Sampling:

A chemical analysis uses only a small fraction of the available sample, the process of sampling is a very important operation. Knowing how much sample to collect and how to further subdivide the collected sample to obtain a laboratory sample is vital in the analytical process. Statistical methods are used to aid in the selection of a representative sample. The analytical sample must be processed in a dependable manner that maintains sample integrity without losing sample or introducing contaminants.

Analytical Samples and Methods:

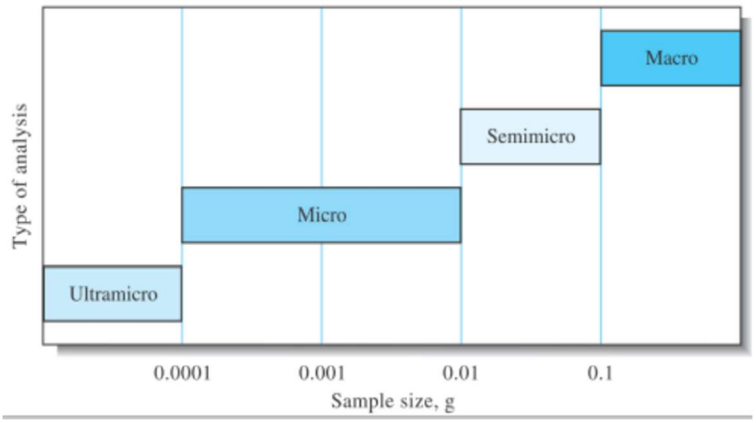
Types of Samples and Methods:

Quantitative methods are traditionally classified as

- gravimetric methods,
- volumetric methods,
- and instrumental methods.

Other methods are based on the size of the sample and the level of the constituents.

Sample Size Techniques for handling very small samples are quite different from those for treating macro samples.



Sample Size	Type of Analysis
> 0.1 g	Macro
0.01 to 0.1 g	Semimicro
0.0001 to 0.01 g	Micro
$< 10^{-4}$ g	Ultramicro

Evaluating Analytical Data: When we use an analytical method, we make three separate evaluations of experimental error. First, before we begin the analysis, we evaluate potential sources of errors to ensure they will not adversely affect our results. Second, during the analysis we monitor our measurements to ensure that errors remain acceptable. Finally, at the end of the analysis we evaluate the quality of the measurements and results, and compare them to our original design criteria. This chapter provides an introduction to sources of error, to evaluating errors in analytical measurements, and to the statistical analysis of data.

i. **Characterizing Measurements and Results:** One way to characterize data from multiple measurements/runs is to assume that the measurements are randomly *scattered around a central value that provides the best estimate of expected*, or "true" value. We describe the distribution of these results by reporting its central tendency and its spread.

ii. **Characterizing Experimental Errors:** Two essential questions arise from any set of data. First, does our measure of central tendency agree with the expected result? Second, why is there so much variability in the individual results? *The first of these questions addresses the accuracy of our measurements* and *the second addresses the precision of our measurements.* In this section we consider the types of experimental errors that affect accuracy and precision.

iii. **Propagation of Uncertainty:** A propagation of uncertainty allows us to estimate the uncertainty in a result from the uncertainties in the measurements used to calculate the result.

iv. **The distribution of Measurements and Results:** *To compare two samples to each other, we need more than measures of their central tendencies and their spreads based on a small number of measurements*. We need also to know how to predict the properties of the broader population from which the samples were drawn; in turn, this requires that we understand the distribution of samples within a population.

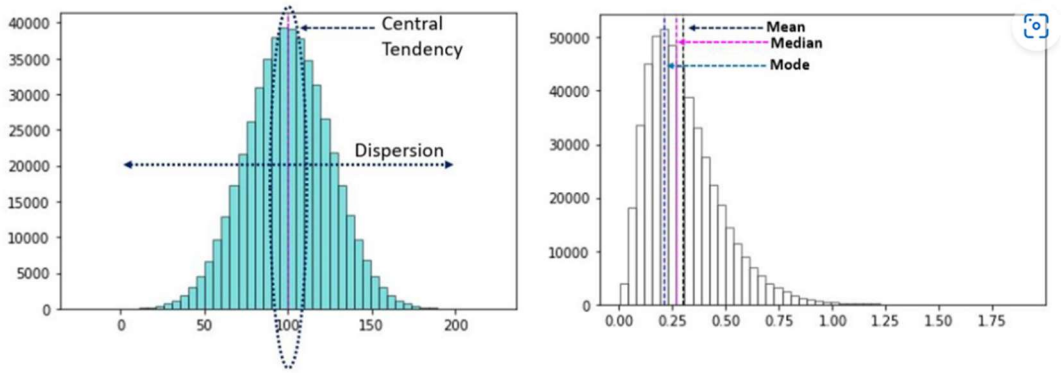
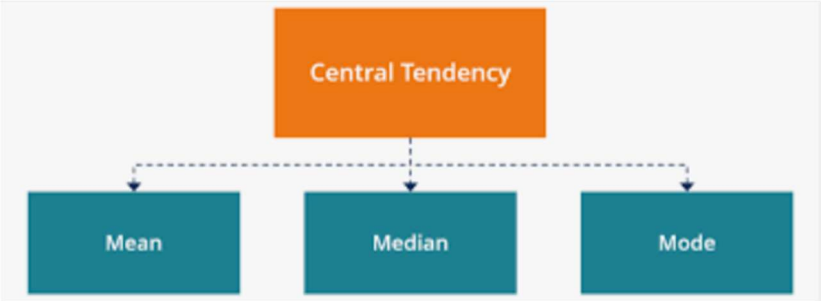


Fig 1: Central Tendency & Dispersion of plotted data in images



v. **Statistical Analysis of Data:** *A confidence interval is a useful way to report the result of an analysis* because it sets limits on the expected result. In the absence of determinate error, a confidence interval based on a sample's mean indicates the range of values in which we expect to find the population's mean. In this section we introduce a general approach to the statistical analysis of data. Specific statistical tests are presented in the next section.

vi. **Statistical Methods for Normal Distributions:** *The most common distribution for our results is a normal distribution*. Because the area between any two limits of a normal distribution curve is well defined, constructing and evaluating significance tests is straightforward.

vii. The International Union of Pure and Applied Chemistry (IUPAC) defines a method's detection limit as the smallest concentration or absolute amount of analyte that has a signal significantly larger than the signal from a suitable blank.

Error:

- Measurements invariably involve errors and uncertainties.
- it is impossible to perform a chemical analysis that is totally free of errors or uncertainties
- We can only hope to minimize errors and estimate their size with acceptable accuracy.
- Errors are caused by faulty calibrations or standardizations or by random variations and uncertainties in results.
- Frequent calibrations, standardizations, and analyses of known samples can sometimes be used to lessen all but the random errors and uncertainties.
  The term error has two slightly different meanings. 1) error refers to the difference between a measured value and the "true" or "known" value. 2) error often denotes the estimated uncertainty in a measurement or experiment. "We can only hope to minimize errors and estimate their size with acceptable accuracy

Individual results from a set of measurements are seldom the same

- Usually, the "best" estimate is considered to be the central value for the set.
- The central value of a set should be more reliable than any of the individual results.
- *Usually, the mean or the median is used as the central value for a set of replicate measurements*. An analysis of the variation in the data allows us to estimate the uncertainty associated with the central value.

Absolute Errors:

The absolute error of a measurement is a numerical difference between the measured value and the true value. The absolute error d is defined as the difference between the measured value (i.e., analytical result) x and the true or accepted value μ

$$d = \mu - x$$

Absolute error can be positive if $\mu > x$ and negative if $\mu < x$. Sometimes absolute error is express by absolute value and the observed result.

Relative Errors:

The criteria of the accuracy of the result, is however, the relative error (e) which is defined as the ratio of the absolute error and the true value, i.e.,

$$e = \frac{d}{\mu} = \frac{\mu - x}{\mu}$$

Relative error is commonly expressed as a percent or in part per thousand (ppt), preferably in ppt to avoid confusion with absolute error which is expressed in percent.

$$e = \frac{\mu - x}{\mu} \times 100$$

***Chemical analyses are affected by at least two types of errors*:**

1. Systematic (or determinate) error, causes the mean of a data set to differ from the accepted value.
2. Random (or indeterminate) error, causes data to be scattered more or less symmetrically around a mean value.

Systematic or Determinate Errors

- Systematic errors
- have a definite value,
- an assignable cause, and
- are of the same magnitude for replicate measurements made in the same way.
- They lead to bias in measurement results.

There are three types of systematic errors:

- Instrumental errors.
- Method errors
- Personal errors

Instrumental Errors

- are caused by non-ideal instrument behaviour, by faulty calibrations, or by use under inappropriate conditions
- Pipets, burrets, and volumetric flasks may hold or deliver volumes slightly different from those indicated by their graduations.
- Calibration eliminates most systematic errors of this type.
- Electronic instruments can be influenced by noise, temperature, pH and are also subject to systematic errors.
- Errors of these types usually are detectable and correctable.

Method Errors:

- The nonideal chemical or physical behavior of the reagents and reactions on which an analysis is based often introduce systematic method errors.
- Such sources of nonideality include the slowness of some reactions, the incompleteness of others, the instability of some species, the lack of specificity of most reagents, and the possible occurrence of side reactions that interfere with the measurement process.
- Errors inherent in a method are often difficult to detect and hence, these errors are usually the most difficult to identify and correct.

Personal Errors:

Result from the carelessness, inattention, or personal limitations of the experimenter.

- Many measurements require personal judgments.
- Examples include estimating the position of a pointer between two scale divisions, the color of a solution at the end point in a titration, or the level of a liquid with respect to a graduation in a pipet or burret.

- Judgments of this type are often subject to systematic, unidirectional errors.
- A universal source of personal error is prejudice, or bias.
- Number bias is another source of personal error that varies considerably from person to person.
- The most frequent number bias encountered in estimating the position of a needle on a scale involves a preference for the digits 0 and 5.
- Also common is a prejudice favouring small digits over large and even numbers over odd.
- Digital and computer displays on pH meters, laboratory balances, and other electronic instruments eliminate number bias because no judgment is involved in taking a reading.

Random or Indeterminate Errors: These are indeterminate and their occurrence does not obey any definite law and this cannot be avoided (systematic error can be avoided). In contrast to systematic error (which is unidirectional), the indeterminate error leads to both positive and negative error with an equal probability.

$E_{total} = E_{systematic} + E_{random} \approx E_{random}$ when the systematic error is negligible or absent.

Accuracy and Precision: In an ordinary sense, these two words are often used synonymously but in a scientific sense, they are different.

Accuracy describes the correctness of a result. In other words, *it gives the measure of closeness of the result to the accepted or true value.* Obviously, we can conclude: *the smaller the error, the greater the accuracy.*

In fact, in absence of any scientific error, the accuracy is high when the precision is high.

Accuracy

- Accuracy indicates the closeness of the measurement to the true or accepted value and is expressed by the error.
- Accuracy measures agreement between a result and the accepted value.
- Accuracy is often more difficult to determine because the true value is usually unknown. An accepted value must be used instead.
- *Accuracy is expressed in terms of either absolute or relative error.*

Precision

- Precision describes the agreement among several results obtained in the same way. Describes the reproducibility of measurements.
- Precision is readily determined by simply repeating the measurement on replicate samples.
- Precision of a set of replicate data may be expressed as *standard deviation, variance, and coefficient of variation.*
- d_i, deviation from mean, is how much x_i, the individual result, deviates from the mean.

$$d_i = |x_i - \bar{x}|$$

Accuracy vs. precision: Random errors determine the precision and it is measured by parameters like *average derivative, range, standard deviation, variance and coefficient of variation.*

If there is a constant systematic error in an analysis, the results may be precise but inaccurate. In such cases, the systematic error will affect each measurement or observation equally and the precision of the data will not be affected.
Thus, we can say that ***precise values are accurate only when there is no systematic error.***
Here it is important to mention that precision of the results can be easily determined by comparing the replicate data but to determine the accuracy, we need the true value which may not be known always.

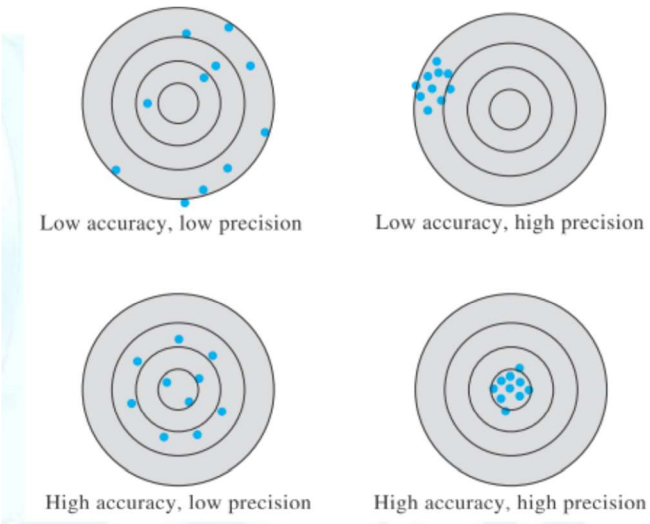


Figure: Note that we can have very precise results (upper right) with a mean that is not accurate and an accurate mean (lower left) with data points that are imprecise.

Methods of expressing Precision:

The Mean and the Median

The *mean*, also called the arithmetic mean or the average, is obtained by dividing the sum of replicate measurements by the number of measurements in the set:

- The symbol Σx_i means to add all of the values xi for the replicates; x_i represents the individual values of x making up the set of N replicate measurements.
- The *median* is the middle value in a set of data that has been arranged in numerical order.
- The median is used advantageously when a set of data contain an outlier. An outlier is a result that differs significantly from others in the set.
- For set of replicate observations, the value which occurs with the ***maximum frequency*** (i.e., highest frequency density) gives the modes.
- An *outlier* can have a significant effect on the mean of the set but has no effect on the median

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

If we consider a series of *n* observations arranged in ascending order magnitude:
$x_1, x_2, x_3, \ldots\ldots x_{n-1}, x_n$.
the arithmetic mean (often called simple the mean) is given by:

$$\bar{x} = \frac{x_1 + x_2 + x_3, \ldots\ldots + x_{n-1} + x_n}{n}$$

The spread of the values is measured most efficiently by the standard deviation defined by:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots\ldots (x_n - \bar{x})^2}{n - 1}}$$

In this equation the denominator is *(n-1) rather than n* when the number values is small.
The equation max also be written as:

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

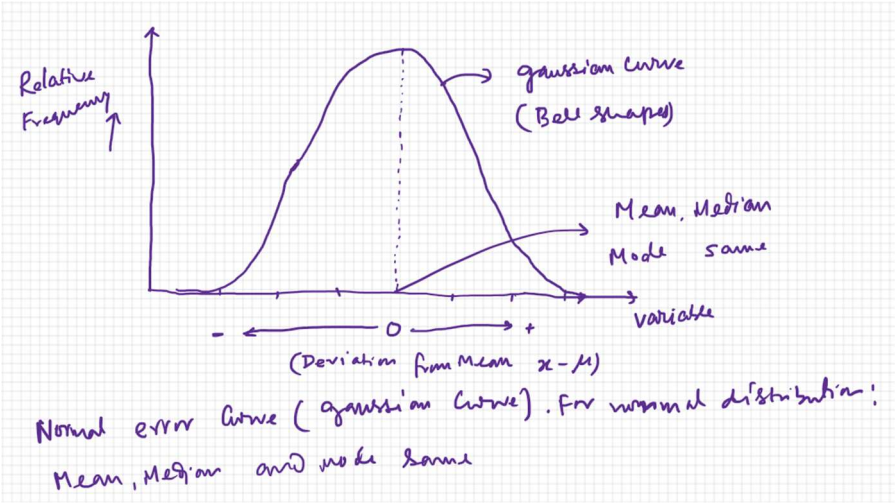
Precision of analytical data may be expressed as. *The square of the standard deviation (s^2) is called the variance*. A further measure of precision, known as the Relative Standard Deviation (R.S.D), is given by

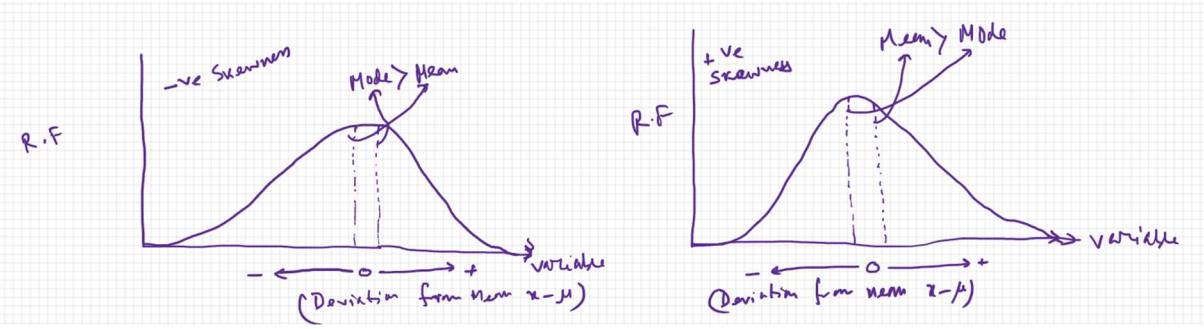
RSD=$\frac{s}{\bar{x}}$

This measure is often expressed as a percentage, known as the **coefficient of variation (C.V.):**

$$C.V. = \frac{s \times 100}{\bar{x}}$$

Spread or Range (R) is used to describe the precision of a set of replicate results. It is the difference between the largest value in the set and the smallest.



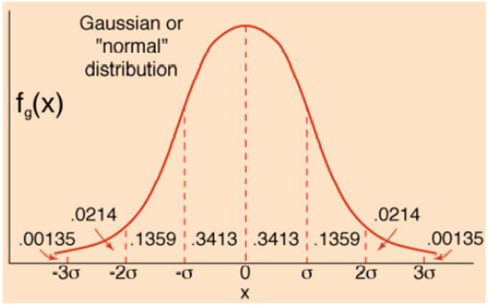


Normal Distribution (i.e., symmetrical distribution): Mean (M) = Median (Me) = Mode (Mo)
Positive Skew Distribution: Mean(M) > Median (Me) > Mode (Mo)
Negative Skew Distribution: Mean(M) < Median (Me) < Mode (Mo)
Skewness (Sk) is measured as $Sk = \frac{Mean - Mode}{Standard\ Deviation} = \frac{M - M_0}{s} = \frac{\bar{x} - M_0}{s}$

Characteristic of the Random or Indeterminate Errors: The values of replicate measurements always differ to some extent even when the measurements are done by the same person with great care under the identical conditions as nearly as possible. The values are found to be distributed within a certain range. ***In other words, due to the accumulated random errors, the values of replicate measurements fluctuate randomly around the mean value***. With respect to the mean value, low and high results (by a certain amount) appear with an equal probability, i.e., ***both the negative and positive deviation from the mean value are symmetrically distributed.*** These symmetric deviations are due to indeterminate or random errors. Because of the symmetrical distribution of the ***replicate data, the random errors tend to cancel one another and the effect of random errors on the mean value is maximum.*** It may be pointed out that in contrast to the indeterminate error, the determinate error is generally unidirectional (i.e., either the positive or negative deviation in a particular method of analysis).

The random error or indeterminate errors can not be corrected in the same way as practised for the determinate errors because the law of variation of indeterminate error is unknown. But it is possible to have the idea regarding the **most probable results** and its **degree of reliability** in terms of statistical treatment.

Normal law of distribution of errors: A normal distribution implies that if you take a large enough number of measurements of the same property for the same sample under the same conditions subject only to random (indeterminate) error, the values will be distributed around the expected value, or mean, and that the frequency with which a particular result. This distribution is sometimes also called a Normal distribution. Although no data follows this ideal mathematical distribution, many kinds of data follow a distribution that is approximately Gaussian.



The mathematical form of the Gaussian distribution is

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

where μ is the mean of the parent population, and σ is the standard deviation. P_G is the probability of a particular value occurring.

~~If we make a measurement, we expect that the measurement will be approximately (but not exactly) equal to the true value. If we make a very large number of measurements, some will be larger and some smaller than the true value. If we took an infinite number of measurements, they would be distributed about the true value. A distribution defines how they will be distributed.~~

Properties of Gaussian Curves: Gaussian curves can be described by an equation that contains two parameters, the population mean μ and the population standard deviation σ. The term parameter refers to quantities such as μ and σ that define a population or distribution. Data values such as x are variables. The term statistic refers to an estimate of a parameter that is made from a sample of data.

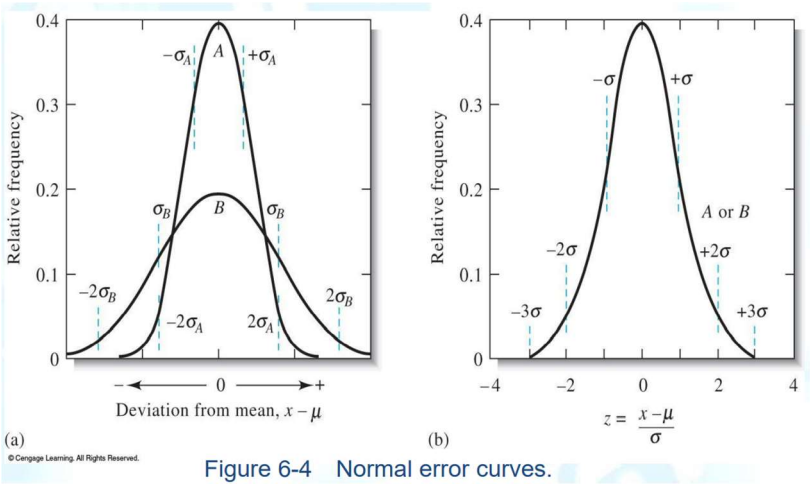


Figure 6-4 Normal error curves.

The equation for a normalized Gaussian curve is as follows:

Statistical Treatment of Random Error:

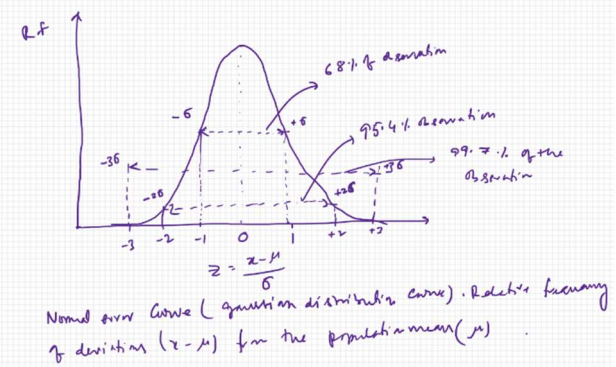
Gaussian Distribution: For data in which the error is truly random, the probability of obtaining a specified value for an individual data point (x_i) is a function of the population mean (μ), and the standard deviation of the analytical method being employed (σ). Equation shows a normal probability distribution function. It presents the symmetrical distribution of data for replicate measurements around the mean value of an infinite set of data.

It is noted that the corresponding curve for a smaller number of observations is called Frequency Polygon which is generally irregular in shape. However, the bell-shaped regular normal curve for an infinite set of data may be well correlated with a mathematical function as follows.

$$Y = A\ exp[-h^2(x-\mu)^2]$$

where x is the value of a particular data point, σ is the standard deviation, μ is the mean of the population and Y is the probability of obtaining a particular value of "x"

Y= frequency of occurrence of a deviation from the population mean μ.



$x - \mu = deviation\ of\ an\ individual\ value\ from\ the\ population\ mean\ \mu.$

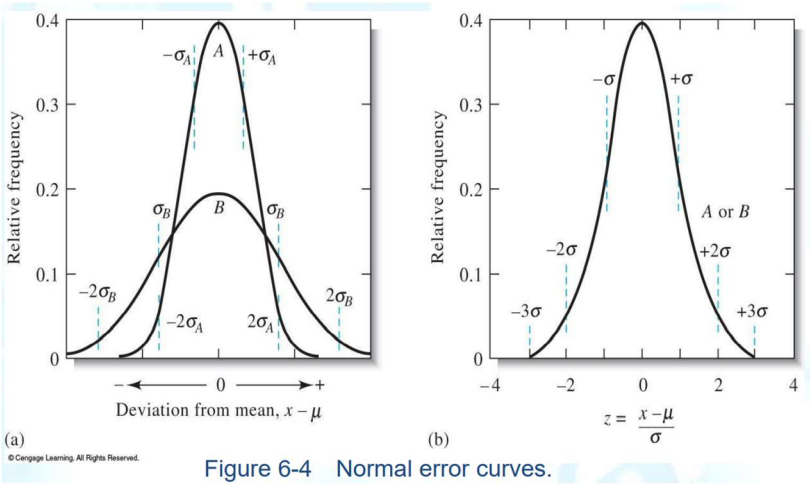


Figure 6-4 Normal error curves.

Normal probability function or Gausssian Distribution function is expressed as

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

The above equation is referred to as a normal probability function (npf) or a Gaussian distribution or colloquially as "a bell curve".

Now, the absolute derivation of the individual values of x from the population mean μ is given by $(x - \mu)$ and it may be measured by the unit of single of variable Z defined by:

$$Z = \frac{x-\mu}{\sigma} \quad i.e., dZ = \frac{dx}{\sigma} \qquad dx = \sigma dZ$$

$$Putting, \quad dZ = \frac{dx}{\sigma}; \quad Ydx = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$= \frac{1}{\sqrt{2\pi}} exp\left[-\frac{Z^2}{2}\right] dZ$$

Now the area under the curve between any two values of $(x - \mu)$ gives the probability of the measure of the fraction of total population of observation having the values between the limits.

$$Area = \int_{-x}^{+x} Ydx = \int_{-x}^{+x} \frac{exp\left[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right]}{\sigma\sqrt{2\pi}} dx = 1$$

For ±σ, $\int_{-\sigma}^{+\sigma} Ydx = \int_{-\sigma}^{+\sigma} \frac{1}{\sqrt{2\pi}} exp - \left[\frac{Z^2}{2}\right] dZ$ of

Therefore, x=μ±2σ $Area = \frac{1}{\sqrt{2\pi}} \int_{-2}^{+2} exp - \left[\frac{Z^2}{2}\right] dZ = 0.954$

x=μ±3σ $Area = \frac{1}{\sqrt{2\pi}} \int_{-3}^{+3} exp - \left[\frac{Z^2}{2}\right] dZ = 0.997$

This indicates that ~68.3% of the total area under the curve is enclosed in the range Z=-1 to Z=+1; i.e., limits of μ±1σ or, -σ to +σ around the population mean. It signifies ~2/3 of the observations for an infinite number of data will lie within the limits μ±1σ.

Significance of s (Root Mean Square Deviation): It gives the **measure of precision.** The smaller value of s indicates that the observed values are clustered closely to $\bar{x}$ i.e., the **higher precision** is reflected. In other words, the magnitude of random error is small for the small values of s.

Significance of σ (Standard Deviation): It is determined by the magnitude of random errors. The *smaller value of σ indicates the higher precision of the results.* The analytical results may be shown as: $\bar{x} \pm \sigma$ which indicates the probability (p) of an observation to lie in the range $\bar{x} - \sigma \; to \; \bar{x} + \sigma$ is 0.683 (68.3%). In other words, the range $\bar{x} \pm \sigma$ covers the 68.3% zone of the total observations. Similarly, significance of other representations of the analytical results can be understood. These are:

$\bar{x} \pm 0.67\sigma \quad p = 0.50 \; i.e., 50\% \; zone$

$\bar{x} \pm 1\sigma \quad p = 0.683 \; i.e., 68.3\% \; zone$

$\bar{x} \pm 2\sigma \quad p = 0.955 \; i.e., 95.5\% \; zone$

$\bar{x} \pm 3\sigma \quad p = 0.997 \; i.e., 99.7\% \; zone$

Assuming $\bar{x} \approx \mu$ for normal error curve

The normal error curve can give another interpretation of random errors in terms of σ (standard deviation). We can say: for a single measurement, the probability of random errors associated with it lie within the range $\pm 1\sigma \; is \; 0.683 \; (i.e., 68.3\%)$; similarly for the random errors within the range $\pm 2\sigma \; and \; \pm 3\sigma$ associated with a single measurement, the corresponding probabilities are 0.955 (i.e., 95.5%) and 0.997 (i.e., 99.7%) respectively.

Thus, the standard deviation (σ) is a useful parameter to measure the random error (i.e., precision) involved in the analytical method.

Confidence Limits: Earlier we learned how to calculate a standard error. Another common statistical tool for reporting the uncertainty (precision) of a measurement is the confidence limit (CL). For example, we might report the percent alcohol in a solution as 13% with a 95% CL of ±2%, where the ±2% represents the CL.

Unless otherwise stated, the reported CL is at the 95% CL and represents the range in which we are 95% certain the "true" answer lies. The reason the 95% CL is the accepted norm is because 95.4% of all data points in a normal distribution is encompassed by a range of approximately ±2σ. It is reported at 95% instead of 95.4% for purposes of simplicity. However, as you will soon see, it is possible to calculate CL values other than the 95% CL.

We define CL using σ. Recall that σ is the standard deviation of the entire population. When we do not know σ we use "s" instead and a fudge factor, which we will describe shortly. If we know the standard deviation for the entire population, then the 95% CL10 is simply 95% CL = ±2σ and we would report the mean as μ ±2σ.

However, we seldom know the mean or the standard deviation of an entire population. All chemical analyses deal with a sampled population. The CL for a sample is given in Equation

$$\textbf{Confidence limit} = \pm t\frac{S}{\sqrt{n}}$$

and we would report the average as

were $\bar{x}$ is the mean of the sample, "s" is the standard deviation of the sample, N is the number of data points in the sample and "t" is a "fudge factor" taken from Table.

Table 22.1: Confidence Limit t-values as a function of (N-1)[12]

N-1	90%	95%	99%	99.5%
2	2.920	4.303	9.925	14.089
3	2.353	3.182	5.841	7.453
4	2.132	2.776	4.604	5.598
5	2.015	2.571	4.032	4.773
6	1.943	2.447	3.707	4.317
7	1.895	2.365	3.500	4.029
8	1.860	2.306	3.355	3.832
9	1.833	2.262	3.205	3.690
10	1.812	2.228	3.169	3.581

Analysing Data Set: Testing for significance & criteria for rejection of an observation:

T-Test: When a particular sample is analysed using two different methods, the mean values obtained may be differ. To view whether, the difference between the mean values significant or not, T-test is a useful one to tell whether the mean values differ significantly or not. It also points out the correctness of the principle of the null hypothesis with a certain confidence level (95%/99%).

$$T = \frac{|\bar{x}_1 - \bar{x}_2|}{s}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

The calculated T value can be compared with the critical T-value for a degree of freedom n_1+n_2-2 at the desired confidence level. If the calculated T-value is smaller than the critical T-value at the desired confidence level, then the null hypothesis is substantiated & there is no significant difference between the mean values of the two sets of data at the desired confidence level.

Relevance and Use of t-Test Formula: It is imperative for a statistician to understand the concept of t-test as it holds significant importance while drawing conclusive evidence about whether or not two data sets have statistics that are not very different. This test is run to check the validity of a null hypothesis based on the critical value at a given confidence interval and degree of freedom. However, please note that the student's t-test is applicable for data set with a sample size of less than 30.

F-Test: This statistical approach points out whether there is any significant difference between S_1 & S_2 which are the sample standard deviation (S.D) of two sets of data.

$$F = \frac{S_1^2}{S_2^2} \quad (S_1 > S_2 \text{ to make } F > 1)$$

If calculated $F > F_{critical}$ $or, F_{tabulated}$ at the desired confidence level $S_1 \& S_2$ differ significantly.

$$if\ F_{calc} < F_{tabulated}\ \ S_1, S_2\ do\ not\ differ\ significantly$$

Q-Test: This statistical approach is used to reject an outlier known as rejection quotient. The Q parameter is calculated as follows:

$$Q_{obs}\ or\ Q_{calc} = \frac{|x - \bar{x}|}{R} = \frac{|x_s - x_n|}{x_{max} - x_{min}}$$

$$where\ Xs = suspected\ results\ and\ Xn = nearest\ neighbours\ Xs,\ \ R = Range$$

The calculated value of Q is to be compared with the critical value of Q at a desired confidence level.
If $Q_{calc} > Q_{crit}$, then the suspected results is to be rejected at that certain CL.

***Ex1*:** A sample of soda ash (Na_2CO_3) was analysed and the % of Na_2CO_3 was found for measurements: 40.55, 40.58, 40.50, 40.62, 40.70. 40.70 is the suspected value. (For n=5, Q_{crit} =0.64 at 90% CL, $Q_{crit} = 0.71\ at\ 95\%\ CL$.

$$\bar{x}\ for\ the\ four\ measurements = 40.56\%$$

Deviation of suspected results from the mean, $d = |40.56 - 40.70|\%$=0.14%

$$\therefore\ Q_{calc} = \frac{40.70 - 40.62}{40.70 - 40.50} = \frac{0.08}{0.20} = 0.4$$

$$\therefore Q_{calc} = 0.4\ < Q_{crit} = \frac{0.64}{0.71} at\ 90\%/95\%\ CL$$

So, the suspected results is not to be rejected at 90% & $95\%\ CL$

Identifying Outliers: Q-Tests

Although the International Organization of Standardization (ISO) now recommends that we use the Grubb's test for identifying outliers, the Q-test still remains a very commonly used method and we introduce it here because you are likely to encounter it in your careers. We will examine the Grubb's test in the next section.

Sometimes you obtain a set of replicate data and there is one (or more) data point that just "seems wrong". For example, Table 22.3 shows the results for the N = 10 replicate analysis of caffeine in tea. The data points tend to cluster around 80 ppm with the exception of Cup #5 which had a lower reading of 72 ppm. The sloppy analyst might be tempted to throw out Cup #5's data based solely on intuition; however it is quite possible that 72 ppm falls within the 95% confidence interval for this distribution of points. It is unethical to simply ignore data that you dislike. You should include all data in a report, even outliers, and if you decide to reject a point in your final analysis, you must have a statistical justification for that decision. A ***Q-test*** is a statistical tool used to identify an outlier within a data set[14]. To perform a Q-test you must first arrange your data in a progressive order (low-to-high or high-to-low) and then using Equation 22.10, you calculate an experimental Q-value (Q_{exp}). If Q_{exp} is greater than the critical Q-value (Q_{crit}) found in Table 22.4 then you are statistically justified in removing your suspected outlier from further consideration.[15] You then recalculate the mean, standard deviation and the 95% CL with the outlier removed from the calculations.

Table 22.3

Cup	ppm Caf
1	78
2	82
3	81
4	77
5	72
6	79
7	82
8	81
9	78
10	83
Avg	**79.3**
StDev	**3.3**
95% C.L.	**2.0**

Table 22.4: Critical Rejection Values for Identifying an Outlier: Q-test

	Q_{crit}		
N	90% CL	95% CL	99% CL
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568

$$Q_{exp} = \frac{|x_q - x_{n+1}|}{w} \qquad \textbf{Eq. 22.10}$$

X_q = suspected outlier
X_{n+1} = next nearest data point
w = range (largest – smallest data point in the set)

Example 22.4- Perform a Q-test on the data set from Table 22.3 and determine if you can statistically designate data point #5 as an outlier within a 95% CL If so, recalculate the mean, standard deviation and the 95% CL

Strategy – Organize the data from highest to lowest data point and use Equation 22.10 to calculate Q_{exp}.

Solution – Ordering the data from Table 22.3 from highest to lowest results in

									$x_{(n+1)}$	x_q
Cup	10	7	2	8	3	6	9	1	4	5
ppm caf	83	82	82	81	81	79	78	78	77	72
Range =	83-72 =	11								

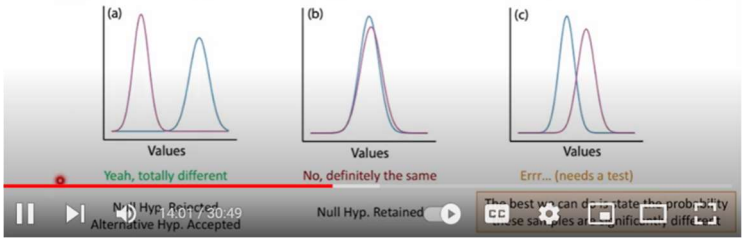
Substitution into Equation 22.10 yields

$$Q_{exp} = \frac{|x_q - x_{n+1}|}{w} = \frac{|72 - 77|}{11} = 0.455$$

Using the Q_{crit} table, we see that Q_{crit}=0.466. Since Q_{exp}<Q_{crit}, you must keep the data point.

Hypotheses

- The *t*-Test for comparing means, and the *F*-test for standard deviations, are framed in terms of the **null hypothesis**
  - **Null hypothesis:** mean values or standard deviation values (depending on what you are testing) are the same within error
  - **Alternative hypothesis**: means or standard deviations are 'significantly' different



Statistical Tests

Picture in picture

- *F* test: Are two **standard deviation** values 'significantly different' from each other when experimental uncertainty is considered? (Section 4-2)
- Student's *t* and the *t* test: (Section 4-3 to 4-4)
  - Are two **mean** values statistically significantly different from each other?
  - Over what range (centered around the measured average $\bar{x}$) is there a 95% probability of 'finding' the true population mean μ?
- Grubbs test for an outlier: Is it fair to call this data point an outlier? (Section 4-5)

Table 4-4 Values of Student's *t*

Degrees of freedom	Confidence level (%) 50	90	95	98	99	99.5	99.9
1	1.000	6.314	12.706	31.821	63.656	127.321	636.578
2	0.816	2.920	4.303	6.965	9.925	14.089	31.598
3	0.765	2.353	3.182	4.541	5.841	7.453	12.924
4	0.741	2.132	2.776	3.747	4.604	5.598	8.610
5	0.727	2.015	2.571	3.365	4.032	4.773	6.869
6	0.718	1.943	2.447	3.143	3.707	4.317	5.959
7	0.711	1.895	2.365	2.998	3.500	4.029	5.408
8	0.706	1.860	2.306	2.896	3.355	3.832	5.041
9	0.703	1.833	2.262	2.821	3.250	3.690	4.781
10	0.700	1.812	2.228	2.764	3.169	3.581	4.587
15	0.691	1.753	2.131	2.602	2.947	3.252	4.073
20	0.687	1.725	2.086	2.528	2.845	3.153	3.850
25	0.684	1.708	2.060	2.485	2.787	3.078	3.725
30	0.683	1.697	2.042	2.457	2.750	3.030	3.646
40	0.681	1.684	2.021	2.423	2.704	2.971	3.551
60	0.679	1.671	2.000	2.390	2.660	2.915	3.460
120	0.677	1.658	1.980	2.358	2.617	2.860	3.373
∞	0.674	1.645	1.960	2.326	2.576	2.807	3.291

The values in the table represent the width of the *t* distribution (in terms of numbers of standard deviations)

∞ is the Gaussian distribution

Picture in picture

95.5% of measurements within ±2σ

17:55 / 30:49

In calculating confidence intervals, σ may be substituted for *s* in Equation 4-4 if you have a great deal of experience with a particular ... and ... therefore ... determine its "true" p... ion standard deviation. If σ is used instead of *s*, ... value ... to use in ... equation ... hes from ... bottom ...

Comparing mean

t-Test for Compari

Picture in picture

- In a *t*-test, we calculate a value for *t* from the $\bar{x}$, s, and n values
- If the calculated *t* is **greater** than the tabulated *t* at the 95% CL, the two means are significantly different.
- If the calculated *t* is **lower** than the table value, difference between means is not significant.
- Two sets of formulas to use, depending on whether the standard deviations of the measurements are significantly different
  - So conduct an *F*-test first to determine this

- If the calculated *t* is **greater** than the tabulated *t* at the 95% CL, the two means are significantly different.
- If the calculated *t* is **lower** than the table

Picture in picture

The case when the standard deviations **are not** significantly different:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{\text{pooled}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$s_{\text{pooled}} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Compare with *t* table for the sum of the degrees of freedom... i.e., d.f. = $(n_1 + n_2 - 2)$

- If the calculated *t* is **greater** than the tabulated *t* at the 95% CL, the two means are significantly different.
- If the calculated *t* is **lower** than the table value, not significant.

The case when the standard deviations **are** significantly different:

$$t_{\text{calculated}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}} \qquad (4\text{-}7)$$

$$\text{degrees of freedom} = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \qquad (4\text{-}8)$$

Round the calculated degrees of freedom to an integer, and look that up in the table

Practice Problem

24. Gács and Ferraroli reported a method for monitoring the concentration of SO_2 in air.[24] They compared their method to the standard method by analyzing urban air samples collected from a single location. Samples were collected by drawing air through a collection solution for 6 min. Shown here is a summary of their results with SO_2 concentrations reported in $\mu L/m^3$.

standard method:	21.62	22.20	24.27	23.54
	24.25	23.09	21.02	
new method:	21.54	20.51	22.31	21.30
	24.62	25.72	21.54	

Using an appropriate statistical test, determine whether there is any significant difference between the standard method and the new method at $\alpha = 0.05$.

22:02 / 30:49

Both means and st.dev.

Solution

Picture in picture

$$F_{\text{exp}} = \frac{s_{\text{new}}^2}{s_{\text{std}}^2} = \frac{(1.92)^2}{(1.28)^2} = 2.25$$

The critical value for $F(0.05,6,6)$ is 5.820.

Since F_{exp} is smaller than F_{table}, the standard deviations are **not** significantly different

$$s_{\text{pool}} = \sqrt{\frac{(6)(1.28)^2 + (6)(1.92)^2}{7 + 7 - 2}} = 1.63$$

Since t_{exp} is smaller than t_{table}, the means are **not** significantly different

$$t_{\text{exp}} = \frac{|\bar{X}_{\text{std}} - \bar{X}_{\text{new}}|}{s_{\text{pool}}} \times \sqrt{\frac{n_{\text{std}} \times n_{\text{new}}}{n_{\text{std}} + n_{\text{new}}}} = \frac{|22.86 - 22.51|}{1.63} \times \sqrt{\frac{7 \times 7}{7 + 7}} = 0.40$$

22:41 / 30:49

The critical value for $t(0.05,12)$ is 2.[illegible].

4-3 Confidence Intervals

In science we very often have a limited number of measurements. Can compute $\bar{x}$ and s but not μ or σ.

The **confidence interval** is a range of values within which there is a specified probability of finding the population mean μ.

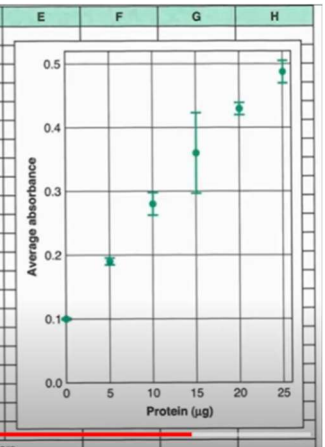
Using the confidence interval: $\mu = \bar{x} \pm \frac{ts}{\sqrt{n}}$

For example, the 95% CI means "there is a 95% probability that the true population mean μ lies within $\pm \frac{ts}{\sqrt{n}}$ of the measured average $\bar{x}$ "

where s is the measured standard deviation, n is the number of observations, and t is Student's t, taken from Table 4-4. In this table, the *degrees of freedom* are $n - 1$.

4-8 Error Bars

	A	B	C	D
1	Adding error bars to a graph			
2	Protein	Mean	Std dev	95%
3	(µg)	absorbance	absorbance	confidence
4	0	0.0993	0.00058	0.0014
5	5	0.1867	0.00153	0.0038
6	10	0.2753	0.00577	0.0143
7	15	0.3613	0.02658	0.0660
8	20	0.4267	0.00289	0.0072
9	25	0.4890	0.00656	0.0163
10				
11	TINV(0.05,2) =	4.303		
12	Student's t (95% conf, 2 deg freedom)			
13	95% confidence interval: D4 = B11*C4/SQRT(3)			
14				
15	To insert y error bars for data in column D:			
16	Click one point on the graph to highlight all points			
17	In Chart Tools, Layout, select Error Bars			
18	Choose More Error Bars Options			
19	Click Custom and click Specify Value			
20	For Positive Error Value, highlight D4:D9			
21	For Negative Error Value, highlight D4:D9			
22	Click OK and error bars will appear on the chart			
23	Click on one x error bar and press Delete to remove all x error bars			



26:29 / 30:49

Figure 4-10 Adding 95% confidence error bars to a graph.

The error bars extend $\frac{ts}{\sqrt{n}}$ above **and** $\frac{ts}{\sqrt{n}}$ below the point, reflecting the meaning of ±

Some graphs give error bars that are ± s
Some other graphs use the 95% CI
Some other graphs use the standard error of the mean

You need to specify which is which in the figure caption or text!

4-5 Grubbs Test for an Outlier

The version for a single suspected outlier:

$$G = \frac{|\text{questionable value} - \bar{x}|}{s}$$

If $G_{calculated} > G_{table}$, reject the questionable point.

Include the questionable point when calculating the average and st. dev. to use in this test

Table 4-6 Critical values of G for rejection of outlier[a, b]

Number of observations	G (95% confidence)
4	1.463
5	1.672
6	1.822
7	1.938
8	2.032
9	2.110
10	2.176
11	2.234
12	2.285
15	2.409
20	2.557

b. $G_{calculated} = |\text{questionable value} - \text{mean}|/s$. If $G_{calculated} > G_{table}$, the value in question can be rejected with 95% confidence. Values in this table are for a one-tail test, as recommended by the American Society for Testing and Materials (ASTM).

The t-test

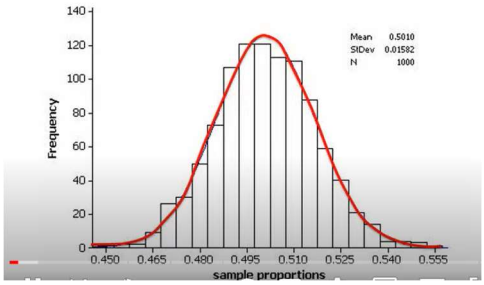
Picture in picture

What can this test tell you?

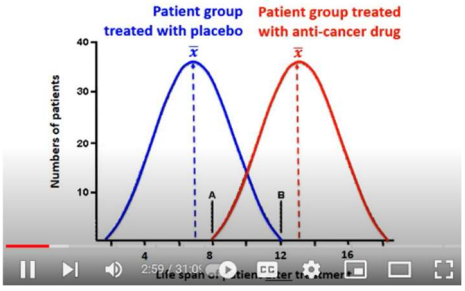
The t-test tells us if there is a **statistically significant difference** between the **mean values** of **two data sets**, when:

The t-test

Picture in picture



The t-test



The t-test

Picture in picture

1. An **independent samples t-test** compares the mean for two groups *(e.g. **effect of drug** vs **placebo**)*
2. A **paired sample t-test** compares means from the same group at different times *(e.g. **effect of drug after 6 months** vs **pre-treatment** of same population)*
3. A **one sample t-test** tests the mean of a single group against a known mean

The t-value

Picture in picture

Measures the size of the difference in mean values relative to the variation in your sample data

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

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

$\bar{x}_1$ = mean of first sample
$\bar{x}_2$ = mean of second sample
s_1 = standard deviation of first sample
s_2 = standard deviation of second sample
n_1 = number of measurements in first sample
n_2 = number of measurements in second sample