


4. Dados bivariados

4.1- Introdução

No módulo de Estatística foi feita referência a dados bidimensionais, de tipo quantitativo. Quando dispomos de uma amostra de dados bivariados, a qual pode ser representada na forma (x_1, y_1) , $(x_2, y_2), \dots, (x_n, y_n)$, apresentamos esta informação através de uma representação gráfica a que se dá o nome de Diagrama de dispersão:

Diagrama de dispersão – É uma representação gráfica para os dados bivariados, em que cada par de dados (x_i, y_i) , é representado por um ponto de coordenadas (x_i, y_i) , num sistema de eixos coordenados.

Já vimos no capítulo 2, a forma de representar, em Excel, dados bivariados, utilizando a opção *XY(Scatter)*. Não apresenta qualquer dificuldade a construção desta representação gráfica, uma vez que basta proceder da seguinte forma:

- Seleccionar as células que contêm os dados, organizados em 2 colunas;
- Carregar no ícone , seleccionar a opção *XY(Scatter)* e o sub-tipo pretendido;
- Formatar convenientemente a representação obtida (retirar a legenda, retirar as linhas de grelha, etc).

Quando se trata de dados qualitativos, não tem sentido proceder à representação gráfica dos dados através de um diagrama de dispersão. No entanto, é possível organizar essa informação na forma de tabelas de contingência (que aliás também podem ser usadas para dados quantitativos, quer discretos, quer contínuos, depois de proceder à sua discretização).

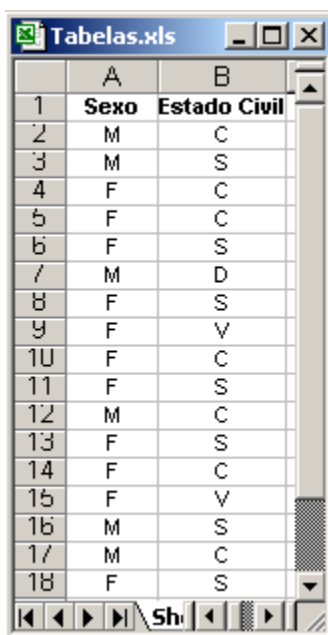
Vamos, neste capítulo, introduzir uma metodologia que utiliza uma ferramenta do Excel, a *PivotTable*, que além de permitir construir tabelas de contingência, também pode ser utilizada para proceder a agrupamentos de dados quantitativos.

4.2 – Tabelas de contingência

Suponhamos que estamos interessados em estudar a associação entre variáveis de tipo qualitativo como, por exemplo, sexo e religião. Uma forma de apresentar os dados, é utilizando tabelas de contingência.

Exemplo 4.2.1 – Uma empresa decidiu estudar o seu pessoal quanto ao estado civil e sexo. Representando por M e F as categorias da variável Sexo, e por C (casado(a)), S (solteiro(a)), D (divorciado(a)) e V (viúvo(a)), obteve a seguinte lista: (M,C), (M,S), (F,C), (F,C), (F,S), (M,D), (F,S), (F,V), (F,C), (F,S), (M,C), (F,S), (F,C), (F,V), (M,S), (M,C), (F,S) (Este exemplo é fictício e serve unicamente para introduzir o estudo das tabelas de contingência, pois os casos interessantes em Estatística envolvem amostras de maior dimensão).

Começámos por introduzir estes dados numa folha de Excel, colocando nas células A1 e B1 os títulos, respectivamente Sexo e Estado Civil, e nas células A2 a A18 a informação sobre o sexo dos 17 elementos e nas células B2 a B18, o respectivo estado civil:

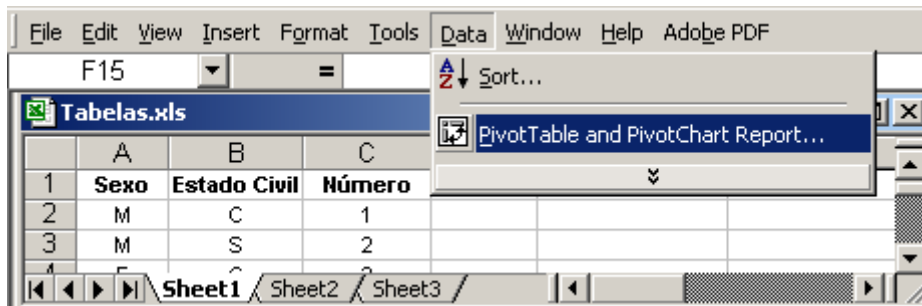


	A	B
1	Sexo	Estado Civil
2	M	C
3	M	S
4	F	C
5	F	C
6	F	S
7	M	D
8	F	S
9	F	V
10	F	C
11	F	S
12	M	C
13	F	S
14	F	C
15	F	V
16	M	S
17	M	C
18	F	S

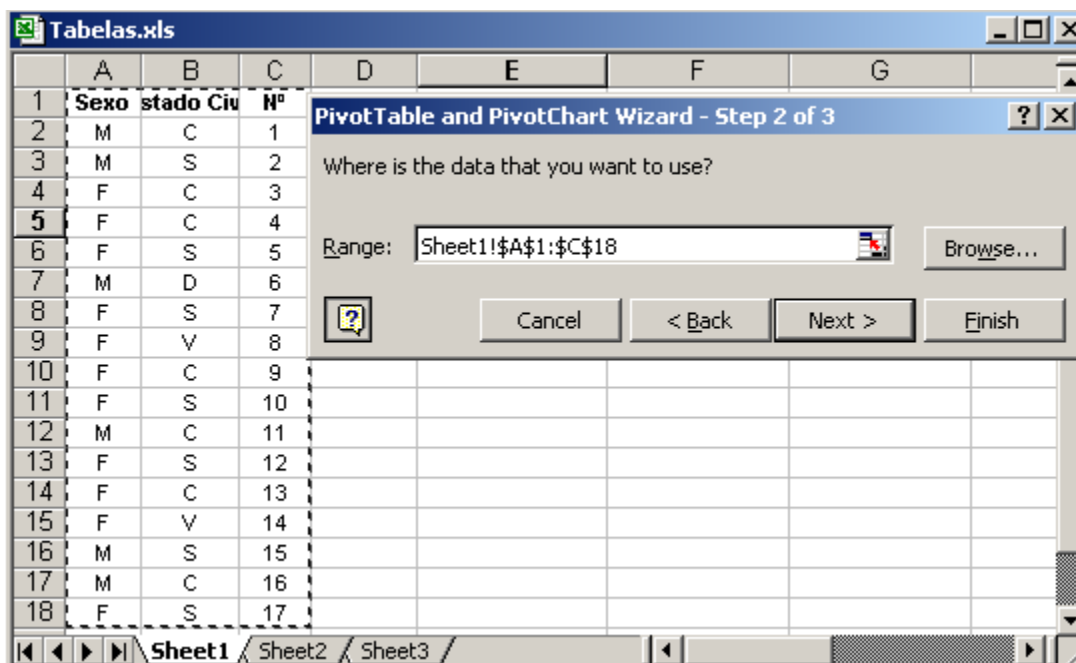
Introduzimos uma coluna auxiliar, a que chamámos N^o, com o número do par, a qual vai ser utilizada para exemplificar a construção de uma tabela de contingência, utilizando as *PivotTable*.

Para criar uma tabela, proceder do seguinte modo:

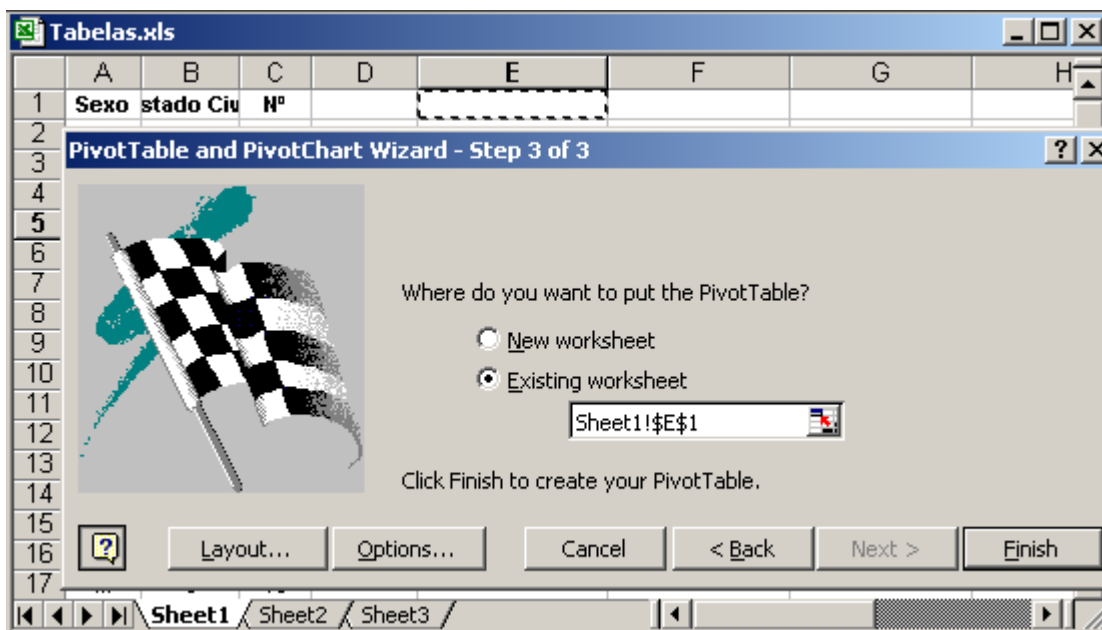
- No menu Data, clicar em *PivotTable and PivotChart Report*:



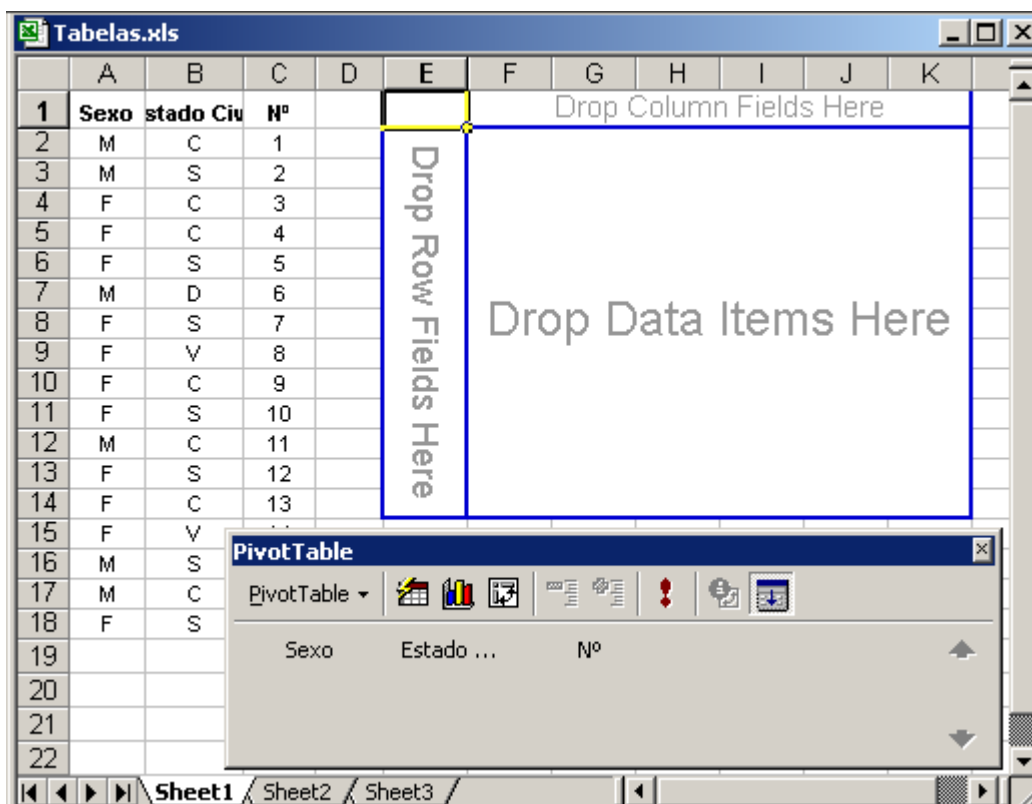
- No passo 1 da *PivotTable and PivotTable Wizard*, seguir as instruções, e clicar *PivotTable* à pergunta *What kind of report do you want to create?*
- No passo 2 seguir as instruções, seleccionando os dados que se pretende usar (não esquecer de seleccionar os títulos):



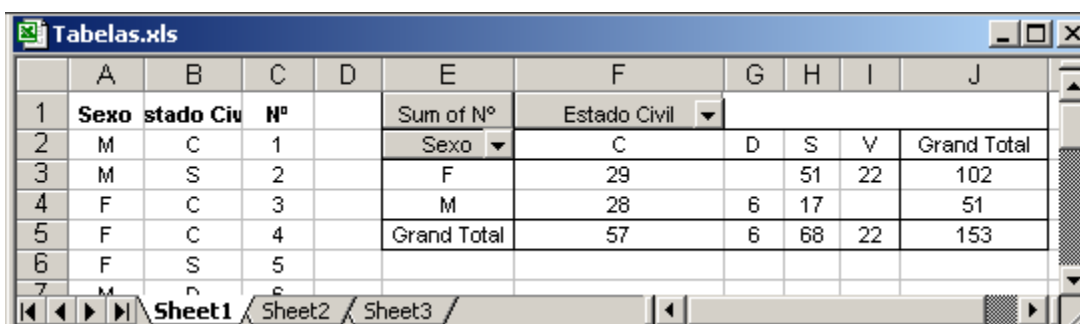
- No passo 3 seleccionar o lugar onde se pretende criar a tabela. Nós optámos por seleccionar a célula E1,



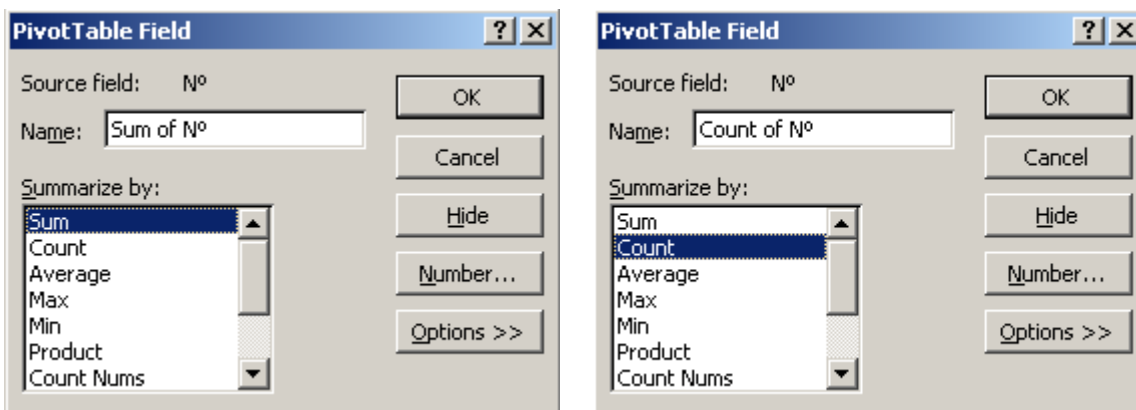
obtendo como resultado:



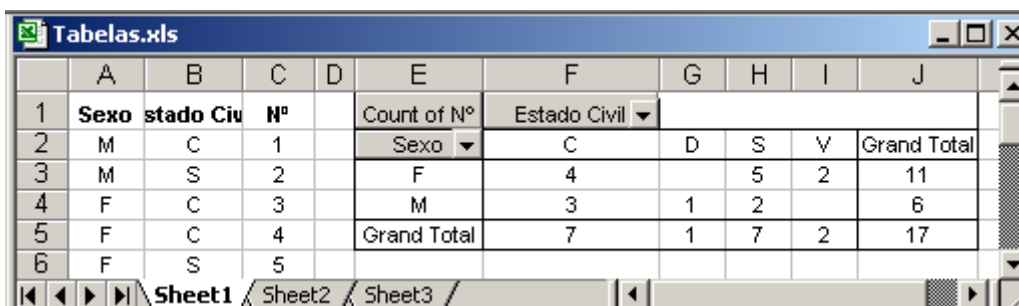
- Arrastar o botão Sexo da barra PivotTable, e colocá-lo (drop it) no campo Row; Arrastar o botão Estado civil da barra PivotTable, e colocá-lo (drop it) no campo Column; Arrastar o botão N° da barra PivotTable, e colocá-lo (drop it) no campo Data:



- Esta tabela, que resulta das operações anteriores, não é a que nos interessa, sendo agora necessário clicar 2 vezes no campo Sum of N.º e seleccionar a opção Count:



Finalmente temos a tabela de contingência desejada, que nos dá a distribuição conjunta (em valores absolutos) do par (Sexo, Estado civil), permitindo obter o número de indivíduos que satisfazem simultaneamente cada uma das modalidades (feminino(a),casado(a)), (feminino(a),divorciado(a)), ...(masculino(a),viúvo(a)):



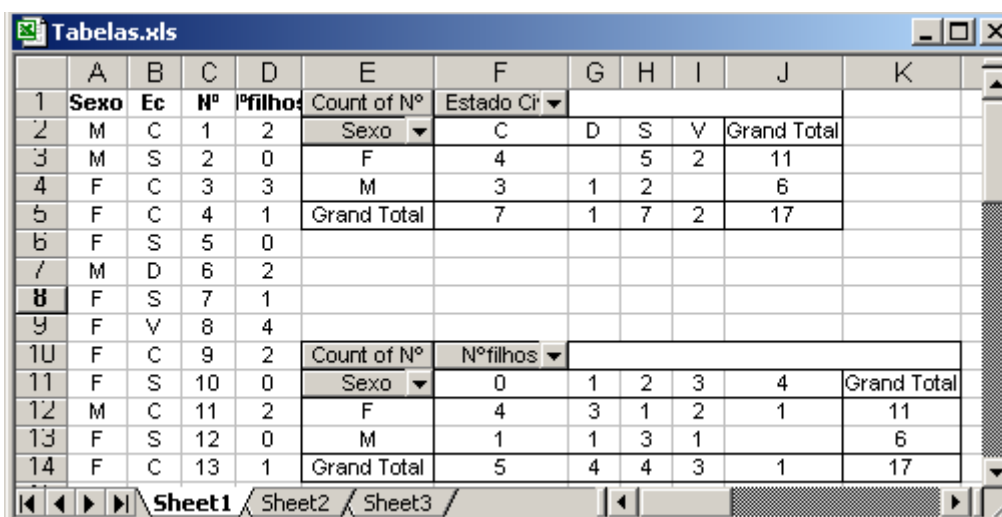
	A	B	C	D	E	F	G	H	I	J
1	Sexo	Estado Civil	Nº		Count of Nº	Estado Civil				
2	M	C	1		Sexo	C	D	S	V	Grand Total
3	M	S	2		F	4		5	2	11
4	F	C	3		M	3	1	2		6
5	F	C	4		Grand Total	7	1	7	2	17
6	F	S	5							

O facto da célula correspondente ao F e D estar vazia, significa que não havia indivíduos do sexo feminino e divorciados. Esta tabela apresenta ainda as distribuições marginais (em valores absolutos) da variável Sexo e Estado civil, respectivamente nas células J3 a J4 e F5 a I5. Efectivamente, através da tabela, pode-se concluir que o número de indivíduos do sexo feminino era 11, enquanto que do sexo masculino eram 6. Analogamente, também podemos tirar conclusões sobre o número de indivíduos em cada modalidade da variável Estado civil.

Exemplo 4.2.1 (cont) - Suponhamos que ao recolher a informação, junto de cada indivíduo, sobre o seu estado civil, também se tinha investigado sobre o número de filhos (esta informação é relevante para o serviço de processamento de salários proceder à retenção do IRS). Construa uma tabela de contingência para o par (Sexo, Estado civil).

Inserimos a informação sobre a variável Nº de filhos, e procedemos à construção da tabela de contingência da mesma forma que anteriormente, com as alterações convenientes, nomeadamente:

- No passo 2 seleccionámos as células de A1 a D18;
- No passo 3 seleccionámos a célula E10, para inserir a tabela;
- No passo seguinte arrastámos o botão Sexo da barra *PivotTable*, e colocámo-lo no campo *Row*; Arrastámos o botão N° de filhos da barra *PivotTable*, e colocámo-lo no campo *Column*; Arrastámos o botão N° de filhos da barra *PivotTable*, e colocámo-lo no campo *Data*;
- Clicámos 2 vezes no campo *Sum of N°* e seleccionámos a opção *Count*:

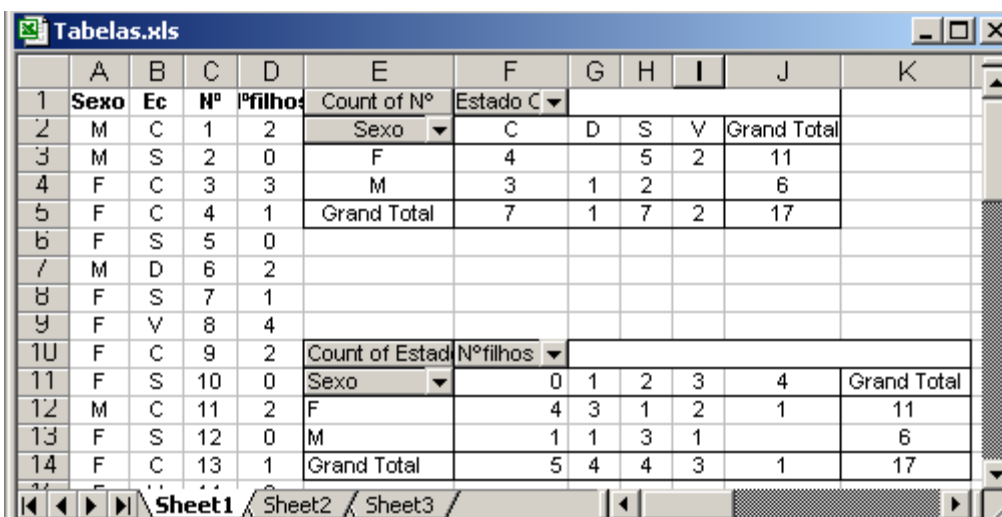


	A	B	C	D	E	F	G	H	I	J	K
1	Sexo	Ec	Nº	Nºfilhos	Count of Nº	Estado Ci					
2	M	C	1	2	Sexo	C	D	S	V	Grand Total	
3	M	S	2	0	F	4		5	2	11	
4	F	C	3	3	M	3	1	2		6	
5	F	C	4	1	Grand Total	7	1	7	2	17	
6	F	S	5	0							
7	M	D	6	2							
8	F	S	7	1							
9	F	V	8	4							
10	F	C	9	2	Count of Nº	Nºfilhos					
11	F	S	10	0	Sexo	0	1	2	3	4	Grand Total
12	M	C	11	2	F	4	3	1	2	1	11
13	F	S	12	0	M	1	1	3	1		6
14	F	C	13	1	Grand Total	5	4	4	3	1	17

Nesta 2ª tabela temos a distribuição conjunta do par (Sexo, N° de filhos).

Exemplo 4.2.1 (cont) – Proceda como no exemplo anterior, excepto no passo seguinte ao passo 3, em que o botão da variável que arrasta para o campo *Data* é o botão da variável Estado civil.

Com este procedimento o resultado é o seguinte:



	A	B	C	D	E	F	G	H	I	J	K
1	Sexo	Ec	Nº	Nºfilhos	Count of Nº	Estado Ci					
2	M	C	1	2	Sexo	C	D	S	V	Grand Total	
3	M	S	2	0	F	4		5	2	11	
4	F	C	3	3	M	3	1	2		6	
5	F	C	4	1	Grand Total	7	1	7	2	17	
6	F	S	5	0							
7	M	D	6	2							
8	F	S	7	1							
9	F	V	8	4							
10	F	C	9	2	Count of Estad	Nºfilhos					
11	F	S	10	0	Sexo	0	1	2	3	4	Grand Total
12	M	C	11	2	F	4	3	1	2	1	11
13	F	S	12	0	M	1	1	3	1		6
14	F	C	13	1	Grand Total	5	4	4	3	1	17

Quando colocámos o botão Estado civil no campo *Data*, imediatamente obtivemos uma tabela igual à anterior, com as contagens, em vez das somas, já que *Count* é a opção que está seleccionada, por defeito, quando colocamos no campo *Data* uma variável não numérica.

4.3 – Utilização das *PivotTables* para agrupar dados

Quando temos um conjunto de dados, já vimos no Capítulo 2 a forma de proceder ao seu agrupamento. Vamos agora ver, como essa tarefa pode ser feita através da utilização da *PivotTable*.

4.3.1 – Dados de tipo qualitativo

Vamos voltar ao ficheiro *Deputados.xls* (de que apresentamos a seguir uma pequena parte)

	A	B	C	D	E	I
1		Nome	Grupo Parl.	Circulo Eleitoral	Sexo	Idade em 31/12/2007
2	1	Abel Lima Baptista	CDS-PP	Viana do Castelo	M	44
3	2	Adão José Fonseca Silva	PSD	Bragança	M	50
4	3	Agostinho Correia Branquinho	PSD	Porto	M	51
5	4	Agostinho Moreira Gonçalves	PS	Porto	M	55
6	5	Agostinho Nuno de Azevedo Fe	PCP	Braga	M	63
7	6	Alberto Arons Braga de Carva	PS	Setúbal	M	58
8	7	Alberto de Sousa Martins	PS	Porto	M	62

para exemplificar a construção de uma tabela de frequências de uma variável qualitativa, utilizando a *PivotTable*.

Exemplo 4.3.1 – Utilizando a *PivotTable*, proceda ao agrupamento de dados da variável Grupo parlamentar, do ficheiro *Deputados.xls*.

- No menu *Data*, clicar em *PivotTable and PivotChart Report*;
- No passo 1 da *PivotTable and PivotTable Wizard*, seguir as instruções, e clicar *PivotTable* à pergunta *What kind of report do you want to create?*;
- No passo 2 seguir as instruções, seleccionando os dados que se pretende usar (não esquecer de seleccionar os títulos). Neste caso seleccionar as células C1:C231;
- No passo 3 seleccionar o lugar onde pretende criar a tabela. Nós optámos por seleccionar a célula A12;
- Arrastar o botão Grupo parlamentar da barra *PivotTable*, e colocá-lo (drop it) no campo *Row*; Arrastar o botão Grupo parlamentar e colocá-lo (drop it) no campo *Data*:

	AH	AI	AJ	AK	AL	AM	AN
1							
2		Count of Grupo Parl.		Tabela de frequências			
3		Grupo Parl.	Total		Classes	Freq. abs.	
4		BE	8		BE	8	
5		CDS-PP	12		CDS-PP	12	
6		PCP	12		PCP	12	
7		PEV	2		PEV	2	
8		PS	121		PS	121	
9		PSD	75		PSD	75	
10		Grand Total	230		Total	230	

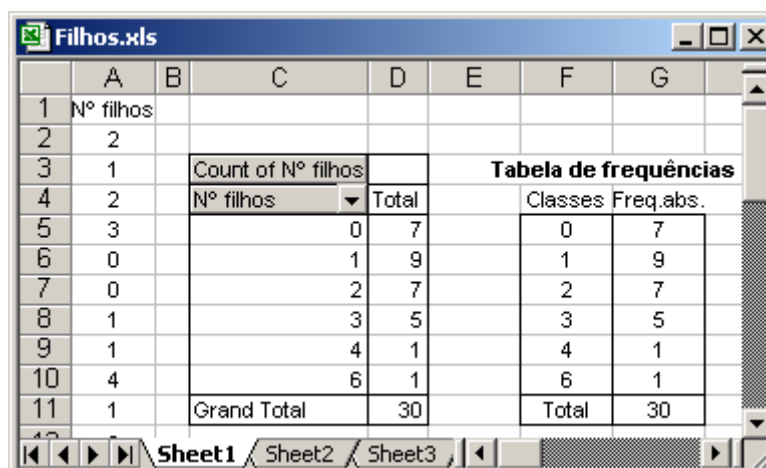
O procedimento anterior conduziu-nos à tabela do lado esquerdo da figura anterior, cujo conteúdo foi copiado para construir a tabela do lado direito, com uma apresentação mais sugestiva.

4.3.2 – Dados de tipo discreto

A organização de dados discretos numa tabela de frequências, utilizando a *PivotTable*, faz-se do mesmo modo que para os dados de tipo qualitativo. Vamos exemplificar procedendo ao agrupamento da variável N.º de filhos dos dados do ficheiro *Filhos.xls*.

Exemplo 4.3.2 - Utilizando a *PivotTable*, proceda ao agrupamento de dados da variável N.º de filhos, do ficheiro *Filhos.xls*.

- No menu Data, clicar em *PivotTable and PivotChart Report*;
- No passo 1 da *PivotTable and PivotTable Wizard*, seguir as instruções, e clicar *PivotTable* à pergunta *What kind of report do you want to create?*;
- No passo 2 seguir as instruções, seleccionando os dados que se pretende usar (não esquecer de seleccionar os títulos). Neste caso seleccionar as células A2 a A31 (que contém o n.º de filhos de uma amostra de 30 deputados);
- No passo 3 seleccionar o lugar onde pretende criar a tabela. Nós optámos por seleccionar a célula C3;
- Arrastar o botão N.º de filhos da barra *PivotTable*, e colocá-lo (*drop it*) no campo *Row*; Arrastar o mesmo botão e colocá-lo (*drop it*) no campo *Data*;
- Clicar duas vezes no botão *Sum of N.º filhos*, da tabela, e seleccionar *Count*:



	A	B	C	D	E	F	G
1	Nº filhos						
2	2						
3	1		Count of Nº filhos		Tabela de frequências		
4	2		Nº filhos	Total	Classes Freq.abs.		
5	3		0	7	0	7	
6	0		1	9	1	9	
7	0		2	7	2	7	
8	1		3	5	3	5	
9	1		4	1	4	1	
10	4		6	1	6	1	
11	1		Grand Total	30	Total	30	

Obtivemos a tabela do lado esquerdo, a qual foi copiada para o lado direito, com um aspecto mais usual.

4.3.3 – Dados de tipo contínuo

Vamos exemplificar o agrupamento de uma variável de tipo contínuo, utilizando a *PivotTable*, mas avisamos desde já, que se os dados não forem inteiros, o processo não é correcto e tem de ser utilizado com as devidas precauções, como veremos oportunamente. O processo que vamos utilizar foi sugerido por um artigo de Neville Hunt, na revista *Teaching Statistics* (Volume 25, Number 2, Summer 2003).

Começaremos por abordar a situação de termos uma variável contínua, mas em que os dados são inteiros.

1ª Parte – Dados em formato de inteiro

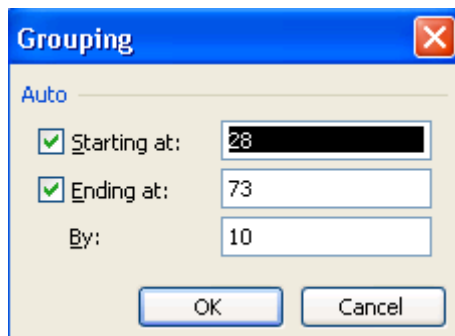
Exemplo 4.3.3 – Considere o ficheiro *Idade.xls*, que contém a idade de 230 deputados. Proceda ao agrupamento em classes, utilizando as *PivotTables*.

Considere o ficheiro *Idade.xls*, em que os dados da variável se encontram nas células C2 a C231 e proceda da seguinte forma:

- No menu Data, clique em *PivotTable and PivotChart Report*;
- No passo 1 da *PivotTable and PivotTable Wizard*, siga as instruções, e clique *PivotTable* à pergunta *What kind of report do you want to create?*;
- No passo 2 siga as instruções, seleccionando os dados que pretende usar. Neste caso seleccione as células C1 a C31 (embora os dados estejam nas células C2 a C231, o título está na C1);
- No passo 3 seleccione o lugar onde pretende criar a tabela. Nós optámos por seleccionar a célula AO4;
- Arraste o botão *Idade* da barra *PivotTable*, e coloque-o (drop it) no campo *Row*; Arraste o mesmo botão e coloque-o (drop it) no campo *Data*;
- Clique duas vezes no botão *Sum of Idade*, da tabela, e seleccione *Count*;

A tabela que aparece depois destas operações, mostra a frequência de cada valor individual (como estamos com dados contínuos, embora inteiros, corremos o risco de termos uma tabela com tantas classes, quantos os dados, todos com frequência igual a 1!). Assim, é necessário proceder a mais algumas operações, para agrupar os dados:

- Clique em algum dos dados da variável Idade e seleccione *Data* → *Group and Outline* → *Group*, que faz surgir o seguinte diálogo:




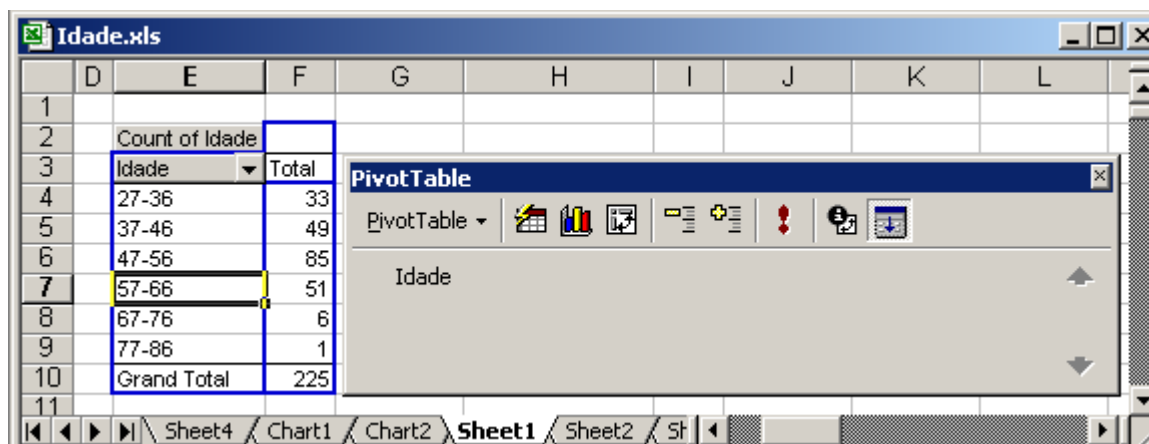
Por defeito, no diálogo anterior é considerado como “*Starting at*” e “*Ending at*” respectivamente, o mínimo e o máximo do conjunto de dados a agrupar. Para “*By*” é considerado, também por defeito, um valor que dependerá do número de dados e da grandeza desses dados.

- Clicando em OK, é produzida a seguinte tabela de frequências:

	AO	AP
3		
4	Count of Idade	
5	Idade	Total
6	28-37	39
7	38-47	62
8	48-57	78
9	58-67	44
10	68-77	7
11	Grand Total	230
12		

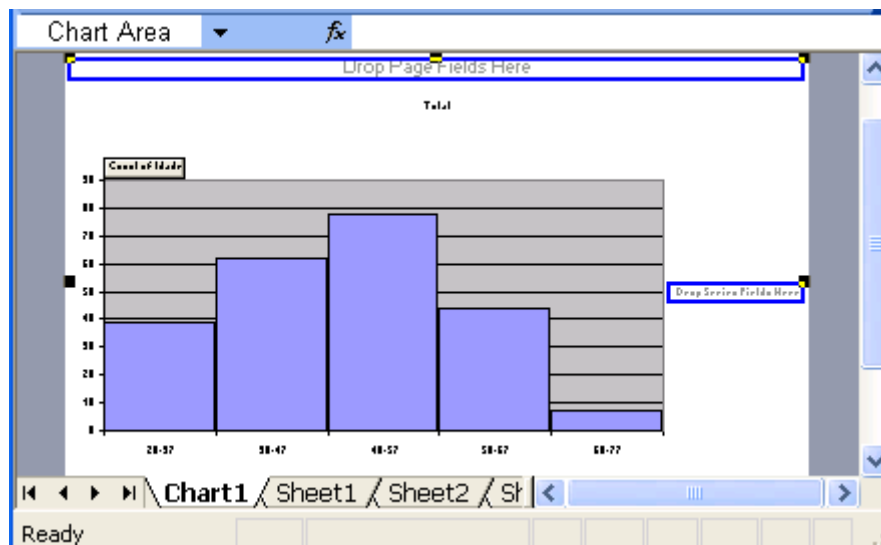
Observação: Repare-se que na construção desta tabela, ao dizer que pretendemos que o agrupamento seja feito *By:10*, não significa que se adicione 10 ao mínimo para formar a 1ª classe e assim por diante. Neste caso 10 é o número de inteiros que vai do limite inferior de cada classe, até ao limite superior e não significa propriamente amplitude de classe, da forma como é definida, isto é, como sendo a diferença entre os limites do intervalo de classe. Se pretendêssemos classes de amplitude 10, teríamos de ter seleccionado, antes de efectuar o agrupamento, *By:11* e obteríamos as classes 28-38, 39-49, 50-60, 61-71 e 72-82.

Para construir o histograma associado a esta tabela, basta carregar em alguma parte da tabela e na barra da *PivotTable* clicar no ícone .

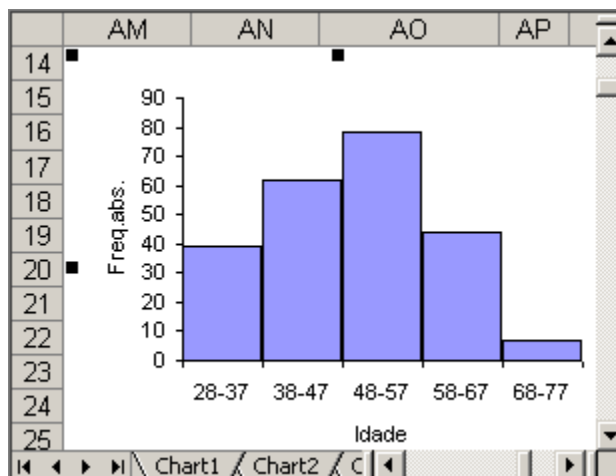


Por defeito aparece a construção de um gráfico de barras, com intervalos entre as barras, que podem ser removidas por um processo idêntico ao já utilizado, aquando da construção do histograma. Assim:

- Clique com o lado direito do rato numa das colunas e seleccione *Format data Series* → *Options* → *Gap width:0*:



- Finalmente podemos esconder os botões clicando com o lado direito do rato num deles e seleccionando *Hide PivotChart Field Buttons* e acrescentando de seguida títulos aos eixos:



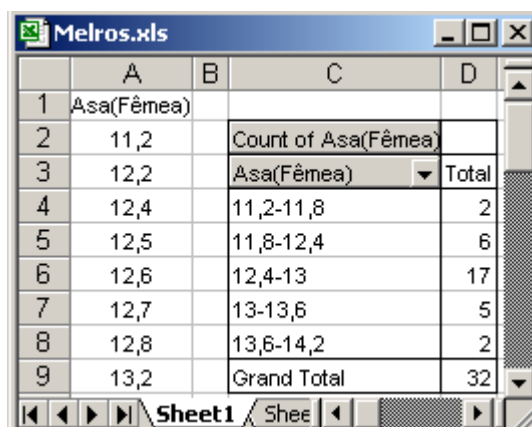
Obs: Para obter o gráfico anterior copiámos a figura obtida numa folha *Chart* do Excel para uma folha normal (*Sheet*).

2ª Parte – Dados em formato decimal

Como vimos na construção das classes da tabela anterior, estas são construídas sem ambiguidade, na medida em que qualquer elemento do conjunto de dados só pode pertencer a uma única classe. O mesmo não acontece se estivermos a trabalhar com dados com casas decimais, como veremos no exemplo seguinte.

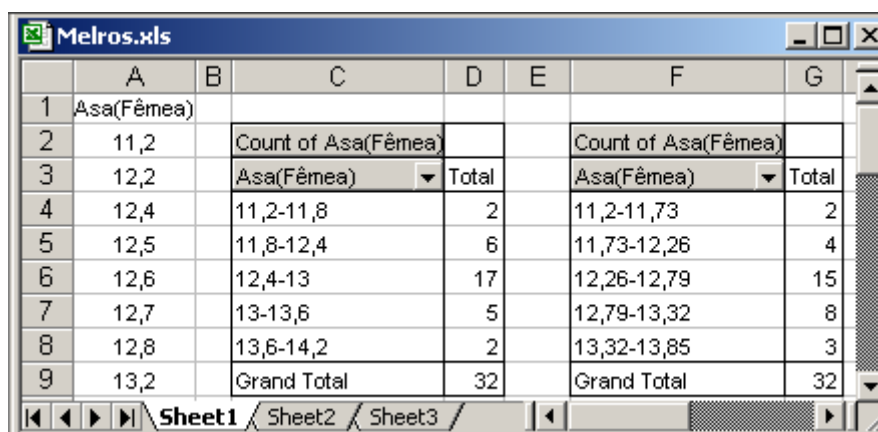
Exemplo 4.3.4 – Considere novamente os dados do exemplo 2.3.8, em que se estudou o comprimento, em centímetros, das asas de melros. Proceda ao agrupamento dos dados correspondentes aos melros-fêmea.

Consideremos a amostra constituída pelas 32 medidas das asas de outros tantos melros-fêmeas que inserimos numa folha de Excel, ocupando as células A2 a A33, reservando a A1 para o título Fêmea. Construímos uma tabela de frequências, utilizando o processo seguido anteriormente, mas escolhendo para amplitude de classe o valor 0,6. O resultado obtido foi a seguinte tabela:



	A	B	C	D
1	Asa(Fêmea)			
2	11,2		Count of Asa(Fêmea)	
3	12,2		Asa(Fêmea)	Total
4	12,4		11,2-11,8	2
5	12,5		11,8-12,4	6
6	12,6		12,4-13	17
7	12,7		13-13,6	5
8	12,8		13,6-14,2	2
9	13,2		Grand Total	32

Como se verifica, ao contrário do que acontecia com a variável Idade, o limite superior de um intervalo é igual ao limite inferior do intervalo seguinte, ficando a dúvida de saber em que classe inserir um elemento igual a um desses limites. Na verdade estes intervalos funcionam como se fossem fechados á esquerda e abertos à direita (excepto a última classe que também é fechada à direita), pelo que um valor igual, por exemplo, a 11,8, será contabilizado na classe 11,8-12,4. Este problema pode ser resolvido, considerando para amplitude de classe um valor decimal, com uma casa decimal a mais dos que os dados. No exemplo anterior, se escolhêssemos como amplitude de classe 0,53, já o problema deixaria de existir, pois não teríamos dúvida em que classe contabilizar qualquer um dos valores do conjunto de dados:



	A	B	C	D	E	F	G
1	Asa(Fêmea)						
2	11,2		Count of Asa(Fêmea)			Count of Asa(Fêmea)	
3	12,2		Asa(Fêmea)	Total		Asa(Fêmea)	Total
4	12,4		11,2-11,8	2		11,2-11,73	2
5	12,5		11,8-12,4	6		11,73-12,26	4
6	12,6		12,4-13	17		12,26-12,79	15
7	12,7		13-13,6	5		12,79-13,32	8
8	12,8		13,6-14,2	2		13,32-13,85	3
9	13,2		Grand Total	32		Grand Total	32

Como diz Neville Hunt no artigo referido anteriormente, página 45, e passamos a citar: *...After reading this article, some teachers will (not unreasonably) decide that Excel is not fit to be used for this type of analysis. However, the universal popularity and availability of Excel are such that students will inevitably try to use it for this purpose at some stage, so it is important that they should be made aware of its limitations and need for vigilance.*



Esta citação vem ao encontro daquilo que pensamos e já referimos neste texto, de que o Excel não é um software de Estatística, mas ao nível elementar resolve muitas situações, desde que ao utilizá-lo se saiba o que se pretende. Por exemplo, quando se pretende um histograma, e ao obter um diagrama de barras, é necessário ter presente que, embora o histograma seja construído à custa de barras, estas têm que estar unidas.