

Dados são mais do que números

Maria Eugénia da Graça Martins

«Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write»

H. G. Wells

Introdução

Nos últimos anos tem-se assistido a um incremento de tópicos de probabilidade e estatística nos currículos de matemática, que pode ser entendido como o reconhecimento de uma profecia de H. G. Wells, que no fim do século 19 escrevia o apontamento com que iniciamos este texto.

Sendo um dos objectivos da escola, a preparação dos jovens para exercerem uma cidadania eficiente, é necessário dotá-los de meios para melhor e mais rapidamente compreenderem o mundo em que estão inseridos. A variabilidade está presente em todas as situações do mundo que nos rodeia, pelo que as conclusões que tiramos a partir da informação que se nos apresenta, na forma de *dados*, tem inerente um certo grau de incerteza. Por exemplo, é natural esperar que um pacote de açúcar que na embalagem tenha escrito um quilograma, não pese exactamente um quilograma e que ao pesar o mesmo pacote duas vezes não obtenhamos exactamente o mesmo valor. A Estatística trata e estuda esta variabilidade — permite-nos a partir dos *dados* retirar conclusões e exprimir o grau de confiança que devemos ter nessas conclusões.

Os imperativos das sociedades de hoje exigem que as pessoas sejam matemática, científica e tecnologicamente aptas para interagir com o mundo que as rodeia. O processo de democratização da matemática foi arrastado pela «quantização» da sociedade, que por outro lado nos conduziu à necessidade de novas ferramentas para melhor a interpretar e avaliar. A democratização do ensino fez com que a Matemática não seja mais uma disciplina para uma elite e tem de ser vista como fazendo parte da educação de qualquer cidadão. Vivendo numa sociedade de informação, não basta preocuparmo-nos com a literacia e a numeracia, sendo importante promovermos a literacia estatística, que abarca as competências necessárias para tratar a informação. Como refere Sheaffer (1986) «Teachers, and then students, must be trained to make intelligent decisions based on numerical information if our society is to grow and prosper» (p. 141).

Tal como foi importante para os nossos avós aprenderem a ler e a contar, faz parte da educação para a cidadania saber ler os números e os gráficos, com que somos confrontados no dia-a-dia (Martins et al, 2010, p. 7).

Estatística e Matemática

Podemos questionar-nos porque é que a Estatística surge, no ensino a nível pré-universitário, como fazendo parte do currículo de Matemática? Uma resposta possível e talvez demasiado simplista é avançada por Sheaffer (2005) «The mathematics community was well organized to make changes to their curriculum as society advanced to the information age and was open to accepting statistics as part of their charge». No entanto este acolhimento nem sempre tem sido pacífico já que, como ainda refere Sheaffer (2005) «Data analysis is about context; mathematics is about pattern and logic free of context. The data analyst must make a decision (inference) based on partial knowledge; a mathematician wants to prove (deduce) results based on a set of general principles. A data analyst is willing to make errors (the chance of which can often be measured), and the mathematician wants to reason without error».

Sendo consensual que a inclusão de Estatística no currículo escolar é fundamental, assistimos a uma reformulação dos currículos de Matemática, em que se procura realçar que um aspecto fundamental na literacia estatística, é compreender e usar o raciocínio estatístico, que é diferente do raciocínio matemático e que a educação estatística não se pode restringir a uma visão da Estatística, como um ramo da Matemática. Na verdade, enquanto o pensamento matemático se refere a relações entre conceitos abstractos, o pensamento estatístico tem sempre presente o contexto que dá origem aos *dados* que, por sua vez, permitem (ou não) responder a certas questões.

A Estatística tal como, por exemplo, a Física e a Economia, embora utilizando a Matemática, tem os seus conceitos e processos próprios. Algumas recomendações da American Statistical Association e da Mathematical American Association vão no sentido de que, no ensino da Estatística (Gal, 1997):

- A ênfase se deva colocar nos conceitos e no trabalho com *dados*, em detrimento de fórmulas e cálculos. Os *dados* devem ser recolhidos para responderem a questões e não unicamente para se ter um conjunto de *dados* para treinar fórmulas ou representações gráficas. A utilização de *dados* reais, relativos a situações interessantes para os alunos, tem a vantagem de os motivar, provocando a formulação de novas questões e eventualmente a recolha de novos *dados*. Pode haver situações em que seja útil a utilização de *dados* hipotéticos, nomeadamente em situações em que se procure explorar propriedades de alguns conceitos de Estatística, mas estes casos devem ser a excepção e não a regra;
- Muitos problemas estatísticos não têm uma solução matemática única. Pelo contrário, têm por base uma questão e terminam com a apresentação de uma opinião que pode ter diferentes graus de «razoabilidade» ...;
- O objectivo principal da educação estatística é tornar os estudantes capazes de exprimirem descrições fundamentadas, juízos, inferências e opiniões acerca de *dados*, ou argumentarem acerca da interpretação de *dados*, usando instrumentos matemáticos, no grau necessário;
- Os juízos e inferências que podemos esperar da parte dos alunos, a maior parte das vezes não poderão ser caracterizados como «verdadeiros» ou «falsos», tendo sobretudo de

ser avaliados em termos da qualidade dos raciocínios e da adequação dos métodos utilizados e da natureza dos *dados* e muitas vezes podem depender do conhecimento limitado que os alunos têm do mundo que os rodeia.

Não obstante esta diferença, a convivência entre a Estatística e a Matemática pode ser «saudável», nomeadamente para a Matemática. É comum os jovens questionarem-se nas aulas da disciplina de Matemática «Mas para que é que isto serve, onde vou utilizar isto?». A utilização, em Estatística, de *dados* reais, contextualizados, pode motivar os jovens na apreensão dos conceitos matemáticos: uma fracção ou percentagem pode ser a frequência relativa com que determinado clube de futebol ganha os jogos num campeonato, um somatório pode ser a média do número de mensagens de telemóvel enviadas pelos alunos de uma turma, uma recta pode ser uma recta de regressão, que explica a relação entre os as alturas e os pesos dos alunos de uma turma, etc.

Dados com contexto

Como se frisou na secção anterior, o mundo que nos rodeia será mais facilmente compreendido se puder ser quantificado. Em todas as áreas do conhecimento é necessário saber «o que medir» e «como medir». A Estatística é a ciência que ensina a recolher *dados* válidos, assim como a interpretá-los.

Perante um conjunto de *dados* podem-se distinguir duas metodologias de aproximação:

- por vezes o estatístico é confrontado com conjuntos de *dados* sem ter qualquer ideia preconcebida sobre o que é que vai encontrar e então procede a uma análise exploratória de *dados*, quase sempre utilizando processos gráficos, análise esta que revelará aspectos do comportamento dos *dados*; neste caso não se fala em *amostras*, mas sim *conjuntos de dados* (Murteira, 1993) e de uma maneira geral a análise exploratória é suficiente para os fins que se têm em vista;
- em outros casos, que é de um modo geral a situação de interesse em Estatística, sendo aquela em que nos posicionamos, procede à análise de *dados* com propósitos bem definidos no sentido de responder a questões específicas. Neste caso os *dados* têm que ser produzidos por meio de técnicas adequadas de forma a que resultem *dados* válidos (*amostras* representativas). Estas técnicas, em que é fundamental a intervenção do acaso, revolucionaram e fizeram progredir a maior parte dos campos da ciência aplicada. Pode-se dizer que hoje em dia não existe área do conhecimento para cujo progresso não tenha contribuído a Estatística.

Não é demais realçar a importância desta fase, a que chamamos de Produção ou Aquisição de *Dados*. Como é referido em Tannenbaum (1998, p. 426): «Behind every statistical statement there is a story, and like a story it has a beginning, middle, an end, and a moral. In this first statistics chapter we begin with the beginning, which in statistics typically means the process of gathering or collecting data. Data are the raw material of which statistical information is made, and in order to get good statistical information one needs good data». Consideremos, por analogia, o que se passa quando se pretende realizar um determinado cozinhado. Começa-se por seleccionar

Ano	Freq. absoluta	Freq. relativa
5°	14	0,28 ou 28%
6°	12	0,24 ou 24%
7°	10	0,20 ou 20%
8°	8	0,16 ou 16%
9°	6	0,12 ou 12%
Total	50	1 ou 100%

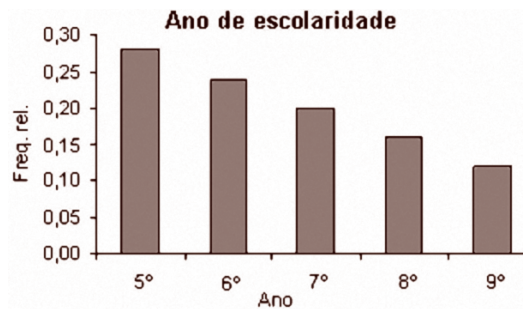


Figura 1. Tabela de frequências e diagrama de barras para a variável Ano de escolaridade

os ingredientes, que serão depois manipulados de acordo com determinada receita. O resultado do cozinhado pode ser um desastre, embora de aspecto agradável. Efectivamente, se os ingredientes não estiverem em condições, resulta um prato de aspecto semelhante ao que se obteria com ingredientes bons, mas de sabor intragável. O mesmo se passa com o procedimento estatístico. Se os dados não forem bons, embora se aplique a técnica correcta, o resultado pode ser desastroso, na medida em que se pode ser levado a retirar conclusões erradas.

Antes de começar a recolha de dados é fundamental, face a determinado problema, identificar correctamente a População sobre a qual se pretende recolher informação e fazer um planeamento de recolha da amostra, onde se decide quais e como devem ser recolhidos os dados. De um modo geral, o trabalho do estatístico deve começar antes de os dados serem recolhidos. Deve planear o modo de os recolher, de forma a que, posteriormente, se possa extrair o máximo de informação relevante para o problema em estudo, ou seja para a população de onde os dados foram recolhidos e de modo a que os resultados obtidos possam ser considerados válidos. Vem a propósito referir a célebre frase de R. A. Fisher: «Ao pedir a um estatístico que diagnostique dados já recolhidos, muitas vezes só se obtém uma autópsia» (Tradução livre de To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of (Presidential Address to the First Indian Statistical Congress, 1938, Calcutta).

Resumindo, existem alguns cuidados a 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.

Apresentamos a seguir algumas situações que concretizam o que acabamos de expor.

Exemplos

EXEMPLO 1.—Suponhamos que estávamos interessados em saber qual o ano de escolaridade predominante nos alunos que frequentam uma Escola Básica do 2.º e 3.º ciclos, com 565 alunos. Como não se tem tempo para fazer a pergunta a todos os alunos, selecciona-se, ao acaso, 50 alunos a quem se pergunta 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

Neste exemplo podemos identificar:

- *População*.—Conjunto dos 565 alunos da escola, pelo que a unidade observacional, objecto do estudo, é o aluno.
- *Característica ou variável em estudo*.—A característica que se está a observar sobre cada aluno é «Ano a que pertence», sendo portanto uma variável de tipo qualitativo.
- *Dados*.—São o resultado da observação sobre cada um dos 50 alunos, seleccionados para a amostra, do ano a que pertence.

Uma vez os dados recolhidos precede-se à sua organização na forma de uma tabela de frequências e de um diagrama de barras (figura 1).

Da tabela e do gráfico na figura 1 verificamos que predominam na amostra os alunos do 5.º ano de escolaridade, pelo que somos levados a «inferir» que na escola se verifica a mesma predominância.

EXEMPLO 2.—Suponhamos agora que fomos a uma Escola Básica do 2.º e 3.º ciclos com 565 alunos e pedimos na secretaria da escola que nos informassem sobre quantos alunos frequentaram cada ano de escolaridade, num determinado ano lectivo à nossa escolha, por exemplo 2010-2011.

Neste exemplo podemos identificar:

- *População*.—Conjunto dos 5 anos de escolaridade leccionados na escola, no ano lectivo 2010-2011, pelo que a unidade observacional, objecto do estudo, é o ano de escolaridade.
- *Característica ou Variável em estudo*.—A característica que se está a observar sobre cada ano é «Número de alunos que frequentaram o ano de escolaridade», sendo portanto uma variável de tipo quantitativo discreto.

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

Figura 2. Tabela com o número de alunos que frequentaram a escola no ano lectivo 2010–2011

Exemplo	Unidade observacional	Variável	Dado
1	Aluno	Ano de escolaridade	Por exemplo: 6.º
2	Ano	Número de alunos	Por exemplo: 68

Figura 3. Tabela para comparação dos Exemplos 1 e 2

- **Dados.**—São o resultado da observação sobre cada um dos 5 anos de escolaridade, do número de alunos.

A secretaria forneceu-nos os dados numa tabela (figura 2).

Repare-se que a tabela (figura 2) não é uma tabela de frequências! 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. 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, que são efectivamente os nossos dados.

Enquanto no exemplo 1, o aluno era o objecto de estudo e a variável que se estava a estudar era o ano de escolaridade, 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, obtendo 50 dados por termos observado 50 alunos, no caso do exemplo 2, o objecto de estudo é o ano de escolaridade e o que se pretende observar ou contar é o número de alunos em cada ano, pelo que o resultado dessa observação 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! A situação resume-se como indicado na figura 3.

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. A falta de uma identificação correcta do que é o dado, pode dar origem a situações que não se deveriam verificar, como as que se exemplificam a seguir, surgidas em livros de texto.

Confusões com dados

Uma situação que aparece frequentemente, é a confusão que se faz entre dado e frequência. Apresentamos um gráfico com barras (figura 4), que se retirou de um texto de Matemática:

De acordo com o título do gráfico, este representará o número de alunos das turmas do 5.º ano de determinada escola, onde ainda se retira a informação de quantos alunos são rapazes e quantos são raparigas em cada uma das turmas. Mas então, o que temos representado no gráfico são os próprios dados e não as frequências, como está indicado no eixo das ordenadas. O gráfico anterior, apesar de ser um gráfico com barras, não é um gráfico de barras.

Outro exemplo, onde também se faz uma confusão idêntica, é o que apresenta o gráfico (figura 5) introduzido da seguinte forma:

«O gráfico representa o número de peixes que cinco amigos pescaram num dia»

Pede-se para classificar a distribuição quanto à moda, no pressuposto de que a «distribuição» é bimodal, confundindo-se dado com frequência. Por definição, a moda é o(s) valor(es) que surge(m) com maior frequência. Então teremos de ver quais os nossos dados, que são os resultados da observação do «número de peixes» que cada um dos cinco amigos pescou e que são 30, 20, 25, 25 e 30. Temos uma variável quantitativa discreta, cujas observações podem ser organizadas numa tabela (figura 6) de frequências (sem grande interesse estatístico devido ao diminuto número de dados).

Agora podemos concluir que a distribuição dos dados é efectivamente bimodal, pois apresenta duas modas.

Figura 4. Gráfico apresentando o número de alunos das turmas A, B, C, D, E e F, discriminados por sexo

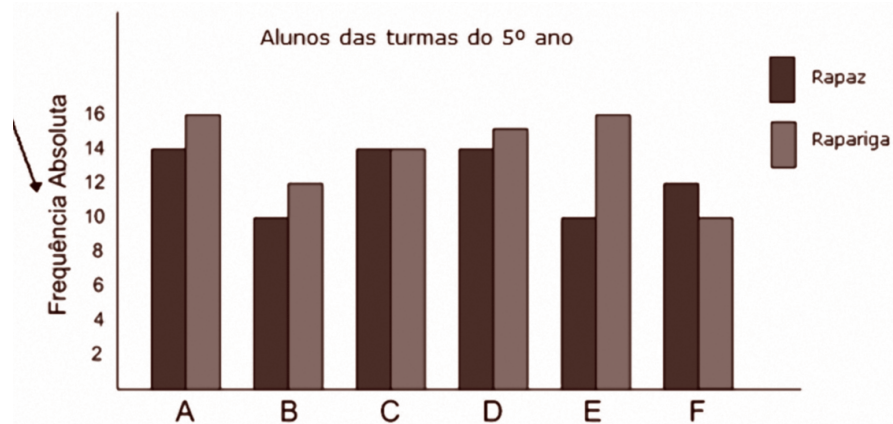




Figura 5. Gráfico apresentando os dados relativos à variável Número de peixes que cada um de cinco amigos pescaram num dia

Classe	Freq. absoluta
20	1
25	2
30	2

Figura 6. Tabela de frequências para a variável Número de peixes que cada um de cinco amigos pescaram num dia

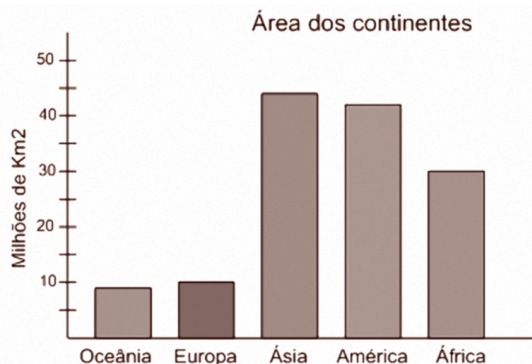


Figura 7. Gráfico apresentando as áreas dos continentes

Continente	Área
Oceânia	9
Europa	10
Ásia	44
América	42
África	30

Figura 8. Tabela apresentando as áreas dos continentes

Vejamos ainda o seguinte exemplo que também foi seleccionado de um texto de matemática, onde se lê:

«O gráfico (figura 7) representa as áreas dos continentes. Constrói uma tabela de frequências absolutas»

A solução apresentada é a tabela na figura 8.

Neste caso a unidade observacional é o continente e a variável objecto de estudo é a «Área de um continente». Assim os nossos dados são 9 km², 10 km², 44 km², 42 km² e 30 km², pelo que nem o gráfico apresentado é um gráfico de barras, nem a tabela é uma tabela de frequências.

Conclusão

Os exemplos anteriores mostram a relevância de termos sempre presente o que se está a estudar e o que são os nossos dados. Caso contrário corremos o risco de por em causa o ditado segundo o qual «Um gráfico vale mais que mil palavras». Mesmo o gráfico mais simples pode induzir em erro, se não soubermos avaliar qual a informação que se pretende transmitir!

Referências

- GAISE (2005). *Guidelines for assessment and instruction in statistics education report. A pre-k-12 curriculum framework*. Available online at www.amstat.org/education/gaise
- Gal, I. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1-13). IOS press (on behalf of the isi) (<http://www.stat.auckland.ac.nz/~iase/publications/assessbkref>)
- Graça Martins, M. E., & Cerveira, A. (1999). *Introdução às Probabilidades e à Estatística*. Lisboa: Universidade Aberta.

- Graça Martins, M. E. (2005). *Introdução à Probabilidade e à Estatística — Com complementos de Excel*. Lisboa: SPE.
- Graça Martins, M. E., Loura, I., & Mendes, F. (2007). *Análise de dados — texto de apoio para os professores do 1º ciclo*. Lisboa: DGIDC/ME.
- Graça Martins, M. E., & Ponte, J. P. (2010). *Organização e tratamento de dados*. Lisboa: DGIDC/ME. Disponível em http://area.dgidc.min-edu.pt/materiais_npmeb/matematicaotd_final.pdf
- Holmes, P. (2002). Some lessons to be learned from curricular developments in statistics. In *Proceedings of the International Conference on Teaching Statistics, ICOTS 6*, Cape Town, South Africa.
- Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistics Review*, 65 (2), 123-165.
- Sheaffer, R. (2005). In a world of data, Statistics counts. Disponível em: http://apcentral.collegeboard.com/apc/members/courses/teachers_corner/22510.html
- Sheaffer, R. (1986). Statistics and Probability in the school mathematics curriculum: A review of the ASA-NCTM Quantitative Literacy Project. In *Proceedings of the International Conference on Teaching Statistics, ICOTS 2*, Victoria, Canada.
- Sheaffer, R. (1990). The ASA-NCTM Quantitative Literacy Project: An overview. In *Proceedings of the International Conference on Teaching Statistics, ICOTS 3*, Dunedin, New Zealand.
- Tannenbaum, P., & Arnold, R. (1998). *Excursions in modern mathematics*. New Jersey: Prentice Hall.
- Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). IOS press (on behalf of the isi) (<http://www.stat.auckland.ac.nz/~iase/publications/assessbkref>)

Maria Eugénia Graça Martins
Faculdade de Ciências da Universidade de Lisboa