

Literacia estatística

João Branco
Maria Eugénia Graça Martins

Pensar à maneira da Estatística será um dia tão necessário para o cidadão eficiente como a habilidade de ler e escrever

G. Wells

Introdução

Quando, há algum tempo atrás, a palavra literacia surge no nosso vocabulário, embora não necessariamente no nosso dicionário, o seu significado é fundamentalmente o seguinte: capacidade do indivíduo para ler, escrever e falar na sua língua materna, efectuar cálculos e resolver problemas do dia-a-dia, de forma a cumprir as tarefas que lhe são exigidas tanto no emprego como na sociedade.

Numa sociedade em transformação e desenvolvimento, cada vez mais exigente com o cidadão, que vive constantemente exposto a grandes massas de informação, é natural que o conceito de literacia também tenha evoluído. No estudo internacional PISA (**P**rogramme for **I**nternational **S**tudent **A**ssessment), levado a cabo em 29 países da OCDE, considerado o maior estudo sobre as competências dos alunos que terminam a escolaridade obrigatória e de que resultou o volume *Measuring Student Knowledge and Skills: The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*, o conceito de literacia aparece de forma mais abrangente e mais exigente, destacando já três vertentes específicas (literacia em leitura, literacia matemática e literacia científica) que, citando o dito estudo, têm as definições que a seguir se apresentam:

Literacia em Leitura — A capacidade de compreender, usar e reflectir sobre textos escritos, com o fim de atingir os nossos objectivos, desenvolver conhecimentos e potencialidades, e participar na sociedade.

Literacia Matemática — A capacidade do indivíduo identificar, compreender, e de se ocupar da Matemática, de ter opiniões bem fundamentadas sobre o papel que a Matemática desempenha, como se torna necessário na sua vida presente e futura, na vida profissional, na vida social com os seus pares e familiares, para viver como um cidadão construtivo, interessado e ponderado.

Literacia Científica — A capacidade de usar conhecimentos científicos, de identificar problemas e de tirar conclusões baseadas em evidências para compreender e tomar decisões sobre o mundo natural e as mudanças que lhe são impostas pela actividade humana.

Vários autores, ver Steen (1997, 2001), falam de literacia quantitativa, também designada por numeracia, e uma definição em voga (Steen, 2001) é:

Literacia Quantitativa — Um conjunto de competências, conhecimentos, convicções e predisposições, hábitos mentais, capacidades de comunicação e jeito para resolver problemas que as pessoas precisam para enfrentar de maneira eficaz situações envolvendo

quantidades que surgem na vida e na actividade profissional.

Para que se perceba bem o que é literacia quantitativa há o cuidado de fazer a distinção entre literacia quantitativa e Matemática, a Matemática que se ensina nos cursos tradicionais. Esta é uma disciplina, com um programa, cujo objectivo é a aplicação de ideias abstractas ao estudo da relação entre objectos ideais. A literacia quantitativa ocupa-se de problemas concretos relativos a objectos ou acontecimentos reais que surgem em contextos determinados. A literacia quantitativa dá ao cidadão a capacidade de interpretar informação quantitativa de natureza muito diversificada, o que é hoje uma necessidade permanente para a tomada de decisões correctas em praticamente todas as actividades da vida corrente. Trata-se mais de uma linguagem do que uma disciplina. Representa um novo tipo de formação e por isso é natural que outros métodos de ensino e aprendizagem, que não os tradicionais, sejam mais adequados para se conseguirem os objectivos para que ela aponta. A literacia quantitativa não dispensa naturalmente conhecimentos de matemática, e muito menos dispensa a Estatística, aquela parte que se ocupa dos problemas ligados a situações de incerteza. Contudo não parece ser com programas (de matemática ou de estatística) mais vastos ou mais exigentes que o ensino tradi-

Unidade = ano

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

Tabela 1.

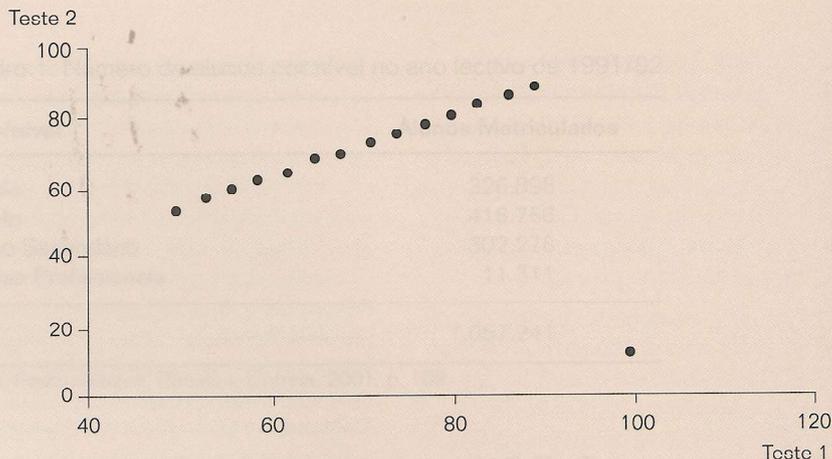


Gráfico 1.

cional leva o estudante a melhorar a sua literacia quantitativa.

O progressivo desenvolvimento da Estatística e a crescente necessidade de conhecimentos estatísticos para enfrentar situações da vida real, levaram à introdução da literacia estatística, à semelhança do que aconteceu com a literacia matemática, exigida por uma “quantização” cada vez mais acentuada da sociedade. Como é referido em Moore (1997), Anne Hawkins define assim a ideia de literacia estatística:

Na sua expressão mais simples, literacia estatística pode ser interpretada como uma habilidade de interagir eficazmente num ambiente de incerteza (não determinístico).

Uma interpretação vaga, mas na qual faz sentido incluir a situação mais concreta que é o frequente contacto com dados e a necessidade da sua análise.

Um aspecto fundamental na literacia estatística é compreender e usar o raciocínio estatístico. Note-se que o tipo de raciocínio estatístico é diferente do raciocínio matemático e a educação estatística não se pode restringir a uma visão da estatística simplesmente como um ramo da matemática (Vere-Jones, 1995). O tipo de raciocínio matemático, eminentemente um raciocínio lógico, em que as proposições ou são verdadeiras ou falsas, não é compatível com o tipo de raciocínio estatístico, em que tratamos com proposições que não podemos dizer que são verdadeiras nem tão pouco falsas, estando numa situação

de incerteza, que pode ser quantificada através da probabilidade:

Verdadeiro?

Incerteza

Falso?

Esta situação de incerteza acompanha-nos no nosso dia-a-dia, nas mais variadas situações.

A educação estatística tem uma dimensão diferente das áreas normalmente consideradas como ramos da Matemática, como por exemplo a Geometria, a Análise e a Álgebra, pelo seu envolvimento directo com o estudo de outras ciências como as ciências médicas e afins, ciências políticas e ciências sociais. É importante ensinar um médico, um sociólogo, um técnico da indústria farmacêutica e todos aqueles que fazem uso da Estatística a utilizá-la correctamente. A utilização incorrecta desta ciência pode levar a decisões erradas com consequências negativas quer para o desenvolvimento das outras ciências quer para o desenrolar da vida do cidadão comum. E, como refere Chatfield (1991), em Estatística é possível cometer erros com maior frequência do que em outras ciências, especialmente pelos não especialistas. Em seguida apresentam-se alguns casos de análises estatísticas que podem levar a interpretações e decisões incorrectas quando não se conhecem bem os conceitos estatísticos.

Reflexo de iliteracia estatística

Esta preocupação com a educação estatística, tem levado à introdução de alguns conceitos básicos de Esta-

tística e Probabilidade no ensino obrigatório e pré-universitário de alguns países, nos quais se inclui Portugal. Não nos iludamos, no entanto, com as facilidades por vezes apregoadas de que estas noções são meras questões de “bom senso” ou do “senso comum” que não trazem nada de novo e que não precisam de ser ensinadas. O certo é que elas são necessárias ao cidadão comum na condução da sua actividade diária e o seu desconhecimento pode acarretar graves inconvenientes e prejuízos. Por isso não estamos de acordo com aquela corrente simplista e desactualizada e apresentamos a seguir algumas situações simples, mas que surgem com demasiada frequência para serem ignoradas e às quais é preciso responder com sabedoria.

A média enganadora

A média é largamente utilizada para sintetizar a informação contida num conjunto de dados. Tratando-se de uma redução tão drástica, é necessário acautelar as situações em que a informação que ela transmite não tem qualquer utilidade ou é falsa. O exemplo que se segue é ilustrativo. Numa região começaram a aparecer pessoas com uma doença desconhecida, tendo os médicos do centro de saúde recolhido informação sobre 35 desses doentes, escolhidos aleatoriamente, e concluído que a média das idades era 32 anos. Conjecturou-se que se tratava de uma doença atacando os adultos jovens. Um médico mais curioso, sabedor que a média nem sempre é uma boa medida para resumir a informação contida nos dados, pediu que

País	tmv	pes/TV
Angola	44	200
Austrália	76.5	2
Cambodja	49.5	177
Canadá	76.5	1.7
China	70	8
Egipto	60.5	15
França	78	2.6
Haiti	53.5	234
Iraque	67	18
Japão	79	1.8
Madagascar	52.5	92
México	72	6.6
Marrocos	64.5	21
Paquistão	56.5	73
Rússia	69	3.2
África do Sul	64	11
Sri Lanca	71.5	28
Uganda	51	191
Reino Unido	76	3
Est. Unidos	75.5	1.3
Vietnam	65	29
Yemen	50	38

Tabela 2.

Ihe fornecessem as idades dos 35 doentes seleccionados, com os quais construiu a seguinte representação em caule-e-folhas:

Através da representação da Tabela 1, foi possível concluir que afinal a doença estava a atacar as crianças e as pessoas da terceira idade.

A utilização incorrecta do coeficiente de correlação

O coeficiente de correlação é largamente utilizado, nomeadamente na comunicação social, para exprimir o maior ou menor grau de associação entre duas variáveis. Nem sempre o uso do coeficiente de correlação é feito de forma correcta, sobretudo se não forem tomadas certas precauções. Veja-se o seguinte exemplo. Um professor decidiu registar as notas que os seus alunos tinham tido em dois testes, para averiguar se se teria verificado consistência entre os resultados dos dois testes, no sentido que um aluno que tenha tido boa (má) nota no primeiro teste, também tenha tido boa (má) nota no segundo teste. Calculou o coeficiente de correlação e ficou desapontado com o valor obtido, 0.04! Resolveu fazer a representação

gráfica dos dados, sob a forma de um diagrama de dispersão e obteve o Gráfico 1.

A representação mostra uma associação linear, quase perfeita, entre os dados, havendo um único valor a fugir desse padrão. Se for retirado o elemento discrepante que aparece no gráfico, já o coeficiente de correlação assume o valor 0.9997.

O exemplo anterior chama a atenção para alguns problemas que podem surgir quando a interpretação do coeficiente de correlação não é acompanhada de uma representação prévia dos dados.

Frequentemente também se esquece que o que o coeficiente de correlação mede é o grau de associação linear entre duas variáveis pelo que, perante um valor deste coeficiente perto de zero, haverá tendência para dizer que as variáveis não se associam, quando na realidade pode existir uma forte associação não linear.

Por outro lado, ao detectar associação entre duas variáveis, nem sempre se toma o devido cuidado com a interpretação que se dá a esta associação. Efectivamente, nem sempre a existência de associação entre duas variáveis significa uma relação de *causa-efeito*. Pode haver outras variáveis, relacionadas com as variáveis em estudo, o que acontece com frequência, que provoquem essa associação, como se exemplifica a seguir.

Para um conjunto de 22 países registou-se o número de pessoas por aparelho de televisão (pes/TV), assim como o tempo médio de vida (tmv), tendo-se obtido os valores que se apresentam na Tabela 2 (Rossmann e Chance, 2001).

A representação dos pontos de coordenadas (Pes/TV, tmv) num diagrama de dispersão permite-nos concluir da existência de uma associação linear negativa, com alguma intensidade, isto é, existe tendência para que quanto menor for o número de pessoas por aparelho de TV, maior será o tempo médio de vida. Só por graça é que se poderia dizer que um modo de aumentar o tempo médio de vida, seria aumentar o número de aparelhos de TV! É evidente que a associação negativa encontrada se deve

à presença de uma terceira variável, que podemos denominar por "nível de vida", que influencia as variáveis observadas.

Assim, uma regra básica a ter em linha de conta, quando se trabalha com o coeficiente de correlação ou a recta de regressão, é efectuar a representação prévia dos dados, num diagrama de dispersão.

O gráfico com eixos inapropriados

Se é bem verdade que um gráfico vale mais do que mil palavras, nem sempre esta "máxima" deve ser seguida, pois podemos estar perante gráficos enganadores. É uma situação que se verifica, nomeadamente, quando os eixos desses gráficos não são escolhidos convenientemente, quer devido a uma má escolha das escalas num ou mais eixos, quer devido à truncatura do eixo das frequências, isto é, fixando o início da escala nesse eixo num valor superior a zero. Os dois exemplos que se seguem esclarecem estes dois problemas.

Suponha que o número de acidentes, por mês, no IP5, foi, no período de Setembro de 1997 a Janeiro de 1998, o seguinte: 8, 9, 12, 13 e 12. Dois jornais hipotéticos apresentaram representações gráficas para transmitir a informação anterior (Gráficos 2 e 3).

Como comentário, podemos dizer que um dos jornais tentaria dramatizar o problema.

O segundo exemplo (Gráfico 4.) refere-se ao resultado de uma sondagem relativa às recentes eleições para o novo líder da Juventude Socialista. Os resultados da sondagem (Jornal Expresso N° 1547 de 22 de Junho de 2002) indicam 215 votos (51%) para Jamila Madeira e 208 votos (49%) para Filipe Costa. A notícia relativa a este evento, propositadamente intitulada "Ilusão de óptica", apresenta um gráfico, semelhante ao que se vê na página seguinte, que faz explodir uma diferença muito reduzida (7 votos, cerca de 2%), entre os desempenhos dos dois candidatos, numa vitória (ou derrota) verdadeiramente impressionante.

O próprio jornalista justifica assim a habilidade do gráfico em transmitir informação deturpada: "O truque para

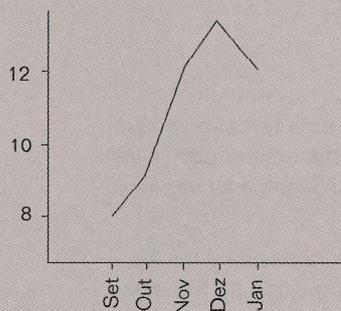


Gráfico 2.

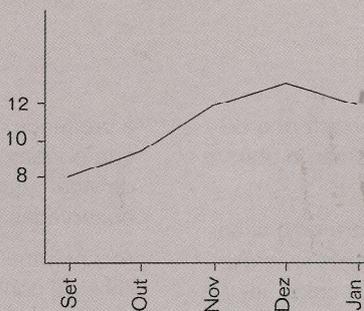


Gráfico 3.

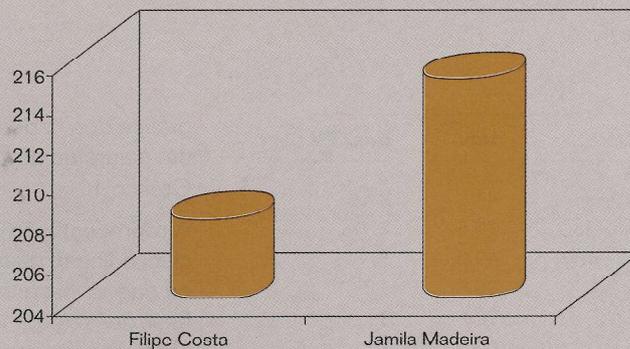


Gráfico 4.

que o resultado de Jamila apareça com o dobro de tamanho da coluna de Filipe foi utilizar, não as percentagens (51% a 49%), mas o número de respostas (215 a 208), e ao desenhar o gráfico, não começar a partir do 0, mas do 204, mostrando apenas o topo da votação. Independências?!!”

Na verdade o truque está, unicamente, na escolha da escala e não no facto de se utilizarem frequências absolutas ou relativas.

Alguns problemas com o cálculo de probabilidades

Assim como se podem cometer erros básicos em Estatística, o mesmo acontece em Probabilidades. Os exemplos seguintes pretendem ilustrar situações probabilísticas em que é comum isso acontecer, já que:

- A intuição é muitas vezes enganadora;
- Em muitas situações uma análise correcta depende da identificação de resultados igualmente possíveis (prováveis ...), o que nem sempre é fácil.

Se perguntar numa turma de alunos qual das sequências MFFMFM, MMMMFM, é mais provável de ocorrer, no nascimento de 6 crianças, onde representamos por M o nascimento de rapaz e F de rapariga e admitimos igual probabilidade para o nascimento de rapaz e rapariga, terá dúvida de que a esmagadora maioria responde MFFMFM? No entanto os dois acontecimentos têm a mesma probabilidade, $1/2^6$.

Se repetir a experiência que consiste em lançar ao ar 6 moedas e deixar cair as moedas sobre uma mesa, experimente perguntar numa turma de alunos o que é mais provável obter:

a) 2 caras e 4 coroas ($15/2^6$)

b) 3 caras e 3 coroas ($20/2^6$)

c) 5 caras e 1 coroa ($6/2^6$)?

Provavelmente a maior parte dos alunos escolheria a resposta correcta b), mas sem ser pela razão certa!

Se numa turma com 30 alunos encontrar 2 alunos a fazer anos no mesmo dia, poderá pensar tratar-se de uma rara coincidência. Efectivamente um resultado que vai contra a intuição é que bastam 23 pessoas para que a probabilidade de haver pelo menos duas a fazer anos no mesmo dia seja superior a 50% (Graça Martins *et al*, 1999a)!

Atente-se na seguinte notícia (Rasfeld, 2001): “Nestes últimos meses, milhares de crianças americanas têm estado a escrever cartas para os soldados americanos estacionados no Golfo Pérsico, seus desconhecidos, para lhes mostrar que eles não foram esquecidos no seu país. Em geral o endereço é: ‘Para um soldado’. O sargento Rory Lomas, de 27 anos de idade, natural de Savannah, na Georgia, recebeu uma tal carta na Arábia Saudita. E por pura coincidência: ‘a carta para um soldado’ foi escrita pela sua própria filha Cetericka de 10 anos de idade”. Perante este relato, pensamos que uma situação destas só pode ser devida a intervenção divina! A probabilidade de isto acontecer deve ser extremamente pequena, diz-nos a nossa intuição. Mas mais uma vez a nossa intuição nos enganou. Efectivamente a situação descrita é uma versão do conhecido problema dos encontros, que pode ser formulado como se explica já a seguir (Graça Martins, *et al*. 1999a). Uma secretária distraída tinha n cartas para enviar a outros tantos destinatários. Meteu aleatoriamente as cartas dentro dos envelopes, sem tomar atenção aos nomes. Qual a pro-

bilidade de pelo menos uma pessoa receber a carta que lhe era dirigida? O valor para esta probabilidade é aproximadamente 0.63, aproximação que já se obtém para $n=4$.

Componentes da formação de uma pessoa estatisticamente literada

Não é pacífico enumerar as componentes da formação exigida pela literacia estatística, já que a própria definição deste conceito não está propriamente estabelecida.

Podemos, no entanto, indicar alguns requisitos básicos que se consideram necessários para que o cidadão possa cumprir o que dele se espera numa sociedade de números e quantidades (Gal, I., 2002):

- Perceber a necessidade de trabalhar com dados (compreendendo que dados não são unicamente números, mas números inseridos num determinado contexto), conhecendo a sua proveniência e a forma de os produzir;
- Estar familiarizado com os termos e ideias básicas de Estatística Descritiva, nomeadamente métodos (medidas, tabelas e gráficos) para reduzir a informação contida nos dados;
- Compreender noções básicas de Probabilidade;
- Entender o mecanismo do processo inferencial, ao tomar decisões estatísticas.

O primeiro tópico considerado, o da origem e produção de dados, é por vezes relegado para segundo plano, sendo no entanto crucial em qualquer procedimento estatístico. Para realçar a importância desta fase consideremos, por analogia, o que se passa quando se realiza um cozinhado

(Graça Martins e Cerveira, 1999b). Começa-se por seleccionar os ingredientes, que serão depois manipulados de acordo com determinada receita. O resultado pode ser desastroso, 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. 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. Ficaram célebres e hoje em dia ainda se verificam, antecipações de resultados eleitorais completamente contraditórios com os resultados após os actos eleitorais, devido essencialmente a uma amostra deficiente, a partir da qual se obtiveram esses resultados, eventualmente com técnicas estatísticas adequadas.

A familiaridade com os termos e ideias básicas da Estatística Descritiva já foi realçada, quando falámos no perigo da sua utilização incorrecta.

A compreensão das noções básicas de probabilidade é importante, pois o termo Probabilidade é utilizado todos os dias, mais ou menos de forma intuitiva, já que nos mais variados aspectos da nossa vida, está presente a incerteza. Neste ponto deve ser realçado o facto de nem todos os resultados enunciados sob a forma probabilística, serem baseados em estudos estatísticos, havendo por vezes lugar a juízos probabilísticos subjectivos ou baseados em situações anedóticas.

Hoje em dia somos confrontados sistematicamente com informação, veiculada pela comunicação social, sobre resultados de sondagens, que são apresentados obrigatoriamente com a ficha técnica, onde se inclui nomeadamente a margem de erro. Ora, para que sejamos consequentes com esta exigência, é necessário dar às pessoas as ferramentas necessárias para poderem compreender e assimilar a informação que lhes está a ser transmitida. A compreensão do processo inferencial significa também que as pessoas ficam alerta para a possibilidade de se cometerem erros quando se procura generalizar para um conjunto vasto de indivíduos, algumas

propriedades verificadas só em alguns deles, mas que estes erros podem ser controlados e quantificados, através da probabilidade, residindo aqui a enorme potencialidade da Estatística.

Conclusão

O desenvolvimento dos computadores, a sua intervenção crescente na sociedade e a produção intensiva de informação, de que eles são os principais agentes, é um fenómeno que se tem vindo a intensificar desde as últimas décadas do século XX. Esta transformação rápida é de certo modo a responsável pelo nascimento da literacia quantitativa, correspondendo à necessidade do homem moderno se adaptar às novas condições de vida, compreendendo e usando com eficácia a informação que lhe chega diariamente. Aliás, os cálculos automáticos e gráficos automáticos, que lançam a controvérsia entre os matemáticos são fundamentais em Estatística. O uso da tecnologia é, hoje em dia, um aspecto fundamental da prática da Estatística e podemos dizer que a literacia estatística arrasta a literacia computacional.

As definições de literacia em geral e de literacia nos vários domínios e níveis particulares não estão ainda estabelecidas mas, dados os objectivos para que elas apontam, começa já a perceber-se que poderá ser necessário criar adequados mecanismos de educação para a literacia.

No caso da literacia estatística o que se pretende não é criar especialistas em estatística, mas sim criar nas pessoas a capacidade de compreenderem os processos elementares da recolha e análise de dados, entenderem o que está por detrás de um raciocínio estatístico, terem a consciência do que é um fenómeno aleatório, sendo capazes de construir modelos simples da realidade.

A literacia estatística, ao nível do cidadão comum, deve permitir a cada um de nós resolver com ligeireza e segurança um rol de problemas que nos dizem directamente respeito ou que nos são apresentados frequentemente pelos media e cuja resolução apela a conhecimentos e raciocínio estatísticos. Interpretar tabelas e gráficos, entender disputas salariais, indi-

ces de preços, oscilações bolsistas, taxas de desemprego, taxas relativas à evolução de doenças, mecanismos e resultados eleitorais e de sondagens, comparar a qualidade e custos de bens ou serviços são apenas algumas solicitações dirigidas ao cidadão e a que ele pode dar resposta fazendo uso da literacia estatística.

Um cidadão com estas competências que lhe dá a literacia estatística é um cidadão bem informado, vive melhor e pode contribuir de forma esclarecedora para uma sociedade mais justa.

Referências

- Chatfield, C. (1991). Avoiding Statistical Pitfalls. *Statistical Science*, 6, 3, 240-268.
- Gal, I. (2002). Adult's Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, 70, 1, 1-51.
- Graça Martins, M. E., Monteiro, C., Viana, J. P. e Turkman, M. A. (1999a). *Probabilidades e Combinatória*. Ministério da Educação. Departamento do Ensino Superior. Lisboa
- Graça Martins, M. E. e Cerveira, A. (1999b). *Introdução às Probabilidades e à Estatística*. Edição Universidade Aberta. Lisboa.
- Moore, D. (1997). New Pedagogy and New Content. The Case of Statistics. *International Statistical Review*, 65, 2, 123-165.
- Rasfeld, P. (2001). The Role of Statistics in School Mathematics Teaching Today. *International Journal for Mathematics Teaching and Learning*.
- Rossman A. e Chance, B. (2001). *Workshop Statistics: Discovery with data*. Key College Publishing. Emeryville, CA.
- Steen, L. A., ed (1997). *Why Numbers Count: Quantitative Literacy for Tomorrow America*. The College Board. New York.
- Steen, L. A., ed (2001). *Mathematics and Democracy: The case for Quantitative Literacy*. National Council on Education and the Disciplines. Princeton.
- Vere-Jones, D. (1995). The Coming of Age of Statistical Education. *International Statistical Review*, 63, 1, 3-23.

João Branco
Centro de Matemática e Aplicações
Departamento de Matemática
Instituto Superior Técnico

Maria Eugénia Graça Martins
Centro de Estatística e Aplicações
Departamento de Estatística
Operacional da Faculdade de Ciências
da Universidade de Lisboa