

## Chapter 11 : Errors And Accuracy

---

### Chapter Objectives

At the completion of this chapter, you should be able to:

- ❖ identify different types of errors;
- ❖ calculate absolute and relative error;
- ❖ estimate maximum absolute error;
- ❖ generate check digit.

### 11.1 Introduction

In building a system, measures must be taken against possible human mistakes. Procedures must be set up to help people to detect any mistakes. The role of data preparation or data entry are prone to errors, therefore they must not be overlooked as the ill effect may affect the entire system seriously.

### 11.2 Mistakes

At the man/machine input interface, i.e. when a person wants to enter data into a computer system, the common mistakes are:

Lose track of the source documents.

Enter data wrongly.

Invalid datatype.

e.g. enter alphabets instead of numeric data.

Out of the valid data range.

e.g. enter 102 while the valid range is 1 to 100.

Transposition of characters.

e.g. enter 23322 while the actual data is 23223.

Interference in data transmission from a remote site.

e.g. 00100101 is transmitted along a cable and on the way due to voltage surge, the data becomes 10100101.

and many more.

Mistakes are due to human carelessness.

Parity bit check.

*Example:*

---

If 'B' is represented as 100 0010, we can use even parity conversion to produce a parity bit.

```

IF odd number of 1's THEN
    parity bit ← 1
ELSE
    parity bit ← 0
ENDIF
    
```

The parity bit may be appended to the left of the 100 0010, 'B' is represented as 0100 0010. Suppose the data is received as 0110 0010, there are three 1's which should give 1 in the parity bit, then the mistake is detected.

Check digit method.

- A Check digit is derived from the true data before transmission, which is then appended to the right of the true data for validation.

*Step 1:* A true data 445, each of the digits is assigned a weight,

$$\begin{array}{r}
 4 \quad 3 \quad 2 \quad \leftarrow \text{Weightage} \\
 4 \quad 4 \quad 5
 \end{array}$$

$$\begin{array}{r}
 5 * 2 = 10 \\
 4 * 3 = 12 \\
 4 * 4 = 16 \\
 \hline
 38
 \end{array}$$

*Step 2:* The sum of 38 is to be divided by a modulus. (In this example, 11 is used as it can detect most of the mistakes).

38/11 gives quotient 3 and the remainder 5.

*Step 3:* Remainder 5 is subtracted from the modulus (11) to obtain the check digit '6'.

$$11 - 5 = 6$$

*Step 4:* The check digit can be appended to the right of 445, to give 4456, which could be validated.

In order to verify the correctness of 4456, we could just multiply 4456 by the set of weight, 6 is assigned the weight of 1,

$$\begin{array}{r}
 4 \quad 3 \quad 2 \quad 1 \quad \leftarrow \text{Weightage} \\
 4 \quad 4 \quad 5 \quad 6 \\
 (4*4) + (4*3) + (5*2) + (6*1) = 44
 \end{array}$$

44 is divisible by 11, thus conclude that 4456 is received correctly.

It has been tested that MOD 11 is most efficient in detecting mistakes, reason being 11 is a prime number. Though bigger number is better, the

small increase in efficiency is overwhelmed by large increase in processing.

*Check point*

List five types of mistakes which may occur during the data preparation and data entry stages of the processing cycle.

Give one other example each of the appropriate use of the hash and batch totals. Describe the principle of a parity check and how even and odd parity differ.

Using mod 7 and weighting 2,3,4,5..., append appropriate check digits to:

v. 439

w. 1058535

Using the same mod 7, determine whether the following are valid:

x. 1794

y. 10545722

## 11.3 Errors

### 11.3.1 Inherent Error

It is error that already exist by itself in the measurement scale. E.g. Take a ruler of measurement, though there are many small divisions, there is still the smallest division that is not sub-divided further. Thus there is uncertainty in reading the length as mentioned above.

If the length fall between 5.1cm and 5.2cm, the smallest division is 0.1cm, then the reported reading may be 5.15 cm and the *inherent error* is 0.05cm.

### 11.3.2 Induced Error

It is the error which is brought in from 'outside' due to external factors.

**Example:** In a fixed point 8-bit register in computer, the implied point is to the right of 5 bits.

If a data 1010.01110 is to be stored, rounding or truncation is necessary.

Rounding the above data is stored as

0	1	0	1	0	1	0	0
---	---	---	---	---	---	---	---

Thus the error is brought in due to the limitation of computer.

*Check point*

What is the difference between error and mistakes?

---

## 11.4 Measurement of Error

Absolute error

- If a true value 2.5 is reported to be 3 (round to full number) an error of 0.5 is induced. This error is called the absolute error.
- Absolute error  
 $E_a = | \text{reported value} - \text{true value} |$

| | means the absolute value is taken, the sign is not considered at all.

Relative error

- Another more meaningful measurement of error is relative error, because it expresses the absolute error as a proportion of the true value (if the true value is not available, the reported value is used.)

$$E_r = \frac{|\text{absolute error} / \text{True Value}|}{|\text{absolute error} / \text{True Value}| * 100\%}$$

Maximum relative error between any TWO values reported as

$$R = 100 * (\text{error bound}/R)$$

**Example:** 3.6 is rounded to 1D, the Maximum error between 3.55 and 3.65 =  $100 * (0.1/3.6) = 2.78 \%$

*Check point*

What is the error bound of the values reported as:

- 3.5cm
- 45.3 seconds

Calculate the maximum relative error on:

- 13.5
- 130.5
- 1300.5

## 11.5 Error Propagation

Inherent errors in data values, may produce further errors as they are operated on arithmetically, these errors may be termed as Induced errors. This spread of errors is called the Error Propagation.

Addition

**Example:** 5+7, each of the data values are rounded to whole number.

$$\begin{array}{ll} 5.5 + 7.5 = 13 & \text{Max} \\ 4.5 + 6.5 = 11 & \text{Min} \end{array}$$

$$\begin{array}{l} E_r = 100 * (1/12) = 8.33\% \\ E_r \text{ of } 5 = 10\% \\ E_r \text{ of } 7 = 7.1 \% \end{array}$$

You may try 13 + 6, a conclusion may be derived.

**Conclusion:** Relative error of the sum is between the relative error of the individual reported values.

$$\begin{array}{l} E_a = \text{absolute error of} \\ E_a(A + B) = E_a(A) + E_a(B) \end{array}$$

Subtraction

**Example:** 8-4, each rounded to whole number.

$$\begin{array}{ll} 8.5 - 3.5 = 5 & \text{Max} \\ 7.5 - 4.5 = 3 & \text{Min} \end{array}$$

$$\begin{array}{l} E_a(8-4) = 1 = E_a(8) + E_a(4) \\ E_r = 100 * (1/4) = 25\% \\ E_r \text{ of } 8 = 100 * (0.5/8) = 6.25\% \\ E_r \text{ of } 4 = 100 * (0.5/4) = 12.5\% \end{array}$$

You may try 3-2

**Conclusion:** Relative error of the subtraction is more than the relative error of the individual reported values.

$$E_a(A-B) = E_a(A) + E_a(B)$$

Multiplication

**Example:** 3\*4\*5, each rounded to whole number.

$$\begin{array}{ll} 3*4*5 = 60, & \\ 3.5*4.5*5.5 = 86.625, & \text{Max} \\ 2.5*3.5*4.5 = 39.375, & \text{Min} \end{array}$$

$$\text{Error bound} = 86.625 - 39.375 = 47.25$$

This error bound is very big indeed, and it is not symmetrical about 60 (the result of multiplying 3,4 and 5). If we quantify the relative error, it would be as follow:

$$\begin{array}{l} \text{Take half of the error bound,} \\ E_r = 100 * (23.625/60) = 39.375\% \\ E_r \text{ of } 3 = 100 * (0.5/3) = 16.667\% \end{array}$$

$$E_r \text{ of } 4 = 100 * (0.5/4) = 12.5\%$$

$$E_r \text{ of } 5 = 100 * (0.5/5) = 10\%$$

You may try  $3*3*3$ , what conclusion you may draw?

**Conclusion:** Relative error of the multiplication is approximately the sum of the relative errors of the individual reported values.

$$E_r(A*B*C) = E_a(A)/A + E_a(B)/B + E_a(C)/C$$

Induced error in multiplication is the sum of the inherent error in each reported value.

Division

**Example:**  $7.5/5.3$ , each is rounded to 1D

$$7.55/5.25 = 1.438, \quad \text{Max}$$

$$7.45/5.35 = 1.393, \quad \text{Min}$$

$$\text{Error bound} = 1.438 - 1.393 = 0.045$$

$$E_a = 0.045/2 = 0.0225$$

$$E_r \text{ of } (7.5/5.3) = 100 * (0.0225/1.415) = 1.59\%$$

$$E_r \text{ of } 7.5 = 100 * (0.05/7.5) = 0.667\%$$

$$E_r \text{ of } 5.3 = 100 * (0.05/5.3) = 0.943\%$$

Try  $3.5/1.5$ ,

**Conclusion:** Relative error of the division is approximately the sum of the relative errors of individual reported values.

$$E_r(A/B) = E_a(A)/A + E_a(B)/B$$

Finally, try a combination of arithmetic operation, multiplying and division.

**Example:**  $(3.5*13.5*1.5)/(7.0*4.5*40.5)$

In what ways the error propagated is maximised?

Multiply the top row and bottom row respectively, then divide the numerator by the denominator.

In what ways the error may be minimised?

$(3.5/7.0) * (13.5/4.5) * (1.5/40.5)$ , the individual divisions will produce values which are relatively small, overflow error may be reduced.

Rounding

- Rounding errors frequently occur when doing manual decimal arithmetic. They occur with even greater frequency in computer arithmetic.

**Example:** If 0.0056 is rounded to 3 decimal places the answer is 0.006. A rounding error of  $0.006 - 0.0056 = 0.0004$  is therefore induced.

If 0.0055 is rounded to 3 decimal places, we may use the following rule:

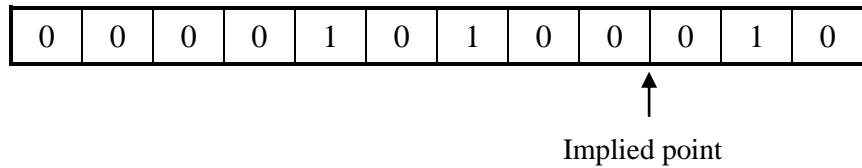
- If the third decimal place is odd, do nothing;
- else add 1 to the third decimal place.

0.0055 is rounded to 0.005.

If 0.0045 is rounded to 3 decimal places, the reported value is 0.005.

Truncation

- If a storage space that can only accommodate 4 digits is used to store a number 67890, the least significant digit '0' will be chopped off giving 6789 as the number stored.
- If 1010.0010111 is to be stored in 12-bit register with 4 bits to the right of the implied point. The stored data will be



Overflow

- Overflow is different from truncation error in that the most significant digit will be drop-out instead of the least significant. Storing 67890 in 4 digits, the most significant digit '6' is chopped off.
- This error is very serious in computer programming, thus many compilers and interpreters are written such that they can deliver 'overflow error' message.

Significant figures

- At times, we can only take a certain number of significant figures. Take for example the population of Singapore, we may say there are 3,500,000 people taking only 2 significant figures. Even though we may have found the population to be 3,547,893, as the less significant digits 47,893 is small compared to 3,500,000.
- A real number 2.03452355 may be expressed in 5 significant figures, as 2.0345.

Converting from one number system to another

- To convert 0.7 decimal to binary number will induce an error as shown below:

.7	*	2
1.4	*	2
0.8	*	2
1.6	*	2
1.2	*	2
0.4	*	2
0.8...		

The series of multiplication is recurring, if we to stop at 7 binary places, we have to accept that an error is induced.

$$0.7 = (0.1011001)_2$$

Rounding, truncation, overflow, significant figures and change of number system will all lead to *induced* errors in data.

*Check point*

What is the effect on error if  $3.5 \times 7.25$  is:

- dd. Calculated to integer precision;
- ee. Cut to 1D;
- ff. Rounded to 1D;
- gg. Calculated to 3 significant figures;
- hh. Worked out 'in full', with no cut or rounding.

### 11.6 To Control Errors

We want to calculate the volume of a sphere of radius 5.5 cm, the formula below can be used:

$$V = (4/3) \times \pi r^3$$

Take  $\pi = 3.141$ , and  $r = 5.5$

		$E_r$	
r	5.5	0.91%	Inherent
V	696.7785	$3 \times 0.91 = 2.73\%$	Induced

This means that V has an absolute error of  $(2.73/100) \times 696.7785 = 19.02$ .

This means that range of Volume is  $696 \pm 19$  cubic centimetres.  
The decimal places are not significant.

To reduce the induced error, we should, if possible, take 'r' to 5D and do the same for p. Raise 'r' to the third power, multiply by n, multiply by 4 and finally divide by 3. Only by taking such precautions, could we minimise the errors.

**Answers to check points**

- 20. There are many possible sources of mistakes at this stage;
  - Document omitted or misread;
  - Transposition errors;
  - Characters omitted;
  - Mis-keying;
  - Out of range;



Invalid datatype;

There are many possible answers:

Hashing the invoice numbers;

Hashing the dates;

Batching totals of amount of money;

Batching totals of number of items withdrawn from store;

Parity checking is testing the make-up of the binary digits in a coded eight-bit byte.

Even parity: by making the leftmost digit a 1, if the total numbers of 1's in the remaining seven bits is even;

Odd parity: by making the leftmost digit a 1, if total number of 1's in the remaining seven bits is odd;

Data	Hash Total	Divide by 7, take remainder	7-remainder	Final data
439	43	1	6	4396
1058535	117	5	2	10585352

Data	Hash Total	Divide by 7, remainder	
1794	47	5	not valid
10545722	105	0	valid

Mistakes are made due to human carelessness in handling data. Errors arise from the data itself, i.e. Inherent error; it may be brought in due to external factors like rounding, limit of computer systems, i.e. Induced error.

Error bound = 3.45cm -- 3.55cm

Error bound = 45.25s -- 45.35s

Data	Relative error
13.5	0.37%
130.5	0.038%
1300.5	0.0038%

The reported answer is 25.375.

ii.	$3 \times 7 = 21$	Relative error = 17.2%
jj.	25.2	Relative error = 0.7%
kk.	$3.5 \times 7.3 = 25.55$ ,	Relative error = 0.7%
ll.	$25.375 = 25.4$ (3 sig.f.)	Relative error = 0.1%
mm.	$3.55 \times 7.255 = 25.75525$ ,	Max
	$3.45 \times 7.245 = 24.99525$ ,	Min
	Error bound = 0.76	
	Relative error = $100 \times (0.38/25.375) = 1.5\%$	

## Points to Remember

To distinguish mistake from error.

To be aware of how to reduce mistakes.

The methods are *hash total*, *batch total*, *parity-bit method* and *check digit method*.

Inherent error.

- It is error that exist by itself in the measurement scale.

Induced error.

- It is error that is brought in from 'outside' by external factors. Usually, induced error is the result of propagation of inherent errors.

To measure error using two quantities.

- Absolute error  $E_a = \text{'reported value - true value'}$
- Relative error  $E_r = E_a/\text{True value}$

Error can spread, or propagate.

- $E_a(A+B) = E_a(A) + E_a(B)$

As a result of addition, the absolute error of (A+B) is the sum of absolute error of A and that of B.

- $E_a(A-B) = E_a(A) + E_a(B)$
- $E_r(A*B*C) = E_a(A)/A + E_a(B)/B + E_a(C)/C$
- $E_r(A/B) = E_a(A)/A + E_a(B)/B$

Types of errors.

- Rounding; Truncation; Overflow; Conversion

Divide a number by another, before multiplying them, is a method to reduce possible overflow error.

## 11.7 Past Year Questions

21. Given that  $X = 5.63$  and  $Y = 7.82$ , where each number has been rounded to 2 decimal places, calculate to 4 significant figures.

mn. Relative error when they are added [ 2 ]

oo. Relative error for  $X * Y$  [ 2 ]

$x = 0.34$  and  $y = 0.55$ , both corrected to 2 decimal places.

Find

pp. the absolute error of  $2x + y$  [ 2 ]

qq. the relative error of  $2x + y$  [ 2 ]

Two values are recorded as 8.7 and -4.3, both correct to 1D.

What is the ABSOLUTE ERROR when they are added together? [ 2 ]

Calculate the absolute and relative error of  $a + b$  given  $a = 3.5$  (corrected to 1D) and  $b = 1.55$  (corrected to 2D). [ 4 ]

Pick the correct completion for each of the following sentences: [ 3 ]

rr. Discarding least significant digits of a number so it will fit in a fixed-size storage location [ 1 ]

i. is a method of implementing floating-point numbers.

ii. is called truncation.

iii. does not happen on modern computers.

ss. The difference between the reported value and the true value of a quantity [ 1 ]

i. is the absolute error.

ii. is zero.

iii. is the inherent error.

tt. Error can arise in computations [ 1 ]

i. only when the initial readings for values are inaccurate.

ii. if the mantissa of a computed result won't fit in the storage space allocated for it without loss of data.

iii. because no real numbers can be represented exactly in a floating-point representation.

—

\_\_\_\_\_ errors are ones which reflect the smallest measurement possible on the scale used. [ 1 ]

Subtraction is \_\_\_\_\_ susceptible to error propagation than is addition. [ 1 ]

Explain briefly what is inherent and induced error. [ 4 ]

Using mod 11 and weighting of 2, 3 and 4 for units, tens, and hundreds columns, calculate a check digit for 834. [ 3 ]

An account number consists of 10 digits, the tenth digit being a CHECK DIGIT, MODULES 11.

uu. Calculate an appropriate check digit for 005611350. [ 4 ]

vv. Which of these numbers would be declared invalid: 0765213141, 0265073147? [ 2 ]

Using mod 7 and weighting of 2, 3, 4 for the units, tens and hundreds columns append an appropriate CHECK DIGIT to each of 928 and 921. [ 3 ]

Using modules 11 and weighting of 2, 3 and 4 for the units, tens and hundreds columns:

ww. append a CHECK DIGITS to the number 758. [ 2 ]

xx. show why there is a problem when generating a check digit for 748, and suggest how to overcome this problem. [ 2 ]

yy. determine whether or not 8752 is a VALID CODE. [ 2 ]

a. Using mod 11 and weighting of 2, 3 and 4 for the units, tens and hundreds columns, append a check digit to 974. [ 2 ]

zz. Show how a transposition error occurring in the number 974 would be detected. [ 2 ]

22. a. What is parity in a parity check? [ 1 ]

aaa. When will a parity check fail? [ 1 ]

23. a. How is data verification performed? [ 2 ]

bbb. Why might data verification be used? [ 1 ]

ccc. How might the data entry be validated? [ 2 ]

Describe what is the purpose of a check digit? [ 2 ]

24. a. Write the decimal fraction 4.0 as binary fraction using only four binary digits after the decimal point. [ 1 ]

ddd. Convert your answer back to denary, and determine the absolute error. [ 1 ]

eee. Calculate the relative error. [ 2 ]