RESEARCHERID
THOMSON REUTERS

[IDSTM: January 2017]
IC™ Value: 3.00

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A COMPARATIVE STUDY USING ANALYSIS OF VARIANCE

Anu Sayal*, Rohit Kandari

* Department of Mathematics, Uttaranchal University. Arcadia grant
P.O. Chandanwari, Prem Nagar. Dehradun-248001. Uttarakhand
Department of Mechanical Engineering, Uttaranchal University. Arcadia grant. P.o. Chandanwari, prem nagar. Dehradun-248001. Uttarakhand

ABSTRACT

In the present paper application of ANOVA has been discussed. A brief introduction of ANOVA & its types has been shown which has been portrayed through suitable examples focusing on the applicability of ANOVA. ANOVA has widespread application in all spheres. It is a powerful statistical tool used for making comparisons and drawing inferences. In this paper we have discussed ANOVA study by using relevant examples.

KEYWORDS: ANOVA , One Way ANOVA , Two Way ANOVA , Mean Square , Randomization , Replication

INTRODUCTION

ANOVA is used for testing more than two treatments at a time using F-test. In this method the total variation in the experimental data is compared with experimental error variation with F-test. If F-test shows significant variation. ANOVA is a powerful statistical procedure for determining if differences in means are significance and for dividing the variance into components. This technique is used in the production and manufacturing of the products in manufacturing firms, where different of factors affecting in different ways are consider and their combined study is presented out.

Meaning and definition of anova:

Analysis of variance is a process of testing simultaneous significance of the differences among the means of several categories (often called classes or groups) . as in mechanical way of approach we just compare and layed the experimental results to get a accurate and précised result so we can more and more benefit with less input. If are comparing one category with a single criteria concept of two way and for more criterias and categories we use two way anova concept.

Concepts used in anova techniques:

- Mean square: the measure of variability used in the analysis of variance is called mean square. Sum of square deviation from mean divided by degrees of freedom.
  {Mean square = sum of squared deviation from mean / degree of freedom }
- Randomization: random allocation of treatments to experimental units is called randomization. The allotment of treatments to experimental units can be done with the help of random number table. It ensures the validity of statistical tests like F-test, t-test,etc,it also provides assumptions of independence of experimental errors and assumption of normality of experimental errors.
- Replication: repetition of treatments to experimental units is known as replication. Replication of treatment reduces experimental errors and also provides experimental or residual variance which is essential for carrying out F-test.

Techniques for analysis of variance:

- One way ANOVA: if one factor is used in testing levels of that factor than it is called one way anova. For example when different modifications on same parts of machines area implemented on a single device then it known as one way classification of anova.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

- Two ways ANOVA : when two factors are used in testing levels within these factors it is known as two way anova. For example, when two parts are tested at different levels of modification with respect to their efficiency oqn machine then two way analysis of variance is a useful tool.

Applications of anova

1. The main application of analysis of variance is to test the homogeneity of observations.
2. The significance of additional terms in a regression equation.
3. The curvy linearity or linearity of fitted regression.
4. The significance of co-relation ratio.
5. Significance of two or more 'principle of calcification'.
6. The significance in relation to multiple regression.

Limitations of method of anova

When the data given are in the form of proportions or percentages then conditions will not be fulfill and the analysis as such cannot be done. Also if during some field experimental error appear to be co-related we should check up on it. In case of one or more assumptions are not satisfied in the data the analysis can be improved by the omission of certain abnormal observation as the case may be or by sub division of the error variance or by proper choice of scale through transformation of data. It is applicable only when numbers of groups are three or more. If number is less than three we cannot apply anova technique.

Steps for one way anova

Calculate variance between the samples

The total variance from one class to another class is called the variance between classes . The sum of square of deviation between the samples is denoted by SSB or SSS or SSC.

The steps for calculating variance between samples are :-

a) calculate the mean of each sample i.e,. $\overline{x_1}$, $\overline{x_2}$, …..etc.
b) calculate the grand average $\overline{\overline{x_1}}$ of all these mean values i.e $\overline{\overline{x_1}} = \frac{\overline{x1} + \overline{x2} \ldots\ldots}{n1+n2+\cdots..}$
   where n_1, n_2 ….. are the no. of observations of classes.
c) take the difference between the means of the various samples and the grand average.
d) square these deviations and obtain the total which will give the sum of squares between the samples.
e) divide the total obtained in steps (d) by the degree of freedom. The degree of freedom will be 1 less than the no. of samples is $\upsilon_1 = c-1$, where c is the no. of samples.

Therefore, $SSS = [\frac{(\Sigma x_1)^2}{n} + \frac{(\Sigma x_2)^2}{n} + .. + \frac{(\Sigma x_n)^2}{n}] - \frac{T^2}{N}$

where , T is the total of all the observations and n is the no. of samples.

1. Correction factor: since the measurement are taken from the mean : Correction factor = $\frac{T^2}{N}$
2. Total sum of squares :- $TSS = \sum x_1^2 + \sum x_2^2 + \ldots\ldots\ldots\ldots \sum x_n^2 - \frac{T^2}{N}$

4. Calculate variance (or sum of square of deviation) within the sample :-
 The variations among the observation of each specific class is called internal variation and the sum total of the internal variation is called variance within classes. It is denoted by SSE or SSW.

a) calculate the mean values of each sample i,e, $\overline{x1}$, $\overline{x2}$,….. and so on.
b) take the deviation of the various items in a sample from the mean values of the respective sample.
c) square the deviation and obtain the total which gives the sum of squares within the sample.
d) divide this total obtained in steps by the degree of freedom. The degrees of freedom obtained by deducting the number of samples from the total number of items i.e, $\upsilon_2 = n - c$, where 'c' is the number of samples and 'n' is the total number of observations. Therefore, SSE (or SSW) = TSS – SSS (or SSB)
e) Now calculate mean square deviation between the samples (MSB)
   Dividing SSS(or SSB) by the degree of freedom gama$_1$ = c- 1
   [$MSB = SSB / \upsilon_1$]

In the same way calculate the mean square deviation within the sample
Dividing SSE(or SSW) by the degree of freedom , $\upsilon_2 = n - c$

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

[$MSW = SSW / \upsilon_2$]

f) calculation of F – statistic or F – ratio : -

Both MSB and MSW are independent unbiased estimates of the same population variances [$F = MSB / MSW$]

(Degree of freedom $\upsilon_1 = c - 1$ and $\upsilon_2 = n - c$)

In general MSB > MSW

If MSB < MSW then $F = MSW / MSB$

g) compare the calculated values of F with the translated values of F for the given degree of freedom at a certain critical levels. Generally , we take 5% level of significance.

If calculated value of F is greater than the tabulated value then reject the null hypothesis but if it is lesser null hypothesis is accepted.

ANALYSIS OF VARIANCE TABLE: ONE WAY CLASSIFICATION

Source of variance	SS(sum of squares)	V(degree of freedom)	MS(mean square)	Variance ratio of F
Between samples	SSB	$\upsilon_1 = c - 1$	MSB	
Within samples	SSW	$\upsilon_2 = n - c$	MSW	MSW
Total	TSS	n – 1		

Where

TSS or SST – Total sum of squares of variation

SSB – sum of squares between samples (column)

SSW – sum of squares within samples (row)

MSB – mean sum of squares between samples

MSW – mean sum of squares within samples

Steps for two way anova :-

Analysis of variation table: (two way classification)

Squares of variation	Sum of squares	Degree of freedom	Mean sum of squares, Ratio of F
Between samples	SSC	c – 1	MSC= SSC/(c-1)
Between rows	SSR	r – 1	MSR=SSR/(r-1)
Residual error	SSE	(c – 1)(r – 1)	MSE=SSE/(r-1)(c-1)
Total	SST	n – 1	

Where, SSC – sum of squares between column
SSR – sum of squares between rows
SSE – sum of squares due to error
SST - total sum of squares

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

The sum of squares for the source (error) “Residual” is obtained by subtracting from the total sum of squares and the sum of squares between column and rows i,e. SSE = SST – (SSC + SSR)
Total number of degree of freedom = n – 1 or cr – 1
Whereas ‘c’ refers to number of column and ‘r’ refers to number of rows

The number of degrees of freedom between columns = c – 1
The number of degrees of freedom between rows = r - 1
The number of degrees of freedom for residual (error) = (c – 1)(r – 1) or (n – 1) –[(c –1) +(r – 1)]
The F values are defined as F = MSC / MSE
The calculated values of F are compared with the tabulated values. If calculated vales of F is greater than the tabulated values at pre assigned values / level of significance , the null hypothesis is rejected otherwise accepted.

ILLUSTRATIVE EXAMPLES

Example 1: A production deptt of a manufacturing company manufactures four machines A,B,C,D and observe their sales in three seasons – summer , winter and monsoon. The figures (in lacs) are given in following table. carry out an analysis of variance.

SEASSONS	**MACHINES**				**SEASONS TOTAL**
	A	B	C	D	
SUMMER	36	36	21	35	128
WINTER	28	29	31	32	120
MONSOON	26	28	29	29	112
MACHINES TOTAL	90	93	81	96	360

Solution:
Null hypothesis : sales of machine and sales of season do not differ .
The above data is classified according to criteria
(i) Machines
(ii) Seasons
In order to simplify calculations we code the data by subtracting 30 from each values. The data in coded form is given below:

SEASSONS	MACHINES				SEASONS TOTAL
	A	B	C	D	
SUMMAR	6	6	-9	5	8
WINTER	-2	-1	1	2	0
MONSOON	-4	-2	-1	-1	-8
MACHINES TOTAL	0	3	-9	6	0

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Now , sum of all items of various samples T = 0

Correction factor $\frac{T^2}{N} = 0$

Sum of squares between machines $(SSB)_{mac} = \frac{0^2}{3} + \frac{3^2}{3} + \frac{(-9)^2}{3} + \frac{6^2}{3} - 0$
$= 0 + 3 + 27 + 12 - 0 = 42$

Number of degrees of freedom = c – 1 = 4 – 1 = 3

Sum of squares between seasons $(SSB)_{sn} = \frac{8^2}{4} + \frac{0^2}{4} + \frac{(-8)^2}{4} - \frac{T^2}{n} - 0$
$= 16 + 0 + 16 - 0 = 32$

Degrees of freedom = r – 1 = 3 – 1 = 2

Total sum of square(TSS) = $6 + (-2)^2 + (-4)^2 + 6^2 + (-1)^2 + (-2)^2 + (-9)^2 + 1^2 + (-1)^2 + 5^2 + (2)^2 + (-1)^2 - \frac{T^2}{N}$
$= 36 + 4 + 16 + 36 + 1 + 4 + 81 + 1 + 1 + 25 + 4 + 1 - 0$
$= 210$

Residual (error) = TSS – SSB(machine) – SSB(season)
= 210 – 42 – 32 = 136

Degree of freedom = (c – 1)(r – 1) = 3 × 2 = 6

ANALYSIS OF VARIANCE TABLE

SOURCE OF VARINCE	SS	Df	MEAN SQUARE
BETWEEN COLUMNS (MACHINE)	42	3	$\frac{42}{3} = 14$
BETWEEN ROWS (SEASON)	32	2	$\frac{32}{2} = 16$
RESIDUALS	136	6	$\frac{136}{6} = 22.67$

Now first compare machine variance with the residual variance , then

$F_{cal} = \frac{22.67}{14} = 1.619$

The tabulated value of F at $\upsilon_1 = 3$, $\upsilon_2 = 6$ and 5 % level of significance = 4.76.

here $F_{cal} < F_{tab}$

Therefore we conclude that the machine do not differs significantly. So therefore null hypothesis H_o is accepted.

Again comparing season variance with residual variance , we get

$$F_{cal} = \frac{22.67}{16} = 1.417$$

The tabulated value of F at $\upsilon_1 = 2$, $\upsilon_2 = 6$ and 5 % level of significance = 5.14, here $F_{cal} < F_{tab}$

Therefore we conclude that the machine do not differs significantly. So therefore null hypothesis (H_o) is accepted.

Example 2: the following data gives the yield on 12 plots of land in 3 samples under 3 varieties of fertilizers

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

A	B	C
25	20	24
22	17	26
24	16	30
21	19	20

Is there any significance difference in the average yields of lands under the three varieties of fertilizers ? (F at df(2 , 9) at 5 % level = 4.26) Sol .- Null hypothesis (Ho) : there is no significant difference in the average yields under the three varieties. Alternative hypothesis (H1) : the difference in average yields is significant . calculations:

SAMPLE A		**SAMPLE B**		**SAMPLE C**	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
25	625	20	400	24	576
22	484	17	289	26	676
24	576	16	256	30	900
21	441	19	361	20	400
$\sum x_1 = 92$	$\sum x_1^2 = 2126$	$\sum x_2 = 72$	$\sum x_2^2 = 1306$	$\sum x_3 = 100$	$\sum x_3^2 = 2552$

T = sum of all items $= \sum x_1 + \sum x_2 + \sum x_3$

$= 92 + 72 + 10 = 264$

Correction factor = $\frac{T^2}{N} = \frac{(264)^2}{(4+4+4)} = 5808$

Total sum of squares (TSS) = $\sum x_1^2 + \sum x_2^2 + \sum x_3^2 - \frac{T^2}{N}$

$= 2126 + 1306 + 2552 - 5808$

$= 176$

Sum of square between the samples (SSB) $= [\frac{(\Sigma x1)2}{n} + \frac{(\Sigma x2)2}{n} + \frac{(\Sigma x3)2}{n}$

$= \frac{92^2}{4} + \frac{72^2}{4} + \frac{100^2}{4} - 5808$

$= 104$

Degree of freedom (between samples) = $\upsilon_1 = c - 1 = 3 - 1 = 2$

Mean square deviation between samples (MSB) = $\frac{SSB}{\upsilon 1} = \frac{104}{2} = 52$

Sum of square within the sample (SSW) = TSS – SSB

= 176 – 104

= 72

Degree of freedom (within sample) = $\upsilon_2 = N - C$

$= (n_1 + n_2 + n_3) - c$

$= (4 + 4 + 4) - 3 = 9$

Where , C = total number of samples

N = total terms of all the samples

Mean square deviation within the samples (MSW) = $\frac{SSW}{\upsilon 2} = \frac{72}{9} = 8$

Anova table for fertilizers

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

SOURCE OF VARIATION	SUM OF SQUARES (SS)	DEGREES OF FREEDON	MEAN SQUARE (MS)	VARIANCE RATIO (TEST F STATISTICS)
BETWEEN SAMPLES	SSB = 104	$\upsilon_1 = 2$	MSB = $\frac{SSB}{\upsilon 1}$ = 52	F = $\frac{MSB}{MSW}$
WITHIN SAMPLES	SSW = 72	$\upsilon_1 = 9$	MSW = $\frac{SSW}{\upsilon 2}$ =8	$\frac{52}{8}$ = 6.5
TOTAL	TSS = 176	N – 1		

Decisions = the calculated value of F at 0.05 at degrees of freedom (2 , 9) is 6.5 where as the tabulated value of F for the same level of significance and degree of freedom is 4. 26 Since the calculated value of F is much greater than the tabulated value of F , then we reject the null hypothesis (Ho) and conclude that the difference in average yields under the three varieties is significant .

CONCLUSION

As discussed in the present paper we observe that ANOVA is a technique to get the desired results in a statistical manner. Also in mechanical engineering we deals with different machines and devices and various factors which affect during machining and working which can be studied in an appropriate manner by using Analysis of Variance. In one way ANOVA we observe the effect of one factor and in two way ANOVA we consider two factors and their corresponding effects. They both have their merits and de-merits that is why they are restricted to their particular fields and criteria.

REFERENCES

[1] Engineering Mathematics-III by N.P.Bali.
[2] Engineering Mathematics-III by H.K. Dass
[3] Gelman, Andrew(2005) : Analysis of variance
[4] The annals of Statistics
[5] American journal of mechanical engineering, 2013 1 (7) pg 256-261
[6] Brown M, Forsythe, A : Robust tests for the equality of variances
[7] Journals of the American Statistical Association, 364-367, 1974.
[8] M.S. Phadke: Quality engineering using robust design, Prentice-hall, Englewood Cliffs, NJ, 1989.
[9] Anscombe, F. J : The validity of Comparative Experiments. Journal of the Royal Statistical Society.
[10] Scheffe, Henry (1959): The analysis of variance.