Evaluation of analytical data

Dr Wathiq Al-Hachami

Types of Error in Experimental Data

- Random (indeterminate) Error

Affects precision only, deviation in both directions, cannot be eliminated. Can be reduced by the use of more precise equipment and techniques.

- Systematic (determinate) Error

Affects accuracy only, deviation in one direction, readings all too high or too low. Can be identified and eliminated.

- Gross Errors

Easily recognised since they involve a major breakdown in the analytical process (samples being spilt, wrong dilution, or wrong using of instrument), it is detectable by carrying out sufficient replicate measurements.

Sources of Systematic Error

- Instrument errors: Need frequent calibration - both for apparatus such as volumetric flasks, burettes etc., but also for electronic devices such as spectrometers.

- Method errors: introduced when the assumptions about the relationship between the signal and the analyte are invalid.

- Sampling errors: occurs when the sampling strategy does not produce a representative sample.

- Personal errors: measurements are subjected to human error.

e.g. going over the equivalence point in a titration.

Systematic errors can be:

- **Constant** (e.g. error in burette reading - less important for larger values of reading) or,
- **Proportional** (e.g. presence of given proportion of interfering impurity in sample; equally significant for all values of measurement).

- **Minimise instrumental errors** by careful recalibration and good maintenance of equipment.

- **Minimise personal errors** by care and self-discipline
- **Minimise method errors** - most difficult. “True” value may not be known.

The Mean and Median

- **Mean**: an average obtained by dividing the sum of the measurement by the number of measurements

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

- **Median**: the middle value when the data is in numerical order from lowest to highest value.

- For an even number of values, the median is the mean of the two central values. For an odd number of data points, the median can be evaluated directly.

Example

- Find the mean of the set {2,5,5,6,8,8,9,11}.

Answer:

$$mean = \frac{2+5+5+6+8+8+9+11}{8} = 6.75$$

- Find the median of the set {2,5,8,11,16,21,30}.

Answer:

There are seven numbers (ODD) in the set, and they are arranged in ascending order. The middle number (the 4^{th} one in the list) is 11. so, the median is 11.

Find the median of the set {3,10,36,255,79,24,5,8}.

Answer:

1. Firstly, arrange the numbers in ascending order,

{3,5,8,10,24,36,79,255}

2. There are (8) numbers in the set (*EVEN*), so we have to find the two central numbers to calculate the *median.*

3. The middle numbers are {10,24}

So $$median = \frac{10+24}{2} = 17$$

Precision and accuracy

- **Precision** describes how closely measurements are to each other and how carefully measurements were made. It can be described in ***two*** expressions:
- Repeatability: the precision of one set of conditions (same analyst, same day, same session, and same equipment used).
- Reproducibility: the precision under different set of conditions (different days or/and different analysts).
- **Precision describes the agreement among several results that have been obtained in the same way.**

Precision and accuracy

- **Accuracy** indicates the closeness of the measurement to its true or accepted value and is expressed by the error.
- Absolute Error
  - The *absolute error E* in the measurement of a quantity x_i is given by the equation

$$E = x_i - x_t$$

- Relative error
  - Often, the *relative error* E_r is a more useful quantity than the absolute error. The percent relative error is given by the expression

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

  - **Accuracy measures agreement between a result and its true value.**

Precision and accuracy

Accuracy: Refers to how far you are from the true value. (True value is yellow circle)

This is accurate but not Precise.

Precision: Refers to how close do you get to the same value if you repeated the exercise several times.

This is **precise but not accurate.**

Precise and accurate: Quality of product will be considered high if results are **Precise and accurate.**

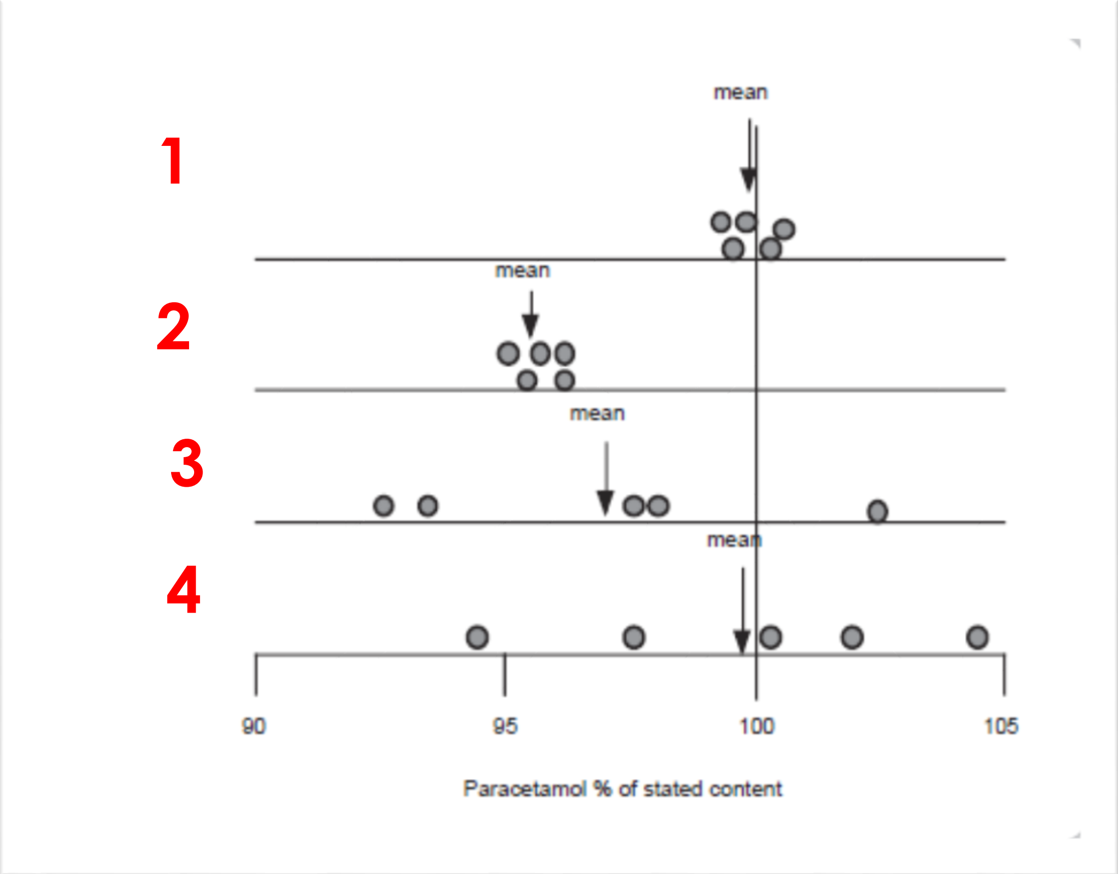
Example about accuracy and precision

- A batch of paracetamol tablets are stated to contain 500 mg of paracetamol per tablet. Four students carry out a spectrophotometric analysis of an extract from the tablets and obtain the following percentages of stated content for the repeat analysis of paracetamol in the tablets:

Student 1	99.5%,	99.9%	100.2%	99.4%	100.5%
Student 2	95.6%	96.1%	95.2%	95.1%	96.1%
Student 3	93.5%	98.3%	92.5%	102.5%	97.6%
Student 4	94.4%	100.2%	104.5%	97.4%	102.1%

Example

1 is precise and accurate
2 is precise and inaccurate (systematic)
3 is imprecise and inaccurate (random)
4 is imprecise and accurate



Measures of Variation

- The **range** of a set of data is the difference between the greatest and least values.
- The **interquartile range** is the difference between the third and first quartiles
- The **variance** is a measure of how data points differ from the mean

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

- **Standard deviation** is a measure of how each value in a data set varies or deviates from the mean

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

- **Relative standard deviation:**

$$RSD = \frac{s}{mean} * 100$$

Example

There are (9) members of the Community Youth Leadership Board. Find the range and interquartile range of their ages: 22, 16, 24, 17, 16, 25, 20, 19, 26.

greatest value – least value = 26 – 16 Find the range.

= 10

median

16 16 17 19 | 20 | 22 24 25 26

$$Q_1 = \frac{(16 + 17)}{2} = 16.5 \qquad Q_3 = \frac{(24 + 25)}{2} = 24.5$$

Find the interquartile range.

$$Q_3 - Q_1 = 24.5 - 16.5 = 8$$

The range is 10 years. The interquartile range is 8 years.

Standard Deviation

Find the mean and the standard deviation for the values 78.2, 90.5, 98.1, 93.7, 94.5.

$$\bar{x} = \frac{(78.2 + 90.5 + 98.1 + 93.7 + 94.5)}{5} = 91$$

Find the mean.

x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
78.2	91	–12.8	163.84
90.5	91	–0.5	.25
98.1	91	7.1	50.41
93.7	91	2.7	7.29
94.5	91	3.5	12.25

Organize the next steps in a table.

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N-1}}$$

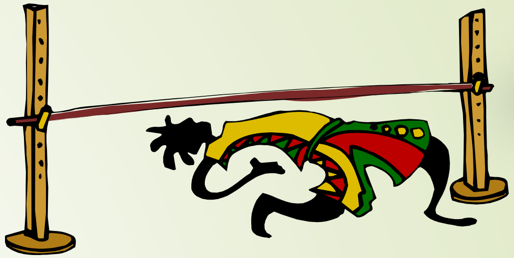
Find the standard deviation.

$$\sqrt{\frac{234.04}{4}} = \approx \quad 7.65$$

The mean is 91, and the standard deviation is about 7.65.

Method Validation- LOD and LOQ

Sensitivity (How low can you go?)



- Limit of detection (LOD) – "the lowest content that can be measured with reasonable statistical certainty.", it can be calculated by the following equation:

$$LOD = \frac{3s}{slope\ of\ the\ calibration\ graph}$$

- Limit of quantitative measurement (LOQ) – "the lowest concentration of an analyte that can be determined with acceptable precision (repeatability) and accuracy under the stated conditions of the test", it can be calculated by:

$$LOQ = \frac{10s}{slope\ of\ the\ calibration\ graph}$$