An Illustration of One-way ANOVA by first principle

To understand how to go about evaluating sampling uncertainty and quality control of sampling, we must be familiar with the basic analysis of variance ANOVA principles.

ANOVA is a statistical technique which allows variations within sets of data to be isolated and estimated.

The simplest ANOVA is a one-way or one-factor ANOVA. It is applicable when data can be subdivided into groups according to a single factor such as analyst, laboratory or a particular experimental condition. It can be used both to test for significant differences between the groups and to estimate within- and between-group variance components.

Let's revisit these ANOVA principles by a simple illustration.

Three laboratory analysts in a training exercise were asked to conduct a series of six replicated analyses on a sample of flour for its protein content in %m/m using the classical Kjeldahl Nitrogen standard method. The results are summarized in Table 1 below.

Table 1: Test results of protein content (%m/m) in a sample of flour conducted by three analysts with means and variance calculated

	Analyst		
Repeat #	A	B	C
1	10.2	11.6	8.1
2	8.5	12.0	9.0
3	8.4	9.2	10.7
4	10.5	10.3	9.1
5	9.0	9.9	10.5
6	8.1	12.5	9.5
Mean $\bar{x}$	9.12	10.92	9.48
Variance s^2	1.006	1.702	0.962

In here, we have a single group (or factor) called Analyst and so we shall carry out a one-way ANOVA with our hypothesis statements as below:

Ho : $\bar{x}_1 = \bar{x}_2 = \bar{x}_3$
H1 : not all $\bar{x}$'s are equal
where $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ are the means for each analyst in this exercise.

The hypothesis test for this ANOVA comprises two types of variations from the samples analyzed, i.e. within- and between-groups 'partitioned' from the total variation in the whole data set.

Within Analysts variation

The first part is the variation of data analyzed by each of the Analyst i, and this variation reflects the random error in the analysis as reflected in its standard deviation, s_i and variance, s_i^2, and evaluated by using the following sum of squares within (SSw) (or also known as error sum of squares SSE) equation:

$$SS_w = \sum_{i=1}^{k} (n_i - 1) s_i^2$$

where k is the number of analysts (or levels) and n_i is the number of replicates of Analyst i. Hence, we have

SS_w = (6-1)(1.006) + (6-1)(1.702) + (6-1)(0.962) = 18.345

The degree of freedom (df) is $k(n-1)$ where n is $\sum n_i$, equal to 3(6-1) = 15

Therefore, the mean square within Analysts, $MS_w = \frac{SS_w}{df}$ = 1.223

Between Analysts variation

The second partition is the variation among the analysts, which is known as sum of squares between (SS_b) by equation:

$$SS_b = \sum_{i=1}^{k} n_i (\overline{x}_i - \overline{\overline{x}})^2$$

where $\overline{\overline{x}}$ is the grand mean value of all the observations which is (9.12+10.92+9.48)/3 = 9.83

Therefore, we have

$SS_b = 6(9.12-9.83)^2 + 6(10.92-9.83)^2 + 6(9.48-9.83)^2 = 10.858$

The degree of freedom for between Analysts is $k-1$, equal to (3-1) = 2

Therefore, the mean square between Analysts $MS_b = \frac{SS_b}{df}$ = 5.429

F-statistic test

$F = \frac{MS_b}{MS_w} = \frac{5.429}{1.223}$=4.439 which is greater than the $F_{(\alpha=0.05,\ v=2, v=15)}$ critical value of 3.682 with a probability P value of 0.031.

Therefore it is concluded that mean values of all the 3 Analysts were significantly different.

The MS Excel® spreadsheet results calculated by its Data Analysis tools are presented below (Table 2). The outcomes are exactly the same as stated above.

Table 2: Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Analyst A	6	54.7	9.117	1.006
Analyst B	6	65.5	10.917	1.702
Analyst C	6	56.9	9.483	0.962

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Analysts	10.858	2	5.429	4.439	0.031	3.682
Within Analysts	18.345	15	1.223			
Total	29.203	17				