Comparing your lab results with the others by one-way ANOVA

You may have developed a new test method and in your method validation process you would like to check the method's ruggedness by conducting a simple cross-check program by comparing your own test results against the results provided by the other collaborating peer laboratories on a similar sample. What you may have done is to send them one or more similar homogeneous samples, request them to follow your method closely and report its repeatability data after their lab analyses.

When all the data are collated, you can carry out an analysis of variance (ANOVA) to see if all the data are significantly comparable or not. In fact, ANOVA is an extremely powerful statistical technique which can be used to separate and estimate the different causes of variation. In this case, the causes of variation are the 'laboratory' factor and the random error in measurement of each participating laboratory. Since there is only one controlled factor, the statistical approach is called one-way ANOVA.

Let's see an example to demonstrate how one-way ANOVA can be carried out.

The following table show the results of 4 similarly prepared samples received by 4 different laboratories in a simple cross-check programme organised for the measurement of zinc content in water in mg/L by the FAAS method:

Trial No.	**Lab 1**	**Lab 2**	**Lab 3**	**Lab 4**
1	103	102	97.4	107
2	99	102	95.3	110
3	101	106	99.5	109
Mean, mg/L	101	103.3	97.4	108.7

Overall Mean : 102.6 mg/L

The above table shows that the mean values for the 4 samples (from a same source) reported by the participating laboratories were *apparently* different. But, were they really so?

We know that because of *random error*, even if the true value which we are trying to measure, say 100 mg Zn/L is unchanged, the sample mean may vary from one laboratory to another.

ANOVA therefore, tests whether the difference between the samples reported by the 4 laboratories is too great to be explained by the random error alone. First of all, let's do a hypothesis testing as below:

H_o : all the samples tested by the laboratories were reporting comparable population mean results with a variance σ_o^2, *i.e.* $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1 : all mean results were not comparable, *i.e.* $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

On the basis of this hypothesis, the data variance σ_o^2 can be estimated in 2 ways:

* one involving the variation ***within*** *each laboratory*, and,
* the other, the variation ***between*** *the laboratories*.

(i) Within lab variation

Remember that variance is the *square* of standard deviation. Hence, upon calculation, we have:
Variance of Lab 1 = 4.00
Variance of Lab 2 = 5.33
Variance of Lab 3 = 4.41
Variance of Lab 4 = 2.33

Averaging these values gives:

Within-sample estimate of σ_o^2 = (4.00+5.33+4.41+2.33)/4 = 4.02

This estimate has 8 degrees of freedom, ν [each sample estimate had (3-1) or 2 degrees of freedom and there were 4 laboratories].

(ii) Between-lab variation

Since the samples analyzed by the 4 laboratories were all drawn from a population (one source) which has population variance σ_o^2 , their average of all the samples analyzed should come from this population with variance $\frac{\sigma_o^2}{n}$ (where square of $\frac{\sigma_o}{\sqrt{n}}$ is the variance of the population).

(*Recall the well known Central Limit Theorem* equation : $\mu = \bar{x} \pm z\frac{\sigma}{\sqrt{n}}$)

Thus, if the null hypothesis H_o is true, the variance of the means of the samples analyzed by these laboratories gives an estimate of σ_o^2/n.

In this example, we have :

Lab 1	101.0
Lab 2	103.3
Lab 3	97.4
Lab 4	108.7
Mean of Labs, mg/L	102.6
Number of labs (n)	4
Degree of freedom	3
Std deviation (*SD*)	4.724
Lab mean variance = (*SD*) squared	22.314

That means : $\sigma_o^2 / 3$ = 22.314 (where n = 3 measurements for each lab)

Therefore, the between-sample (laboratory) estimate of σ_o^2 (mean square) = 22.314 x 3 = 66.94

Note : this estimate of σ_o^2 *does not depend on the variability within each laboratory, because it is calculated from the lab means.*

In this case, the number of degrees of freedom ν is (4 – 1) or 3 and the mean square is 66.94 and so the sum of squared terms is 3 x 66.94 = 200.8

Now, let's summarize our calculations so far:

Within-lab mean square = 4.02 with df ν = 8

Between-lab mean square = 66.94 with df ν = 3

If our H_o is correct, these 2 estimates of σ_o^2 should not differ significantly, or else, for H_1, the between-lab estimate of σ_o^2 will be greater than the within-lab estimate because of between-lab variation.

To test if it is significantly greater, a one-tailed F-test is used:

$$F_{3,8} = 66.94 / 4.02 = 16.7$$

From the F-table, we find that the critical value at $F_{3,8}$ is 4.07 with 95% confidence. Since the calculated value of F is greater than this critical value, the null hypothesis Ho is rejected, i.e. the laboratory means *do* differ significantly. In other words, the proposed analytical method may not be 'rugged' enough and does not seem to be reproducible by all the participating laboratories.

If we were to use the Data Analysis Tool of Microsoft Excel® spreadsheet under the *ANOVA Single Factor* to analyse the above data, we should get the following results which are exactly the same as described above based on the first principle:

ANOVA: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Lab 1	3	303.000	101.000	4.000
Lab 2	3	310.000	103.333	5.333
Lab 3	3	292.200	97.400	4.410
Lab 4	3	326.000	108.667	2.333

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Labs	200.827	3	66.942	16.656	0.00084	4.07
Within Labs	32.153	8	4.019			
Total	232.98	11				

Another question: which lab or labs were the root cause for non-comparable results? Were they all significantly different amongst themselves?

To answer this, we can rearrange the means of the 4 labs in increasing order and compare the difference between adjacent values with a quantity called the **least significant difference (*LSD*)**, which has the following expression:

$$LSD = t_{h(n-1)} \times s \times \sqrt{\frac{2}{n}}$$

where s is the within-lab estimate of σ_o^2

h = number of factor (labs in this case)

n = number of repeats in each factor

and $h(n-1)$ is the number of degrees of freedom of this estimate.

In the above example, the sample means arranged in increasing order of size are:

Lab 3	Lab 1	Lab 2	Lab 4
97.4	101.0	103.3	108.7

The degree of freedom v = 4(3-1) = 8 and hence $t_{v=8}$ = 2.306.

The least significant difference therefore is 2.306 x $\sqrt{4.02}$ x $\sqrt{(2/3)}$ with 95% confidence, giving 3.78.

Now, difference of Lab 3 and Lab 1 = 101.0 – 97.4 = 3.6

difference of Lab 1 and Lab 2 = 103.3 – 101.0 = 2.3

difference of Lab 2 and Lab 4 = 108.7 – 103.3 = 5.4

Therefore when we compare this *LSD* value with the differences between the means, we note that :

- the mean of Lab 4 differs significantly from those of Lab 1,2 and 3;
- the mean values of Lab 1, 2 and 3 do not differ significantly from each other.