Analytical Methods of Chemistry :-

- An 'analysis' provides chemical or physical information about a sample.
- The components of interest in the sample are called analytes and the remainder of the sample is called matrix.

In an analysis we determine the identity, concentration we measure one or more of the analyte's chemical or physical properties.

Techniques, Methods, Procedure and Protocols :-

A technique is any chemical or physical principle that can be used to study an analyte. Many techniques have been used to determine concentration of lead in drinking water. eg, in graphite furnace atomic absorption spectroscopy lead is atomized, and the ability of the free atoms to absorb light is measured. Thus both a chemical (atomization) and a physical principle (absorption of light) are used in this technique.

A method is the application of a technique for determination of a specific analyte in a specific matrix. The graphite furnance atomic absorption spectroscopy for determination of lead level in water is different from the method of determination of lead level in soil or in blood. Though both have same technique.

A procedure is a set of written direction detailing how to apply a method to a perticular sample, including information on proper sampling, handling of interferents and validating results.

A protocol is a set of written guidlines detailing the procedures that must be followed.

A method is the application of a technique to a specific matrix for a specific analyte. Methods for determining the concentration of lead in drinking water can be developed using any of the techniques — insoluble lead salt such as $PbSO_4$, $PbCrO_4$ can be determined by gravimetric method. Lead form several soluble complexes that can be determined by complexometric titration or if the complexes are highly absorbing, in a spectrophotometric method can be used. Lead in gaseous free atom state (like fuel exhaust) can be measured by an atomic absorption spectroscopy. Lead can have multiple oxidation states (Pb, Pb^{2+}, Pb^{4+}) makes coulometric, potentiometric, and voltametric methods feasible.

The requirement of the analytes the best method by choosing its following criteria: accuracy, precision, sensativity, selectivity, robustness, ruggedness, scale of operation, analysis time, availability of equipment and cost.

Accuracy: Accuracy is a measure of how closely the result of an experiment agrees with expected result.

eg: We have to weight 1.000 g of $K_2Cr_2O_7$.

Student 1: Weight 1.001 g $K_2Cr_2O_7$

Student 2: " 1.005 g "

Student 3: " 0.996 g "

Now experimental result of student-1 have more accuracy than student 2 or student 3 because, the result is more close to the expected result. Now accuracy of a set of experiments can be determined by Error. Laser the error more accurate the result.

$$\% \text{ Error} = \frac{\text{obtained result} - \text{expected result}}{\text{expected result}} \times 100.$$

ie. Student 1 % of Error $= \frac{1.001 - 1.000}{1.000} \times 100 = 0.1\%$

" 2 " $= \frac{1.005 - 1.000}{1.000} \times 100 = 0.5\%$

" 3 " $= \frac{0.996 - 1.000}{1.000} \times 100 = 0.4\%$

So the student-1 have more accurate result.

Precision:- When a sample is analyzed several times, the individual result are rarely the same. Instead, the results are randomly scattered. Precision is a measure of the variability. The closer the agreement between individual analyses, the more precise the result.

eg: 3 students have to weight 1.000 g of $K_2Cr_2O_7$.

Student 1	Student 2	Student 3
Weighing 1 : 1.001	Weighing 1 : 1.002	
" 2 : 1.005	" 2 : 1.001	
" 3 : 0.998	" 3 : 1.001	

Here for student 2, three experments have very close result whereas student 1 have random result. So, student 2 results have more precision.

Sensitivity: It is the ability of an instrument to differentiate small variation of data of two parrallel experiment.

eg we have to weight 1.006 g of $K_2Cr_2O_7$. There two balance one is three digit and another 4 digit. First one can not differentiate two values 1.0062 and 1.0063, it will show 1.006 for both of cases. But the second one can. So the second weighing machine is more sensative than first one.

On the other hand, sensativity of method also arise. Suppose estimation of lead by graviimetric method gives two digit data (eg x.0x) where spectrophotometric method give three digit data (eg x.00x) then spectrophotometric method will be called more sensative.

Robustness and Ruggedness:-

For a method to be useful it must provide reliable result. Unfortunately, methods are subject to a variety of chemical and physical interferences that contribute uncertainity to the analysis. When a method is relatively free from chemical interference, it can be applied to the determination of analytes in a wide variety of sample matrices. Such methods are considered robust.

Random variations in experimental conditions also introduce uncertainity. If a method's sensitivity is highly dependent on experimental condition such a temperature, pH, reaction time etc, the slight change in those conditions may lead to significantly different result. A rugged method is relatively insensative to changes in experimental condition.

"Selectivity":-

An analytical method is selective if its signal is a function of only the amount of analyte present in the sample.

Suppose signal of an analyte denotated as S_A and signal of interference as S_I. The signal strength 'S' is dependent on amount of analyte n (mole) or concentration (c) only.

i.e $S = kn$

or $S = kC$.

So any analyte $S_A = k_A n_A$ k_A = sensativity of analyte

or $S_A = k_A C_A$

But if some interference present there then -

$$S_{sample} = S_A + S_I = k_A n_A + k_I n_I$$

or $$= k_A C_A + k_I C_I$$

So the method should minimum S_I (ie $k_I n_I$ or $k_I C_I$) then it will be more selective.

Selectivity of a method is defined by its selectivity coefficient ($K_{A,I}$).

$$K_{A,I} = \frac{k_I}{k_A}$$

Lesser the value of $K_{A,I}$, the method will be more selective and applicable.

Similarly scale of operation also important, in which scale the result we need (mg scale micro gram scale) according to presence of analyte, we have to choose the method.

Some method needs for long time whereas some method for analysis are quicker. Some method needs ~~costly~~ costly equipement whereas some need common aparatus. Availibity of instruments and cost of experiment also important for choosing method for determination or analysis of analytes.

Calibration and standerization :-

After choosing method, next be very important task of chemist is calibration of aparatus or instrument and standarization of method.

Calibration ensures that the equipment or instrument used to measure the signal is operating correctly by using a standard known to produce an exact signal.

Standarization is the process of experimentally determining the relationship b/w the signal and the amount of analyte.

Sampling:- Proper choosing of method doesn't always guarantee to accurate result-. Error in collection of sample may lead to error in result. Biased or nonrepresentative and contamination of sample are two sources of sampling error.

Validation:- Before a procedure can provide useful analytical information, it is necessary to demostrate that it is capable to providing acceptable result. Validation is an evaluation of wheather the precision or accuracy obtained by following the procedure.

Characterizing Measurement and Results:-

Measurement of Central tendency

Mean: The mean $\bar{X}$, is the numerical average obtained by dividing the sum of the individual measurement by the number of measurement.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

eg suppose the mass of 7 coins having weight-

3.080, 3.094, 3.107, 3.056, 3.112, 3.174, 3.198

$$\bar{X} = \frac{3.080 + 3.094 + 3.107 + 3.056 + 3.112 + 3.174 + 3.98}{7}$$

$$= 3.117g$$

The mean is the most common estimator of central tendency.

Median: The median, X_{med}, is the middle value when data are ordered from the smallest to largest value. When the number of measurement is 'n'. Then Median will be $\frac{n}{2}$ when n = even and $\left(\frac{n+1}{2}\right)$ for n = odd.

Here in set of 7 data, (lowest to highest) ie n=7 and the median will $\frac{7+1}{2}$ is 4th data. ~~ie 3.05~~

ie 3.056, 3.080, 3.094, (3.107), 3.112, 3.174, 3.198

Measurement of Spread:-

If the mean and median provides an estimate of ~~coin~~ coin's true mass, then the ~~mass~~ spread of the individual measurement

must provides an estimate of the variability in the masses of individual coins.

Range : The range, w, is the difference b/w the largest and smallest values in the data set.

$$\text{Range} = w = X_{\text{largest}} - X_{\text{smallest}}.$$

eg In the data set of coin's set $w = (3.198 - 3.056)\,g$
$= 0.142\,g$.

The range provides information about the total variability in the data set, but does not provides any information about the distribution of individual measurement.

Standard Deviation :=

Standard deviation 's' is defined as —

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

- Where x_i is one of the n individual measurement, $\bar{x}$ is the mean.

Relative standard deviation

$$s_r = \frac{s}{\bar{x}}$$

From standard deviation we can get how measurement for a group are spread out from average value (mean) or expected value. A low standard deviation means that most of the numbers close to average. A high standard deviation means that the numbers are more spread out.

eg For the set of weight of the coins

$$s = \sqrt{\frac{(3.080-3.117)^2 + (3.094-3.117)^2 + (3.107-3.117)^2 + (3.056-3.117)^2 + (3.112-3.117)^2 + (3.174-3.117)^2 + (3.198-3.117)^2}{(7-1)}}$$

$$= \sqrt{\frac{0.01556}{6}} = 0.051$$

Variance:- Variance is defined as the square of standard deviation.

"Characterizing Experimental Errors":-

Realizing that our data for the mass of coin's can be characterized by a measure of central tendency and a measure of spread suggest two questions. 1st, does our measure of central tendency agree with the true value? second, why are our data scattered around the central value? Errors associated with central tendency reflects the accuracy of the analysis, but the precision of the analysis is determined by those error associated with spread.

Accuracy is a measure of how close a measure of central tendency is to the true or expected value μ. Accuracy is usually expressed as either an absolute error

$$E = \bar{X} - \mu.$$

or a percentage of relative error.

$$Er = \frac{\bar{X} - \mu}{\mu} \times 100$$

Determinate error:- Any systematic error that cause a measurement or result to always to be high or too small; can be traced or identifiable source.

Determinate error are in four categories-

(i) Sampling error: An error introduced during the process of collecting a sample for analysis.

(ii) Method error: An error due to limitations in the analytical method used to analyze a sample.

$$S_{meas} = kn + S_{rea}$$

S_{meas} = Signal of the analyte ; k = sensitivity n = amount of analyte

S_{rea} = Signal of interference. Method error comes when $S_{rea} \neq 0$.

(iii) Measurement error:- An error due to limitations in the equipment and instruments used to make measurement.

(iv) Personal error:- An error due to to biases introduced by the analyst.

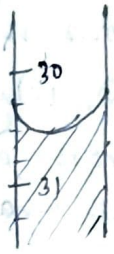
Identifying Determinate Error:-

Some determinate error can be detected experimentally, if repeated experiments error is constant then it for all sample.

Indeterminate Error:-

Any random error that causes some measurement or results to be too high while others are too low. It effect the distribution of measurement around a central value and are characterised by a random variation in both magnitude and direction.

eg a buret with scale divisions every 0.1 ml has an indeterminate error of ±0.1 - 0.03 ml when estimating volume to the hundredth of a milliliter. Due to the maniscus.



Evaluating Indeterminate error:-

Although it is impossible to eliminate indeterminate error, its effect can be minimized if the sources and relative magnitude of the indeterminate error are known. Indeterminate error may be estimated by an appropriate measure of spread. Typically, a standard deviation is used.

Uncertainity :- ~~It is express~~ It expresses the range of possible values that a measurement or result might reasonably be expected to have. The precicision of an analysis, whether reported as a range or a standard deviation, is calculated from experimental data and provides an estimation of indeterminate error affecting measurement, that might effect our result. Uncertainity accounts for both determinate error and indeterminate error.

Suppose we buy a pipet of 10 ml. we used it without calibration. The pipet's tolerance value ±0.02 ml. Now you used it ten times for volume transfer and what you get the result you have standard deviation 0.006 ml which is lower than 0.02 ml. So the precision is 0.006 ml which lower than the value of uncertainty 0.02 ml.

Propagation of uncertainity : If you use this pipet for 100 ml volume transfer by 10 time each then in every measurement your error from standard deviation ±0.006 ml. That is for ten times it becomes ~~10 × 0.006 = 0.06~~ larger. ml. Thus the uncertainity propagates with replication of experiment.

Uncertainity when adding or substracting

When measurements are added or substracted, the absolute uncertainity in the result is the square root of sum of the square of absolute uncertainity of individual measurement.

Suppose $R = A + B + C$, then

absolute uncertainity of R

$$S_R = \sqrt{S_A^2 + S_B^2 + S_C^2}$$

eg for two times measurement of pipet -

$$S_R = \sqrt{(0.006)^2 + (0.006)^2} = 0.0085.$$

Similarly for equn $R = A \times B \times C$

$$\frac{S_R}{R} = \sqrt{\left(\frac{S_A}{A}\right)^2 + \left(\frac{S_B}{B}\right)^2 + \left(\frac{S_C}{C}\right)^2}$$

~~The Precision of Measurement and Result :-~~

Rounding Off :-

If the digit following the last significant figure is greater than 5, the number is rounded up to the next higher digit. If it is less than 5, the number is rounded to the present value of the last significant figure :

$$9.47 = 9.5$$
$$9.43 = 9.4$$

If the last digit is a 5, the number is rounded off to the nearest even digit.

$8.65 = 8.6$ (6 is even)

$8.75 = 8.8$ (7 is odd so it rounded to 8)

The Confidence Limit

Calculation of the standard deviation for a set of data provides an indication of precision inherent in a particular procedure, or analysis. But unless there is a large amount of data, it

does not by itself give any information about how close the experimentally determined mean $\bar{x}$ might be the true mean value μ (mean of infinite number of measurement). Statistically, though, allow us to estimate the range within which the true value might fall, within given probability, defined by the experimental mean and standard deviation.

The range is called the confidence interval and the limits of this range is called the confidence limit.

Likely, the true value falls within the range is called the probability or confidence level.

$$\text{confidence limit} = \bar{x} \pm \frac{ts}{\sqrt{N}}$$

Where t is a statistical factor that depends on the number of degree of freedom and the confidence level desired. The number of degree of freedom is one less than the number of measurement.

s = standard deviation N = no. of measurement.

Eg A soda ash sample is analyzed in the analytical chemistry lab by titration with standard hydrochloride acid. The analysis is in triplicate with 95% confidence level and the results: 93.50, 93.58 and 93.43 wt% Na_2O_3. What range the true value lies?

Ans: Mean $\bar{x} = \frac{93.50 + 93.58 + 93.43}{3} = 93.50$ wt%

Standard deviation $s = 0.075$ wt% (Calculate!)

At 95% confidence level at degree of freedom $\nu = N - 1 = 3 - 1 = 2$ is $t = 4.303$ (to be given by chart)

$$\text{Confidence limit} = \bar{x} \pm \frac{ts}{\sqrt{N}}$$

$$= 93.50 \pm \frac{4.303 \times 0.075}{\sqrt{3}}$$

$$= 93.50 \pm 0.19 \text{ wt\%}$$

So at 95% confidence level the actual value lie in the range 93.31 to 93.69 wt%.

Significance Test:

In developing new analytical method, it is often desirable to compare the result of that method with those of an accepted (perhaps standard) method. How can one can tell if there is a significant difference b/w the new method and the accepted one? The F-test evaluates difference b/w the spread of result, while t-test looks at differences b/w means.

The F-test:-

F-test is defined as the ratio of variance of two method. Variance is the square of standard deviation.

$$F = \frac{s_1^2}{s_2^2}$$

eg You are developing a new colometric method for estimation of glucose in blood. Folin-Wu is a standard method of it. Then you have to prove that your method is better than the standard method.

Your method (mg/dL)	Folin-Wu method (mg/dL)
127	130
125	128
123	131
130	129
131	127
126	125
129	

mean: $\bar{x}_1 = 127$ $\bar{x}_2 = 128$

$$s_1^2 = \frac{\sum (x_{i_1} - \bar{x}_1)^2}{N_1 - 1} = \frac{50}{7-1} = 8.3$$

$$s_2^2 = \frac{\sum (x_{i_2} - \bar{x}_2)^2}{N_2 - 1} = \frac{24}{6-1} = 4.8$$

$$\therefore F = \frac{8.3}{4.8} = 1.73$$

i.e. $F > 1$.

But its limit of F in b/w degree of freedom $\nu_1 = 6$ and $\nu = 5$ is 4.95 (there is a chart in literature). So the value of F (1.73) is much less than literature value

So we can conclude there is no significance difference in precision of this two method.

t-Test :

The t-test is used to determine if two sets of measurement are statistically different. If $t_{calc} > t_{table}$ then the two data sets are significantly different at the chosen confidence level.

Pooled standard deviation:- It is used to ~~determin for~~ performing t-test.

$$S_p = \sqrt{\frac{\sum(x_{i_1} - \bar{x}_1)^2 + \sum(x_{i_2} - \bar{x})^2 + \cdots + \sum(x_{ik} - \bar{x}_k)^2}{N-k}}$$

Where $\bar{x}_1, \bar{x}_2, \cdots \bar{x}_k$ are means of each of k set and $x_{i_1}, x_{i_2}, x_{i_k}$ are the individual values of each set. N is the total number of measurement and k is the number of set of data.

(i) t-Test with a Reference value is known:-

The true value of a result is μ.

Then $\mu = \bar{x} \pm \frac{ts}{\sqrt{N}}$ as like confidence limit

$\therefore t_{calc} = (\bar{x} - \mu)\frac{\sqrt{N}}{s}$

(ii) Comparison of Means of two method of same set of data

where $\frac{\sqrt{N}}{s}$ is replaced by

$$\frac{\sqrt{\frac{N_1 N_2}{N_1 + N_2}}}{S_p}$$

S_p = Pooled standard deviation.

$$\therefore t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{s_p} \sqrt{\frac{N_1 N_2}{N_1 + N_2}}$$

(iii) Paired t-test :- It is for comparison of series of different samples by two different method.

Here
$$t = \frac{|\bar{D}|}{s_d} \sqrt{N}$$

$$s_d = \sqrt{\frac{\Sigma (D_i - \bar{D})^2}{N-1}}$$

D_i = individual difference b/w two different method.

$\bar{D}$ = mean of all individual differences.

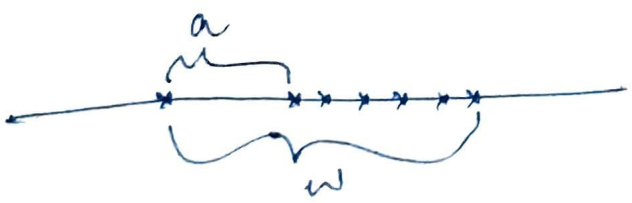
Rejection of a Result : The Q-test

When a series of replicate analyses is performed, it is not uncommon that one of the results will appear to differ markedly from the others. A decision will have to be made whether to reject the result or to retain it. Unfortunately, there are no ~~significant~~ uniform criteria that can be used to decide if a suspect result can be ascribed to accidental error rather chance variation. It is tempting to delete extreme values from data set because they alter the calculated statistics in an unfavourable way, that is increase the standard deviation and variance and they may alter the reported mean.

Q-test is the test, most statistically correct for a fairly small number of observation to ~~dis~~ reject the absurd result.

The ratio 'Q' is calculated by arranging the data in increasing or decreasing order of numbers, then difference between suspected result with nearest neighbour (a) by the rang (w).

$$\therefore Q = \frac{a}{w}$$



If the value of Q is greater than the rejection quotient in table of certain confidence level. Then the data can be rejected.

Eg The following set of chloride determination on separate aliquots of a pooled serum were reported 103, 106, 107 and 114 meq/L. One value appears suspect. Determine if it can be ascribed to accidental error at 95% confidence level.

⇒ The suspect result 114 meq/L. The difference between nearest neighbour 107 meq/L is (114-107) = 7 meq/L. The range (w) = (114 − 103) = 11 meq/L.

So $$Q = \frac{a}{w} = \frac{7}{11} = 0.64.$$

Now in the chart of 95% confidence level the value is 0.829.

$$\therefore Q_{calc} < Q_{table}$$

So the suspected value can be retained. Not a accidental error but indeterminate error.