VOLUME-17 Part B and C

CONTENTS

VII. Experimental Techniques and data analysis

VII. 1. Data Interpretation and Analysis

Tables are one of the commonest and earliest way to storing and tabulating data. Tabulation is the systematic and scientific presentation of quantitative data in such a form that it is able to elucidate the problem under consideration. An aspirant is expected to be able to go through a table, quickly and effectively i.e. with understanding. Essentials of a table: there are six essential elements of a table –

1. Title 2. Stub (section containing row heading)
3. Body 4. Column caption (heading of each column)
5. Footnotes 6. Source

Error Analysis

The knowledge we have of the physical world is obtained by doing experiments and making measurements. It is important to understand how to express such data and how to analyze and draw meaningful conclusions from it.

If the result of a measurement is to have meaning it cannot consist of the measured value alone. An indication of how accurate the result is must be included also. Indeed, typically more effort is required to determine the error or uncertainty in a measurement than to perform the measurement itself. Thus, the result of any physical measurement has two essential components:

1. A numerical value (in a specified system of units) giving the best estimate possible of the quantity measured.

2. The degree of uncertainty associated with this estimated value. For example, a measurement of the width of a table would yield a result such as 95.3 +/ – 0.1 cm.

Significant Figures

The significant figures of a (measured or calculated) quantity are the meaningful digits in it. There are conventions which you should learn and follow for how to express numbers so as to properly indicate their significant figures.

- Any digit that is not zero is significant. Thus 549 has three significant figures and 1.892 has four significant figures.

- Zeros between non zero digits are significant. Thus 4023 has four significant figures.

- Zeros to the left of the first non-zero digit are not significant. Thus 0.000034 has only two significant figures. This is more easily seen if it is written as 3.4×10^{-5}.

- For numbers with decimal points, zeros to the right of a non zero digit are significant. Thus 2.00 has three significant figures and 0.050 has two significant figures. For this reason it is important to keep the trailing zeros to indicate the actual number of significant figures.

- For numbers without decimal points, trailing zeros may or may not be significant. Thus, 400 indicates only one significant figure. To indicate that the trailing zeros are significant a decimal point must be added. For example, 400 has three significant figures, and 4×10^2 has one significant figure.

- Exact number have an infinite number of significant digits. For example, if there are two oranges on a table, then the number of oranges is 2.000… Defined numbers are also like this. For example, the number of centimeters per inch (2.54) has an infinite number of significant digits, as does the speed of light (299792458 m/s).

There are also specific rules for how to consistently express the uncertainty associated with a number. In general, the last significant figure in any result should be of the same order of magnitude (i.e. in the same decimal position) as the uncertainty. Also, the uncertainty should be rounded to one or two significant figures. Always work out the uncertainty after finding the number of significant figures for the actual measurement.

Classification of Error

Generally, errors can be divided into two broad and rough but useful classes: systematic and random. Systematic errors are errors which tend to shift all measurements in a systematic way so their mean value is displaced. This may be due to such things as incorrect calibration of equipment, consistently improper use of equipment or failure to properly account for some effect. In a sense, a systematic error is rather like a blunder and large systematic errors can and must be eliminated in a good experiment. But small systematic errors will always be present. For instance, no instrument can ever be calibrated perfectly.

Other sources of systematic errors are external effects which can change the results of the experiment, but for which the corrections are not well known. In science, the reasons why several independent confirmations of experimental results are often required (especially using different techniques) is because different apparatus at different places may be affected by different systematic effects. Aside from making mistakes, the reason why experiments sometimes yield results which may be far outside the quoted errors is because of systematic effects which were not accounted for.

Random errors are errors which fluctuate from one measurement to the next. They yield results distributed about some mean value. They can occur for a variety of reasons.

- They may occur due to lack of sensitivity. For a sufficiently small change an instrument may not be able to respond to it or to indicate it or the observer may not be able to discern it.

- They may occur due to noise. There may be extraneous disturbances which cannot be taken into account.

- They may be due to imprecise definition.

- They may also occur due to statistical processes such as the roll of dice.

Random errors displace measurements in an arbitrary direction whereas systematic errors displace measurements in a single direction. Some systematic error can be substantially eliminated (or properly taken into account). Random errors are unavoidable and must be lived with. Many times you will find results quoted with two errors. The first error quoted is usually the random error, and the second is called the systematic error. If only one error is quoted, then the errors from all sources are added together. (In quadrature as described in the section on propagation of errors).

A good example of "random error" is the statistical error associated with sampling or counting. For example, consider radioactive decay which occurs randomly at a same (average) rate. If a sample has, on average, 1000 radioactive decays per second then the expected number of decays in 5 seconds would be 5000. A particular measurement in a 5 second interval will, of course, vary from this average but it will generally yield a value within 5000 +/ –. Behaviour like this, where the error,

$$\Delta n = \sqrt{n_{\text{expected}}}$$

is called a Poisson statistical process. Typically if one does not know n_{expected} it is assumed that, $n_{\text{measured}} = n_{\text{expected}}$ in order to estimate this error.

A. Mean Value: Suppose an experiment were repeated many, sat N, times to get,

$$X_1, X_2, ...X_k ....X_N$$

N measurements of the same quantity *x*. If the errors were random then the errors in these results would differ in sign and magnitude. So if the average or mean value of our measurements were calculated,

$$\overline{X} = \frac{X_1 + X_2 + ...X_k + ... + X_N}{N} = \frac{\sum_{k=1}^{N} X_k}{N}$$

some of the random variations could be expected to cancel out with others in the sum. This is the best that can be done to deal with random errors: repeat the measurement many times, varying as many "irrelevant" parameters as possible and use the average as the best estimate of the true value of *x*. (It should be pointed out that this estimate for a given *N* will differ from the limit as $N \to \infty$ the true mean value; though, of course for *N* it will be closer to the limit.) In the case of the previous example: measure the height at different times of day, using different scales, different helpers to read the scale, etc.

Doing this should give a result with less error than any of the individual measurements. But it is obviously expensive, time consuming and tedious. So, eventually one must compromise and decide that the job is done. Nevertheless, repeating the experiment is the only way to gain confidence in and knowledge of its accuracy. In the process an estimate of the deviation of the measurements from the mean value can be obtained.

B. Measuring Error: These are several different ways the distribution of the measured values of a repeated experiment such as discussed above can be specified.

- **Maximum Error:** The maximum and minimum values of the data set, X_{mx} and X_{mi}, could be specified. In these terms, the quantity

$$\Delta x_{mx} = \frac{X_{mx} - X_{mi}}{2}$$

is the maximum error. And virtually no measurements should ever fall outside $\bar{x} \pm \Delta x_{mx}$.

- **Probable Error:** The probable error, Δx_{prob}, specifies the range $\bar{x} \pm \Delta x_{mx}$ which contains 50% of the measured values.

- **Average Deviation:** The average deviation is the average of the deviations from the mean, $\Delta x_{ky} = \dfrac{\sum_k |x_k - \bar{x}|}{N}$

For a Gaussian distribution of the data, about 58% will lie within $\bar{x} \pm \Delta x_{ky}$.

- **Standard Deviation:** For the data to have Gaussian distribution means that the probability of obtaining the result x is,

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-x_0)^k}{k}}$$

where x_0 is most probable value and σ, which is called the standard deviation, determines the width of the distribution. Because of the law of large numbers this assumption will tend to be valid for random errors. And so it is common practice to quote error in terms of the standard deviation of a Gaussian distribution fit to the observed data distribution. This is the way you should quote error in your reports.

It is just as wrong to indicate an error which is too large as one which is too small. In thhe measurement of the height of a person, we would reasonably expect the error to be +/– ¼" if a careful job was done, and may be +/– 3/4" if we did a hurried sample measurement. Certainly saying that a person's height is 5'8.250" +/–0.002" is ridiculous (a simple jump will compress your spine more than this) but saying that a person's height is 5'8" +/–6" implies that we have at best, made a very rough estimate!

C. Standard Deviation: The means is the most probable value of a Gaussian distribution. In terms of the mean, the standard deviation of any distribution is,

$$\sigma = \sqrt{\frac{\sum_k (x_k - \bar{x})^2}{N}}$$

Continued with...Page 5 Onwards.... It's So Goooood!!!, Buy it now...!