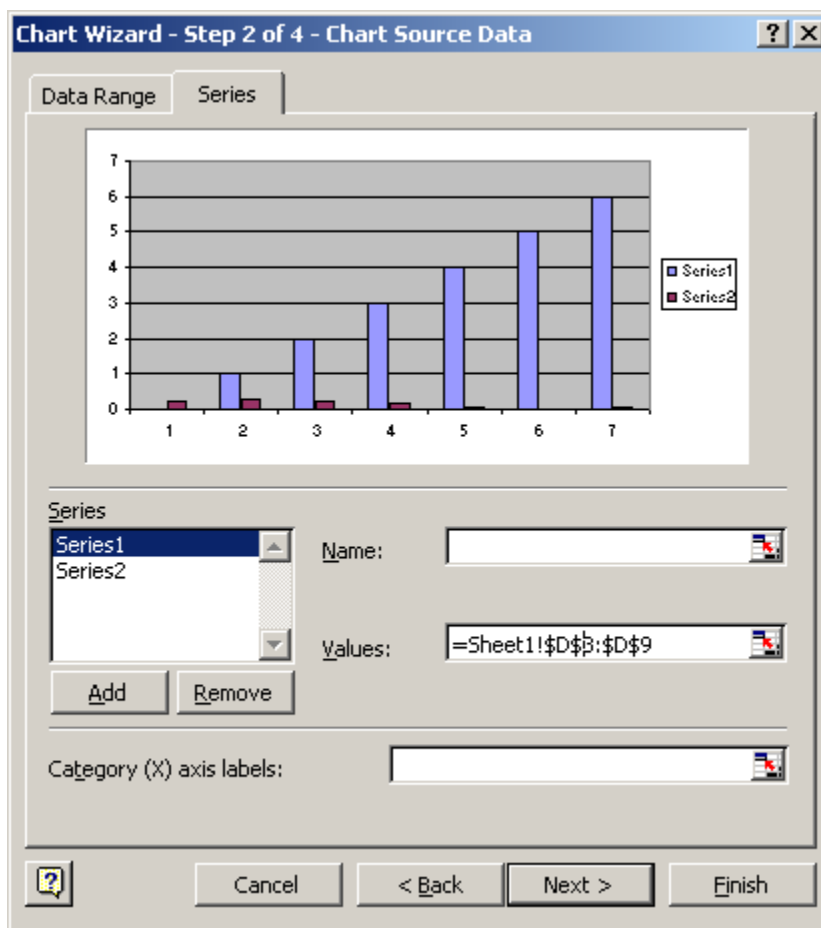
Tabela de frequências

Classes	Freq.abs.	Freq.rel.
0	7	0,233
1	9	0,300
2	7	0,233
3	5	0,167
4	1	0,033
5	0	0,000
6	1	0,033

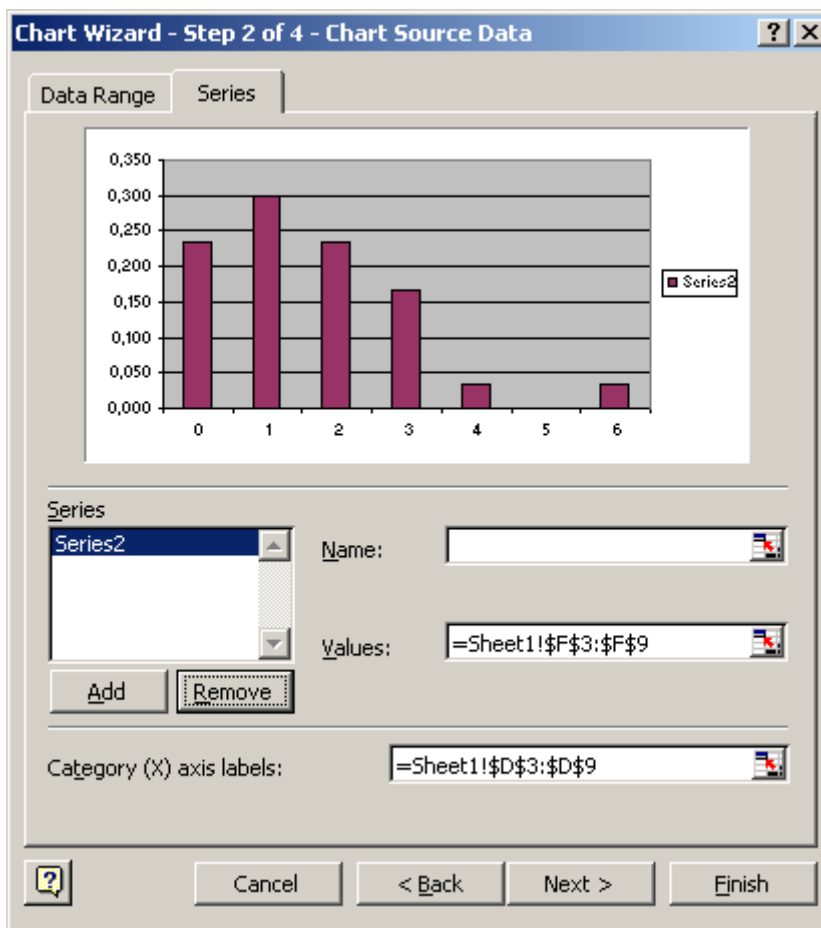
30

2ª parte – Procedimento para a construção do diagrama de barras:

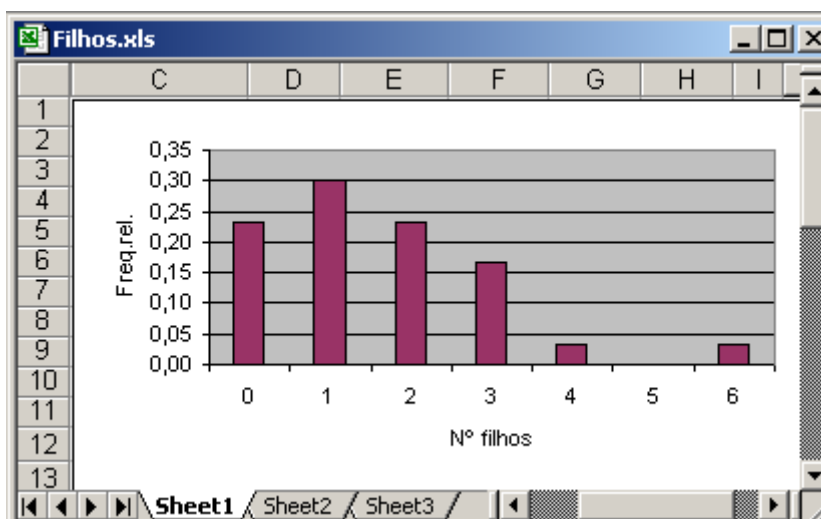
- Seleccionar as células que contêm as classes e as que contêm as frequências relativas (para seleccionar as células que contêm as frequências relativas, como não são adjacentes às que contêm as classes, depois de seleccionar estas, tem que se pressionar a tecla CTRL e com ela pressionada, seleccionar aquelas);
- Seleccionar na barra de ferramentas a opção *Chart*  e a seguir a opção *Column*, tal como se fez para os dados de tipo qualitativo;
- Seleccionar *Next* e de seguida *Series*, como se apresenta a seguir:



- Copiar a *Series1*, dada pelos valores =Sheet1!\$D\$3:\$D\$9, que constituem as classes, para *Category (X) axis labels* e remover *Series1* de *Series*:



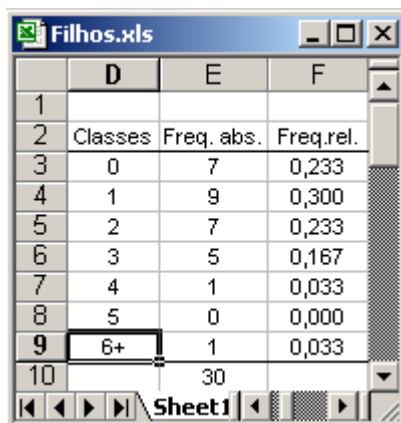
- Seleccionar *Next*. Nas *Chart Options* seleccionar *Legend* e retirar a selecção de *Show Legend*. Seleccionar *Titles* e colocar títulos adequados. Carregar em *Finish*:



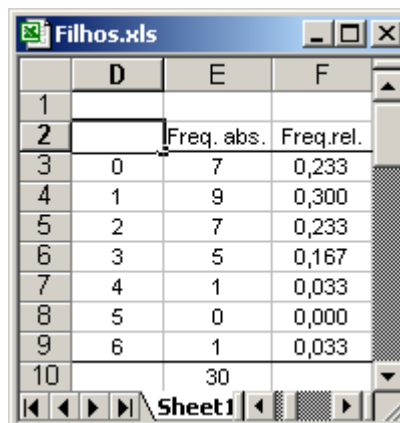
Nota (Esta nota foi sugerida pela leitura do artigo de *Neville Hunt – Charts in Excel, in Teaching Statistics, Volume 26, Number 2, pags 49-53*): Como vimos na descrição que acabámos de fazer para a construção de um diagrama de barras em Excel, o procedimento está longe de ser natural, já que o que seria de esperar era que, uma vez seleccionada a opção Column, nos surgisse a representação gráfica anterior, aparte pequenas alterações de “cosmética”.

Suponhamos, no entanto, que na última classe tínhamos considerado 6+, para significar 6 ou mais filhos. Então, ao fim dos dois primeiros passos da descrição anterior temos a representação gráfica pretendida. O facto é que agora o Excel interpretou as classes como categorias e fez a representação esperada.


Suponhamos ainda, que em vez de modificarmos o 6 para 6+, apagamos o conteúdo de D2:

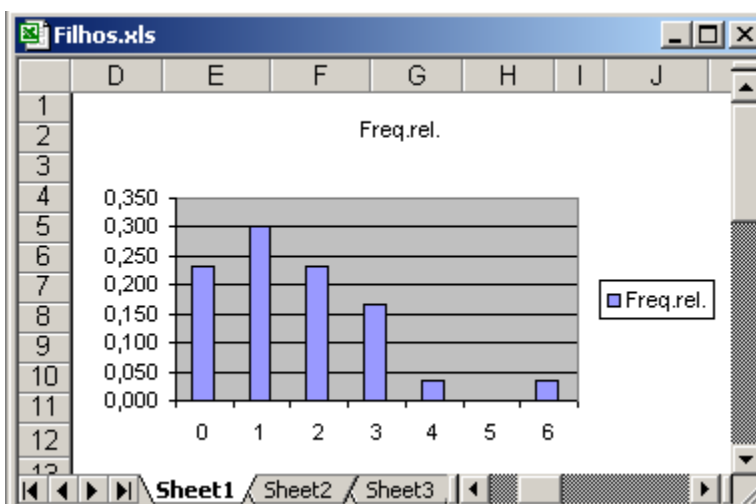


	D	E	F
1			
2	Classes	Freq. abs.	Freq.rel.
3	0	7	0,233
4	1	9	0,300
5	2	7	0,233
6	3	5	0,167
7	4	1	0,033
8	5	0	0,000
9	6+	1	0,033
10		30	



	D	E	F
1			
2		Freq. abs.	Freq.rel.
3	0	7	0,233
4	1	9	0,300
5	2	7	0,233
6	3	5	0,167
7	4	1	0,033
8	5	0	0,000
9	6	1	0,033
10		30	

Seleccionando agora as células D2 a D9 e F2 a F9 e novamente no Chart  a opção Column, então a representação que se obtém é, imediatamente, a seguinte:



Depois de apagar a legenda e inserir os títulos de forma conveniente, temos a representação final do exemplo anterior, sem grandes complicações.

2.3.1.2.2 – Função cumulativa

A função cumulativa é uma função definida para todo o valor real x , e que para cada x dá a soma das frequências dos valores da amostra menores ou iguais a x .

Quando temos uma variável de tipo discreto, a função cumulativa é uma função em escada, isto é, é uma função que cresce por degraus, mudando de degrau nos pontos em que a frequência é diferente de 0, e em que a altura do degrau é igual à frequência respectiva. Vamos exemplificar a sua construção com o exemplo apresentado na secção anterior para a construção do diagrama de barras.

Exemplo 2.3.2 (cont) – Construa a função cumulativa para os dados do número de filhos da amostra dos 30 deputados.

Retomando a tabela de frequências do exemplo 2.3.2, vamos acrescentar uma coluna com as frequências relativas acumuladas:

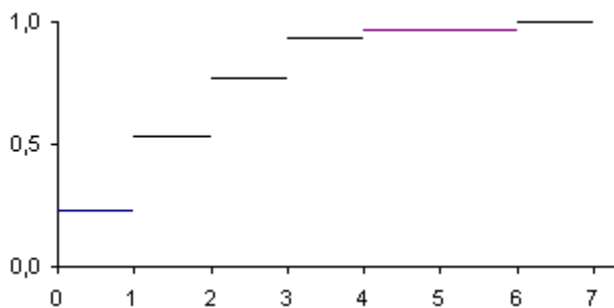
Classes	Freq.abs.	Freq.rel.	Freq.rel.acum.
0	7	0,233	0,233
1	9	0,300	0,533
2	7	0,233	0,767
3	5	0,167	0,933
4	1	0,033	0,967
5	0	0,000	0,967
6	1	0,033	1,000

30

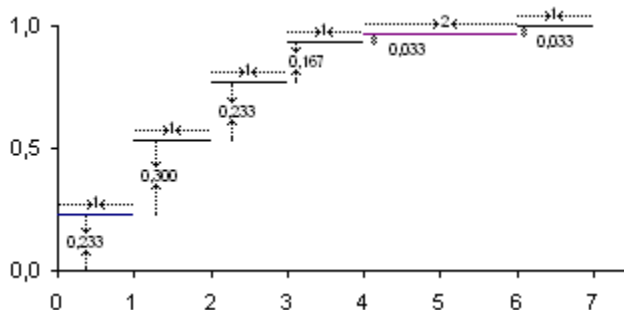
A função cumulativa há-de ser tal que:

- Para valores de $x < 0$, será nula;
- Para valores de $0 \leq x < 1$, será igual a 0,233;
- Para valores de $1 \leq x < 2$, será igual a 0,533;
- Para valores de $2 \leq x < 3$, será igual a 0,767;
- Para valores de $3 \leq x < 4$, será igual a 0,933;
- Para valores de $4 \leq x < 6$, será igual a 0,967;
- Para valores de $x \geq 6$, será igual a 1;

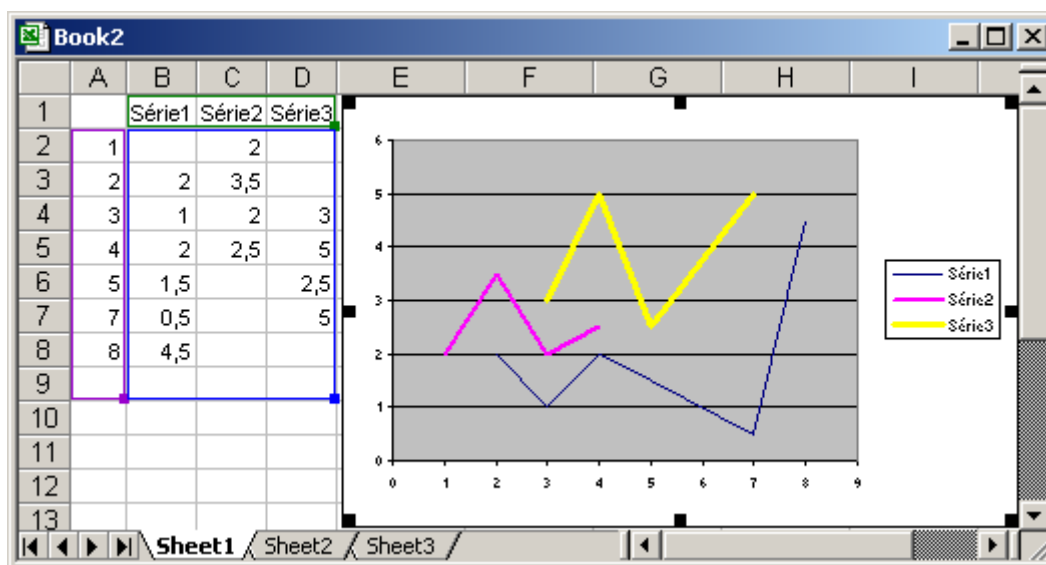
O Excel não dispõe de uma representação imediata para a função anterior, pelo que temos de utilizar um pequeno artifício. Suponhamos, para já, que por algum processo tínhamos conseguido construir o gráfico da função cumulativa, que tem o seguinte aspecto:



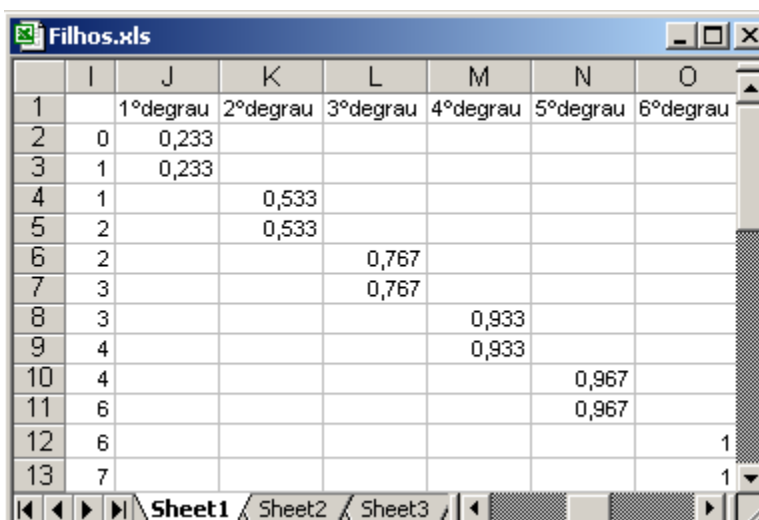
Esta função é constituída por 6 degraus, em que a altura do degrau é, em cada ponto, igual à frequência relativa respectiva e a dimensão do patamar é igual à diferença entre os pontos consecutivos, com frequência relativa diferente de zero:



O Excel dispõe de uma representação gráfica, o *Scatter* (Diagrama de dispersão), em que no último subtipo apresentado para as opções, une os pontos, por ordem crescente das abcissas, simultaneamente de tantas séries (conjuntos de pontos) quantas as desejadas. Exemplifiquemos com os pontos da seguinte tabela, em que pretendemos representar 3 conjuntos de dados a que chamámos Série1, Série2 e Série3:

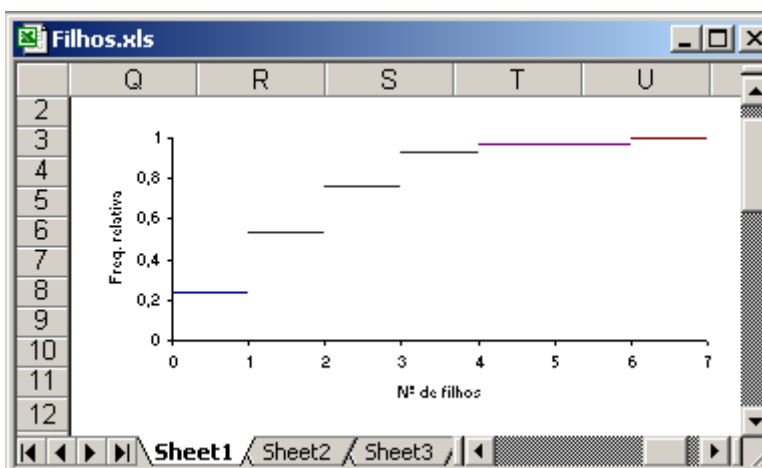


Vamos utilizar esta função *Scatter* para construir os sucessivos degraus da função cumulativa, em que cada degrau corresponde a uma série - união de dois pontos, e em que temos tantas séries a representar, quantos os degraus. Assim, o artifício está em representar, numa tabela do Excel, os degraus pretendidos através das coordenadas dos pontos, como exemplificamos a seguir:



	I	J	K	L	M	N	O
1		1ºdegrau	2ºdegrau	3ºdegrau	4ºdegrau	5ºdegrau	6ºdegrau
2	0	0,233					
3	1	0,233					
4	1		0,533				
5	2		0,533				
6	2			0,767			
7	3			0,767			
8	3				0,933		
9	4				0,933		
10	4					0,967	
11	6					0,967	
12	6						1
13	7						1

Agora basta seleccionar as células I2 a O13 e fazer o diagrama de dispersão, como indicado anteriormente. Proceda como na construção do diagrama de barras, para retirar a legenda e acrescentar títulos:



2.3.2 – Variáveis quantitativas contínuas

2.3.2.1 – Histograma

2.3.2.1.1 – Tabela de frequências com as classes com a mesma amplitude

No caso de um conjunto de dados contínuos, já vimos anteriormente a forma de obter a tabela de frequências. Como se viu, as classes são intervalos e a representação gráfica adequada é o *histograma*, já apresentado no módulo de Estatística:

Histograma - é um diagrama de áreas, formado por uma sucessão de rectângulos adjacentes, tendo cada um por base um intervalo de classe e por área a frequência relativa (ou frequência absoluta). Por conseguinte, a área total coberta pelo histograma é igual a 1 (ou igual a n , a dimensão do conjunto de dados a representar).

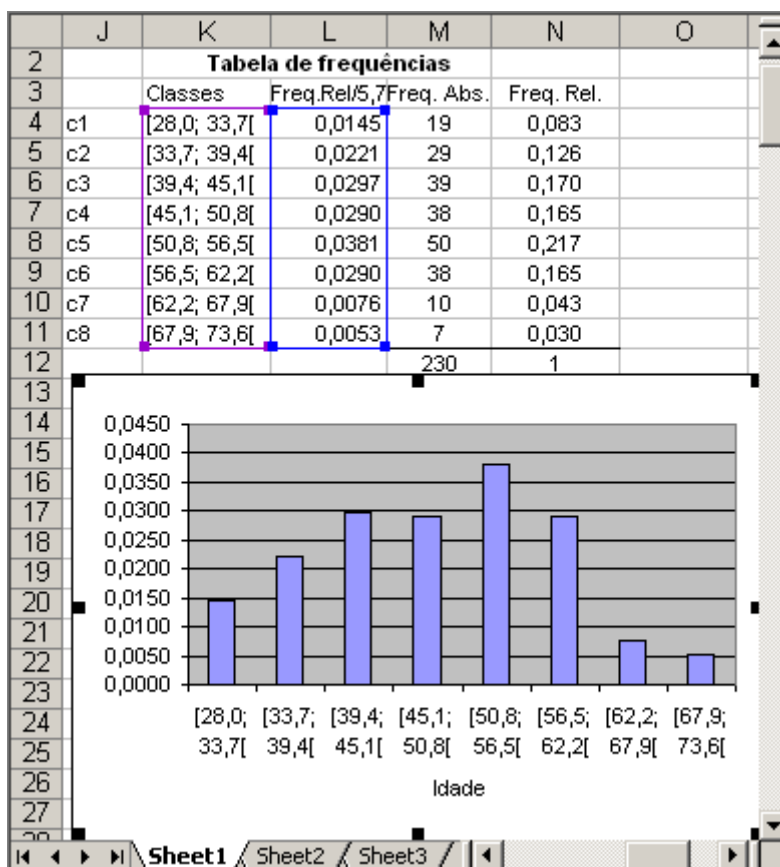
Para construir o histograma de forma correcta, isto é, de modo a que as áreas dos rectângulos sejam iguais às frequências, a altura do rectângulo correspondente a determinada classe, deverá ser igual à frequência da classe a dividir pela respectiva amplitude. Contudo, se as classes tiverem todas a mesma amplitude, é usual construir os rectângulos com alturas iguais às frequências relativas (absolutas) das respectivas classes, vindo as áreas dos rectângulos proporcionais e não iguais às frequências. A constante de proporcionalidade é a amplitude de classe. No entanto, se se pretender comparar amostras através de histogramas, embora o histograma não seja a representação mais adequada para a comparação de amostras, deve-se ter o cuidado de os construir da forma indicada inicialmente, e utilizando as frequências relativas, de modo que a área total ocupada por cada um dos histogramas seja igual a 1.

Exemplificamos, de seguida, a construção de um histograma utilizando o Excel.

Exemplo 2.3.3 – Considerando a tabela de frequências construída em 2.3 para a variável idade, construa o histograma adequado.

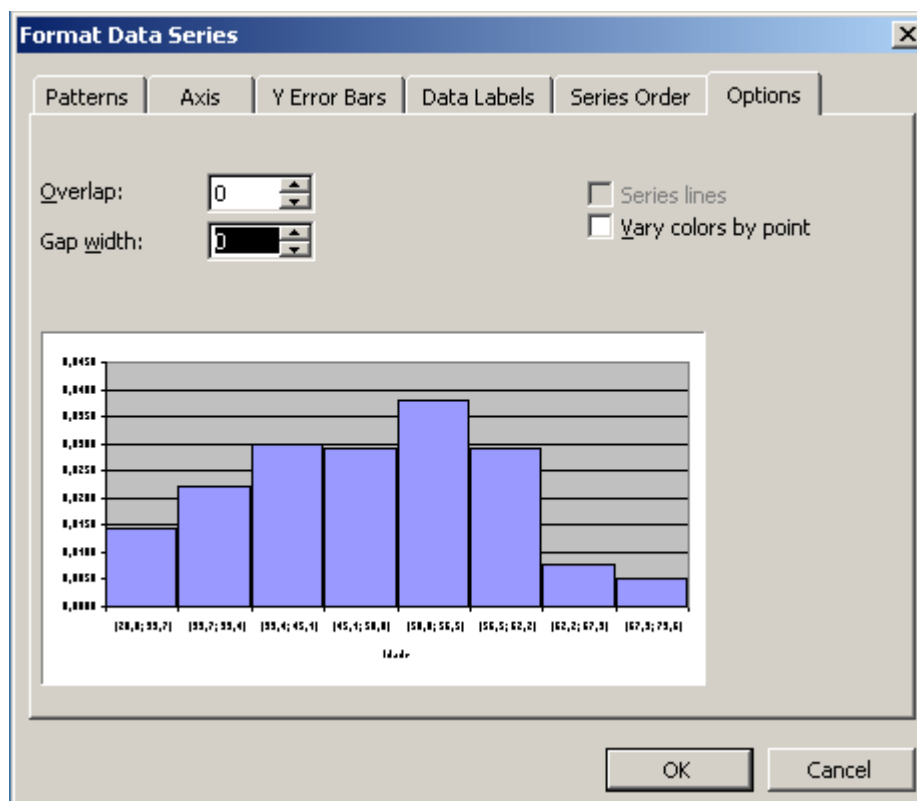
Processo utilizado para obter o histograma:

- Acrescentar, à tabela considerada, uma outra coluna com a frequência relativa a dividir pela amplitude de classe (igual a 5,7). No caso presente, inserimos estas células adjacentes às células que contêm as classes. No entanto, não é necessário ter esta preocupação, já que se se pretender seleccionar células não adjacentes, basta seleccionar as células da primeira coluna e se a coluna seguinte não for adjacente, começar por carregar a tecla *CTRL* e com ela pressionada seleccionar, então, as células pretendidas;
- Seleccionar as células de K4 a L11 (que contêm as classes e as frequências relativas a dividir pela amplitude de classe);
- Proceder como em 3.1 para construir um diagrama de barras, para obter a figura que se apresenta a seguir;

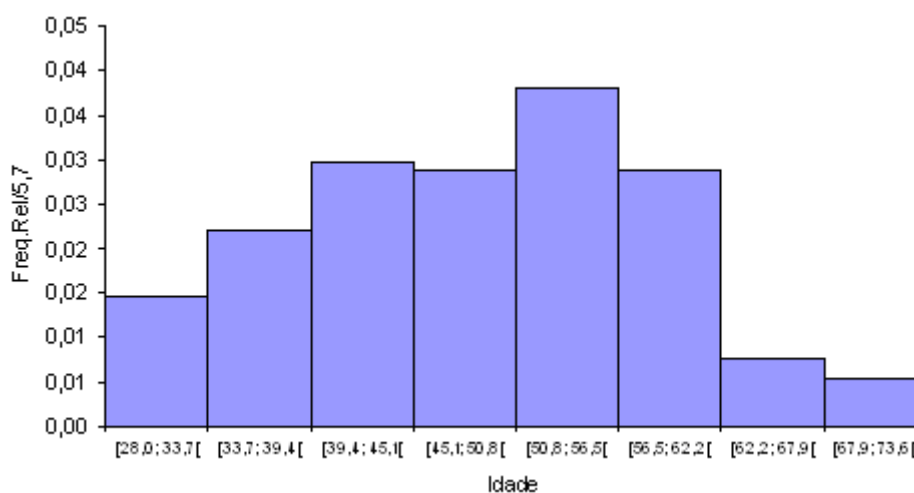


Para obter o histograma, já que o que se nos apresenta na figura anterior não é um histograma pois não tem as barras adjacentes, terá de:

Clicar duas vezes sobre as barras, de forma a que apareça o menu *Format Data Series* ou *Format data Points*.; Seleccionar *Options* e em *Gap Width* seleccionar 0;OK:



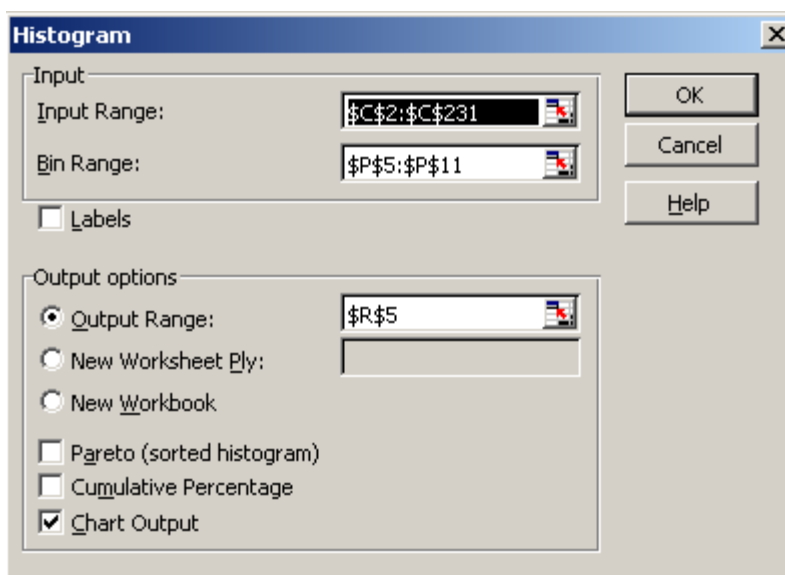
Finalmente pode-se melhorar esteticamente o histograma, diminuindo o número de casas decimais nos valores apresentados no eixo dos YY, retirando as linhas, etc.



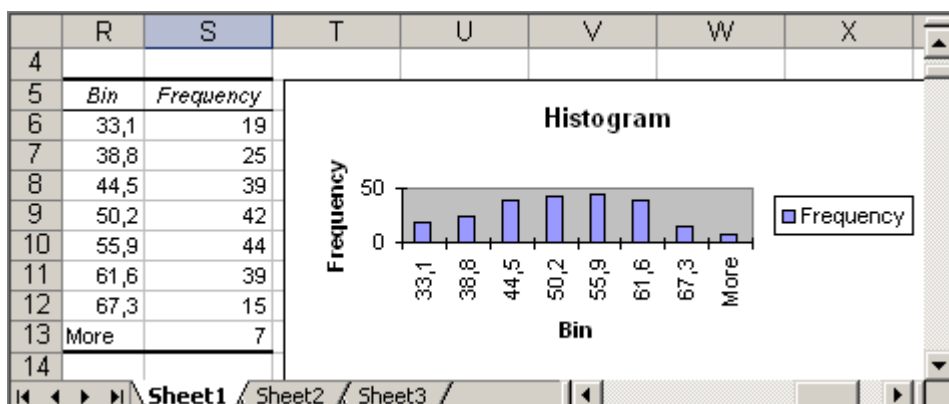
2.3.2.1.2 – Função *Histogram*

No Excel existe uma função, idêntica à função *Frequency*, a função *Histogram*, a que se acede seleccionando *Tools*→*Data*→*Analysis*→*Histogram*→*Ok*. Vamos exemplificar a sua utilização para o conjunto de dados da variável *Idade*, anteriormente considerado:

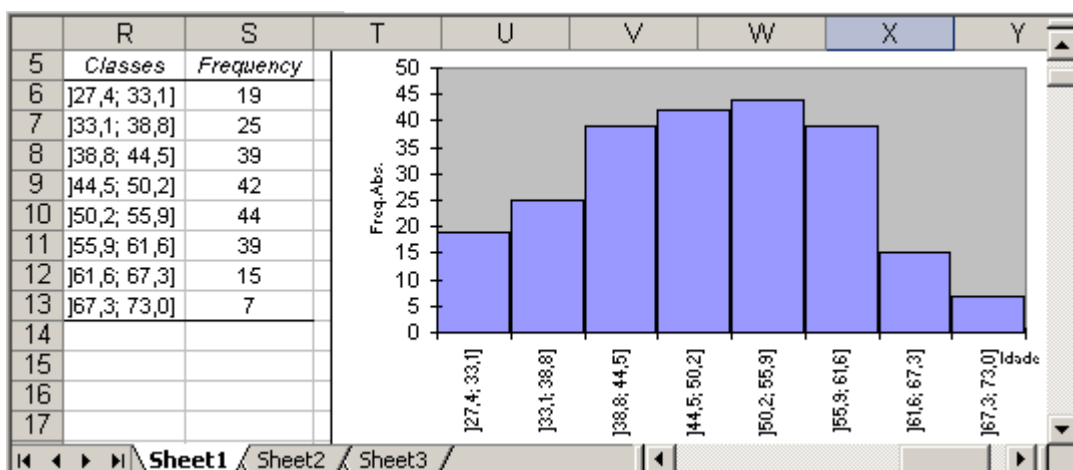
- Definir a coluna de separadores ou limites de classes, que constituirá o *Bin Range*: No nosso caso contruímos as classes subtraindo a amplitude de classe sucessivamente ao máximo, obtendo os valores {33,1, 38,8, 44,5, 50,2, 55,9, 61,6, 67,3} (tal como para a função *Frequency*, as classes são fechadas à direita e abertas à esquerda), que colocámos nas células P4:P10;
- Seleccionar *Tools*→*Data*→*Analysis*→*Histogram*→*Ok*:



- Em *Input Range*, indicámos o local dos dados e seleccionámos ainda a opção *Chart Output* e clicámos *OK*. Como resultado obtivemos o seguinte:



- Substituímos os limites das classes pelos intervalos das classes e arranjámos convenientemente o gráfico, já que a representação que se obtém, ao contrário do que é indicado no título, não é um histograma:



Nota: Ao considerar a função *Histogram*, tem a possibilidade de não indicar os separadores de classe, deixando vazio o espaço denominado *Bin Range*, uma vez que serão considerados, por defeito, classes. Contudo, não aconselhamos que se deixe esta escolha ao Excel, uma vez que, por exemplo, a primeira classe que é considerada, é constituída pelos valores menores ou iguais ao mínimo, o que não tem qualquer sentido.

2.3.3.1.3 - Tabela de frequências com as classes com amplitudes diferentes

Por vezes a organização e redução de um conjunto de dados contínuos, através de uma tabela de frequências, pressupõe que os intervalos, que constituem as classes, tenham limites escolhidos pelo utilizador, sem obedecerem a um critério estritamente resultante da aplicação de uma regra matemática. É o caso, por exemplo, da variável idade, em que poderá ser interessante escolher determinadas classes etárias.

Tendo em conta a definição de histograma, como sendo um diagrama de áreas, constituído por uma série de rectângulos adjacentes, em que a área de cada rectângulo é igual ou proporcional à frequência de classe, no caso de a tabela de frequências não apresentar as classes todas com a mesma amplitude, já o histograma não se pode reduzir a um diagrama de barras, em que as barras tenham a mesma amplitude e as alturas sejam iguais às frequências.

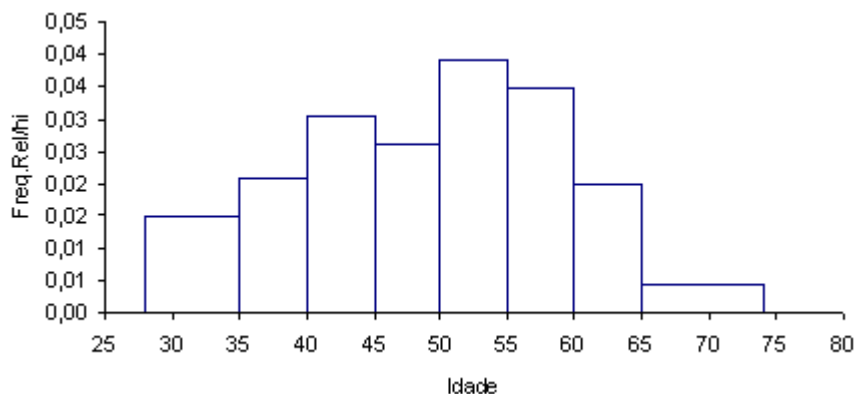
Não sendo o Excel um software de Estatística, não apresenta uma solução imediata para a construção do histograma nestas condições, sendo necessário recorrer a um artifício. Exemplificaremos a seguir a aplicação de uma técnica possível para a resolução do problema, recorrendo à representação gráfica *Scatter*.

Exemplo 2.3.4 – Consideremos ainda a variável idade dos deputados. Organize os dados segundo uma tabela de frequências, considerando as seguintes classes [28, 35[, [35, 40[, [40, 45[, [45, 50[, [50, 55[, [55, 65[, [65, 75[, [75, 78].

A construção da tabela de frequências pode ser feita utilizando a função *Frequency*, como vimos na secção anterior. No entanto, vai ser necessário acrescentar uma nova coluna onde, para cada classe, se considera a frequência relativa (ou absoluta) a dividir pela amplitude de classe. Será esta coluna que irá fornecer as alturas dos rectângulos que constituirão o histograma. Com esta precaução, garantimos que as áreas destes rectângulos são iguais às frequências relativas (ou absolutas). Apresenta-se a seguir a tabela de frequências obtida, segundo a descrição anterior:

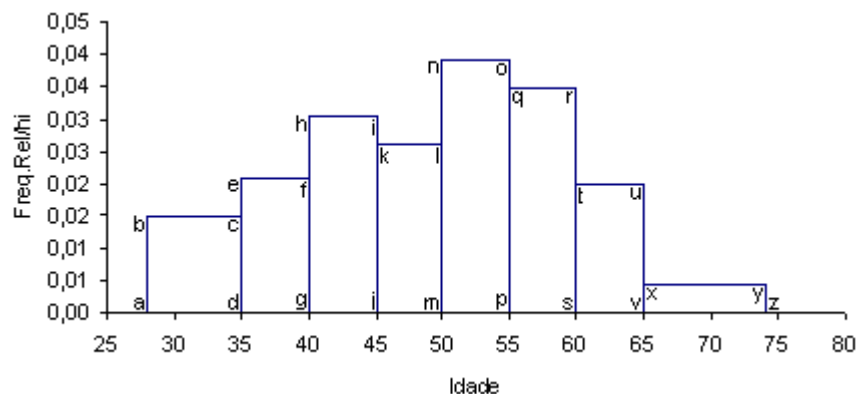
	K	L	M	N	O
2	Tabela de frequências				
3	Classes	Amplitude hi	Freq.Rel/hi	Freq. Abs.	Freq. Rel.
4	[28; 35[7	0,0149	24	0,104
5	[35; 40[5	0,0209	24	0,104
6	[40; 45[5	0,0304	35	0,152
7	[45; 50[5	0,0261	30	0,130
8	[50; 55[5	0,0391	45	0,196
9	[55; 60[5	0,0348	40	0,174
10	[60; 65[5	0,0200	23	0,100
11	[65; 74[9	0,0043	9	0,039
12				230	1,000

O histograma correspondente a esta tabela de frequências, com cuja construção não nos vamos preocupar para já, terá o seguinte aspecto:

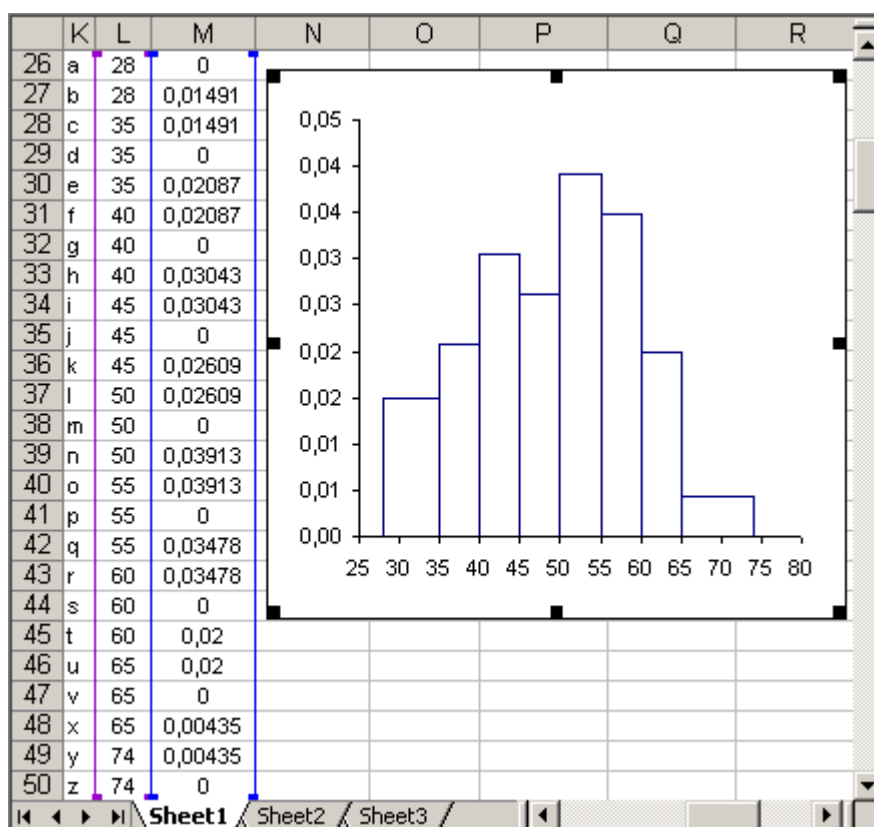


Temos um histograma correctamente construído, em que as áreas dos rectângulos são iguais às frequências relativas, ocupando o histograma uma área total igual a 1.

Na figura anterior, vamos marcar alguns pontos com letras:



Repare que se unir o ponto **a** com **b**, de seguida com **c**, até esgotar todos os pontos, obtém o histograma. Então, para obter a representação gráfica desejada, basta construir uma tabela, numa folha de Excel, com as coordenadas dos pontos que pretendemos unir e utilizar a representação *Scatter*, tal como foi feito para representar a função cumulativa em 3.1.2.2:

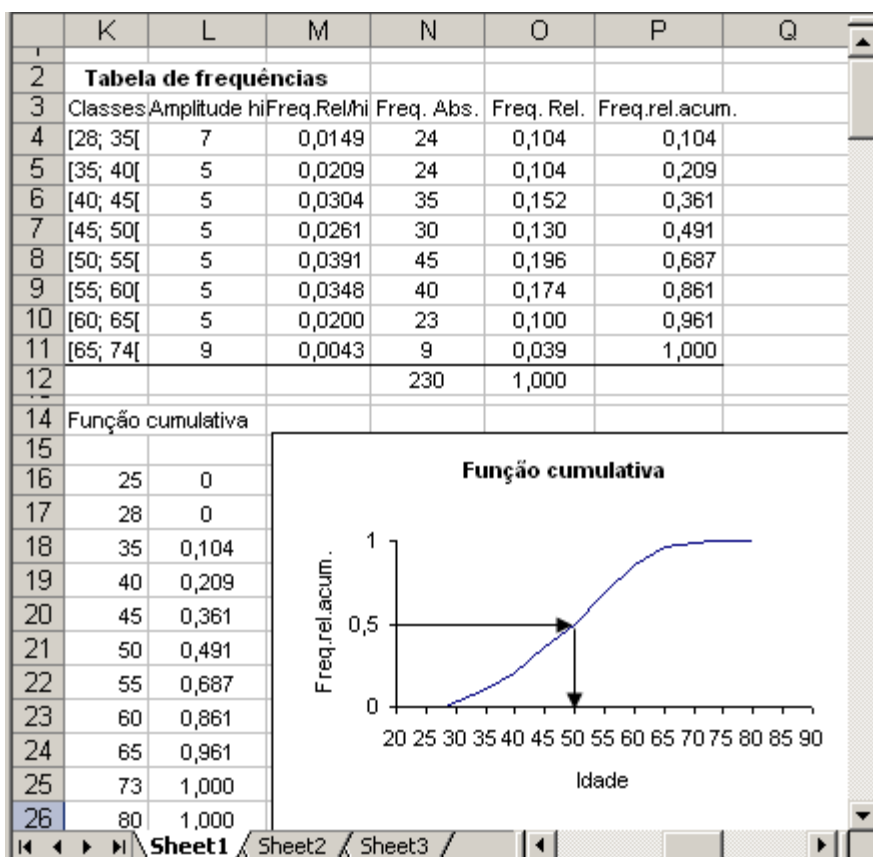


2.3.2.2 – Função cumulativa

Para representar graficamente as frequências acumuladas, considera-se a função cumulativa, que se obtém utilizando a seguinte metodologia:

- Antes do limite inferior da 1ª classe, l_1 , a frequência acumulada é nula, pelo que se traça um segmento sobre o eixo dos xx, até esse ponto;
- No limite inferior da 2ª classe, l_2 , a frequência acumulada é a frequência da classe anterior, f_1 . Admitindo que a frequência se distribui uniformemente no intervalo de classe, unimos os pontos de coordenadas $(l_1, 0)$ e (l_2, f_1) ;
- No limite inferior da 3ª classe, l_3 , a frequência acumulada é a soma das frequências das duas classes anteriores, (f_1+f_2) . Então unimos os pontos de coordenadas (l_2, f_1) e $(l_3, (f_1+f_2))$;
- Quando chegarmos à última classe, temos a garantia que a frequência acumulada, correspondente ao seu limite superior, é igual a 1, pelo que nesse ponto marcamos 1 e continuamos com um segmento de recta paralelo ao eixo dos xx.

Exemplo 2.3.4 (continuação) – Construa a função cumulativa, a partir da tabela de frequências apresentada no exemplo 2.3.4. Para obter a função cumulativa, basta acrescentar à tabela de frequências uma nova coluna com as frequências relativas acumuladas. De seguida utiliza-se a representação *Scatter*, para unir os pontos, tais como foram definidos nas indicações dadas, anteriormente, para a construção da função cumulativa:

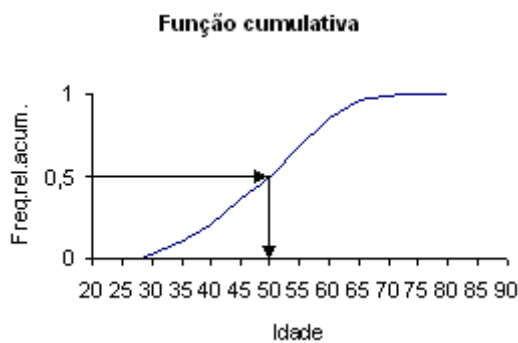


Da maneira como foi construída, a função cumulativa tem algumas propriedades importantes, nomeadamente:

- Está definida para todo o x real (na representação gráfica anterior escolhemos arbitrariamente o valor da abcissa igual a 25 para começar a construir a função cumulativa);
- É sempre não decrescente;
- Só assume valores no intervalo $[0, 1]$;
- Permite obter informação sobre qual o valor da abcissa a que corresponde determinada frequência acumulada.

Vamos explorar um pouco mais esta última propriedade.

Suponhamos que se pretendia saber, a partir da representação gráfica da função cumulativa, obtida para o exemplo anterior, qual o valor aproximado para a idade a que corresponde uma frequência relativa acumulada de 50%. De acordo com a figura, este valor deve estar na classe $[50, 55[$.



Uma vez que se admite que a frequência se distribui uniformemente sobre a amplitude de classe, isto é a frequência $0,196 (=0,687-0,491)$ distribui-se uniformemente sobre o intervalo de amplitude 5, através da resolução de uma equação de proporcionalidade, obtém-se o valor que andávamos à procura:

$$\frac{0,196}{0,009} = \frac{5}{x} \quad x = \frac{0,009 \times 5}{0,196} = 0,22$$

onde $0,009=0,5-0,491$. Então o valor pretendido é $50 + 0,22 = 50,22$ anos, ou seja 50 anos.

Ao valor obtido anteriormente, a que corresponde uma frequência acumulada de 50%, chamamos *mediana*. A mediana, que já foi objecto de estudo no módulo de Estatística, divide a distribuição das frequências em duas partes iguais. Recordamos que a técnica utilizada permitiu-nos obter um valor aproximado para a mediana, cujo valor exacto só poderia ter sido

determinado a partir dos dados originais, antes de proceder ao agrupamento. Aliás, veremos mais à frente a determinação desta e de outras medidas, utilizando o Excel.

Se em vez de pretendermos determinar o valor a que corresponde a percentagem de 50%, procurássemos os valores a que correspondem as percentagens de 25% ou 75%, obteríamos os chamados quartis, respectivamente 1.º e 3.º quartil, e a metodologia utilizada para os determinar a partir da função cumulativa seria idêntica à utilizada para determinar a mediana.

2.3.3 – Outras representações gráficas

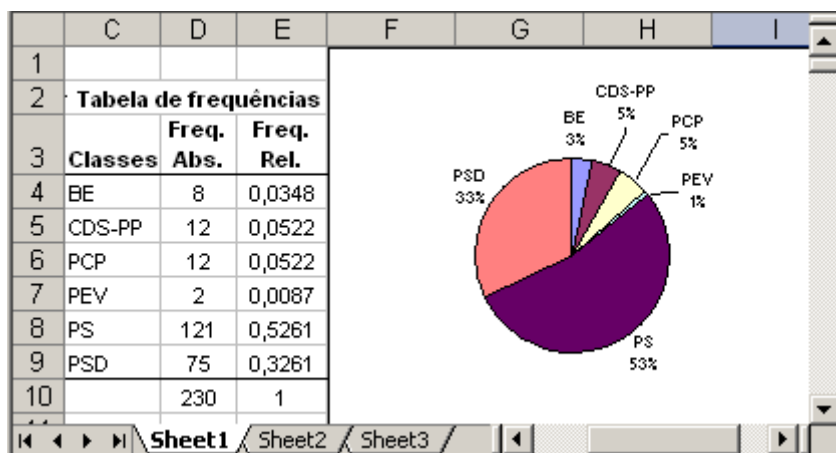
Além das representações gráficas consideradas anteriormente, em que destacamos o diagrama de barras para dados discretos e o histograma para dados contínuos, existem ainda outras representações que podem ser utilizadas para dados qualitativos ou quantitativos – diagrama circular, ou dados quantitativos – caule-e-folhas e diagrama de extremos e quartis. Todas estas representações já foram objecto de estudo no módulo de Estatística, pelo que privilegiaremos aqui a forma de os construir utilizando o Excel.

2.3.3.1 – Diagrama circular

Esta representação, utilizada essencialmente para dados qualitativos, é constituída por um círculo, em que se apresentam vários sectores circulares, tantos quantas as classes consideradas na tabela de frequências da amostra em estudo. Os ângulos dos sectores são proporcionais às frequências das classes. A representação deste diagrama, em Excel, é imediata, apresentando várias modalidades.

Exemplo 2.3.5 – Apresente sob a forma de um diagrama circular a distribuição dos deputados do ficheiro *Deputados.xls* segundo o grupo parlamentar.

Esta variável já foi objecto de estudo num exemplo anterior, de forma que recorreremos à tabela de frequências já calculada, para obter a representação gráfica pretendida. Seleccionam-se as células com as classes e as respectivas frequências absolutas ou relativas e no menu *Chart* seleccionassem *Pie*, a modalidade desejada:



2.3.3.2 – Caule-e-folha

Esta representação, como se sabe, é uma representação que se pode considerar entre a tabela e o gráfico, uma vez que são apresentados os verdadeiros valores da amostra, mas de forma sugestiva, que faz lembrar um histograma. Antes de abordarmos a forma de construir um caule-e-folhas utilizando o Excel, vamos apresentar um exemplo, que nos poderá ajudar a compreender os passos necessários para essa construção.

Exemplo 2.3.6 – Consideremos a seguinte amostra constituída pela idade de 30 deputados, escolhidos aleatoriamente da tabela de deputados do ficheiro *Deputados.xls*:

63 59 31 51 51 61 42 65 48 63 57 43 54 42 52
 51 57 34 38 44 61 60 56 66 63 52 47 33 46 52

Uma representação possível em caule-e-folhas é a que se apresenta a seguir:

```

3 | 1 3 4 8
4 | 2 2 3 4 6 7 8
5 | 1 1 1 2 2 2 4 6 7 7 9
6 | 0 1 1 3 3 3 5 6
    
```

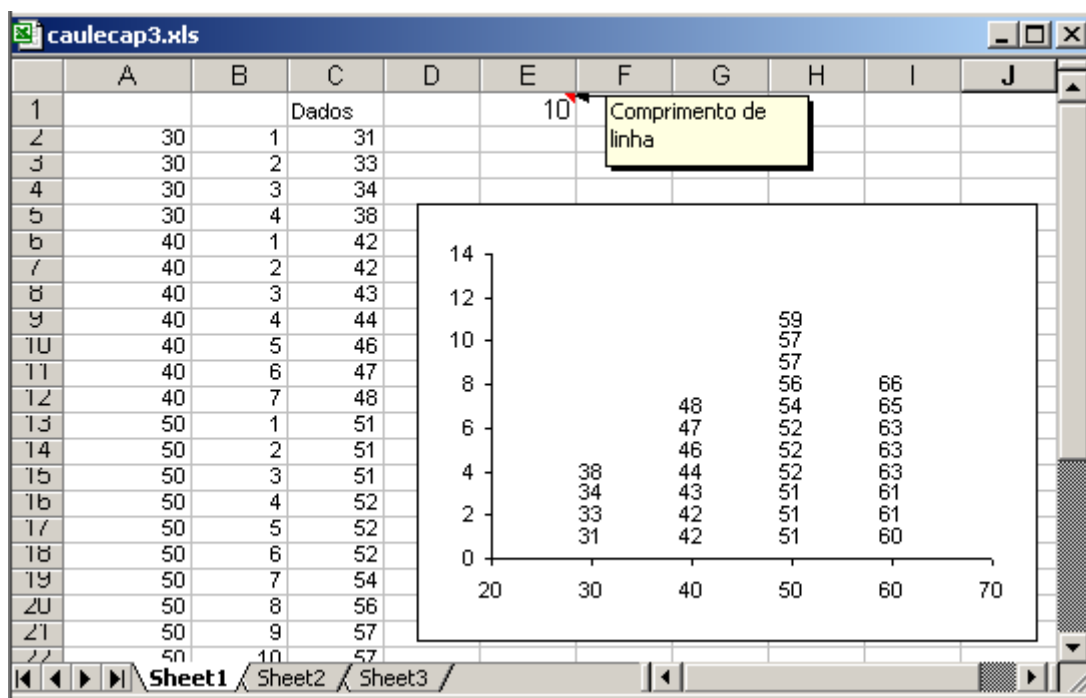
Nesta representação considerámos 4 caules e o intervalo entre caules sucessivos é de 10 unidades. No caule 3 pendurámos todas as folhas deste caule e o mesmo foi feito com todos os outros caules. É como se tivéssemos considerado as classes [30, 40[, [40, 50[, [50, 60[e [60, 70[para agrupar os dados. Suponhamos que em vez de considerar estas classes, de amplitude 10, estávamos interessados em considerar classes de amplitude 5, a saber [30, 35[, [35, 40[, [40, 45[, [45, 50[, [50, 55[, [55, 60[, [60, 65[e [65, 70[. Então a representação anterior teria o seguinte aspecto:

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

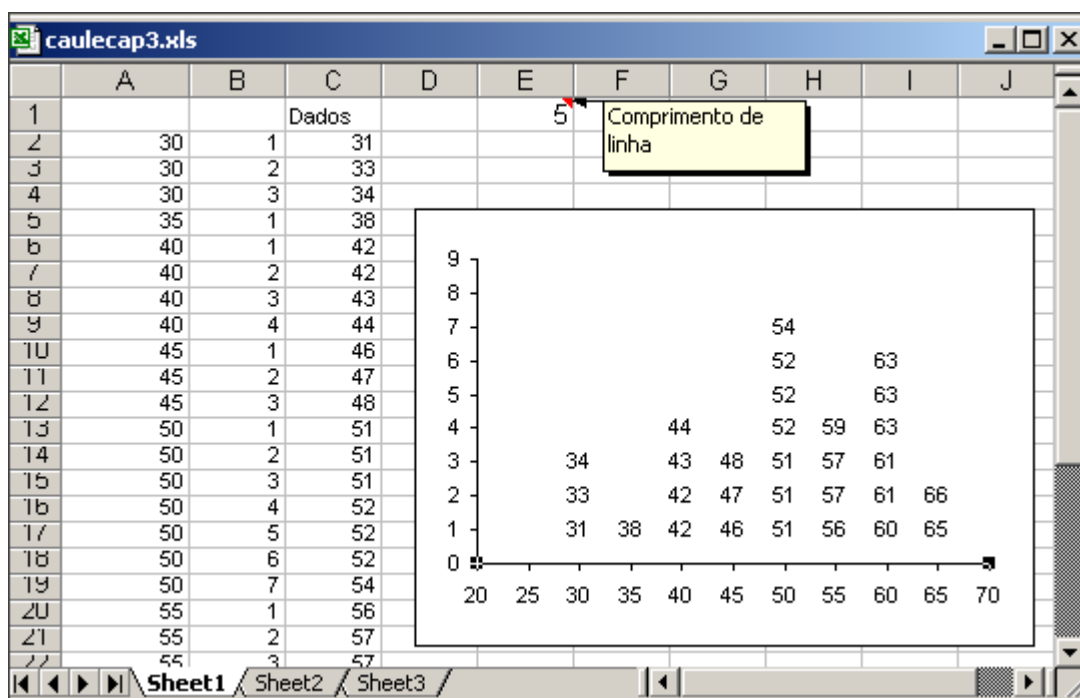
Qualquer que seja a representação considerada, qualquer caule tem sempre a possibilidade de ter penduradas o mesmo número de folhas. No exemplo anterior, no primeiro sub caule 3 (ou 4, ou 5, ou 6) aparecem penduradas as folhas 0, 1, 2, 3 e 4, enquanto que no segundo sub caule 3 (ou 4, ou 5, ou 6) aparecem penduradas as folhas 5, 6, 7, 8 e 9). Uma outra possibilidade seria considerar classes de amplitude 2, fazendo cada caule dividido em 5 sub caules e cabendo a cada sub caule 2 folhas (repare-se com a analogia com a construção do histograma, em que considerámos as classes com igual amplitude). A esta amplitude de classe é usual chamar *comprimento de linha*.

Não existe no Excel uma representação imediata para a construção de um caule-e-folhas, pelo que vamos utilizar um processo desenvolvido por Neville Hunt (Hunt, 2001), para o Excel:

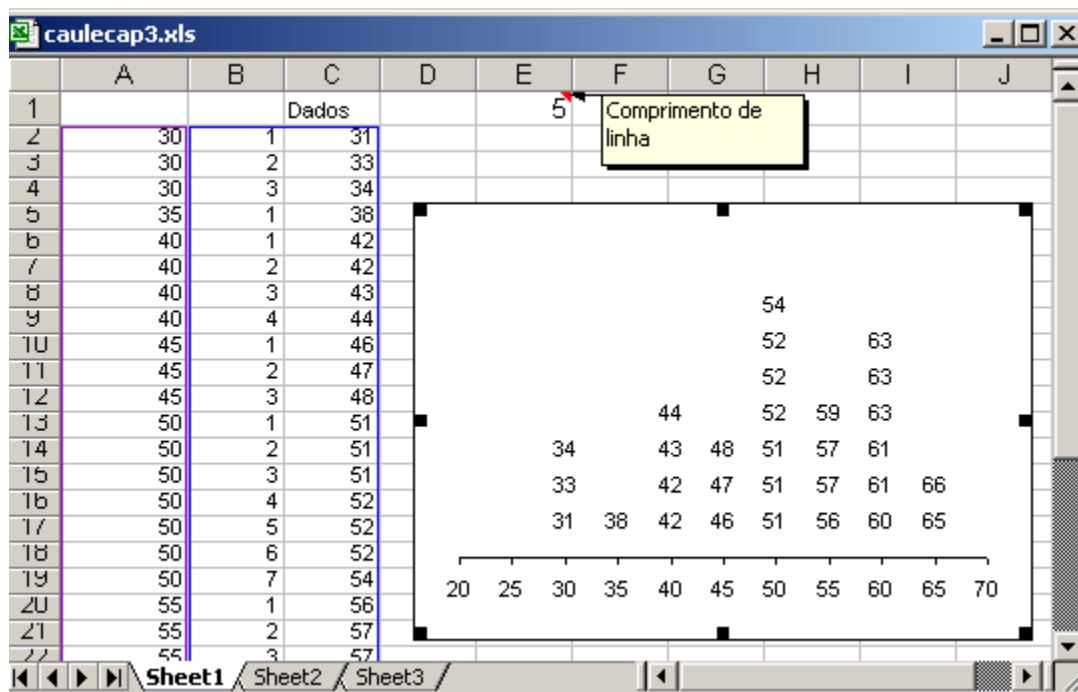
- 1º passo – Insira os dados na coluna C, começando na célula C2; se não estiverem ordenados, ordene-os por ordem crescente;
- 2º passo – Insira na célula E1 o valor que deseja para o comprimento de linha: 10, 5 ou 2 ou uma potência de 10, destes valores;
- 3º passo – Na célula A2 escreva a seguinte fórmula = $INT(C2/E\$1)*E\1 e replique-a tantas vezes quantos os dados inseridos no 1º passo, na coluna C;
- 4º passo – Na célula B2 escreva o valor 1. Na célula B3 escreva a fórmula = $IF(A3=A2; B2+1; 1)$ e replique a fórmula, tantas vezes quantos os dados inseridos no 1º passo, na coluna C;
- 5º passo – Selecciona as células das colunas A, B e C com os resultados obtidos nos passos anteriores e no módulo *Chart Wizard* (Assistente de Gráficos) escolha *Bubble*;
- 6º passo – Faça um duplo clique numa das bolas representadas e na janela *Format data Series* (ou clique com o botão direito do rato e seleccione *Format data Series*) seleccione *Patterns*:
Border: None
Area: None
Data Labels: Show bubbles sizes
 OK;
- 7º passo – Faça um duplo clique numa das “*Data labels*” (ou clique com o botão direito do rato e seleccione *Format Data Labels*), e na janela *Format Data Labels*, em *Alignment*:
Label Position: Centre
 OK;
- 8º passo – Clique numa das linhas horizontais que atravessam o gráfico e apague-as com a tecla Delete. Faça o mesmo ao fundo cinzento, seleccionando-o e carregando na tecla Delete. Apague também a legenda.
- 9º passo – Formate convenientemente os eixos.



Na folha de Excel, se mudarmos o valor do comprimento de linha para 5, aparece de imediato a seguinte representação (aparte uma formatação adequada do eixo dos xx):



Repare-se que, embora as notações usadas para os caules e as folhas não sejam idênticos aos da representação inicialmente considerada, feita sem o recurso ao Excel, o aspecto gráfico é o mesmo. Para uma maior semelhança, seleccionámos o eixo dos yy e fizemos *Delete*:



2.3.3.3 – Diagrama de extremos e quartis

Esta representação, muito simples, mas bastante elucidativa ao realçar a informação contida nos dados, no que diz respeito à simetria e variabilidade, pressupõe que se calculem algumas estatísticas necessárias para a sua construção.

Mais uma vez estamos perante uma representação gráfica cuja construção, por meio do Excel, necessita de alguns “truques”. Assim, o primeiro passo para uma dessas construções, consiste em representar, adequadamente, numa folha de Excel, as estatísticas Mínimo, Máximo, 1.º e 3.º quartis e mediana.

Exemplo 2.3.7 – Construa um diagrama de extremos e quartis para a variável idade dos deputados do ficheiro Deputados.xls.

Construção do diagrama de extremos e quartis, em Excel:

1. Utilizando o Excel, começam por se calcular as estatísticas necessárias¹, que se apresentam da seguinte forma:

¹ No capítulo 3 abordaremos a determinação das estatísticas descritivas, utilizando o Excel.

	E	F
1		
2	1º quartil	=QUARTILE(\$C\$2:\$C\$231;1)
3	Mínimo	=MIN(\$C\$2:\$C\$231)
4	Mediana	=MEDIAN(\$C\$2:\$C\$231)
5	Máximo	=MAX(\$C\$2:\$C\$231)
6	3º quartil	=QUARTILE(\$C\$2:\$C\$231;3)

	C	D	E	F
1	Idade			
2	53	1º quartil	41	
3	32	Mínimo	28	
4	61	Mediana	50	
5	51	Máximo	73	
6	48	3º quartil	56	

2. Seleccionar as células que contêm as estatísticas, assim como as suas etiquetas: E2 a F6;

3. No módulo Chart Wizard (Assistente de Gráficos) seleccionar:

Line

Seleccionar *Line with markers displayed at each data value*

Clicar *Next*

Seleccionar *Series in Rows*

Clicar *Finish*

4. Clicar com o botão direito do rato num dos pontos. Seleccionar:

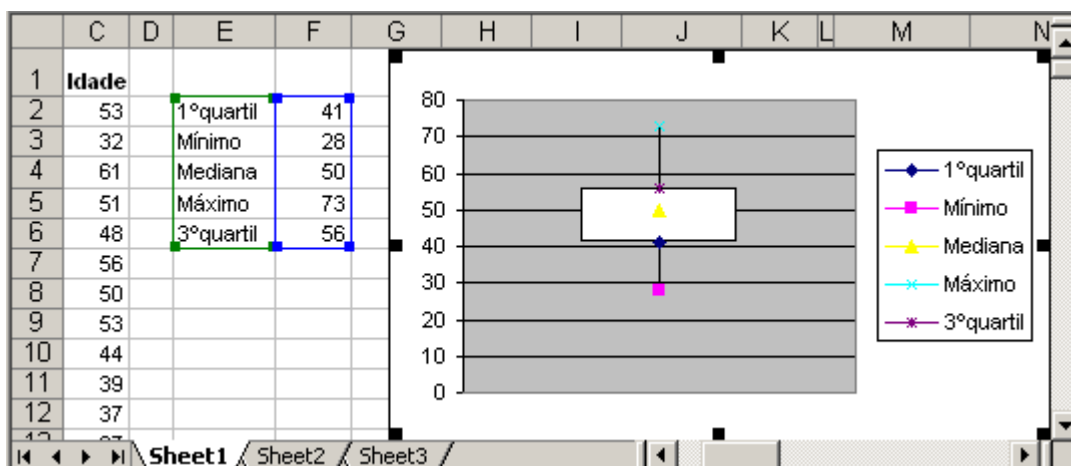
Format Data Series

Seleccionar *Options*

Escolher *High-low lines* e *Up-down bars*; Ajuste à sua escolha *Gap width*;

OK

5. Arranjar “esteticamente” o gráfico:



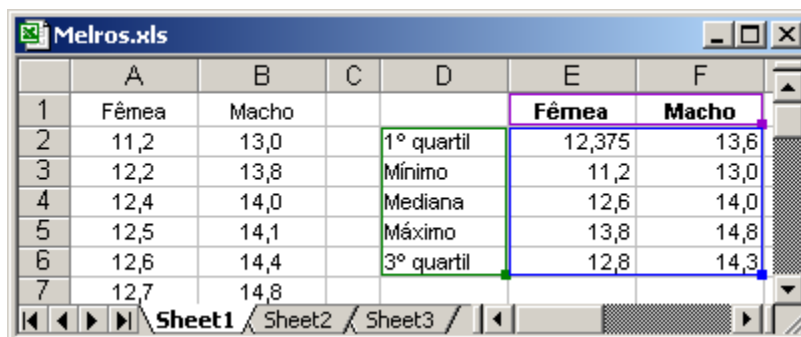
Esta representação de um conjunto de dados, num diagrama de extremos e quartis, é especialmente indicada para comparação de várias amostras, como se exemplifica a seguir:

Exemplo 2.3.8 – Registou-se o comprimento, em centímetros, das asas de 32 melros-fêmeas e 25 melros-macho, tendo-se obtido os seguintes resultados:

Melro-fêmea -	11,2	11,7	12,0	12,1	12,2	12,2	12,3	12,3	12,4	12,4	12,4	12,4
	12,5	12,5	12,5	12,5	12,6	12,6	12,7	12,7	12,7	12,8	12,8	12,8
	12,8	13,0	13,1	13,1	13,2	13,5	13,6	13,8				
Melro-macho -	13,0	13,4	13,5	13,5	13,5	13,6	13,6	13,7	13,8	13,8	13,8	13,9
	14,0	14,0	14,1	14,1	14,1	14,2	14,3	14,3	14,4	14,4	14,4	14,4
												14,8

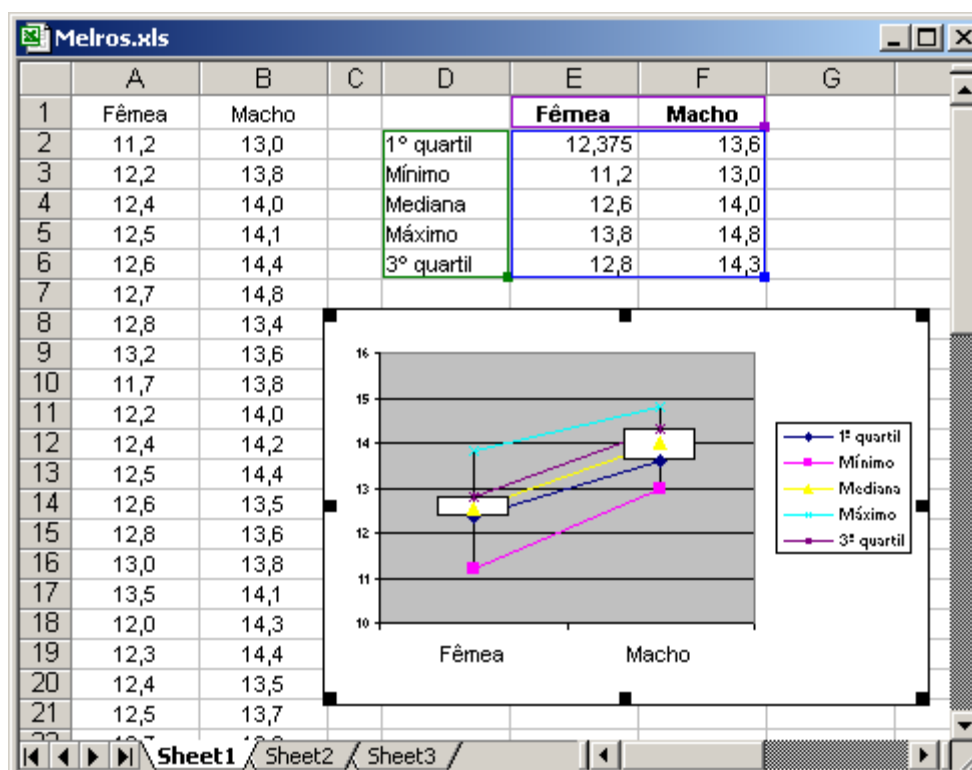
Utilizando uma representação adequada, compare os dois conjuntos de dados.

Começamos por introduzir os dados numa folha de Excel, calculando de seguida as características amostrais relevantes para a construção de um diagrama de extremos e quartis:



	A	B	C	D	E	F
1	Fêmea	Macho			Fêmea	Macho
2	11,2	13,0		1º quartil	12,375	13,6
3	12,2	13,8		Mínimo	11,2	13,0
4	12,4	14,0		Mediana	12,6	14,0
5	12,5	14,1		Máximo	13,8	14,8
6	12,6	14,4		3º quartil	12,8	14,3
7	12,7	14,8				

Para proceder à construção do diagrama de extremos e quartis comece por seleccionar as células que contêm os valores das características amostrais, assim como as etiquetas (células D1 a F6), e proceda de acordo com as instruções dadas no exemplo anterior. Depois de formatar convenientemente o eixo dos yy, obterá a seguinte representação:



As linhas a unir as caixas podem ser removidas, seleccionando cada uma, com o botão direito do rato e seleccionando sucessivamente:

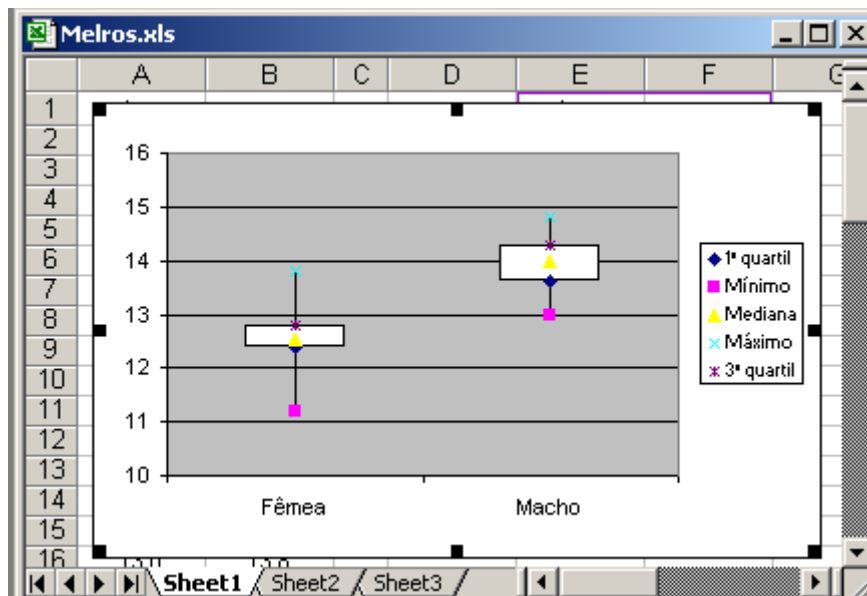
Format Data Series

Patterns

Line

None

Ok



O gráfico anterior é bastante elucidativo na medida em que mostra que o tamanho das asas do melro-macho é, de um modo geral superior ao do melro-fêmea, apresentando ainda uma maior variabilidade.

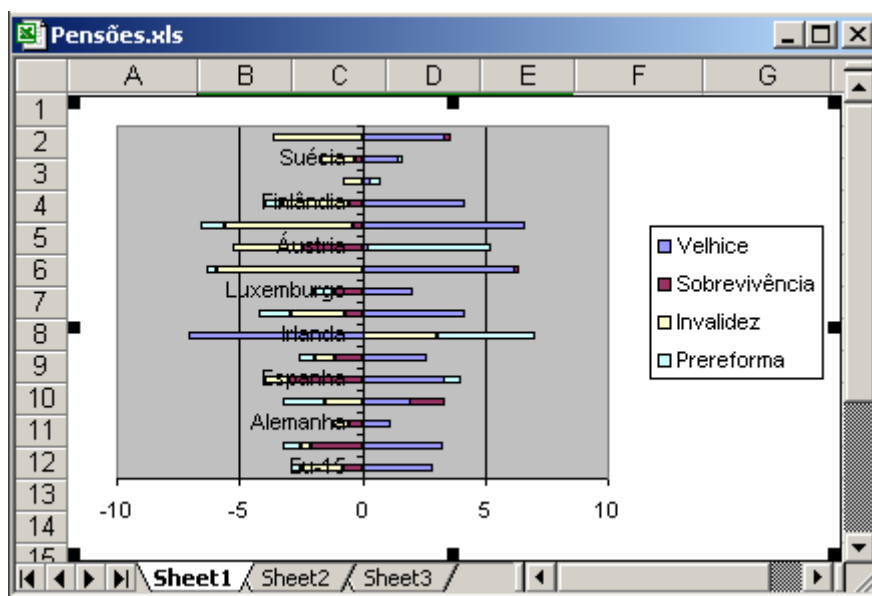
2.4 – Alguns exemplos

A seguir apresentamos alguns exemplos, sobre a forma de projectos, para os quais podemos utilizar vários tipos de representações gráficas, algumas já referidas anteriormente, outras introduzidas pela primeira vez, mas que apresentam realização imediata com o Excel.

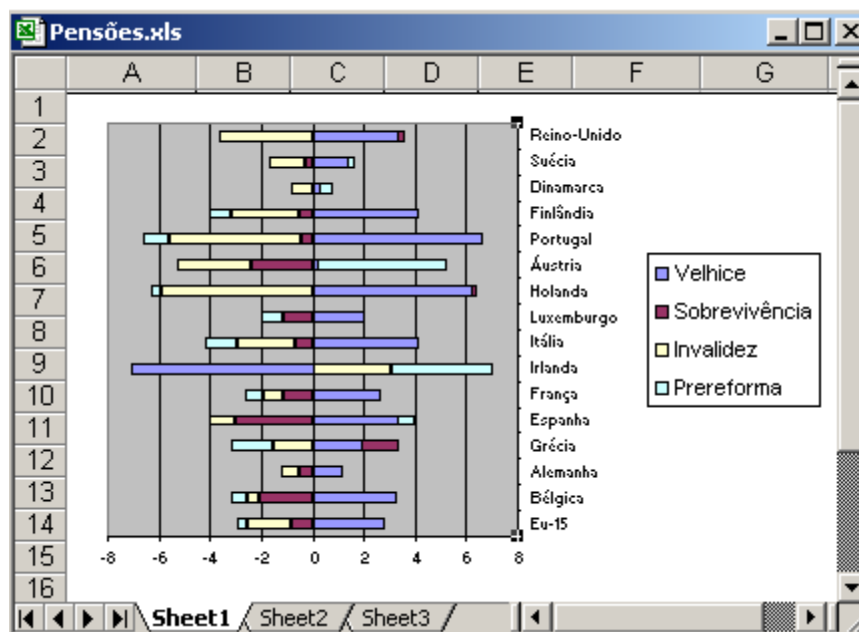
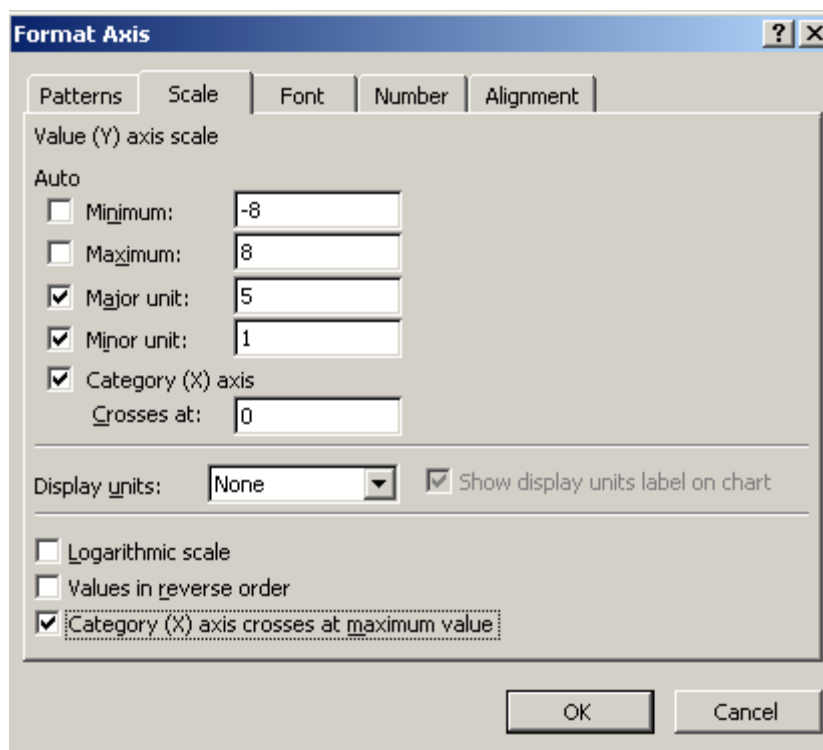
Projecto 1 - Neste projecto são apresentados alguns dados relativamente à Modificação da Estrutura das Categorias de Pensões entre 1993 e 2001 (em pontos percentuais) (Eurostat – Statistiques en bref – Population et conditions sociales, 8/2004):

	Velhice	Sobrevivência	Invalidez	Prereforma
Eu-15	2,8	-0,8	-1,7	-0,4
Bélgica	3,2	-2,1	-0,4	-0,7
Alemanha	1,1	-0,5	-0,7	0,1
Grécia	1,9	1,5	-1,5	-1,7
Espanha	3,3	-3	-1	0,7
França	2,6	-1,1	-0,8	-0,7
Irlanda	-7,1	0	3,1	4
Itália	4,1	-0,7	-2,2	-1,3
Luxemburgo	2	-1,1	0	-0,9
Holanda	6,2	0,2	-5,9	-0,4
Áustria	0,2	-2,4	-2,9	5,1
Portugal	6,6	-0,4	-5,2	-1
Finlândia	4,1	-0,5	-2,7	-0,8
Dinamarca	0,3	0	-0,8	0,5
Suécia	1,4	-0,3	-1,4	0,3
Reino-Unido	3,3	0,3	-3,6	0

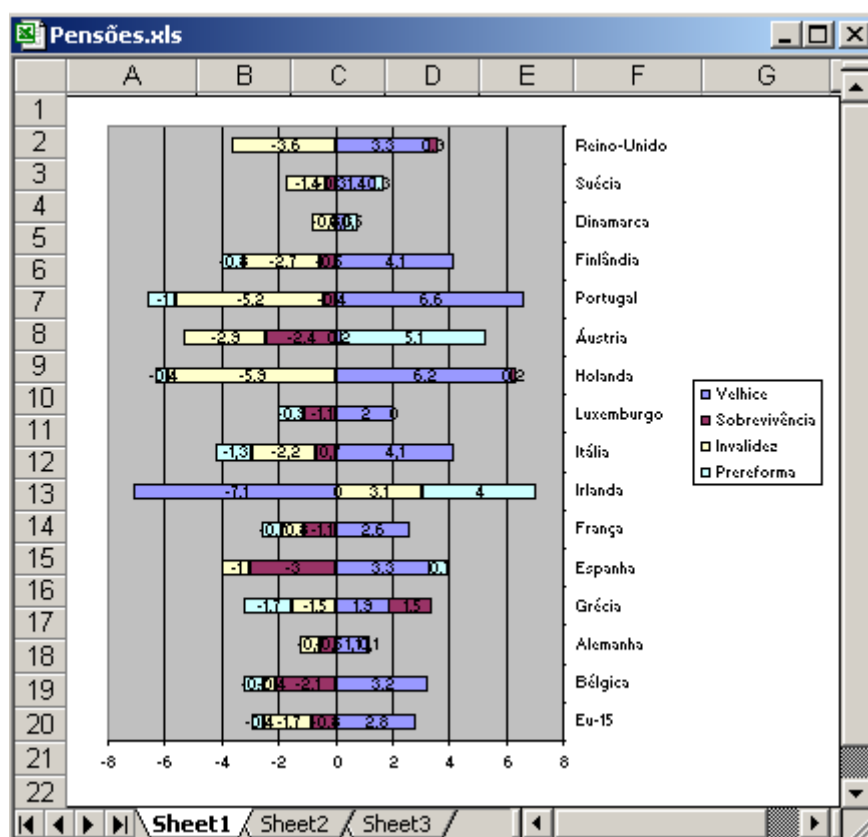
Uma forma adequada para representar estes dados, é através de um diagrama de barras, nomeadamente barras horizontais, seleccionando na opção *Chart* o 2º tipo da opção *Bar*.



Vamos fazer alguma “cosmética” na representação gráfica anterior, nomeadamente mudando a escala para -8 a 8 e fazendo com que as legendas não se sobreponham ao gráfico:



Podemos ainda acrescentar sobre o gráfico os valores quando houver conveniência em disponibilizar esta informação. Para isso basta seleccionar *Chart Options* → *Data labels* → *Show Value*:




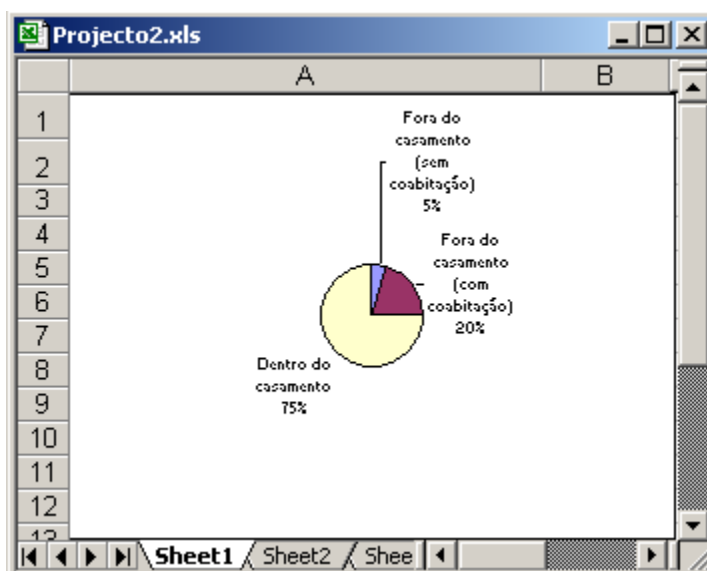
Projecto 2 – Entre os dois últimos recenseamentos da população portuguesa, os Censos 91 e os Censos 2001, realizados, respectivamente, em 15 de Abril de 1991 e 12 de Março de 2001, verificou-se que a população residente no território nacional passou de 9.867.147 para 10.356.117 habitantes, a que corresponde um acréscimo de 4.8%. Na generalidade das regiões verificou-se um aumento da população, com excepção das regiões do Alentejo e Madeira. Partindo dos resultados censitários definitivos, estimou-se a população residente em 31 de Dezembro de 2002 em 10.407.500 indivíduos, dos quais 5.030.200 do sexo masculino.

Apresentam-se a seguir algumas tabelas e gráficos com alguns indicadores (www.ine.pt):

1. Nados-vivos segundo a filiação – 2002


	A	B
1	Fora do casamento (sem coabitação)	5,10%
2	Fora do casamento (com coabitação)	20,40%
3	Dentro do casamento	74,50%

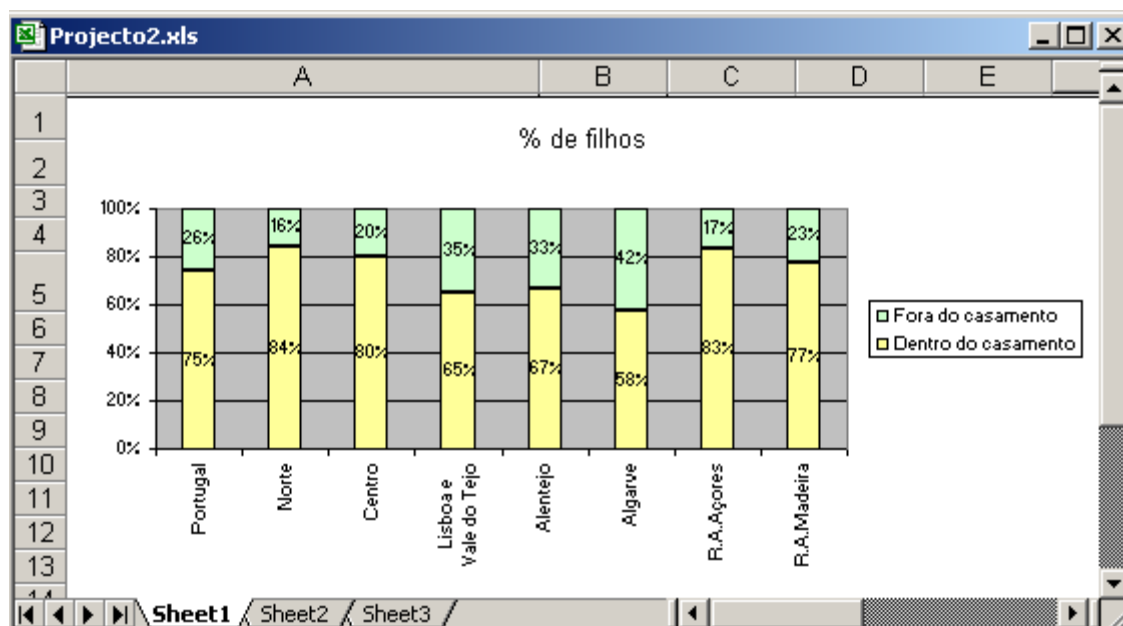
Uma representação adequada para a tabela anterior é o diagrama circular. Assim, vamos seleccionar **Chart**  → **Pie** → 1º **subtipo** → **Next** → **Next** → **Data labels** → **Show label and percent** → **Finish**:



Nados-vivos segundo a filiação, por regiões:

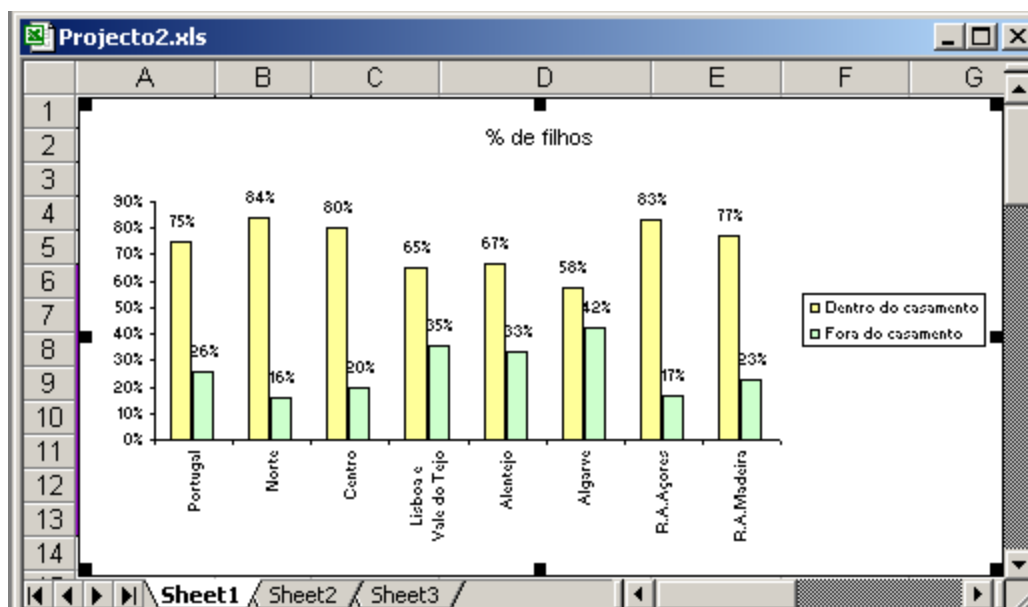
Região	Porcentagem
Dentro do casamento	
Portugal	74,50%
Norte	83,80%
Centro	80,00%
Lisboa e Vale do Tejo	64,70%
Alentejo	66,70%
Algarve	57,60%
R.A. Açores	83,10%
R.A. Madeira	77,30%

Acrescentámos à tabela anterior uma outra coluna – células C6 a C13, com os filhos fora do casamento e decidimos aqui optar por uma representação em barras verticais. Assim, depois de seleccionar as células A5 a C13, fizemos **Chart**  → **Column** → 3º **subtipo** → **Next** → **Next** → **Data labels** → **Show value** → **Titles** → **Chart title** → **% de filhos** → **Finish**:



Observação: Foi possível optarmos pela representação gráfica anterior, uma vez que os dados das duas características em estudo somavam 100%.

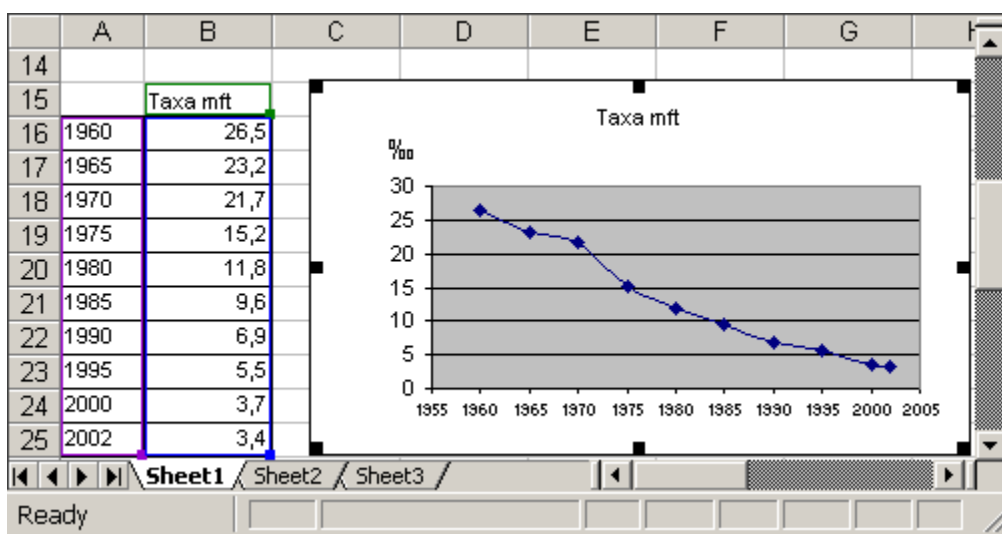
Outra representação possível obtém-se seleccionando *Chart* → *Column* → *1º subtipo* → *Next* → *Next* → *Data labels* → *Show value* → *Titles* → *Chart title* → *% de filhos* → *Finish*:



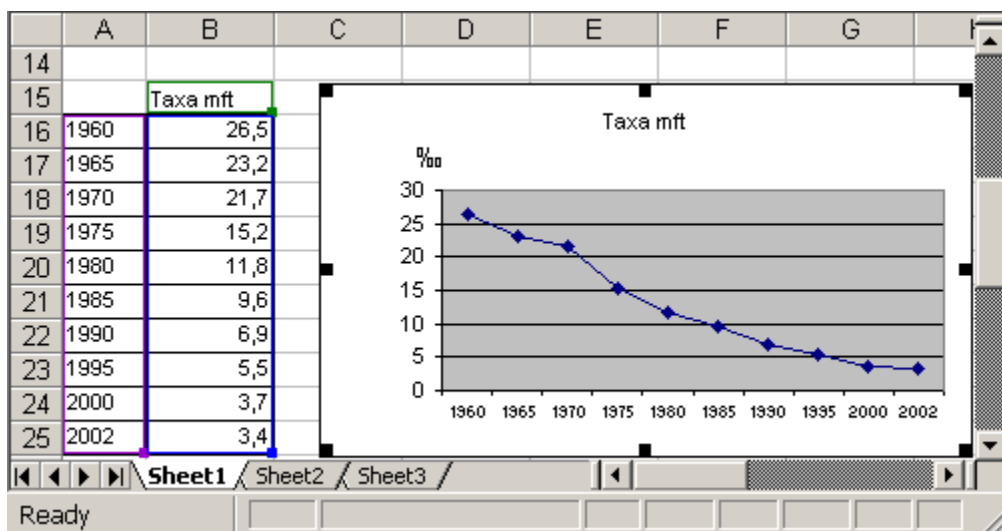
2. Taxa de mortalidade fetal tardia (Taxa mft) (28 ou mais semanas de gestação):

1960	26.5‰
1965	23.2‰
1970	21.7‰
1975	15.2‰
1980	11.8‰
1985	9.6‰
1990	6.9‰
1995	5.5‰
2000	3.7‰
2002	3.4‰

Introduzimos a tabela anterior numa folha de Excel e antes de procedermos a uma representação gráfica passámos os pontos para vírgulas e retirámos a permilagem, não reconhecida no Excel. Seguidamente depois de seleccionar as células A15 a B25, seleccionámos *Chart* → *XY(Scatter)* → *2ºsubtipo* → *Next* → *Next* → *Legend:Retirar a selecção de Show Legend* → *Titles* → *‰ em Value(Y)* → *Finish*:



Chamamos a atenção para o facto de ser possível obter uma representação aparentemente semelhante à anterior utilizando a opção *Chart* → *Line* → *4ºsubtipo* → *Next* → *Next* → *Legend* → *Retirar a selecção de Show Legend* → *Titles* → *‰ em Value(Y)* → *Finish*:



Repare-se, no entanto, que a representação anterior não está correcta, pois a variável tempo do eixo dos xx está a ser interpretada como uma variável qualitativa e não quantitativa como deveria ser. Assim, o intervalo entre 1995 e 2000 é igual ao intervalo entre 2000 e 2002, o que obviamente não está correcto.

3. Taxa de mortalidade infantil

1960	77.5‰
1965	64.9‰
1970	58.0‰
1975	38.9‰
1980	24.3‰
1985	17.8‰
1990	10.9‰
1995	7.5‰
2000	5.5‰
2002	5.0‰

A representação gráfica dos dados desta tabela pode ser idêntica à do ponto anterior.

4. Casamentos segundo a forma de celebração

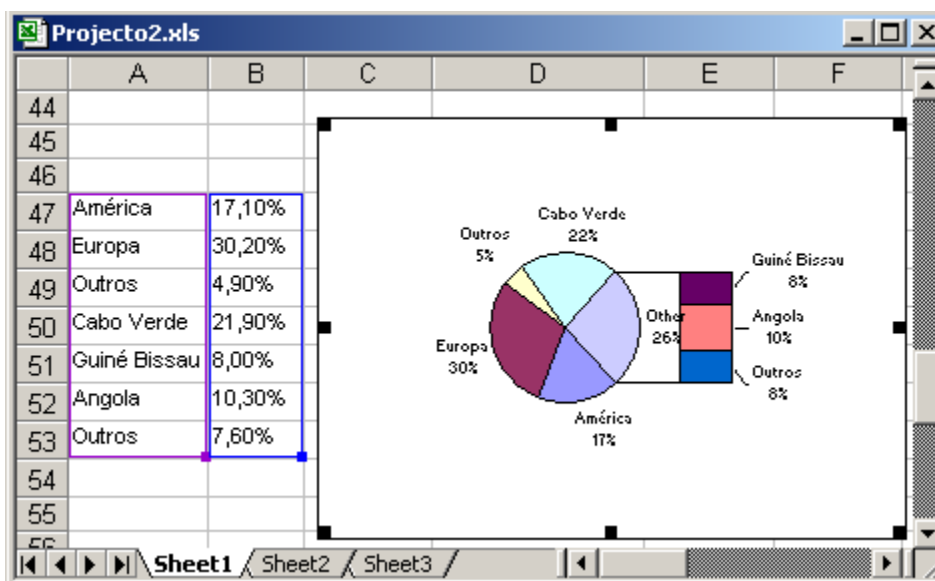
Unidade %	Civil	Católico
1960	9.2	90.8
1965	11.8	88.2
1970	13.4	86.6
1975	20.0	80.0
1980	25.3	74.7
1985	25.9	74.1
1990	27.5	72.5
1995	31.2	68.8
2000	35.2	64.8
2002	37.5	62.5

Para esta tabela pode-se usar uma representação gráfica idêntica à usada no ponto 1, para mostrar a percentagem de filhos dentro e fora do casamento.

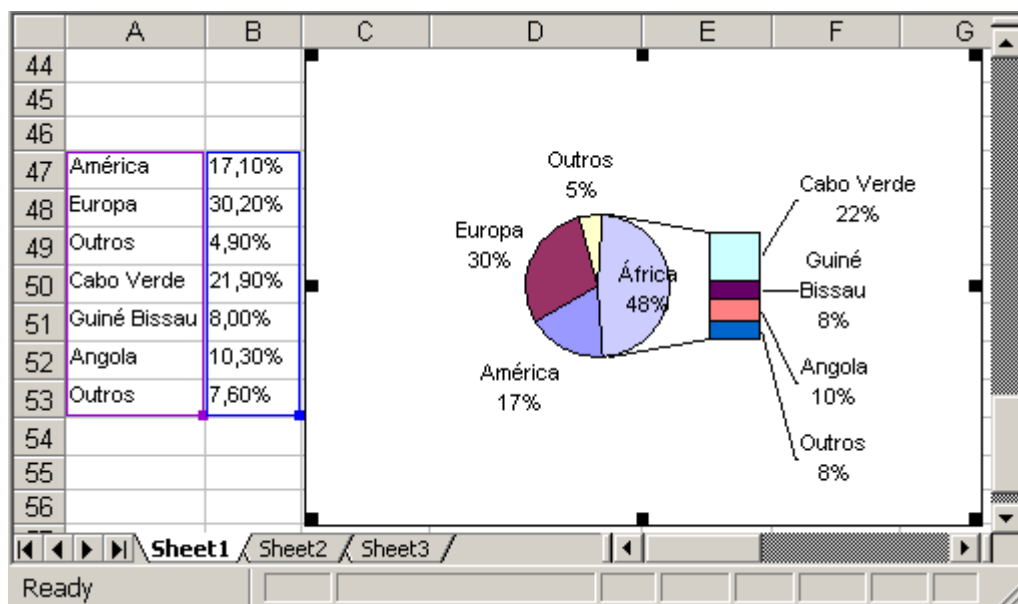
5. População estrangeira com estatuto legal de residente segundo a nacionalidade

América	17,1%		Angola	10,3%
Europa	30,2%	África	Cabo Verde	21,9%
África	47,8%		Guiné Bissau	8,0%
Outros	4,9%		Outros	7,6%

Para fazer uma representação destes dados recorreremos a um diagrama em *Pie* (circular), mas num subtipo especial que permite visualizar a forma como África está repartida. Assim considere-se a seguinte tabela em Excel, ocupando as células A47 a B53 e seleccione-se **Chart** → *Pie* → 6º subtipo → *Next* → *Next* → *Data labels* → *Show label and percent* → *Legend* → *Retirar a selecção de Show Legend* → *Finish*:



Para incluir Cabo Verde na parte direita do gráfico carregar com o botão direito do rato em qualquer parte do gráfico e seleccionar *Format Data Series* → *Second plot contains the last: 4* → *Finish*. Finalmente substituir *Other* (com 48%) por África:

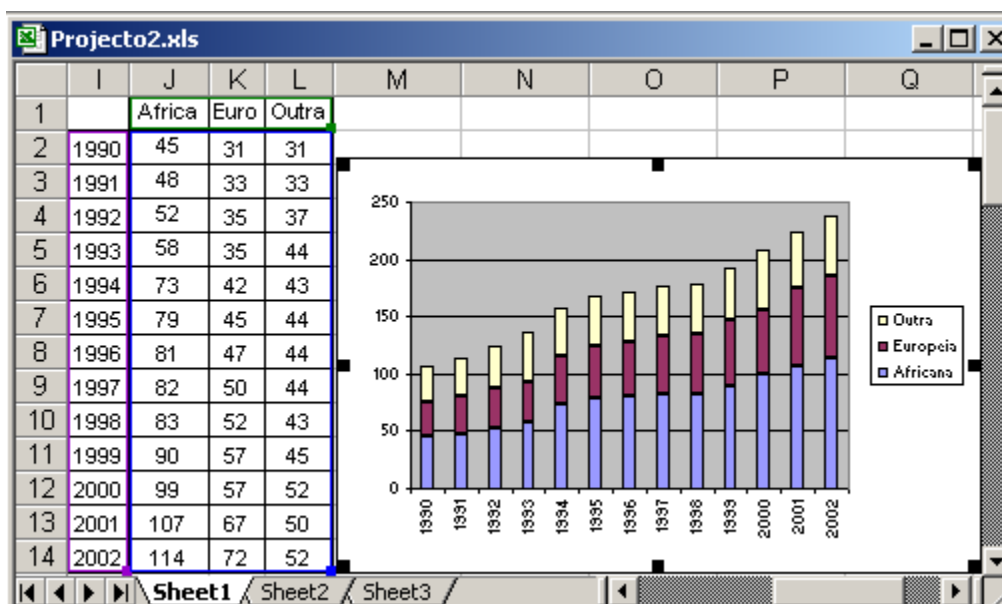


Para representar os dados da tabela seguinte

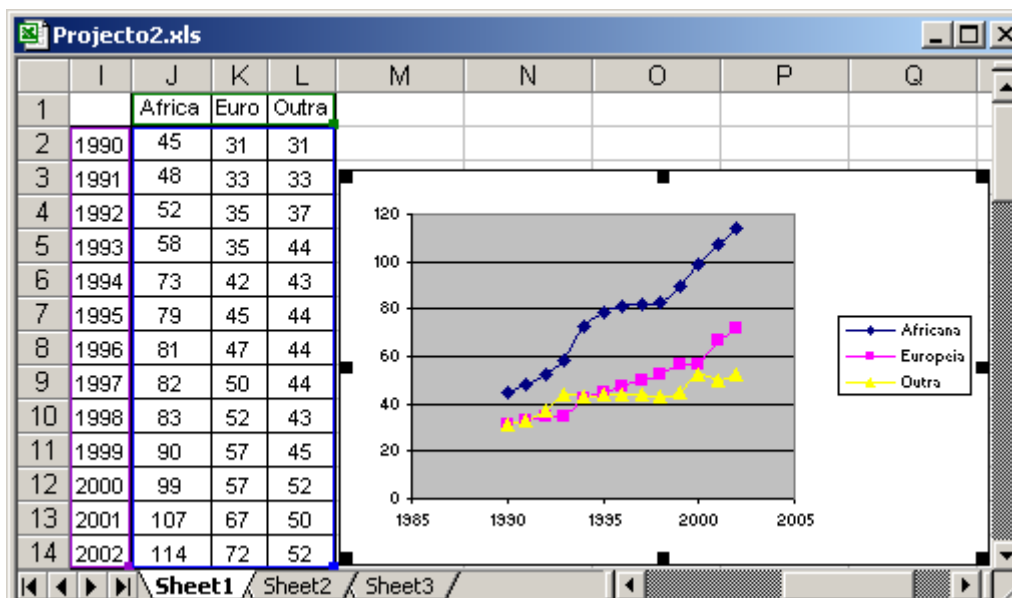
	Africana(1)	Europeia	Outra
1990	45	31	31
1991	48	33	33
1992	52	35	37
1993	58	35	44
1994	73	42	43
1995	79	45	44
1996	81	47	44
1997	82	50	44
1998	83	52	43
1999	90	57	45
2000	99	57	52
2001	107	67	50
2002	114	72	52

(1)Unidade 10³

podemos considerar o 2.º *subtipo* de *Column* (chama-se a atenção para que neste caso não seria correcto utilizar o 3.º *subtipo* de *Column*, uma vez que estamos os dados estão em número absoluto e não em percentagem):



ou o 2.º subtipo de XY(Scatter):



Como vimos há várias representações gráficas para os dados de uma mesma tabela, umas mais sugestivas do que outras. Desde que a representação escolhida esteja correcta, deixa-se a liberdade da escolha ao “artista” que está a organizar e a reduzir os dados.

