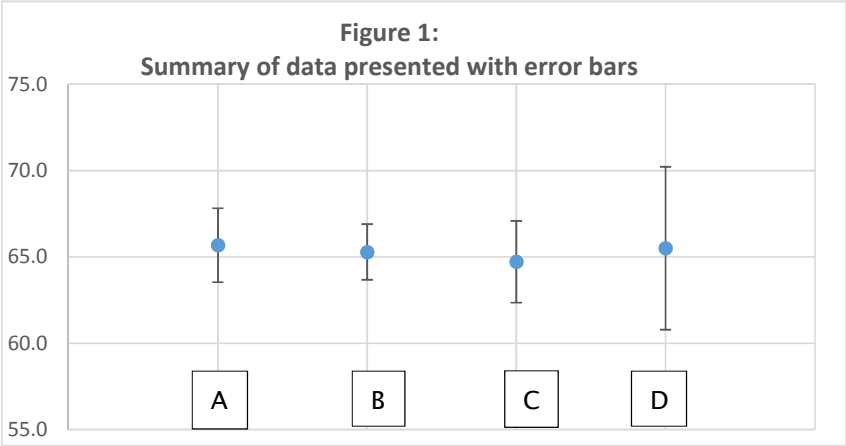
Analysis of Variance (ANOVA) revisited

Analysis of variance (ANOVA) is an important and powerful statistical technique which allows us to isolate and estimate the different causes of variations within sets of data collected.

Consider an example of laboratory training exercise which involved four analysts to check their individual technical competence in carrying out a series of testing for the determination of lead in a water sample using the same standard test method and the same FAAS instrument. Each analyst analyzed five portions of the water sample. Their results are summarized in Table 1 and Figure 1 below.

Table 1: Data of a training exercise on determination of Lead in a water sample (mg/L)

Analyst	A	B	C	D
X1	64.6	65.6	64.0	68.6
X2	65.5	67.5	62.6	68.0
X3	69.2	65.9	64.8	60.0
X4	65.6	63.6	63.5	70.1
X5	63.5	63.8	68.7	60.8
Mean	65.7	65.3	64.7	65.5
SD	2.142	1.615	2.364	4.727



In Table 1, we see two sources of variation within the data set.

First, we take note of the standard deviation of the results reported by each analyst. This will represent the inherent random variability associated with the test method used to measure the lead concentration in the water (i.e. variation *within* analysts).

Secondly we could determine the standard deviation of the mean values of the four sets of results from the analysts. This standard deviation will have a contribution from the random variation mentioned previously, plus any additional variation caused by the different analysts, i.e. variation *between* analysts.

Analysis of variance （ANOVA） allows us to separate and estimate these sources of variation and then make statistical comparisons.

So in the above example, we can use ANOVA to determine whether there is a significant difference among the means of the data sets produced by the analysts, or, whether the variation in the mean values can be accounted for by variation in the measurement system alone.

In statistical terms, the hypotheses being tested are:

H_o : the population means of the groups of data are all equal
H_1 : the population means of the groups of data are not all equal

ANOVA can thus be applied to any data set which can be grouped by particular ***factors***. In the example above, one factor – the analyst – has changed from one set of data to the next. The particular analysts, Analyst A to Analyst D, are the ***levels*** for the factor in this study.

So, one-way (also known as single-factor) ANOVA is used when data can be grouped by a single factor. When there are two factors influencing the results, two-factor ANOVA is used.

We may further subdivide the factors into either ***controlled*** (or '***fixed***') ***effects***, such as the reaction times and temperatures involved in an experiment, or ***random effects***, such as operator or 'run' effects.

The basic calculations for ANOVA remains the same whether effects are controlled (fixed) or random, but there exists more complex situations where the nature of the effects can sometimes influence the interpretation.

Of course, two-factor ANOVA or two-way ANOVA as commonly called is used when there are two factors that can influence the result of a measurement. If the above training example involved measurements of the concentration of lead using three different instruments available in the laboratory, we could ask each analyst to repeat their measurements on all three instruments. Then, the two factors for the ANOVA would therefore be analyst and instrument.

We can then use two-way ANOVA to investigate whether either (or both) of the factors have a significant influence on the results obtained. It is also possible to determine whether there is any interaction between the factors, i.e. whether one factor changes the effect of another.

There are two types of experimental design which lend themselves to analysis by two-way ANOVA – factorial (or cross-classified) designs and hierarchical (or nested) designs. In fact, two-way ANOVA is virtually always carried out using statistical software or a spreadsheet with statistical functions. We shall return to these subjects in subsequent blogs.