DOE – Two-Factor Factorial Design with replication

Two-factor (or two-way) factorial design is the simplest factorial design and has two or more levels for each factor of interest. To simplify discussion, we shall restrict to only two levels for each factor treatment. Designs which contain more than two levels of a factor are logical extensions of the two-level case. In here, we will also only consider the case where there are equal numbers of replicates (n) for each combination of the levels of factor *A* with those of factor *B*. Assumptions of normality and equal variance are taken valid here without outlier data and this also leads to assume that the error variables are independent.

This factorial design is sometimes called a 2^2 or 2x2 (read as 2 by 2) design. When the "2x2" terminology is used, the first number refers to the number of levels of the first factor, whilst the second number refers to the number of levels of the second factor. On the other hand, if "2^2" terminology is used, the first number refers to 2 levels whilst the power of 2 refers to the number of factors. In this manner, we have $2^2 = 4$ treatment combinations of factor A and factor B.

It is clear that as the number of levels of each factor increases and the number of replications in each factor also increases, we will encounter quite a complex calculation. Hence, it is assume that in practice, a statistical software such as SPSS, MINITAB or Excel spreadsheet package will be used when analyzing data from such experimental design models.

Let us discuss a conceptual approach for the decomposition of the total variation for the two-factor factorial design model with equal replication.

First of all, we need to define the following terms:

r = the number of levels of factor *A*
c = the number of levels of factor *B*
n' = the number of values in replication for each cell
n = the total number of observations in the experiment, i.e $n = rcn'$
X_{ijk} = the value of the k^{th} observation for level i of factor A and level j of factor *B*

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n'} X_{ijk}}{rcn'}$$ is the overall or grand mean

$$\bar{X}_i = \frac{\sum_{j=1}^{c}\sum_{k=1}^{n'} X_{ijk}}{cn'}$$, which is the mean of the ith level of factor *A* (where i = 1,2, …, r)

$$\bar{X}_i = \frac{\sum_{i=1}^{r}\sum_{k=1}^{n'} X_{ijk}}{rn'}$$, which is the mean of the jth level of factor *B* (where j = 1.2. …., *c*)

$$\bar{X}_{ij} = \sum_{k=1}^{n'} \frac{X_{ijk}}{n'}$$, which is the mean of the cell ij, the combination of the jth level of factor *A* and the jth level of factor *B*

In our previous blogs, we have noted that in the completely randomized design model,

SST (Sum of squares total or total variation) = SSTr (Sum of squares among treatments or groups) + SSE (Sum of squares within treatments or group for random error)

and in the randomized block design model, we have:

SST (Sum of squares total) = SSTr (Sum of squares among treatments or groups) + SSB (Sum of squares due to blocks) + SSE (Sum of squares of random error within blocks)

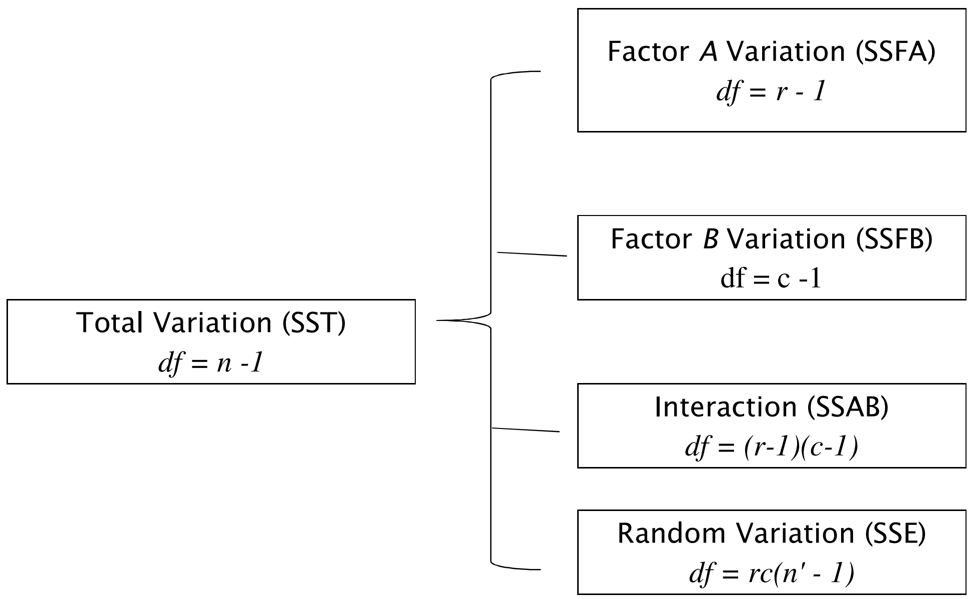
Similarly in this two-factor factorial design model with equal replication in each cell, we will have:

SST (Sum of squares total) = SSTr (Sum of squares among treatments or groups) + SSFA (Sum of squares due to factor A) + SSFB (Sum of squares due to factor B +SSE (Sum of squares due to inherent random error)

This decomposition of the total variation (SST) is best summarized in the following display:

Partitioning the Total Variation

SST = SSFA + SSFB + SSAB + SSE



The total variation for factors *A* and *B* is:

$$SST = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n'} (X_{ijk} - \bar{\bar{X}})^2$$

The sum of squares due to factor *A* (or SSFA) represents the differences between the various levels of factor *A* and the grand mean. SSFA is computed by the following equation:

$$SSFA = cn' \sum_{i=1}^{r} (\bar{X}_i - \bar{\bar{X}})^2 \qquad [1]$$

Similarly the sum of squares due to factor *B* (or SSFB) represents the differences between the various levels of factor *B* and the grand mean. SSFB is computed by the following equation:

$$SSFB = rn'\sum_{j=1}^{c}(\overline{X}_j - \overline{\overline{X}})^2 \quad [2]$$

The sum of squares due to the effect of the interaction between *A* and *B* (or SSAB) represents the effect of the combinations of levels of factor *A* and factor *B*. SSAB is computed by the following equation:

$$SSAB = n'\sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij} - \overline{X}_i - \overline{X}_j + \overline{\overline{X}})^2 \quad [3]$$

And, lastly the sum of square error (or SSE) represents the differences among the observations within each cell and the corresponding cell mean. SSE is computed by the following equation:

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n'}(X_{ijk} - X_{ij})^2 \quad [4]$$

The respective degrees of freedom of components are listed in the above figure.

If each of the sums of squares is divided by its associated degrees of freedom, we obtain the four mean squares (MSFA, MSFB, MSAB, and MSE) as shown by the following equations:

$$MSFA = \frac{SSFA}{r-1} \quad [5]$$

$$MSFB = \frac{SSFB}{c-1} \quad [6]$$

$$MSAB = \frac{SSAB}{(r-1)(c-1)} \quad [7]$$

$$MSE = \frac{SSE}{rc(n'-1)} \quad [8]$$

So, with these equations in mind, we can proceed to build a two-factor ANOVA model which calls for three distinct tests to be performed.

If we assume that the levels of factor A and factor B have been specifically selected for analysis (rather than being randomly selected from a population of possible levels), we can make the following three tests of hypotheses:

1. To test the hypothesis of no difference due to factor A

$H_o : \mu_1 = \mu_2 = \ldots. = \mu$
H_1 : not all μ_i are equal

We perform a F–test statistic and obtain:

$$F = \frac{MSFA}{MSE} \qquad [9]$$

The null hypothesis is rejected at the α level of significance if

$$F = \frac{MSFA}{MSE} > F_c \qquad [10]$$

the upper–tail critical value from an F distribution with (r-1) degrees of freedom in the numerator and $rc(n'-1)$ degrees of freedom in the denominator.

2. To test the hypothesis of no difference due to factor *B*

$H_o : \mu_1 = \mu_2 = \ldots. = \mu$
H_1 : not all μ_j are equal

We perform a F–test statistic and obtain:

$$F = \frac{MSFB}{MSE} \qquad [11]$$

The null hypothesis is rejected at the α level of significance if

$$F = \frac{MSFB}{MSE} > F_c \qquad [12]$$

the upper–tail critical value from an F distribution with (c-1) degrees of freedom in the numerator and $rc(n'-1)$ degrees of freedom in the denominator.

3. To test the hypothesis of no interaction of factors A and B

H_o : The interaction of *A* and *B* = 0

H_1 : The interaction of *A* and $B \neq 0$

We perform a F-test statistic and obtain:

$$F = \frac{MSAB}{MSE} \qquad [13]$$

The null hypothesis is rejected at the α level of significance if

$$F = \frac{MSAB}{MSE} > F_c \qquad [14]$$

the upper-tail critical value from an F distribution with $(r-1)(c-1)$ degrees of freedom in the numerator and $rc(n'-1)$ degrees of freedom in the denominator.

In summary, the two-factor ANOVA with replication can be presented in the table below:

Source	Degrees of freedom df	Sum of squares SS	Mean square (Variance) MS	F value
A	$r-1$	$SSFA$	$MSFA = \frac{SSFA}{r-1}$	$\frac{MSFA}{MSE}$
B	$c-1$	$SSFB$	$MSFB = \frac{SSFB}{c-1}$	$\frac{MSFB}{MSE}$
AB	$(r-1)(c-1)$	$SSAB$	$MSAB = \frac{SSAB}{(r-1)(c-1)}$	$\frac{MSAB}{MSE}$
Error	$rc(n'-1)$	SSE	$MSE = \frac{SSE}{rc(n'-1)}$	
Total	$n-1$	SST		

We shall discuss an example for the two-factor (two-way) model with replication in the next blog.