Errors in quantitative analysis

Key terms and concepts

- Errors, accuracy, and precision
- Define absolute and relative uncertainties
- Statistics of repeated measurements
  - ✓ Mean, standard deviation, RSD
  - ✓ Pooled standard deviation
  - ✓ Gaussian distribution
  - ✓ Confidence limits of the mean for large samples
  - ✓ Confidence limits of the mean for small samples

- Significance tests
  - ✓ Comparison of an experimental mean with a known value
  - ✓ Comparison of two experimental means
  - ✓ Paired t-test
  - ✓ F-test for the comparison of standard deviations
  - ✓ Grubb's test for an outlier

Errors, accuracy, and precision

- We are always uncertain to some extent about a measurement and **this uncertainty is** called **experimental error.** Conclusions can be expressed with a high or a low degree of confidence, but never with complete certainty. Experimental errors are classified as either systematic or random

Systematic errors

- Systematic error, also called determinate errors, arise from a flaw in equipment/instrumental defects or malfunctioning; the design of an experiment-impurities in reagents, personal errors, method errors etc.,.

- These types of errors tend to give results that are alwayes higher or lower than the true value, that is they are unidirectional.
- They are not random in nature and occur when something is wrong with the measurement. Thus, If you conduct the experiment again in exactly the same manner, the errors are reproducible
- Systematic errors will, thus, bias the results and influence the accuracy of the analysis. With care and cleverness, you can detect and correct a systematic error, although this may not be easy
- Some examples of such errors are those caused by incorrectly calibrated pH meter, partial loss of a volatile analyte, recording of titration endpoints, etc.,

Random Errors

- Random error, also called indeterminate error, arises from causes over which the analyst has no control. They arises from the **effects of uncontrolled** (and maybe uncontrollable) **variables** in the measurement. They are the slight variations that occur when successive measurements are made by the same observer under identical conditions.
- For instance, there is random error associated with reading a scale. One person reading the same instrument several times might report several different readings

- Another random error results from electrical noise in an instrument. Positive and negative fluctuations occur with approximately equal frequency and cannot be completely eliminated
- These types of errors arise from individual results falling on both sides of the true response, that is they are bidirectional in nature that affect the results irregularly
- They have an **equal chance of causing** a measurement to **be positive or negative.** It is always present and cannot be corrected. However, it might be reduced by changing the experiment, but it cannot be eliminated

- Comparison of systematic errors and random errors is shown in table 1.2.

Table 1.2 Random and systematic errors

Random errors	Systematic errors
Affect **precision** – repeatability or reproducibility	Produce **bias** – an overall deviation of a result from the true value even when random errors are very small
Cause replicate results to fall on either side of a mean value	Cause all results to be affected in one sense only – all too high or all too low
Can be estimated using replicate measurements	Cannot be detected simply by using replicate measurements
Can be minimized by good technique but not eliminated	Can be corrected, e.g. by using standard methods and materials
Caused by both humans and equipment	Caused by both humans and equipment

Precision and Accuracy

- **Precision** describes how closely together a set of measurements are to each other. In other words, it is **the degree of agreement among a series of replicate** measurements of the same quantity. It describes the reproducibility of a result.
- If you measure a quantity several times and the values agree closely with one another, your measurement is precise. If the values vary widely, your measurement is not precise
- The precision of a measurement can be expressed by calculating statistical parameters such standard deviation, coefficient of variation, etc.,

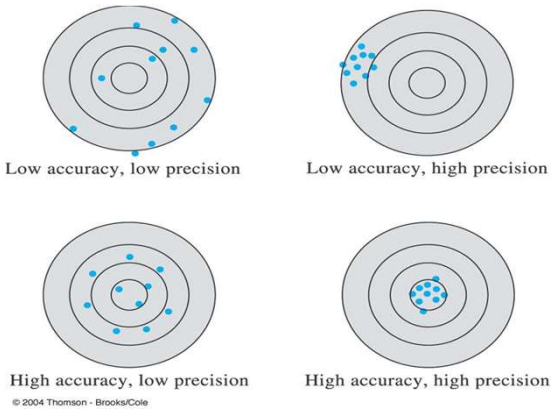
- **Accuracy** refers to the **degree of agreement between a measured value and the "true value"** or accepted or most probable value. In other words, it describes how close the average of a measured value is to the "true" value. It is also defined as nearness to the "true" value
- Determining the accuracy of a measurement usually requires comparing the results obtained by the analytical method with a known standard or reference material (such as a Standard Reference Material). So accuracy is how close your value is to the known standard value.
- It is usually expressed by the absolute or relative errors

How do one's get the "true" value?

- Somebody must measure the "true" value, and there is error associated with every measurement
- The "true" value is best obtained by an experienced person using a well-tested procedure. It is desirable to test the result by using different procedures, because, even though each method might be precise, systematic error could lead to poor agreement between methods
- Good agreement among several methods affords us confidence, but never proof, that results are accurate

- A measurement might be precise, but inaccurate
  - If you made a mistake preparing a solution for a titration, you might do a series of reproducible titrations but report an incorrect result because the concentration of the titrating solution was not what you intended. In this case, the precision is good but the accuracy is poor
- Conversely, a measurement might be accurate, but imprecise
  - if make poorly reproducible measurements clustered around the correct value, the precision is poor but the accuracy is good.
- An **ideal procedure** is both precise and accurate

Relationship between accuracy and precision



Absolute and Relative Uncertainty

Absolute uncertainty expresses the margin of uncertainty associated with a measurement. If the estimated uncertainty in reading a calibrated buret is ±0.02 mL, we say that ±0.02 mL is the absolute uncertainty associated with the reading.

Relative uncertainty compares the size of the absolute uncertainty with the size of its associated measurement. The relative uncertainty of a buret reading of 12.35 ± 0.02 mL is a dimensionless quotient:

Relative uncertainty:

$$\text{Relative uncertainty} = \frac{\text{absolute uncertainty}}{\text{magnitude of measurement}} \qquad (3\text{-}2)$$

$$= \frac{0.02\ \text{mL}}{12.35\ \text{mL}} = 0.002$$

The percent relative uncertainty is simply

Percent relative uncertainty:

$$\text{Percent relative uncertainty} = 100 \times \text{relative uncertainty} \qquad (3\text{-}3)$$

$$= 100 \times 0.002 = 0.2\%$$

If the absolute uncertainty in reading a buret is constant at ±0.02 mL, the percent relative uncertainty is 0.2% for a volume of 10 mL and 0.1% for a volume of 20 mL.

Propagation of uncertainty from random error

- We can usually estimate or measure the random error associated with a measurement, such as the length of an object or the temperature of a solution
- The uncertainty might be based on how well we can read an instrument or on our experience with a particular method. ***If possible, uncertainty is expressed*** as the ***standard deviation or as a confidence interval,*** which are discussed later
- In an experiment, we usually need to perform mathematical operations on several numbers, each of which has some uncertainty as a result of random error (we will assume for the moment that there is no systematic error)

- However, we **can not simply add the individual uncertainties** because it is likely that some of the errors are positive and others negative. That is we expect some of the errors will cancel

Addition and Subtraction

Suppose you wish to perform the following arithmetic, in which the experimental uncertainties, designated e_1, e_2, and e_3, are given in parentheses.

$$\begin{array}{r} 1.76\ (\pm 0.03) \leftarrow e_1 \\ +\ 1.89\ (\pm 0.02) \leftarrow e_2 \\ -\ 0.59\ (\pm 0.02) \leftarrow e_3 \\ \hline 3.06\ (\pm e_4) \end{array} \qquad (3\text{-}4)$$

The arithmetic answer is 3.06. But what is the uncertainty associated with this result?

For addition and subtraction, the uncertainty in the answer is obtained from the *absolute uncertainties* of the individual terms as follows:

Uncertainty in addition and subtraction:

$$e_4 = \sqrt{e_1^2 + e_2^2 + e_3^2} \qquad (3\text{-}5)$$

For the sum in Equation 3-4, we can write

$$e_4 = \sqrt{(0.03)^2 + (0.02)^2 + (0.02)^2} = 0.04_1$$

The absolute uncertainty e_4 is ±0.04, and we express the answer as 3.06 ± 0.04. Although there is only one significant figure in the uncertainty, we wrote it initially as 0.04_1, with the first insignificant figure subscripted. We retain one or more insignificant figures to avoid introducing round-off errors into later calculations through the number 0.04_1. The insignificant figure was subscripted to remind us where the last significant figure should be at the conclusion of the calculations.

To find the percent relative uncertainty in the sum of Equation 3-4, we write

$$\text{Percent relative uncertainty} = \frac{0.04_1}{3.06} \times 100 = 1_3\%$$

The uncertainty, 0.04_1, is 1_3% of the result, 3.06. The subscript 3 in 1_3% is not significant. It is sensible to drop the insignificant figures now and express the final result as

3.06 (±0.04) (absolute uncertainty)
3.06 (±1%) (relative uncertainty)

Multiplication and Division

For multiplication and division, first convert all uncertainties into percent relative uncertainties. Then calculate the error of the product or quotient as follows:

Uncertainty in multiplication and division:

$$\%e_4 = \sqrt{(\%e_1)^2 + (\%e_2)^2 + (\%e_3)^2} \qquad (3\text{-}6)$$

For example, consider the following operations:

$$\frac{1.76\ (\pm 0.03) \times 1.89\ (\pm 0.02)}{0.59\ (\pm 0.02)} = 5.64 \pm e_4$$

First convert absolute uncertainties into percent relative uncertainties.

$$\frac{1.76\ (\pm 1.7\%) \times 1.89\ (\pm 1.1\%)}{0.59\ (\pm 3.4\%)} = 5.64 \pm e_4$$

Then find the percent relative uncertainty of the answer by using Equation 3-6.

$$\%e_4 = \sqrt{(1.7)^2 + (1.1)^2 + (3.4)^2} = 4.0\%$$

The answer is $5.6_4\ (\pm 4.0\%)$.

To convert relative uncertainty into absolute uncertainty, find 4.0% of the answer.

$$4.0\% \times 5.6_4 = 0.04_0 \times 5.6_4 = 0.2_3$$

The answer is $5.6_4\ (\pm 0.2_3)$. Finally, drop the insignificant digits.

$5.6\ (\pm 0.2)$ (absolute uncertainty)
$5.6\ (\pm 4\%)$ (relative uncertainty)

The denominator of the original problem, 0.59, limits the answer to two digits.

Mixed Operations

Now consider a computation containing subtraction and division:

$$\frac{[1.76\ (\pm 0.03) - 0.59\ (\pm 0.02)]}{1.89\ (\pm 0.02)} = 0.619_0 \pm ?$$

First work out the difference in the numerator, using absolute uncertainties. Thus,

$$1.76\ (\pm 0.03) - 0.59\ (\pm 0.02) = 1.17\ (\pm 0.03_6)$$

because $\sqrt{(0.03)^2 + (0.02)^2} = 0.03_6$.

Then convert into percent relative uncertainties. Thus,

$$\frac{1.17\ (\pm 0.03_6)}{1.89\ (\pm 0.02)} = \frac{1.17\ (\pm 3.1\%)}{1.89\ (\pm 1.1\%)} = 0.619_0\ (\pm 3.3\%)$$

because $\sqrt{(3.1\%)^2 + (1.1\%)^2} = 3.3\%$.

The percent relative uncertainty is 3.3%, so the absolute uncertainty is $0.03_3 \times 0.619_0 = 0.02_0$. The final answer can be written as

$0.619\ (\pm 0.02_0)$ (absolute uncertainty)
$0.619\ (\pm 3.3\%)$ (relative uncertainty)

Because the uncertainty begins in the 0.01 decimal place, it is reasonable to round the result to the 0.01 decimal place:

$0.62\ (\pm 0.02)$ (absolute uncertainty)
$0.62\ (\pm 3\%)$ (relative uncertainty)

Exponents and Logarithms

For the function $y = x^a$, the percent relative uncertainty in y ($\%e_y$) is equal to a times the percent relative uncertainty in x ($\%e_x$):

Uncertainty for powers and roots:

$$y = x^a \Rightarrow \%e_y = a(\%e_x) \qquad (3\text{-}7)$$

For example, if $y = \sqrt{x} = x^{1/2}$, a 2% uncertainty in x will yield a $(\frac{1}{2})(2\%) = 1\%$ uncertainty in y. If $y = x^2$, a 3% uncertainty in x leads to a $(2)(3\%) = 6\%$ uncertainty in y (Box 3-2).

If y is the base 10 logarithm of x, then the absolute uncertainty in y (e_y) is proportional to the relative uncertainty in x, which is e_x/x:

Uncertainty for logarithm:

$$y = \log x \Rightarrow e_y = \frac{1}{\ln 10}\frac{e_x}{x} \approx 0.434\ 29\frac{e_x}{x} \qquad (3\text{-}8)$$

You should not work with percent relative uncertainty $[100 \times (e_x/x)]$ in calculations with logs and antilogs, because one side of Equation 3-8 has relative uncertainty and the other has absolute uncertainty.

The **natural logarithm** (ln) of x is the number y, whose value is such that $x = e^y$, where e (= 2.718 28 . . .) is called the base of the natural logarithm. The absolute uncertainty in y is equal to the relative uncertainty in x.

Uncertainty for natural logarithm:

$$y = \ln x \Rightarrow e_y = \frac{e_x}{x} \qquad (3\text{-}9)$$

Now consider $y = \text{antilog}\, x$, which is the same as saying $y = 10^x$. In this case, the relative uncertainty in y is proportional to the absolute uncertainty in x.

Uncertainty for 10^x:

$$y = 10^x \Rightarrow \frac{e_y}{y} = (\ln 10)e_x \approx 2.302\,6\,e_x \qquad (3\text{-}10)$$

If $y = e^x$, the relative uncertainty in y equals the absolute uncertainty in x.

Uncertainty for e^x:

$$y = e^x \Rightarrow \frac{e_y}{y} = e_x \qquad (3\text{-}11)$$

Table 3-1 summarizes rules for propagation of uncertainty. You need not memorize the rules for exponents, logs, and antilogs, but you should be able to use them.

Example Uncertainty in H^+ Concentration

Consider the function pH = $-\log[H^+]$, where $[H^+]$ is the molarity of H^+. For pH = 5.21 ± 0.03, find $[H^+]$ and its uncertainty.

Solution First solve the equation pH = $-\log[H^+]$ for $[H^+]$: Whenever $a = b$, then $10^a = 10^b$. If pH = $-\log[H^+]$, then $\log[H^+] = -\text{pH}$ and $10^{\log[H^+]} = 10^{-\text{pH}}$. But $10^{\log[H^+]} = [H^+]$. We therefore need to find the uncertainty in the equation

$$[H^+] = 10^{-\text{pH}} = 10^{-(5.21 \pm 0.03)}$$

In Table 3-1, the relevant function is $y = 10^x$, in which $y = [H^+]$ and $x = -(5.21 \pm 0.03)$. For $y = 10^x$, the table tells us that $e_y/y = 2.302\,6\,e_x$.

$$\frac{e_y}{y} = 2.302\,6\,e_x = (2.302\,6)(0.03) = 0.069\,1 \qquad (3\text{-}12)$$

The relative uncertainty in y ($= e_y/y$) is 0.069 1. Inserting the value $y = 10^{-5.21} = 6.17 \times 10^{-6}$ into Equation 3-12 gives the answer:

$$\frac{e_y}{y} = \frac{e_y}{6.17 \times 10^{-6}} = 0.069\,1 \quad \Rightarrow \quad e_y = 4.26 \times 10^{-7}$$

The concentration of H^+ is $6.17\ (\pm 0.426) \times 10^{-6} = 6.2\ (\pm 0.4) \times 10^{-6}$ M. An uncertainty of 0.03 in pH gives an uncertainty of 7% in $[H^+]$. Notice that extra digits were retained in the intermediate results and were not rounded off until the final answer.

Table 3-1 Summary of rules for propagation of uncertainty

Function	Uncertainty	Function[a]	Uncertainty[b]
$y = x_1 + x_2$	$e_y = \sqrt{e_{x_1}^2 + e_{x_2}^2}$	$y = x^a$	$\%e_y = a\%e_x$
$y = x_1 - x_2$	$e_y = \sqrt{e_{x_1}^2 + e_{x_2}^2}$	$y = \log x$	$e_y = \frac{1}{\ln 10}\frac{e_x}{x} \approx 0.434\,29\frac{e_x}{x}$
$y = x_1 \cdot x_2$	$\%e_y = \sqrt{\%e_{x_1}^2 + \%e_{x_2}^2}$	$y = \ln x$	$e_y = \frac{e_x}{x}$
$y = \frac{x_1}{x_2}$	$\%e_y = \sqrt{\%e_{x_1}^2 + \%e_{x_2}^2}$	$y = 10^x$	$\frac{e_y}{y} = (\ln 10)e_x \approx 2.302\,6\,e_x$
		$y = e^x$	$\frac{e_y}{y} = e_x$

a. x represents a variable and a represents a constant that has no uncertainty.

b. e_x/x is the relative error in x and $\%e_x$ is $100 \times e_x/x$.

3-15. Find the absolute and percent relative uncertainty and express each answer with a reasonable number of significant figures.

(a) $6.2\ (\pm 0.2) - 4.1\ (\pm 0.1) = ?$

(b) $9.43\ (\pm 0.05) \times 0.016\ (\pm 0.001) = ?$

(c) $[6.2\ (\pm 0.2) - 4.1\ (\pm 0.1)] \div 9.43\ (\pm 0.05) = ?$

(d) $9.43\,(\pm 0.05) \times \{[6.2\,(\pm 0.2) \times 10^{-3}] + [4.1\,(\pm 0.1) \times 10^{-3}]\} = ?$

Describing a small set of data

- Simple techniques which can be used to summarize a set of data includes
- The **mean,** defined by:

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

Where x_i = individual values of x and N = number of replicate measurements

Median

- The middle result when data are arranged in order of size. With an even number of determination, take the average of the middle two.

- **The range** of the sample can be obtained by subtracting the smallest value from the largest value
- **Sample standard deviation, s,** measures how closely the data are clustered about the mean
- It can be calculated as follows for small samples of data, i.e. small N

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

- The quantity N-1 is known as degrees of freedom

- For a population (a data set containing every possible data point)mean is symbolized μ and standard deviation is symbolized σ
- **Population Standard Deviation (σ)** - defined as a measure of precision of a population of data, given by:

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

Where μ = population mean; N is very large.

- Since we virtually never can sample all of a population in chemistry, we rarely use μ and σ

Variance: This is the square of the standard deviation

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

Coefficient of variance (CV) (or Relative standard deviation): express the sample standard deviation as a percentage of the sample mean

$$CV = (\frac{s}{\bar{x}}) \times 100\% \qquad \text{RSD} = \frac{s}{\bar{x}} \times 100\%$$

- ❖ Less than1% is usually considered very good for routine measurements which are more often in the 1–5% range.

Standard Error of a Mean

- The standard error of the mean, is defined as follows

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

Pooled standard deviation

- To achieve a value of s which is a good approximation to σ, i.e. N ≥ 20, it is sometimes necessary to pool data from a number of sets of measurements (all taken in the same way)

The equation for computing a pooled standard deviation from several sets of data takes the form

$$s_{\text{pooled}} = \sqrt{\frac{\sum_{i=1}^{N_1}(x_i - \bar{x}_1)^2 + \sum_{j=1}^{N_2}(x_j - \bar{x}_2)^2 + \sum_{k=1}^{N_3}(x_k - \bar{x}_3)^2 + \cdots}{N_1 + N_2 + N_3 + \cdots - N_t}} \quad (6\text{-}7)$$

where N_1 is the number of results in set 1, N_2 is the number in set 2, and so forth. The term N_t is the total number of data sets that are pooled.

- Alternatively the pooled standard deviation can also be calculated by using the following expression:

$$s_p = \sqrt{\frac{s_1^2(N_1-1) + s_2^2(N_2-1) + \ldots + s_n^2(N_n-1)}{N_1 + N_2 + \ldots + N_n - n}}$$

where s_i is the standard deviation of N_i set of measurements.

- The pooled standard deviation is sometimes used to obtain an improved estimate of the precision of a method and it is used for calculating the precision of two sets of data in a paired 't' test.

Practice exercise:

Activity 2.1: 1) Why replicate samples analysis is better than single sample analysis?
2) Analyses of sample of iron are gave the following percentages values for the iron content: 7.08, 7.21, 7.12, 7.09, 7.16, 7.14, 7.07, 7.14, 7.18 and 7.11 Calculate the mean, standard deviation, RSD, variance, median, mode, and coefficient of variation for these values.
3) Calculate the pooled standard deviation from the following data obtained for volumetric determination of cadmium in drinking water samples.

Sample	Cd^{+2}, in ppm
1	12.3, 16.5, 11.8, 15.3
2	28.6, 25.7, 23.7
3	35.1, 29.9, 33.2, 36.4

Describing a large set of data

- **For example,** we will take sixty determinations of Copper content (in ppm) in a reference sample as presented in Table 3.1

Table 3.1 Sixty determinations of Cu content (in ppm) in a reference sample

61.0	65.4	60.0	59.2	57.0	62.5	57.7	56.2	62.9	62.5
56.5	60.2	58.2	56.5	64.7	54.5	60.5	59.5	61.6	60.8
58.7	54.4	62.2	59.0	60.3	60.8	59.5	60.0	61.8	63.8
64.5	66.3	61.1	59.7	57.4	61.2	60.9	58.2	63.0	59.5
56.0	59.4	60.2	62.9	60.5	60.8	61.5	58.5	58.9	60.5
61.2	57.8	63.4	58.9	61.5	62.3	59.8	61.7	64.0	62.7

Data taken from Caulcutt and Boddy [3].

- With this larger set of data it is difficult to get a feel for the spread of the determinations and the need to summarize the sixty determination is very clear

- With such a large set of data, calculation of the mean, standard deviation is rather tedious unless a suitable calculator is available. Using such calculators the sample mean and the median is 60.37and 60.50 respectively; the sample standard deviation is equal to 2.541 ppm.
- A simple blob chart is shown in figure

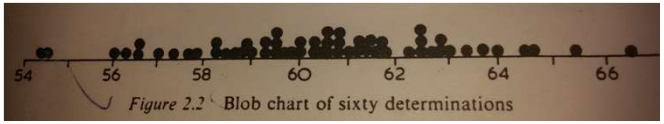


Figure 2.2 Blob chart of sixty determinations

- With such a larger set of data, there is a need to group the determinations before attempting a pictorial representation. This grouping has been carried out in Table 2.2.

Table 2.2 Frequency distribution of sixty determinations

Table 2.2 Frequency distribution of sixty determinations

Copper concentration (p.p.m.)	54.0 to 54.9	55.0 to 55.9	56.0 to 56.9	57.0 to 57.9	58.0 to 58.9	59.0 to 59.9	60.0 to 60.9	61.0 to 61.9	62.0 to 62.9	63.0 to 63.9	64.0 to 64.9	65.0 to 65.9	66.0 to 66.9	Total
No. of determinations	2	0	4	4	6	8	12	9	7	3	3	1	1	60

- Table 2.2 is known as a frequency distribution. It indicates how the determinations are spread or distributed by telling us how many determinations fall into each of the thirteen groups. The numbers in the bottom row of the table are often referred to as frequencies.
- As can be seen in table 2.2 that two of the sixty determinations lie between 54.0 and 54.9 whilst there are none between 55.0 and 55.9 etc. It is very clear that a large percentage of the determinations lie in the middle five groups extending from 58.0 to 62.9.

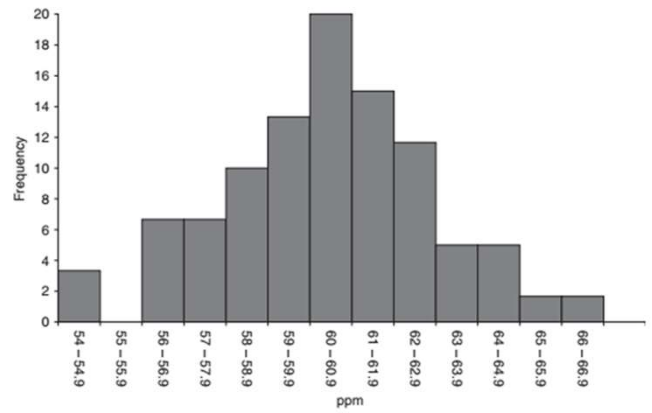


Figure 3.3 Frequency histogram

- It is often useful to have the frequencies of a frequency distribution expressed as percentages of the total frequency. If the numbers in the bottom raw of table 2.2 are divided by 60 and multiplied by 100 we get the percentage frequencies in Table 2.3.

Table 2.3 percentage frequency distribution of sixty determinations

Table 2.3 Percentage frequency distribution of sixty determinations

Copper concentration (p.p.m.)	54.0 to 54.9	55.0 to 55.9	56.0 to 56.9	57.0 to 57.9	58.0 to 58.9	59.0 to 59.9	60.0 to 60.9	61.0 to 61.9	62.0 to 62.9	63.0 to 63.9	64.0 to 64.9	65.0 to 65.9	66.0 to 66.9	Total
% of determinations	3.3	0.0	6.7	6.7	10.0	13.3	20.0	15.0	11.7	5.0	5.0	1.7	1.7	100%

- Table 2.3 tells us the percentage of the sixty determinations which fall into each group. The same information is presented in a graphical form in figure 2.3.

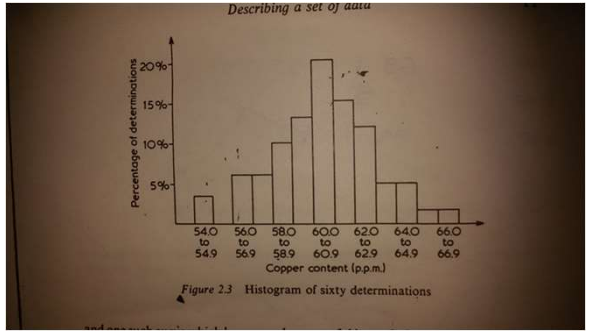
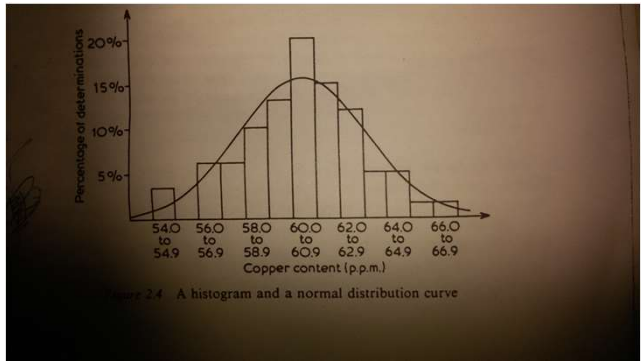


Figure 2.3 Histogram of sixty determinations

- Each vertical bar in the histogram has a height which is proportional to the percentage of determinations in the group represented by the bar
- As shown in the histogram the bulk of the distribution is bunched in the center close to the true concentration of 60.0 ppm. It has two 'tails' which represent the highest and lowest determinations. The histogram tells us very little about the largest or smallest determination we are likely to get since most of the information is bunched in the centre. One way round this difficulty is to replace the histogram with a smooth curve and one such curve is known as the normal distribution curve.

The normal distribution

- The histogram in Figure 2.3 has been reproduced below together with a normal distribution curve.



A histogram and a normal distribution curve

- The normal curve in figure 2.4 has the same mean (60.37) and the same standard deviation (2.542) as the sixty determinations represented by the histogram.
- The histogram and the normal curve are not identical but they are very similar. It is not unreasonable to suggest that the histogram would get closer in shape to the normal curve if you took more and more determinations on the standard solutions. So it would be reasonable to assume that the normal curve is a representation of the whole population of determinations from which the sample was taken. Having made this assumption we could speak with more confidence about the small percentage of very high or very low determinations.

- Suppose **for example** we wanted to **estimate the percentage of determinations which would be greater** than 65.0. the required percentage can be obtained from the **normal distribution table** but to use this table you must first standardize the value (65.0) in which we are interested using the following formula

  Standardize value=(value-mean)/(standard deviation)

  So standardize value=(65.0-60.37)/2.541

  =1.822

- This result is telling us that the value (65.0) is 1.822 standard deviations above the mean
- Using a standard value of 1.822 we can see from table that ~ 3.4% of determination on the standard solution will exceed 65.0 ppm.

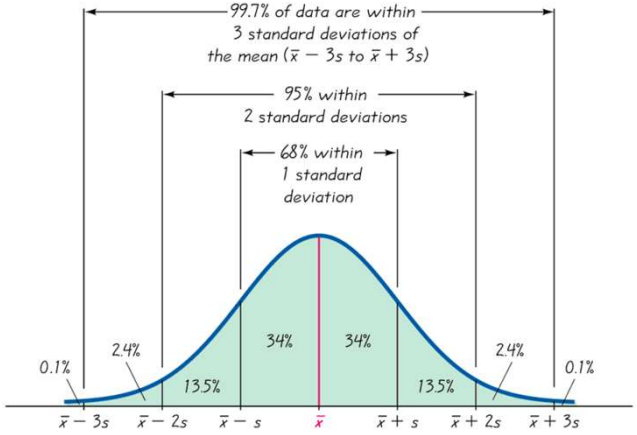
The normal distribution

- All normal distributions have some common features regardless of the mean or standard deviations. For a normal distribution:
  - The frequency curve is bell-shaped and symmetrical about the mean
  - The mean, median and mode are equal
  - The size and shape of the bell are determined by the mean and standard deviation of the distribution.
  - The "68-95-99.7 rule," states that approximately:

~68% will be within the range $(\bar{x}-s, \bar{x}+s)$

~95% will be within the range $(\bar{x}-2s, \bar{x}+2s)$

~99.7% will be within the range $(\bar{x}-3s, \bar{x}+3s)$



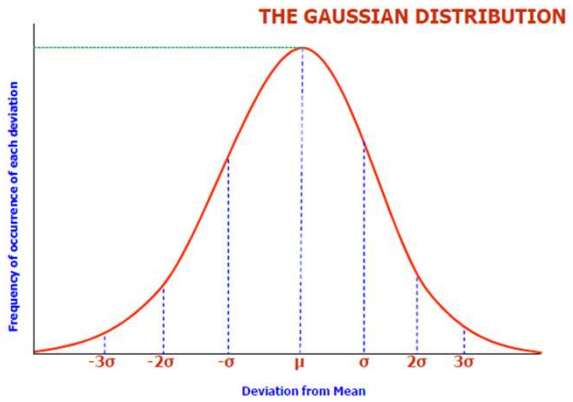
- The formula for area under the normal distribution curve is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

This expression is often called a probability density function (pdf) and is proportional to the chance that a reading has a value of x if a dataset has a true mean of μ and standard deviation of σ. The function is scaled so that the area under the curve between plus and minus infinity is 1. The centre of the normal distribution curve is at $x = \mu$ and the standard deviation σ is reached when the height of the curve is $e^{-1/2}$ or 0.6065 of that of the centre. This is illustrated in Figure 3.6.

- The integral of f(x) with respect to x b/n limits x=a and x=b is the probability of finding a result in that range. The equation gives us a curve shown in figure 1 below

Figure 1. Gaussian distribution curve



- The Gaussian distribution curve is characterized by two parameters: **μ the mean for the infinite population** of data that define the probability density function, and **σ the population standard deviation**
- The maximum of the function is when x = μ, and the larger the value of σ, the more spread out the function is, as shown in figure 2.3.
- If the data were results of repeated analyses of a sample, then σ is a measure of the precision of the analysis.

- The curve is symmetrical about μ and the greater the value of σ the greater the spread of the curve

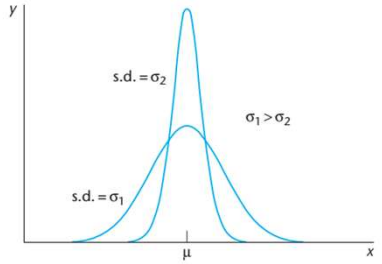


Figure 2.3 Normal distributions with the same mean but different values of the standard deviation.

- It is impractical to repeat an experiment so many times that we obtain a Gaussian distribution. Instead, we typically perform an experiment 3-5 times and use statistical analysis to estimate parameters for the large set (the population)

Confidence intervals

- **Student's t-** is a statistical tool often **used to find confidence intervals** and **compare results from different experiments**
- From **a finite set of measurements**, we cannot obtain μ or σ; we obtain $\bar{X}$ and s
  - We can then calculate a confidence interval within which we are certain to some level of confidence, that contains μ
  - **Confidence interval** = $\bar{x} \pm t_{n-1} s/\sqrt{n}$

where $\bar{X}$ - sample mean

s - sample standard deviation

n - sample size

t -is taken from the two-sided t-table with (n-1) degrees of freedom

- The t-value depends on the **probability level** required (α) and also the **degrees of freedom**, that is, the number of repeat measurements minus one.
- The difference (100%-confidence level) is defined as the error probability and gives the probability that μ falls outside the confidence interval.

In general confidence intervals for small samples and large samples are calculated as follows.

For small samples, the confidence limits of the mean are given by

$$\bar{x} \pm t_{n-1} s/\sqrt{n}$$

For large samples, the confidence limits of the mean are given by

$$\bar{x} \pm zs/\sqrt{n}$$

where the value of z depends on the degree of confidence required.

For 95% confidence limits, $z = 1.96$
For 99% confidence limits, $z = 2.58$
For 99.7% confidence limits, $z = 2.97$

For example

The sodium ion content of a urine specimen was determined by using an ion-selective electrode. The following values were obtained: 102, 97, 99, 98, 101, 106 mM. What are the 95% and 99% confidence limits for the sodium ion concentration?

The mean and standard deviation of these values are 100.5 mM and 3.27 mM respectively. There are six measurements and therefore 5 degrees of freedom. From Table A.2 the value of t_s for calculating the 95% confidence limits is 2.57 and from equation (2.9) the 95% confidence limits of the mean are given by:

$$100.5 \pm 2.57 \times 3.27/\sqrt{6} = 100.5 \pm 3.4 \text{ mM}$$

Similarly the 99% confidence limits are given by:

$$100.5 \pm 4.03 \times 3.27/\sqrt{6} = 100.5 \pm 5.4 \text{ mM}$$

Calculate the 95% and 99% confidence limits of the mean for the nitrate ion concentration measurements in Table 2.1.

We have $\bar{x} = 0.500$, $s = 0.0165$ and $n = 50$. Using equation (2.8) gives the 95% confidence limits as:

$$\bar{x} \pm 1.96 s/\sqrt{n} = 0.500 \pm 1.96 \times 0.0165/\sqrt{50} = 0.500 \pm 0.0046 \text{ μg ml}^{-1}$$

and the 99% confidence limits as:

$$\bar{x} \pm 2.58 s/\sqrt{n} = 0.500 \pm 2.58 \times 0.01651/\sqrt{50} = 0.500 \pm 0.0060 \text{ μg ml}^{-1}$$

Comparison of results using statistical tests of significance

- The comparison of the values obtained from a set of results with either the true value or other sets of data makes it possible to determine whether the analytical method is accurate and/or precise, or if it is superior or inferior to other method. The two common methods used for comparing results are student's t-test and F-test. A basic hypothesis made in such statistical tests is the **null hypothesis**, which assumes that any difference in numerical quantities compared like $\bar{X}$ and μ, or $\overline{X1}$ and $\overline{X2}$, or s1 and s2 is due only to random errors.

- So **checking the correctness** of a null hypothesis involves computing the probability that the observed difference are the results of random errors. Note that we cannot ever prove the null hypothesis is true; we can only disprove it
- If the **observed differences is equal to or less than** the difference expected at a certain probability level (95%, 99%, etc.,), then **the null hypothesis is considered acceptable,** and the difference is regarded as statistically insignificant and the difference is attributed to indeterminate errors
- If the **difference is greater than the** theoretically predicted value at a given probability level, then **the null hypothesis is reject** and the difference is **considered to be significant** or attributed to determinate errors

Let's examine three cases that are handled differently

(I) We measure some quantity multiple times and an average and standard deviation. We want to compare our value with an accepted value. Our average is not the same as the accepted value, but does it agree with the accepted value within the experimental error? This is the case of **comparison of experimental mean with the true value**

(II) We measure some quantity multiple times by two different methods. Thus, we have two average values and two standard deviations. Do our results agree with each other within experimental error? This the **case of comparison of means of two methods or samples, or analyst etc.,**

(III) We have a number of samples designated A, B, C, and D so on (up to n). We measure each sample once with method 1 and once with method 2 and obtained different values. Do our results agree with each other within experimental error? This is **the case of comparing individual differences using paired t-test.**

Comparing a measured result with a 'known' value

❑ The comparison of $\bar{X}$ and μ, can be done as follows.

If $\bar{X} - \mu > t\frac{s}{\sqrt{N}}$, then $\bar{X}$ is judged **significantly different from μ,** at a predetermined confidence level

On the other hand,

If $\bar{X} - \mu < t\frac{s}{\sqrt{N}}$, then $\bar{X}$ is judged **not significantly different from μ,** at a predetermined confidence level.

- The comparison of $\bar{X}$ and μ, can also be done as follows. First compute test statistic, which is given by

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

- This experimental t, designated texp, is then compared with the theoretical value of t, which is computed statistically for a predetermined confidence level and can be read from t-tabel.
- If $t_{exp} > t_{theor}$, there **is significant difference** between $\bar{X}$ and μ. If $t_{exp} < t_{theor}$, the difference between $\bar{X}$ and μ is insignificant.

Example-1

- Suppose we purchase a coal sample certified by NIST to contain **3.19%(w/w) sulfur** and want to test a new analytical method to determine if it can reproduce the value provided by NIST. The measured values are **3.29, 3.22, 3.30, and 3.23 %(w/w) sulfur**, giving a mean of **3.26** and a standard deviation of **0.041**. Does our method agree with NIST?

- We have 4 measurements, giving us 3 degrees of freedom. From the t-table, t value at 95% CL= 3.182 for 3 df. Thus, the 95% confidence interval is

$$= \bar{x} \pm \frac{ts}{\sqrt{N}} = 3.260 \pm \frac{3.182\ x\ 0.041}{2}$$

$= 3.260 \pm 0.065\%$ S = 3.195 – 3.325% S

- Thus, the NIST value of 3.19% is outside our confidence interval, so with 95% confidence we can say that our method does not agree with the NIST value

Example-2

The absorbance scale of a spectrometer is tested at a particular wavelength with a standard solution which has an absorbance given as 0.470. Ten measurements of the absorbance with the spectrometer give $\bar{x} = 0.461$, and $s = 0.003$. Find the 95% confidence interval for the mean absorbance as measured by the spectrometer, and hence decide whether a systematic error is present.

The 95% confidence limits for the absorbance as measured by the spectrometer are [equation (2.9)]:

$$\bar{x} \pm t_{n-1}s/\sqrt{n} = 0.461 \pm 2.26 \times 0.003/\sqrt{10} = 0.461 \pm 0.002$$

(The value of t_9 was obtained from Table A.2.)

Since the confidence interval does not include the known absorbance of 0.470, it is likely that a systematic error has occurred.

Example -3

A method for the determination of mercury by atomic absorption spectrometry gave values of 400, 385 and 382 ppm for a standard known to contain 400 ppm. Does the mean value differ significantly from the true value, or is there any evidence of systematic error (bias)?

$$\bar{x} = 389 \text{ ppm} \qquad s = 9.64 \text{ ppm} \qquad \mu = 400 \text{ ppm}$$

Using equation (4) to evaluate t_{exptl}

$$t_{exptl} = \frac{(\bar{x} - \mu)}{s} \times N^{1/2} = \frac{(389 - 400)}{9.64} \times 3^{1/2} = \mathbf{1.98}$$

For 2 degrees of freedom, the two-tailed t_{tab} value is 4.30 at the 95% probability level. As t_{exptl} is **less** than the two-tailed value of t_{tab}, the mean is not significantly different from the true value. There is, therefore, no evidence of a systematic error, or bias.

Example-4

In a new method for determining selenourea in water, the following values were obtained for tap water samples spiked with 50 ng ml^{-1} of selenourea:

50.4, 50.7, 49.1, 49.0, 51.1 ng ml^{-1}

Is there any evidence of systematic error?

The mean of these values is 50.06 and the standard deviation is 0.956. Adopting the null hypothesis that there is no systematic error, i.e. $\mu = 50$, and using equation (3.1) gives

$$t = \frac{(50.06 - 50)\sqrt{5}}{0.956} = 0.14$$

From Table A.2, the critical value is $t_4 = 2.78$ ($P = 0.05$). Since the observed value of $|t|$ is less than the critical value the null hypothesis is retained: there is no evidence of systematic error. Note again that this does not mean that there are no systematic errors, only that they have not been demonstrated.

Practice exercise

In an analysis to determine the ethanol content of a wine by gas chromatography, an internal standard of isopropanol is used to account for the variability in the volume injected between tests. In the measurement of a four-point calibration curve and the repeated analysis of the wine sample, six injections in all are performed. Each injection contained 1% v/v of the internal standard. The isopropanol peak area, in arbitrary units, for each of the six injections were 2957398, 3733127, 2900811, 3010190, 2810196, 2084063.

Calculate the injection precision (i.e., the standard deviation of the measurements).

Calculate the mean and 95% confidence limits for the data

How many injections are required to reduce the relative error of the mean to 5% with 95% confidence?

Comparison of two experimental means

- Another way in which the results of a new analytical method may be tested is by comparing them with those obtained by using a second (perhaps a reference) method.
- In this case we have two sample means x_1 and x_2. . Taking the null hypothesis that the two methods give the same result, that is H_0: $\mu_1 = \mu_2$, we need to test whether $(\bar{x}_1 - \bar{x}_2)$ differs significantly from zero.

Comparing the means of two samples whose population standard deviations are equal

In order to decide whether the difference between two sample means $\bar{x}_1$ and $\bar{x}_2$ is significant, that is to test the null hypothesis, H_0: $\mu_1 = \mu_2$, the statistic t is calculated:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (3.2)$$

where s is calculated from:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \qquad (3.3)$$

and t has $n_1 + n_2 - 2$ degrees of freedom.

This method assumes that the samples are drawn from populations with equal standard deviations.

For example

In a comparison of two methods for the determination of chromium in rye grass, the following results (mg kg^{-1} Cr) were obtained:

Method 1: mean = 1.48; standard deviation 0.28
Method 2: mean = 2.33; standard deviation 0.31

For each method five determinations were made.
(Sahuquillo, A., Rubio, R. and Rauret, G. 1999. *Analyst* 124: 1)

Do these two methods give results having means which differ significantly?

The null hypothesis adopted is that the means of the results given by the two methods are equal. From equation (3.3), the pooled value of the standard deviation is given by:

$$s^2 = ([4 \times 0.28^2]) + [4 \times 0.31]^2)/(5 + 5 - 2) = 0.0873$$

$$s = 0.295$$

From equation (3.2):

$$t = \frac{2.33 - 1.48}{0.295\sqrt{\frac{1}{5} + \frac{1}{5}}} = 4.56$$

There are 8 degrees of freedom, so (Table A.2) the critical value $t_8 = 2.31$ ($P = 0.05$): since the experimental value of $|t|$ is greater than this the difference between the two results is significant at the 5% level and the null hypothesis is rejected. In fact since the critical value of t for $P = 0.01$ is about 3.36, the difference is significant at the 1% level. In other words, if the null hypothesis is true the probability of such a large difference arising by chance is less than 1 in 100.

Practice exercise

In a series of experiments on the determination of tin in foodstuffs, samples were boiled with hydrochloric acid under reflux for different times. Some of the results are shown below:

Refluxing time (min)	Tin found (mg kg^{-1})
30	55, 57, 59, 56, 56, 59
75	57, 55, 58, 59, 59, 59

(Analytical Methods Committee. 1983. *Analyst* 108: 109)

Does the mean amount of tin found differ significantly for the two boiling times?

The mean and variance (square of the standard deviation) for the two times are:

30 min $\bar{x}_1 = 57.00$ $s_1^2 = 2.80$

75 min $\bar{x}_2 = 57.83$ $s_2^2 = 2.57$

The null hypothesis is adopted that boiling has no effect on the amount of tin found. By equation (3.3), the pooled value for the variance is given by:

$$s^2 = (5 \times 2.80 + 5 \times 2.57)/10 = 2.685$$

$$s = 1.64$$

From equation (3.2):

$$t = \frac{57.00 - 57.83}{1.64\sqrt{\frac{1}{6} + \frac{1}{6}}}$$

$$= -0.88$$

There are 10 degrees of freedom so the critical value is $t_{10} = 2.23$ ($P = 0.05$). The observed value of $|t|$ (= 0.88) is less than the critical value so the null hypothesis is retained: there is no evidence that the length of boiling time affects the recovery rate.

Example-2

Two methods for the determination of polyaromatic hydrocarbons in soils were compared by analyzing a standard with the following results:

No. of determinations by each method:	10	
No. of degrees of freedom:	18	
UV spectrophotometry:	$\bar{x} = 28.00$ mg kg^{-1}	$s = 0.30$ mg kg^{-1}
Fluorimetry:	$\bar{x} = 26.25$ mg kg^{-1}	$s = 0.23$ mg kg^{-1}

Do the mean results for the two methods differ significantly?

Equation (2) is first used to calculate a pooled standard deviation:

$$s_{pooled} = \left\{ \left[(N-1)s_A^2 + (M-1)s_B^2 \right] / \left[N + M - 2 \right] \right\}^{1/2} = \{(9 \times 0.3^2 + 9 \times 0.23^2)/18\}^{1/2}$$

$$s_{pooled} = 0.267 \text{ mg kg}^{-1}$$

Then equation (1) is used to evaluate t_{exptl}

$$t_{exptl} = \frac{(\bar{x}_A - \bar{x}_B)}{s_{pooled}} \times \left(\frac{NM}{N+M} \right)^{1/2} = \{(28.0 - 26.25)/0.267\} \times 5^{1/2} = \mathbf{14.7}$$

For 18 degrees of freedom, the two-tailed value of t_{tab} at the 95% probability level is 2.10, and at the 99% level it is 2.88.

Since tcalc(14.7) > ttable(2.10), we reject the null hypothesis at the 95% confidence level and conclude that there is significant difference in the mean results of the two methods

Practice exercise-1

An analytical laboratory analyses the glucose levels in soft drinks using a spectroscopic enzyme assay and is considering using an enzyme electrode instead. To ascertain whether the spectroscopic assay and the enzyme electrode give means that are not significantly different, a soft drink was analyzed six times by the same analyst using each method. The concentration of glucose (units: mM) determined for the 12 test portions were:

- Using the spectroscopic assay: 1.90, 1.82, 1.70, 1.94, 1.85, 1.90
- Using the enzyme electrode: 1.35, 1.65, 1.76, 1.41, 1.80, 1.33

Practice exercise-2

A new flame atomic-absorption spectroscopic method of determining antimony in the atmosphere was compared with the recommended calorimetric method. For samples from an urban atmosphere the following results were obtained:

Sample no.	Antimony found (mg m^{-3})	
	New method	Standard method
1	22.2	25.0
2	19.2	19.5
3	15.7	16.6
4	20.4	21.3
5	19.6	20.7
6	15.7	16.8

(Castillo, J. R., Lanaja, J., Marinez, M. C. and Aznárez, J. 1982. *Analyst* 107: 1488)

Do the results obtained by the two methods differ significantly?

Comparing the means of two samples whose population standard deviations are unequal

- If the population standard deviations are unlikely to be equal then it is no longer appropriate to pool sample standard deviations in order to give an overall estimate of standard deviation. An approximate method in these circumstances is given below:

In order to test H_0: $\mu_1 = \mu_2$ when it cannot be assumed that the two samples come from populations with equal standard deviations, the statistic t is calculated, where

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (3.4)$$

with

$$\text{degrees of freedom} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_1}\right)^2}{\left(\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}\right)} \qquad (3.5)$$

with the value obtained being truncated to an integer.

For example

The data below give the concentration of thiol (mM) in the blood lysate of the blood of two groups of volunteers, the first group being 'normal' and the second suffering from rheumatoid arthritis:

Normal: 1.84, 1.92, 1.94, 1.92, 1.85, 1.91, 2.07
Rheumatoid: 2.81, 4.06, 3.62, 3.27, 3.27, 3.76

The null hypothesis adopted is that the mean concentration of thiol is the same for the two groups.

The reader can check that:

$n_1 = 7 \quad \bar{x}_1 = 1.921 \quad s_1 = 0.076$

$n_2 = 6 \quad \bar{x}_2 = 3.465 \quad s_2 = 0.440$

Substitution in equation (3.4) gives $t = -8.48$ and substitution in equation (3.5) gives 5.3, which is truncated to 5. The critical value is $t_5 = 4.03$ ($P = 0.01$) so the null hypothesis is rejected: there is sufficient evidence to say that the mean concentration of thiol differs between the groups.

Practice exercise-1

Suppose that you are to examine both flame atomic absorption(FAA) and graphite furnace atomic absorption(GAA) methods in order to quantify a metal in drinking water in the neighborhood of 5 ppm in order to decide which techniques is more suitable. An NBS standard (5ppm) is tested by both methods and the data are summarized in Table 2.2.

Table 2.2 analysis of a 5 ppm standard by FAA and GAA

sample	FAA	GAA
1	4.99	5.01
2	4.99	4.95
3	5	5.03
4	5.01	4.96
5	5	4.99
6	5	5.02
Mean	5	4.99
Variance	0.0000567	0.00107

Is the two methods are significantly different or are the two means significantly different?

Paired t test for comparing individual differences

- Suppose we have a number of samples designated A, B, C so on (up to n). We measure **each sample once with method 1 and once with method 2 and obtained** different values. Do our results agree with each other within experimental error?
- In this case, each **sample has been analyzed a single time with each method** (no measurement has been duplicated)

Comparison of two methods for measuring A			
Sample number	Method-1	Method-2	Difference (di)
1	17.2	14.2	-3
2	23.1	27.9	4.8
3	28.5	21.2	-7.3
4	15.3	15.9	0.6
5	23.1	32.1	9
6	32.5	22	-10.5
7	39.5	37	-2.5
8	38.7	41.5	2.8
9	52.5	42.6	-9.9
10	42.6	42.8	0.2
11	52.7	41.1	-11.6

- To determine whether there is a statistically significant difference between the two methods, we use a paired t test

- We first calculate the difference, d_i, between the two methods for each sample
- We then calculate the mean and the standard deviation for the differences
- The standard deviation for the differences is calculated by

  $s_d = \sqrt{\Sigma(di - d)2|(n-1)}$, where n is the number of pairs

  $s_d = 6.7$
- We can then calculate the t_{calc} using

$$t_{calc} = \frac{|\bar{x}_d|\sqrt{n}}{s_d} = \frac{2.4\sqrt{11}}{6.7} = 1.2$$

- The value for t-table at 95% confidence for 10 df is 2.228. Since **tcalc < t-table, the null hypothesis is retained** and there is no evidence that the two methods are significantly different. There is greater than 5% chance that the difference observed between the two methods is due to random error

A new high performance liquid chromatographic method for the determination of pseudoephedrine in a pharmaceutical product at two different levels was compared with an established method with the following results:

Pseudoephedrine per dose (mg)	
Method 1	Method 2
59.9	58.6
59.3	58.3
60.4	60.5
30.7	29.4
30.2	30.4
30.1	28.9

Do the means of the two methods differ significantly?

Because the two levels of pseudoephedrine differ considerably, equation (3) for a paired t-test is used to calculate t_{expt}. The differences between the pairs of values are 1.3, 1.0, –0.1, 1.3, –0.2 and 1.2 mg per dose, and the estimated standard deviation of the differences from their mean of 0.750 mg per dose is 0.706 mg per dose. Substitution of these values into the equation gives

$$t_{expt} = \frac{\bar{x}_d}{s_d} \times N^{1/2} = (0.750/0.706) \times 6^{1/2} = \mathbf{2.60}$$

For 5 degrees of freedom, the two-tailed value of t_{tab} at the 95% probability level is 2.57. As t_{expt} is **greater** than t_{tab}, there is a significant difference between the means of the two methods. (Note: using equation (1) would give a t_{expt} value of 0.08 and an incorrect conclusion.)

Practice exercise

The amount of calcium in different samples of milk powder (in mg of calcium per g of milk powder) were analyzed by two methods, one employing extraction followed by analysis using atomic absorption spectroscopy, the other using a complexometric titration method. The results of nine analyses are shown in table 3.4.

Table 3.4 Replicate results for the measurement of calcium in milk by two methods

Test material	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
AAS[a] ($mg\,g^{-1}$)	3.01	2.58	2.52	1.00	1.81	2.83	2.13	5.14	3.20
CT[b] ($mg\,g^{-1}$)	2.81	3.20	3.20	3.20	3.35	3.86	3.88	4.13	4.86

[a] Atomic absorption spectroscopy.
[b] Complexometric titration.

Determine whether the two analytical methods give equivalent analytical results.

F-test for the comparison of standard deviations

- The significance tests described so far are used for **comparing means, and hence for detecting systematic errors.** In many cases it is also important to **compare the standard deviations, i.e. the random errors of two** sets of data
- As with tests on means, this comparison can take two forms. Either we may wish to test whether Method A is more precise than Method B (i.e. a one-sided test) or we may wish to test whether Methods A and B differ in their precision (i.e. a two-sided test)

- The F-test considers the ratio of the two sample variances, i.e. the ratio of the squares of the standard deviations, $s1^2/s2^2$.

In order to test whether the difference between two sample variances is significant, that is to test H_0: $\sigma_1^2 = \sigma_2^2$, the statistic F is calculated:

$$F = s_1^2/s_2^2 \qquad (3.7)$$

where 1 and 2 are allocated in the equation so that F is always ≥1.

The numbers of degrees of freedom of the numerator and denominator are $n_1 - 1$ and $n_2 - 1$ respectively.

The test assumes that the populations from which the samples are taken are normal.

- If the null hypothesis is true then the variance ratio should be close to 1. Differences from 1 can occur because of random variation, but if the difference is too great it can no longer be attributed to this cause.

- If the calculated value of **F exceeds** a certain critical value (obtained from tables) then **the null hypothesis is rejected.** This critical value of F depends on the size of both samples, the significance level and the type of test performed.

A proposed method for the determination of the chemical oxygen demand of wastewater was compared with the standard (mercury salt) method. The following results were obtained for a sewage effluent sample:

	Mean (mg l^{-1})	Standard deviation (mg l^{-1})
Standard method	72	3.31
Proposed method	72	1.51

For each method eight determinations were made.
(Ballinger, D., Lloyd, A. and Morrish, A. 1982. *Analyst* 107: 1047)

Is the precision of the proposed method significantly greater than that of the standard method?

We have to decide whether the variance of the standard method is significantly greater than that of the proposed method. F is given by the ratio of the variances:

$$F = \frac{3.31^2}{1.51^2} = 4.8$$

This is a case where a one-sided test must be used, the only point of interest being whether the proposed method is more precise than the standard method. In Table A.3 the number of degrees of freedom of the denominator is given in the left-hand column and the number of degrees of freedom of the numerator at the top. Both samples contain eight values so the number of degrees of freedom in each case is 7. The critical value is $F_{7,7} = 3.787$ ($P = 0.05$), where the subscripts indicate the degrees of freedom of the numerator and denominator respectively. Since the calculated value of F (4.8) exceeds this, the variance of the standard method is significantly greater than that of the proposed method at the 5% probability level, i.e. the proposed method is more precise.

Practice exercise

Are the two standard deviations of table 2.2 data in practice exercise-1 significantly different from each other?

Example 3

A proposed new method for the determination of sulfate in an industrial waste effluent is compared with an existing method, giving the following results:

Method	Mean/g dm^{-3}	No. of replicates	No. of degrees of freedom	s/mg dm^{-3}
Existing	72	8	7	3.38
New	72	8	7	1.50

Is there a significant difference between the precisions of the two methods?

$$F_{expt} = s^2_{existing} \Big/ s^2_{new} = \frac{(3.38)^2}{(1.50)^2} = 5.08$$

The two-tailed tabular value for F with 7 degrees of freedom for both the numerator and the denominator is

$$F_{7,7} = 5.00 \text{ at the } 95\% \text{ probability level}$$

As F_{expt} is **greater** than F_{tab}, the null hypothesis is rejected; the two methods are giving significantly different precisions.

Grubbs' test for an outlier

- Every experimentalist is familiar with the situation in which one (or possibly more) of a set of results appears to differ unreasonably from the others in the set. Such a measurement is called an outlier. In some cases an outlier may be attributed to a human error
- The ISO recommended test for outliers is Grubbs' test. This test compares the **deviation of the suspect value from the sample mean** with the standard deviation of the sample. And then If the **calculated value of G exceeds the critical** value, the **suspect value is rejected.** The suspect value is the value that is furthest away from the mean

In order to use Grubbs' test for an outlier, that is to test H_0 : all measurements come from the same population, the statistic G is calculated:

$$G = |\text{suspect value} - \bar{x}|/s \quad (3.8)$$

where $\bar{x}$ and s are calculated with the suspect value *included*.

The test assumes that the population is normal.

For example,

The following values were obtained for the nitrite concentration (mg l^{-1}) in a sample of river water:

0.403, 0.410, 0.401, 0.380

The last measurement is suspect: should it be rejected?

The four values have $\bar{x} = 0.3985$ and $s = 0.01292$, giving

$$G = |0.380 - 0.3985|/0.01292 = 1.432$$

From Table A.5, for sample size 4, the critical value of G is 1.481 ($P = 0.05$). Since the calculated value of G does not exceed 1.481, the suspect measurement should be retained.

If three further measurements were added to those given in the example above so that the complete results became:

0.403, 0.410, 0.401, 0.380, 0.400, 0.413, 0.408

should 0.380 still be retained?

The seven values have $\bar{x} = 0.4021$ and $s = 0.01088$. The calculated value of G is now

$$G = |0.380 - 0.402|/0.01088 = 2.031$$

The critical value of G ($P = 0.05$) for a sample size 7 is 2.020, so the suspect measurement is now rejected at the 5% significance level.

Practice exercise

Decide whether the value 216 should be rejected from the set of results 192, 216, 202, 195, and 204.

Dixon's test (sometimes called the Q-test)

- It is another test for outliers which is popular because the calculation is simple. This test assesses a suspect measurement by **comparing the difference between it and the measurement nearest to** it in size after he data have been arranged in order of increasing value.
- If the calculated value of Q exceeds the critical value, the suspect value is rejected

In order to use Dixon's test for an outlier, that is to test H_0 : all measurements come from the same population, the statistic Q is calculated:

$$Q = |\text{suspect value} - \text{nearest value}|/(\text{largest value} - \text{smallest value}) \quad (3.9)$$

This test is valid for samples size 3 to 7 and assumes that the population is normal.

Apply Dixon's test to the data from the previous example.

$$Q = |0.380 - 0.400|/(0.413 - 0.380) = 0.606$$

The critical value of Q ($P = 0.05$) for a sample size 7 is 0.570. The suspect value 0.380 is rejected (as it was using Grubbs' test).

Practice exercise

Using the Q test, decide whether the value 216 should be rejected from the set of results 192, 216, 202, 195, and 204.

Thank you