Check Digits

It's the little things that tax us and put us on the rack.
You can sit upon a mountain but not upon a tack
-*Anonymous*

Identification numbers are used in many everyday transactions. Some examples are:

- Credit Card Number
- ISBN
- NM Drivers License Number
- Social Security Number
- U. S. Post Office Money Order
- UPC (Bar Codes)
- Vehicle Identification Number (VIN)
- Zip Code (12 digit)

All of these have some common features:

Except for Social Security Numbers, all have check digits. Except for NM Drivers Licenses, all contain coded information. Why use check digits? Machines can make errors. There are errors in transmission. People can make errors. Students typically "bubble in" the wrong Social Security number on scantrons. Sometimes they even write it incorrectly.

The following table is taken from [4] page 5.

Error Type	**Form**	**Relative Frequency**	**Actual**	**Error**
Single Digit	a → b	79.1%	1234	1284
Transposition of adjacent digits	ab → ba	10.2%	1234	1324
Jump Transposition	abc → cba	0.8%	1234	3214
Twin	aa → bb	0.5%	1776	1446
Phonetic	a0 → 1a	0.5%	12403	12143
Jump Twin	aca → bcb	0.3%	12325	15355

Notice that single digit and transposition errors occur almost 90% of the time. Combined, they are the most frequent errors.

Let us look at some examples of identification numbers that are used frequently. The following will be examined: Zip Code, UPC, ISBN, and a scheme developed by IBM.

Zip Code (12 digit)

The zip code was developed by the U. S. Post Office to automate their sorting procedures. The bar code is intended to be read by a scanner used in automated mail sorting facilities to route pieces of mail. Each digit is represented by five vertical bars. Two of which are taller than the others. There are exactly ten combinations of three short and two long bars. If a printer prints incorrectly and there are not exactly two tall bars and three short, the scanner will detect an error. When this type of error is detected for one digit, it can be corrected automatically. The form of the zip code is:

gssll-rrrr-hhc

The first digit, g, represents one of ten geographical areas.
The next two, ss, identify a sectional center. This is a central mail distribution point.
Thje next two, ll, represent the local town or local post office.
The next four locate the route.
The next two the house.
The last is a check digit.

To check the zip code for errors, the twelve digits are added. If the zip code is correct the result will be a multiple of ten. I.e. the sum is 0 mod(10).

It is easy to calculate the check digit. Add the first eleven digits. The check digit is the number needed to make the sum a multiple of ten.

This scheme has some advantages:

1. This will catch all single digit errors.
2. It is easy to calculate

The scheme has some disadvantages:

1. It will not catch transposition errors

The following examples illustrate the computations and errors:

88003-8001-011 (Correct)

$$8 + 8 + 0 + 0 + 3 + 8 + 0 + 0 + 1 + 0 + 1 + 1 = 30$$

88003-7001-011 (Incorrect)

$$8 + 8 + 3 + 7 + 1 + 1 + 1 = 29$$

88003-8010-011 (Incorrect)

$$8 + 8 + 0 + 0 + 3 + 8 + 0 + 1 + 0 + 0 + 1 + 1 = 30$$

Notice that the first incorrect number is detected. The second incorrect number is not detected.

UPC (12 digit, Version A)

There are several versions of the UPC code. The one discussed here is the 12 digit version known as Version A. The bar code is intended to be read by a scanner. The width of the bars and spacing determine the machine readable code. The code is divided into left and right halves. The bar codes for digits in the left half are mirror images of those in the right half. This allows the code to be read either from right to left or vice versa.

The following table is taken from [2] page 347.

Binary UPC Coding		
Digit	Left Half	Right Half
0	0001101	1110010
1	0011001	1100110
2	0010011	1101100
3	0111101	1000010
4	0100011	1011100
5	0110001	1001110
6	0101111	1010000
7	0111011	1000100
8	0110111	1001000
9	0001011	1110100

The form of the UPC code is:

P LLLLL RRRRR C

The first digit,P, is the "product type." A code of 0, indicates a general product and is put on by the manufacturer. A code of 2 indicates a random weight item, 3 indicates drug and certain health related items, 4 indicates items marked for price reduction by the retailer, and 5 indicates cents off coupons. If the first three digits are 978, the next eight digits are part of an ISBN number.

The next five, LLLLL, with the product type form the left half of the number. This includes the code for the manufacturer. Each manufacturer is given a code by the Uniform Code Council. It may or may not be five digits long. It could be longer or shorter. Companies with a large number of products want short identification numbers. Digits seven through eleven, RRRRR, represent the code for the product. This is assigned by the manufacturer. The twelfth digit is the check digit.

To check the UPC code, add the odd numbered digits, multiply this by three, then add the even numbered digits. If the number is correct, the sum will be an even multiple of ten. I.e. it will be 0 mod(10).

To calculate the check digit. add the odd numbered digits of the base number, multiply this by three, add the even numbered digits. The check digit is the number that makes the sum an even multiple of ten.

Another way of looking at the calculation is that it is the dot product of two vectors:

$$(3\ 1\ 3\ 1\ 3\ 1\ 3\ 1\ 3\ 1\ 3\ 1)\ (a_1\ a_2\ a_3\ a_4\ a_5\ a_6\ a_7\ a_8\ a_9\ a_{10}\ a_{11}\ a_{12})$$

This scheme has some advantages:

1. This will catch all single digit errors.
2. This will catch most transposition errors.
3. It is easy to calculate.

This scheme has some disadvantages:

1. It will not catch transposition errors when the difference between the digits is five.
2. It will not catch Jump Transposition errors.

The following examples show correct and incorrect codes:

0-78910-29403-1 (Correct)

$$3(0 + 8 + 1 + 2 + 4 + 3) + (7 + 9 + 0 + 9 + 0 + 1)$$
$$3(18) + 26 = 54 + 26 = 80$$

0-78910-29<u>14</u>3-1 (Incorrect)

$$3(0 + 8 + 1 + 2 + 1 + 3) + (7 + 9 + 0 + 9 + 4 + 1)$$
$$3(15) + 30 = 75$$

0-78910-2<u>49</u>03-1 (Incorrect)

$$3(23) + 21 = 90$$

Notice the scheme detects the first error but not the second. The second error has a transposition where the difference is five.

The UPC code is scheduled to be replaced on January 1, 2007 withEAN-13. This is a European system that has 13 digits. The current UPC numbers will simply have a 0 as the first digit. The check digit will be computed by using the weight matrix (1 3 1 3 1 3 1 3 1 3 1 3 1). [8]

<u>International Standard Book Number (ISBN)</u>

This was developed to uniquely identify books. It was not developed to be scanned. It is an international standard.

The ISBN has the form: L-PPPPP-BBBBB-C

The ISBN is shown in four groups of numbers. The first number identifies the country, or language group where the book was published, the second the publisher, the third the book, and the fourth is a check digit. A first digit of 0 or 1 indicates the book was published in an English speaking country, a 2 indicates a French speaking country, a 3 indicates a German speaking country. Some countries have a code with more than one digit, Czechoslovakia is indicated by 80 and Denmark by 87. The ISBN is a ten digit number, the groups may have any length provided the total length of the number is ten digits. There is a plan to change the length of the ISBN to 13 digits. Books with current numbers will have 978 put in front of them. The 13 digit code will begin to be used on January 1, 2005. The implementation date is January 1, 2007. [9] "The check digit for the 13-digit ISBN will be updated based on the research currently being carried out by the Society of Motion Picture and Television Engineers (SMPTE—an official American National Standards Institute standards development body) for the ISAN standard." [9] The check digit will be similar to the one used in the UPC 12 digit code.

To check the ISBN number use the following algorithm:

1. Each digit has a "place" value. The first digit has a place value of ten, the next a nine, etc. until the tenth digit which has a place value of one.
2. Multiply each digit by its place value and add. The sum will be a multiple of eleven.

To compute the check digit, do the following:

1. Multiply each digit in the "base" by its place value and add.
2. The check digit will be the number that makes the sum a multiple of eleven.
3. A "digit" of ten is represented by X.

The ISBN number has the following advantages:

1. It will catch all single digit errors and all transposition errors.
2. It is not difficult to calculate.

The ISBN number has the following disadvantages:

1. It uses a "strange" character (X) to represent a "digit" (ten).
2. It works on ten digit numbers.

The following examples show correct and incorrect codes:

0-88385-720-0 (Correct)

$$10(0)+9(8)+8(8)+7(3)+6(8)+5(5)+4(7)+3(2)+2(0)+0$$
$$0 + 72 + 64 + 21 + 48 + 25 + 28 + 6$$
$$264 = 11(24)$$

0-88<u>83</u>5-720-0 (Incorrect)

$$72 + 64 + 56 + 18 + 25 + 28 + 6 = 269$$

269 is not a multiple of eleven.

<u>IBM Scheme</u>

This was developed to work with any length identification number. It uses a permutation to develop the check digit.

The form of the identification number is $a_1a_2\ldots a_{n-1}$-c

a_i is a digit.

c is the check digit.
The length of the number may be even or odd and of any length.

To check a number, apply the permutation, σ, to each digit as follows
For an even length identification number compute $\sigma(a1) + a2 + \ldots + \sigma(an\text{-}1) + an$

For an odd length identification number compute $a1 + \sigma(a2) + \ldots + \sigma(an\text{-}1) + an$

The permutation σ expressed in cyclic notation is (0)(124875)(36)(9)

Another way to calculate the permutation is to double the number and use the result modulo nine.

If the number is correct, the sum will be an even multiple of ten.

To calculate the check do the following:

1. Apply the permutation to the base number.
2a. If the length of the ID is to be even, apply the permutation to the odd numbered digits, the check will be the digit that makes the sum a multiple of ten.
2b. If the length of the ID is to be odd, apply the permutation to the even numbered digits, the check will be the digit that makes the sum a multiple of ten.

The IBM scheme has the following advantages:

1. May be used with any length number.
2. It catches all single digit errors.
3. It catches most transposition errors.
4. It does not use special characters.

The IBM scheme has the following disadvantages:

1. It will not catch transposition errors where nine and zero are transposed.

The following examples show correct and incorrect codes:

12345-5 (Correct)

2 + 2 + 6 + 4 + 1 + 5 = 20

5432-0 (Correct)

5 + 8 + 3 + 4 + 0 = 20

12<u>4</u><u>3</u>5-5 (Incorrect)

2 + 2 + 8 + 3 + 1 + 5 = 21

54<u>2</u>2-0 (Incorrect)

5 + 8 + 2 + 4 + 0 = 19

<u>Verhoeff Scheme</u>
Any check digit scheme should attempt to meet the following goals:

1. Catch all single digit and transposition errors.
2. Only use nine digits for the check digit. (No special characters.)
3. Work for identification numbers of any length.

The Verhoeff scheme was developed by J. Verhoeff in 1969. [4] It uses the following concepts:

Group
Permutation
Composition of functions
Cayley Table

A group is a mathematical entity consisting of a set of elements and a binary operation defined on the elements of the set. The binary operation will be designated by *. A group has the following properties:

1. *Closure*: For any two elements, a, b, the element a*b is also in the set.
2. *Identity*: There is an element, designated e, such that for any element, a, a*e = e*a = a.
3. *Associative*: For any three elements, a, b, c, (a*b)*c = a*(b*c)
4. *Inverse*: For any element, a, there exists an element b such that a*b = b*a = e. The element b is called the inverse of a. The element a is also the inverse of b.

Notice that a group need not be commutative. The group used in the Verhoeff scheme is not commutative.

A Cayley table is a table that shows the result of the group binary operation on the elements of a group. An example in "normal arithmetic" would be a multiplication table.

The Verhoeff scheme combines a permutation with a group. It uses the permutation σ = (0)(14)(23)(56789). It also does calculations using the Dihedral Group D_{10}.

The dihedral group D_{10} is the collection of symmetries of a regular pentagon. The elements of this set are the rotations and reflections of a regular pentagon. There are four rotations and five reflections that leave the pentagon unchanged. The identity is the movement that leaves the pentagon as it was to start. These are the ten elements of D_{10}.

The binary operation, *, is the composition of two symmetries. Each symmetry is represented by a digit.

The following notation will be used:
The symbol, ∘ , is used to denote composition.

$\sigma \circ \sigma = \sigma^2$

$\sigma \circ \sigma \circ \ldots \circ \sigma = \sigma^n$

The symbol, * , is used to denote the binary operation in D_{10}.

3 * 8 = 6

8 * 3 = 5

The Cayley Table for D_{10} is:

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

The form of the identification number is: $a_1a_2\ldots a_{n-1}$-c

a_i is a digit

c is the check digit.

The number may be any length.

In order to check the number use the following calculation:

$\sigma^{n-1}(a_1) * \sigma^{n-2}(a_2) * \ldots * \sigma(a_{n-1}) * a_n = 0$

To calculate the check digit apply the above calculation to the base number, then select the inverse from D_{10}.

This has the following advantages:

1. Detects all errors listed in the table.
2. Works with any length number.
3. Only uses digits (no special characters).

It has the following disadvantages:

1. Uses unfamiliar operations.
2. Not easily calculated manually.

<u>Bibliography</u>

1. Gallian, Joseph, *Assigning Drivers License Numbers*, Mathematics Magazine, Vol. 64, No. 1, February 1991
2. Gallian, Joseph, *For All Practical Purposes*, *Part III*, W. H. Freeman and Company, 2001
3. Gallian, Joseph, www.d.umn.edu/~jgallian/fapp5
4. Kirtland, Joseph, *Identification Numbers and Check Digit Schemes,* The Mathematical Association of America, 2001, tenth printing
5. Mohr, Jonathan, http//www.augustana.ab.ca/~mohrj/algorithms/checkdigit.html
6. Paris, Bernd-Peter, *http://thalia.spec.gmu.edu/~pparis/classes/notes_101/node44.html*, 1998
7. Paris, Bernd-Peter, *http://thalia.spec.gmu.edu/~pparis/classes/notes_101/node45.html*, 1998
8. *http://www.barcodeisland.com/ean13.phtml*
9. *http:/www.isbn.org*, R. R. Bowker