Two way ANOVA:

In this session we discuss Two way Analysis of variance model for fixed effects . The entire topic is divided into the following subdivisions.

1. Two way layout
2. Model description
3. Decomposition of the total sum of squares
4. Statistical Analysis
5. ANOVA Table
6. Comparing pairs of treatment means and block means
7. Conclusion

1. Two way layout

Consider the following example: The yield of milk may be affected by different treatments(diet), as well as the different breed of cows. Suppose that n cows are divided into k different groups according to their breed and each breed containing p cows and p diets are given to these cows in each group. Now our interest is to see the effect of p different type of diet(food) given at random to these cows on the yield of milk. Here the response variable (yield of milk) is affected by two factors A(diet) and B(breed). Such an experiment is called two factor experiment or (two way classification).

Let y_{ij} represent yield of milk from the cow of the jth breed fed with diet (food i),i= 1….p, j=1…k

The data can be tabulated as follows.

	Breed of cows				Totals	Averages
Diet	1	2		k		
1	y_{11}	Y_{12}	…	Y_{1n}	$y_{1.}$	$\overline{y_{1.}}$
2	y_{21}	Y_{22}.	…	Y_{2n}	$y_{2.}$	$\overline{y_{2.}}$
.	.	.	…	.	.	.
..	.	.	…	.	.	.

.	.	.	…	.	.	.
p	y_{p1}	y_{p2}	…	y_{pn}	$y_{p.}$	$\overline{y_{p.}}$
Total	y.1	y.2	…	y.n	y..	
Averages	$\overline{y_{.1}}$	$\overline{y_{.2}}$	$\overline{y_{.3}}$	$\overline{y_{.n}}$		

Let $y_{i.}$ represent the total of the observations and $\overline{y_{i.}}$ average of observations under the ith level of factor A(generally called treatments), Let $y_{.j}$ represent the total of the observations and $\overline{y_{.j}}$ average of observations under the jth level of factor B(generally referred as Blocks or replicates or groups). Similarly $y_{..}$ represent the grand total of all the observations and $\overline{y_{..}}$ represent grand average of all the observations, and are expressed as follows.

$$y_{i.} = \sum_{j=1}^{k} y_{ij}, \qquad \overline{y_{i.}} = \frac{y_{i\cdot}}{k}, \quad i=1,2\ldots p$$

$$y.j = \sum_{i=1}^{p} y_{ij}, \qquad \overline{y.j} = \frac{y.j}{p}, \quad j=1,2\ldots k$$

$$y_{..} = \sum_{i=1}^{p} \sum_{j=1}^{k} y_{ij}, \quad \overline{y_{..}} = \frac{y_{.\cdot}}{N},$$ where N is total number of observations N= p *x* k

Y_{ij} represent the jth observation taken from treatment i.

2. Model description:

We define the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \tag{4}$$

Where μ general effect

α_i – ith treatment effect;

βj – jth block effect;

ϵij- error term which is independently and identically distributed random variables with mean 0 and variance σ^2

Observe that the response variable y^{ij} is a liner function of the model parameters. Equation (4) is called the two way analysis of variance model because two factors namely treatments and blocks whose effects on the response is investigated.

3. **Decomposition of the total sum of squares**:

$$\sum_{i=1}^{p}\sum_{j=1}^{k}(y_{ij}-\overline{y}_{..})^2=\sum_{i=1}^{p}\sum_{j=1}^{k}((y_{ij}-\overline{y_{i.}}-\overline{y_{.j}}+\overline{y_{..}})+(\overline{y_{i..}}-\overline{y_{..}})+(\overline{y_{..j.}}-\overline{y_{..}}))^2$$

$$=k\sum_{i=1}^{p}(\overline{y_{i.}}-\overline{y_{..}})^2+p\sum_{j=1}^{k}(\overline{y_{.j}}-\overline{y_{..}})^2+\sum_{i=1}^{p}\sum_{j=1}^{k}(y_{ij}-\overline{y_{i.}}-\overline{y_{.j}}+\overline{y_{..}})^2 + 0$$

Observe that all the cross products becomes zero 6

Or

SST = SSTR+ SSB+ SSE where

SST = $\sum_{i=1}^{p}\sum_{j=1}^{k}(y_{ij}-\overline{y_{..}})^2$ (Total sum of squares)

= $\sum_{i=1}^{p}\sum_{j=1}^{k}y_{ij}^2-\frac{y_{..}^2}{N}$ (simplified formula)

SSTR = $k\sum_{i=1}^{p}(\overline{y_{i.}}-\overline{y_{..}})^2$ (sum of squares due to treatments)

= $\sum_{i=1}^{p}\frac{y_{i.}^2}{k}-\frac{y_{..}^2}{N}$

SSB = $p\sum_{j=1}^{k}(\overline{y_{.j}}-\overline{y_{..}})^2$ (Sum of squares due to groups)

= $\sum_{i=1}^{k}\frac{y_{.j}^2}{p}-\frac{y_{..}^2}{N}$

SSE = $\sum_{i=1}^{p}\sum_{j=1}^{k}(y_{ij}-\overline{y_{i.}}-\overline{y_{.j}}+\overline{y_{..}})^2$ (Sum of squares due to error)

= SST-SSTR-SSB

The total corrected sum of squares can be partitioned into a sum of squares of the differences between the treatment averages and the grand average, plus , a sum of squares of the differences between the block averages and the grand average plus a sum of squares due to random error.

4. Statistical Analysis

Here we are interested in testing the equality of p treatment means as well as equality of k group means

Hence the appropriate hypotheses is

1.H_A: $\alpha_1 = \alpha_2 = \ldots \alpha_p = 0$

And the alternative hypothesis is

H_{A1}: atleast one $\alpha_i \neq \alpha_j$ for all i, j

2. H_B: $\beta_1 = \beta_2 = \ldots \beta_p = 0$

And the alternative hypothesis is

H_{B1}: atleast one $\beta_i \neq \beta_j$ for all i, j

Degrees of freedom of various sum of squares:

The TSS is computed with N=pxk quantities, which will carry N-1 degree of freedom, SSTR will have p-1 degree of freedom, SSB will have K-1 degree of freedom since

$$\sum_{i=1}^{p} \sum_{j=1}^{k} (y_{ij} - \overline{y_{..}}) = \sum_{i=1}^{p} (y_{i.} - \overline{y_{..}}) = \sum_{j=1}^{k} (y_{.j} - \overline{y_{..}}) = 0$$

and SSE will carry (N-1) – (p-1) –(k-1) = (p-1) (k-1) degrees of freedom

We now investigate a formal test of the hypothesis

1. H_A: $\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$ And
2. H_B: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$

Against the respective alternatives

We have assumed that the errors ϵ_{ij} are normally and independently distributed with mean zero and variance σ^2. The observations y_{ij} are normally and independently distributes with mean $\mu + \alpha i + \beta j$ and variance σ^2. Thus SST is a sum of squares in normally distributed random variables hence can be shown that SST/σ^2 is distributes as chi-square with N-1 degrees of freedom, and that SSTR $/\sigma^2$ is chi - square variate with p-1 degrees of freedom, SSB/σ^2 is chi - square variate with k-1 degrees of freedom if the null hypothesis true. Further we can shown that SSE/σ^2 is chi-square (k-1) x (p-1) degrees of freedom. It also implies that SSTR/ σ^2, SSB/σ^2 and SSE/σ^2 independently distributed chi-square random variables. Therefore if the null hypothesis H_A is true, the ratio

$$F_A = \frac{SSTR/(p-1)}{SSE/(p-1)(k-1)} = \frac{MSSTR}{MSSE} \tag{8}$$

is distributed as F with p-1 and (p-1)(k-1) degrees of freedom.

. Similarly if the null hypothesis H_B is true, the ratio

$$F_B = \frac{SSB/(K-1)}{SSE/(p-1)(k-1)} = \frac{MSSB}{MSSE} \tag{9}$$

is distributed as F with k-1 and (p-1)(k-1) degrees of freedom.

Equation (8) is the test statistics to test for H_A i.e.no differences in treatment means.

Equation (9) is the test statistics to test for H_B i.e.no differences in group means

5. Analysis of variance Table(ANOVA Table)

Sources of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F-ratio
Treatments	p-1	SSTR	MSSTR = SSTR/(p-1)	F_A = MSSTR/ MSSE
replicates	k-1	SSB	MSSB = SSB/(k-1)	F_B = MSS**B**/ MSSE
Error	(p-1) (k-1)	SSE	MSSE = SSE/(p-1) (k-1)	
total	N-1	TSS		

We reject H_A and conclude that there are differences in the treatment means if $F_A > F_{\alpha,p-1,(p-1)(k-1)}$ Where F_A is computed from equation 8 and $F_{\alpha, p-1, (p-1)(k-1)}$, is the table value referring to F table at α level significance corresponding to p-1and (p-1)(k-1)degrees freedom.

We reject H_B and conclude that there are differences in the group means if $F_B > F_{\alpha, k-1, (p-1)(k-1)}$ Where F_B is computed from equation 9 and $F_{\alpha, k-1, (p-1)(k-1)}$, is the table value referring to F table at α level significance corresponding to k-1and (p-1)(k-1)degrees of freedom.

6. **Comparing the pairs of treatment means and block means:**

Case of rejection of H_A.:

if $F_A > F_{\alpha, p-1, (p-1)(k-1)}$, then H_A: $\alpha_1 = \alpha_2 = \ldots \alpha_p = 0$ is rejected. This means that at least one α_i is different from other effects which is responsible for the rejection of the H_A. Now our interest is to find out such α_i and divide the population into groups such that the means of the population within the group are same. This can be done by pairwise testing of α_i 's.

i.e. we test H_{0A}: $\alpha_i = \alpha_j$, (i≠j) against H_{1A}: $\alpha i \neq \alpha j$

This can be tested using the following t-statistics:

T = $\dfrac{\overline{y_{i.}} - \overline{y_{j.}}}{\sqrt{\dfrac{2MESS}{k}}}$ which follows t distribution with (p-1)(k-1) degrees of freedom under H_{0A}.

Thus the decision rule to reject H_{0A} at α if the observed difference

$\overline{y_{i.}} - \overline{y_{j.}} > t_{\alpha,(p-1)(k-1)} \sqrt{\dfrac{2MESS}{k}}$ The quantity $\overline{y_{i.}} - \overline{y_{j.}} > t_{\alpha,(p-1)(k-1)} \sqrt{\dfrac{2MESS}{k}}$ is called critical difference.

Similarly Case of rejection of H_β :

If $F_B > F_{\alpha, k-1, (p-1)(k-1)}$, then H_B: $\beta_1 = \beta_2 = \ldots \beta_p = 0$is rejected. This means that at least one β_j is different from other effects which is responsible for the rejection of the H_B. Now our interest is to find out such β_j and divide the population into groups such that the means of the population within the group are same. This can be done by pairwise testing of β_j's.

i.e. we test H_{0B}: $\beta_i = \beta_j$, (i≠j) against H_{1B}: $\beta_i \neq \beta_j$

This can be tested using the following t-statistics:

T = $\dfrac{\overline{y_{.i}} - \overline{y_{.j}}}{\sqrt{\dfrac{2}{p}MESS}}$ which follows t distribution with (p-1)(k-1) degrees of freedom under

H_{0B}.

Thus the decision rule to reject H_{0B} at α if the observed difference

$\overline{y_{i.}} - \overline{y_{.j}} > t_{\alpha,(p-1)(k-1)} \sqrt{\frac{2MESS}{p}}$ The quantity $\overline{y_{i.}} - \overline{y_{.j}} > t_{\alpha,(p-1)(k-1)} \sqrt{\frac{2MESS}{p}}$ is called critical difference.

7. **Conclusion**: We have discussed two way layout with an example. We have presented the two way analysis of variance model, stated the hypothesis to be tested and discussed the analysis . The entire analysis procedure we summarized in ANOVA Table.. We have discussed procedure to Compare pairs of treatment means when null hypothesis for treatments is rejected. We have also discussed procedure to Compare pairs of block means when null hypothesis stated for blocks is rejected **..**