Analysis of Variance

LECTURE - 3

FACTORIAL EXPERIMENTS

Dr. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

2^2 Factorial Experiment :

Factors : A, B

Two levels of A, they are denoted as a_0 and a_1 or 0 and 1

Two levels of B are represented as b_0 and b_1 or 0 and 1

2^2 Factorial Experiment : 4 Treatment combinations

$$a_0 b_0 \equiv 0 \;\; 0 \equiv I$$

$$a_0 b_1 \equiv 0 \;\; 1 \equiv a$$

$$a_1 b_0 \equiv 1 \;\; 0 \equiv b$$

$$a_1 b_1 \equiv 1 \;\; 1 \equiv ab.$$

Sometimes 0 is referred to as 'low level' and 1 is referred to 'high level'.

I denotes that both factors are at lower levels (a_0 b_0) or (0 0).

This is called as **control treatment**

Treating (ab) as $(a)(b)$ symbolically (mathematically and conceptually, it is incorrect), we can now express all the main effects, interaction effect and general mean effect as follows:

$$\text{Main effect of } A = \frac{(a)+(ab)}{2} - \frac{(1)+(b)}{2}$$

$$= \frac{1}{2}\left[(ab)-(b)+(a)-(1)\right]$$

$$= \frac{(a-1)(b+1)}{2}$$

$$\text{Main effect of } B = \frac{(b)+(ab)}{2} - \frac{(1)+(a)}{2}$$

$$= \frac{1}{2}\left[(ab)-(a)+(b)-(1)\right]$$

$$= \frac{(a+1)(b-1)}{2}$$

Interaction effect of A and $B = \frac{(ab)-(b)}{2} - \frac{(a)-(1)}{2}$

$$= \frac{1}{2}\left[(ab)-(a)+(1)-(b)\right]$$

$$= \frac{(a-1)(b-1)}{2}$$

General mean effect $(M) = \frac{(1)+(a)+(b)+(ab)}{4}$

$$= \frac{1}{4}\left[(1)+(a)+(b)+(ab)\right]$$

$$= \frac{(a+1)(b+1)}{4}.$$

Main effect of $A = \frac{(a-1)(b+1)}{2}$

Main effect of $B = \frac{(a+1)(b-1)}{2}$

Interaction effect of A and $B = \frac{(a-1)(b-1)}{2}$

General mean effect $(M) = \frac{(a+1)(b+1)}{4}$.

- Notice the roles of + and – signs as well as the divisor.
- There are two effects related to A and B.

Main effect of $A = \frac{(a-1)(b+1)}{2}$

Main effect of $B = \frac{(a+1)(b-1)}{2}$

Interaction effect of A and $B = \frac{(a-1)(b-1)}{2}$

General mean effect $(M) = \frac{(a+1)(b+1)}{4}$.

- To obtain the effect of a factor, write the corresponding factor with – sign and others with + sign.

For example, in the main effect of A, *a* occurs with – sign as in $(a - 1)$ and b occurs with + sign as in $(b + 1)$.

Main effect of $A = \frac{(a-1)(b+1)}{2}$ Main effect of $B = \frac{(a+1)(b-1)}{2}$

Interaction effect of A and $B = \frac{(a-1)(b-1)}{2}$ General mean effect $(M) = \frac{(a+1)(b+1)}{4}$.

•In AB, both the effects are present so a and b both occur with + signs as in $(a+1)(b+1)$.

Main effect of $A = \frac{(a-1)(b+1)}{2}$

Main effect of $B = \frac{(a+1)(b-1)}{2}$

Interaction effect of A and $B = \frac{(a-1)(b-1)}{2}$

General mean effect $(M) = \frac{(a+1)(b+1)}{4}$.

•Also note that the main and interaction effects are obtained by considering the typical differences of averages, so they have divisor 2 whereas general mean effect is based on all the treatment combinations and so it has divisor 4.

• There is a well defined statistical theory behind this logic but this logic helps in writing the final treatment combination easily.

$$A = \frac{1}{2}\left[(ab) + (a) - b - (1)\right] \qquad AB = \frac{1}{2}\left[(ab) + (1) - a - (b)\right]$$

$$B = \frac{1}{2}\left[(ab) + (b) - a - (1)\right] \qquad M = \frac{1}{4}\left[(ab) + (a) + (b) + (1)\right].$$

Factorial effects	Treatment combinations				Divisor
	(1)	(a)	(b)	(ab)	
M	+	+	+	+	4
A	-	+	-	+	2
B	-	-	+	+	2
AB	+	-	-	+	2

The model corresponding to 2^2 factorial experiment is

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk}, \ i = 1,2, \ j = 1,2, \ k = 1,2,\ldots,n$$

where *n* observations are obtained for each treatment combinations.

When the experiments are conducted factor by factor, then much more resources are required in comparison to the factorial experiment.

For example, if we conduct *RBD* for three level of nitrogen N_0, N_1 and N_2 and two levels of irrigation I_0 and I_1, then to have 10 degrees of freedom for the error variance, we need

- 6 replications on nitrogen,
- 11 replications on current.

So total number of plots needed are 40.

For the factorial experiment with 6 combinations of 2 factors, total number of plots needed are 18 for the same precision.

We have considered up to now by assuming only one observation for each treatment combination, i.e., no replication.

If *r* replicated observations for each of the treatment combinations are obtained, then the expressions for the main and interaction effects can be expressed as

$$A = \frac{1}{2r}\left[(ab) + (a) - b - (1)\right]$$

$$B = \frac{1}{2r}\left[(ab) + (b) - a - (1)\right]$$

$$AB = \frac{1}{2r}\left[(ab) + (1) - a - (b)\right]$$

$$M = \frac{1}{4r}\left[(ab) + (a) + (b) + (1)\right].$$

Now we detail the statistical theory and concepts related to these expressions.

Let Y_* =((1), *a*, *b*, *ab*)' be the vector of total response values.

Then

$$A = \frac{1}{2r}\ell'_A Y_* = \frac{1}{2r}(-1 \quad 1 \quad -1 \quad 1)Y_*$$

$$B = \frac{1}{2r}\ell'_B Y_* = \frac{1}{2r}(-1 \quad -1 \quad 1 \quad 1)Y_*$$

$$AB = \frac{1}{2r}\ell'_{AB} Y_* = \frac{1}{2r}(1 \quad -1 \quad -1 \quad 1)Y_* \ .$$

Note that A, B and AB are the linear contrasts.

Note that A, B and AB are the linear contrasts.

Recall that a linear parametric function is estimable only when it is in the form of linear contrast.

Moreover, A, B and AB are the linear orthogonal contrasts in the total response values $(1), a, b, ab$ except for the factor $1/2r$.

The sum of squares of a linear parametric function $\ell' y$ is given by $\dfrac{(\ell' y)^2}{\ell' \ell}$.

If there are r replicates, then the sum of squares is $\dfrac{(\ell' y)^2}{r\ell' \ell}$.

It may also be recalled under the normality of *y*'s, this sum of squares has a Chi-square distribution with one degree of freedom (χ_1^2). Thus the various associated sum of squares due to A, B and AB are given by following:

$$SSA = \frac{(\ell'_A Y_*)^2}{r\ell'_A \ell_A} = \frac{1}{4r}(ab + a - b - (1))^2$$

$$SSB = \frac{(\ell'_B Y_*)^2}{r\ell'_B \ell_B} = \frac{1}{4r}(ab + b - a - (1))^2$$

$$SSAB = \frac{(\ell'_{AB} Y_*)^2}{r\ell'_{AB} \ell_{AB}} = \frac{1}{4r}(ab + (1) - a - b)^2.$$

Each of *SSA*, *SSB* and *SSAB* has χ_1^2 under normality of *y*.

The sum of squares due to total is computed as usual

$$TSS = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{r} y_{ijk}^2 - \frac{G^2}{4r} \quad \text{where} \quad G = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{r} y_{ijk}$$

is the grand total of all the observations.

The *TSS* has χ^2 distribution with $(2^2 r - 1)$ degrees of freedom

The sum of squares due to error is also computed as usual as

$$SSE = TSS - SSA - SSB - SSAB$$

which has χ^2 distribution with

$$(4r-1)-1-1-1 = 4(r-1)$$

degrees of freedom.

The *F*-statistic corresponding to A, B and AB are

$$F_A = \frac{SSA/1}{SSE/4(r-1)} \sim F(1, 4(r-1) \text{ under } H_0,$$

$$F_B = \frac{SSB/1}{SSE/4(r-1)} \sim F(1, 4(r-1) \text{ under } H_0,$$

$$F_{AB} = \frac{SSAB/1}{SSE/4(r-1)} \sim F(1, 4(r-1) \text{ under } H_0.$$

The ANOVA table is case of 2^2 factorial experiment is given as follows:

Source	Sum of squares	Degrees of freedom	Mean squares	F
A	SSA	1	$MSA = SSA/1$	$F_A = \frac{MSA}{MSE}$
B	SSB	1	$MSB = SSB/1$	$F_B = \frac{MSB}{MSE}$
AB	$SSAB$	1	$MSAB = SSAB/1$	$F_{AB} = \frac{MSAB}{MSE}$
$Error$	SSE	$4(r-1)$	$MSE = SSE/4(r-1)$	
Total	TSS	$4r-1$		

The decision rule is to reject the concerned null hypothesis

when the value of concerned *F* statistic

$$F_{effect} > F_{1-\alpha}(1, 4(r-1)).$$