

# Estudo da distribuição normal com a calculadora

José António Fernandes

Para Batanero (2004), a distribuição normal é uma das ideias fundamentais da estocástica e, como tal, deve ser ensinada nas aulas de probabilidades.

A relevância da distribuição normal deve-se ao facto de muitos fenómenos físicos, biológicos, psicológicos e sociológicos poderem ser por ela modelados. Além disso, esta distribuição é uma boa aproximação de outras distribuições (e.g., binomial, de Poisson e *t* de Student) para certos valores dos seus parâmetros, a adequada aplicação de muitos métodos estatísticos requer a condição de normalidade e o teorema do limite central garante que a média e outras estatísticas de amostras aleatórias, de suficiente dimensão, seguem uma distribuição aproximadamente normal, inclusive quando são extraídas de populações não normais.

Em termos dos novos programas, a importância deste tema parece reflectir-se no tratamento que lhe é dado nos programas das disciplinas de Matemática A (12º ano) e de Matemática Aplicada às Ciências Sociais e Humanas (11º ou 12º ano). Já no caso do programa da disciplina de Matemática B (12º ano), o tema é tratado com menor desenvolvimento.

Seguidamente, iremos explorar com a calculadora<sup>[1]</sup> a representação gráfica e propriedades da distribuição normal, aproximações sucessivas à distribuição normal e a avaliação da normalidade de uma distribuição dada.

## A distribuição normal e suas propriedades

Em termos teóricos, a distribuição normal é definida por uma função matemática que depende dos parâmetros  $\mu$  e  $\sigma$ , os quais correspondem ao valor esperado e ao desvio padrão da população<sup>[2]</sup>. Esta função, designada por *função densidade de probabilidade* da distribuição normal, traduz-se graficamente por uma curva em forma de *sino*.

Recorrendo a uma calculadora gráfica podemos obter representações gráficas de distribuições normais correspondentes a diferentes valores dos parâmetros (ver figuras 1, 2, 3 e 4). No caso da figura 1, em que  $\mu = 0$  e  $\sigma = 1$ , a distribuição toma a designação de *distribuição normal padronizada*.

Por observação dos diferentes gráficos verificamos que todos eles são simétricos em relação à recta de equação  $X = \mu$  e têm um máximo absoluto no ponto de abscissa  $\mu$ . Além disso, pode demonstrar-se que os gráficos têm dois

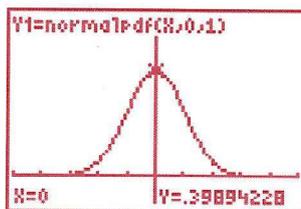


Figura 1. [ $\mu = 0$  e  $\sigma = 1$ ]

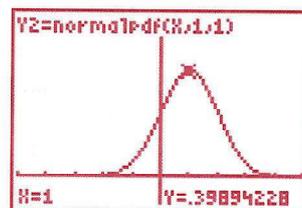


Figura 2. [ $\mu = 1$  e  $\sigma = 1$ ]

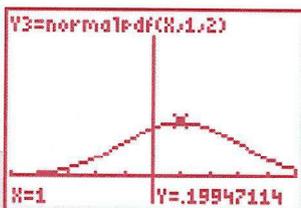


Figura 3. [ $\mu = 1$  e  $\sigma = 2$ ]

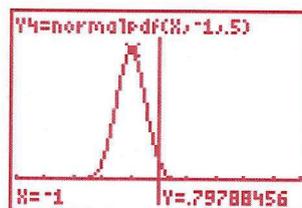


Figura 4. [ $\mu = -1$  e  $\sigma = 0,5$ ]

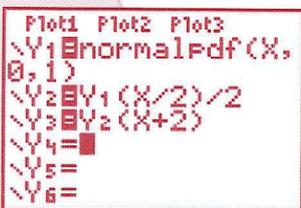


Figura 5.

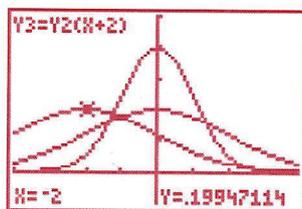


Figura 6.

pontos de inflexão, correspondentes aos pontos de abcissas  $\mu - \sigma$  e  $\mu + \sigma$ , e uma assíntota horizontal bilateral de equação  $Y = 0$ .

Por outro lado, o gráfico de qualquer distribuição normal pode ser obtido a partir do gráfico da distribuição normal padronizada (ver figura 1), *esticando-o* ou *encolhendo-o* simultaneamente na horizontal e na vertical pelo factor  $1/\sigma$  e/ou, seguidamente, deslocando este último na horizontal de  $\mu$  unidades. Nas figuras 5 e 6 podemos observar a transformação do gráfico da distribuição normal padronizada no gráfico da distribuição normal de  $\mu = -2$  e  $\sigma = 2$ .



Figura 7.

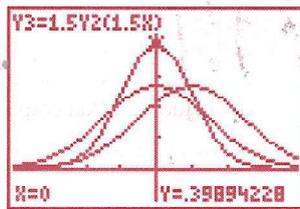


Figura 8.

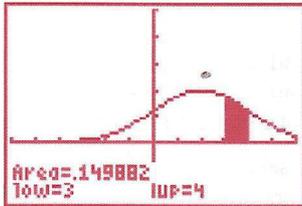


Figura 9.

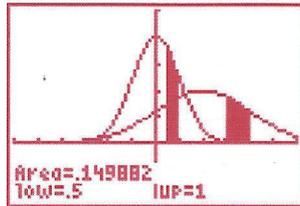


Figura 10.

Reciprocamente, a execução das inversas destas transformações, por ordem inversa, permite transformar o gráfico de uma distribuição normal de valor esperado  $\mu$  e de desvio padrão  $\sigma$  no gráfico da distribuição normal padronizada. Nas figuras 7 e 8 podemos observar a transformação do gráfico da distribuição normal de  $\mu = 1$  e  $\sigma = 1,5$  no gráfico da distribuição normal padronizada.

Esta última transformação evidencia a possibilidade de determinar probabilidades de uma variável aleatória  $X$ , com distribuição normal qualquer, a partir da correspondente variável aleatória reduzida  $Z = (X - \mu)/\sigma$ , com distribuição normal padronizada. No caso da distribuição normal de  $\mu = 2$  e  $\sigma = 2$ , recorrendo ao comando `ShadeNorm(3,4,2,2)`, verificamos que a probabilidade de  $X$  tomar valores entre 3 e 4 é 0,149882 (ver figura 9). Transformando  $X$  em  $Z$ , a execução do comando `ShadeNorm(0.5,1,0,1)` dá-nos a probabilidade da variável reduzida  $Z$  tomar valores compreendidos entre 0,5 e 1 (0,5 é o valor de  $Z$  que corresponde a  $X = 3$  e 1 o que corresponde a  $X = 4$ ) é também 0,149882 (ver figura 10).

Centremo-nos, agora, na distribuição normal padronizada ( $\mu = 0$  e  $\sigma = 1$ ). Considerando que a probabilidade da variável aleatória  $X$  tomar valores no intervalo  $[x_1, x_2]$  é dada pela área limitada pelo gráfico da distribuição, pelas rectas verticais  $X = x_1$  e  $X = x_2$  e pelo eixo dos  $XX$ , conclui-se que a probabilidade de  $X$  tomar valores:

- inferiores ou iguais a  $\mu$ , isto é,  $P(X \leq \mu)$ , é 0,5 ou 50% (ver figura 11);
- no intervalo centrado em  $\mu$  e raio  $\sigma$ , isto é,  $P(\mu - \sigma \leq X \leq \mu + \sigma)$ , é 0,683 (3 c.d.) ou 68,3% (ver figura 12);

— no intervalo centrado em  $\mu$  e raio  $2\sigma$ , isto é,  $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$ , é 0,955 (3 c.d.) ou 95,5% (ver figura 13);

— no intervalo centrado em  $\mu$  e raio  $3\sigma$ , isto é,  $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$ , é 0,997 (3 c.d.) ou 99,7% (ver figura 14).

Além das propriedades anteriores, verifica-se que área limitada pelo gráfico e pelo eixo dos  $XX$ , o que corresponde à probabilidade da variável aleatória  $X$  tomar qualquer valor real, é 1 ou 100%.

Todas as propriedades, referidas anteriormente para a distribuição normal padronizada, também se verificam para qualquer outra distribuição normal. Assim, estas propriedades revestem-se de uma grande importância em duas situações: na primeira, estas propriedades constituem critérios para avaliar em que medida uma distribuição dada se aproxima ou não da distribuição normal; na segunda, partindo do conhecimento de que uma dada distribuição é normal, estas propriedades permitem fazer inferências acerca da distribuição.

### Aproximações sucessivas à distribuição normal

Considerem-se sucessivos lançamentos de uma ou mais moedas ao ar.

Aumentando o número de moedas lançadas ao ar e considerando as frequências relativas dos valores da variável aleatória *número de caras* (ou *número de coroas*) obtidas podemos observar que a distribuição correspondente se vai aproximando da distribuição normal.

Consideremos 400 lançamentos de uma, três e seis moedas ao ar. Recorrendo a histogramas, em cada caso, representar graficamente as distribuições das frequências relativas da variável aleatória *número de caras* e avaliar a aproximação à correspondente curva normal.

Para simular o lançamento de uma moeda 400 vezes, definir e representar graficamente a distribuição das frequências

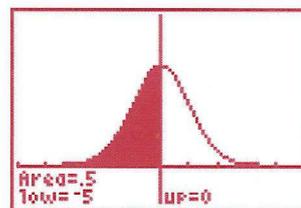


Figura 11.

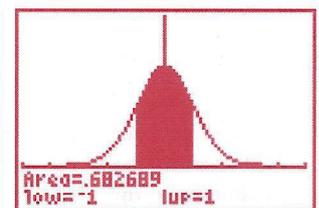


Figura 12.

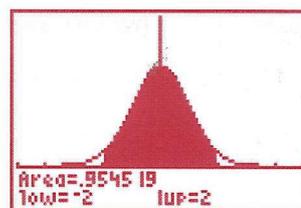


Figura 13.

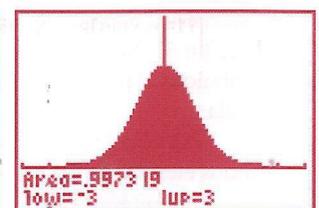


Figura 14.

relativas obtidas e representar graficamente a distribuição normal correspondente, definimos e executamos na calculadora os comandos seguintes:

- `randInt(0,1,400)→L1` simula o número de caras obtidas em cada um dos 400 lançamentos de uma moeda ao ar e armazena os resultados na lista L1;
- `{0,1}→L1U` atribui à lista L1U os valores 0 (0 caras) e 1 (1 cara);
- `"{sum(L1=0),sum(L1=1)}/400"→L1UF` calcula as frequências relativas dos valores 0 e 1 e armazena os resultados na lista L1UF (observe-se que as aspas significam que a lista está indexada a uma fórmula);
- `1-VarStats L1U,L1UF` calcula estatísticas da distribuição definida pelas listas L1U e L1UF, designadamente  $\bar{x}$  e  $\sigma x$  (observe-se que estas estatísticas são necessárias na definição de Y1 da etapa seguinte);
- Escrever `Y1=normalpdf(X, $\bar{x}$ , $\sigma x$ )` no menu Y=, substituindo  $\bar{x}$  e  $\sigma x$  pelos valores determinados na etapa anterior (observe-se que Y1 é a função densidade de probabilidade da distribuição normal com média  $\bar{x}$  e desvio padrão  $\sigma x$ );
- Activar o Plot1, seleccionar o histograma e especificar as listas U e UF;
- Definir a janela `[-1.5,7]×[-0.25,1]` e premir a tecla GRAPH. Obtém-se, assim, o histograma da distribuição e a correspondente curva normal (ver figura 15).

No caso do lançamento de três moedas 400 vezes, definimos e executamos na calculadora os comandos seguintes:

- `L1+randInt(0,1,400)→L2` simula o número de caras obtidas em cada um dos 400 lançamentos de duas moedas ao ar e armazena os resultados na lista L2 (recorde-se que antes estava armazenado em L1 o número de caras obtidas em cada um dos 400 lançamentos da moeda);
- `L2+randInt(0,1,400)→L2` simula o número de caras obtidas em cada um dos 400 lançamentos de três moedas ao ar e armazena os resultados na lista L2;
- `{0,1,2,3}→L2T` atribui à lista L2T os valores 0 (0 caras), 1 (1 cara), 2 (2 caras) e 3 (3 caras);
- `"{sum(L2=0),sum(L2=1),sum(L2=2),sum(L2=3)}/400"→L2TF` calcula as frequências relativas dos valores 0,1,2,3 e armazena os resultados na lista L2TF;

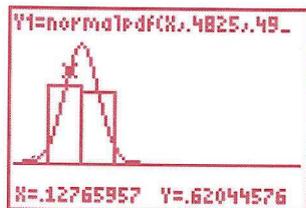


Figura 15.

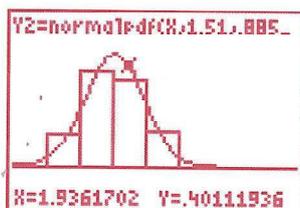


Figura 16.

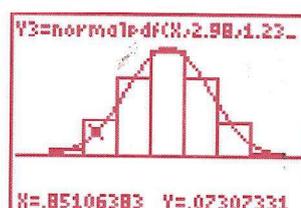


Figura 17.

- `1-VarStats L2T,L2TF` calcula estatísticas da distribuição definida pelas listas L2T e L2TF, designadamente  $\bar{x}$  e  $\sigma x$  (observe-se que estas estatísticas são necessárias na definição de Y2 da etapa seguinte);
- Escrever `Y2=normalpdf(X, $\bar{x}$ , $\sigma x$ )` no menu Y=, substituindo  $\bar{x}$  e  $\sigma x$  pelos valores determinados na etapa anterior (não esquecer de desactivar Y1);
- Activar o Plot2, seleccionar o histograma e especificar as listas T e TF (não esquecer de desactivar o Plot1);
- Definir a janela `[-1.5,7]×[-0.15,0.6]` e premir a tecla GRAPH. Obtém-se, assim, o histograma da distribuição e a correspondente curva normal (ver figura 16).

Finalmente, para simular o lançamento de seis moedas 400 vezes, definimos e executamos na calculadora os comandos seguintes:

- `L2+randInt(0,1,400)→L3` simula o número de caras obtidas em cada um dos 400 lançamentos de quatro moedas ao ar e armazena os resultados na lista L3 (recorde-se que antes estava armazenado em L2 o número de caras obtidas em cada um dos 400 lançamentos de três moedas);
- `L3+randInt(0,1,400)→L3` simula o número de caras obtidas em cada um dos 400 lançamentos de cinco moedas ao ar e armazena os resultados na lista L3;
- `L3+randInt(0,1,400)→L3` simula o número de caras obtidas em cada um dos 400 lançamentos de seis moedas ao ar e armazena os resultados na lista L3;
- `{0,1,2,3,4,5,6}→L3S` atribui à lista L3S os valores 0 (0 caras), 1 (1 cara), 2 (2 caras), 3 (3 caras), 4 (4 caras), 5 (5 caras) e 6 (6 caras);
- `"{sum(L3=0),sum(L3=1),sum(L3=2),sum(L3=3),sum(L3=4),sum(L3=5),sum(L3=6)}/400"→L3SF` calcula as frequências relativas dos valores 0,1,2,3,4,5,6 e armazena os resultados na lista L3SF;
- `1-VarStats L3S,L3SF` calcula estatísticas da distribuição definida pelas listas L3S e L3SF, designadamente  $\bar{x}$  e  $\sigma x$  (observe-se que estas estatísticas são necessárias na definição de Y3 da etapa seguinte);
- Escrever `Y3=normalpdf(X, $\bar{x}$ , $\sigma x$ )` no menu Y=, substituindo  $\bar{x}$  e  $\sigma x$  pelos valores determinados na etapa anterior (não esquecer de desactivar Y2);

- Activar o Plot3, seleccionar o histograma e especificar as listas 5 e SF (não esquecer de desactivar o Plot2);
- Definir uma janela  $[-1.5, 7] \times [-0.15, 0.4]$  e premir a tecla GRAPH. Obtém-se, assim, o histograma da distribuição e a correspondente curva normal (ver figura 17).

Observando os gráficos das figuras 15, 16 e 17, verifica-se um ajustamento cada vez melhor entre o histograma que representa a distribuição e a correspondente curva normal, o que é justificado teoricamente pelo facto da distribuição binomial se aproximar da distribuição normal com o aumento do número de moedas consideradas.

### Avaliar a normalidade de uma distribuição

Conforme foi referido antes, sabendo-se que uma dada distribuição é normal, ficamos a conhecer a probabilidade da variável aleatória pertencer a cada um dos seguintes intervalos: para o intervalo  $[\mu - \sigma, \mu + \sigma]$  a probabilidade é 68,3%, para o intervalo  $[\mu - 2\sigma, \mu + 2\sigma]$  a probabilidade é 95,5% e para o intervalo  $[\mu - 3\sigma, \mu + 3\sigma]$  a probabilidade é 99,7%. Ora, este conhecimento permite-nos avaliar em que medida uma distribuição se aproxima ou não da distribuição normal.

A partir dos três intervalos considerados, estudemos em que medida a distribuição definida no exemplo seguinte se aproxima ou afasta da distribuição normal.

Contou-se o número de pessoas, incluindo o condutor, em 40 automóveis à entrada de um parque de estacionamento, tendo-se obtido os seguintes valores:

#### Nº de pessoas por automóvel

1 2 2 3 1 1 2 4 3 1 3 1 1 2 2 4 5 3 3 2  
2 2 1 2 3 4 1 5 4 4 2 1 3 3 1 2 5 3 2 4

Determinando a média e o desvio padrão, obtém-se os valores:  $\bar{x} = 2,5$  e  $s = 1,2$  (1 c.d.). Substituindo os valores obtidos para  $\bar{x}$  e para  $s$  nos respectivos intervalos, tem-se:

- $[\bar{x} - s, \bar{x} + s]$  dá origem ao intervalo  $[2,5 - 1,2; 2,5 + 1,2]$ , ou seja,  $[1,3; 3,7]$ .

Observando a tabela dada, verifica-se que pertencem ao intervalo os valores 2 e 3, a que corresponde um

total de 21 (12+9) dados. Ou, em percentagem,  $21/40=52,5\%$  do total dos dados.

Definindo e executando o comando  $(\text{sum}(L1 \geq \bar{x} - \sigma \text{ and } L1 \leq \bar{x} + \sigma) / 40) \times 100$  na calculadora, obtém-se a percentagem de dados que pertencem ao intervalo  $[\bar{x} - s, \bar{x} + s]$  (ver figura 18).

- $[\bar{x} - 2s, \bar{x} + 2s]$  dá origem ao intervalo  $[2,5 - 2 \times 1,2; 2,5 + 2 \times 1,2]$ , ou seja,  $[0,1; 4,9]$ . Observando a tabela dada, verifica-se que pertencem ao intervalo os valores 1, 2, 3 e 4, a que corresponde um total de 37 (10+12+9+6) dados. Ou, em percentagem,  $37/40=92,5\%$  do total dos dados.

Definindo e executando o comando  $(\text{sum}(L1 \geq \bar{x} - 2\sigma \text{ and } L1 \leq \bar{x} + 2\sigma) / 40) \times 100$  na calculadora, obtém-se a percentagem de dados que pertencem ao intervalo  $[\bar{x} - 2s, \bar{x} + 2s]$  (ver figura 19).

- $[\bar{x} - 3s, \bar{x} + 3s]$  dá origem ao intervalo  $[2,5 - 3 \times 1,2; 2,5 + 3 \times 1,2]$ , ou seja,  $[-1,1; 6,1]$ . Observando a tabela dada, verifica-se que pertencem ao intervalo os valores 1, 2, 3, 4 e 5, a que corresponde todos os dados ou 100%.

Definindo e executando o comando  $(\text{sum}(L1 \geq \bar{x} - 3\sigma \text{ and } L1 \leq \bar{x} + 3\sigma) / 40) \times 100$  na calculadora, obtém-se a percentagem de dados que pertencem ao intervalo  $[\bar{x} - 3s, \bar{x} + 3s]$  (ver figura 20).

Assim, comparando os valores obtidos para a percentagem de dados nos três intervalos com as percentagens correspondentes à distribuição normal, verifica-se que a distribuição do número de pessoas que viajavam nos 40 automóveis se afasta consideravelmente da distribuição normal, especialmente no caso do intervalo  $[\bar{x} - s, \bar{x} + s]$ .

Na figura 21 podemos verificar o ajustamento entre o histograma da distribuição dada e a curva normal correspondente.

Pela figura observa-se um considerável desajustamento entre a curva normal e o histograma da distribuição dada. Pode ainda verificar-se que se trata de uma distribuição assimétrica positiva ou desviada para a direita.

Por outro lado, sabemos que a média  $\bar{x} = 2,5$  e, determinando o valor da mediana, obtém-se  $\tilde{x} = 2$ . Ora, como  $\bar{x} > \tilde{x}$ , tal significa que a distribuição é assimétrica positiva ou desviada para a direita.

Assim, a evidência obtida a partir de diferentes estratégias — os intervalos centrados na média (diferindo de  $s$ ,  $2s$  e  $3s$ ), o ajustamento entre o histograma da distribui-

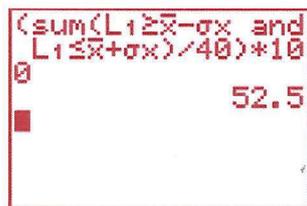


Figura 18.

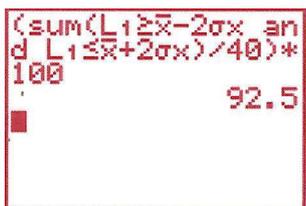


Figura 19.

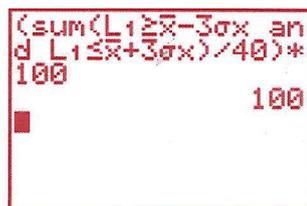


Figura 20.

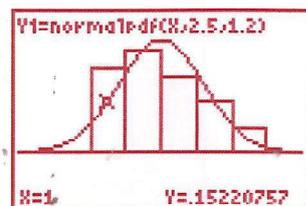


Figura 21.

ção dada e a curva normal correspondente e a comparação entre os valores da média e da mediana —, porque apontam no mesmo sentido, permite-nos afirmar com maior segurança que a distribuição dada se afasta da normal.

### Considerações didáticas

Para Batanero (2004), a realização de experiências empíricas sobre a convergência seriam desejáveis desde a escola básica, pois facilitariam a compreensão dos teoremas matemáticos que serão abordados posteriormente na universidade. No entanto, estas experiências são limitadas e as sucessões aleatórias geradas na sala de aula convergem lentamente e, por vezes, falham quando são necessárias a uma demonstração, o que pode ser contraproducente.

Face a estas dificuldades, em geral, a simulação de experiências aleatórias recorrendo a computadores e calculadoras reveste-se de uma importância evidente uma vez que esta tecnologia permite, num tempo realista, gerar uma quantidade de dados suficientemente grande de modo a evitar discrepâncias com as propriedades teóricas que se pretende que emirjam ou se concretizem a partir dos dados.

Por outro lado, a importância de uma abordagem empírica à distribuição normal, do tipo da que aqui se propõe, resulta também das dificuldades que os alunos revelam neste tema, não garantido uma abordagem teórica e tardia a aquisição de significados e interpretações adequadas. Neste sentido, um estudo realizado por Batanero, Tauber e Sánchez (2001), com alunos universitários de uma disciplina introdutória de análise de dados, revelou dificuldades várias dos alunos neste tema. De entre essas dificuldades, salientam-se dificuldades com o cálculo de valores tipificados e seus inversos, em distinguir entre distribuição teórica e distribuição empírica e em decidir se uma variável aleatória se distribui ou não normalmente.

Em relação à avaliação da normalidade de uma distribuição, muitos alunos chegaram a conclusões erradas porque a sua decisão partiu da representação gráfica ou resultado da aplicação ou prova de apenas uma propriedade (normalmente a simetria), confundindo, assim, uma condição necessária de normalidade com uma condição suficiente. Deste resultado pode concluir-se a conveniência de se aplicar ou verificar, simultaneamente, várias propriedades na avaliação da normalidade de uma distribuição dada, até porque as propriedades estudadas ao nível do ensino secundário são, geralmente, menos robustas.

Tal como foi exemplificado antes, podemos reunir evidência a partir dos intervalos centrados na média, com base na simetria (recorrendo à média, mediana e moda) e sobre o ajustamento entre a representação gráfica da distribuição e a correspondente curva normal. Neste caso, a confirmação de uma mesma conclusão, a partir de várias estratégias, permite afirmá-la com maior convicção, e simultaneamente diminuir a insegurança e as dúvidas que o recurso a uma única estratégia geralmente acarreta.

### Notas

[1] Na exploração dos exemplos foi utilizada uma calculadora gráfica TI-83 Plus da Texas Instruments.

$$\begin{aligned} [2] \quad P(x_1 \leq X \leq x_2) &= \\ &= \int_{x_2}^{x_1} n(x; \mu, \sigma) dx = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_2}^{x_1} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx. \end{aligned}$$

### Referências bibliográficas

- Batanero, C. (2004). Ideas estocásticas fundamentales: ¿Qué contenidos se debe enseñar en la clase de probabilidad? In J. A. Fernandes, M. V. Sousa e S. A. Ribeiro (Orgs.), *Ensino e aprendizagem de probabilidades e estatística — Actas do I Encontro de Probabilidades e Estatística na Escola*. Braga: Centro de Investigação em Educação da Universidade do Minho.
- Batanero, C., Tauber, L. M. e Sánchez, V. (2001). Significado y comprensión de la distribución normal en un curso introductorio de análisis de datos. *Cuadrante*, 10(1), 59–91.
- Martins, M. E., Monteiro, C., Viana, J. P. e Turkman, M. A. (1999). *Probabilidades e estatística — 12º ano de escolaridade*. Lisboa: Ministério da Educação.
- Ministério da Educação (2001). *Programa de Matemática aplicada às Ciências Sociais*. Lisboa: Autor.
- Ministério da Educação (2002a). *Matemática A — Programa do 12º ano*. Lisboa: Autor.
- Ministério da Educação (2002b). *Matemática B — Programa do 12º ano*. Lisboa: Autor.
- Wonnacott, T. H. e Wonnacott, R. J. (1990). *Introductory statistics* (5th ed.). New York, NY: John Wiley & Sons.

José António Fernandes  
Universidade do Minho