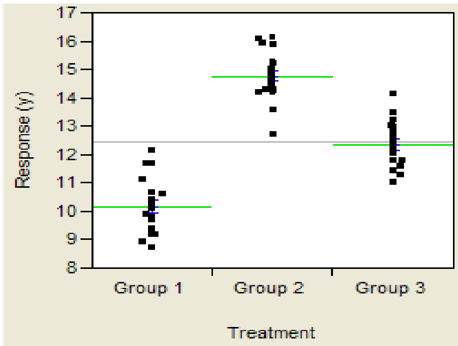
Analysis of Variance (ANOVA)



We have been using the Student's t-test and other techniques for comparing mean performance levels under two different conditions such as two different test methods or laboratories or laboratory analysts. However, we know that in analytical work we often have more than two means to be considered. Some possible situations are:

- comparing the mean concentration of protein in food product samples stored under different condition;
- comparing the mean results obtained for an analyte concentration by several different methods;
- comparing the mean titration results obtained by several different analysts using the same apparatus.

In all these examples, there are two possible sources of variation.

The first, which is always present, is due to the random error in measurement, leading to the issue of repeatability of the data. It is this error which causes a different result to be obtained each time a measurement is repeated under the same conditions.

The second possible source of variation is due to what is called as a controlled or fixed-effect factor. In the context of experimental design, a factor is a set of related conditions thought to affect performance of its measurements in terms of its mean value during the experiment. So, for the examples above, the controlled factors are respectively the conditions under which the sample was stored, the method of analysis used, and the analysts carrying out the titration.

The variance of different set of test results is analyzed and hence the statistic is called **analysis of variance** (frequently abbreviated to **ANOVA**). ANOVA is indeed an extremely powerful statistical technique which can be used to separate and estimate the different causes of variation.

For the above examples, it can be used to separate any variation which is caused by changing the controlled factor from the variation due to random error. It can thus test whether altering the controlled factor leads to a significant difference between the mean values obtained.

ANOVA can also be used in situations where there is more than one source of random variation. Consider, for example, the active ingredient testing of a drum of pesticide formulation. Samples are taken from different part of the drum chosen at random and replicate analyses performed on these samples. In addition to the random error in the measurement of the active ingredient, there may also be variation in the active ingredient contents of the samples from different parts of the drum. Since the samples were chosen at random, this variation will also be random and is thus sometimes known as a random-effect factor. Again, ANOVA can be used to separate and estimate the sources of variation.

Both types of statistical analysis described above are known as ***one-way* ANOVA** as there is one factor, either controlled or random, in addition to the random error in measurements. The arithmetical procedures are similar in the fixed- and random-effect factor cases.

When there are two factors affecting the results of an experiment, ***two-way* ANOVA** must be used to study their effects.

Experiments with two or more factors are known as factorial experiments. In the simplest case, there can be a different sample of participants for each possible combination of conditions. This arrangement is known as a between subjects (or completely randomized) factorial experiment. The discussion of two-way ANOVA can get a little bit more complicated than the one-way ANOVA does.

In the next few blogs, we shall explain the ANOVA techniques in plain language as much as possible and shall illustrate the power of ANOVA not only in the analysis of means in the design of experiment (DOE) but also in solving the straight line regression or linear calibration problems as well as to curvilinear regression.