A Brief Guide to Laboratory Statistics

CONTENTS

[The Basics 2](#_Toc376177062)

[Data 2](#_Toc376177063)

[Descriptive Statistics 2](#_Toc376177064)

[Standard Deviation 2](#_Toc376177065)

[Properties of standard deviations 4](#_Toc376177066)

[z score 5](#_Toc376177067)

[Pooled Standard Deviation 5](#_Toc376177068)

[Pooled variance 5](#_Toc376177069)

[Determining Standard Deviation in the Laboratory. 6](#_Toc376177070)

[Effect Size 6](#_Toc376177071)

[Type I and Type II errors 6](#_Toc376177072)

[Standard error of the mean 7](#_Toc376177073)

[Duplicate Repeatability 7](#_Toc376177074)

[Interim Precision 8](#_Toc376177075)

[Horwitz’s Trumpet 8](#_Toc376177076)

[Robust Statistics 9](#_Toc376177077)

[IQR etc 10](#_Toc376177078)

[Robust z score 10](#_Toc376177079)

[Standard Difference 11](#_Toc376177080)

[Curve Fitting and Least squares 11](#_Toc376177081)

[Proficiency programs 11](#_Toc376177082)

[Control charts 12](#_Toc376177083)

[Two IRM/SRMs in a batch 14](#_Toc376177084)

[MU 15](#_Toc376177085)

[Precision 15](#_Toc376177086)

[Bias 15](#_Toc376177087)

[UoBias 16](#_Toc376177088)

[LOD & LOQ 17](#_Toc376177089)

[Comparing Data 18](#_Toc376177090)

[t test 18](#_Toc376177091)

[F test 19](#_Toc376177092)

[Anova 21](#_Toc376177093)

[Rounding 23](#_Toc376177094)

[Applying the Basics. 23](#_Toc376177095)

[Before you do anything else 23](#_Toc376177096)

[Excel Hints & Tips 24](#_Toc376177097)

[Examples 25](#_Toc376177098)

[References 26](#_Toc376177099)

# The Basics

There is nothing here to scare you. (Other than this document being 52.6% larger than expected.)

Statistics are only a way of describing the properties of data.

# Data

Most collections of data follow a Gaussian distribution. Commonly known as a ‘Bell’ curve or a ‘’Normal’ curve.

Some data, notably microbiological data, does not follow the Gaussian distribution. In order to make it do so, you must transform the data by techniques such using logarithms.

# Presenting Data for statistical analysis

1. Each variable you measure should be in one column
2. Each different observation of that variable should be in a different row
3. There should be one table for each "kind" of variable
4. If you have multiple tables, they should include a column in the table that allows them to be linked.

# Descriptive Statistics

Using Excel’s descriptive statistics function you can get an indication of a dataset’s ‘normality’ from the skewness and kurtosis figures.

**Skewness**: Skewness quantifies how symmetrical the distribution is.

* A symmetrical distribution has a skewness of zero.
* An asymmetrical distribution with a long tail to the right (higher values) has a positive skew.
* An asymmetrical distribution with a long tail to the left (lower values) has a negative skew.

**Kurtosis**: Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution.

* Excel sets a normal curve to a kurtosis of 3.
* A result less than 3 indicates a curve flatter than a normal distribution.
* A result greater than 3 indicates a curve sharper than a normal distribution.

# Standard Deviation

The standard deviation is just a way of describing the spread of the curve. The spread is described by the variance. The standard deviation is the square root of the variance, which means that it has the same units as the mean.

A small standard deviation means a narrow shaped ‘bell’, a large standard deviation means a low flat ‘bell’.

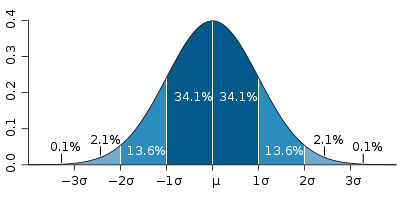


Figure 1.

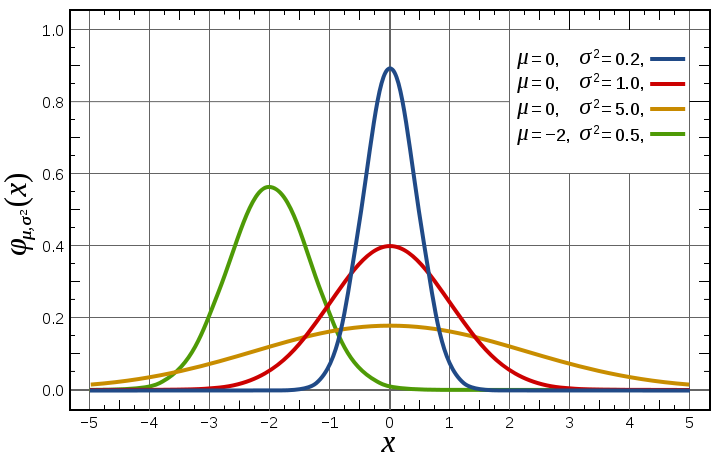


Figure 2.

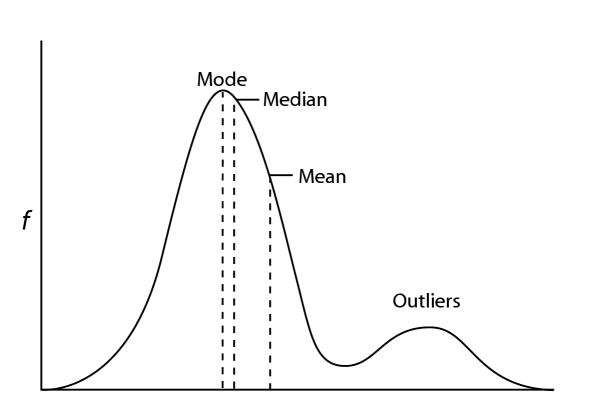


Figure 3.

Standard Deviation shows how much variation there is from the "average" (mean). It may be thought of as the average difference of the scores from the mean of distribution, how far they are away from the mean.

A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values.

The most common estimator for σ used is an adjusted version, the **sample standard deviation,** denoted by "s" and defined as follows:

s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \overline{x})^2},

where {x1, x2,... xn}is the sample and x-bar is the mean of the sample.

This correction (the use of N − 1 instead of N) is known as Bessel's correction. The reason for this correction is that s2 is an unbiased estimator for the variance of the underlying population, if that variance exists and the sample values are drawn independently with replacement. However, s is not an unbiased estimator for the standard deviation σ; it tends to underestimate the population standard deviation.

# Properties of standard deviations

For constant *c* and random variables *X* and *Y*:

 \operatorname{stdev}(X + c) = \operatorname{stdev}(X) \,, 

 \operatorname{stdev}(cX) = |c|\,\operatorname{stdev}(X) \,, 

 \operatorname{stdev}(X + Y) = \sqrt{\operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y)} \,, 

where var() and cov() stand for variance and covariance, respectively.

NOTE: Usually, covariance is not an issue for chemists as the variables are usually independent. Usually. An example where they are not is in the calculation of the MU of Energy. The value for carbohydrate is not independent of the values for fat and protein and allowance must be made for covariance. This is done most easily by using the Monte Carlo method.

(A ± a) + (B ± b) = (C ± c)

A + B = C and c = sqrt( a2 + b2 )

For a complex product (and division) equation:

(A ± a) \* (B ± b) / (C ± c) = (D ± d) where \* indicates multiplication   
  
D = A \* B / C and

d = D \* sqrt ((a/A)2 + (b/B)2 + (c/C)2)

# z score

A z-score is a normalised standard deviation where the mean is zero and the standard deviation is 1. This permits the comparison of distributions.

Conceptually, this is analogous to a signal to noise ratio.

# Pooled Standard Deviation

Pooled standard deviation is the square-root of the pooled variance.

# Pooled variance

In statistics, many times, data are collected for a dependent variable, y, over a range of values for the independent variable, x.

For example, the observation of fuel consumption might be studied as a function of engine speed while the engine load is held constant. If, in order to achieve a small variance in y, numerous repeated tests are required at each value of x, the expense of testing may become prohibitive. Reasonable estimates of variance can be determined by using the principle of pooled variance after repeating each test at a particular x only a few times. Pooled variance is a method for estimating variance given several different samples taken in different circumstances where the mean may vary between samples but the true variance (equivalently, precision) is assumed to remain the same.

It is calculated by

s_p^2=\frac{\sum_{i=1}^k (n_i - 1)s_i^2}{\sum_{i=1}^k(n_i - 1)}

or with simpler notation,

s_p^2=\frac{(n_1 - 1)s_1^2+(n_2 - 1)s_2^2+\cdots+(n_k - 1)s_k^2}{n_1+n_2+\cdots+n_k - k}

where sp2 is the pooled variance, ni is the sample size of the i'th sample, si2 is the variance of the ith sample, and k is the number of samples being combined. n − 1 is used instead of n for the same reason it may be used in estimating variances from samples (i.e. Bessel's correction).

# Determining Standard Deviation in the Laboratory.

Where a sample is tested multiple times, such as a control sample, the standard deviation is easily determined by the normal equation or by the function STDEV(range) in Excel.

If you do not have access to multiple testing of a single sample, the standard deviation can be estimated from the difference of a series of duplicates.

**Note**: The data needs to be of comparable matrices and comparable levels of analyte.

where x1 and x2 are the duplicates and n is the number of duplicate pairs.

Repeatability standard deviation is determined from duplicates and reproducibility standard deviation is determined from retests.

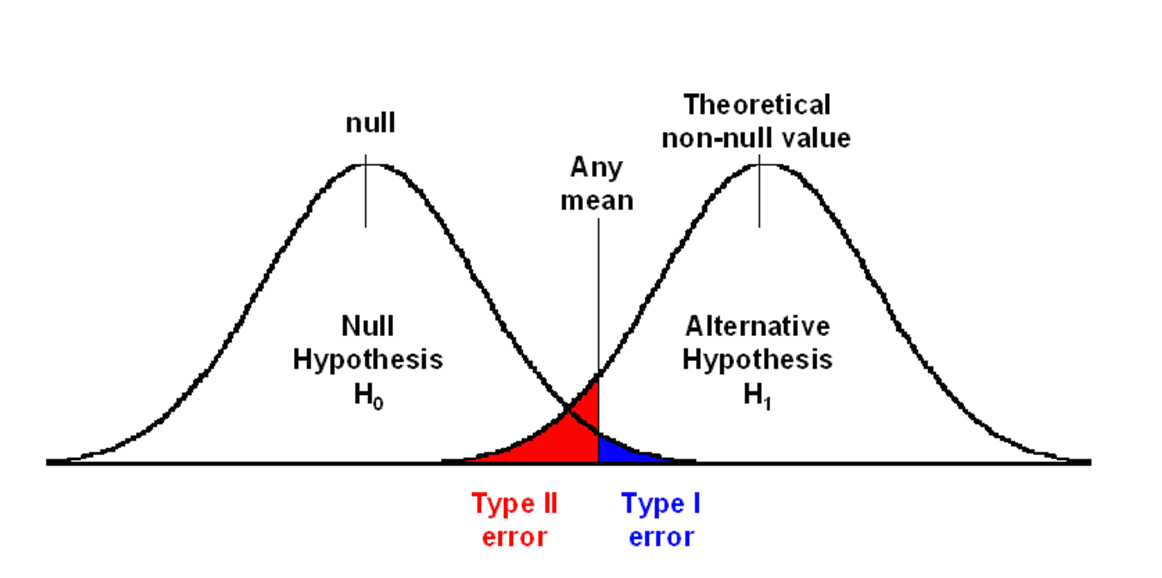
# Effect Size

Calculate the effect size by subtracting the control group mean from the final group mean and dividing the result by the standard deviation of the control group. (Sometimes an arithmetic mean of the two sd’s is used.) The greater the effect, the greater the impact of the ‘treatment’. A result near zero indicates no effect.

# Type I and Type II errors

Type I error: the null hypothesis is rejected, even though it is true.

Type II error: a null hypothesis is accepted, even though it is false.



# Standard error of the mean

The standard error of the mean (SEM) is the standard deviation of the sample-mean's estimate of a population mean.

SEM is usually estimated by the sample estimate of the population standard deviation (sample standard deviation) divided by the square root of the sample size (assuming statistical independence of the values in the sample):

SEM =

where

s = the sample standard deviation (i.e., the sample-based estimate of the

standard deviation of the population), and

n = the size (number of observations) of the sample.

Consequently, the SEM is always smaller than the sample standard deviation and this is why an average is always a better estimate than a single test.

# Duplicate Repeatability

Repeatability, in the laboratory sense, is how close you expect two duplicate results to be. Same sample, tested in duplicate, together.

where sr = standard deviation for repeatability duplicates.

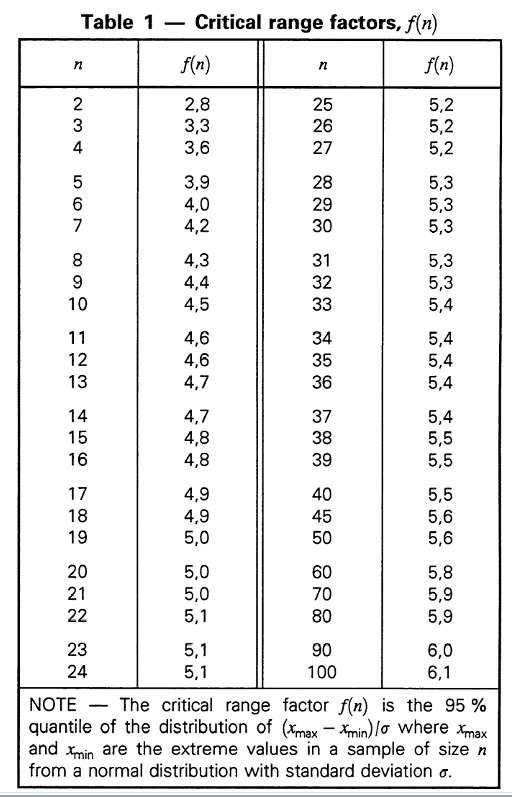
# Interim Precision

Interim Precision means reproducibility, in the laboratory sense, and is how close you expect two replicate results to be. Same sample, different batches (and possibly days, analysts and instruments.

where sR = standard deviation for replicates.

True reproducibility is a sign of the variance that can be expected between laboratories and is reflected in proficiency programs.

Reproducibility > Interim Precision > Repeatability.



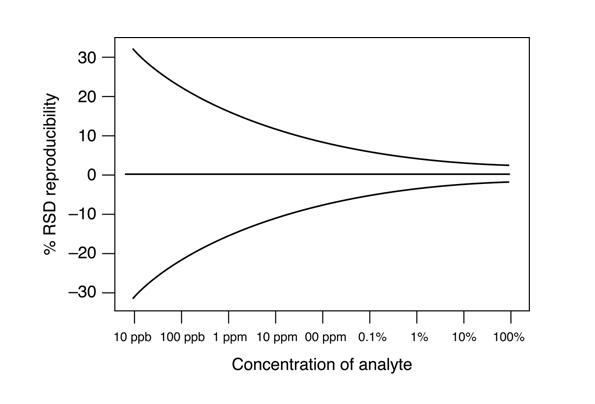
From ISO 5725-6

# Horwitz’s Trumpet

Dr William Horwitz looked at the relative standard deviation for several thousand proficiency programs and found that there was a relationship between the level of analyte being determined and the relative standard deviation of the results of the participating laboratories. He could then predict the RSD of another progam.

where c is the concentration of analyte, expressed in g/g. (eg 1 mg/kg = 0.000001g/g)

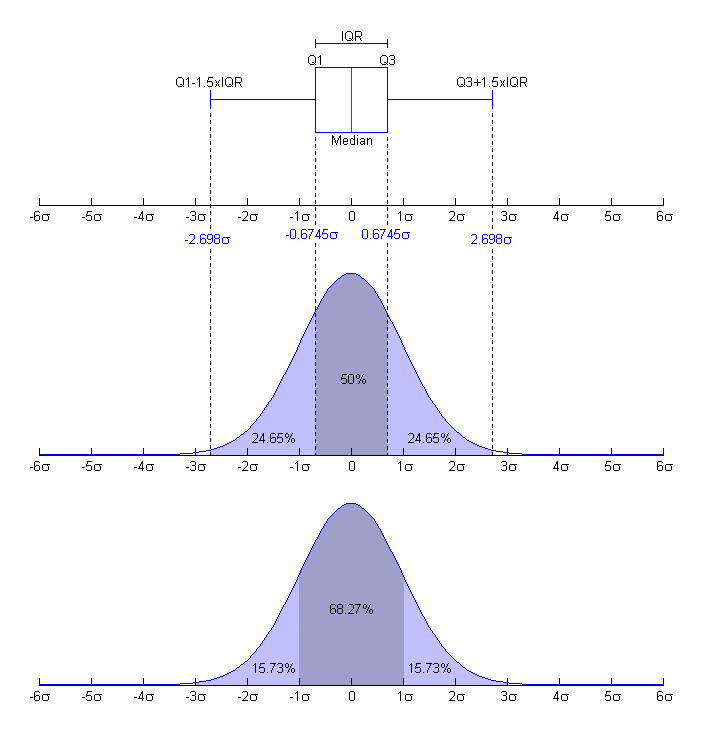
A Horrat Value is the ratio of a proficiency program’s RSD to Horwitz’s predicted value.



# Robust Statistics

Robust statistics are not influenced markedly by outliers or extreme results. For example the median will remain unchanged if one result is extreme where the mean can be dramatically affected.

Proficiency programs, such as Global Proficiency, use robust statistics because the participation group is often so small.



# IQR etc

sd ~ 0.7413 x IQR

where IQR is the interquartile range of the sample, is a consistent estimate of σ if the population is normally distributed. The interquartile range IQR is the difference of the 3rd quartile of the data and the 1st quartile of the data. The asymptotic relative efficiency (ARE) of this estimator with respect to the one from sample standard deviation is 0.37. Hence, for normal data, it is better to use the one from sample standard deviation; when data is with thicker tails, this estimator can be more efficient.

# Robust z score

# Standard Difference

Conceptually, a standard difference is similar to a z-score.

The value for the limit is an agreed value for the test. It is, in effect, a pooled standard deviation for the test and most useful when only a few laboratories are participating in a program.

# Curve Fitting and Least squares

r

R2

In statistics, the coefficient of determination, R2 is used in models for the prediction of future outcomes on the basis of other related information. It is the proportion of variability in a data set that is accounted for by the statistical model. It provides a measure of how well future outcomes are likely to be predicted by the model.

In this case, R2 is simply the square of the sample correlation coefficient between the outcomes and their predicted values, or in the case of simple linear regression, between the outcome and the values being used for prediction. In such cases, the values vary from 0 to 1.

"Approximately seventy percent of the variation in the response variable can be explained by the explanatory variable. The remaining thirty percent can be explained by unknown, lurking variables or inherent variability."

Significance of R2

# Proficiency programs

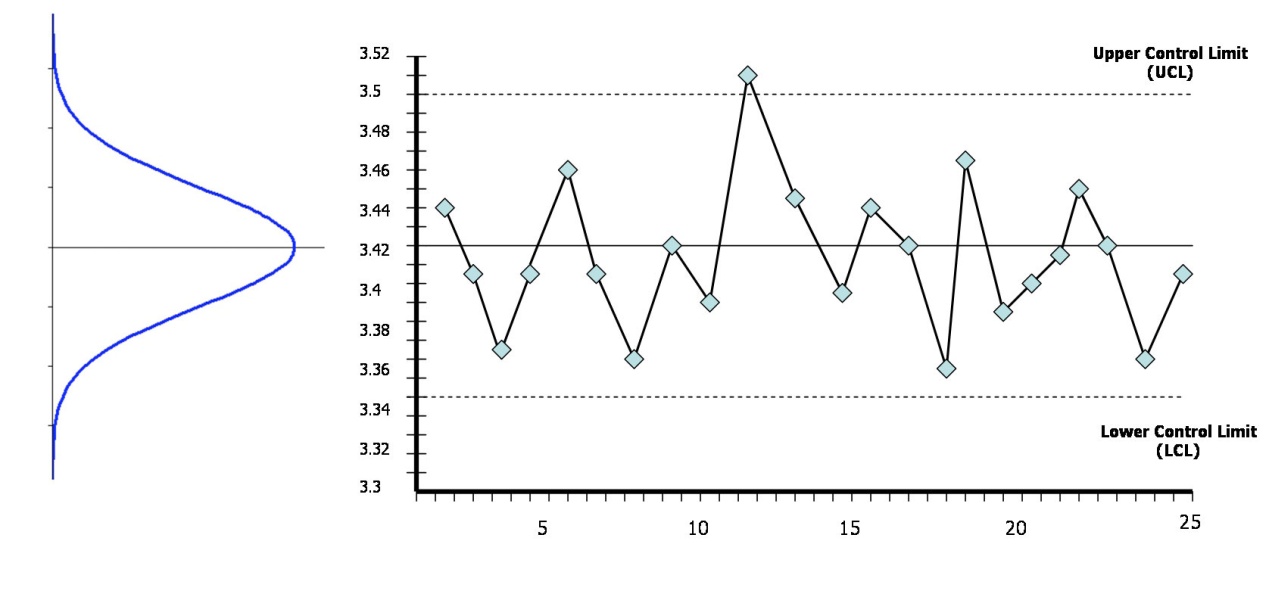
En-Score

where

# Control charts

Control Charts, as the name suggests, help determine whether a process is under control.

Conceptually, they are a 3-D Normal curve. The majority of points should fall in the middle, evenly across both sides of the middle, with no trending.



Trending Rules:

|  |  |  |  |
| --- | --- | --- | --- |
| **Rule** | **Description** | **Chart Example** | Problem Indicated |
| Rule 1 | One point is more than 3 standard deviations from the mean. | Rule 1 - Control Charts for Nelson Rules.svg | One sample (two shown in this case) is grossly out of control. |
| Rule 2 | Nine (or more) points in a row are on the same side of the mean. | Rule 2 - Control Charts for Nelson Rules.svg | Some prolonged bias exists. |
| Rule 3 | Six (or more) points in a row are continually increasing (or decreasing). | Rule 3 - Control Charts for Nelson Rules.svg | A trend exists. |
| Rule 4 | Fourteen (or more) points in a row alternate in direction, increasing then decreasing. | Rule 4 - Control Charts for Nelson Rules.svg | This much oscillation is beyond noise. |
| Rule 5 | Two (or three) out of three points in a row are more than 2 standard deviations from the mean in the same direction. | Rule 5 - Control Charts for Nelson Rules.svg | This is directional and the position of the mean and size of the standard deviation do not affect this rule. |
| Rule 6 | Four (or five) out of five points in a row are more than 1 standard deviation from the mean in the same direction. | Rule 6 - Control Charts for Nelson Rules.svg | There is a medium tendency for samples to be mediumly out of control. |
| Rule 7 | Fifteen points in a row are all within 1 standard deviation of the mean on either side of the mean. | Rule 7 - Control Charts for Nelson Rules.svg | The side of the mean for the third point is unspecified. |
| Rule 8 | Eight points in a row exist with none within 1 standard deviation of the mean and the points are in both directions from the mean. | Rule 8 - Control Charts for Nelson Rules.svg | There is a strong tendency for samples to be slightly out of control. |

# Two IRM/SRMs in a batch

|  |  |  |
| --- | --- | --- |
| Criteria | Response | Rationale |
| Both IRMs within the ±2sd range. | System in control – no action |  |
| One IRM outside the ±2sd range, one within the ±2sd range | No action but monitor. | There is a 5% chance that this will happen randomly. |
| Both IRMs outside the ±2sd range. | Action required. | There is only a 0.25% chance of this happening randomly. |
| Either IRM is outside the 3sd range. | Action required. | There is only a 1% chance of this happening randomly. |

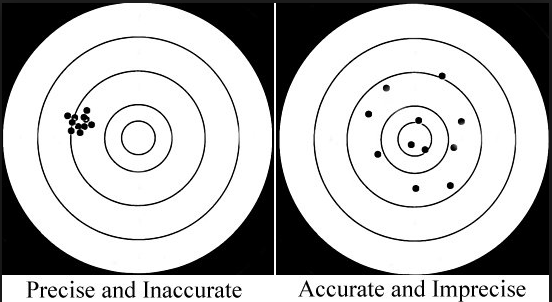
Note: the IRMs should not be run sequentially. With things like fat and ash it is not a big issue as the samples are all, in effect, tested separately but, in an instrumental batch, it is preferable that the IRMs are at each end of the batch.

# MU

Measurement Uncertainty is a value range around a laboratory result that the lab is 95% confident contains the true value.

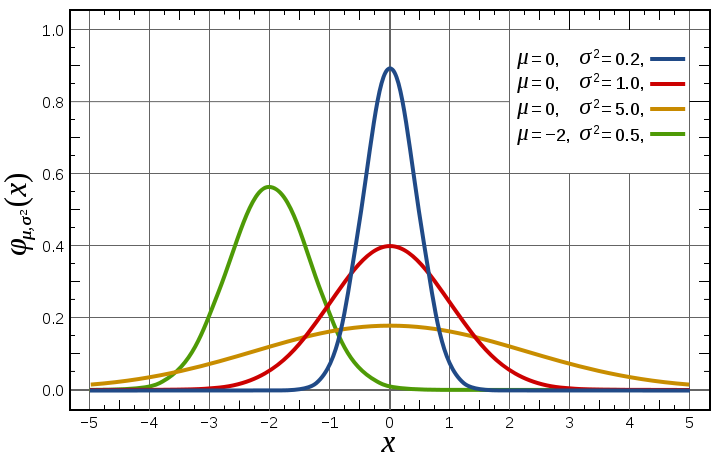
# Precision

How close the laboratory results are to each other.



# Bias

How close the laboratory results are to the ‘true’ value.

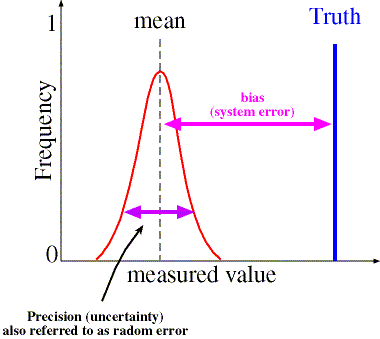


Bias = laboratory result minus the expected result.

# UoBias

Even when no bias appears to exist in a sample, there will be uncertainty as to whether any bias exists.

The diagram below compares the laboratory value to the ‘Truth’. The laboratory value has the appropriate uncertainty spread around it, however in reality there is no absolute ‘Truth’. Both the determined value and the certified “truth” value will have an uncertainty around them.



where

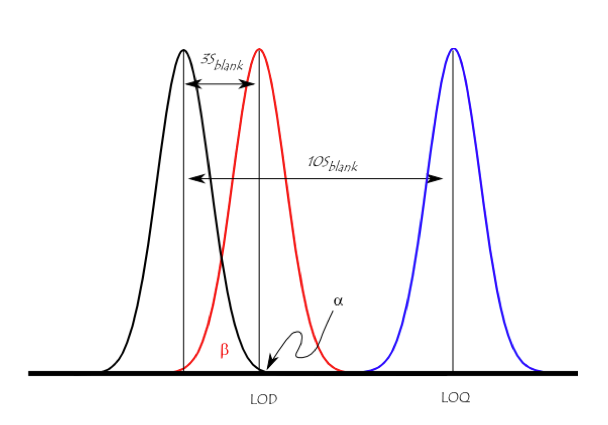
u(y) is the standard uncertainty of your result, and

u(yexp) is the standard uncertainty of the expected result.

The ‘Truth’ may be a CRM, a proficiency sample, a spike or even a result from another method or lab. It will have an uncertainty associated with its result.

If the magnitude of the bias, |b|, is greater than t(0.05, n-1) \* ub, then the bias is significant and must be included in the MU estimation. Generally t is taken to be equal to 2.

# LOD & LOQ



A general rule of thumb:

LOD = 3 x sd

LOQ = 10 x sd

LOR = LOQ

Where:

LOD = Limit of detection

LOQ = Limit of quantitation

LOR = Limit of reporting

sd = standard deviation of replicate tests near the LOR.

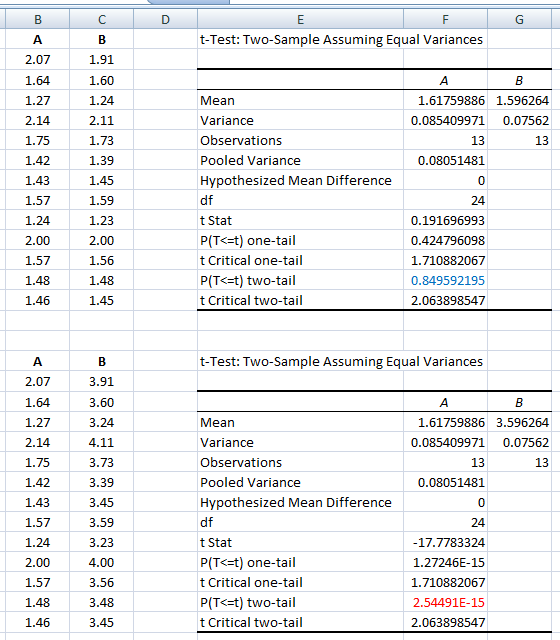
# Comparing Data

# t test

For a single set of data, the equation is

The t-test is used for comparing two means and their associated standard deviations, to determine whether there is a significant difference between the two. Or, put another way, are they likely to have come from the same population or not?

In the following two examples, the first sets are closely related and the second sets are not. This is reflected in the p(T<=t) probability figure. If p(T<=t) < 0.05 then the sets are considered likely to be from different populations.



Typical report wording (first example): (t=0.192, 24 d.f., P=0.85)

# F test

The F-test is designed to test if two population variances are equal. It does this by comparing the ratio of two variances. So, if the variances are equal, the ratio of the variances will be 1.

Assumptions / Notes

* The larger variance should always be placed in the numerator
* The test statistic is F = s1^2 / s2^2 where s1^2 > s2^2
* Divide alpha by 2 for a two tail test and then find the right critical value
* If standard deviations are given instead of variances, they must be squared
* When the degrees of freedom aren't given in the table, go with the value with the larger critical value (this happens to be the smaller degrees of freedom). This is so that you are less likely to reject in error (type I error)
* The populations from which the samples were obtained must be normal.
* The samples must be independent

**F-Test Example**

If we are not interested in whether one method is better compared to another, but were simply trying to determine if the variances of were the same or different, we would need to use a 2-tailed test. For instance, assume we made two sets of measurements of ethanol concentration in a sample of vodka using the same instrument, but on two different days. On the first day, we found a standard deviation of s1 = 9 ppm and on the next day we found s2 = 2 ppm. Both datasets comprised 6 measurements. We want to know if we can combine the two datasets, or if there is a significant difference between the datasets, and that we should discard one of them.

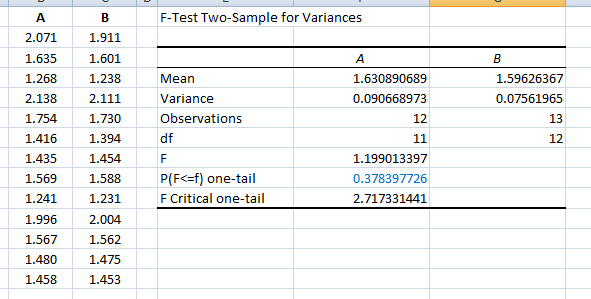
As usual, we begin by defining the null hypothesis, H0: σ12 = σ22, and the alternate hypothesis, HA: σ12 ≠ σ22. The "≠" sign indicates that this is a 2-tailed test, because we are interested in both cases: σ12 > σ22 and σ12 < σ22.

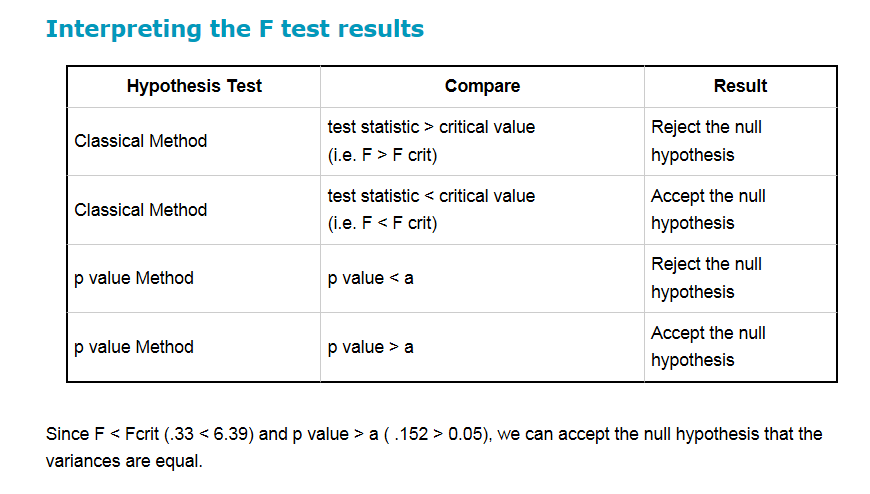
For the F-test, you can perform a 2-tailed test by multiplying the confidence level P by 2, so from a table for a 1-tailed test at the P = 0.05 confidence level, we would perform a 2-tailed test at P = 0.10, or a 90% confidence level.

For this dataset, s2 > s1, Fcalc = s12/ s22 = 92/22 = 20.25.

The tabulated value for ν = 5 at 90% confidence is F5,5 = 5.050.

Since Fcalc > F5,5, we reject the null hypothesis, and can say with 90% certainty that there is a difference between the standard deviations of the two methods.

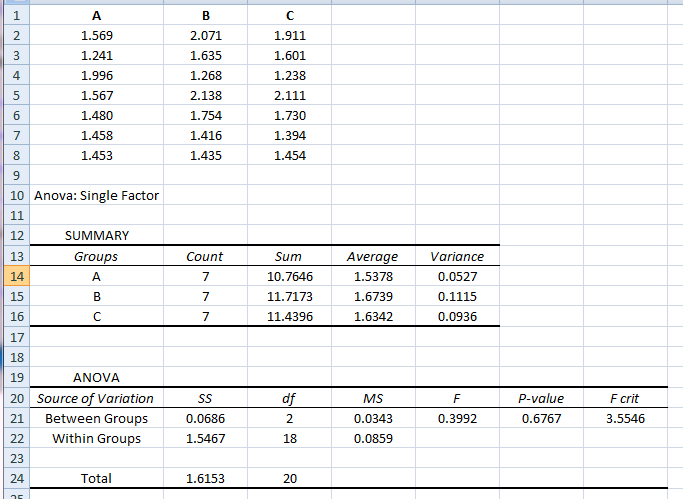




# Anova

ANOVA permits the comparison of two or more groups to compare their means. It will tell you that they are not all the same but will not tell you which one is different.

Typical statement: "There were no statistically significant differences between group means as determined by one-way ANOVA (F(2,18) = 0.399, p = 0.677)"



An estimate of repeatability standard deviation, sr, is obtained by the square root of the Within Groups Mean Square value. ( sqrt(0.0859) = 0.293, a variance of 0.0864, cf the three individual variance results).

An estimate of the between group standard deviation is given by

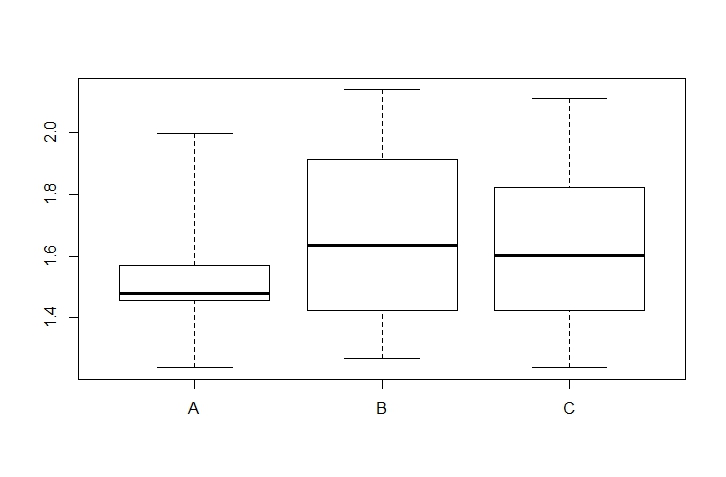
where n = the number of data points in a sets (=7 in the example).

An estimate of the intermediate precision standard deviation is given by

If the between variance is smaller than the within variance, then the means are really close to each other. When F<1 or F=1, between-group variance is smaller than within group variance, suggesting that the groups differ as much as one would expect due to sampling error. If there is no effect due to the different groups, the two MS values should be the same.

If differences across sample means are "large" relative to differences within samples, then we should reject the null hypothesis that the samples are all drawn from the same population. If the relationship is the other way, differences within samples is greater than differences between samples, we cannot say the means are different.

In the example data, the boxplot of the data shows that they are very similar:



# Rounding

The correct level of rounding for an analysis is indicated by the standard deviation of that analysis. If need be you can round to a less accurate level that the level indicated by this rule but not to a higher level of accuracy unless it is at the client’s specific request (eg project work).

**Rounding Rule**: A result should be rounded to the power of ten that is immediately less than half the standard deviation of that analysis.

**Example 1**: If the sd of the analysis is 0.5, half the sd is 0.25 and the power of ten immediately less than this is 0.1 so the results should be rounded to one decimal place.

**Example 2**: If the sd of the analysis is 3, half the sd is 1.5 and the power of ten immediately less than this is 1 so the results should be rounded to the nearest whole figure.

# Applying the Basics.

# Before you do anything else

Determine what do you want to show.

Correlation

Differences

Determine what data you need to collect.

Plot your data.

It is always a good reality check to plot your data. Nothing gives you a better idea of how your data sits than a graphical depiction of it.

* Regular plots
* X-Y plots
* Box Plots

Once you have had a look at your data, decide whether it is complete and whether it needs to be ‘cleaned’.

Incomplete data may have areas where your feel that more data is needed to present a complete picture.

Cleaning data is not a bad thing and can involve removing blanks, removing ‘less thans’ (<), and removing obvious outliers.

Is your data normally distributed? Do you need to transform it to a format that is normally distributed? Check skewness and kurtosis.

Are there trends?

Trending data can sometimes be normalised by plotting the point to point changes.

Are there outliers?

Grubbs’s Outlier Test

<http://www.graphpad.com/quickcalcs/Grubbs1.cfm>

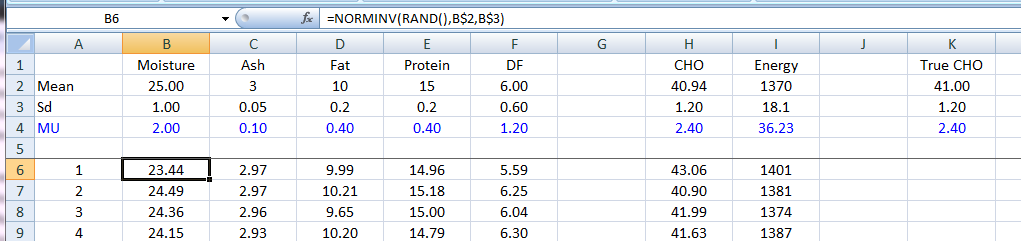
# Excel Hints & Tips

To generate random but normally distributed numbers use:

=NORMINV(RAND(), mean, standard\_dev)

Scratch Pad:

Monte Carlo & MU(Energy).



Spiking & recovery

# Examples

**A one-sample t-test with Excel.**

Excel does not have a single sample formula.

It is assumed you have the results of a single measure on a single sample: I.e. the mean and standard deviation. The measure might just be a single outcome, or else the outcome could be a difference score, or a change score calculated as a before-after difference.

For example, we take the Anorexia study cited in Howell, which gave a mean change score of 7.26 (in pounds) with standard deviation of 7.16. Sample size (i.e. pairs of data) was 17, giving 16 degrees of freedom.

It is important to do the t-test with the built-in Excel function TDIST(t,df,tails) where df is the degrees of freedom and tails = 2 (for a two-tailed test). The TTEST built-in function should not be used.

Calculate the standard error from the sample standard deviation with the Excel formula:

=7.16 / SQRT(17)

This gives the value 1.737

Enter the formula for t:

=7.26 / 1.737

This should give the value 4.18

Enter the function =TDIST(4.18, 16, 2)

This should give the probability .00071

Note that the calculated probability corresponds to the smaller part of the distribution.

# References

1. **ISO.** *Accuracy (trueness and precision) of measurement methods and results - Part 3.* s.l. : Australian Standards, 1994. ISO 5725-3.

2. **FAO Corporate Document Repository.** 8 Internal Quality Control of Data. *Guidelines for quality management in soil and plant laboratories. (FAO Soils Bulletin 74).* [Online] 1998. [Cited: 24 May 2013.] http://www.fao.org/docrep/W7295E/W7295E00.htm.