SRI AKILANDESWARI WOMEN'S COLLEGE, WANDIWASH

DATA ANALYSIS

Class : II UG Chemistry

Dr. S. RAJESH
Assistant Professor
Department of Chemistry

SWAMY ABEDHANADHA EDUCATIONAL TRUST, WANDIWASH

INTRODUCTION

- The role of analytical chemist is to obtain results from experiments and know how to explain data significantly.
- Statistics are necessary to understand the significance of the data that are collected and therefore to set limitations on each step of the analysis.

Types of errors

- 1. Determinate errors (Systematic)

- 2. Indeterminate errors (Random)

Characteristics of determinate errors:

- 1. Cause of error is known.
- 2. Consistency, that is the values are almost the same.
- 3.Difficult to correct for with statistics
- 4. Due to poor technique, faulty calibration, poor experimental design
- 5.Controls the **accuracy** of the method
- 6. Can be avoided

TYPES OF DETERMINATE ERRORS

Types of determinate or systematic errors:

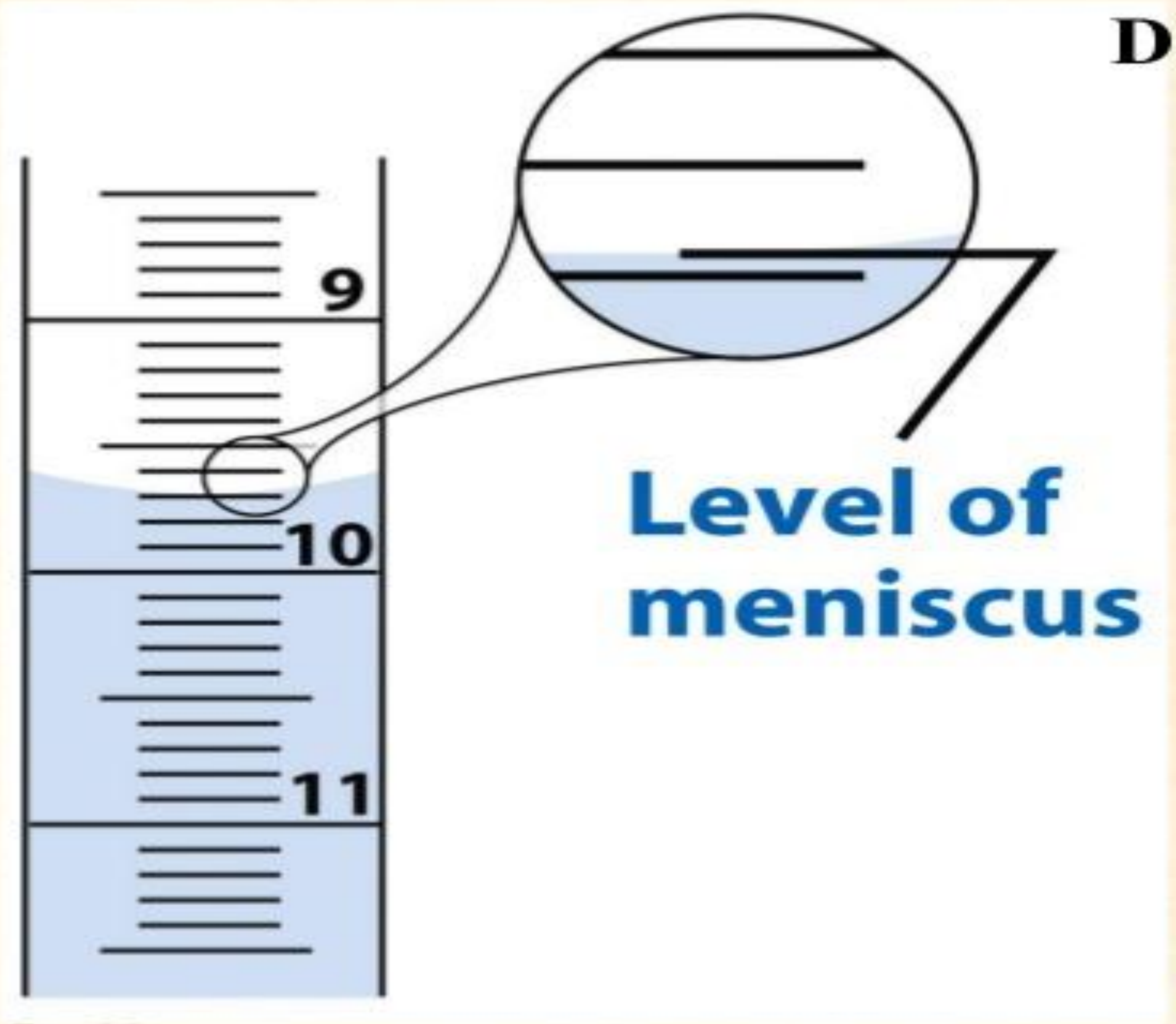
- 1.Operative errors
- 2.Instrumental errors
- 3.Method errors
- 4.Error of the reagent.

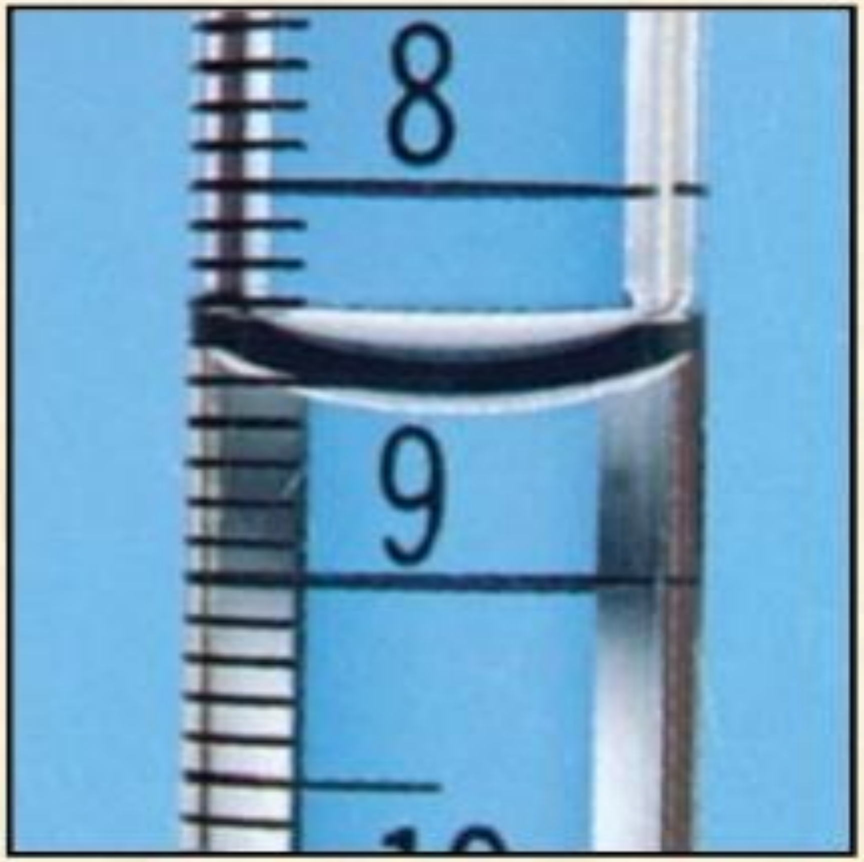
(i) Operative errors

It is also known as observation error and is caused by carelessness, clumsiness or not using the right techniques by the operators.

For example, recording wrong burette reading as 29.38 mL, whereas the correct reading is 29.35 mL. To avoid this error is by reading the volume correctly.

Determinate Error





(ii) Instrumental errors

- The faulty equipments, uncalibrated weight and glasswares used may cause instrumental errors. Example, using instruments that are not calibrated and this can be corrected by calibrating the instruments before using.

(iii) Method errors

- Method errors are caused by the nature of the methods used.
- This error cannot be omitted by running the experiments several times.
- It can be recognized and corrected by calculations or by changing to a different methods or techniques.
- This type of error is present in volumetric analysis that is caused by reagent volume.
- An excess of volume used as compared to theory will result in the change of colours. Example, mistakes made in determining end point caused by coprecipitation.

- (iv) Errors of the reagent
- This error will occur if the reagents used are not pure. Correction is made by using pure reagents or by doing back-titration

INDETERMINATE OR RANDOM ERRORS

- Also known as accidental or random error.
- This represent experimental uncertainty that occurs in any measurement.
- This type does not have specific values and are unpredictable.

Charateristics of indeterminate errors:

- a) Cause of error is unknown.
- b) Spreads randomly around the middle value

(follow the normal distribution/Gaussian curve).

- c) Usually small.
- d) Have effects on precision of measurement
- e) cannot be avoided
- f) Easily treated with statistical methods

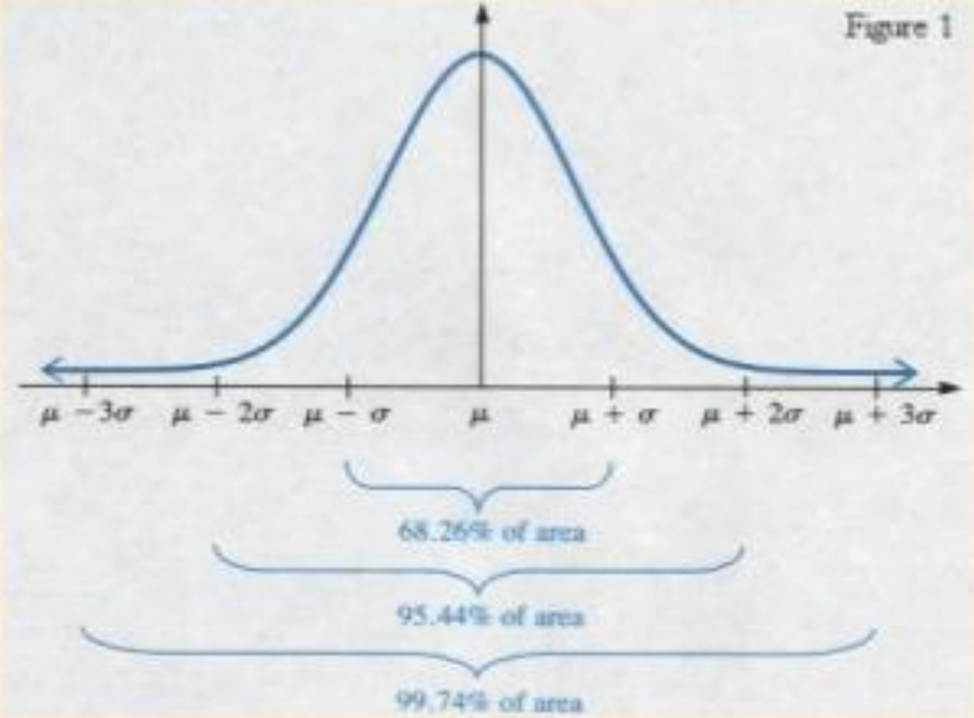
Example of indeterminate errors: change in humidity and temperature in the room that cannot be controlled.

Normal Distribution

Bell-shapes curve/normal curve

Mean (μ)= located in the center

The σ symbols represent the standard deviation of an infinite population of measurement, and this measure of precision defines the spread of the normal population distribution



Variation due to indeterminate error

- Assume the same object is weighed repeatedly without determinate error on the same balance.
- Examine the data on next page for replicate weighings of a 1990 penny.

Observed mass (g) for 1990

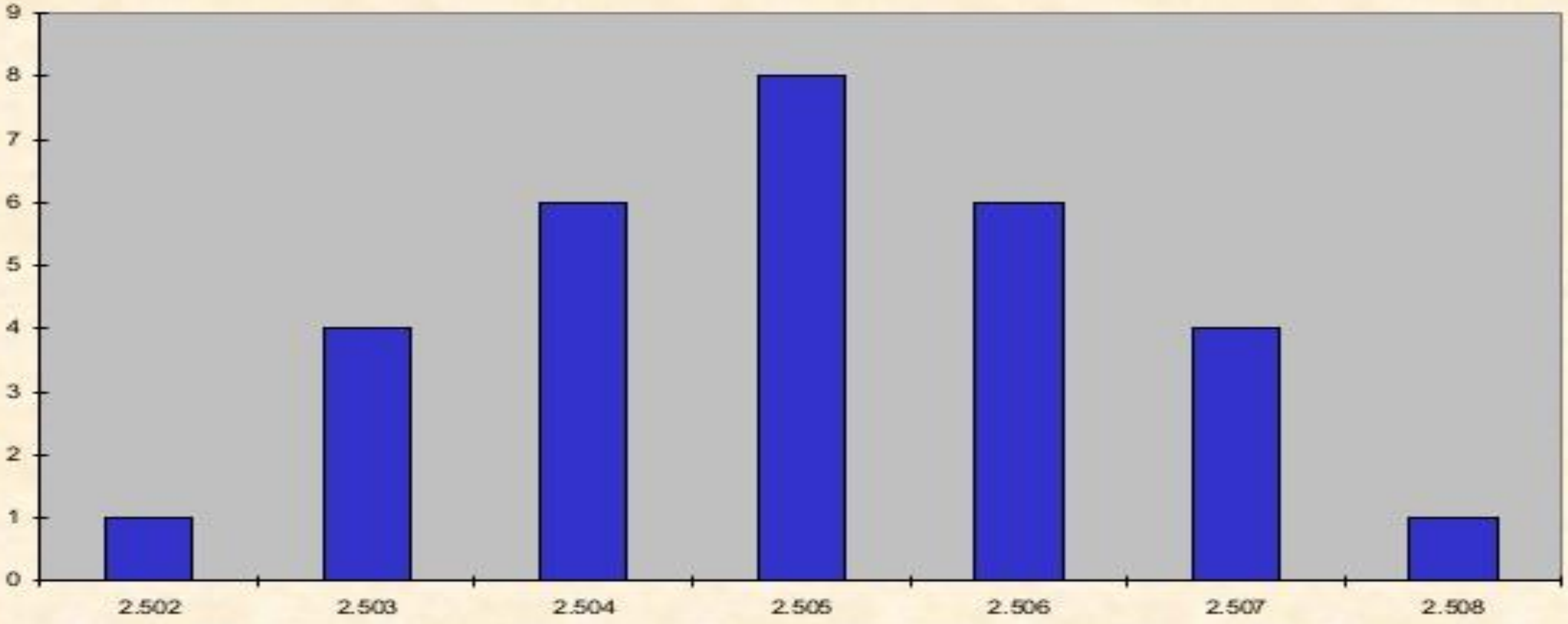
2.505	2.503	2.507	2.506	2.502
2.508	2.504	2.505	2.507	2.504
2.503	2.506	2.507	2.505	2.503
2.504	2.506	2.505	2.506	2.504
2.504	2.507	2.506	2.506	2.505
2.503	2.505	2.504	2.505	2.505

The values are not identical, and not all values occur with equal frequency.

Value	Frequency
2.502	1
2.503	4
2.504	6
2.505	8
2.506	6
2.507	4
2.508	1

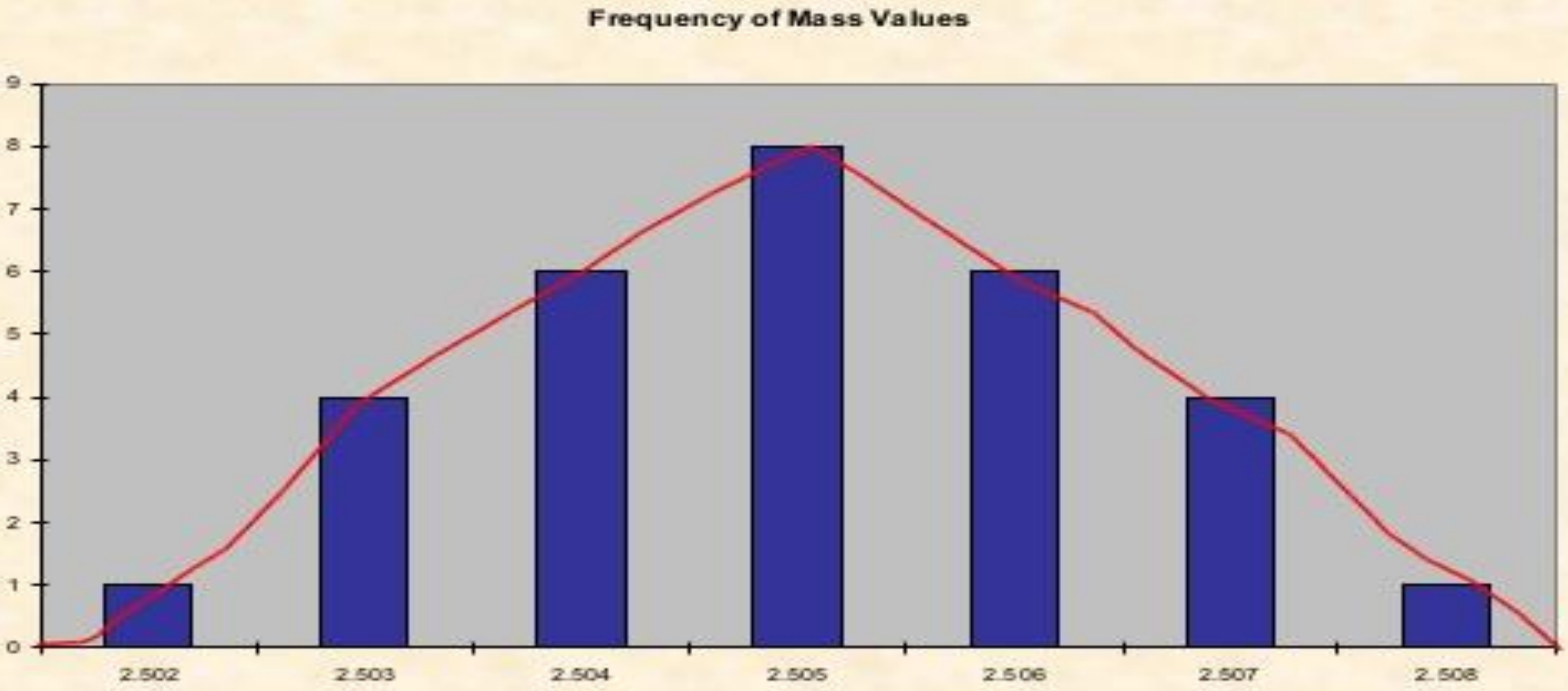
- The next slide is a histogram of the frequency data – notice the symmetrical arrangement of the values around one which occurs most frequently.
- The *mode* of a data set is the value which occurs most commonly or frequently.

Frequency of Mass Values

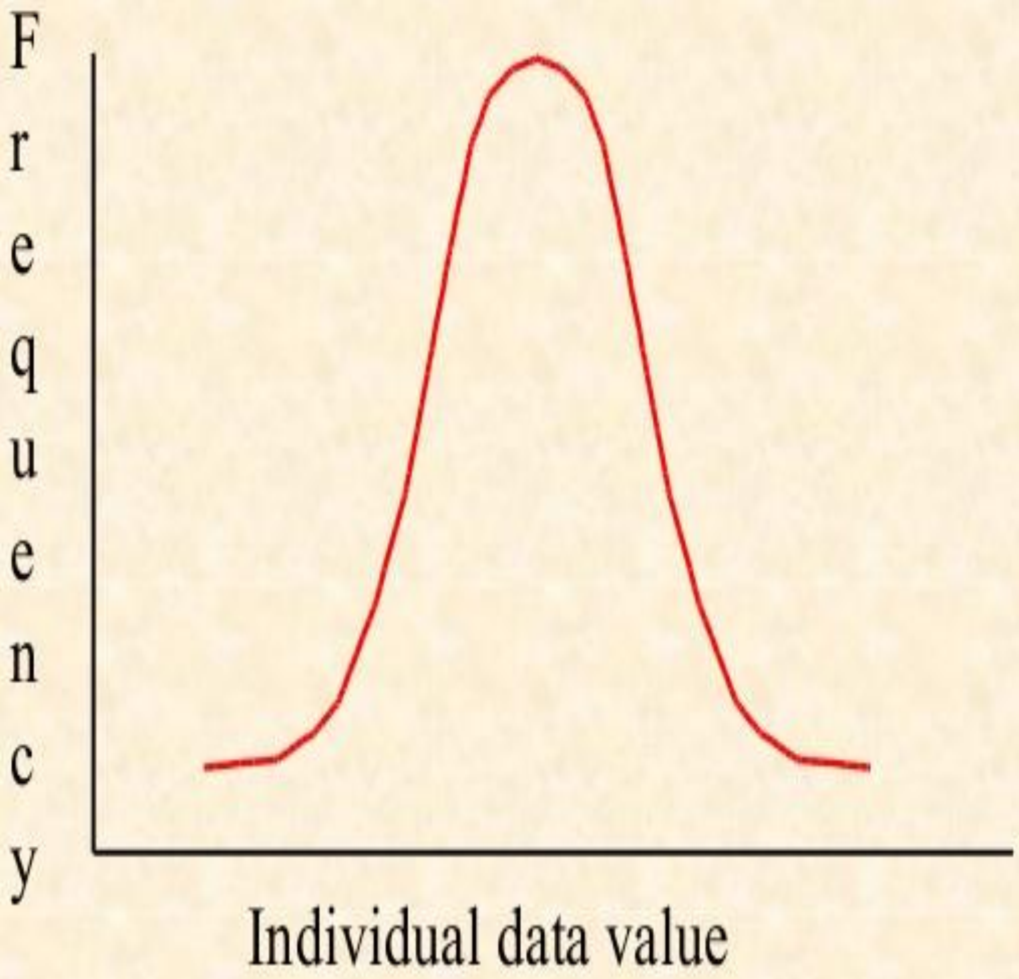


- The *mean* or *average* of a data set is the sum of all the values divided by the number of values in the set.
- The *median* is the value for which half the data is larger and half the data is smaller.

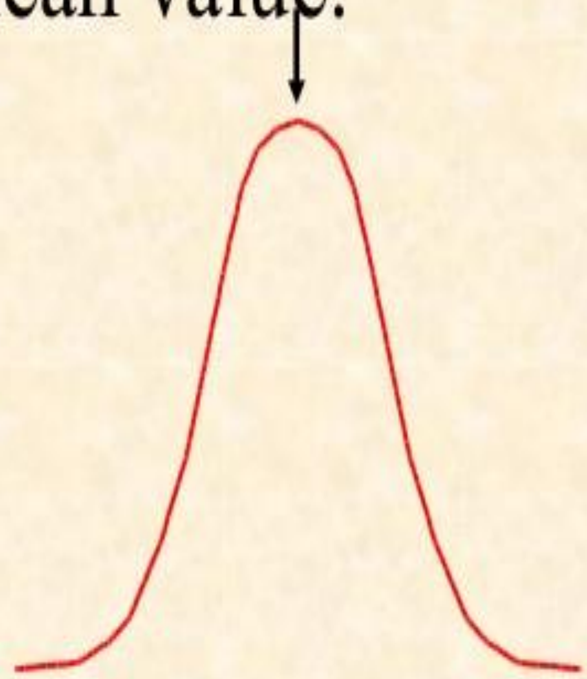
- A large enough data set of replicate values has the same number for mode, mean, and frequency.
- If the data set is very large it is called a population. When the data is grouped into smaller and smaller intervals, the histogram approaches a shape called a Gaussian distribution or “bell curve.”


Frequency of Mass Values
0
1
2
3
4
5
6
7
8
9
2.502
2.503
2.504
2.505
2.506
2.507
2.508

Or, closer to the ideal Gaussian shape:



In a population of data, the peak of the Gaussian distribution falls at the mean value.



Precision and Accuracy in Measurements

- **Precision** – how closely repeated measurements approach one another.
- **Accuracy** – closeness of measurement to "true" (accepted) value.

Ways of expressing Accuracy

- There are various ways and unit in which the accuracy of a measurement can be expressed:
- A) Absolute error
- B) Relative error

Absolute error

- **Recall:** Accuracy is expressed as absolute error or relative error.
- Difference between the true value and the measured value.
- Absolute errors: E = O – A
- O = observed error
- A = accepted value

- Example:If a 2.62g sample of material is analyzed to be 2.52g, the absolute error is -0.10g
- The positive and negative sign is assigned to show whether the errors are high or low.
- If the measured value is the average of several measurements, the error is called mean error

Relative error

- The absolute or mean error expressed as a percentage of the true value are called ***relative error.***
- Example:(-0.10/2.62) X 100% = -3.8%
- For ***relative accuracy*** expressed as measured value as apercentage of true value.
- ***Example: (2.52/2.62) X 100 = 96.2%***

- In very accurate work, we are usually dealing with relative errors of less than 1%.
- 1% error = 1 part in 100 = 10 part in 1000
- Part per thousand (ppt) is often used in expressing precision of measurement

Example

- The result of an analysis are 36.97g, compared with the accepted value of 37.06g. What is the relative error in part per thousands?

Solution:

Absolute error =36.97g – 37.06g = - 0.09g

Relative error = $\frac{-0.09}{37.06}$ X 1000%= -2.4

Ways of expressing precision

- **Recall:** Precision can be expressed as standard deviation, deviation from the mean, deviation from the median, and range or relative precision.

Example 1: The analysis of chloride ion on samples A, B and C gives the following result

Sample	%Cl-	Deviation from mean	Deviation from median
A	24.39	0.10	0.11
B	24.20	0.09	0.08
C	24.28	0.01	0.00
	$\bar{X}$ = 24.29	$\bar{d}$ = 0.07 (d bar)	0.06

Deviation from mean

- Deviation from mean is the difference between the values measured and the mean.
- For example, deviation from mean for sample A= 0.10%
- Q1: How to get X ?
- A1: $= \frac{24.39 + 24.20 + 24.28}{3}$

  $= 24.29$

Deviation from median

- Deviation from median is the difference between the values measured and the median.
- Example, deviation from the median for sample A = 0.11%.
- Now, identify median = 24.28
- Therefore, deviation from median for sample A = 24.39- 24.28 = 0.11%

- Range is the difference between the highest and the lowest values.
- For example, range = 24.39- 24.20 = 0.19 %.

- **Sample standard deviation**

For N (number of measurement) < 30

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N-1}} \quad \text{or} \quad \sqrt{\frac{\sum d^2}{N-1}}$$

For N > 30

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}} \quad \text{or} \quad \sqrt{\frac{\sum d^2}{N}}$$

THANK YOU