IJESMR
International Journal OF Engineering Sciences & Management Research

METHOD OF IMPLEMENTING ANALYSIS OF VARIANCE (ANOVA) IN RESEARCH STUDIES

Asha Singh* & Balkeshwar Singh
*Department of Business Management, Mewar University, Chittorghar, Rajasthan, India
Mechanical Engineering Section, Salalah College of Technology, Salalah, Sultanate of Oman

KEYWORDS: Research, methodology, analysis of variance, means, population.

ABSTRACT
Research is an art of scientific investigation. It is known as voyage of discovery. Analysis of variance abbreviated as ANOVA use in Research. It is considered to be an important tool of analysis in the hands of a researcher. It is a statistical technique specially designed to test whether the means of more than two quantitative populations are equal. The analysis of variance technique developed by R. A. Fisher in 1920 is capable of fruitful application to diversity practical problems. ANOVA is an extremely useful technique concerning researches in the fields of economics, biology, education, psychology, sociology, business industries and in researches of several other disciplines. The main aim of this paper is to provide the students with an opportunity to comprehend a broad area of research.

INTRODUCTION
The analysis of variance originated in agrarian research and its language is thus located with such agricultural terms as blocks (referring to land) and treatments (referring to populations or samples) which are differentiated in terms of varieties of seed, fertilizers or cultivation methods. The word treatment in analysis of variance is used to refer to any factor in the experiment that is controlled at different levels or values. The treatments can be different point of sale displays assembly line technique, sales training programmes or, in short any controlled factor deliberately applied to the elementary units observed in the experiment.
Today, procedure of this analysis finds application in nearly every type of experimental decision in natural sciences as well as social sciences. ANOVA technique, in general, investigates any number factors which are hypothesized or said to influence the dependent variable. One may as well investigate the difference amongst various categories within each of these factors which may have a large number of possible values.

OBJECTIVES OF ANALYSIS OF VARIANCE
- Understand the need for analyzing data from more than two sample;
- Understand the underlying models to ANOVA;
- Understand when and be able to carry out a one-way ANOVA;
- Understand when and be able to carry out a two-way ANOVA;

TECHNIQUE OF ANALYSIS OF VARIANCE
For the sake of clarity the technique of analysis of variance has been divided into two groups as given below
(a) ONE-WAY ANOVA –The One-Way ANOVA is used when comparing the means from three or more groups. Specifically, a One -Way ANOVA is used when there is a single independent variable that has three or more categories. The One-Way ANOVA tells us if the three groups differ from one another on a dependent variable. In one-way ANOVA, we consider only one factor. The technique involves the following steps:

1. Calculate the mean of each sample i.e.
   $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots \bar{X}_k$
   Where k are samples.
2. Take out the mean of the sample means as follows:
   $$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \cdots + \bar{X}_k}{No.of\ sample(k)}$$
3. Calculate the deviations of the sample means from the mean of the sample means and calculate the square of such deviations which may be multiplied by the number of items in the corresponding sample, and then obtain their total. This is known as the sum of squares for variance between the samples (or SS between).
   $$\text{SS between} = n_1\left(\bar{X}_1 - \bar{\bar{X}}\right)^2 + n_2\left(\bar{X}_2 - \bar{\bar{X}}\right)^2 + \ldots + n_k\left(\bar{X}_k - \bar{\bar{X}}\right)^2$$
4. Divide the result of the (3) step by the degrees of freedom between the samples to obtain variance or mean square (MS) between samples.
   $$Ms\ between = \frac{SS\ between}{(k-1)}$$
   Where (k-1) represents degrees of freedom between samples.

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

5. Calculate the deviations of the values of the sample items for all the samples from corresponding means of the sample and calculate the squares of such deviations and then obtain their total. This total is known as the sum of squares for variance between the samples (or SS within).

$$\text{SS within} = \sum\left(X_{1i} - \overline{X}_1\right)^2 + \sum\left(X_{2i} - \overline{X}_2\right)^2 + \ldots.. + \sum\left(X_{ki} - \overline{X}_k\right)^2$$

i = 1,2,3,………

6. Divide the result of (5) step by the degrees of freedom within samples to obtain the variance or mean square (MS) within samples.

$$Ms\ within = \frac{SS\ within}{(n-k)}$$

Where (n-k) represents degrees of freedom within samples,
n= total number of items in all the samples i,e. $n_1 + n_2 + \ldots\ldots\ldots\ldots + n_k$
k= number of samples

7. Calculate the ratio F as follows:

$$F - ratio = \frac{MS\ between}{MS\ within}$$

8. Finally compare the calculated value of F with the table value of F for the degrees of freedom at a certain level. Generally we take 5% level of significance. If the calculated value of F is greater than the table value, it is concluded that the difference in sample means is significant. If the calculated value of F is greater than the table value, it is concluded that the difference in sample means is not significant.

Analysis of Variance Table: It is also known as ANOVA table. The specimen of ANOVA table is given below:

ANOVA Table: One – way ANOVA

Source of variation	Sum of Squares(SS)	v Degree of freedom(d.f.)	Mean Square(MS)	F-ratio
Between Samples	SSC	$v_1 = (k-1)$	$MSE = \frac{SSE}{(n-k)}$	$\frac{MS\ between}{MS\ within}$
Within Samples	SSE	$v_2 = (n-k)$	$MSC = \frac{SSC}{(k-1)}$	
Total	SST	(n-1)		

SST – Total sum of squares of variances.
SSC – Sum of squares between samples (columns).
SSE – Sum of squares within samples (rows).
MSC – Mean sum of squares between samples.
MSE - Mean sum of squares within samples.

EXAMPLE OF ONE-WAY ANOVA

The three samples below have been obtained from normal populations with equal variances. Test the hypothesis that the sample means are equal:

8	7	12
10	5	9
7	10	13
14	9	12
11	9	14

The table value of F at 5%level of significance for $v_1 = 2$ and $v_2 = 12$ is 3.88.

Solution: Let us taken the null hypothesis that there is no significant difference in the means of three samples.

	X_1	X_2	X_3
	8	7	12
	10	5	9
	7	10	13
	14	9	12

IJESMR
International Journal OF Engineering Sciences & Management Research

	11	9	14
Total	50	40	60
$\overline{X}$	10	8	12

$$\overline{\overline{X}} = \frac{10+8+12}{3} = 10$$

Variance between samples

$\left(\overline{X}_1 - \overline{\overline{X}}\right)^2$	$\left(\overline{X}_2 - \overline{\overline{X}}\right)^2$	$\left(\overline{X}_3 - \overline{\overline{X}}\right)^2$
0	4	4
0	4	4
0	4	4
0	4	4
0	4	
0	20	20

Sum of squares between samples = 0 + 20 + 20 = 40

Variance within samples

X_1	$\left(X_1 - \overline{X}_1\right)^2$	X_2	$\left(X_2 - \overline{X}_2\right)^2$	X_3	$\left(X_3 - \overline{X}_3\right)^2$
8	4	7	1	12	0
10	0	5	9	9	9
7	9	10	4	13	1
14	16	9	1	12	0
11	1	8	1	14	4
	30		16		14

Sum of squares within samples = 30 + 16 +14 = 60

ANOVA Table

Source of variation	**Sum of squares**	**v**	**Mean square**
Between	40	12	20
Within	60	12	5
Total	100	14	

F = 20 /5 =4, for $v_1 = 2$ and $v_2 = 12$, $F_{0.05} = 3.88$

The calculated value of F is greater than the table value. The hypothesis is rejected. Hence there is significant difference in the sample means.

(b) TWO-WAY ANOVA: This technique is used when the data are classified on the basis of two factors. Example - In a factory, the various units of a product produced during a certain period may be classified on the basis of different varieties of machines used and also on the basis of different grades of labour. In a two-way ANOVA the analysis of variance table takes the following form:

Source of Variation	**Sumof Squares**	**v Degree of freedom**	**Mean sum of Squares**	**Ratio of F**
Between Samples	SSC	(c-1)	$MSC = \frac{SSC}{(C-1)}$	$\frac{MSC}{MSE}$
Between Rows	SSR	(r-1)	$MSR = \frac{SSR}{(r-1)}$	$\frac{MSR}{MSE}$
Residual or Error	SSE	(c-1)(r-1)	$MSE = \frac{SSE}{(r-1)(c-1)}$	
Total	SST	(n-1)		

IJESMR
International Journal OF Engineering Sciences & Management Research

SST – Total sum of squares
SSC – Sum of squares between columns
SSE – Sum of squares due to error
SSR - Sum of squares between rows

The sum of squares for the source 'Residual' is obtained by subtracting from the total sum of squares the sum of squares between columns and rows, i.e.,

SSE = SST – (SSC + SSR)

The total number of degrees of freedom = n-1 or cr-1

Where c refers to number of columns, r refers to number of rows,

Number of degrees of freedom between columns = (c-1)
Number of degrees of freedom between rows = (r-1)
Number of degrees of freedom for residual = (c-1) (r-1)
Residual or error sum of square = Total sum of squares – sum of squares between columns – sum of square between rows.
After calculating the F value those are compared with the table values. If calculated value of F is greater than the table value at the level of significance, the null hypothesis is rejected otherwise accepted.

EXAMPLE OF TWO-WAY ANOVA - To test the significance of the variation of the retail prices of a commodity in three principle cities, Bombay, Calcutta and Delhi four shops were chosen at random in each city and prices observed in rupees were as follows:

Bombay	**Calcutta**	**Delhi**
16	14	4
8	10	10
12	10	8
14	6	8

Do the data indicate that the prices in the three cities are significantly different?
Solution: Let us take the hypothesis that there is no significant difference in the prices in the three cities.

	Bombay	**Calcutta**	**Delhi**
	16	14	4
	8	10	10
	12	10	8
	14	6	8
Total	50	40	30
$\overline{X}$	12.5	10	7.5

$$\overline{\overline{X}} = \frac{12.5+10+7.5}{3} = 10$$

Variance between samples

$\left(\overline{X}_1 - \overline{\overline{X}}\right)^2$	$\left(\overline{X}_2 - \overline{\overline{X}}\right)^2$	$\left(\overline{X}_3 - \overline{\overline{X}}\right)^2$
6.25	0	6.25
6.25	0	6.25
6.25	0	6.25
6.25	0	6.25
25	0	25

Sum of squares between samples = 25 + 0 + 25 = 50

Variance within samples

$\left(X_1 - \overline{X}_1\right)^2$	$\left(X_2 - \overline{X}_2\right)^2$	$\left(X_3 - \overline{X}_3\right)^2$
12.25	16	12.25
20.25	0	6.25
0.25	0	0.25

IJESMR
International Journal OF Engineering Sciences & Management Research

2.25	16	0.25
35	32	19

Sum of squares within samples = 35 +32 + 19 =86

ANOVA Table

Source of variation	Sum of squares	v	Mean square
Between	50	2	25.00
Within	86	9	9.56
Total	136	11	

F = 25 / 9.56 = 2.62, for $v_1 = 2$, $v_2 = 9$, $F_{0.05} = 4.26$

The calculated value is less than the table value. The hypothesis holds true. Hence the prices in the three cities are not significantly different.

ASSUMPTIONS OF ANOVA: There are some assumptions of ANOVA which are given below:

- Normality: The values in each group are normally distributed.
- Homogeneity: The variance within each group should be equal for all groups ($\sigma_1{}^2 = \sigma_2{}^2 = \sigma_3{}^2 = ... = \sigma_c{}^2$).This assumption is needed to combine or pool the variances within the groups into a single "within groups" source of variation.
- Independence of error: The error (variation of each value around its own group mean) should be independent for each value.

CONCLUSION

Variance is an important statistical measure and is described as the mean of the squares of deviations taken from the mean of the given series of data. The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples. The analysis of variance is not only intended to serve the purpose of testing for the significance of the difference between two sample variances rather its purpose is to test for the significance of the differences among sample means. It can give us answers as to whether different sample data classified in terms of a single variable area meaningful. It can also provide us with meaningful comparisons of sample data which are classified according to two or more variables.

REFERENCES

1. Pervez Ghauri &Kjell Gronhaug, Research Methods in Business Studies, Third Edition, Pearson Education, 2005, ISBN – 0-27368156-7.
2. Ranjit Kaur, Research Methodology, Third Edition, sage publication, 2011, ISBN – 978-184920-301-2.
3. Dr. Prasant Sarangi, Research Methodology, Taxmann Publication, 2010, ISBN – 978-817194-737-9.
4. C.R. Kothari, Research Methodology, Second Edition, New age international publishers, 2004, ISBN (13):978-81-2242488-1.
5. Dawson, Catherine, Practical Research Methods, New Delhi, UBS Publishers' Distributors, 2002.
6. Dr. S.P. Gupta, Statistical Methods, Thirty-Eighth revised edition, Sultan Chand & Sons publisher,2009,ISBN81-8054-739-6.
7. Creswell, J. W. (2008). Educational Research: Planning, conducting, and evaluating quantitative and qualitative research (3rd ed.). Upper Saddle River: Pearson.