With this type of distribution about 68 % of all values will fall within one standard deviation on either side of the mean, 95 % will fall within two standard deviations, and 99.7% within three standard deviations.

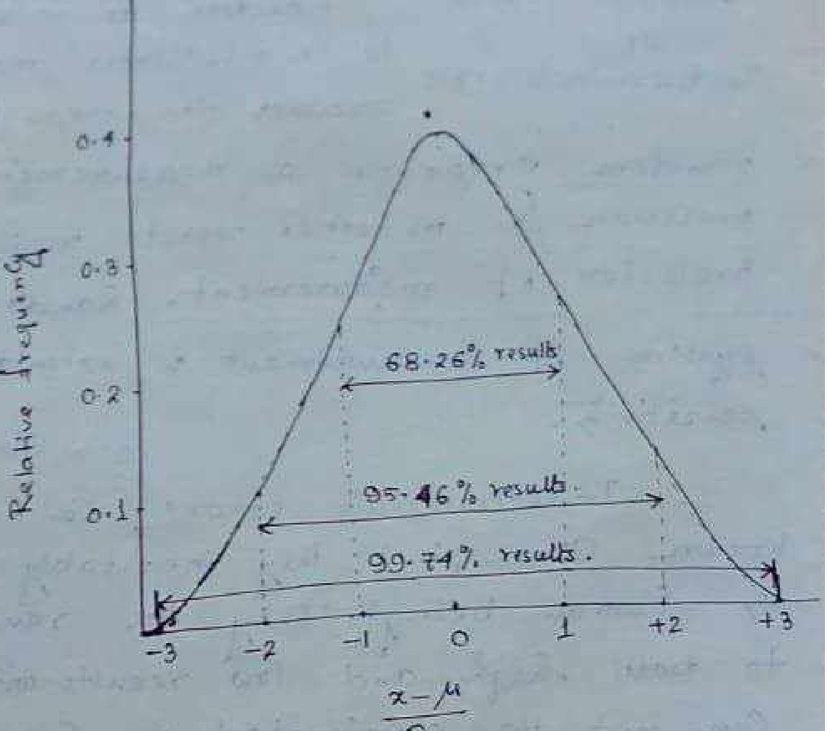


Fig: Normal distribution Curve; relative frequencies of deviations from the mean for a normally distributed infinite population; deviations $(x-\mu)$ are in units of σ.

* Statistical treatment of finite samples:

The term parameter refers to quantities such as μ and σ that define a population or distribution. This is in contrast to quantities such as the data values x that are variables. The term statistic refers to an estimate of a parameter that is made from a sample of data, as discussed below.

Mean: The mean of a finite number of measurements, $x_1, x_2, x_3, \ldots x_n$, is often designated $\bar{x}$ to distinguish it from μ. Of course, $\bar{x}$ approaches μ as a limit when n, the number of measured values, approaches infinity.

The sample mean $\bar{x}$ is found from
$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$
Where N is the number of measurements in the sample set.

The same equation is used to calculate the population mean, μ
$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}.$$
Where N is now the total number of measurements in the population.

* Error :

Chemical analyses are affected by at least two types of errors.

Type I (Random or indeterminate error) : affect measurement precision,
Causes data to be scattered more or less symmetrically around a mean value.

Type II (Systematic or determinate error) : affect the accuracy of results.
Causes the mean of a data set to differ from accepted value.

* A third type of error is **gross error**. Gross errors differ from indeterminate and determinate errors. They usually occurs only occasionally, are often large and may cause a result to be either high or low. They are often the product of human errors.

* **Outlier** : An outlier is an occasional result in replicate measurements that differs significantly from the rest of the results.

Type I (Systematic error) :

Systematic errors have a definite value and an assignable cause, and are of the same magnitude for replicate measurements made in the same way. Systematic errors lead to bias in measurement results. Bias affects all of the data in a set in the same way and that it bears a sign.

Bias measures the systematic error associated with an analysis. It has a negetive sign if it causes the results to be low and a positive sign otherwise. (Cases the result to be high).

There are three types of systematic errors :

1. Instrumental errors : Caused by nonideal instrument behaviour, instabilities in their power supplies, by faulty calibrations, or by use under inappropriate conditions (uncalibrated glassware)

2. Method error : Arise from nonideal chemical and physical behavior of analytical systems. eg → incompleteness of m^ns, in gravimetry error may arise owing to appreciable solubility of ppt.

* Coprecipitation of impurities, slight solubility of a ppt, side reactions, incomplete m^ns, impurities in reagents. etc.

3. Personal error & operational error. : result from the carelessness, inattention, or personal limitations of the experimenter.
not connected with the method of analysis, mechanical loss of materials underweighing, overweighing, observation, colour change etc.

Random error or Indeterminate error : → Caused by the many unControllable variables.

Random or indeterminate error Causes data to be sCattered more or less systematiCally around the mean value. In general random error in a measurement is reflected by its precision or in other words random error affects the precision of measurement. Random error arise when a system of measurement is extended to its maximum sensitivity.

as is Certain to happen or unavoidably

Indeterminate errors Can not be attributed to any known Cause, but they inevitably attend measurements made by human beings. They are random in nature and lead to both high and low results with equal probability. They Can not be eliminated or Corrected and are the ultimate limitation on the measurement.

* Distribution of Normal Errors :

The spread of a series of results obtained from a given set of measurements Can be ascertained from the value of the standard deviation. But it gives no indication ~~about how results are distributed.~~

If a large number of repliCate readings, at least 50, are taken of a Continuous variable, eg. titrimatric end point, the results obtained will usually be distributed about the mean in a roughly symmetriCal manner. The mathematiCal model that best satisfies such a distribution of random errors is Called Normal (or Gaussian) distribution.

* A Gaussian or normal error Curve is a Curve that shows the symmetriCal distribution of data around the mean of an infinite set of data. also Called bell-shaped Curve.

The Curve satisfies the equation :

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

It is important to know that the Greek letter σ and μ refer to the standard deviation and mean respectively of a total population.

* **Median** : The median of an odd number of results is simply the middle value when the results are listed in order; for an even number of results, the median ~~are~~ is the average of the two middle ones.

Range : For a finite number of values, the simplest measure of variability is the range; which is the difference between the largest and smallest values.

Average Deviation : For a large group of data, which are normally distributed, the average deviation approaches 0.8σ. To Calculate the average or mean deviation, one simply finds the differences between individual results and the mean, regardless of sign, adds these individual deviations and devides by the number of results :

$$\text{Average Deviation} = \bar{d} = \frac{\sum_{i=1}^{i=n} |x_i - \bar{x}|}{n}$$

Relative average deviation : Average deviation is expressed relative to the magnitude of the measured quantity, for example, as a percentage, ppt...

$$R.A.D\ (\%) = \frac{\bar{d}}{\bar{x}} \times 100 = \frac{\sum_{i=1}^{i=n} |x_i - \bar{x}|/n}{\bar{x}} \times 100$$

$$R.A.D\ (ppt) = \frac{\sum_{i=1}^{i=n} |x_i - \bar{x}|/n}{\bar{x}} \times 1000.$$

Standard Deviation : The standard deviation is much more meaningful statistically than is the average deviation. The symbol 's' is used for the standard deviation of a finite number of values; 'σ' is reserved for the population parameter.

The population standard deviation σ; which is a measure of the precision of a population of data, is given by the equ^n:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

N, is the number of data points making up the population.

When it is applied to a small sample of data. Thus, the sample standard deviation 's' is given by the equ^n.

$$s = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}}.$$

$(x_i - \bar{x})$ represent the deviation d_i of value x_i from the mean $\bar{x}$.
$(N-1)$ is the number of degrees of freedom.

If N is large (say 50 or more); then of course, it is immaterial whether the term in the denominator is $(N-1)$ (which is strictly correct) or N.

When $(N-1)$ is used instead of N, s is said to be an unbiased estimator of the population standard deviation σ. If this substitution is not used, the calculated s will be less on average than the true standard deviation σ; that is s will have a negative bias.

* Variance : The square of standard deviation is called variance, s^2.

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}.$$

* Coefficient of variance (v) or, Percent relative standard deviation :

When the standard deviation is expressed as a percentage of the mean, it is called the coefficient of variance, v.

$$v = \frac{s}{\bar{x}} \times 100.$$

** The normality of a solution is determined by four separate titrations, the results being 0.2041, 0.2049, 0.2039 and 0.2043. Calculate the mean, median, range, average deviation, relative average deviation, standard deviation and coefficient of variation.

$$\text{Mean } \bar{x} = \frac{0.2041 + 0.2049 + 0.2039 + 0.2043}{4}$$

$$= 0.2043.$$

$$\text{Median } M = \frac{0.2041 + 0.2043}{2} = 0.2042.$$

$$\text{Range } R = (0.2049 - 0.2039) = 0.0010.$$

$$\text{Average deviation } \bar{d} = \frac{(0.0002) + (0.0006) + (0.0004) + (0.0000)}{4}.$$

$$= 0.0003.$$

$$\text{Relative Average Deviation } \frac{\bar{d}}{\bar{x}} \times 1000 \text{ ppt.}$$

$$= \frac{0.0003}{0.2043} \times 1000 = 1.5 \text{ ppt.}$$

Standard Deviation : $s = \sqrt{\dfrac{(0.0002)^2 + (0.0006)^2 + (0.0004)^2 + (0.0000)^2}{4-1}}$

$= 0.0004.$

Coefficient of variance $v = \dfrac{0.0004}{0.2043} \times 100$

or % R.S.D.

$= 0.2\ \%.$

**. Standard Deviation of Calculated Results :

Standard deviation of a Sum or Difference

Consider the summation.

$$\begin{array}{rl} +0.50 & (\pm 0.02) \\ +4.10 & (\pm 0.03) \\ -1.97 & (\pm 0.05) \\ \hline 2.63 & \end{array} \left.\vphantom{\begin{array}{c}a\\a\\a\end{array}}\right\} \text{Absolute standard deviations.}$$

If the three individual s.d happen by chance to have the same sign, the s.d of the sum could be as large as +0.02+0.03+0.05 = +0.10, or −0.10. On the other hand, it is possible that the three s.d could combine to give an accumulated value of zero: (−0.02−0.03+0.05 or +0.02+0.03−0.05). More likely, however the s.d. of the sum will lie between these two extremes.

The variance of a sum or difference is equal to the sum of the individual variances. The most probable value for a standard deviation of a sum or difference can be found by taking the square root of the sum of the square of the individual absolute standard deviations. So for the computation

$$y = a(\pm s_a) + b(\pm s_b) + c(\pm s_c).$$

The variance of y, s_y^2 is given by

$$s_y^2 = s_a^2 + s_b^2 + s_c^2$$

Hence the s.d of the result s_y is

$$s_y = \sqrt{s_a^2 + s_b^2 + s_c^2}.$$

Where s_a, s_b, s_c are the standard deviations of the three terms making up the result.

** $\therefore\ s_y = \sqrt{(0.02)^2 + (0.03)^2 + (0.05)^2} = 0.06$

and the sum should be reported as $2.63\ (\pm 0.06)$.

(4) Analytical.

Standard deviations in exponential Calculation:

Consider the relationship $y = a^x$.

Where the exponent x can be considered free of uncertainty.

So, the r.s.d in y resulting from the uncertainty in a is

$$\frac{S_y}{y} = x\left(\frac{S_a}{a}\right).$$

Thus the r.s.d of the square of a number is twice the r.s.d of the number, the r.s.d of the Cube root of a number is one third that of the number.

** The s.d in measuring the diameter d of a sphere is ±0.02 Cm. What is the s.d in the Calculated volume V of the sphere if d = 2.15 Cm?

From the equation for the volume of a sphere, we have.

$$V = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi\left(\frac{d}{2}\right)^3 = \frac{4}{3}\pi\left(\frac{2.15}{2}\right)^3 = 5.20 \text{ Cm}^3.$$

we may write; $\frac{S_v}{V} = 3\,\frac{S_d}{d} = 3 \times \frac{0.02}{2.15} = 0.0279.$

The absolute s.d in V is then

$$S_v = 5.20 \times 0.0279 = 0.145$$

Thus $V = 5.2\,(\pm 0.1)$ Cm³ (rounded).

** The solubility product Ksp for the silver salt AgX is $4.0\,(\pm 0.4) \times 10^{-8}$. The molar solubility of AgX in water is

Solubility $= (K_{sp})^{1/2} = (4.0 \times 10^{-8})^{1/2} = 2.0 \times 10^{-4}$ M.

What is the uncertainty in the Calculated solubility of AgX in water?

$y = (a)^x$;

$\therefore \frac{S_a}{a} = \frac{0.4 \times 10^{-8}}{4.0 \times 10^{-8}}$.

$\frac{S_y}{y} = \frac{1}{2} \times \frac{0.4}{4.0} = 0.05$

$\therefore S_y = 2.0 \times 10^{-4} \times 0.05 = 0.1 \times 10^{-4}$

$\therefore$ Solubility $= 2.0\,(\pm 0.1) \times 10^{-4}$ M.

** It is important to note that the error propagation in taking a number to a power is different from the error propagation in multiplication.

Uncertainty in the square of $4.0\,(\pm 0.2)$

relative error $\frac{S_y}{y} = 2\left(\frac{0.2}{4}\right) = 0.1$ or 10%.

if $y = a_1 a_2$ i.e. $a_1 = 4.0\,(\pm 0.2)$, $a_2 = 4.0\,(\pm 0.2)$

(Product of two numbers)

$\frac{S_y}{y} = \sqrt{\left(\frac{0.2}{4}\right)^2 + \left(\frac{0.2}{4}\right)^2} = 0.07$ or 7%.

So measurement is independent of one another.

* Standard deviation of a Product or Quotient :

In this Case, the r.s.d of two of the numbers in the Calculation are ~~the~~ larger than the result itself. Evidently ; we need a different approach for multiplication and division. In this situation relative standard deviation ~~of a~~ is introduced. The r.s.d of a product ~~and~~ or quotient is determined by the r.s.d of the ~~sum~~ numbers forming the Computed result. For example, in the Case of

$$y = \frac{a \times b}{c}.$$

We obtain the r.s.d. S_y/y of the result by summing the squares of the r.s.d of a, b, and c and extracting the square root of the sum :

$$\frac{S_y}{y} = \sqrt{\left(\frac{S_a}{a}\right)^2 + \left(\frac{S_b}{b}\right)^2 + \left(\frac{S_c}{c}\right)^2}$$

**
$$\frac{4.10(\pm 0.02) \times 0.0050(\pm 0.0001)}{1.97(\pm 0.04)} = 0.010406(\pm ?).$$

Applying the equ^n :

$$\frac{S_y}{y} = \sqrt{\left(\frac{0.02}{4.10}\right)^2 + \left(\frac{0.0001}{0.0050}\right)^2 + \left(\frac{0.04}{1.97}\right)^2}$$

$$= \sqrt{(0.0049)^2 + (0.0200)^2 + (0.0203)^2} = 0.0289$$

To Complete the Calculation, we must find the absolute standard deviation of the result.

$$S_y = y \times (0.0289)$$
$$= 0.0104 \times 0.0289$$
$$= 0.000301$$

and we Can write the answer and its uncertainty as $0.0104(\pm 0.0003)$

Note that if y is a negative number, we should treat S_y/y as an absolute value.

** Calculate the s.d of the result of

$$\frac{[14.3(\pm 0.2) - 11.6(\pm 0.2)] \times 0.050(\pm 0.001)}{[820(\pm 10) + 1030(\pm 5)] \times 42.3(\pm 0.4)} = 1.725(\pm ?) \times 10^{-6}.$$

Ans: rounded to $1.7(\pm 0.2) \times 10^{-6}$

** Standard deviations of Logarithm and Antilogarithms :

For $y = \log a$; $S_y = 0.434 \frac{S_a}{a}$.

For $y = \text{Antilog}\, a$: $\frac{S_y}{y} = 2.303\, S_a$.

Thus, the absolute s.d of the logarithm of a number is determined by the relative s.d of the number; Conversely the r.s.d of the antilogarithm of a number is determined by the absolute s.d of the number.

** Calculate the a.s.d of the results of the following calculations. The a.s.d for each quantity is given in parentheses.

a) $y = \log\left[2.00\,(\pm 0.02) \times 10^{-4}\right] = -3.6990 \pm ?$

b) $y = \text{antilog}\left[1.200\,(\pm 0.003)\right] = 15.849 \pm ?$

c) $y = \text{antilog}\left[45.4\,(\pm 0.3)\right] = 2.5119 \times 10^{45} \pm ?$

a) We must multiply the r.s.d by 0.434.

$$S_y = 0.434 \times \frac{0.02 \times 10^{-4}}{2.00 \times 10^{-4}} = 0.004$$

Thus $Y = \log\left[2.00\,(\pm 0.02) \times 10^{-4}\right] = -3.699\,(\pm 0.004)$. (rounded)

b)

$$\frac{S_y}{y} = 2.303 \times (0.003) = 0.0069.$$

$$S_y = 0.0069 \cdot y = 0.0069 \times 15.849 = 0.11.$$

Thus $y = \text{antilog}\left[1.200\,(\pm 0.003)\right] = 15.8 \pm 0.1$. (rounded)

c)

$$\frac{S_y}{y} = 2.303 \times (0.3) = 0.69.$$

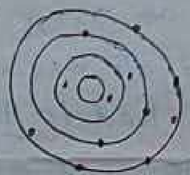
$$S_y = 0.69 \times 2.5119 \times 10^{45} = 1.7 \times 10^{45}.$$

Thus $y = \text{antilog}\left[45.4\,(\pm 0.3)\right] = 2.5\,(\pm 1.7) \times 10^{45}$.

* **Precision** : The term precision refers to the agreement among a group of experimental results. i.e the closeness of result that have been obtained in exactly the same way. it implies nothing about their relation to the true value.

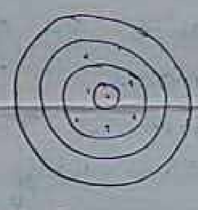
Accuracy : Accuracy indicate the closeness of measurement to its true or accepted value and is expressed by the error (the smaller the error, the greater the accuracy).

* The basic difference between precision and accuracy is that, where as accuracy measures agreement between a result and the accepted/true value, but Precision on the other hand, describes the agreement among several results obtained in the same way. We can determine precision just by measuring replicate samples. Accuracy is often more difficult to determine because the true value is usually unknown. An accepted value must be used instead.



Low accuray, low precision

Low accuray, high precision



High accuray, low precision.

High accuray, high precision.

Accuracy is expressed in terms of either absolute or relative error.
~~Precision is commonly stated in terms of the standard deviation, average deviation, or range.~~

Absolute error $E = x_i - x_t$. where x_i = measured quantity
x_t = true value/accepted value.

Relative error $\dot{E} = \frac{x_i - x_t}{x_t} \times 100\%$.

Precision is commonly stated in terms of the standard deviation; variance; and Coefficient of variation. These three are functions of how much an individual result x_i differs from the mean, which is called the deviation from the mean d_i

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