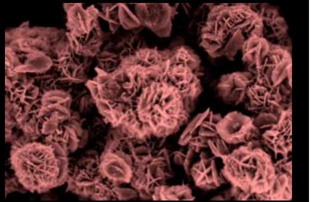




Chemical Engineering

UNIVERSITI TEKNOLOGI PETRONAS

Data Handling, Errors and Statistics in Analytical Chemistry

CEB 4032: ANALYTICAL CHEMISTRY
CFB3032: ANALYTICAL INSTRUMENTATION

TS. CHM. DR. MOHD DZUL HAKIM WIRZAL

Inspiring Potential · Generating Futures

Outline

- Analytical Objectives
- Qualitative Analysis
- Quantitative Analysis
- Analytical Methodology
- Type of equipment which includes Gravimetry, Spectrophotometry, Spectroscopy and Chromatography.
- Basic Tools and Operation
- Data Handling and Statistic
- Errors

Learning Outcomes:

At the end of this chapter:

(1) Objectives of analytical chemistry.

(2) Definition of analytical chemistry, qualitative anlaysis and quantitative analysis.

(3) Type of equipment and basic tools used in analytical chemistry and laboratory safety.

(4) Basic calculation, statistical method and errors in data handling for reliability and significant derived results.

Error

- Measurements always involve errors and uncertainties → due to the mistakes on the part of the experiment.

- Error → refer to the difference between a <u>measured value and the ‘true’ or ‘known’ value</u>.

- Often gives the estimated uncertainty in a measurement of experiment.

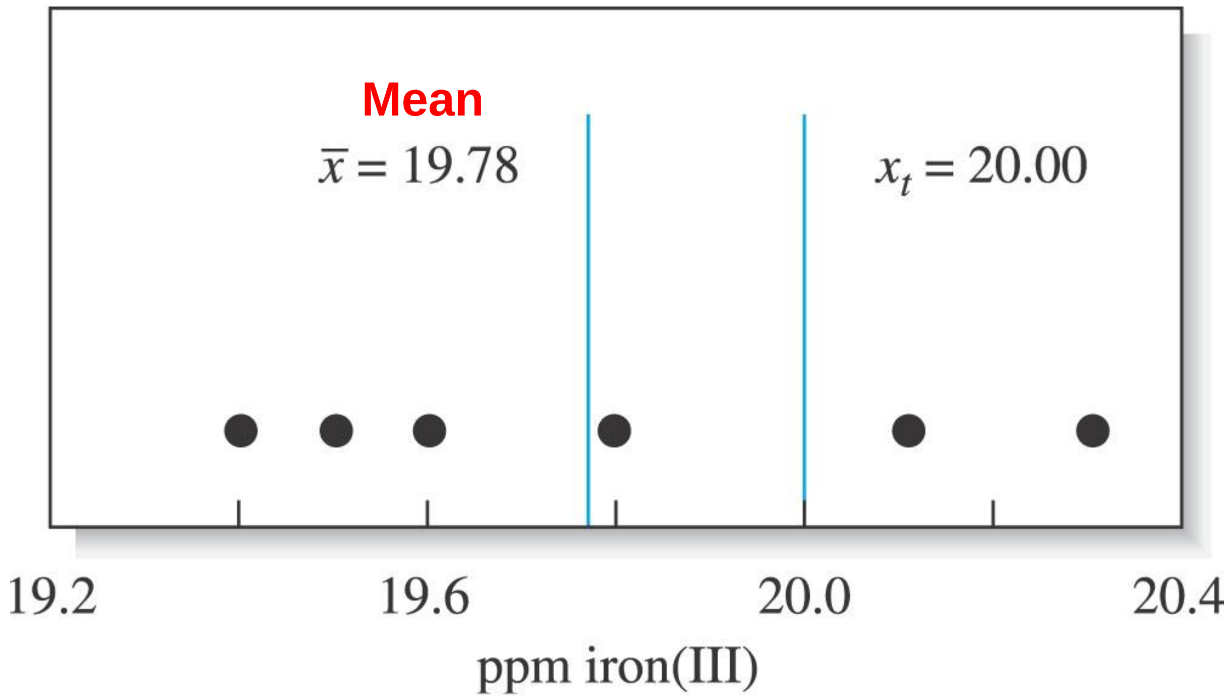
- Caused by faulty calibration or standardizations or random variations and uncertainties in results.

- It is impossible to perform a chemical analysis that is totally free of errors or uncertainties.

- Only hope to minimize errors and estimate their size with acceptable accuracy.

Example Error

Result six replicate determinations of iron in aqueous samples of a standard solution containing 20 ppm iron (III)



- The results range from a low of 19.4 ppm to a high of 20.3 ppm.

Important Terms

1) Replicates

2) Mean

3) Median

4) Precision

5) Accuracy

1) Replicates

→samples of about the same size that are carried through an analysis in exactly the same way.

Example:

One student measures Fe (III) concentration for six times. The results are listed below:

19.4,19.5,19.6,19.8,20.1,20.3 ppm

6 replicates = 6 measurements

2) The Mean, $\tilde{x}$

→ Or the average, obtained by dividing the sum of replicate measurements by the number of measurements in the set.

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Where x_i = individual value of x making up the set of N replicates measurements

3) **The Median**

→The middle result when replicate data are arranges according to increasing or decreasing value.

→For an odd number results, the median can be evaluated directly.

→For an even number, the mean of the middle pair is used.

Chemical Engineering
Inspiring Potential · Generating Futures

Example 1

Calculate the mean and median for the data shown below:

19.4,19.5,19.6,19.8,20.1,20.3 ppm

Answer

$$\text{Mean} = \frac{19.4+19.5+19.6+19.8+20.1+20.3}{6}$$

= 19.78 ppm = 19.8 ppm

6 replicates

An **even number** of replicates !!!

$$\text{Median} = \frac{19.6+19.8}{2} = 19.7 \text{ ppm}$$

Example 2

Calculate the mean and median for the data shown below:

19.4,19.5,19.6,19.8,20.1 ppm

Answer

$$\text{Mean} = \frac{19.4+19.5+19.6+19.8+20.1}{5}$$

$= 19.68$ ppm $= 19.7$ ppm

5 replicates

An **odd number** of replicates !!!

Median = 19.6 ppm

4) Precision

→ The reproducibility of measurements.

→ The closeness of results that have been obtained in exactly the same way

→ Three terms are widely used to describe the precision of a set of replicate data:

- standard deviation
- variance
- coefficient of variation

5) Accuracy

→Closeness of the measurement to the true or accepted value.

→expressed in terms of absolute or relative error.

Absolute Error: the difference between the measured value and the true value. Has a sign (-) or (+) and unit.

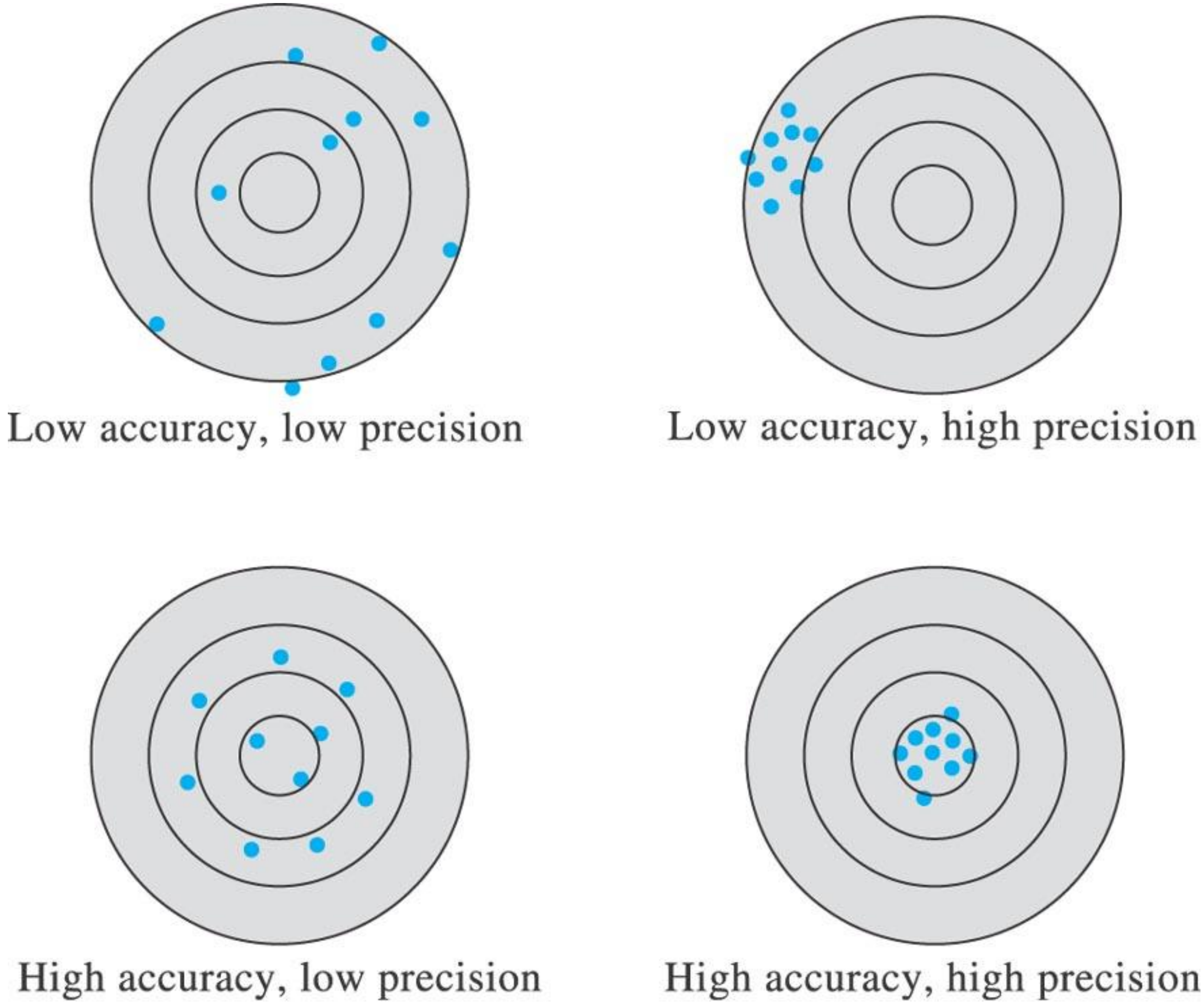
$$E = x_i - x_t$$

Where x_t is the true or accepted value.

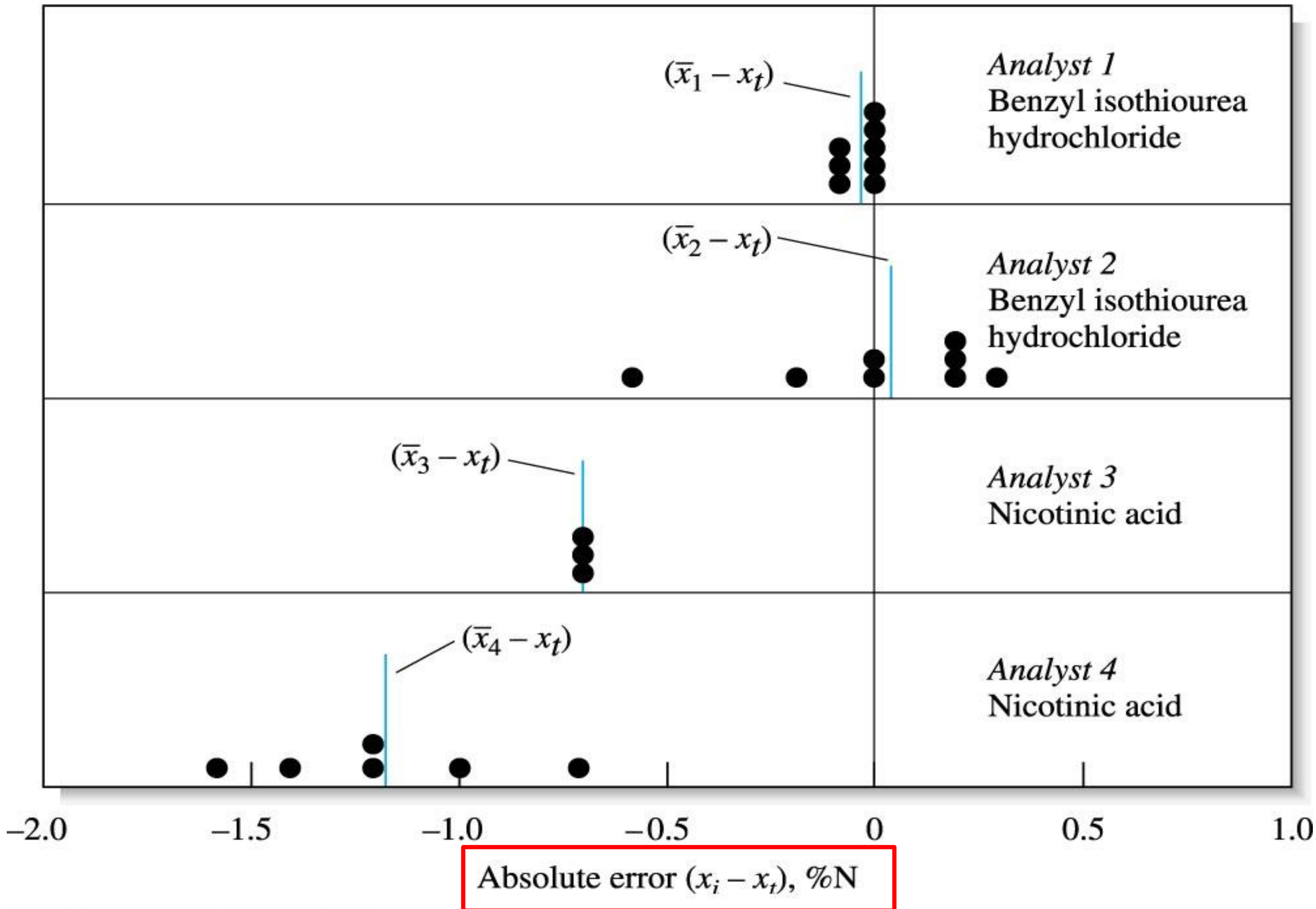
Relative Error: absolute error divided by true value. Expressed in %, ppt or ppm. Has sign but no unit.

$$E_r = \frac{x_i - x_t}{x_t} \text{ x } 100\%$$

Chemical Engineering
Inspiring Potential·Generating Futures



Error Analysis : Precision or Accuracy?



Example 3

Given that the true value of the data set below is 20 ppm. Determine the absolute error and relative error for each of the measurement.

19.4,19.5,19.6,19.8,20.1,20.3 ppm

Answer

E = 19.4 - 20.0 = -0.6 ppm	Er =(-0.6/20)x100% = - 3%
E = 19.5 - 20.0 = -0.5 ppm	Er =(-0.5/20) x100%= -2.5% ~ -3%
E = 19.6 - 20.0 = -0.4 ppm	Er =(-0.4/20) x100% = -2%
E = 19.8 - 20.0 = -0.2 ppm	Er =(-0.2/20) x100% = -1%
E = 20.1 - 20.0 = 0.1 ppm	Er=(0.1/20)x100% = 0.5%
E = 20.3 - 20.0 = 0.3 ppm	Er = (0.3/20)x100% =1.5% ~ 2%

$$E = x_i - x_t$$

$$E_r = \frac{x_i - x_t}{x_t} \times 100\%$$

Chemical Engineering

Inspiring Potential · Generating Futures

UNIVERSITI TEKNOLOGI PETRONAS

Example 4

A method of analysis yields weights for gold that are low by 0.3 mg. Calculate the relative error caused by this uncertainty if the weight of gold in the sample is 500 mg.

Answer

$$E = x_i - x_t$$ Absolute Error

$$E_r = \frac{x_i - x_t}{x_t} \text{ x } 100\%$$ Relative Error

Given:

E = -0.3 mg

Er = (-0.3 mg/500 mg) x100% = -0.06%

Chemical Engineering
Inspiring Potential · Generating Futures

Type of Errors in Experimental Data

1)Systematic (or determinate) error → causes the mean of a set to differ from the accepted value → Affect accuracy.

2)Gross error →occur only occasionally, are often large, and may cause a result to be either high or low. Product from human error → Lead to outliers.

3)Random (or indeterminate) error → causes data to be scattered more or less symmetrically around a mean value → Affect precision.

1)Systematic/Determinate Error

i) Instrument Errors

- caused by non ideal instrument behaviors, by faulty calibrations or by use under inappropriate conditions.

ii) Method Errors

-arise from nonideal chemical or physical behavior of analytical systems. For e.g. completeness and speed of reaction, interfering side reactions and sampling problems.

iii) Personal Errors

- results from the carelessness, inattention, or personal limitations of the experimental.

i) Instrument Error

- Variation in temperature.
- Contamination of the equipment.
- Power fluctuations.
- Component failure

All these can be corrected by calibration or proper instrumentation maintenance.

ii) Method Error

- Slow or incomplete reactions.
- Unstable species.
- Nonspecific reagents.
- Side reactions.

These can be corrected with proper method development.

iii) Personal Error

- Misreading of data.
- Improper calibration
- Poor technique/sample preparation.
- Personal bias.
- Improper calculation of results.

These are mistakes that can be minimized or eliminated with proper training and experience.

2) Gross Error

- Caused an experimental value to be discarded.
- Lead to outlier's and require statistical techniques to be rejected.
- Examples: "overrun end point" in titration, instrument breakdown, loss of crucial sample, and discovery that a "pure" reagent was actually contaminated.

We do not use the data obtained when gross error has occurred during collection.

3) Random/Indeterminate Error

- Caused by the many uncontrollable variables that are an inevitable part of every analysis.
- Most contributors to random error can not be identified.
- Random errors give rise to a normal or Gaussian curve.
- Gaussian curve→ a curve that shows the symmetrical distribution of data around the mean of an infinite set of data.
- Results can be evaluated using statistics → assumes a normal distribution.

The Statistical Treatment of Random Error

- Statistical analysis reveals only information that is already present in a data set.

- No new information is created by statistical treatment.

- To categorize and characterize data in different ways and to make objective and intelligent decisions about data quality and interpretation.

UNIVERSITI TEKNOLOGI PETRONAS

Chemical Engineering

Inspiring Potential·Generating Futures

Samples and Populations

- Sample → subset of measurements selected from the population.
- Population → collection of all measurements of interest to the experimenter.

Example: Normal Distribution of a Set Data

Chemical Engineering
Inspiring Potential · Generating Futures

UNIVERSITI TEKNOLOGI PETRONAS

TABLE 6-2

	A	B	C	D	E	F	G	H
1	Replicate Data for the Calibration of a 10-mL Pipet*							
2	**Trial**	**Volume, mL**		**Trial**	**Volume, mL**		**Trial**	**Volume, mL**
3	1	9.988		18	9.975		35	9.976
4	2	9.973		19	9.980		36	9.990
5	3	9.986		20	9.994		37	9.988
6	4	9.980		21	9.992		38	9.971
7	5	9.975		22	9.984		39	9.986
8	6	9.982		23	9.981		40	9.978
9	7	9.986		24	9.987		41	9.986
10	8	9.982		25	9.978		42	9.982
11	9	9.981		26	9.983		43	9.977
12	10	9.990		27	9.982		44	9.977
13	11	9.980		28	9.991		45	9.986
14	12	9.989		29	9.981		46	9.978
15	13	9.978		30	9.969		47	9.983
16	14	9.971		31	9.985		48	9.980
17	15	9.982		32	9.977		49	9.984
18	16	9.983		33	9.976		50	9.979
19	17	9.988		34	9.983			
20	*Data listed in the order obtained							
21	Mean	9.982		Maximum	9.994			
22	Median	9.982		Minimum	9.969			
23	Std. Dev.	0.0056		Spread	0.025			

UNIVERSITI TEKNOLOGI PETRONAS

Chemical Engineering

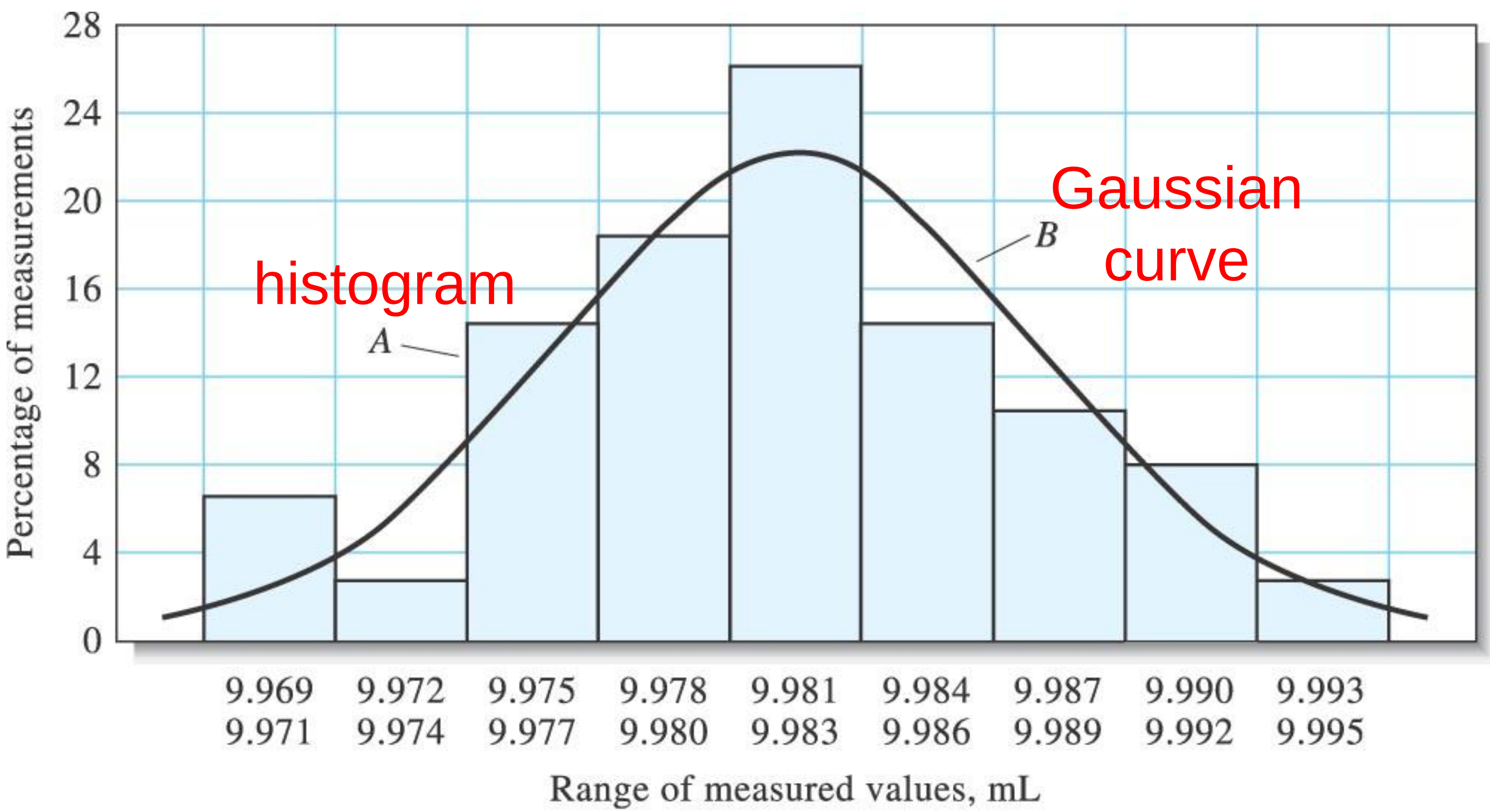
Inspiring Potential · Generating Futures

TABLE 6-3

Frequency Distribution of Data from Table 6-2

Volume Range, mL	Number in Range	% in Range
9.969–9.971	3	6
9.972–9.974	1	2
9.975–9.977	7	14
9.978–9.980	9	18
9.981–9.983	13	26
9.984–9.986	7	14
9.987–9.989	5	10
9.990–9.992	4	8
9.993–9.995	1	2
	Total = 50	Total = 100%

Normal or Gaussian Curve



Properties of Gaussian Curves

- Contains two parameters: population mean,μ and population standard deviation,σ .

- The sample mean and sample standard deviation are example statistics that estimate parameters μ and σ, respectively.

Sample Mean, $\tilde{x}$

→the arithmetic average of limited sample drawn from a population data.

→Sum of the measurement values divided by the number of measurements.

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Population Mean, μ

→True mean for the population.

$$\mu = \frac{\sum_{i-1}^{N} x_i}{N}$$

→Where *N* is now the total number measurements in the population.

Sample Standard Deviation, s

→ Measure of the precision of a sample data.

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

→ Where $(x_i - \bar{x})$ represent the deviation of xi form the mean.

Population Standard Deviation, σ

→ Measure of the precision of a population data.

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

→ Where N is the number of data points making up the population.

Example 5

The following results were obtained in the replicate determination of the lead content of a blood sample: 0.752,0.756,0.752,0.751 and 0.760 ppm Pb. Calculate the mean and the standard deviation of this set of data.

Chemical Engineering
Inspiring Potential·Generating Futures

UNIVERSITI TEKNOLOGI PETRONAS

Answer

The mean, $\bar{x}$:

(0.752 + 0.756 + 0.752 + 0.751 + 0.760)/5 = 0.7542

Sample	x_i	$(x_i - \bar{x})^2$
1	0.752	$(0.752-0.7542)^2 = 0.00000484$
2	0.756	$(0.756-0.7542)^2 = 0.00000324$
3	0.752	$(0.752-0.7542)^2 = 0.00000484$
4	0.751	$(0.751-0.7542)^2 = 0.00001024$
5	0.760	$(0.760-0.7542)^2 = 0.00003364$
	$\sum_{i=1}^{N}(x_i - \bar{x})^2$	(0.00000484+0.00000324 +0.00000484+0.00001024 +0.00003364)=0.0000568

$$s=\sqrt{\frac{\sum_{i-1}^{N}\left(x_i-\bar{x}\right)^2}{N-1}}=\sqrt{\frac{0.000408}{5-1}}$$

= 0.0038 ppm

Chemical Engineering

Inspiring Potential · Generating Futures

UNIVERSITI TEKNOLOGI PETRONAS

Variance and Other Measurement

Variance S^2

→square of the standard deviation.

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

Relative Standard Deviation (RSD)

→dividing the standard deviation by the mean value of the set data.

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

RSD in ppt

$$RSD\,(ppt) = \frac{s}{\bar{x}} \times 1000\ ppt$$

Coefficient of Variation (CV)

→dividing the standard deviation by the mean value of the set data.

$$CV = \frac{s}{\bar{x}} \times 100\%$$

Spread or Range (W)

-Used to describe the precision of a set of replicate results.

-The difference between the largest value in the set and the smallest.

Standard Error of a Mean or Standard Deviation of a Mean

-Inversely proportional to the square root of the number of data, N

$$s_m = \frac{s}{\sqrt{N}}$$

Example 6

The following results were obtained in the replicate determination of the lead content of a blood sample: 0.752,0.756,0.752,0.751 and 0.760 ppm Pb. Calculate (a) the variance, (b) the relative standard deviation in ppt, (c) the coefficient of variation, (d) the spread and the standard error of a mean.

Example 7

A batch of nuclear fuel pellets was weighed to determine if they fell within control guidelines. The weights were: 127.2, 128.4, 127.1, 129.0 and 128.1 g. Calculate

(a) The mean

(b) The median

(c) The range

Pooled Standard Deviation

UNIVERSITI TEKNOLOGI PETRONAS

- Combine standard deviation from different experiments to obtain a reliable estimate of the precision of a method.

$$s_{pooled} = \sqrt{\frac{\sum_{i=1}^{N_1}(x_i - \overline{x_1})^2 + \sum_{j=1}^{N_2}(x_j - \overline{x_2})^2 + \sum_{k=1}^{N_3}(x_k - \overline{x_3})^2 + \ldots}{N_1 + N_2 + N_3 + \ldots - N_t}}$$

Chemical Engineering
Inspiring Potential · Generating Futures

Example 8

The mercury in samples of seven fish taken from Chesapeake Bay was determined by a method based upon the absorption of radiation by gaseous elemental mercury. Calculate a pooled estimate of the standard deviation for the method, based upon the first three columns of data:

Specimen	Number of Samples Measured	Hg Content, ppm	Mean, ppm Hg	Sum of Squares of Deviations from Mean
1	3	1.80, 1.58, 1.64	1.673	0.0258
2	4	0.96, 0.98, 1.02, 1.10	1.015	0.0115
3	2	3.13, 3.35	3.240	0.0242
4	6	2.06, 1.93, 2.12, 2.16, 1.89, 1.95	2.018	0.0611
5	4	0.57, 0.58, 0.64, 0.49	0.570	0.0114
6	5	2.35, 2.44, 2.70, 2.48, 2.44	2.482	0.0685
7	4	1.11, 1.15, 1.22, 1.04	1.130	0.0170

Number of measurements = 28

Sum of squares = 0.2196

Chemical Engineering

Inspiring Potential · Generating Futures

UNIVERSITI

Answer

The values in the last two columns for specimen 1 were computed as follows:

x_i	$\lvert(x_i - \bar{x})\rvert$	$(x_i - \bar{x})^2$
1.80	0.127	0.0161
1.58	0.093	0.0086
1.64	0.033	0.0011
5.02	Sum of squares =	0.0258

$$\bar{x} = \frac{5.02}{3} = 1.673$$

Repeat for each set of data

The other data in columns 4 and 5 were obtained similarly. Then

$$s_{\text{pooled}} = \sqrt{\frac{0.0258 + 0.0115 + 0.0242 + 0.0611 + 0.0114 + 0.0685 + 0.0170}{28 - 7}}$$

$$= 0.10 \text{ ppm Hg}$$

Chemical Engineering
Inspiring Potential · Generating Futures

UNIVERSITI TEKNOLOGI PETRONAS

Summary of Equations

1) Mean
2) Median
3) Absolute Error
4) Relative Error
5) Sample Mean
6) Population Mean
7) Sample Standard Deviation

8) Population Standard Deviation
9) Variance
10) Relative Standard Deviation
11) Coefficient of Variation
12) Spread or Range
13) Standard Error of a Mean /Standard Deviation of a Mean
14) Pooled Standard Deviation

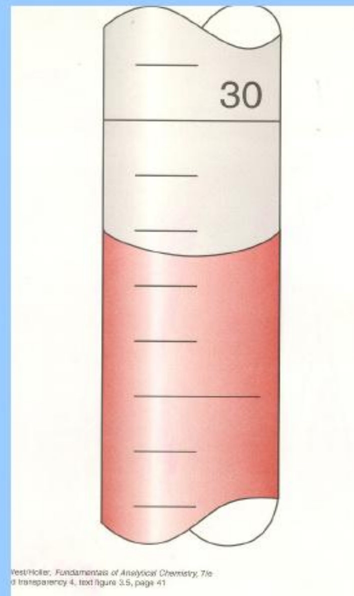
Significant Figures

- The number of digits reported in a measurement reflect the accuracy of the measurement and the precision of the measuring device.

- Significant figures in a number are all of the certain digits plus the first uncertain digit.

- In any calculation, the results are reported to the fewest significant figures (for multiplication and division) or fewest decimal places (addition and subtraction).

Significant Figures



30.2? mL

30.24 mL

30.25 mL

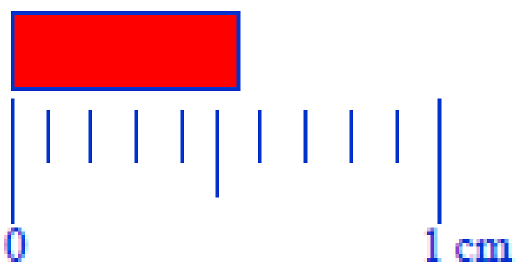
30.26 mL

the first uncertain digit

0. 55 cm?

0. 56 cm?

2 signif. figs



0. 55 cm implies an error of at least 0.55 ± 0.01 cm

Importance of Significant Figures

Type of balance	Precision (s) (mg)	Price ($)
Macrobalance	± 0.1 mg	~$3,000
SemimicroAnalytical balance	± 0.01 mg	~ $7,000
Analytical balance	± 0.001 mg	~ $13,000
Analytical balance	± 0.0003 mg	~ $16,000

a) 1.23 g b) 1.230 g c) 1.2300 g

Rules of thumb for determining the number of significant figures

1) Disregard all initial zeros (*Leading zeros* are *not significant*).

2) Disregard all final zeros unless they follow a decimal point (*Trailing zeros* are *significant*).

3) All remaining digits including *zeros between nonzero integers* are *significant*.

(A) 0.002 has ________ significant figure.	(A) 1
(B) 0.0202 has ________ significant figure.	(B) 3
(C) 0.0020 has ________ significant figure	(C) 2
(D) 24.00 has ________ significant figure.	(D) 4

Chemical Engineering

Inspiring Potential · Generating Futures

UNIVERSITI TEKNOLOGI PETRONAS

Examples

1) 1, 20 and 300 have ? significant figures

2) 123.45 has ? significant figures

3) 1001 has ? significant figures

4) 100.02 has ? significant figures

5) 0.00001 has ? significant figures

6) 1.100 has ? significant figures

7) 0.00100 has ? significant figures

Answers

1) 1, 20 and 300 have 1 significant figures

2) 123.45 has 5 significant figures

3) 1001 has 4 significant figures

4) 100.02 has 5 significant figures

5) 0.00001 has 1 significant figures

6) 1.100 has 4 significant figures

7) 0.00100 has 3 significant figures

Rules Regarding Significant Figures in Calculation

- For Addition and Subtraction, the result should contain the same number of <u>decimal places</u> as the number with the <u>smallest number of decimal places</u>.

Example:

C_1 = 6.31 ppm *(smallest number of decimal places, 2 decimal places)*

+ C_2 = 8.736 ppm

Therefore, C_T = 15.046 ppm

≅ <u>15.05 ppm</u>

(smallest number of decimal places = 2)

Example

3.76 g + 14.83 g + 2.1 g = 20.69 g

2.1 shows the smallest number of decimal places = 1

We must round our answer, 20.69, to one decimal place. Therefore, our final answer is <u>20.7 g</u>.

- For Multiplication and Division, the result should contain the same number of significant figures in the number with the smallest number of significant figures.

Example:

Calculate the value of

$$22.37\ \text{cm} \times 3.10\ \text{cm} \times 85.75\ \text{cm} = 5946.50525\ \text{cm}^3$$

22.37 shows 4 significant figures

3.10 shows 3 significant figures

85.75 shows 4 significant figures

The smallest number of significant figures is 3

Therefore, the final answer becomes $5950\ \text{cm}^3$

When dropping non-significant figures

0.0148 → 0.015 (2 s.f)

→ 0.01 (1 s.f)

Rounding Data

- Round up for digits > or = 5, and round down for digits < 5.
- In rounding a number ending in 5, always round so that the result ends with an even number.
- For example: 0.635 rounds to 0.64 and 0.625 rounds to 0.62.
- Remember not to round off calculations until the final result is obtained!

Example 12

How many significant figures does each of the following numbers have?

(1)200.06

(2)6.030 x 10^{-4}

(3)7.80 x 10^{10}

Answer

(1)200.06 → 5 s.f.

(2)6.030×10^{-4} → 4 s.f.

(3)7.80×10^{10} → 3 s.f.

Example 13

Give the answer to the following maximum number of significant figures: 50.00 x 27.8 x 0.1167.

Answer

50.00 x 27.8 x 0.1167 = 162.213

50.00 shows 4 significant figures

27.8 shows 3 significant figures

0.1167 shows 4 significant figures

The maximum number of significant figures = 4, so, the final answer become = 162.2

End of Chapter

Chemical Engineering
Inspiring Potential · Generating Futures

UNIVERSITI TEKNOLOGI PETRONAS

Recap: Outline for Introduction Chapter

- Analytical Objectives
- Qualitative Analysis
- Quantitative Analysis
- Analytical Methodology
- Type of equipment which includes Gravimetry, Spectrophotometry, Spectroscopy and Chromatography.
- Basic Tools and Operation
- Data Handling and Statistic
- Errors

Exercise 2

Q1:

Consider the following set of replicate measurements:

0.812, 0.792, 0.794, 0.900

Calculate the (i) mean; (ii) median; (iii) spread; (iv) standard deviation and (v) coefficient of variation.

The accepted value is 0.830. By using the mean value, calculate (i) the absolute error and (ii) the relative error in parts per thousand (ppt).

Q2:

Write each answer with the correct number of digits.

a. $1.021 + 2.69 =$
b. $12.3 - 1.63 =$
c. $4.34 \times 9.2 =$
d. $0.0602 / (2.113 \times 10^4) =$

UNIVERSITI TEKNOLOGI PETRONAS

Q3:

Round each numbers as indicated.

a. 1.2367 to 4 significant figures.
b. 1.2384 to 4 significant figures.
c. 0.1352 to 3 significant figures.
d. 2.051 to 2 significant figures.
e. 2.0050 to 3 significant figures.

Chemical Engineering
Inspiring Potential · Generating Futures