Identification Numbers

Zip codes, bar codes for items in a grocery store, and ISBN numbers for books are three common ways numerical information is encoded for optical scanning. When a machine reads information, there is always the possibility that the information will be read incorrectly. For example, moisture or dirt on the scanner used by a grocery store clerk can prevent an item's code from being read correctly. It would be unacceptable if, because of a scanning error, a customer is charged for caviar when they are buying tuna fish. So these codes (i.e. numbers) are designed to detect the most frequently encountered scanning errors: if an error is detected, the item is then rescanned until the correct code is read.

Examples of Identification Numbers

The USPS ZipCode:

The United States Postal Service uses a bar code to read zip codes on mail. The zip code for the Department of Mathematical Sciences at NMSU is

$$88003 - 8001.$$

However, the bar code associated to this zip code has the number 2 appended to the end so that the sum of all 10 digits results in a multiple of 10.

Exercise 1 *Check this.*

The 2 appended to the end is called the check digit.

The Universal Product Code (UPC):

The Universal Product Code, or UPC code appears on virtually all grocery items. It is a twelve digit code consisting of two blocks of five digits preceded and followed with a single digit. For example

0 4900001134 0

is a typical UPC number. The first six digits identify the country and the manufacturer of the product, and the next five identify the product itself. Again, the final digit is the check digit.

Later on, when we discuss identification schemes in general, it will be convenient to write a UPC code as a 12-tuple, i.e. as

$$(a_1, \ldots, a_{12}).$$

The 12− tuple $(a_1, \ldots, a_{12})$ represents a valid product provided that

$$3a_1 + a_2 + 3a_3 + a_4 + 3a_5 + a_6 + 3a_7 + a_8 + 3a_9 + a_{10} + 3a_{11} + a_{12}$$

is a multiple of 10. The example above is thus written as

$$(0, 4, 9, 0, 0, 0, 0, 1, 1, 3, 4,\ 0)$$

where $a_1 = 0, a_2 = 4, a_3 = 9, ...a_{11} = 4, a_{12} = 0$ and the computation for this code is

$$0 + 4 + 27 + 0 + 0 + 0 + 0 + 1 + 3 + 3 + 12 + 0 = 50.$$

Since this sum is evenly divisible by 10, the number is valid.

The International Standard Book Number (ISBN):

Up until 2007, books had been identified by a ten digit number, abbreviated ISBN (now ISBN-10). For example, the book *Topics In Algebra*, Second Edition, published by Wiley, has ISBN

$$0471010901.$$

The first digit identifies the language in which the book is written, the second block of digits identifies the publisher, the third block identifies the book itself, and the final digit is the check digit. In this scheme, each digit can be a numeral $0, \ldots, 9$, or X, which represents 10. A ten digit number, this time represented as a 10-tuple $(a_1, a_{2,} a_{3,} a_{4,} a_{5,} a_6, a_7, a_8, a_9, a_{10})$ is a valid ISBN provided that

$$10a_1 + 9a_2 + 8a_3 + 7a_4 + 6a_5 + 5a_6 + 4a_7 + 3a_8 + 2a_9 + a_{10}$$

is evenly divisible by 11. In the number above, we have

$$0 + 36 + 56 + 7 + 0 + 5 + 0 + 27 + 0 + 1 = 132 = 11 \times 12,$$

so the number is indeed valid.

More recently ISBN-10 has been replaced by ISBN-13, a 13 digit number which uses almost the same computation as the UPC. The digits are multiplied by 1,3,1,3,...,1 and then added. Often both ISBN-10 and ISBN-13 appear on the book.

Exercise 2 *The text* Calculus *Second Edition by Jon Rogawski published by Freeman, have the following ISBN's:*

1. ISBN-13: 978-1-4292-6009-1
2. ISBN-10: 1-4292-6009-2.

Check that I didn't make an error in copying these numbers from the book's cover.

When we discuss the error detecting abilities of the various ID number schemes, you might wonder why the change to ISBN-13 was made.

In the case of zip codes, the check digit is not printed, but you can determine its value. For the zip code $88003 - 8001$, we look for the smallest positive integer x so that $8 + 8 + 0 + 0 + 3 + 8 + 0 + 0 + 1 + x$ is evenly divisible by 10. Since this sum is $28 + x$, the only choice for x is to be 2.

Exercise 3 *Find the check digit for the following zip codes:*

- $11210 - 2014$
- $21284 - 0301$

The purpose of the check digit is to detect errors in reading the code. For example, suppose that the zip code $88003-8001$ was incorrectly read as $88008-8001$ by reading the fifth digit as an 8 instead of as a 3. This is what we call a single digit error (in the fifth position). The sum of the digits (including the check digit 2) would then be $8+8+8+8+1+2 = 35$, which is not divisible by 10. Therefore, the postal service's scanners would detect an error, and the zip code would have to be read again.

Exercise 4 *Suppose that the zip code was incorrectly read as* $88003 - 8071$*. Would the error be recognized? Explain why or why not.*

Exercise 5 *Explain why the zip code check digit scheme will detect every single digit error in any position.*

For UPC as with the zip code, given the first eleven digits, there is enough information to uniquely determine the check digit. For example, given the partial UPC 0 7114200001, if the check digit is x, then

$$3+7+3+1+3+2+3+0+3++0+3 = 28+x,$$

which forces $x = 2$.

Exercise 6 *Determine the check digit x for the UPC* 1 4110510627 x.

Exercise 7 *What value of x will make* 1 $411x510627$ 4 *a valid UPC?*

As with the zip code scheme, the UPC adds the check digit to help detect errors.

Exercise 8 *For example, if the code* 0 7114200001 2 *was incorrectly read as* 0 7134200001 2 *by reading the fourth digit as a* 3 *instead of as a* 1*, make the relevant computation to see that the error is detected.*

Exercise 9 *Suppose the UPC number* 0 7114200001 2 *was incorrectly read as* 0 7114200051 2 *. Would the error be recognized? Explain why or why not.*

Exercise 10 *Test the sensitivity of the UPC to single digit errors in a few positions by making single digit errors one at a time in the UPC of a product around you. Does UPC detect every one of them? Explain your answer.*

For the ISBN-10, the digit X, given its usual value of 10, is only used for the check digit. As with the previous two examples, the check digit can be determined uniquely, given that it is between 0 and 10. For example, for the book *A Classical Introduction to Modern Number Theory,* published by Springer, whose ISBN starts with $0 - 387 - 97329$, the check digit, let's call it x, must result in

$$0+27+64+49+54+35+12+6+18+x = 265+x$$

being a multiple of 11. Since $0 \le x \le 10$, and $275 = 11 \times 25$, then $x = 10$. Thus, the check digit for this book is X, and so the ISBN is $0 - 387 - 97329 - X$.

Exercise 11 *Check if the following numbers are valid ISBNs:*

1. $0 - 8218 - 2169 - 5$
2. $0 - 201 - 01361 - 9$

Exercise 12 *Suppose a UPC is read, but the third digit is left out, and the result is* $07y172381751$*, where* y *represents the missing digit. Calculate, in terms of* y*, the sum needed to check if this is a valid number. Then write down the condition on* y *required for the number to be valid, and determine* y*.*

Exercise 13 *The number* $0 - 8176 - 3165 - 1$ *is an invalid ISBN (check this!). It was created by taking the ISBN number of a book and changing one digit. Can you tell which digit was changed? Explain why or explain why not not by giving two examples of a valid ISBN that differs from this one in exactly one digit.*

The ISBN scheme will detect every single digit error . Suppose that the valid ISBN

$$(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10})$$

has been misread in the i^{th} position. Then a_i has been read as $a_i + e$ for some number e with $0 < e < 11$.

Exercise 14 *Make the ISBN calculation for the misread number and observe how it differs from that of the correct number. Note the role played by the index* i *and by* e*. Then conclude that every single digit error will be detected.*