

A Matemática escondida nos livros

Helena Rocha^[1] e Isabel Oitavem^[2]



O ISBN é o International Standard Book Number. Até 2007 este número era constituído por dez algarismos e, desde então, passou a incluir o código EAN (*European Article Number*) e a ser constituído por 13 algarismos. O ISBN é um número reconhecido internacionalmente e que permite rapidamente identificar qualquer livro. No nosso país o sistema existe desde 1988 estando a sua gestão a cargo da APEL — Associação Portuguesa de Editores e Livreiros. O ISBN encontra-se dividido em cinco partes:

- O EAN — um código de três algarismos que permite saber de que tipo de artigo estamos a falar (livro, *software*, etc.). No caso dos livros este código será o 978 ou o 979.
- O grupo de registo — um código que pode ter de 1 a 5 algarismos e que identifica o país ou a língua em que o livro está escrito.

- A editora — um conjunto de até 7 algarismos que identifica a editora que publicou o livro.
- O identificador do título — um código até 6 algarismos que identifica especificamente o livro.
- O algarismo de controlo — o algarismo final que valida matematicamente o ISBN e que tem como função detectar eventuais enganos quando o ISBN é digitado.

O ISBN surge por norma na ficha técnica do livro e na sua contracapa, em geral junto de um código de barras destinado a leitura óptica. O ISBN pode ser apresentado de duas maneiras, como se ilustra na figura 1. A primeira é mais conveniente quando é possível recorrer a um instrumento de leitura óptica do código de barras, e a segunda tem a vantagem de permitir desde logo identificar as cinco partes que constituem o número.

Figura 1



ou

ISBN 978-972-41-4427-6

algarismo	1º algarismo	1º grupo de 6 algarismos		2º grupo de 6 algarismos	
0	111111		0001101		0100111
1	110100		0011001		0110011
2	110010		0010011		0011011
3	110001		0111101		0100001
4	101100		0100011		0011101
5	100110		0110001		0111001
6	100011		0101111		0000101
7	101010		0111011		0010001
8	101001		0110111		0001001
9	100101		0001011		0010111
paridade		ímpar		par	par

Figura 2

Quando o ISBN é representado juntamente com um código de barras o primeiro algarismo surge à esquerda das barras e os restantes surgem por baixo destas, divididos em dois grupos de seis algarismos separados e delimitados por separadores (umas barras em regra mais compridas que as restantes). As barras consistem numa sequência de tiras claras e escuras de diferentes grossuras que representam os algarismos. Cada conjunto de sete módulos brancos ou pretos representa um algarismo. Por exemplo, o algarismo 2 do segundo conjunto de seis algarismos é representado por dois módulos pretos, um módulo branco, dois módulos pretos e dois módulos brancos. Durante a leitura óptica o computador atribui a cada módulo branco o valor zero e a cada módulo preto o valor um, obtendo assim uma sequência binária de comprimento sete (que no caso do nosso exemplo seria 1101100) a que posteriormente associa um algarismo de 0 a 9, de acordo com a última coluna da tabela da figura 2.

Os algarismos do primeiro conjunto de seis algarismos podem ser representados de duas maneiras diferentes. Uma das representações possíveis corresponde ao que podemos chamar o inverso da representação do segundo conjunto de seis algarismos em que as barras pretas passam a brancas e vice-versa (na figura 2, comparar a primeira coluna de código de barras com a última). Neste caso o algarismo 2 seria representado por dois módulos brancos, um módulo preto, dois módulos brancos e dois módulos pretos a que corresponderia o número 0010011.

Outra representação possível é a imagem no espelho da codificação do segundo grupo de seis algarismos (na figura 2, comparar as duas últimas colunas de código de barras). E neste caso o algarismo 2 seria representado por duas barras brancas, duas barras pretas, uma barra branca e duas barras pretas, a que corresponderia a sequência 0011011. Estas duas representações são designadas de par ou ímpar em função do número de 1's que incluem. Assim, 0010011 será a representação ímpar do algarismo 2 porque tem um número ímpar (três) de 1's e 0011011 será a representação par do algarismo 2 porque tem um número par (quatro) de 1's.

A escolha, para cada um dos algarismos do primeiro grupo de seis algarismos, entre a sua representação ímpar ou a sua representação par, determina uma sequência de seis dígitos («1» se for escolhida a representação ímpar e «0» se for escolhida a representação par) que permite codificar o primeiro algarismo do ISBN (aquele que surge à esquerda do código de barras).

A tabela da figura 2 apresenta para cada algarismo as codificações para o primeiro algarismo do ISBN, para o primeiro grupo de seis algarismos com codificações ímpares e pares e para o segundo grupo de seis algarismos. Estas codificações podem ser inferidas a partir da análise do código de barras de diversos ISBN's ou, pelo contrário, podem ser utilizadas para construir códigos de barras. Vamos usá-la para reproduzir o código de barras da figura 1.



Figura 3

O primeiro algarismo do ISBN da figura 1 é 9, que é codificado por 100101. Isto significa que no primeiro grupo de seis algarismos iremos escolher sucessivamente as representações ímpar (1), par (0), par (0), ímpar (1), par (0) e ímpar (1). Ou seja, que para o algarismo 7 iremos escolher a representação ímpar, para o 8 a par, para o 9 a par, para o 7 a ímpar, para o 2 a par e para o 4 a ímpar (figura 3).

A segunda parte do código de barras terá por base os códigos do segundo conjunto de seis algarismos (figura 4).

R Matemática do ISBN: elementar ou não?

O ISBN, podendo ser algo a que os alunos nunca deram atenção, está em certo sentido presente na sala de aula, pois surge no manual de todas as disciplinas. Além disso, alguns alunos podem achar interessante perceber um pouco do seu funcionamento e encontrar um exemplo de uma situação concreta em que a utilidade da Matemática é visível, ao disponibilizar uma forma de detectar eventuais enganos na digitalização dos algarismos. Acresce ainda que esta é uma situação simples, cuja compreensão está ao alcance de alunos mesmo de níveis mais elementares, mas que pode depois ser estendida e aprofundada, tornando-a relevante também para alunos de níveis mais avançados. A aritmética modular é aliás uma das áreas da Matemática cujo estudo autores como Justesen & Hoholdt (2004) consideram só estar ao alcance de alunos com alguma maturidade, como é o caso daqueles que já se encontram a frequentar um 3.º ou 4.º ano no ensino superior.

A um nível mais elementar, esta pode ser uma boa ocasião para proporcionar aos alunos um contacto com diferentes representações duma mesma entidade, abordando uma representação numérica, uma representação geométrica e uma representação binária, que sendo obviamente numérica é diferente daquela a que os alunos estão habituados e permite lançar diversas questões e estabelecer pontes com áreas como a informática.

A forma como são constituídos os ISBN, a partir das cinco componentes base, das quais três têm um número variável de algarismos, pode também constituir um ponto de partida para um trabalho exploratório e de natureza organizativa. Se não vejamos:

O grupo de registo dos livros publicados em Portugal tem três algarismos. No Canadá encontramos livros em que o grupo de registo tem um único algarismo. Sendo o grupo de registo estabelecido a nível mundial, que critérios podem ser estabe-

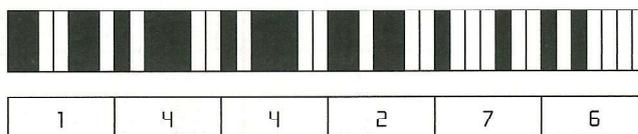


Figura 4

lecidos para decidir que países e idiomas ficam com grupos de registo de um, dois, três ou mais algarismos?

E em Portugal, que critérios podem ser estabelecidos para a atribuição dos códigos às editoras? Que implicações têm essas decisões sobre o número máximo de livros que cada uma pode publicar?

O número de controlo é outra situação interessante do ISBN que, do ponto de vista matemático, pode originar oportunidades de exploração ricas. A abordagem pode ser feita a um nível relativamente elementar, como veremos no ponto seguinte, mas pode também envolver conhecimentos matemáticos que vão para além do elementar. Nomeadamente, conhecimentos de teoria dos corpos finitos relevantes para a teoria dos códigos.

O número de controlo do ISBN no passado e no presente

Como já referimos, até 2007 o ISBN era constituído por dez algarismos: nove algarismos mais um de controlo. Considerando o número ISBN como $x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10}$, onde cada x_i representa um algarismo, o número de controlo x_{10} era determinado de modo a que

$$x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 + 6x_6 + 7x_7 + 8x_8 + 9x_9 + 10x_{10} \equiv 0 \pmod{11}.$$

Tabela 1

Nº de controlo ?	Valor de $234 + 10 \times ?$	Resto da divisão por 11
0	234	3
1	244	2
2	254	1
3	264	0
4	274	10
5	284	9
6	294	8
7	304	7
8	314	6
9	324	5
10	334	4

Actual ISBN		x_2									
		0	1	2	3	4	5	6	7	8	9
x_1	0		3-1	6-2	9-3	2-4	5-5	8-6	1-7	4-8	7-9
	1	1-3		7-5	0-6	3-7	6-8	9-9	2-0	5-1	8-2
	2	2-6	5-7		1-9	4-0	7-1	0-2	3-3	6-4	9-5
	3	3-9	6-0	9-1		5-3	8-4	1-5	4-6	7-7	0-8
	4	4-2	7-3	0-4	3-5		9-7	2-8	5-9	8-0	1-1
	5	5-5	8-6	1-7	4-8	7-9		3-1	6-2	9-3	2-4
	6	6-8	9-9	2-0	5-1	8-2	1-3		7-5	0-6	3-7
	7	7-1	0-2	3-3	6-4	9-5	2-6	5-7		1-9	4-0
	8	8-4	1-5	4-6	7-7	0-8	3-9	6-0	9-1		5-3
	9	9-7	2-8	5-9	8-0	1-1	4-2	7-3	0-4	3-5	

Tabela 2

Ou seja, o décimo algarismo era escolhido de modo a que quando multiplicado por dez e somado a $x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 + 6x_6 + 7x_7 + 8x_8 + 9x_9$, o resultado fosse um múltiplo de 11.

Por exemplo, se pretendêssemos determinar o algarismo de controlo do livro de ISBN 972662792?, começaríamos por calcular $9 + 2 \times 7 + 3 \times 2 + 4 \times 6 + 5 \times 6 + 6 \times 2 + 7 \times 7 + 8 \times 9 + 9 \times 2$, que dá 234. O número de controlo «?» seria então um valor entre 0 e 10 tal que

$$234 + 10 \times ? \text{ é múltiplo de } 11. \text{ (Tabela 1)}$$

Então o número de controlo seria 3 e consequentemente o ISBN completo seria 9726627923.

A passagem do número de algarismos do ISBN de dez para treze veio, contudo, conduzir ao abandono desta forma de determinar o algarismo de controlo. Na verdade, já existia um pequeno problema com o algoritmo associado ao ISBN de dez algarismos. É que o algarismo de controlo podia variar de 0 a 10, o que fazia com que o algarismo de controlo por vezes fosse 10, o que não é um algarismo. Neste caso o problema foi contornado com um pequeno artifício. Recorreu-se a X, que em numeração romana representa o valor 10, e embora não se tivesse obviamente um algarismo tínhamos pelo menos um único símbolo. A passagem do ISBN para treze algarismos obrigava necessariamente a alterações no algoritmo para determinar o algarismo de controlo e levou à adopção de um novo algoritmo.

Num ISBN actual os primeiros doze algarismos são seguidos de um algarismo de controlo. Considerando o ISBN como $x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} x_{11} x_{12} x_{13}$, onde cada x_i representa um algarismo, o algarismo de controlo x_{13} é determinado de modo a que

$$x_1 + 3x_2 + x_3 + 3x_4 + x_5 + 3x_6 + x_7 + 3x_8 + x_9 + 3x_{10} + x_{11} + 3x_{12} + x_{13} \equiv 0 \pmod{10}.$$

Ou seja, o décimo terceiro algarismo é escolhido de modo a que quando somado a $x_1 + 3x_2 + x_3 + 3x_4 + x_5 + 3x_6 + x_7 + 3x_8 + x_9 + 3x_{10} + x_{11} + 3x_{12}$, o resultado seja um múltiplo de 10.

Por exemplo, se pretendérmos determinar o algarismo de controlo do livro de ISBN 978972662792?, começaríamos por calcular $9 + 3 \times 7 + 8 + 3 \times 9 + 7 + 3 \times 2 + 6 + 3 \times 6 + 2 + 3 \times 7 + 9 + 3 \times 2$ que dá 140. O algarismo de controlo «?» seria então um valor entre 0 e 9 tal que $140 + ?$ é múltiplo de 10. Ou seja, 0. E, consequentemente, o ISBN completo seria 9789726627920.

A grande função deste algarismo de controlo é naturalmente detectar eventuais enganos ao digitar o número de ISBN. De entre os diversos enganos possíveis os mais comuns são, segundo Buescu (2002), os que envolvem a inserção de um algarismo errado ou a troca entre dois algarismos adjacentes, sendo que, segundo Picado (2001), a ocorrência de mais de um erro na introdução de um mesmo ISBN é muito reduzida. É fácil ver que tanto o actual como o antigo algoritmo para determinar o algarismo de controlo do ISBN detecta os casos em que é introduzido um algarismo errado. E a troca entre dois algarismos adjacentes!

As tabelas 2 e 3 apresentam as somas dos dois primeiros algarismos em módulo 10 ou em módulo 11 (respectivamente para o ISBN actual e antigo) quando estes são introduzidos pela ordem correcta ($x_1 x_2$) ou pela ordem incorrecta ($x_2 x_1$).

Uma análise das tabelas permite concluir que o algoritmo antigo detecta sempre a troca na ordem de introdução dos dois algarismos, mas o mesmo já não acontece com o algoritmo actual. Neste caso, sempre que a diferença entre os dois algarismos trocados é de cinco unidades, o erro não é detectado. Esta conclusão pode ser induzida a partir da tabela, mas também pode ser deduzida matematicamente. Tal deve-se ao facto do $\text{mdc}(2, 10)$ ser 2 e não 1. Com efeito, se considerarmos a diferença entre as duas somas de controlo (a do número com erro e a do número correcto) veremos que esta é o dobro da diferença entre os dois algarismos trocados e que, portanto, será

Antigo ISBN		X_2									
		0	1	2	3	4	5	6	7	8	9
X_1	0		2-1	4-2	6-3	8-4	10-5	1-6	3-7	5-8	7-9
	1	1-2		5-4	7-5	9-6	0-7	2-8	4-9	6-10	8-0
	2	2-4	4-5		8-7	10-8	1-9	3-10	5-0	7-1	9-2
	3	3-6	5-7	7-8		0-10	2-0	4-1	6-2	8-3	10-4
	4	4-8	6-9	8-10	10-0		3-2	5-3	7-4	9-5	0-6
	5	5-10	7-0	9-1	0-2	2-3		6-5	8-6	10-7	1-8
	6	6-1	8-2	10-3	1-4	3-5	5-6		9-8	0-9	2-10
	7	7-3	9-4	0-5	2-6	4-7	6-8	8-9		1-0	3-1
	8	8-5	10-6	1-7	3-8	5-9	7-10	9-0	0-1		4-3
	9	9-7	0-8	2-9	4-10	6-0	8-1	10-2	1-3	3-4	

Tabela 3

um múltiplo de 10 (e conseqüentemente o erro não será detectado) sempre que a diferença entre os dois algarismos trocados seja 5.

Parece assim que o actual sistema não é tão eficaz como o anterior, mas a verdade é que, como destaca Buescu (2002), não é possível detectar todas as trocas na ordem de introdução de dois algarismos adjacentes e todas as introduções incorrectas de um algarismo tendo para algarismo de controlo apenas as nove hipóteses de 0 a 9. No entanto, esse também não era o objectivo. A crescente divulgação dos mecanismos de leitura óptica tornou muito mais raras as situações em que os algarismos são digitados directamente e, além disso, a grande fiabilidade destes instrumentos de leitura também tornou menos frequentes determinados erros e, conseqüentemente, menos necessários determinados processos para a sua detecção (Picado, 2001).

Em aberto deixamos aqui as questões relativas à base dos algoritmos:

- Porque razão considerar os múltiplos de 11, no algarismo de controlo do antigo ISBN?
- E porquê a escolha do 3 para multiplicar por metade dos algarismos, no algarismo de controlo do actual ISBN?

Número de controlo noutras situações do quotidiano

Existem muitas outras situações com que nos deparamos regularmente e em que também é usado algum tipo de número de controlo (embora os algoritmos para o determinar variem). A que mais se assemelha ao ISBN é a que podemos encontrar quando vamos às compras, junto ao código de barras de qualquer produto. Nesta situação, o último algarismo tem precisamente a função de detectar alguns dos possíveis enganos que podem ocorrer quando o código do artigo é inserido manualmente na caixa registadora. Uma outra situação, talvez mesmo das mais conhecidas, ocorre no bilhete de identidade. Aquele número que surge isolado num quadradinho pretende ser precisamente

um número de controlo. O número de cada nota de euro também inclui um número de controlo. O NIB da conta bancária engloba igualmente um sistema que pretende detectar eventuais enganos, mas neste caso estão envolvidos dois algarismos e não apenas um. E, de uma maneira geral, sempre que nos deparamos com situações em que estão envolvidos números de vários algarismos que tornam natural a ocorrência de erros de digitação, a Matemática está também presente, dando o seu contributo para tentar detectar erros e assim minimizar a ocorrência de situações desagradáveis. Afinal ninguém quer depositar dinheiro na conta bancária de um desconhecido, só porque se enganou no NIB, nem pagar o dobro por um produto, devido a um engano do funcionário que nos atendeu ao digitar o código do artigo. É para minimizar situações como estas que todos contamos com a Matemática escondida nos diversos números de controlo.

Bibliografia

- Buescu, J. (2002). *O mistério do bilhete de identidade e outras histórias*. Lisboa: Gradiva.
- Justesen, J., Hoholdt, T. (2004). *A course in error — correcting codes*. Zürich: European Mathematical Society.
- Masunaga, D. (1994). Zips and strips. *NCTM Student Math Notes*, January.
- Picado, J. (2001). A álgebra dos sistemas de identificação. *Boletim da SPM*, 44, 39-73.

Notas

- 1 Com o apoio da FCT (SFRH/BD/37905/2007) e do ME
- 2 Com o apoio da FCT e da European Science Foundation, projectos PTDC/MAT/104716/2008 e LogICCC/0001/2007

Helena Rocha

Bolsista da FCT/ME

Isabel Dítavem

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa