Study of one-way ANOVA with a fixed-effect factor

In the last blog on "Introduction to ANOVA", we mentioned that in the one-way ANOVA study, the factor contributing to a possible source of variation that would affect the final outcome of mean comparison could be a fixed-effect or a random-effect one. A fixed-effect factor (sometimes known as a controlled factor) can be the laboratory participating in an inter-laboratory cross-check exercise, the lab chemists carrying out the testing or the different drugs for curing a certain disease.

To understand how a one-way ANOVA with a fixed effect factor is done, let's consider the following worked example:

Six analysts each made 5 determinations of the paracetamol contents of the same batch of tablets. The results are shown below:

Table 1: Paracetamol contents (%m/m) reported by 6 analysts

Analyst	A	B	C	D	E	F
Paracetamol content (% m/m)	84.38	84.24	84.29	84.14	84.50	84.70
	84.52	84.25	84.40	84.22	83.88	84.17
	84.63	84.41	84.68	84.02	84.49	84.11
	84.61	84.13	84.28	84.48	83.91	84.36
	84.64	84.00	84.40	84.27	84.11	84.61

We shall test whether there is any significant difference between the means obtained by the six analysts by the one-way ANOVA.

Let the population (batch) variance be σ_o^2 and the significance hypotheses are:

H_o : within-sample mean square estimate of σ_o^2 and between-sample mean square estimate of σ_o^2 do not differ significantly
H_1 : within-sample mean square estimate of σ_o^2 and between-sample mean square estimate of σ_o^2 differ significantly

First, calculate the means and variances of the analysts:

Table 2: The means and variances of results reported by the 6 analysts

Analyst	A	B	C	D	E	F
Mean	84.556	84.206	84.410	84.226	84.178	84.390
Variance	0.01193	0.02323	0.02610	0.02908	0.09157	0.06805

The overall mean = 84.328
The overall variance of 6 mean results reported = 0.022068

1. Within-analyst variation

Within-analyst (generally speaking within-sample) variation looks into the performance of each analyst in their variances with a degree of freedom as (5-1) or 4, and calculate the within-sample estimate of σ_o^2 by averaging these 6 variance values to give 0.04416.

In general, if we designate the factor as a list of h samples with n replicates, we have a generalized table below:

Table 3: Generalization of Table 1

Under a factor					Mean
Sample 1	$x_{1,1}$	$x_{1,2}$ ….	$x_{1,j}$ ….	$x_{1,n}$	$\bar{x}_1$
Sample 2	$x_{2,1}$	$x_{2,2}$ ….	$x_{2,j}$ ….	$x_{2,n}$	$\bar{x}_2$
	….	…	….	….	…
Sample i	$x_{i,1}$	$x_{i,2}$ ….	$x_{i,j}$ ….	$x_{i,n}$	$\bar{x}_i$
	….	….	….	….	…
Sample h	$x_{h,1}$	$x_{h,2}$ ….	$X_{h,j}$ ….	$X_{h,n}$	$\bar{x}_h$

Then,

Within-sample estimate of $\sigma_o^2 = \sum_{i=1}^{h}\sum_{j}^{n}(x_{i,j} - \bar{x}_i)^2 / h(n-1)$ Eq (1)

The Eq(1) is a summation of squares over j and division by degree of freedom (n-1) to give the variance of each sample and then a summation over i and division by h averages these sample variances. It is known as a **mean square** (MS) since it involves a sum of squared (SS) terms divided by the number of degrees of freedom.

In this case, the number of degrees of freedom is 6 x (5-1) or 24, and the mean square is 0.02207, so the sum of the squared terms is 24 x 0.04416 or 0.9998.

2. Between-analyst variation

From Table 2, we stated the overall mean value of 84.328 %m/m based on mean result of each analyst and the associated overall variance of 0.02207, based on 5 degrees of freedom because there were 6 analysts involved.

As the samples were all drawn from a population which had a variance σ_o^2, then their means come from this population with variance σ_o^2/n (with reference to the sampling distribution of the mean). That is: σ_o^2/n = 0.022068.

So, the between-sample estimate of σ_o^2 is 0.022068 x 5 or 0.11034 as the number of repeats, n was 5.

Note that this estimate of σ_o^2 does not depend on the variability *within* each sample because it is calculated from the sample mean. But if, for example, the mean of sample from analyst A was changed, then this estimate of σ_o^2 would also be changed.

In general, we have:

$$\text{Between-sample estimate of } \sigma_o^2 = n\sum_{i=1}^{h}(\bar{x}_i - \bar{\bar{x}})^2/(h-1) \qquad \text{Eq (2)}$$

where $\bar{\bar{x}}$ is the overall mean of all analysts (samples).

The expression Eq(2) again is a 'mean square' involving a sum of squared terms divided by the number of degrees of freedom. In this case, the number of degrees of freedom is 5 as there were 6 analysts participated, and the mean square is 0.11034, so the sum of squared terms is 0.11045 x 5 = 0.55170.

In summary, we have so far:

Within-sample mean square MS = 0.04166 with 24 d.f.
Between-sample mean square MS = 0.11034 with 5 d.f.

If the null hypothesis H_o is correct, then these two estimates of σ_o^2 should not differ significantly. If it is incorrect, the between-sample estimate of σ_o^2 will be *greater than* the within-sample estimate because of between-sample variation.

We use a *one-side* F-test for significance testing:

$$F_{(\alpha=0.05, v1=5, v2=24)} = 0.11034/0.04166 = 2.649$$

From the F table or by Excel® function "=FINV(0.05,5,24)", the F critical value at d.f. 5 and 24 is found to be 2.621.

Since 2.649 > 2.621, we reject the null hypothesis H_o, i.e. the sample means from the 6 analysts do differ significantly.

We may want to ask what the reasons are for the difference. Is it one mean differed from all the others? Could all the means differ from each other? Might there be a case that the means fell into two distinct groups?

To answer this, we can use a simple way known as **the least significant difference method** to decide the reason for a significant difference. We shall discuss this method in the next blog.

In the meantime, by using the Excel Add-in "Data Analysis" "Anova: single factor", we have the following analysis results which match very well with the calculations from the first principle

Anova: Single Factor

SUMMARY

Analyst	Count	Sum	Average	Variance
A	5	422.78	84.556	0.01193
B	5	421.03	84.206	0.02323
C	5	422.05	84.41	0.0261
D	5	421.13	84.226	0.02908
E	5	420.89	84.178	0.09157
F	5	421.95	84.39	0.06805

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Analyst	0.551697	5	0.110339	2.648568	0.048192	2.620654
Within Analyst	0.99984	24	0.04166			
Total	1.551537	29				