One Way ANOVA – A Worked Example

Here is a very simple example that I created to show you one-way ANOVA. Remember that ANOVA stands for the **AN**alysis **O**f **VAR**iance. We will be analyzing the variances of three groups (both within each group and between the groups) to see if there is evidence that the theoretical or population means of the groups are equal. The way that we estimate the variances we will be using is to take the data subtract the sample mean, square the result, add thing up and then divide by the "right" number.

Assume there are three groups have the following data

t	x	y	z
1	2	3	2
2	4	0	1
3	4	3	1
4	2	2	0

This is certainly a small sample. There are only K = 3 groups and only 4 observations for each group, meaning N = 12 observations altogether. However, our method of one-way ANOVA works for any N and any K. Of course, to have a good statistical outcome you will probably want N to be large, say N = 100. This will depend on the situation.

Next, calculate the sample means for each group. This gives $\bar{x} = 3, \bar{y} = 2,$ and $\bar{z} = 1$. We want to use ANOVA to test the hypothesis that

$$H_o\text{: } \mu_X = \mu_Y = \mu_Z$$

which says that the theoretical or population means are all equal. Obviously, we do not mean the sample means. We know they are unequal, since 3≠2≠1. Instead, we want to test that the unknown and unknowable means for random variables X, Y, and Z are equal. Note we cannot use the t-test since that only allows us to test for equality of means in two samples, not three or more. The null hypothesis above actually involves two restrictions: $\mu_X = \mu_Y$ and $\mu_Y = \mu_Z$. These two logically imply that $\mu_X = \mu_Z$ so it is not an independent restriction.

Now, let's compute the sums of squares in different ways within and between groups.

(1) Get the grand mean which is merely the average over the whole data set.

This mean is 2 = {2+ 4+ 4+ 2+ 3+ 0+ 3+ 2+ 2+ 1+ 1+ 0}/12 an call it G = 2.

(2) Get the sums of squares between. This is

$$\mathbf{SSB} = \{4(\bar{x}-G)^2 + 4(\bar{y}-G)^2 + 4(\bar{z}-G)^2\} = 4*1 + 4*0 + 4*1 = 8 \quad \text{Thus, } \mathbf{SSB = 8.}$$

(3) Get the sums of squares within. This is

SSW = SS for Group X + SS for Group Y + SS for Group Z

$$= \{(2-3)^2 + (4-3)^2 + (4-3)^2 + (2-3)^2\} +$$

$$\{(0-2)^2 + (3-2)^2 + (2-2)^2 + (2-2)^2\} +$$

$$\{(2-1)^2 + (1-1)^2 + (1-1)^2 + (0-1)^2\} = 12 \quad \text{Thus, } \mathbf{SSW = 12.}$$

(4) Get the numerator degrees of freedom $df_N = K - 1$ Thus, $df_N = 3-1 = 2$

Get the denominator degrees of freedom $df_D = N - K$ Thus, $df_D = 12 - 3 = 9$

(5) Now compute our test statistic assuming the null is true.

$$\mathbf{F_{df_N, df_D} = F_{2,9} = \frac{SSB / df_N}{SSW / df_D} = \frac{8/2}{12/9} = 72/24 = 3}$$

(6) The pdf of an $F_{2,9}$ looks like the following (using wolfram alpha) Th p-value for $F_{2,9} = 3$ is p = 0.1004 which is significant at the 10% level but not at the 5% level. We therefore say that we can not reject H_o at the 5% level. The means look equal.