

Introduction

Preface :

In college, we not only focus on learning professional knowledge but also cultivating good experimental attitudes and habits. During the class, students can obtain the last knowledge as well as learn how to use the instruments correctly.

Pursuing the truth is the virtue of science. In the lab, we should realize what we are doing exactly and record all original data and that is pursuing the truth. Of course, checking whether the data are reasonable is no doubt important, but it is worthy of paying more attention to how to figure out and analyze the questions. After the experience, we should learn which factors cause the results respectively. We should know what leads to the correct answer and what leads to the wrong one. Those abilities mentioned above will benefit you in the future and that is the purpose of this class.

Take Coulomb's law as example, according to Field and Wave Electromagnetics (2/e) wrote by David K. Cheng: 「The force between two charged bodies, q_1 and q_2 , that are very small in comparison with the distance of separation, R_{12} , is proportional to the product of the charges and inversely proportional to the square of the distance, the direction of the force being along the line connecting the charges.」 Present it mathematically:

$$\vec{F}_{12} = \hat{a}_{R_{12}} k \frac{q_1 q_2}{R_{12}^2}$$

Although it is the result from experiments, there is still a question: How small should the charge bodies be in order to be very small in comparison with the distance of separation, R_{12} ? We should keep in mind that a point charge is an ideal assumption as well as a point mass, and they do not exist in reality. We can separate two charged bodies as far as possible to fit “very small in comparison with the distance of separation” as long as the volumes are defined. However, a new question comes. Coulomb force could be too weak compared with external force from environment if the bodies are separated too far, and the data would be not accurate. Moreover, we do the experiment in the limited space. The designers and experimenters should make compromise with the current conditions. From above, inaccuracy is not avoidable.

However it does not mean experiments are worthless due to inaccuracy. Experiments are essential to examine the truth. The wonderful theory could be abandoned without experimental support.

How can we overcome inaccuracy during the experiment? Practicing and being cautious lead to excellent consequence.

Here are some basic concepts of pursuing the truth during the experiment.

1. All data should be recorded without modification. Even the data are wrong due to clerical error, just draw two lines over them. Remember, keep all row data.
2. Record the data in the form with unit.
3. Single datum is meaningless because it could contain serious error. Therefore every experiment should be recorded five times.

Unit (SI unit) :

We should put a space between value and unit like 3 kg 、 4 K 、 5 cm ...etc.

When a name is use as unit, follow the rule below:

Newton (name) → newton (unit) → N (unit)

Hertz (name) → hertz (unit) → Hz (unit)

Ampere (name) → ampere (unit) → A (unit)

When the unit is not a name, write it in small letter:

meter = meter = m, hour = hr, second = s; but liter = liter = l = L is the only exception.

In any measurement, we could not get the exact value even with fine instruments; therefore, estimation, significant figure, and error come. Estimation is not mentioned since we have already learnt it in high school. Here we only describe rounding, significant figure, and error.

Scientific Notation :

Scientific notation (also referred to as scientific form, standard form or standard index form) is a way of expressing numbers that are too big or too small to be conveniently written in decimal form. It is commonly used by scientists, mathematicians and engineers, in part because it can simplify certain arithmetic operations. On scientific calculators it is known as "SCI" display mode

In scientific notation all numbers are written in the form $a \times 10^n$. (m times ten raised to the power of n)

Most calculators and many computer programs present very large and very small results in scientific notation, typically invoked by a key labelled 「EXP」 or 「E」 (for exponent), depending on vendor and model.

[example-1] 1.632E-19 to 1.632×10^{-19}

[example-2] 3.00E8 to 3.00×10^8

Significant Figure :

23.76 are four significant figures. We can express the length in term of 2.376×10^{-4} km, 0.0002376 km, 0.2376 m, 23.76 cm, 237.6 mm, 237600 m, ...etc. There are all four significant figures in those mentioned above, because those “scarin” 0 are the location of decimal point. Therefore unit conversion does not affect the digit of significant figure.

If 0 does not stand for the decimal point, then boss “zero” and “nonzero” are significant figures. For example, 0.50006 and 34.209 are five significant figures. There are some rules for “0”:

1. The leftmost nonzero digit is the most significant digit.
2. All digits between the least and most significant digits are significant digits.
3. If there is a decimal point, the rightmost digit is the least significant digit, even if it is 0.

A. Arithmetic of Significant

(a) Addition and Subtraction Rules

If uncertainty exists in any row, then the answer is uncertain after addition and subtraction.

[Example-1] $49.57 + 2903.4050 + 9.679 + 5.08 = 2967.7340 \rightarrow 2967.73$

[Example-2] $123.579 - 12.41 = 111.169 \rightarrow 111.17$

(b) Division, Multiplication and Square Root Rules

The significant digit of product is determined by less significant digit of dividend, divisor, multiplicand or multiplier.

[Example-1] $9500635 \times 0.58 = 5510368.30 \rightarrow 5500000 = 5.5 \times 10^6$

[Example-2] $36.94 \times 28.55 = 1054.6370 \rightarrow 1055$

[Example-3] $9357.98 \div 508 = 18.421 \rightarrow 18.4$

B. Banker's Rounding

Rounding financial records is extremely important. For accuracy reasons, computations are often performed with more digits than just hundredths of a dollar. The results are then rounded to hundredths of a dollar. This may all seem boring to you, but adding a penny here and there adds up real fast to large financial institutions.

1. If the part of value that will be aborted, when its first digit is equal to or greater than 6, it should round to upward.

[Example] $30.29 \rightarrow 30.3$

2. If the part of value that will be aborted, when its first digit is equal to or less than 4, remove the value directly.
[Example] $30.24 \rightarrow 30.2$

3. If the part of value that will be aborted is just one digit and the value is 5. Consider the previous digit was odd or even.
 - (a) If there are any non-zero digits beyond the thousandths digit, then add one to the hundredths.
[Example] $30.256 \rightarrow 30.3$

 - (b) Now we know the thousandths digit = 5 and all other digits beyond thousandths are 0. The number is exactly between two pennies. We now "round towards even," that is, if the hundredths digit is odd, then add one. Otherwise, the hundredths digit is even, so leave it alone.
[Example] $30.350 \rightarrow 30.4$
[Example] $30.850 \rightarrow 30.8$

Error Representation :

No measurements are exactly accurate. Error always exists. In order to minimize it, we should know where the error from.

Error can be deviated into system error, artificial error, and random error:

A. Systematic Error

- (a) **Imperfect Calibration of Measurement Instruments:** It comes from improper design or aging of instruments. We should calibrate the instruments before experiment.

- (b) **Changes in Environment:** For example, thermal expansion and contraction will cause error during length measurement; therefore, an air conditioner is a solution.

- (c) **Imperfect Methods of Observation:** It comes from inaccurate theory and improper measurement setup.

B. Random Error

It is always present in a measurement. It is caused by inherently unpredictable fluctuations in the readings of a measurement apparatus or in the experimenter's interpretation of the instrumental reading. The only solution is to increase times and figure the data statistically.

Measurement is limited intrinsically; we must admit that any physics, order, and even empirical formula are not "equal" to the truth we try to express. We just describe it personally. Here we are going to discuss how to express the data statistically.

C. Artificial Error

It is often caused by carelessness like calculation error and clerical error. The solution is doing the experiment repeatedly (In this class, 5 times are needed and it is better to do it in tern) to find out unreasonable data. Do not falsify the wrong data. Repeats are the only solution and remember do not take the data solely, because the new data does not belong to the original ones. Additionally, poor habits are also the possible reason for the artificial error.

Data Representation :

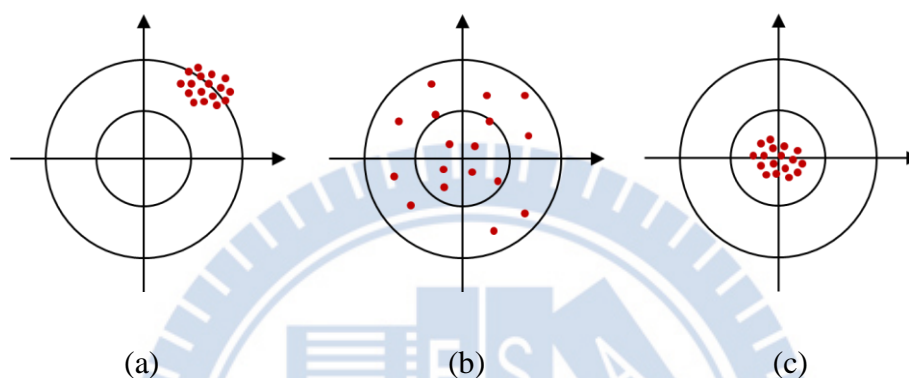


Figure 1. Precision and accuracy illustration

In order to describe precisely physical quantity measured from the experiment, quantity, precision and accuracy, and unit should be included:

- A. Quantity :** Every quantity (estimation included) should be written in form of significant figure or scientific notation except for some physically-defined value and numerical constant (ex: e ...etc.)
- B. Precision and Accuracy :** Usually expressed in form of " $a \pm b$ ". Strictly speaking, precision and accuracy are different.
- (a) **Precision** – It is a measure of how well the result has been determined, without reference to its agreement with the true value, and it also reflects the random error. In figure 1(a), the data are precise and random error is small.
- (b) **Accuracy** – It is a measure of the correctness of the result and it reflects the systematic error. In figure 1(b), the data are accurate and systematic error is small.
- (c) **Precision and Accuracy** – The synthetic index is random error and systematic error. In figure 1 (c) the data are both precise and accurate. Random error and systematic error are both small.

Statistical Analysis :

Statistical analysis is often applied to data analyzing and it is a logistical and powerful instrument.

A. Mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \sum_{i=1}^n \frac{x_i}{n}$$

To be notice, mean does not represent the real value and even not the most possible value. We can state it nothing more than “representative”.

B. Deviation

$$d_i = x_i - \bar{x}$$

C. Average Deviation

$$D = \frac{\sum_{i=1}^n |d_i|}{n}$$

D. Average Deviation Percentage

$$\frac{D}{\bar{x}} \times 100\%$$

E. Standard Deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^n d_i^2}{n-1}}$$

The variance σ_2 is defined as the limit of the average of the squares of the deviations from the mean and standard deviation is the square root of σ_2 . When numbers of data are enough (infinity is ideal), they become normal distribution. Standard deviation shows how much variation or "dispersion" there is from the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values.

Additionally, the formula we learnt in senior high school $\sigma = \sqrt{\frac{\sum_{i=1}^n d_i^2}{n}}$ is called group standard deviation strictly, The difference is in the denominator, group standard deviation strictly, and we won't pay attention to it. We should just aware when dealing with census, group standard deviation strictly should be applied. When we deal with the experiment data, standard deviation is needed.

F. Standard Deviation of the Mean

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n d_i^2}{n(n-1)}}$$

If we do some more measurements, the standard deviation would not change appreciably. On the other hand, the standard deviation of mean would slowly decrease we increase n . However, the factor \sqrt{n} grow rather slowly as we increase n . For example, if we want to improve our precision by a factor of 10, that means we will have to increase n by a factor of 100! Thus, in practice, if we want to increase our precision appreciably, we will probable do better increase our technique than to rely merely on increased numbers of measurements.

G. Coefficient of Correlation

$$\gamma = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$\gamma = 0$	Non-correlation
$0.0 < \gamma < 0.3$	Low correlation
$0.3 < \gamma < 0.7$	Moderate correlation
$0.7 < \gamma < 1.0$	Highly correlation
$\gamma = 1$	Fully correlation

H. Experiment Result

The data should be record in form of $\bar{x} \pm \sigma_x$ with unit.

Mean and Error Transfer :

Two or more experimental results of the arithmetic, should be consider the error transfer.

Let $x = \bar{x} \pm \sigma_x$ 且 $y = \bar{y} \pm \sigma_y$

A. Error Transfer of Addition and Subtraction

$$\overline{x \pm y} = \bar{x} \pm \bar{y} \quad , \quad \sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$$

$$x \pm y = (\bar{x} \pm \bar{y}) \pm \sigma_{x+y} \quad , \quad \sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2}$$

In general form :

$$\sigma_N^2 = \sum_{i=1}^n \sigma_i^2$$

B. Error Transfer of Addition and Subtraction of Multiplication and Division

$$\overline{xy} = \bar{x} \times \bar{y} \quad ; \quad \sigma_{xy} = \sqrt{\left[\left(\frac{\sigma_x}{x} \right)^2 + \left(\frac{\sigma_y}{y} \right)^2 \right]} \times (\bar{x} \times \bar{y})$$

$$x \times y = (\bar{x} \times \bar{y}) \pm \sigma_{xy}$$

$$\frac{\bar{x}}{\bar{y}} = \frac{\bar{x}}{\bar{y}} \quad ; \quad \sigma_{\frac{x}{y}} = \sqrt{\left[\left(\frac{\sigma_x}{x} \right)^2 + \left(\frac{\sigma_y}{y} \right)^2 \right]} \times \frac{\bar{x}}{\bar{y}}$$

$$\frac{x}{y} = \left(\frac{\bar{x}}{\bar{y}} \right) \pm \sigma_{\frac{x}{y}}$$

In general form :

$$\left(\frac{\sigma_N}{y} \right)^2 = \left(\frac{\sigma_1}{y_1} \right)^2 + \left(\frac{\sigma_2}{y_2} \right)^2 + \dots + \left(\frac{\sigma_n}{y_n} \right)^2$$

Where y is the average derived and y_1, y_2, \dots, y_n are average of every element in the calculation.

C. The Relation of Standard Deviation before and after Calculation

$$\overline{x^l \times y^m} = (\overline{x^l} \times \overline{y^m}) ; \left(\frac{\sigma_{x^l y^m}}{\overline{x^l} \times \overline{y^m}} \right)^2 = l^2 \left(\frac{\sigma_x}{\overline{x}} \right)^2 + m^2 \left(\frac{\sigma_y}{\overline{y}} \right)^2$$

$$x^l \times y^m = (\overline{x^l} \times \overline{y^m}) \pm \sigma_{x^l y^m}$$

D. General Standard Deviation

Set $N = f(x, y)$, and then

$$\sigma_N = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 \sigma_y^2}$$

Least Square Regression Analysis :

It is an analysis instrument in common use. By fitting the optimized regression we are able to minimize the sum of square of vertical distance to the regression.

Be given the data set:

$$(x_i, y_i), i = 1, 2, 3, \dots, n$$

A. Linear Regression

If the optimized regression is the form of $y = f(x) = Ax + B$ with A and B unknown, then the sum of square of vertical distance to the regression is:

$$D(A, B) = \sum_{i=1}^n [f(x_i) - y_i]^2 = \sum_{i=1}^n [Ax_i + B - y_i]^2$$

The so-called optimization is to decrease $D(A, B)$:

$$\frac{\partial D}{\partial A} = 0 ; \frac{\partial D}{\partial B} = 0$$

$$\begin{aligned} \frac{\partial D}{\partial A} &= \sum_{i=1}^n \frac{\partial}{\partial A} [Ax_i + B - y_i]^2 = \sum_{i=1}^n 2[Ax_i + B - y_i]x_i \\ &= 2 \sum_{i=1}^n [Ax_i^2 + Bx_i - x_i y_i] = 2 \left[A \sum_{i=1}^n x_i^2 + B \sum_{i=1}^n x_i - \sum_{i=1}^n x_i y_i \right] = 0 \end{aligned}$$

$$\begin{aligned}\frac{\partial D}{\partial B} &= \sum_{i=1}^n \frac{\partial}{\partial B} [Ax_i + B - y_i]^2 = 2 \left[\sum_{i=1}^n Ax_i + \sum_{i=1}^n B - \sum_{i=1}^n y_i \right] \\ &= 2 \left[A \sum_{i=1}^n x_i + nB - \sum_{i=1}^n y_i \right] = 0\end{aligned}$$

And the solutions are

$$A = \frac{n \left[\sum_{i=1}^n x_i y_i \right] - \left[\sum_{i=1}^n x_i \right] \cdot \left[\sum_{i=1}^n y_i \right]}{n \left[\sum_{i=1}^n x_i^2 \right] - \left[\sum_{i=1}^n x_i \right]^2} ; B = \frac{\left[\sum_{i=1}^n y_i \right] - A \left[\sum_{i=1}^n x_i \right]}{n}$$

We can just input (x_i, y_i) to engineering calculator or software like MS-excel, and then A and B can be obtained.

B. Exponential Regression

It describes the data in form of $y = A \times e^{Bx}$. Also it can be applied through calculator or software.

C. Logarithmic Regression

It describes the data in form of $y = A \times \ln x + B$. Also it can be applied through calculator or software.

D. Power Regression

It describes the data in form of $y = A \times x^B$. Also it can be applied through calculator or software.