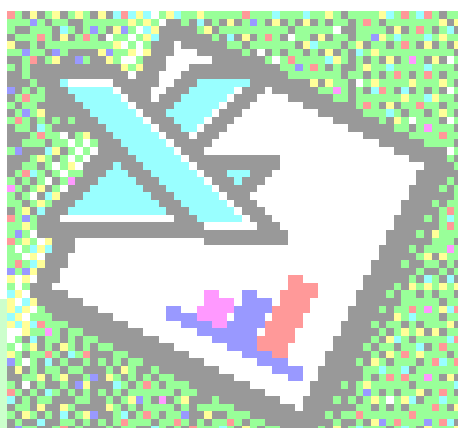




<http://alea-estp.ine.pt>



Dossiers Didáticos

IV – Estatística com Excel

Uma aplicação das noções.

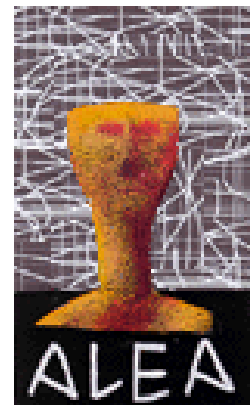
Luís Miguel Cunha

IV – Estatística com Excel

Uma aplicação das noções.

1. Nota Introdutória

O projecto ALEA - Acção Local de Estatística Aplicada - constitui-se como um contributo para a elaboração de novos suportes de disponibilização de instrumentos de apoio ao ensino da Estatística para os alunos e professores do Ensino Básico e Secundário. Este projecto nasceu de uma ideia conjunta da Escola Secundária Tomaz Pelayo e do INE, assente nas necessidades e estruturas que os intervenientes possuem. Melhorar a literacia estatística é, assim, uma condição importante para garantir uma melhor prestação de um serviço de utilidade pública. O Ensino da Estatística no Ensino Básico e Secundário constitui um dos instrumentos mais importantes para cumprir esse objectivo. A página Internet do ALEA está no endereço: <http://alea-estp.ine.pt>.



A área Dossiers Didácticos foi concebida para apoiar a elaboração de materiais didácticos sobre temáticas variadas (População e Demografia, Inquéritos, Inflação e Preços, Gráficos em Estatística, etc.). Vários números destes dossiers vão estar disponíveis neste local, à medida que forem sendo elaborados.

Números anteriores:

Dossier I - População e Demografia -
Quantos Somos e Como somos..

Dossier II – Ambiente e Recursos

Dossier III –A Inflação e o índice de
preços no consumidor

Neste momento apresenta-se o **Dossier IV – Estatística com Excel**, que teve a **colaboração da Prof^a Maria Eugénia Graça Martins, com formato legível em ambiente browser (ex. Internet Explorer, Netscape) e em papel.** Neste *dossier* poderá encontrar uma descrição de como aplicar os conhecimentos adquiridos nas páginas das "Noções de Estatística" (<http://alea-estp.ine.pt>) com o Microsoft Excel. Entendemos que alguns leitores poderão estar menos familiarizados com o Excel, pelo que se inicia este *dossier* com uma breve abordagem ao ambiente Excel e às suas folhas de cálculo.

No final, a rubrica Ver Também contém links para outros estudos de interesse relacionados com as temáticas em causa (artigos, web sites, etc.).





2. Trabalhando com o Excel

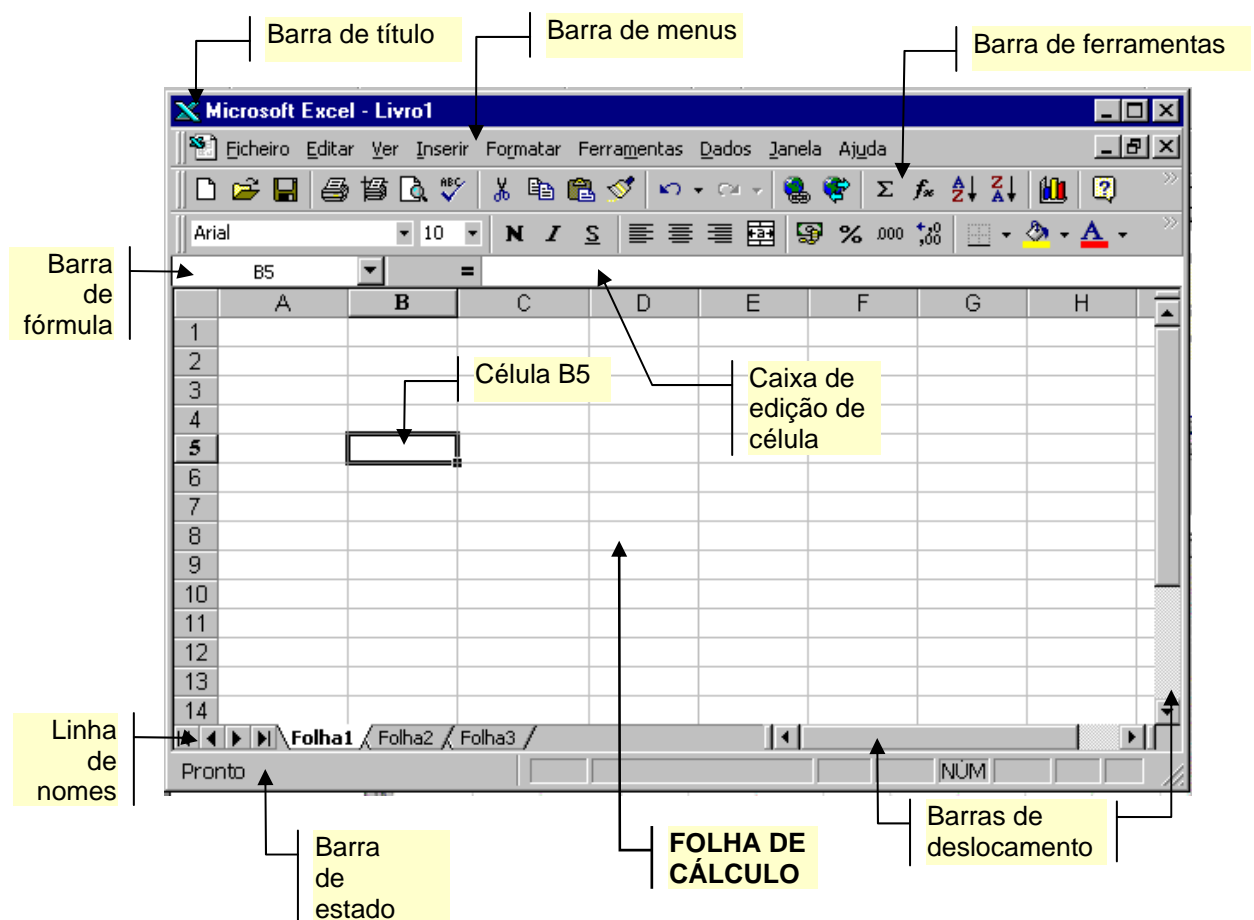
O Excel é um programa de folha de cálculo que foi desenvolvido pela Microsoft para operar em ambiente Windows. Tal como a maioria das aplicações informáticas, tem vindo a ser aperfeiçoado ajustando-se às novas capacidades dos computadores pessoais. Este *dossier* tem como base

a versão 97 em Português, pois apesar de não ser a mais recente (Excel 2000), acreditamos que seja actualmente a de maior disseminação. De forma a facilitar a leitura deste *dossier*, os diferentes comandos e funções do Excel estão destacados em formato **Arial**. Para aqueles que possuem a versão em Inglês, também se apresentam, dentro de parêntesis rectos, os comandos/funções correspondentes.

Sumário:

1. Nota introdutória
2. Trabalhando com o Excel
 - 2.1. O ambiente de trabalho
 - 2.2. Operações básicas sobre a folha de cálculo
 - 2.3. Instalação de funções avançadas
3. Representação gráfica de dados
4. Medidas de localização
5. Medidas de dispersão
6. Dados bivariados
7. Ver também...

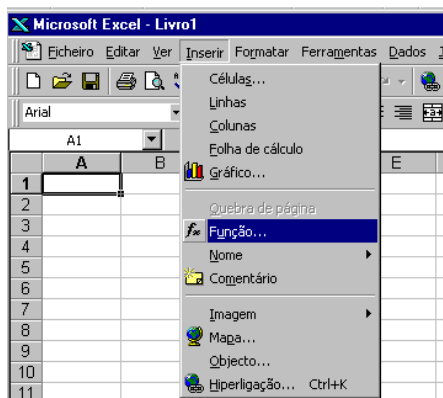
2.1. O ambiente de trabalho



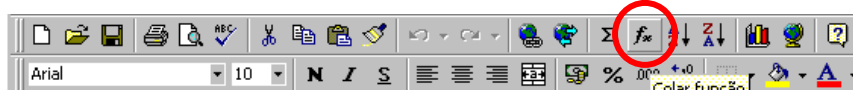
2.2. Operações básicas sobre folhas de cálculo

São várias as opções que o Excel apresenta para aceder às diferentes funções e comandos:

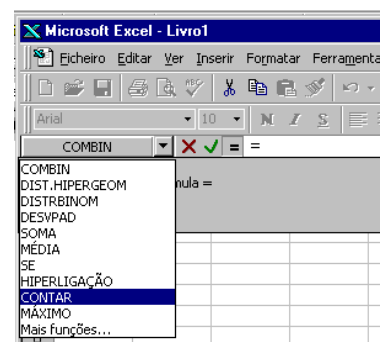
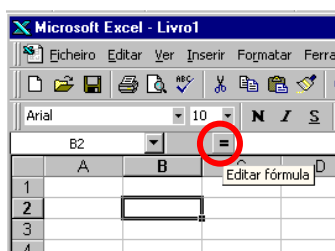
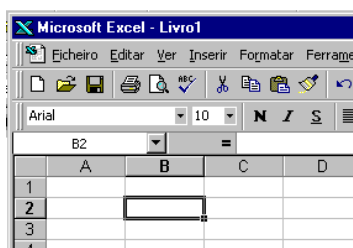
- 1) Através dos menus ou submenus, os quais podem ser seleccionados na **Barra de menus**, clicando com o rato, ou recorrendo a teclas de atalho.



- 2) Através da **Barra de ferramentas**, clicando com o rato no ícone correspondente.



- 3) Através da **Barra de fórmulas**: Selecciona a **célula** de destino, clique na tecla de **edição de fórmulas**, abra o menu das **funções**.

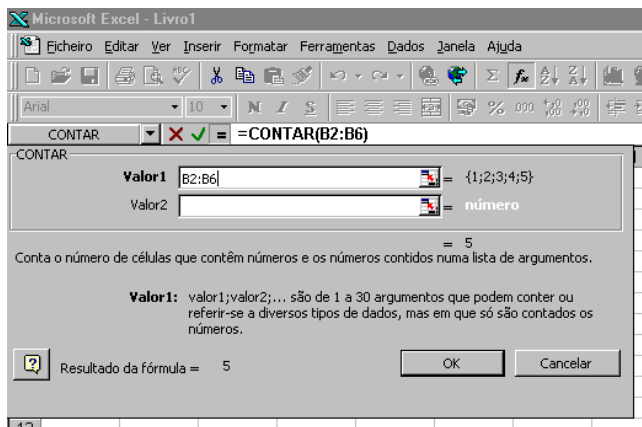
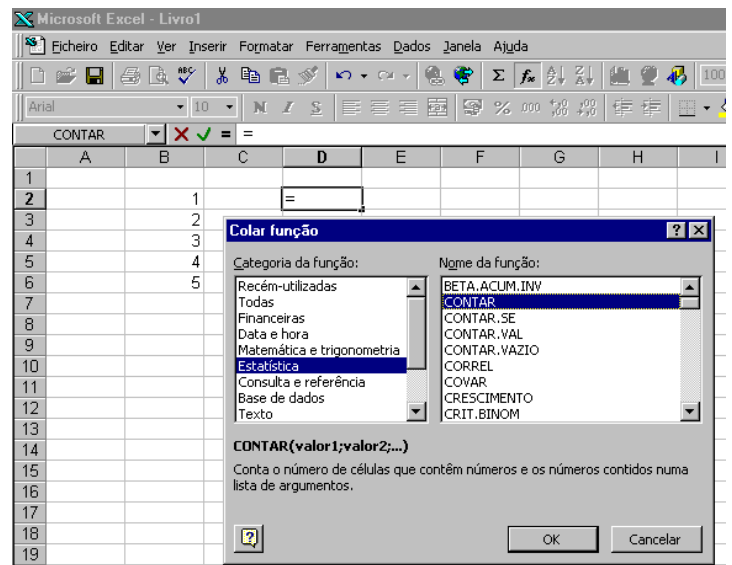


Através dos três métodos citados acederá ao comando **Colar função** [*Function Wizard*], o qual apresenta as diferentes funções agrupadas em 11 categorias:



| Categoria da função | [Function Category] |
|----------------------------|---------------------|
| Financeiras | Finantial |
| Data e Hora | Date & Time |
| Matemática e trigonometria | Math & Trig |
| Estatística | Statistical |
| Consulta e referência | Lookup & Reference |
| Base de dados | Database |
| Texto | Text |
| Lógica | Logical |
| Informação | Information |
| Definidas pelo utilizador | User Defined |
| Engenharia | Engineering |

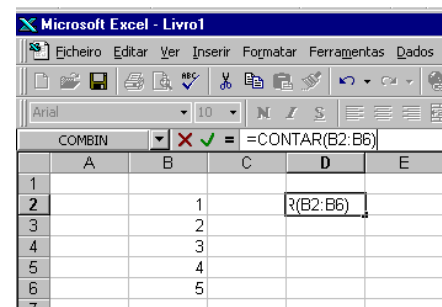
Para cada categoria escolhida, o menu apresenta as diferentes funções, bem como uma breve descrição da função escolhida e da sua sintaxe.



Após a selecção da função desejada, aparece um segundo menu que permite a definição dos diferentes argumentos da função.

Quando estiver familiarizado com as diferentes funções, poderá digitá-las, sem recorrer ao menu.

4) Através da **Caixa de edição de célula** ou directamente na **célula**, ambas seleccionadas com o clicar do rato.

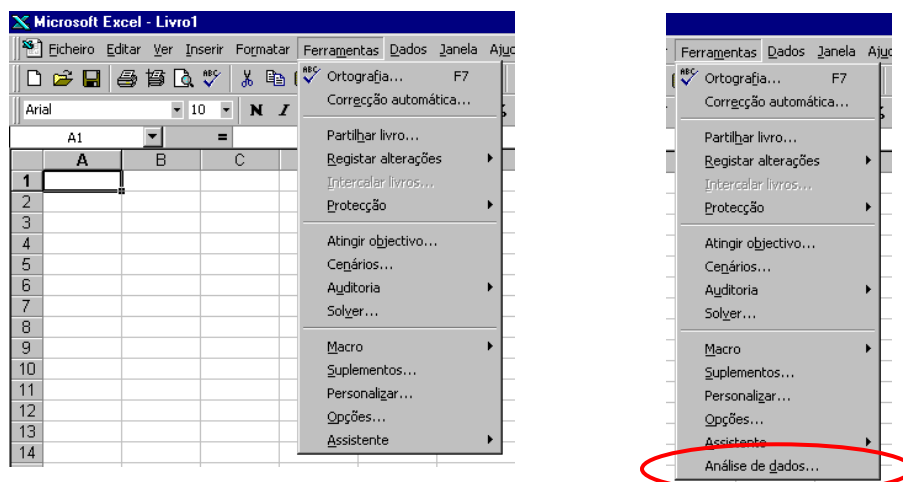


2.3. Instalação de funções avançadas

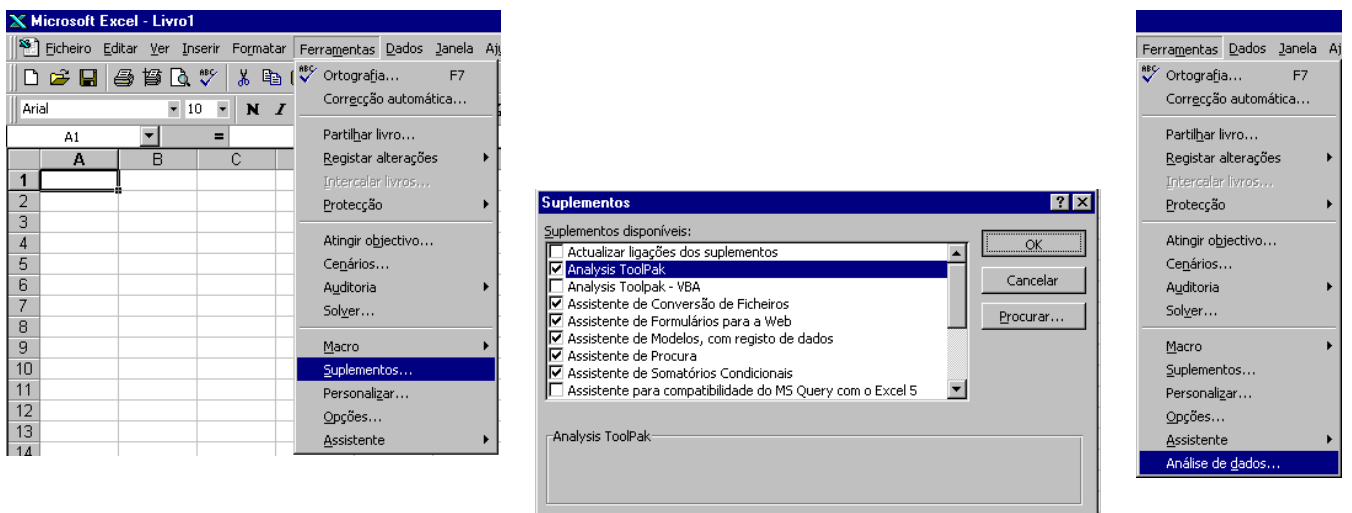
O Excel apresenta um conjunto de funções e procedimentos avançados para a análise estatística de dados, os quais se encontram sob o comando **Análise de dados** [*Data analysis*].

Antes de dar início à apresentação das diferentes funções estatísticas, verifique se estas já foram instaladas no seu computador:

- 1) Aceda ao menu **Ferramentas** [*Tools*] e verifique se o comando **Análise de dados** [*Data analysis*] já consta do menu.



- 2) Se não constar, escolha o comando **Suplementos** [*Add-Ins*] do menu **Ferramentas** [*Tools*], marque a opção **Analysis ToolPak** e clique **OK**, para tornar disponíveis as funções e os procedimentos avançados. O comando **Análise de dados** passou a constar do menu **Ferramentas**.



3. Representação gráfica de dados

Para exemplificar os diferentes processos de representação gráfica de dados, vamos considerar dois conjuntos de dados quantitativos.

Representam a informação resultante de características susceptíveis de serem medidas, apresentam-se com diferentes intensidades, que podem ser de natureza discreta - dados discretos, ou contínua - dados contínuos.

amostra
irmãos
1, 0, 3,
Comece
Da
valores

| | A | B |
|----|---------------|---|
| 1 | N.º de irmãos | |
| 2 | 1 | |
| 3 | 1 | |
| 4 | 2 | |
| 5 | 1 | |
| 6 | 0 | |
| 7 | 3 | |
| 8 | 4 | |
| 9 | 2 | |
| 10 | 3 | |
| 11 | 1 | |
| 12 | 0 | |
| 13 | 2 | |
| 14 | 1 | |
| 15 | 1 | |
| 16 | 0 | |
| 17 | 1 | |
| 18 | 1 | |
| 19 | 0 | |
| 20 | 3 | |
| 21 | 2 | |

Exemplo 1 (Dados

discretos): Suponha-se uma constituída pelo número de dos 20 alunos de uma determinada turma: **1, 1, 2, 4, 2, 3, 1, 0, 2, 1, 1, 0, 1, 1, 0, 3, 2.**

por digitar na célula **A1** o título **N.º de irmãos**. célula **A2** à célula **A21**, digite cada um dos apresentados. Grave o ficheiro para o disco e prepare-se para as novas funções!

Dados quantitativos que só podem tomar um número finito, ou infinito numerável, de valores distintos.

Exemplo
Suponha-
amostra
constituída
de
12,1; 8,9;
8,8; 12,4;
11,0; 10,5.
Comece

| | A | B |
|----|-----------------|---|
| 1 | Nota Matemática | |
| 2 | 12,1 | |
| 3 | 8,9 | |
| 4 | 16,2 | |
| 5 | 8,2 | |
| 6 | 9,8 | |
| 7 | 15,1 | |
| 8 | 14,5 | |
| 9 | 13,4 | |
| 10 | 14,7 | |
| 11 | 7,5 | |
| 12 | 8,8 | |
| 13 | 12,4 | |
| 14 | 16,1 | |
| 15 | 15,2 | |
| 16 | 13,5 | |
| 17 | 14,6 | |
| 18 | 15,5 | |
| 19 | 7,8 | |
| 20 | 12,5 | |
| 21 | 13,2 | |
| 22 | 11 | |
| 23 | 10,5 | |

2 (Dados
se uma
Dados quantitativos que podem tomar todos os valores numéricos, compreendidos no seu intervalo de variação.

pelas notas obtidas por 22 alunos num ponto Matemática de uma determinada turma: **16,2; 8,2; 9,8; 15,1; 14,5; 13,4; 14,7; 7,5; 16,1; 15,2; 13,5; 14,6; 15,5; 7,8; 12,5; 13,2;**

por digitar na célula **A1** o título **Nota Matemática**. Da célula **A2** à célula **A23**,



digite cada um dos valores apresentados. Grave o ficheiro para o disco e prepare-se para as novas funções!

3.1. Frequência absoluta e frequência relativa

Frequência absoluta

Número de elementos pertencentes a uma determinada classe.

Determina-se a frequência absoluta através da função **CONTAR.SE** [*COUNTIF*].

CONTAR.SE(sequência de números ou endereços de célula; critérios de selecção ou endereço de célula contendo critérios)

Esta função calcula o número de células que não estejam em branco e que obedecem a um determinado critério.

Para determinar o número de alunos com 2 irmãos:

- Desloque o apontador para a célula de destino **C3**.
- Introduza a função: **=CONTAR.SE(A2:A21;"2")**.

| | A | B | C | D | E |
|---|---------------|---|---|---|---|
| 1 | N.º de irmãos | | | | |
| 2 | 1 | | | | |
| 3 | 1 | | 4 | | |
| 4 | 2 | | | | |

O procedimento utilizado para o cálculo da frequência absoluta de classes de dados qualitativos, é idêntico ao apresentado para os dados discretos, substituindo-se nos critérios, o cardinal pelo nome da modalidade.

Representam a informação que identifica alguma qualidade, categoria ou característica, não susceptível de medida, mas de classificação, assumindo várias modalidades.

Para determinar o número de alunos que tiveram uma nota inferior a 10 no ponto de Matemática:

- Desloque o apontador para a célula de destino **C2**.

| | A | B | C | D | E |
|---|-----------------|---|---|---|---|
| 1 | Nota Matemática | | | | |
| 2 | 12,1 | | 6 | | |
| 3 | 8,9 | | | | |
| 4 | 16,2 | | | | |

- Introduza a função: **=CONTAR.SE(A2:A23;"<10")**.

| | A | B | C | D | E | F | G |
|---|-----------------|---|---|---|---|---|---|
| 1 | Nota Matemática | | | | | | |
| 2 | 12,1 | | | | | | |
| 3 | 8,9 | | 8 | | | | |
| 4 | 16,2 | | | | | | |



Para determinar o número de alunos que tiveram uma nota igual ou superior a 10 e menor do que 14 no ponto de Matemática:

- Desloque o apontador para a célula de destino **C3**.
- Introduza a função:

=CONTAR.SE(A2:A23;"<14") - CONTAR.SE(A2:A23;"<10").

Frequência relativa

Para o cálculo da frequência relativa, é necessário determinar a dimensão da amostra através das funções **CONTAR [COUNT]** ou **CONTAR.VAL [COUNTA]**.

Razão entre o número de elementos pertencentes a uma determinada classe e o número total de elementos do conjunto de dados a analisar.

CONTAR(sequência de números ou endereços de célula)

Esta função calcula o número de células no conjunto que não estejam em branco e que apresentem valores numéricos.

CONTAR.VAL(sequência de números ou endereços de célula)

Esta função calcula o número de células no conjunto que não estejam em branco, sendo adequada para o cálculo da dimensão de amostras contendo variáveis qualitativas.

| | A | B | C | D | E | F |
|---|---------------|---|---|---|-----|---|
| 1 | N.º de irmãos | | | | | |
| 2 | 1 | | | | | |
| 3 | 1 | | | 4 | 0,2 | |
| 4 | 1 | | | | | |

Para determinar a fracção de alunos com 2 irmãos:

- Desloque o apontador para a célula de destino **C4**.
- Introduza a função: **=CONTAR.SE(A2:A21;"2")/CONTAR(A2:A21)**.

3.2. Tabela de frequências

Para construir uma tabela de frequências, teremos que combinar as funções apresentadas anteriormente.

Tabela que apresenta a distribuição da variável, isto é, quais as classes ou modalidades que a variável assume, assim como a frequência com que assume essas classes.

Vejamos os dados do Exemplo 1, os quais podem ser agrupados em 5 classes distintas: 0, 1, 2, 3 ou



<http://alea-estp.ine.pt>

| | A | B | C | D |
|---|---------------|---------|-----------------------|----------------|
| 1 | N.º de irmãos | | Tabela de frequências | |
| 2 | 1 | Classes | Freq. Absoluta | Freq. Relativa |
| 3 | 1 | | 0 | |
| 4 | 2 | | 1 | |
| 5 | 1 | | 2 | |
| 6 | 0 | | 3 | |
| 7 | 3 | | 4 | |
| 8 | 4 | Total | | |

4 irmãos. Comece por introduzir os diferentes títulos e valores das classes.

Cálculo das frequências absolutas:

- Desloque o apontador para a célula de destino **C3**.
- Introduza a função: **=CONTAR.SE(A2:A21;B3)** .
- Repita o procedimento para as células **C4 a C7**.
- Na célula **C8** introduza a função:

=CONTAR(A2:A21).

| | A | B | C | D |
|---|---------------|---------|-----------------------|----------------|
| 1 | N.º de irmãos | | Tabela de frequências | |
| 2 | 1 | Classes | Freq. Absoluta | Freq. Relativa |
| 3 | 1 | 0 | 4 | |
| 4 | 2 | 1 | 8 | |
| 5 | 1 | 2 | 4 | |
| 6 | 0 | 3 | 3 | |
| 7 | 3 | 4 | 1 | |
| 8 | 4 | Total | 20 | |

| | A | B | C | D |
|---|---------------|---------|-----------------------|----------------|
| 1 | N.º de irmãos | | Tabela de frequências | |
| 2 | 1 | Classes | Freq. Absoluta | Freq. Relativa |
| 3 | 1 | 0 | 4 | 0,2 |
| 4 | 2 | 1 | 8 | 0,4 |
| 5 | 1 | 2 | 4 | 0,2 |
| 6 | 0 | 3 | 3 | 0,15 |
| 7 | 3 | 4 | 1 | 0,05 |
| 8 | 4 | Total | 20 | 1 |

Cálculo das frequências relativas:

- Desloque o apontador para a célula de destino **D3**.
- Introduza a função: **=C3/C8**.
- Repita o procedimento para as células **C4 a C7**.
- Na célula **D8** introduza a função:

=SOMA(D3:D7) [SUM].

Enquanto que no caso dos dados discretos, a construção da tabela de frequências não apresenta qualquer dificuldade, no caso das variáveis contínuas o processo é um pouco mais elaborado, distinguindo-se certas etapas principais, as quais se passam a descrever, e que serão ilustradas com o Exemplo 2:

1º Definição das classes

- a) Determinar a amplitude da amostra (ver secção 5.3).
- b) Dividir essa amplitude pelo número **k** de classes pretendido **{ARRED.PARA.CIMA(LOG(n;2);0)}**; tomar para amplitude de classe **h** um valor aproximado por excesso do quociente anteriormente obtido.
- c) Construir as classes C_1, C_2, \dots, C_k de modo que tenham todas a mesma amplitude e cuja união contenha todos os elementos da amostra:

Para uma amostra de dimensão n , k é o menor inteiro tal que $2^k \geq n$ $\{k \geq \log_2(n)\}$

De forma genérica

$$C_i = [\text{min. da amostra} + (i-1) \times h, \text{min. da amostra} + i \times h[$$

| | A | B | C | D | E | F |
|----|-----------------|--------------|---|-----------------------|----------------|----------------|
| 1 | Nota Matemática | Mínimo | | Tabela de frequências | | |
| 2 | 12,1 | 7,5 | | Classes | Freq. Absoluta | Freq. Relativa |
| 3 | 8,9 | Máximo | | [7,5; 9;3[| | |
| 4 | 16,2 | 16,2 | | [9,3; 11,1[| | |
| 5 | 8,2 | Amplitude | | [11,1; 12,9[| | |
| 6 | 9,8 | 8,7 | | [12,9; 14,7[| | |
| 7 | 15,1 | n | | [14,7; 16,5[| | |
| 8 | 14,5 | 22 | | [7,5; 9;3[| | |
| 9 | 13,4 | k | | [7,5; 9;3[| | |
| 10 | 14,7 | 5 | | Total | 0 | 0 |
| 11 | 7,5 | h | | | | |
| 12 | 8,8 | 1,74 | | | | |
| 13 | 12,4 | h aproximado | | | | |
| 14 | 16,1 | 1,8 | | | | |
| 15 | 15,2 | | | | | |



com $i=1, 2, \dots, k$

2º Contagem dos elementos das classes

Determina-se o número de elementos da amostra que pertencem a cada classe. Analogamente ao que foi considerado no caso dos dados discretos, esses valores serão as frequências absolutas das classes. No caso do exemplo apresentado:

| Nota Matemática | Mínimo | Classes | Freq. Absoluta | Freq. Relativa |
|-----------------|-----------|--------------|----------------|----------------|
| 12,1 | 7,5 | [7,5; 9,3[| 5 | 0,227272727 |
| 8,9 | Máximo | [9,3; 11,1[| 3 | 0,136363636 |
| 16,2 | 16,2 | [11,1; 12,9[| 3 | 0,136363636 |
| 8,2 | Amplitude | [12,9; 14,7[| 5 | 0,227272727 |
| 9,8 | 8,7 | [14,7; 16,5[| 6 | 0,272727273 |
| 15,1 | n | Total | 22 | 1 |
| 14,5 | 22 | | | |
| 13,4 | k | | | |

| Nota Matemática | Mínimo | Classes | Freq. Absoluta | Freq. Relativa |
|-----------------|-----------|--------------|----------------|----------------|
| 12,1 | 7,5 | [7,5; 9,3[| 5 | 0,227272727 |
| 8,9 | Máximo | [9,3; 11,1[| 3 | 0,136363636 |
| 16,2 | 16,2 | [11,1; 12,9[| 3 | 0,136363636 |
| 8,2 | Amplitude | [12,9; 14,7[| 5 | 0,227272727 |
| 9,8 | 8,7 | [14,7; 16,5[| 6 | 0,272727273 |
| 15,1 | n | Total | 22 | 1 |
| 14,5 | 22 | | | |
| 13,4 | k | | | |

Por último, calculam-se as frequências relativas.

Explorando as diversas funções de formatação, poderá obter as seguintes tabelas:

| Tabela de frequências | | |
|-----------------------|----------------|----------------|
| Classes | Freq. Absoluta | Freq. Relativa |
| 0 | 4 | 0,2 |
| 1 | 8 | 0,4 |
| 2 | 4 | 0,2 |
| 3 | 3 | 0,15 |
| 4 | 1 | 0,05 |
| Total | 20 | 1 |

| Tabela de frequências | | |
|-----------------------|----------------|----------------|
| Classes | Freq. Absoluta | Freq. Relativa |
| [7,5; 9,3[| 5 | 0,23 |
| [9,3; 11,1[| 3 | 0,14 |
| [11,1; 12,9[| 3 | 0,14 |
| [12,9; 14,7[| 5 | 0,23 |
| [14,7; 16,5[| 6 | 0,27 |
| Total | 22 | 1 |

3.3. Diagrama de barras

Após a obtenção da tabela de frequências pode-se obter com facilidade o diagrama de barras, através do **Assistente de Gráficos** [Chart Wizard].

Comece por seleccionar as células contendo os dados e os respectivos títulos e clique no ícone

da Barra de ferramentas.

É uma representação adequada para dados discretos ou discretos quantitativos, que consiste em marcar num sistema de eixos coordenados, no eixo dos xx, pontos que representem as classes, e nesses pontos, barras verticais de altura igual ou proporcional à frequência absoluta ou relativa.

| Tabela de frequências | | |
|-----------------------|----------------|----------------|
| Classes | Freq. Absoluta | Freq. Relativa |
| 0 | 4 | 0,2 |
| 1 | 8 | 0,4 |
| 2 | 4 | 0,2 |
| 3 | 3 | 0,15 |
| 4 | 1 | 0,05 |



<http://alea-estp.ine.pt>

Tipos-padrão Tipos personalizados

Tipo de gráfico: Colunas

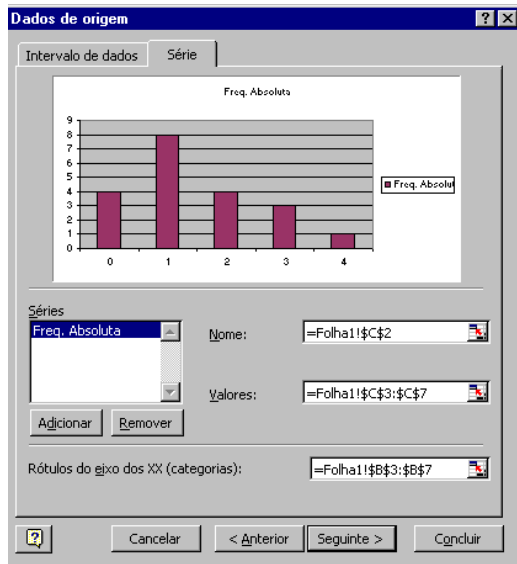
Subtipo de gráfico: Colunas agrupadas. Compara valores ao longo de várias categorias.

Manter premido para ver exemplo

Cancelar < Anterior Seguinte > Concluir

Na primeira **Caixa de diálogo** seleccione a opção **Colunas** [Column].

Para continuar a construção do gráfico, e para passar ao **Passo** [Step] seguinte, clique no botão **Seguinte >** [Next >].



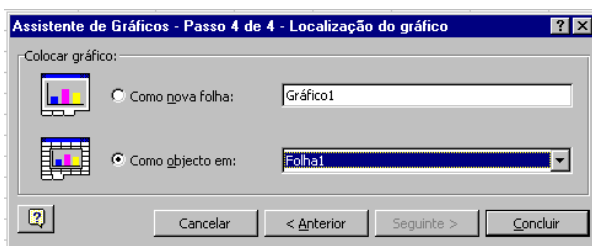
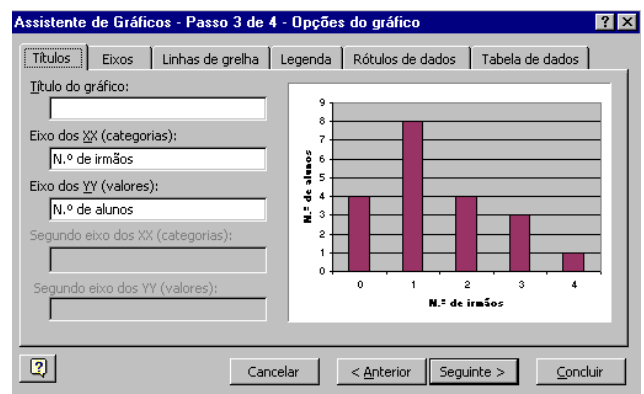
Para continuar a construção do gráfico, e para passar ao **Passo** [Step] seguinte, clique no botão **Seguinte >** [Next >].

No terceiro passo, a **Caixa de diálogo** apresenta várias opções que

permitem formatar o gráfico:

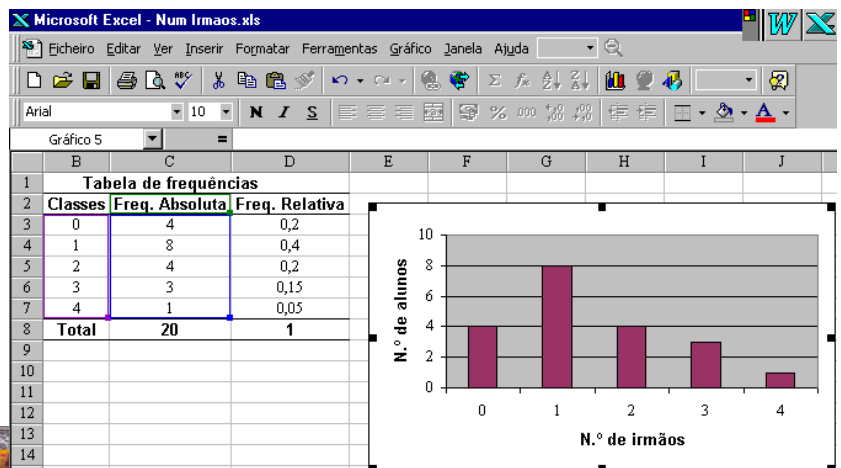
- Em **Títulos** [Titles] siga o exemplo apresentado.
- Em **Legenda** [Legend], desactive a opção da legenda.

Para continuar a construção do gráfico, e para passar ao **Passo** [Step] seguinte, clique no botão **Seguinte >** [Next >].



No último passo pode escolher se o gráfico é colocado numa **nova folha de cálculo** ou numa folha já existente. Clique em **Concluir** [Finish] e obterá o seguinte resultado:

O mesmo procedimento poderá ser aplicado à construção de um diagrama de barras com frequências relativas. Nesse caso, comece por seleccionar as duas colunas, contendo os dados e respectivos títulos, relativos às classes e à frequência relativa. Uma vez que as



colunas não são contíguas, para as seleccionar, comece por seleccionar a primeira coluna e premindo a tecla <ctrl> seleccione a segunda coluna.

3.4. Histograma

Antes de iniciar a construção do histograma, deverá definir, o número de classes, a sua amplitude e os seus limites, tal como se descreveu na secção 3.2, para a construção da tabela de frequências no caso dos dados contínuos.

Representação gráfica de dados contínuos, formado por uma sucessão de rectângulos adjacentes, tendo cada um por base um intervalo de classe e por área a frequência relativa (ou a frequência absoluta). Deste modo a área total coberta pelo histograma é igual a 1 (respectivamente igual a n, a dimensão da amostra).

No entanto, se se pretender utilizar a opção **Histograma [Histogram]** do Excel, faz-se notar que a frequência absoluta de cada classe é calculada, considerando o limite inferior da classe como aberto e o limite superior como fechado. Assim, para construir as classes de modo que tenham todas a mesma amplitude e cuja união contenha todos os elementos da amostra, deve-se começar pelo valor máximo, considerando as seguintes classes:

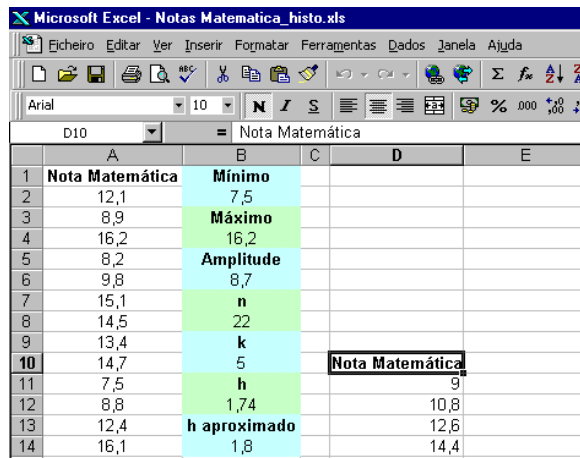
$$C_{k-i+1} =]\max. \text{ da amostra} - i \times h, \max. \text{ da amostra} - (i-1) \times h]$$

com $i=1, 2, \dots, k$

Retomando o Exemplo 2, obtêm-se as seguintes classes:]7,2; 9,0],]9,0; 10,8],]10,8; 12,6],]12,6; 14,4] e]14,4; 16,2]

Crie uma tabela com os limites superiores das classes:

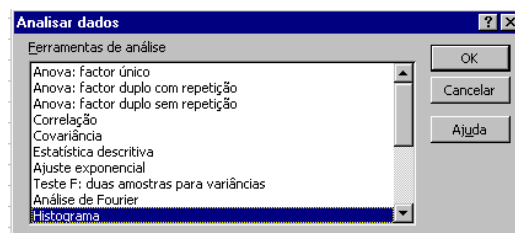
Comece por digitar na célula **D10** o título **Nota Matemática**. Da célula **D11** à célula **D14**, digite cada um dos



| | A | B | C | D | E |
|----|-----------------|--------------|---|-----------------|------|
| 1 | Nota Matemática | Mínimo | | | |
| 2 | 12,1 | 7,5 | | | |
| 3 | 8,9 | Máximo | | | |
| 4 | 16,2 | 16,2 | | | |
| 5 | 8,2 | Amplitude | | | |
| 6 | 9,8 | 8,7 | | | |
| 7 | 15,1 | n | | | |
| 8 | 14,5 | 22 | | | |
| 9 | 13,4 | k | | | |
| 10 | 14,7 | 5 | | Nota Matemática | |
| 11 | 7,5 | h | | | 9 |
| 12 | 8,8 | 1,74 | | | 10,8 |
| 13 | 12,4 | h aproximado | | | 12,6 |
| 14 | 16,1 | 1,8 | | | 14,4 |

valores dos limites superiores de cada uma das classes, com excepção do limite da última classe.

Através do comando **Análise de dados [Data analysis]**, presente no menu **Ferramentas [Tools]**. (Verifique se efectuou o procedimento descrito na secção 2.3). Ao aceder ao comando **Análise de**



dados [Data analysis], seleccione a opção **Histograma** [Histogram] e clique **OK**.

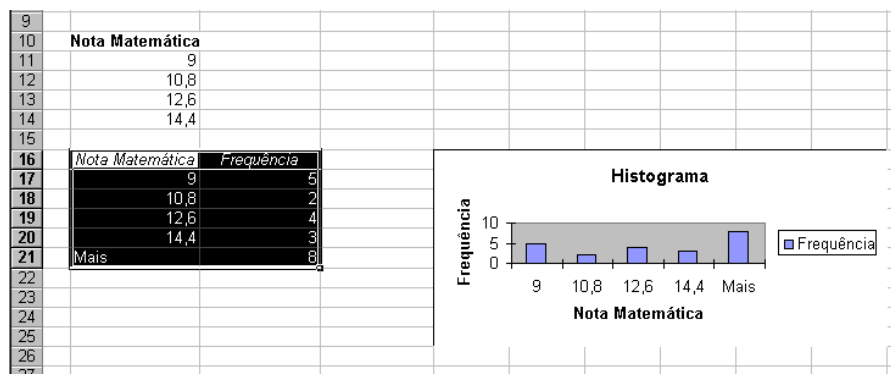
Localização dos dados

Seleccionar quando incluídos no intervalo de entrada

Seleccionar para obter o histograma

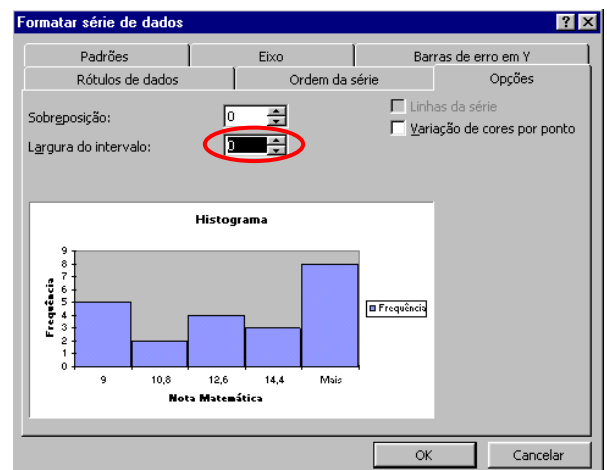
Localização da tabela com os limites das classes

Após preenchimento da **Caixa de diálogo**, clique **OK**.



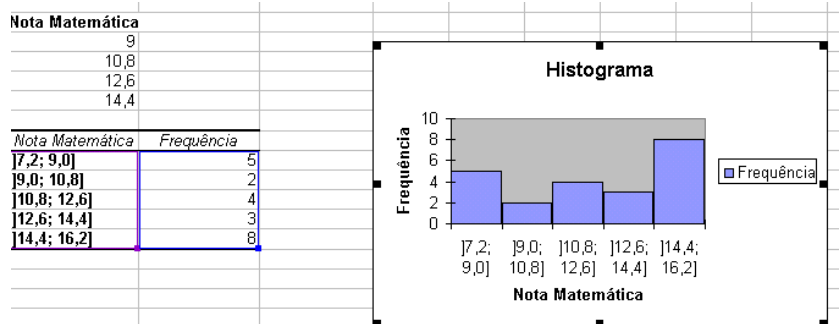
De facto, o que o Excel oferece, é mais um diagrama de barras do que um verdadeiro histograma. Pode-se minimizar esta falha com os seguintes passos:

a) Seleccione o gráfico e clique duas vezes sobre as barras de frequências, de forma a que se abra o menu **Formatar Série de Dados** [Format Data Series]. Seleccione o sub-menu **Opções** [Options] e ajuste para **0** o valor da **Largura do intervalo** [Gap width].



b) Substitua na **folha de cálculo**, os valores do limite superior das classes, pela expressão dos intervalos de classe.

A opção **Histograma** [*Histogram*] não permite, a construção (*desenho*) de histogramas com amplitudes de classe variáveis. Para visualizar um histograma correctamente elaborado em Excel, aceda à página da **DISCUS** e abra o módulo de estatística descritiva (DESC.XLS).



(http://www.mis.coventry.ac.uk/research/discus/discus_home.html)

3.7. Diagrama de extremos e quartis

Para a construção do diagrama de extremos e quartis, são necessárias cinco estatísticas: a mediana, os 1º e 3º Quartis, o máximo e o mínimo (consulte os capítulos 4 e 5 para ver como as obter).

É um tipo de representação gráfica, em que se realçam algumas características da amostra. O conjunto dos valores da amostra compreendidos entre o 1º e o 3º quartis, Q_{25} e Q_{75} é representado por um rectângulo (caixa) com a mediana indicada por uma barra. Consideram-se seguidamente duas linhas que unem os lados dos rectângulos com os valores máximo e mínimo, respectivamente.

Para ilustrar a elaboração do diagrama, vamos considerar as estatísticas relativas a duas amostras de notas obtidas por alunos de duas turmas num ponto de

| | A | B | C | D |
|---|-------------|---------|---------|---|
| 1 | Estatística | Turma A | Turma B | |
| 2 | 1º Quartil | 9,6 | 9,3 | |
| 3 | Mínimo | 6,3 | 2,9 | |
| 4 | Mediana | 13,1 | 12,2 | |
| 5 | Máximo | 16,2 | 18,1 | |
| 6 | 3º Quartil | 14,6 | 13,1 | |

Matemática, **Turma A**: mediana = 13,1; 1º quartil = 9,6; 3º quartil = 14,6; mínimo = 6,3 e máximo = 16,2.

Turma B: mediana = 12,2; 1º quartil = 9,3; 3º quartil = 14,6; mínimo = 2,9 e máximo = 18,1.

Comece por digitar nas células **A1**, **A2** e **A3** os títulos **Estatísticas**, **Turma A** e **Turma B**, respectivamente.


Para cada série introduza os valores das **estatísticas de acordo com a seguinte ordem**: 1º quartil, mínimo, mediana, máximo e 3º quartil. Digite cada um dos valores apresentados. Grave o ficheiro para o disco.

Após a digitação da tabela pode-se obter com facilidade o diagrama de extremos e quartis, através do **Assistente de Gráficos** [*Chart Wizard*].



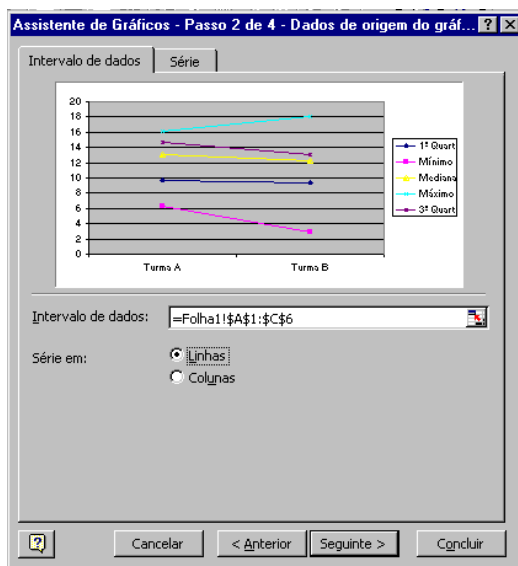
<http://alea-estp.ine.pt>

| | A | B | C | D | E |
|---|-------------|---------|---------|---|---|
| 1 | Estatística | Turma A | Turma B | | |
| 2 | 1º Quartil | 9,6 | 9,3 | | |
| 3 | Mínimo | 6,3 | 2,9 | | |
| 4 | Mediana | 13,1 | 12,2 | | |
| 5 | Máximo | 16,2 | 18,1 | | |
| 6 | 3º Quartil | 14,6 | 13,1 | | |
| 7 | | | | | |

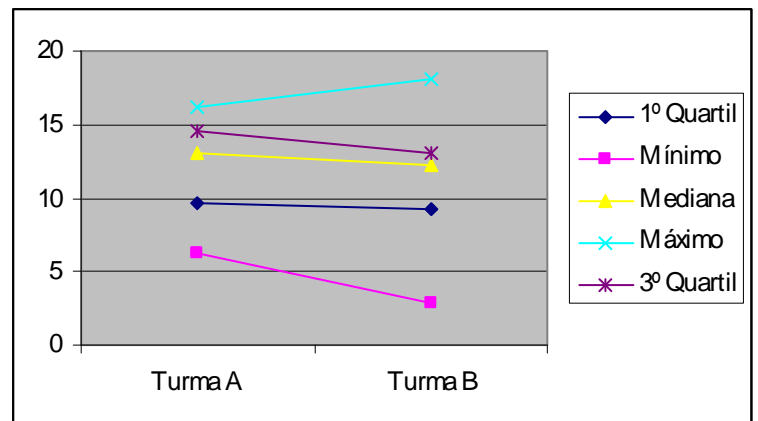
Comece por seleccionar as células contendo os dados e os respectivos títulos e clique no ícone  da **Barra de ferramentas**.

Na primeira **Caixa de diálogo** seleccione a opção **Linhas [Line]**.

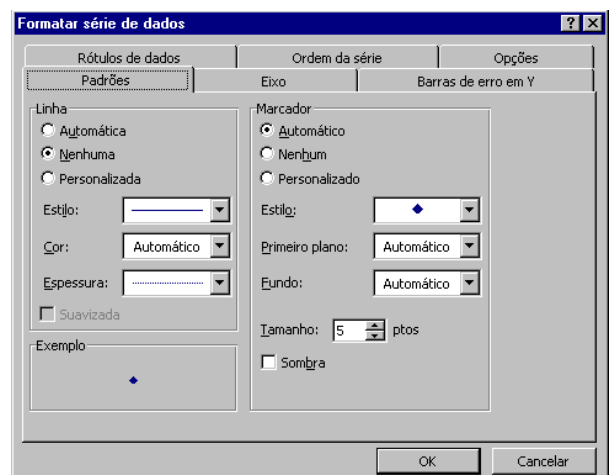
Para continuar a construção do gráfico, e para passar ao **Passo [Step]** seguinte, clique no botão **Seguinte > [Next >]**.



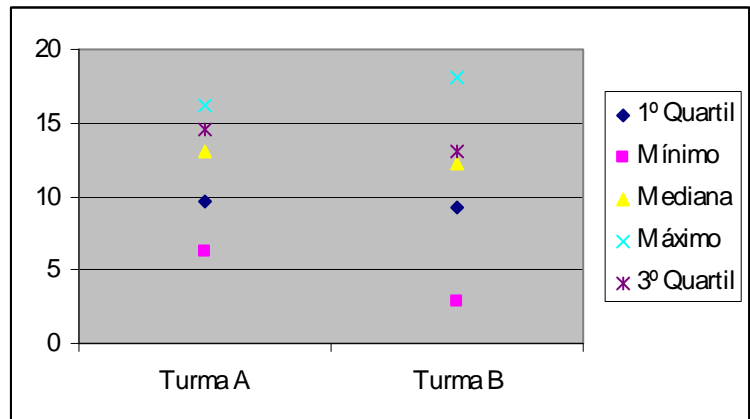
Para continuar a construção do gráfico, seleccione a opção **Série em: Linhas [Series in: Lines]** e clique no botão **Concluir [Finish]**.



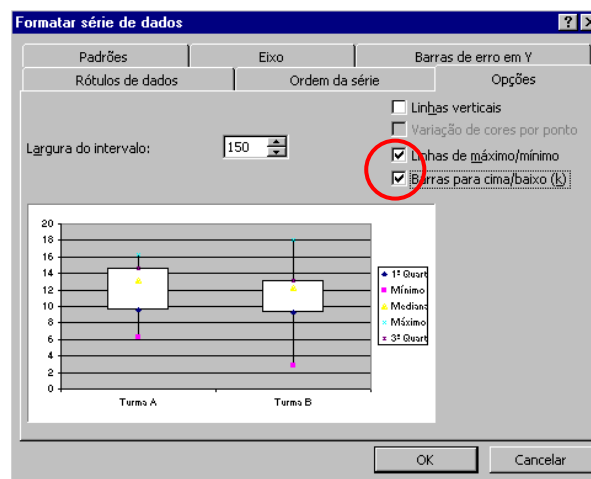
Selecione o gráfico e clique duas vezes sobre uma das linhas, de forma a que se abra o menu **Formatar Série de Dados [Format Data Series]**. Seleccione o sub-menu **Padrões [Patterns]** e ajuste **Linha**



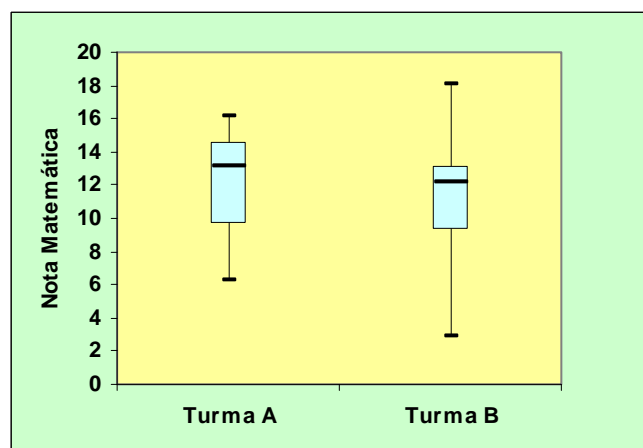
[Line] para a opção **Nenhuma** [None]. Repita o procedimento para todas as séries.



Novamente, seleccione o gráfico e clique duas vezes sobre uma das séries, de forma a que se abra o menu **Formatar Série de Dados** [Format Data Series]. Seleccione o sub-menu **Opções** [Options] e marque as opções: **Linhas de máximo/mínimo** [High-Low lines] e **Barras para cima/baixo** [Up-Down bars].




Explorando as diversas funções de formatação, poderá obter o seguinte gráfico:



3.8. Diagrama circular

À semelhança do apresentado na elaboração do diagrama de barras, após a obtenção da tabela de frequências pode-se obter com facilidade o diagrama circular, através do **Assistente de Gráficos** [*Chart Wizard*].

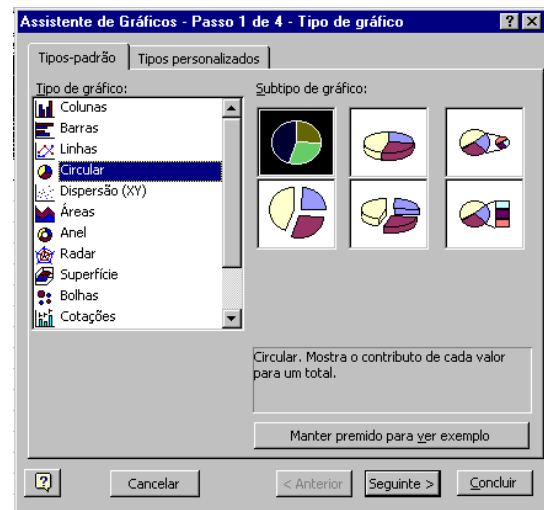
| Microsoft Excel - Num Irmaos.xls | | | |
|---|-----------------------|------------------|----------------|
| Ejchreiro Editar Ver Inserir Formatar Ferrament | | | |
| Arial 10 N I S | | | |
| | D2 | = Freq. Relativa | |
| | B | C | D |
| 1 | Tabela de frequências | | |
| 2 | Classes | Freq. Absoluta | Freq. Relativa |
| 3 | 0 | 4 | 0,2 |
| 4 | 1 | 8 | 0,4 |
| 5 | 2 | 4 | 0,2 |
| 6 | 3 | 3 | 0,15 |
| 7 | 4 | 1 | 0,05 |
| 8 | Total | 20 | 1 |

Comece por seleccionar as células contendo os dados e o respectivo título e clique no ícone  da Barra de ferramentas.

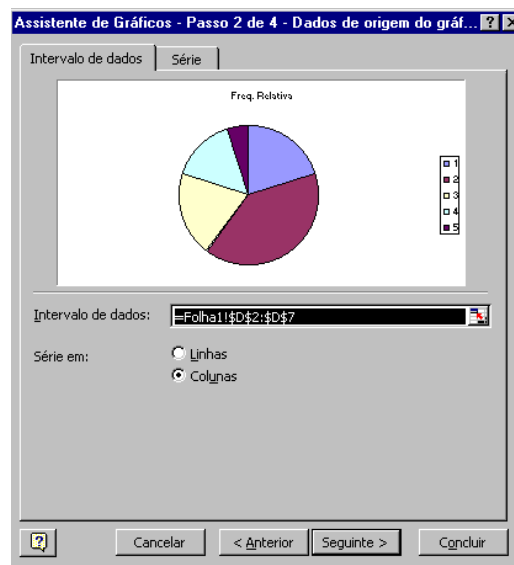
Como o nome sugere esta representação é constituída por um círculo, em que se apresentam vários sectores circulares, tantos quanto as classes consideradas na tabela de frequências da amostra em estudo. Os ângulos dos sectores são proporcionais às frequências das classes.

Na primeira **Caixa de diálogo** seleccione a opção **Circular** [*Pie*].

Para continuar a construção do gráfico, e para passar ao **Passo** [*Step*] seguinte, clique no botão **Seguinte >** [*Next >*].



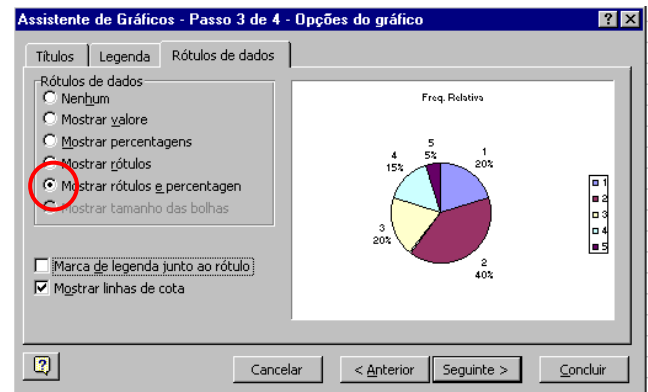
Para continuar a construção do gráfico, e para passar ao **Passo** [*Step*] seguinte, clique no botão **Seguinte >** [*Next >*].



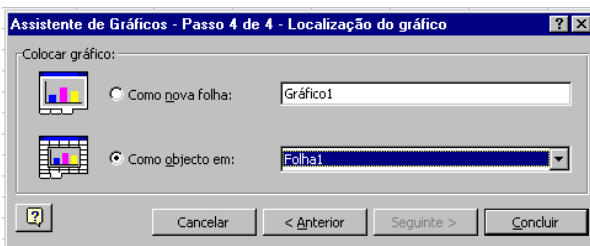
No terceiro passo, a **Caixa de diálogo** apresenta várias opções que permitem formatar o gráfico:

- Em **Rótulos de dados** [*Data labels*] seleccione a opção **Mostrar rótulos e percentagem** [*Show labels and percentages*].

- Em **Legenda** [*Legend*], desactive a opção da legenda.

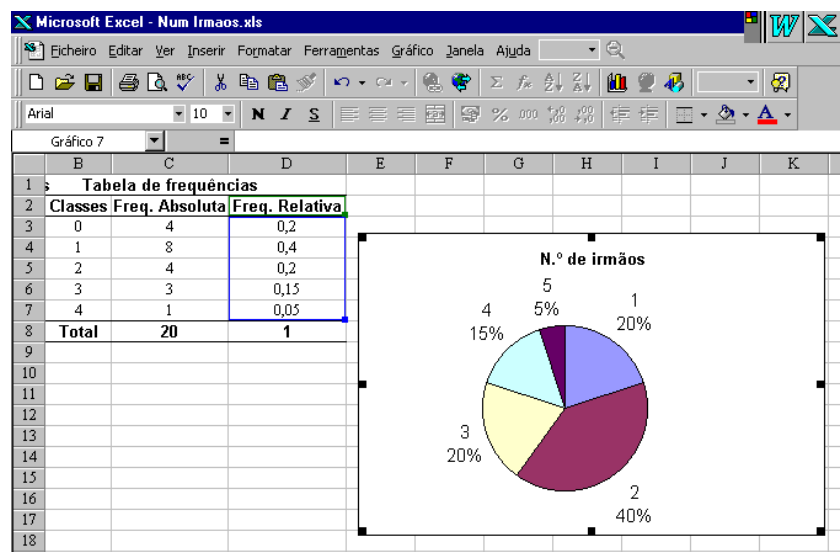


Para continuar a construção do gráfico, e para passar ao **Passo** [*Step*] seguinte, clique no botão **Seguinte >** [*Next >*]

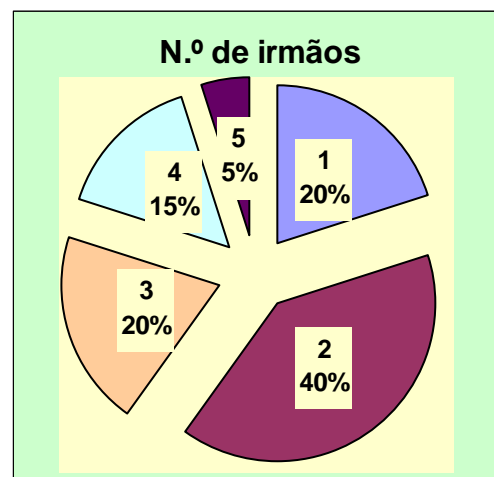


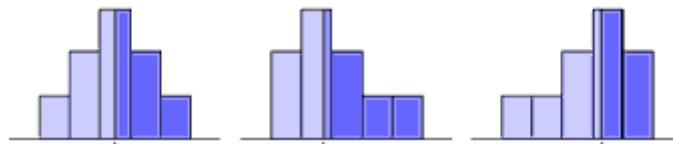
No último passo pode escolher se o gráfico é colocado numa **nova folha de cálculo** ou numa folha já existente.

Clique em **Concluir** [*Finish*] e obterá o seguinte resultado



Explorando as diversas funções de formatação, poderá obter o seguinte gráfico:





4. Medidas de localização

Para ilustrar o cálculo das diferentes medidas de localização e de dispersão, vamos considerar o seguinte conjunto de dados, correspondente à altura, em cm, de uma amostra de 10 alunos de uma turma de Educação Física: 175, 170, 185, 154, 165, 175, 182, 178, 175 e 173.

Comece por digitar na célula **A1** o título **Altura (cm)**. Da célula **A2** à célula **A11**, digite cada um dos valores apresentados. Grave o ficheiro para o disco e prepare-se para as novas funções!

| | A | B | C | D |
|----|-------------|---|---|---|
| 1 | Altura (cm) | | | |
| 2 | 175 | | | |
| 3 | 170 | | | |
| 4 | 185 | | | |
| 5 | 154 | | | |
| 6 | 165 | | | |
| 7 | 175 | | | |
| 8 | 182 | | | |
| 9 | 178 | | | |
| 10 | 175 | | | |
| 11 | 173 | | | |
| 12 | | | | |

4.1. Média

Determina-se a média através da função **MEDIA** [**AVERAGE**].

É a medida de localização do centro da amostra, mais utilizada e obtém-se somando todos os elementos da amostra e dividindo o resultado da soma pela dimensão da amostra.

MEDIA(sequência de números ou endereços de célula)

O resultado desta função é a média aritmética (\bar{x}) dos valores introduzidos como argumento:

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$$

- Desloque o apontador para a célula **B2** e digite: **Média (cm)**.
- Desloque o apontador para a célula **C2** e introduza a função: **=MEDIA(A2:A11)**.

| | A | B | C | D |
|---|-------------|------------|-------|---|
| 1 | Altura (cm) | | | |
| 2 | 175 | Média (cm) | 173,2 | |
| 3 | 170 | | | |
| 4 | 185 | | | |
| 5 | 154 | | | |



4.2. Moda

Determina-se a moda através da função **MODA** [*MODE*].

Valor que surge com mais frequência num conjunto de dados, se estes são discretos, ou a classe com maior frequência se os dados estão agrupados.

MODA(sequência de números ou endereços de célula)

- Desloque o apontador para a célula **B3** e digite: **Moda (cm)**.
- Desloque o apontador para a célula **C3** e introduza a função: **=MODA(A2:A11)**.

The screenshot shows the Excel interface with the following data table:

| | A | B | C | D |
|---|-------------|------------|-------|---|
| 1 | Altura (cm) | | | |
| 2 | 175 | Média (cm) | 173,2 | |
| 3 | 170 | Moda (cm) | 175 | |
| 4 | 185 | | | |
| 5 | 154 | | | |
| 6 | 165 | | | |

The formula bar shows the formula `=MODA(A2:A11)` entered in cell C3.

4.3. Mediana

Determina-se a mediana através da função **MED** [*MEDIAN*].

É uma medida de localização do centro da distribuição dos dados, correspondente ao valor que divide a amostra ao meio, isto é, metade dos elementos do conjunto de dados são menores ou iguais à mediana, enquanto que os restantes são superiores ou iguais.

MED(sequência de números ou endereços de célula)

- Desloque o apontador para a célula **B4** e digite: **Mediana (cm)**.
- Desloque o apontador para a célula **C4** e introduza a função: **=MED(A2:A11)**.

The screenshot shows the Excel interface with the following data table:

| | A | B | C | D |
|---|-------------|--------------|-------|---|
| 1 | Altura (cm) | | | |
| 2 | 175 | Média (cm) | 173,2 | |
| 3 | 170 | Moda (cm) | 175 | |
| 4 | 185 | Mediana (cm) | 175 | |
| 5 | 154 | | | |

The formula bar shows the formula `=MED(A2:A11)` entered in cell C4.



4.4. Quantis

Os quantis determinam-se através das funções

QUARTIL [*QUARTILE*] e **PERCENTIL** [*PERCENTILE*].

Chama-se quantil de ordem p , $0 < p < 1$, ou percentil $100p\%$ ao valor Q_p tal que $100p\%$ dos elementos da amostra são menores ou iguais a Q_p e os restantes são maiores ou iguais a Q_p . Os quantis de ordem $0,25$ e $0,75$ chamam-se, respectivamente, 1° e 3° quantis.

QUARTIL(sequência de números ou endereços de célula; 0 ou 1 ou 2 ou 3 ou 4)

{0 = mínimo, 1 = 1° quantil, 2 = mediana, 3 = 3° quantil, 4 = máximo}

- Desloque o apontador para a célula **B5** e digite: **1º Quartil (cm)**.
- Desloque o apontador para a célula **C5** e introduza a função: **=QUARTIL(A2:A11;1)**.
- Desloque o apontador para a célula **B6** e digite: **3º Quartil (cm)**.
- Desloque o apontador para a célula **C6** e introduza a função: **=QUARTIL(A2:A11;3)**.

| Microsoft Excel - Estatística com Excel.xls | | | |
|---|-------------|--------------------|--------|
| Ficheiro Editar Ver Inserir Formatar Ferramentas Da | | | |
| Arial 10 | | | |
| C6 | | =QUARTIL(A2:A11;3) | |
| A | B | C | D |
| 1 | Altura (cm) | | |
| 2 | 175 | Média (cm) | 173,2 |
| 3 | 170 | Moda (cm) | 175 |
| 4 | 185 | Mediana (cm) | 175 |
| 5 | 154 | 1º Quartil (cm) | 170,75 |
| 6 | 165 | 3º Quartil (cm) | 177,25 |
| 7 | 175 | | |

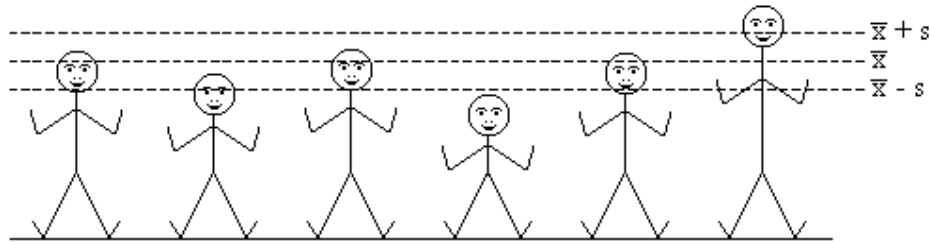
PERCENTIL(sequência de números ou endereços de célula; número entre 0 e 1 inclusivé)

{percentil $100p\%$, introduza o número p }

- Desloque o apontador para a célula **B7** e digite: **Percentil 85% (cm)**.
- Desloque o apontador para a célula **C7** e introduza a função: **=PERCENTIL(A2:A11;0.85)**.

| Microsoft Excel - Estatística com Excel.xls | | | |
|--|-------------|-------------------------|--------|
| Ficheiro Editar Ver Inserir Formatar Ferramentas Dados | | | |
| Arial 10 | | | |
| C7 | | =PERCENTIL(A2:A11;0.85) | |
| A | B | C | D |
| 1 | Altura (cm) | | |
| 2 | 175 | Média (cm) | 173,2 |
| 3 | 170 | Moda (cm) | 175 |
| 4 | 185 | Mediana (cm) | 175 |
| 5 | 154 | 1º Quartil (cm) | 170,75 |
| 6 | 165 | 3º Quartil (cm) | 177,25 |
| 7 | 175 | Percentil 85% (cm) | 180,6 |
| 8 | 182 | | |





5. Medidas de dispersão

5.1. Variância

Determina-se a variância através da função **VAR** [VAR]

Medida que se obtém somando os quadrados dos desvios dos dados relativamente à **média** e dividindo o valor da soma pelo número de observações menos uma.

VAR(sequência de números ou endereços de célula)

O resultado desta função é a variância amostral (s^2) dos valores introduzidos como argumento:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

- Desloque o apontador para a célula **B8** e digite: **Variância (cm2)**.
- Desloque o apontador para a célula **C8** e introduza a função: **=VAR(A2:A11)**.

| Microsoft Excel - Estatística com Excel.xls | | | |
|---|-------------|--------------------|----------|
| Estatística com Excel.xls | | | |
| =VAR(A2:A11) | | | |
| A | B | C | D |
| 1 | Altura (cm) | | |
| 2 | 175 | Média (cm) | 173,2 |
| 3 | 170 | Moda (cm) | 175 |
| 4 | 185 | Mediana (cm) | 175 |
| 5 | 154 | 1º Quartil (cm) | 170,75 |
| 6 | 165 | 3º Quartil (cm) | 177,25 |
| 7 | 175 | Percentil 85% (cm) | 180,6 |
| 8 | 182 | Variância (cm2) | 77,26889 |
| 9 | 178 | | |

5.2. Desvio-padrão

Determina-se o desvio-padrão através da função **DESVPAD** [STDEV].

Medida da variabilidade de uma amostra, relativamente ao seu valor médio, correspondente à raiz quadrada do valor da variância e que é expressa nas mesmas unidades que os dados originais.

DESVPAD(sequência de números ou endereços de célula)

- Desloque o apontador para a célula **B9** e digite: **Desvio-padrão (cm)**.
- Desloque o apontador para a célula **C9** e introduza a função: **=DESVPAD(A2:A11)**.

| Microsoft Excel - Estatística com Excel.xls | | | |
|---|-------------|--------------------|----------|
| Estatística com Excel.xls | | | |
| =DESVPAD(A2:A11) | | | |
| A | B | C | D |
| 1 | Altura (cm) | | |
| 2 | 175 | Média (cm) | 173,2 |
| 3 | 170 | Moda (cm) | 175 |
| 4 | 185 | Mediana (cm) | 175 |
| 5 | 154 | 1º Quartil (cm) | 170,75 |
| 6 | 165 | 3º Quartil (cm) | 177,25 |
| 7 | 175 | Percentil 85% (cm) | 180,6 |
| 8 | 182 | Variância (cm2) | 77,26889 |
| 9 | 178 | Desvio-padrão (cm) | 8,79141 |



5.3. Amplitude

Determina-se a amplitude recorrendo às funções **MÁXIMO** [MAX] e **MÍNIMO** [MIN].

Medida da variabilidade de uma amostra, correspondente à diferença entre o valor máximo e o valor mínimo do conjunto de dados.

MÁXIMO(sequência de números ou endereços de célula)

MÍNIMO(sequência de números ou endereços de célula)

- Desloque o apontador para a célula **B10** e digite:
Amplitude(cm).

- Desloque o apontador para a célula **C9** e introduza a função: **=MÁXIMO(A2:A11)-MÍNIMO(A2:A11).**

| | A | B | C | D | E |
|----|-------------|------------------------------|----------|---|---|
| 1 | Altura (cm) | | | | |
| 2 | 175 | Média (cm) | 173,2 | | |
| 3 | 170 | Moda (cm) | 175 | | |
| 4 | 185 | Mediana (cm) | 175 | | |
| 5 | 154 | 1º Quartil (cm) | 170,75 | | |
| 6 | 165 | 3º Quartil (cm) | 177,25 | | |
| 7 | 175 | Percentil 85% (cm) | 180,6 | | |
| 8 | 182 | Variância (cm ²) | 77,28889 | | |
| 9 | 178 | Desvio-padrão (cm) | 8,79141 | | |
| 10 | 175 | Amplitude (cm) | 31 | | |
| 11 | 173 | | | | |

5.4. Amplitude inter-quartil

Determina-se a amplitude inter-quartil recorrendo à função **QUARTIL** [QUARTILE], descrita na secção 4.4.

Medida da variabilidade da amostra, correspondente à diferença entre os valores do terceiro e do primeiro quartis, dando-nos informação sobre a amplitude do intervalo em que se encontram 50% das observações centrais.

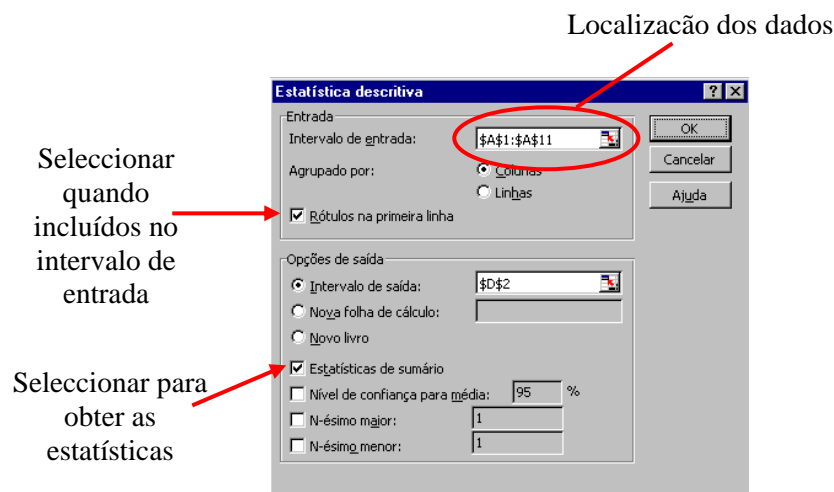
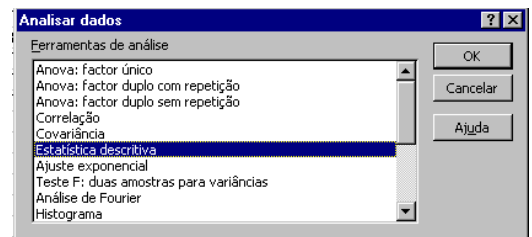
- Desloque o apontador para a célula **B11** e digite:
Amplitude inter-quartil(cm).

- Desloque o apontador para a célula **C9** e introduza a função:
=QUARTIL(A2:A11;3)- QUARTIL(A2:A11;1).

| | A | B | C | D | E |
|----|-------------|------------------------------|----------|---|---|
| 1 | Altura (cm) | | | | |
| 2 | 175 | Média (cm) | 173,2 | | |
| 3 | 170 | Moda (cm) | 175 | | |
| 4 | 185 | Mediana (cm) | 175 | | |
| 5 | 154 | 1º Quartil (cm) | 170,75 | | |
| 6 | 165 | 3º Quartil (cm) | 177,25 | | |
| 7 | 175 | Percentil 85% (cm) | 180,6 | | |
| 8 | 182 | Variância (cm ²) | 77,28889 | | |
| 9 | 178 | Desvio-padrão (cm) | 8,79141 | | |
| 10 | 175 | Amplitude (cm) | 31 | | |
| 11 | 173 | Amplitude inter-quartil (cm) | 6,5 | | |
| 12 | | | | | |



O Excel permite calcular as diferentes medidas de localização e de dispersão, de uma forma mais expedita, através do comando **Análise de dados** [*Data analysis*], presente no menu **Ferramentas** [*Tools*]. Ao aceder ao comando **Análise de dados** [*Data analysis*], seleccione a opção **Estatística descritiva** [*Descriptive statistics*] e clique **OK**.



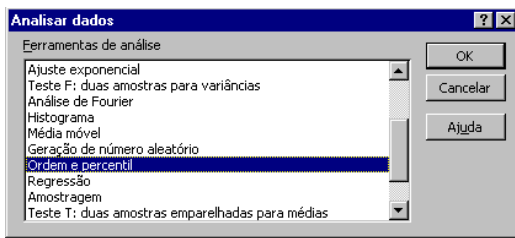
Após preenchimento da **Caixa de diálogo**, clique **OK**.

Para evidenciar os resultados deste comando, formatou-se a tabela dos resultados

Conforme se pode verificar, a lista de estatísticas fornecida pela opção **Estatística descritiva** [*Descriptive statistics*] inclui medidas não descritas anteriormente (ex: curtose, assimetria), mas não inclui os quantis (excepção para a mediana).

| | A | B | C | D | E |
|----|-------------|------------------------------|----------|----------------------|----------|
| 1 | Altura (cm) | | | | |
| 2 | 175 | Média (cm) | 173,2 | Altura (cm) | |
| 3 | 170 | Moda (cm) | 175 | | |
| 4 | 185 | Mediana (cm) | 175 | Média | 173,2 |
| 5 | 154 | 1º Quartil (cm) | 170,75 | Erro-padrão | 2,780088 |
| 6 | 165 | 3º Quartil (cm) | 177,25 | Mediana | 175 |
| 7 | 175 | Percentil 85% (cm) | 180,6 | Moda | 175 |
| 8 | 182 | Variância (cm ²) | 77,28889 | Desvio-padrão | 8,79141 |
| 9 | 178 | Desvio-padrão (cm) | 8,79141 | Variância da amostra | 77,28889 |
| 10 | 175 | Amplitude (cm) | 31 | Curtose | 1,742909 |
| 11 | 173 | Amplitude inter-quartil (cm) | 6,5 | Assimetria | -1,06484 |
| 12 | | | | Intervalo | 31 |
| 13 | | | | Mínimo | 154 |
| 14 | | | | Máximo | 185 |
| 15 | | | | Soma | 1732 |
| 16 | | | | Contagem | 10 |
| 17 | | | | | |

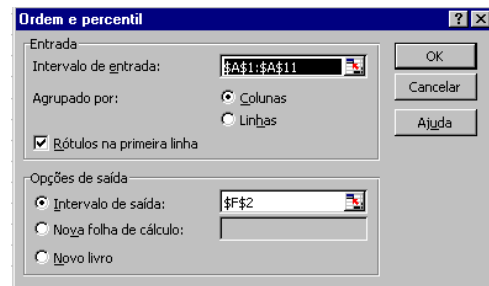




Pode-se obter a ordem e o quantil (percentil) de cada um dos dados apresentados, através da opção **Ordem e percentil** [*Rank and Percentile*],

do comando **Análise de dados** [*Data analysis*]

Após preenchimento da **Caixa de diálogo**, clique **OK**

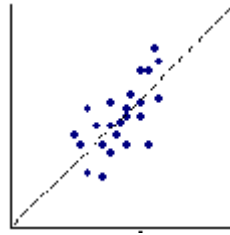


Microsoft Excel - Estatística com Excel.xls

Ficheiro Editar Ver Inserir Formatar Ferramentas Dados Janela Ajuda

| | B | C | D | E | F | G | H | I | |
|----|---|------------------------------|--------|----------------------|----------|-------------|-------|-------------|--------|
| 1 | | | | | | | | | |
| 2 | | Média (cm) | 173,2 | Altura (cm) | Ponto | Altura (cm) | Ordem | Percentagem | |
| 3 | | Moda (cm) | 175 | | 3 | 185 | 1 | 100,00% | |
| 4 | | Mediana (cm) | 175 | Média | 173,2 | 7 | 182 | 2 | 88,80% |
| 5 | | 1º Quartil (cm) | 170,75 | Erro-padrão | 2,780088 | 8 | 178 | 3 | 77,70% |
| 6 | | 3º Quartil (cm) | 177,25 | Mediana | 175 | 1 | 175 | 4 | 44,40% |
| 7 | | Percentil 85% (cm) | 180,6 | Moda | 175 | 6 | 175 | 4 | 44,40% |
| 8 | | Variância (cm2) | 77,289 | Desvio-padrão | 8,79141 | 9 | 175 | 4 | 44,40% |
| 9 | | Desvio-padrão (cm) | 8,7914 | Variância da amostra | 77,28889 | 10 | 173 | 7 | 33,30% |
| 10 | | Amplitude (cm) | 31 | Curtose | 1,742909 | 2 | 170 | 8 | 22,20% |
| 11 | | Amplitude inter-quartil (cm) | 6,5 | Assimetria | -1,06484 | 5 | 165 | 9 | 11,10% |
| 12 | | | | Intervalo | 31 | 4 | 154 | 10 | ,00% |
| 13 | | | | Mínimo | 154 | | | | |
| 14 | | | | Máximo | 185 | | | | |
| 15 | | | | Soma | 1732 | | | | |
| 16 | | | | Contagem | 10 | | | | |
| 17 | | | | | | | | | |





6. Dados bivariados

Para ilustrar o estudo de dados bivariados, recorreu-se ao exemplo apresentado no início do capítulo 4, tendo-se acrescentado o peso, em kg, para cada um dos 10 alunos

Par de valores correspondente a um dado indivíduo ou resultado experimental.

| | A | B |
|----|-----------|-------------|
| 1 | Peso (kg) | Altura (cm) |
| 2 | 72 | 175 |
| 3 | 65 | 170 |
| 4 | 80 | 185 |
| 5 | 57 | 154 |
| 6 | 60 | 165 |
| 7 | 77 | 175 |
| 8 | 83 | 182 |
| 9 | 79 | 178 |
| 10 | 67 | 175 |
| 11 | 68 | 173 |

descritos anteriormente: 72, 65, 80, 57, 60, 77, 83, 79, 67 e 68, respectivamente.

Repita o procedimento de introdução de dados. Comece por digitar na célula **A1** o título **Peso (kg)** e na célula **B1** o título **Altura (cm)**. Da célula **A2** à célula **A11**, digite cada um dos novos valores apresentados e da célula **B2** à célula **B11**, digite os correspondentes valores da altura. Grave o ficheiro para o disco e prepare-se para as novas funções!


6.1. Diagrama de dispersão

Pode-se obter com facilidade a representação gráfica de dados bivariados, através do **Assistente de Gráficos [Chart Wizard]**.

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.

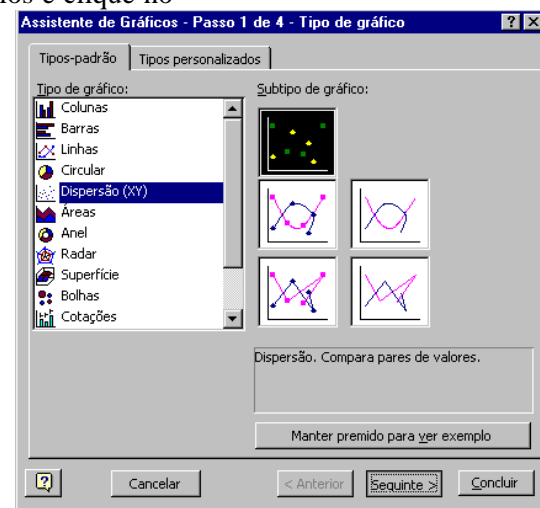
| | A | B |
|----|-----------|-------------|
| 1 | Peso (kg) | Altura (cm) |
| 2 | 72 | 175 |
| 3 | 65 | 170 |
| 4 | 80 | 185 |
| 5 | 57 | 154 |
| 6 | 60 | 165 |
| 7 | 77 | 175 |
| 8 | 83 | 182 |
| 9 | 79 | 178 |
| 10 | 67 | 175 |
| 11 | 68 | 173 |

Comece por seleccionar as células contendo os dados e os respectivos títulos e clique no

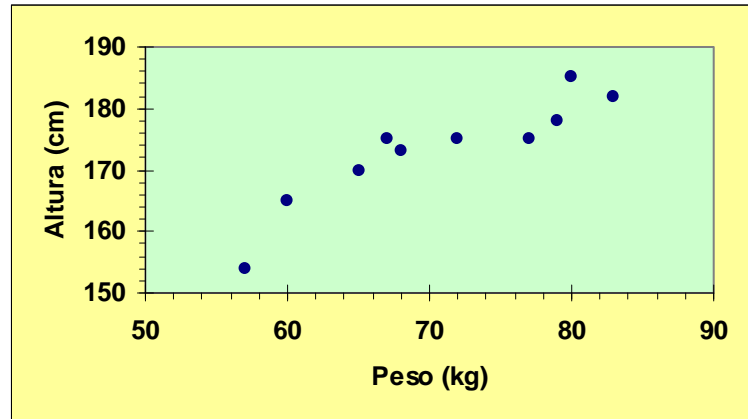
ícone  da **Barra de ferramentas**.

Na primeira **Caixa de diálogo** seleccione a opção **Dispersão (xy)** [XY

(Scatter)].



São múltiplas as opções de formatação para os gráficos de Excel, desde o aspecto geral, aos tipos de letras, à formatação dos eixos, etc. Eis um exemplo do que poderá obter.



6.2. Coeficiente de correlação

Quando, através do diagrama de dispersão, se verifica a existência de uma associação linear entre as variáveis, pode-se medir a maior ou menor força com que as variáveis se associam através do coeficiente de correlação, que no Excel corresponde à função **CORREL**[CORREL].

Medida do grau e do tipo (negativa ou positiva) de associação linear entre duas variáveis.

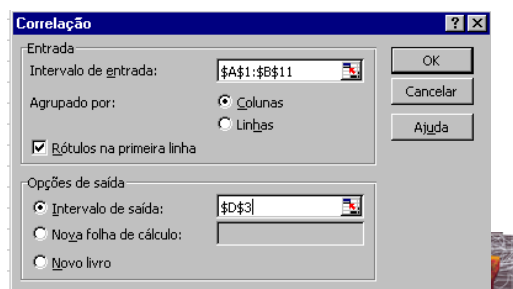
CORREL(sequência de números ou endereços de célula referentes à 1ª variável; sequência de números ou endereços de célula referentes à 2ª variável)

O resultado desta função é o coeficiente de correlação (r) entre os dois conjuntos de valores introduzidos como argumento:

$$r = \frac{s_{xy}}{\sqrt{s_{xx}} \sqrt{s_{yy}}}, \quad s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- Desloque o apontador para a célula **C3** e digite: **Correlação**
- Desloque o apontador para a célula **C4** e introduza a função:
=CORREL(A2:A11;B2:B11)

Alternativamente aceda ao comando **Análise de dados** [Data analysis] e seleccione a opção **Correlação** [Correlation]. Preencha a **Caixa de diálogo** e clique **OK**.



<http://alea-estp.ine.pt>

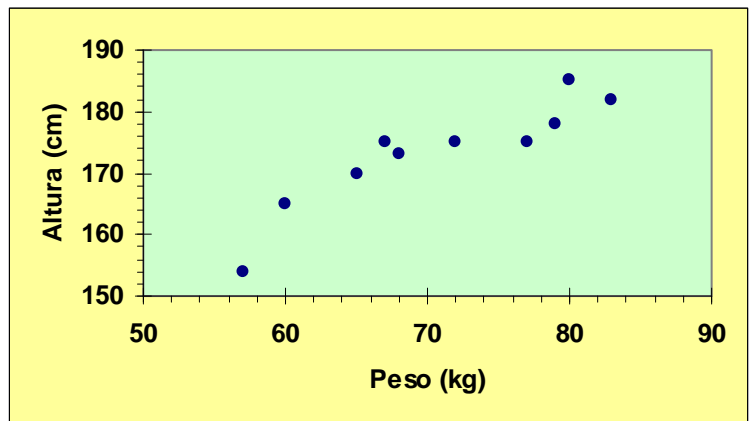
| | A | B | C | D | E |
|---|-----------|-------------|------------|---|---|
| 1 | Peso (kg) | Altura (cm) | | | |
| 2 | 72 | 175 | | | |
| 3 | 65 | 170 | Correlação | | |
| 4 | 80 | 185 | 0,90681871 | | |

| | A | B | C | D | E | F |
|----|-----------|-------------|------------|-------------|------------|---|
| 1 | Peso (kg) | Altura (cm) | | | | |
| 2 | 72 | 175 | Correlação | | | |
| 3 | 65 | 170 | 0,90681871 | | | |
| 4 | 80 | 185 | | Peso (kg) | 1 | |
| 5 | 57 | 154 | | Altura (cm) | 0,90681871 | 1 |
| 6 | 60 | 165 | | | | |
| 7 | 77 | 175 | | | | |
| 8 | 83 | 182 | | | | |
| 9 | 79 | 178 | | | | |
| 10 | 67 | 175 | | | | |
| 11 | 68 | 173 | | | | |
| 12 | | | | | | |
| 13 | | | | | | |

6.3. Recta de regressão

Quando dado um conjunto de dados bivariados, a sua representação gráfica sob a forma de um diagrama de dispersão sugerir uma relação linear entre as variáveis apresentadas, então é possível ajustar uma recta a esse conjunto de pontos.

Voltando ao exemplo das alturas e dos pesos dos alunos de Educação Física e ao seu diagrama de dispersão, pode-se observar uma associação linear entre o peso e a altura. **Será que é possível inferir a altura de um aluno que pese 70 kg?**



Perante uma situação análoga, em que tenhamos um conjunto de dados bivariados (x_i, y_i) , $i=1, \dots, n$, que seguem um padrão linear, vamos tentar ajustar uma recta de equação:

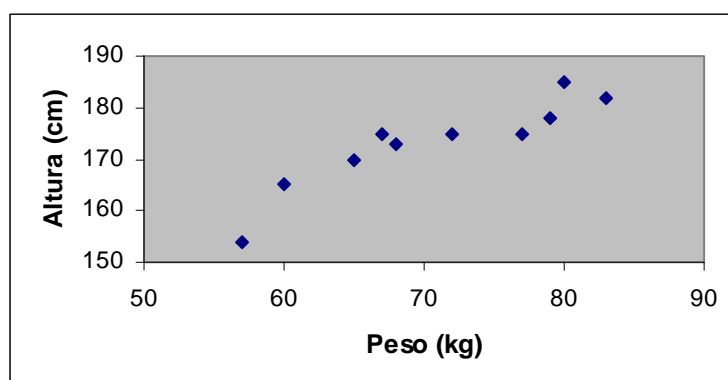
$$y = a + bx$$

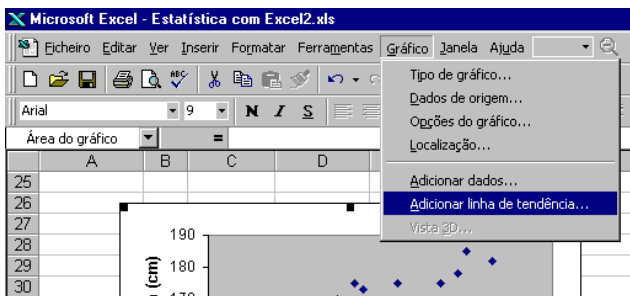
que permite descrever como se reflectem em y – variável resposta, as modificações processadas em x – variável explicativa.

Um dos métodos mais conhecidos de ajustar uma recta a um conjunto de dados é o método dos mínimos quadrados, que consiste em determinar a recta que minimiza, precisamente, a soma dos quadrados dos desvios $[y_i - (a+bx_i)]$ entre os verdadeiros valores de y e os obtidos a partir da recta que se pretende ajustar. A esta recta dá-se o nome de **recta de regressão ou dos mínimos quadrados**.

Vamos exemplificar de seguida a obtenção da recta de regressão com o Excel.

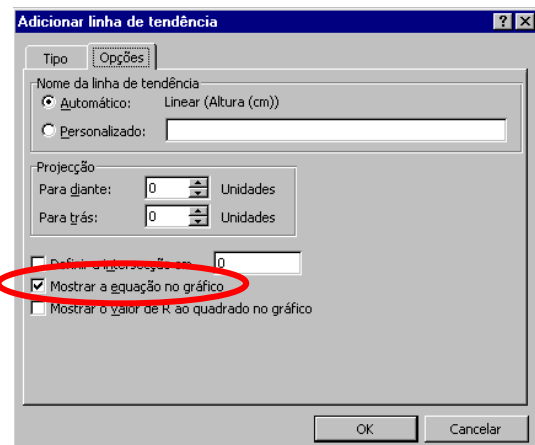
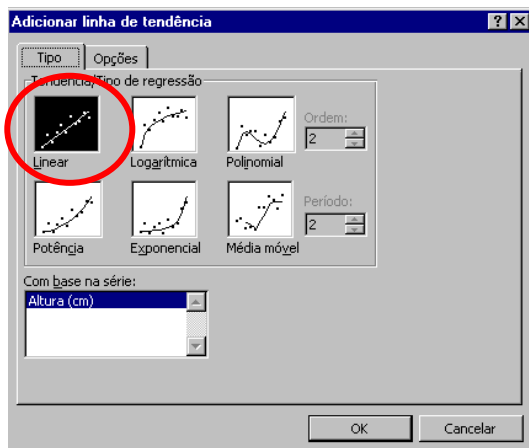
Construa novamente o diagrama de dispersão.





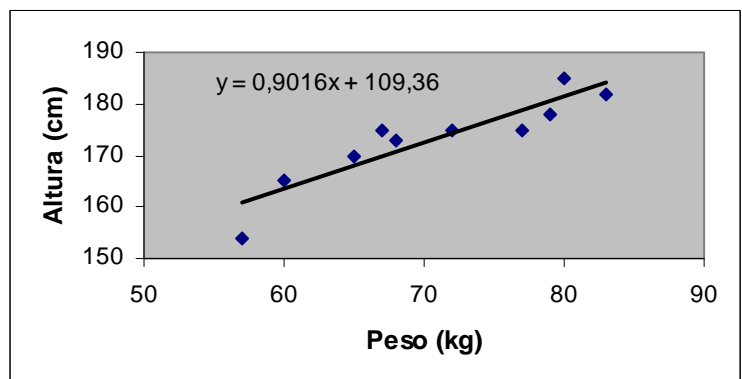
Seleccionando o diagrama, clique no menu **Gráfico** [Chart], seleccione o comando **Adicionar linha de tendência** [Add Trendline].

e siga as opções.



A equação desta recta traduz-se em:

$$\text{Altura} = 109,36 + 0,9016 \times \text{Peso}$$



Substituindo na equação o Peso por 70, obtém-se o valor de 172,472, pelo que a **altura esperada para um aluno que pese 70 kg , é de cerca de 172,5 cm.**

Para obter uma recta ajustada, não necessariamente a recta de regressão, pode-se começar por ajustar uma recta “a olho”, obtendo os coeficientes a partir de dois



pontos seleccionados de forma conveniente. Para verificar se esse ajustamento é razoável calcula-se a soma dos quadrados dos desvios.

Retomando o exemplo, descreve-se o processo a seguir, pelo que sugerimos que prepare uma tabela idêntica à que se apresenta. Para a recta ajustada considerámos os coeficientes $a=100$ e $b=1$.

Os valores do Ajuste (valores da recta ajustada para os x_i dados, isto é, os valores $100+x_i$), do Desvio ($y_i - \text{Ajuste}$) e do Desvio², poderão ser calculados com as seguintes expressões:

- Ajuste (y)

1º valor (célula E2)
= $\$A\$3+C2*\$A\6

Copie esta expressão para as células E3 a E11.

- Desvio

1º valor (célula F2)
= $D2-E2$


Copie esta expressão para as células F3 a F11.

- Desvio²

1º valor (célula G2)
= $F2^2$

Copie esta expressão para as células G3 a G11.

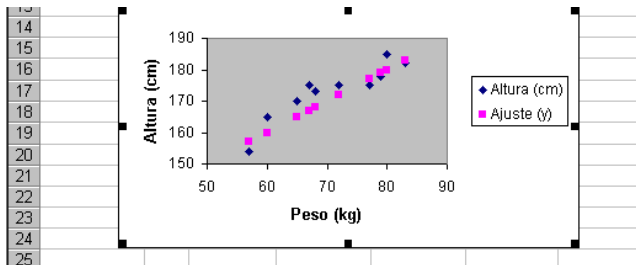
| | A | B | C | D | E | F | G |
|----|---------------|---|-----------|-------------|------------|--------|---------------------|
| 1 | | | Peso (kg) | Altura (cm) | Ajuste (y) | Desvio | Desvio ² |
| 2 | Constante (a) | | 72 | 175 | 172 | 3 | 9 |
| 3 | 100 | | 65 | 170 | 165 | 5 | 25 |
| 4 | | | 80 | 185 | 180 | 5 | 25 |
| 5 | Declive (b) | | 57 | 154 | 157 | -3 | 9 |
| 6 | 1 | | 60 | 165 | 160 | 5 | 25 |
| 7 | | | 77 | 175 | 177 | -2 | 4 |
| 8 | | | 83 | 182 | 183 | -1 | 1 |
| 9 | | | 79 | 178 | 179 | -1 | 1 |
| 10 | | | 67 | 175 | 167 | 8 | 64 |
| 11 | | | 68 | 173 | 168 | 5 | 25 |
| 12 | | | | | Soma | 24 | 188 |

Seleccione as células das três primeiras colunas contendo os dados e os respectivos títulos e clique no ícone  da Barra de ferramentas.

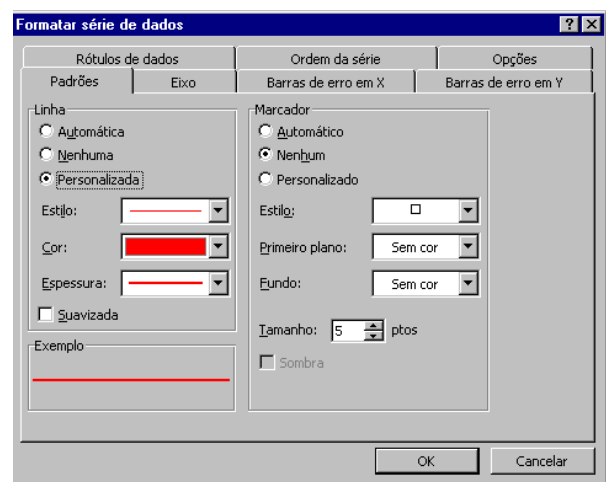
| | A | B | C | D | E | F | G |
|----|---------------|---|-----------|-------------|------------|--------|---------------------|
| 1 | | | Peso (kg) | Altura (cm) | Ajuste (y) | Desvio | Desvio ² |
| 2 | Constante (a) | | 72 | 175 | 172 | 3 | 9 |
| 3 | 100 | | 65 | 170 | 165 | 5 | 25 |
| 4 | | | 80 | 185 | 180 | 5 | 25 |
| 5 | Declive (b) | | 57 | 154 | 157 | -3 | 9 |
| 6 | 1 | | 60 | 165 | 160 | 5 | 25 |
| 7 | | | 77 | 175 | 177 | -2 | 4 |
| 8 | | | 83 | 182 | 183 | -1 | 1 |
| 9 | | | 79 | 178 | 179 | -1 | 1 |
| 10 | | | 67 | 175 | 167 | 8 | 64 |
| 11 | | | 68 | 173 | 168 | 5 | 25 |
| 12 | | | | | Soma | 24 | 188 |



Siga os procedimentos anteriormente descritos e construa um diagrama de dispersão.

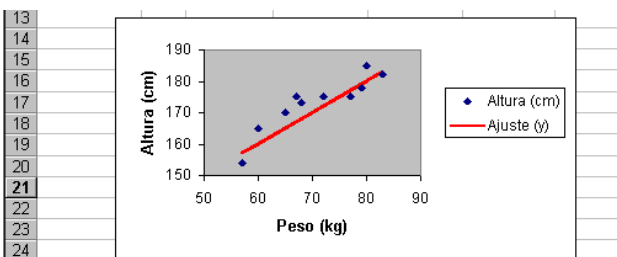


Selecione a série de dados correspondente ao "Ajuste (y)" e clique duas vezes, para abrir o menu **Formatar série de dados** [*Format Data Series*].



Na opção **Linha** [*Line*], personalize de acordo com o exemplo.

Na opção **Marcador** [*Marker*], selecione: **Nenhum** [*None*]

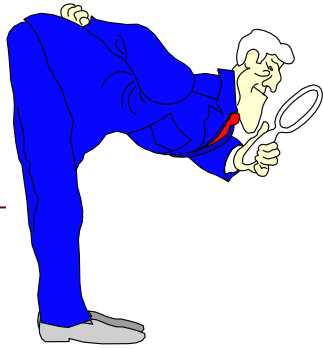


Experimente agora alterar os valores da "Constante(a)" e do "Declive (b)" e observe como se comporta a recta e a soma dos quadrados dos Desvios.

Verifique que para os valores de $a=109,36$ e $b=0,9016$, obtém uma soma dos quadrados dos desvios menor do que qualquer uma das obtidas anteriormente.

| | A | B | C | D | E | F | G |
|----|---------------|---|-----------|-------------|-----------|---------|---------------------|
| 1 | | | Peso (kg) | Altura (cm) | Ajuste(y) | Desvio | Desvio ² |
| 2 | Constante (a) | | 72 | 175 | 174,2752 | -0,7248 | 0,525335 |
| 3 | 109,36 | | 65 | 170 | 167,964 | -2,036 | 4,145296 |
| 4 | | | 80 | 185 | 181,488 | -3,512 | 12,33414 |
| 5 | Declive (b) | | 57 | 154 | 160,7512 | 6,7512 | 45,5787 |
| 6 | 0,9016 | | 60 | 165 | 163,456 | -1,544 | 2,383936 |
| 7 | | | 77 | 175 | 178,7832 | -3,7832 | 14,3126 |
| 8 | | | 83 | 182 | 184,1928 | -2,1928 | 4,808372 |
| 9 | | | 79 | 178 | 180,5864 | 2,5864 | 6,689465 |
| 10 | | | 67 | 175 | 169,7672 | -5,2328 | 27,3822 |
| 11 | | | 68 | 173 | 170,6688 | -2,3312 | 5,434103 |
| 12 | | | | | Soma | -0,0672 | 123,5945 |
| 13 | | | | | | | |





7. Ver Também...

Sobre **Estatística com Excel** existe um vasto conjunto de materiais disponíveis (artigos, revistas, outras publicações, *web sites*, etc). Destacamos os seguintes:

Publicações:

(Livros, algumas revistas com artigos de interesses)

- LEVINE, DAVID M.; BERENSON, MARK L. and STEPHAN, DAVID (1999): "**Statistics for Managers Using Microsoft Excel**", 2nd Ed., Prentice Hall, New Jersey, EUA
- DIAS CURTO, J.J. (1995): **Excel para Economia e Gestão**, Edições Sílabo, Lisboa
- GAENG, PETER (1993): "**Excel for Science and Technology**", Abacus, Grand Rapids, MI, EUA
- PELOSI, MARILYN K.; SANDIFER, THERESA M. and LETKOWSKI, JERZY J. (1998): "**Doing Statistics with Excel 97**", John Wiley & Sons, Inc. New York, USA.
- **Teaching Statistics** (encontra informação e cópia de alguns artigos desta revista no *web site*: <http://science.ntu.ac.uk/rsscse/ts/>)
- **Spreadsheet User** (encontra informação e cópia de alguns artigos desta revista no *web site*: <http://www.shu.ac.uk/mathssuser/>)



Web sites:**Discovering Important Statistical Concepts Using SpreadSheets**

<http://home.edu.coventry.ac.uk/Volume/>

http://www.mis.coventry.ac.uk/research/discus/discus_home.html



A página **DISCUSS** apresenta um conjunto de folhas de cálculo interactivas, preparadas em Excel, que foram desenhadas para o apoio ao ensino da Estatística no Ensino Secundário e/ou num curso Universitário introdutório.

The Association of Statistics Specialists Using Microsoft Excel

<http://www.mailbase.ac.uk/lists/assume/files/welcome.html>

A página da **ASSUME**, apresenta ligações para vários recursos relativos ao uso do Excel para a prática e o ensino da Estatística.

São vários os programas que suplementam as capacidades do Excel na Estatística, apresentam-se os nomes de alguns e os sites onde poderá obter informação detalhada e versões de demonstração dos mesmos.

- **PHStat Excel Add-In.** Software incluído no livro *Statistics for Managers Using Microsoft Excel*
<http://cw.prenhall.com/bookbind/pubbooks/levine/>
- **Analyse-it General statistics software for Microsoft Excel**
 **Analyse-it** <http://www.analyse-it.com/info/genstat.htm>
- **WinSTAT Statistics Add-In for Microsoft Excel**
 <http://www.winstat.com/>
- **XLSTAT an Excel Add-in for statistics and data analysis + several Excel utilities + XLControl**
<http://www.xlstat.com/>
- **XLStatistics Home Page - Excel, Statistics, Data Analysis**
<http://www.man.deakin.edu.au/rodneyc/XLStats.htm>

