COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH
UNIVERSITY GRANTS COMMISSION

CHEMICAL SCIENCES CODE:01

2.14. Data analysis

At a Glance

Accuracy, Precision, Relative error, absolute error, significant figure, median, mean deviation.

teachinns Text with Technology

2.14. Data analysis

2.14.1. Error: When an analyst obtains a result as close to the true hello as possible by the correct application of the analytical procedure employed, he is always interested in knowing the cause and management of errors in measurement which is always present. When a series of parallel measurements are carried out on a perfectly homogeneous sample, the results of 10^{th} may mutualism differ they are said to be inaccurate.

2.14.2. Classification of errors: Errors are mainly two types.

- Systematic or determinate error.
- Random or indeterminate errors.

2.14.3. Systematic error: These types of error can be avoided and their magnitude can be determined. The most important of them are as follows:

Operational and personal errors:

- Only individual analyst is responsible for this type of errors.
- These types of error are not connected with the method or procedure.
- This is usually physical in nature and occurs when sound analytical technique is not followed.
- Example: Over washing or under washing the precipitation, ignition of precipitate at improper temperature, insufficient cooling of crucible before weighing, mechanical loss of materials in various type of analysis and use of reagents containing harmful impurities.
- Personal errors arise due to the inability of a person to make certain observations accurately.

Instrumental and reagent errors:

- Due to the faulty construction of chemical balances, improperly calibrated weights, burette and other glass wares and instruments, the attack of reagents upon porcelain, glass wares etc. and use of the reagents containing impurities, this type of error occurs.

Errors of method:

- Due to incorrect sampling and due to incomplete of reaction this type of errors arises.
- In titrimetric analysis errors arisedue to failure of reactions to proceed to completion, occurrence of induced or side reactions and a difference between the observed end point and stoichiometric equivalent points of a reaction.
- In gravimetric analysis errors arise due to co-precipitation, post precipitation, appreciable solubility of precipitate and decomposition or vitalization of precipitate during English and precipitation of undesired substances.

Additive and proportional errors:

- The absolute value of this type error is independent of the amount of the substance being determined.
  Example: loss in weight of crucible in which precipitate is ignited, errors in weight and solubility of precipitate.
- Absolute value of proportional error depends upon the amount of substances present.
  Example: presence of impurities in a standard substance results in wrong value of normality of the standard solution.

teachinns Text with Technology

2.14.4. Random or indeterminate errors:

- These errors are usually small and irregular.
- These errors are due to the limitations in the equipment used and the limited skill of the observer.

2.14.5. Accuracy: Accuracy is the agreement of a particular value to the true value of the result.

Example: if the true value of a result is 3.00 g and a student takes two measurements and reports the result as 3.01 g and 2.99 g, these values are accurate.

it is expressed in terms of either absolute or relative error.

2.14.6. Absolute error: It is the difference between the measured value and true value.

Absolute error, $E = x_i - x_t$

Where, x_i = measured value and x_t = true value of the quantity.

- A low value produces a negative error.
- A high value produces a positive error.

2.14.7. Relative error:

The percent relative error,

$E_R = (x_i - x_t)/x_t \times 100$

It is also expressed in parts per thousand (ppt).

2.14.8. Precision: It refers to the closeness of various measurements for the same quantity. It describes the reproducibility of measurements. These terms are widely used to describe the position of a set of replicate data. The terms are standard deviation, variance and coefficient of variation. All the terms are a function of the deviation from the mean which is defined as,

Deviation from the mean = $d_i = | x_i - \bar{x} |$

Where, x_i = the individual value.

x **bar** = the mean value of replicate measurements.

- Precision always accompanies accuracy but high degree of precision does not imply accuracy.

2.14.9. Minimization of errors: Systematic errors can be minimized by the following methods:

- Calibration operators and application of collections.
- Running a blank determination.
- Running a control determination using primary standard.
- Use of independent method of analysis.
- Standard addition.
- Running parallel determinations.
- Internal standards.
- Amplification method.
- Isotope dilution.

2.14.10. Significant figures: The total number of digits in a number is called the number of significant figures (SF). The following rules are followed in counting the significant figure:

- All digits are significant except zero at the beginning of the number.
  Example:
  101 → SF = 3
  0.101 → SF = 3
  0.0101 → SF = 3
- The zeros to the right of the decimal point are significant.
  Example:
  101.0 → SF = 4
  101.00→ SF = 5
- The above rules prepurpose bad the numbers are expressed in scientific notation. In this term every number is written as $N \times 10^n$.
  Where, N = a number with a single non-zero digit to the left of the decimal point.
  n = an integer
  Example:
  We can write 2000 in scientific notation as
  2×10^3 → SF = 1
  2.0×10^3 → SF = 2
  2.00×10^3 → SF = 3
- Zero at the end of a number and before the decimal point may or may not be significant.

2.14.11. Rounding off the numerical results: To 3 significant figures, we should express 15.453 as 15.5 and 14755 as 1.48×10^4. It is called "Rounding off" the results.

- If the first digit we remove is less than 5, round down by dropping it and all following disease. Thus 5.663507 becomes 5.66 when rounded off to three significant figures.
- If the first digit remove is 6 or greater than 6 ground off by adding 1 to the digit on the left.
  5.663507 = 5.7 → for 2 SF.

- If first digit we remove is 5 and there are more non zero disits following rounded up.
  5.663507 = 5.664 → for 4 SF
- If the digit we remove is 5 and there is no digit after, then add 1 to the preceding digit if it is odd, otherwise write as such if it is even.
  4.7475(odd digit before 5) = 4.748 → for 4 SF
  4.7465(even digit before 5) = 4.746 → for 4 SF

2.14.12. Mean: It is the numerical value obtained by dividing the sum of a set of measurements by the total number of measurements. It is also known as arithmetic mean. It is given by,

$$\bar{x} = \frac{x_1 + x_1 + x_1 + \cdots + x_{(n-1)} + x_n}{n}$$

n is the number of observations.

x_1, x_2, ….. are the magnitude of the measurements.

2.14.13. Median: It is defined as the middle most or central value of the variate when the observations are arranged in ascending or descending order of their magnitudes.

- If n(the number of observations) is odd then,
  M_e = (n+1/2) th term
- If n is even then,
  M_e = [(n/2)th term + (n/2+1) th term] / 2

The mean and median may or may not be the same. If data is irregular near median then it would not be a true representative.

teachinns Text with Technology

2.14.14. Mean deviation: Mean deviation or average deviation is the arithmetic mean of deviation of all the values taken from a statistical average (mean, median or mode) of the series.

Mean deviation,

$$MD = \frac{1}{n}\sum_{i=1}^{n}|x_i - M|$$

Where,

M = any average of the series.

x_i = i^{th} observation of the series

n = number of observations

2.14.15. Relative mean deviation = Mean deviation/ Mean

2.14.16. Standard deviation: It is the square root of the arithmetic mean of the squared deviations of various values from their arithmetic mean.

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

Where, n = number of observation

$\bar{x}$ = arithmetic mean

x_i = i^{th} observation of the series.

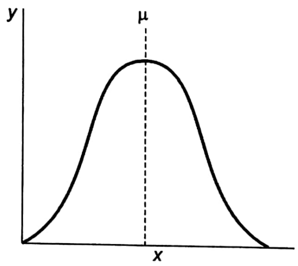
2.14.17. Variance: The square of the standard deviation is called variance.

2.14.18. Coefficient of the standard deviation = $\frac{\sigma}{\bar{x}}$

2.14.19. Coefficient of variation (CV) = $\frac{\sigma}{\bar{x}} \times 100$

2.14.20. Normal or Gaussian distribution: The mathematical model that best satisfies distribution of random errors is called Normal or Gaussian distribution. The bell shaped curve satisfied the equation,

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



Normal distribution

Where, σ = standard deviation

μ = mean of total population

s and x **bar** are used for samples of populations.

2.14.21. Pearson's correlation coefficient: Correlation indicates the degree of relationship between two variables; the movements in one variable are accompanied by corresponding movement in the other variables. In fact correlation is an analysis of correlation between two or more variables.

$$r = \frac{n\sum x_1y_1 - \sum x_1y_1}{\sqrt{[n\sum x_1^2 - (\sum x_1)^2][n\sum y_1^2 - (\sum y_1)^2]}}$$

Where, x1 and y1 are two variables and n is the number of data points. The value of r must lie between +1 and -1.

2.14.22. Probable error of correlation coefficient = $1+r^2/\sqrt{n}$

2.14.23. Linear regression: The equation of straight line is

y = ax + b

Where, y is dependent variable and x is the independent variable. To obtain the regression line y on x , the slope of the line (a)

$$= \frac{n\sum x_1y_1 - \sum x_1y_1}{n\sum x_1^2 - (\sum x_1)^2}$$

And the intercept on the y axis(b) = $\bar{y} = a\bar{x}$, where, $\bar{x}$ and $\bar{y}$ are mean of all values of x_1 and y_1 respectively.

teachinns Text with Technology

2.14.24. Coefficient of determination: It is a measure that is used to determine with certainty with which predictions with the line of best fit can be made.

It is obtained by squaring the value of correlation coefficient (r) and it can be represented by symbol r^2.

- Values of r^2 close to 1 suggest that the model explain most of the variations in dependent variable.
- If the values of r^2 are close to zero, it implies that the model explains little of the variation in the dependent variable.

teachinns Text with Technology

NET JUNE 2016

Q. In a titration, the percentage uncertainties in the measured aliquot volume and the measured titre are ±x and ±y respectively. The percentage error in the calculated concentration of aliquot is

(a) $x+y$ (b) xy (c) $(xy)^{1/2}$ (d) $(x^2+y^2)^{1/2}$

Ans. In a titration the % uncertainties in the measured aliquot volume and the measured titre volume are ±x and ±y respectively. The percentage error in concentration of aliquot is $(x^2+y^2)^{1/2}$.

teachinns Text with Technology