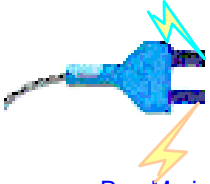


## N.º 11 – DADOS !!!???



Por: Maria Eugénia Graça Martins  
Departamento de Estatística e Investigação Operacional da FCUL  
memartins@fc.ul.pt

Esta ActivALEA tem como objectivo esclarecer algumas dúvidas que, por vezes, existem sobre a forma como se obtêm e o que são **dados!**

Antes de proceder à recolha de dados, temos que ter, à partida, um determinado objectivo:

Estudar algumas *características* comuns de um conjunto de *indivíduos* (ou objectos, ou “coisas”).



Em Deus confiamos.  
Todos os outros têm de trazer dados.

- Cada um dos *indivíduos* desse conjunto chama-se **unidade observacional** ou **caso**.
- À característica comum que se observa em cada unidade observacional ou caso, à qual se possa atribuir uma categoria ou um número, chama-se **variável**. O nome “variável” vem do facto de o resultado da observação da característica *poder variar* de indivíduo para indivíduo.
- Ao resultado da observação da variável em cada indivíduo chamamos **dado**.

### Exemplo 1

Supõe que estavas interessado em saber qual o ano de escolaridade dos alunos que frequentam uma Escola Básica do 2.º e 3.º ciclos, com 565 alunos. Como não tens tempo para fazer a pergunta a todos os alunos, seleccionas ao acaso 50 alunos a quem perguntas qual o ano que frequentam. Um conjunto de respostas possíveis poderia ser:

6.º 7.º 5.º 8.º 9.º 7.º 8.º 5.º 6.º 7.º 5.º 9.º 8.º 6.º 5.º  
 7.º 8.º 9.º 6.º 6.º 9.º 5.º 5.º 8.º 5.º 6.º 5.º 5.º 7.º 6.º  
 5.º 6.º 6.º 7.º 6.º 5.º 5.º 9.º 7.º 6.º 8.º 7.º 7.º 8.º 9.º  
 5.º 7.º 5.º 6.º 8.º

Algumas questões:

1. Qual a **unidade observacional** ou **indivíduo** objecto do estudo?
2. Qual a característica ou **variável** que se está a estudar (sobre o indivíduo objecto do estudo)? Essa variável é de tipo qualitativo ou quantitativo?
3. Do conhecimento que tens sobre a variável em estudo, quais as categorias possíveis que a variável pode assumir?
4. Que nome dás ao resultado da observação da variável sobre os alunos observados?
5. A partir do enunciado do exemplo identifica a população e a amostra.
6. No conjunto de dados considerado anteriormente, o 6.º aparece 12 vezes. Diz-se, então, que a **frequência absoluta** da categoria 6.º é 12. Calcula as frequências absolutas para as outras categorias e constrói uma tabela de frequências com essas frequências absolutas.
7. Junta uma nova coluna à tabela de frequências e preenche essa coluna com as frequências relativas das categorias. Para calculares as frequências relativas, basta dividires as frequências absolutas pelo total de elementos da tua amostra.
8. Constrói uma representação gráfica adequada para os dados.
9. A partir da tabela anterior ou da representação gráfica que construístes, responde às seguintes questões:
  - a) Nos dados que recolheste, concluis que a percentagem de alunos do 2.º ciclo é ligeiramente superior à do 3.º ciclo. Este resultado leva-te a esperar que, nos 565 alunos da escola, haja um pouco mais de alunos no 2.º ciclo do que no 3.º ciclo?
  - b) Qual o ano em que esperas que haja menos alunos, naquela escola? Justifica as respostas a estas questões.
10. Tinha sido fácil tirares as conclusões anteriores olhando para as 50 respostas que te deram no início deste exemplo?

## Exemplo 2

Supõe que foste a uma Escola Básica do 2.º e 3.º ciclos com 565 alunos e pediste na secretaria da escola que te informassem sobre quantos alunos frequentaram cada ano de escolaridade, num determinado ano lectivo à tua escolha. A secretaria forneceu-te os dados na seguinte tabela:

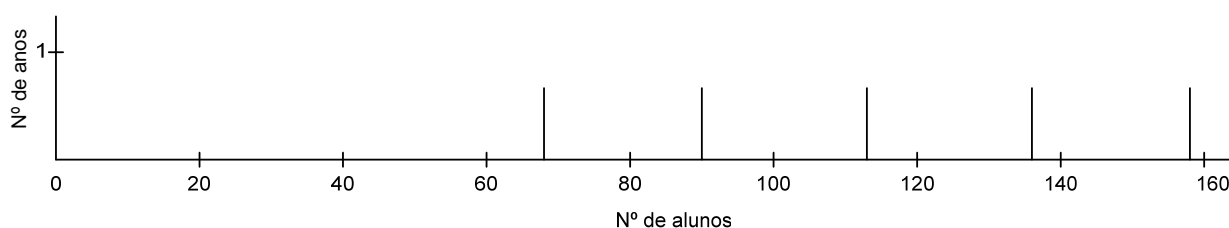
Ano	N.º de alunos
5.º	158
6.º	136
7.º	113
8.º	90
9.º	68

Toma atenção ao seguinte: a tabela anterior não é uma tabela de frequências! Naquela tabela a secretaria limitou-se a registar o valor observado pela *variável* que se está a estudar, que é o **número de alunos**, relativo à *unidade observacional* que é o **ano** (de escolaridade). Os valores que esta variável pode assumir são todos os inteiros, com um limite superior que é a capacidade da escola! A secretaria poderia ter dado a mesma informação da seguinte forma: O número de alunos que frequentaram o 5.º, 6.º, 7.º, 8.º ou 9.º ano foram, respectivamente, 158, 136, 113, 90 e 68.

Se pretendêssemos construir uma tabela de frequências, que neste caso não teria qualquer interesse estatístico, a tabela a construir seria a seguinte:

N.º de alunos	N.º de anos
68	1
90	1
113	1
136	1
158	1

Uma representação gráfica possível para os dados da tabela anterior é o seguinte gráfico de barras:



**Toma nota:**

Estes dois exemplos que acabámos de apresentar chamam a atenção para o cuidado que se tem de ter quando precisamos de recolher dados:

- Em primeiro lugar, é necessário identificar, sem origem para dúvidas, qual o **conjunto de indivíduos** (ou **objectos**, ou “**coisas**”) que se pretende observar;
- Em segundo lugar, é necessário identificar qual a característica ou **variável** que se pretende observar em cada um dos elementos do conjunto anterior; o resultado dessa observação é o **dado**.

Se estas etapas não ficarem muito bem definidas desde o princípio, pode ser complicado saber o que são os nossos dados.

- Em Estatística, os dados têm que ter um contexto.

Os dois exemplos anteriores exemplificam essa necessidade:

Repara que, no exemplo 1, o aluno era o objecto de estudo e a variável que se estava a estudar era o ano, pelo que os nossos dados são o resultado de observar qual é o ano de escolaridade de cada um dos 50 alunos, ou seja 6.º, 7.º, 5.º,..., 6.º, 8.º. Temos 50 dados porque observámos 50 alunos.

No caso do exemplo 2, o objecto de estudo é o ano e o que se pretende observar ou contar é o número de alunos em cada ano, pelo que o resultado dessa observação, que neste caso é uma contagem, é que é o dado. Os dados que obtivemos foram 68, 90, 113, 136 e 158. Só obtivemos 5 dados, porque só tínhamos 5 “coisas” a observar!

Resumindo:

<b>Exemplo</b>	<b>Unidade observacional</b>	<b>Variável</b>	<b>Dado</b>
1	Aluno	Ano	Por exemplo: 6.º
2	Ano	Número de alunos	Por exemplo: 68

Estes dois exemplos tiveram como objectivo mostrar que a mesma entidade tanto pode ser a unidade que se pretende observar como a variável que se está a observar sobre essa unidade. Depende do contexto em que estamos.

### **Exemplo 3**

Supõe que, no recreio da escola, viste um grupo de cinco alunos que estavam a jogar ao berlinde. Ficaste curioso em saber com quantos berlindes estavam a jogar, pelo que resolveste perguntar a cada um o número de berlindes que tinha e registar a informação recolhida na seguinte tabela, onde já tinhas escrito os nomes dos teus colegas:

<b>Nome</b>	<b>N.º de berlindes</b>
Miguel	5
Tiago	6
Pedro	5
Bernardo	4
Guilherme	7

Neste caso, o objecto do nosso estudo é o *aluno* jogador de berlinde, e a *variável* que se está a estudar sobre o aluno é o **número de berlindes**. Sobre os mesmos alunos, poderíamos estar interessados em estudar outras variáveis, como a **cor dos olhos**, o **número de irmãos** ou o **animal doméstico preferido**.

Este exemplo tem pouco interesse porque o número de dados é tão pequeno que não tem que ser sujeito a nenhum tratamento estatístico. Olhando para a tabela anterior concluis imediatamente que há 2 alunos com o mesmo número de berlindes (5), um que tem 4 berlindes, outro que tem 6 e, finalmente, que quem tem o maior número de berlindes tem 7. Obviamente, que não tem qualquer interesse estatístico estar a construir a seguinte tabela de frequências para agrupar os dados:

Tabela de frequências

<b>N.º de berlindes</b>	<b>N.º de alunos</b>
4	1
5	2
6	1
7	1

Só teria interesse se tivéssemos muitos mais casos a estudar!

### Exemplo 4

Supõe que estavas interessado em saber quantos alunos tinham, em média, as turmas da Escola Básica do 2.º e 3.º ciclos “Alegria”. Esta escola dispõe de uma lista com os nomes das turmas, como a que se apresenta a seguir:

5.ºA 5.ºB 5.ºC 5.ºD 5.ºE 5.ºF 5.ºG 5.ºH 5.ºI 5.ºJ 5.ºK 5.ºL 5.ºM 5.ºN 5.ºO 5.ºP 6.ºA  
6.ºB 6.ºC 6.ºD 6.ºE 6.ºF 6.ºG 6.ºH 6.ºI 6.ºJ 6.ºK 6.ºL 6.ºM 6.ºN 6.ºO 6.ºP 7.ºA 7.ºB  
7.ºC 7.ºD 7.ºE 7.ºF 7.ºG 8.ºA 8.ºB 8.ºC 8.ºD 8.ºE 8.ºF 8.ºG 9.ºA 9.ºB 9.ºC 9.ºD 9.ºE  
9.ºF

Como são muitas turmas e não tens tempo para obter a informação desejada sobre todas, decides seleccionar 17 turmas ao acaso e contar quantos alunos tem cada uma, registando os valores observados:

18, 18, 19, 20, 19, 18, 19, 20, 20, 21, 18, 19, 20, 23, 21, 20, 20

Responde às seguintes questões:

1. Neste estudo, quais são as unidades observacionais?
2. Qual a variável que se pretende estudar? Essa variável é de tipo qualitativo ou quantitativo?
3. Sobre as mesmas unidades observacionais podes estudar outra(s) variável(eis)? Dá um exemplo e diz se é de tipo qualitativo ou quantitativo.
4. Resume a informação dos 17 **dados** recolhidos, numa tabela de frequências.
5. A partir da tabela anterior, constrói um diagrama de barras. Na construção deste gráfico tiveste que ter algum cuidado especial?
6. Calcula a média dos teus dados.
7. Podes estimar qual o valor médio do número de alunos, por turma, da Escola “Alegria”? Justifica a tua resposta.

## Respostas

### Exemplo 1

1. A unidade observacional objecto do estudo é o *Aluno*.
2. A variável que se pretende “observar” sobre cada um dos alunos é o *Ano de escolaridade*. Esta variável, embora se exprima por números, é de tipo *qualitativo*, pois não se pode contar nem medir, mas unicamente observar, podendo assumir várias modalidades ou categorias.
3. As categorias possíveis são 5.º, 6.º, 7.º, 8.º ou 9.º.
4. Ao resultado da observação da variável sobre o indivíduo, chamamos *dado*.
5. A população é constituída pelos 565 alunos da escola e a amostra é constituída pelos 50 alunos que se vai observar. Ao resultado da observação da variável sobre os 50 indivíduos, também se dá o nome de amostra, pelo que podemos dizer que a nossa amostra é constituída pelos seguintes elementos:

6.º 7.º 5.º 8.º 9.º 7.º 8.º 5.º 6.º 7.º 5.º 9.º 8.º 6.º 5.º  
 7.º 8.º 9.º 6.º 6.º 9.º 5.º 5.º 8.º 5.º 6.º 5.º 5.º 7.º 6.º  
 5.º 6.º 6.º 7.º 6.º 5.º 5.º 9.º 7.º 6.º 8.º 7.º 7.º 8.º 9.º  
 5.º 7.º 5.º 6.º 8.º

6. A partir da amostra anterior, construímos a seguinte tabela de frequências:

Ano	Freq. Abs. ou N.º de alunos
5.º	14
6.º	12
7.º	10
8.º	8
9.º	6

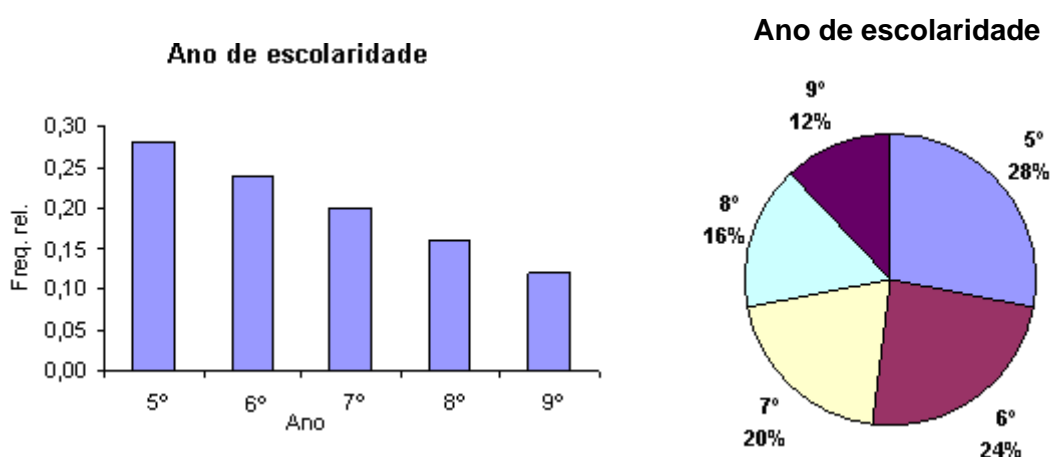
À tabela anterior é costume acrescentar uma linha onde colocamos o total das frequências absolutas, que devem somar 50, que é o número de elementos da amostra, a que também se dá o nome de *dimensão da amostra*:

Ano	Freq. Abs. ou N.º de alunos
5.º	14
6.º	12
7.º	10
8.º	8
9.º	6
Total	50

7. À tabela anterior, juntámos uma nova coluna com as frequências relativas:

Ano	Freq. Abs. ou N.º de alunos	Freq. Relativa
5.º	14	$14/50=0,28$ ou 28%
6.º	12	$12/50=0,24$ ou 24%
7.º	10	$10/50=0,20$ ou 20%
8.º	8	$8/50=0,16$ ou 16%
9.º	6	$6/50=0,12$ ou 12%
Total	50	1 ou 100%

8. Para representar graficamente os dados, podemos utilizar o gráfico de barras ou o diagrama circular:



Das duas representações gráficas anteriores, preferimos o gráfico de barras, uma vez que transmite melhor a informação sobre a diferença entre as frequências das diferentes categorias que a variável *Ano de escolaridade* assume.

9. a) Dos 50 alunos estudados, no 2.º ciclo estão 52%, ou seja, um pouco mais de metade dos alunos. Assim, esperamos que, na totalidade dos alunos, um pouco mais de metade seja do 2.º ciclo.
- b) Dos 50 alunos estudados, só 12% estão no 9.º ano de escolaridade, que é o ano com menos alunos. Assim, também esperamos que, na totalidade dos alunos da escola, o ano com menos alunos seja o 9.º.
10. Não tinha sido fácil, porque, ao olharmos para 50 números, não conseguiríamos retirar a informação que conseguimos retirar com os dados agrupados.

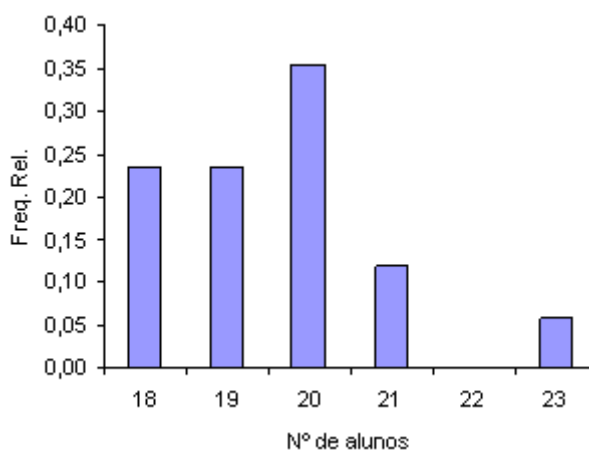
#### Exemplo 4

- As unidades observacionais são as turmas.
- A variável que se pretende estudar é o “Número de alunos da turma”. Esta variável é de tipo quantitativo discreto, já que o resultado da observação é uma contagem.
- Sobre a turma, poderíamos ainda estudar outra variável, o “Ano de escolaridade”, que é de tipo qualitativo.
- Vamos construir uma tabela com as frequências absolutas e relativas

Classes	Freq. Abs.	Freq. Rel.
18	4	0,235
19	4	0,235
20	6	0,353
21	2	0,118
23	1	0,059
Total	17	1



5. A partir da tabela anterior, construímos o seguinte diagrama de barras:



Para construir o gráfico anterior, tivemos de ter o cuidado de considerar o valor 22 com frequência nula. Embora este valor não exista na tabela, sempre que se construir um gráfico de barras é necessário considerar todas as classes, mesmo com frequência nula, desde que estejam entre classes de frequência não nula.

Esta chamada de atenção é particularmente importante quando se utiliza o Excel para construir o gráfico de barras (ver ActivALEA N.º7).

6. Para calcular a média dos dados, podemos utilizar o facto de estarem agrupados

$$\bar{x} = \frac{4 \times 18 + 4 \times 19 + 6 \times 20 + 2 \times 21 + 1 \times 23}{17} = 19,6$$

Chamamos a atenção para o facto de termos calculado a média com os dados agrupados, porque obtemos o mesmo valor que se a tivéssemos calculado com os dados sem estarem agrupados (costuma-se dizer “em bruto”). O mesmo não se passaria se fossem dados contínuos e estivessem agrupados em classes na forma de intervalos, pois neste caso obteríamos um valor aproximado para a média.

7. A partir da média obtida anteriormente, estimamos que o número médio de alunos por turma seja próximo dos 20 alunos.