UNIT 3 ONE WAY ANALYSIS OF VARIANCE

Structure

3.0 INTRODUCTION

In the foregoing unit you have learned about how to test the significance of a mean obtained on the basis of observations taken from a group of persons and the test of significance of the differences between the two means. No doubt the test of significance of the difference between the two means is a very important technique of inferential statistics, which is used to test the null hypothesis scientifically and help to draw concrete conclusion. But its scope is very limited. It is only applicable to the two sets of scores or the scores obtained from two samples taken from a single population or from two different populations.

Now imagine if we have to compare the means of more than two populations or the number of groups, then what would happen? Can we apply successfully the Critical Ratio Test (CR) or the t test? The answer is yes, but not convenient to apply CR test or t test. The reason can be stated with an example. Suppose we have three groups A,B & C and we want to compare the significance difference in the means of the three groups, then first we have to make the pairs of groups e.g. A and B, then B and C, and then A and C and apply C.R. test or t test as the conditions required. In such condition we are to calculate three C.R. values or t values instead of one.

Now suppose we have eight groups and want to compare the difference in the means of the groups, in such condition we have to calculate 28 C.R. or t values as the condition may require.

It means when there are more than two groups say 3, 4, 5 ….. and k, it is not easy to apply 'C.R.' or 't' test of significance very conveniently.

Further 'C.R.' or 't' test of significance simply consider the means of two groups and test the significance of difference exists between the two means. It has no concern in the variance that exist in the scores of the two groups or variance of the scores from the mean value of the groups.

For example let us say that A reaction time test was given to 5 boys and 5 girls of age group 15+ yrs. The scores were obtained in milliseconds are as given in the table below.

Groups	**Reaction time in M. Sec**					**Sum**	**Mean**
Girls	15	20	5	10	35	85	17M.Sec.
Boys	20	15	20	20	10	85	17M.Sec.

From the mean values shown in the table we can say that the two groups are equal in their reaction time and the average reaction time is 17 M. Sec. In this example, if we apply 't' test of significance, we will find, the difference in the two means insignificant and our null hypothesis is retained.

But if we look carefully to the individual scores of the reaction time of boys and girls, we will find that there is a difference in the two groups. The group of girls is very heterogeneous in their reaction time in comparison to the boys.

As the variation between the scores is ranging from 5 to 30 and deviation of scores from mean varies from 12 M. Sec. to 18 M. Sec.

While the group of boys is more homogeneous in their reaction time, as the variation in the individual scores is ranging from 5 to 10 and deviation of the scores from mean is 3 M. Sec to 7 M. Sec therefore group B is much better in their reaction time in comparison to the group A.

From, this example, you have seen that the test of significance of difference between the two means, some time lead us to draw wrong conclusion and we may wrongly retain the null hypotheses, though it should be rejected in real conditions.

Therefore, when we have more than two, say three or four or so forth and so on, the 'CR' or 't' test of significance are not very useful. In such condition, 'F' test is more suitable and it is known as one way analysis of variance. Because we are testing the significance difference in the average variance exists between the two or more than two groups, instead to test the significance of the difference of the means of the groups.

In this unit we will be dealing with F test or the analysis of variance.

3.1 OBJECTIVES

After going through this unit, you will be able to:

- Define variance;
- Differentiate between variance and standard deviation;
- Define analysis of variance;
- Explain when to use the analysis of variance;

- Describe the process of analysis of variance;
- Apply analysis of variance to obtain ‘F’ Ratio and to solve related problems;
- Analyse inferences after having the value of ‘F’ Ratio;
- Elucidate the assumptions of analysis of variance;
- List out the precautions while using analysis of variance; and
- consult the ‘F’ table correctly and interpret the results.

3.2 ANALYSIS OF VARIANCE

The analysis of variance is an important method for testing the variation observed in experimental situation into different part each part assignable to a known source, cause or factor.

In its simplest form, the analysis of variance is used to test the significance of the differences between the means of a number of different populations. The problem of testing the significance of the differences between the number of means results from experiments designed to study the variation in a dependent variable with variation in independent variable.

Thus the analysis of variance, as the name indicates, deals with variance rather than with standard deviations and standard errors. It is a method of dividing the variation observed in experimental data into different parts, each part assignable to a known source, cause or factor therefore

$$F = \frac{\text{Variance between the groups}}{\text{Variance within the groups}} = \frac{\sigma^2 \text{Between the groups}}{\sigma^2 \text{Within the groups}}$$

In which σ^2 is the population variance.

The technique of analysis of variance was first devised by Sir Ronald Fisher, an English statistician who is also known as the father of modern statistics as applied to social and behavioural sciences. It was first reported in 1923 and its early applications were in the field of agriculture. Since then it has found wide application in many areas of experimentation.

3.2.1 Meaning of the Variance

Before to go further the procedure and use of analysis of variance to test the significance difference between the means of various populations or groups at a time, it is very essential, first to have the clear concept of the term variance.

In the terminology of statistics the distance of scores from a central point i.e. Mean is called deviation and the index of variability is known as the mean deviations or standard deviation (σ).

In the study of sampling theory, some of the results may be some what more simply interpreted if the variance of a sample is defined as the sum of the squares of the deviation divided by its degree of freedom (N-1) rather than as the mean of the squares deviations.

The variance is the most important measure of variability of a group. It is simply the square of S.D. of the group, but its nature is quite different from standard deviation, though formula for computing variance is same as standard deviation (S.D.)

$$\therefore \text{ Variance} = \text{S.D.}^2 \text{ or } \sigma^2 = \frac{\Sigma(X-M)^2}{N}$$

Where X : are the raw scores of a group, and

M : Mean of the raw scores.

Thus we can define variance as **"the average of sum of squares of deviation from the mean of the scores of a distribution."**

3.2.2 Characteristics of Variance

The following are the main features of variance:

- The variance is the measure of variability, which indicates the among groups or between groups difference as well as within group difference.
- The variance is always in plus sign.
- The variance is like an area. While S.D. has direction like length and breadth has the direction.
- The scores on normal curve are shown in terms of units, but variance is a area, therefore either it should be in left side or right side of the normal curve.
- The variance remain the some by adding or subtracting a constant in a set of data.

Self Assessment Questions

1) Define the term variance.

....

....

....

....

2) Enumerate the characteristics of Variance.

....

....

....

....

3) Differentiate between standard deviation and variance.

....

....

....

....

4) What do you mean by Analysis of Variance? Why it is preferred in comparison to 't' test while determining the significance difference in the means.

....

....

....

....

3.2.3 The Procedure of Analysis of Variance (ANOVA)

In its simplest form the analysis of variance can be used when two or more than two groups are compared on the basis of certain traits or characteristics or different treatments of simple independent variable is studied on a dependent variable and having two or more than two groups.

Before we discuss the procedure of analysis of variance, it is to be noted here that when we have taken a large group or a finite population, to represent its total units the symbol 'N' will be used.

When the large group is divided into two or more than two sub groups having equal number of units, the symbol 'n' will be used and for number of groups the symbol 'k' will be used.

Now, suppose in an experimental study, three randomly selected groups having equal number of units say 'N' have been assigned randomly, three kinds of reinforcement viz. verbal, kind and written were used. After a certain period, the achievement test was given to three groups and mean values of achievement scores were compared. The mean scores of three groups can then be compared by using ANOVA. Since there is only one factor i.e. type of reinforcement is involved, the situation warrants a single classification or one way ANOVA, and can be arranged as below:

Table 3.2.1

S.N.	**Group - A** Scores of Verbal Reinforcement		**Group - B** Scores of Kind Reinforcement		**Group – C** Scores of Written Reinforcement	
	X_a	X_a^2	X_b	X_b^2	Xc_1	X_{c1}^2
	Xa_1	(X_{a1}^2)	X_{b1}	(X_{b1}^2)	Xc_1	(X_{c1}^2)
	Xa_2	(X_{a2}^2)	X_{b2}	(X_{b2}^2)	Xc_2	(X_{c2}^2)
	Xa_3	(X_{a3}^2)	X_{b3}	(X_{b3}^2)	Xc_3	(X_{c3}^2)
	Xa_4	(X_{a4}^2)	X_{b4}	(X_{b4}^2)	Xc_4	(X_{c4}^2)
	Xa_5	(X_{a5}^2)	X_{b5}	(X_{b5}^2)	Xc_5	(X_{c5}^2)
	.	.	.	.	.	.
	.	.	.	.	.	.
	Xa_n	(X_{an}^2)	X_{bn}	(X_{bn}^2)	Xc_n	(X_{cn}^2)
Sum	$\sum Xa$	$\sum X_a^2$	$\sum X_b$	$\sum X_b^2$	$\sum Xc1$	$\sum X_{c1}^2$
Mean	$\frac{\sum Xa}{n}$ =Ma		$\frac{\sum Xb}{n}$ =Mb		$\frac{\sum Xc}{n}$=Mc	

To test the difference in the means i.e. MA, MB and MC, the one way analysis of variance is used. To apply one way analysis of variance, the following steps are to be followed:

Step 1 Correction tem Cx = $\frac{(\sum x)^2}{N} = \frac{(\sum x_a + \sum x_b + \sum x_x)^2}{n_1 + n_2 + n_3}$

Step 2 Sum of Squares of Total $SS_T = \sum x^2 - Cx$

$$= \sum x^2 - \frac{(\sum x)^2}{N}$$

$$= \left(\sum x_a^2 + \sum x_b^2 + \sum x_c^2\right) - \frac{(\sum x_3)}{N}$$

Step 3 Sum of Squares Among the Groups $SS_A = \frac{(\sum x)^2}{N} - Cx$

$$= \frac{(\sum x_a)^2}{n_1} + \frac{(\sum x_b)^2}{n_2} + \frac{(\sum x_c)^2}{n_3} - \frac{(\sum x)^2}{N}$$

Step 4 Sum of Squares Within the Groups $SS_W = SS_T - SS_A$

Step 5 Mean Scores of Squares Among the Groups $MSS_A = \frac{SS_A}{k-1}$

Where k = number of groups.

Step 6 Mean Sum of Squares Within the Groups $MSS_W = \frac{SS_W}{n-k}$

Where N = Total number of units.

Step 7 F Ratio i.e. F = $\frac{MSS_A}{MSS_W}$

Step 8 Summary of ANOVA

Table 3.2.2: Summary of ANOVA

Source of variance	Df	S.S.	M.S.S.	F Ratio
Among the Groups	k-1	SS_A	$\frac{SS_A}{K-1}$	$\frac{MSS_A}{MSS_W}$
Within the groups (Error Variance)	N-K	SS_W	$\frac{SS_w}{N-k}$	
Total	N-1			

The obtained F ratio in the summary table, furnishes a comprehensive or overall test of the significance of the difference among means of the groups. A significant F does not tell us which mean differ significantly from others.

If F-Ratio is not significant, the difference among means is insignificant. The existing or observed differences in the means is due to chance factors or some sampling fluctuations.

To decide whether obtain F-Ratio is significant or not we are taking the help of F table from a statistics book.

The obtained F-Ratio is compared with the F value given in the table keeping in mind two degrees of freedom k-1 which is also known as greater degree of freedom or df_1 and N-k, which is known as smaller degree of freedom or df_2. Thus, while testing the significance of the F ratio, two situations may arise.

The obtained F Ratio is Insignificant:

When the obtained F ratio is found less than the value of F ratio given in F table for corresponding lower degrees of freedom df_1 that is, k-1 and higher degree of freedom df that is, (df=N-K) (See F table in a Statistics Book) at .05 and .01 level of significance it is found to be significant or not significant. Thus the null hypothesis is rejected retained. There is no reason for further testing, as none of the mean difference will be significant.

When the obtained 'F Ratio' is found higher than the value of F ratio given in F table for its corresponding df_1 and df_2 at .05 level of .01 level, it is said to be significant. In such condition, we have to proceed further to test the separate differences among the two means, by applying 't' test of significance. This further procedure of testing significant difference among the two means is known as post-hoc test or post ANOVA test of difference.

To have clear understanding, go through the following working examples very carefully.

Example 1

In a study of intelligence, a group of 5 students of class IX studying each in Arts, Commerce and Science stream were selected by using random method of sample selection. An intelligence test was administered to them and the scores obtained are as under. Determine, whether the three groups differ in their level of intelligence.

Table 3.2.3

S.No.	Arts Group Intelligence scores	Comm. Group Intelligence scores	Science Group Intelligence scores
1	15	12	12
2	14	14	15
3	11	10	14
4	12	13	10
5	10	11	10

Solution: In the example k = 3 (i.e. 3 groups), n =5 (i.e. each group having 5 cases), n = 15 (i.e. the total number of units in the group)

Null hypothesis $H_0 = \mu_1 = \mu_2 = \mu_3$

i.e. the students of IX class studying in Arts, Commerce and Science stream do not differ in their level of intelligence.

Thus

Table 3.2.4

Arts Group		Commerce Group		Science Group	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
15	225	12	144	12	144
14	196	14	196	15	225
11	121	10	100	14	196
12	144	13	169	10	100
10	100	11	121	10	100
$\Sigma X_1 = 62$	$\Sigma X_1^2 = 786$	$\Sigma X_2 = 60$	$\Sigma X_2^2 = 730$	$\Sigma X_2 = 61$	$\Sigma X_3^2 = 765$
5		5		5	
12.40		12.00		12.20	

Step 1 : Correction term

$$Cx = \frac{\Sigma(x)^2}{N} = \frac{(\Sigma x_1 + \Sigma x_2 + \Sigma x_3 \ldots\ldots\ldots \Sigma x_k)^2}{n_1 + n_2 + n_3 \ldots\ldots n_k} = \frac{(62+62+61)^2}{5+5+5} = \frac{(183)^2}{15}$$

Or $\quad Cx = 2232.60$

Step 2 : SS_T (Sum of squares of total) = $\Sigma x^2 - Cx$

Or $$= \left(\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 \ldots\ldots\ldots \Sigma x_k^2\right) - \frac{(\Sigma x)^2}{N}$$

$= (786+730+765) - 2232.60$

$= 2281.00 - 2232.60$

$SS_T = 48.40$

Step 3 : SS_A (Sum of squares among the groups) = $\Sigma \frac{(\Sigma x)^2}{N} - Cx$

Or $$= \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} + \ldots\ldots\ldots\ldots + \frac{(\Sigma x_k)^2}{n_k} - Cx$$

$$= \frac{(62)^2}{5} + \frac{(60)^2}{5} + \frac{(61)^2}{5} - 2232.60$$

$= 2233.00 - 2232.60$

Or $SS_A = 0.40$

Step 4 : SS_W (Sum of squares within the groups) = $SS_T - SS_A$

Or $= 48.40 - 0.40$

$SS_W = 48.00$

Step 5 : MSS_A (Mean sum of squares among the groups)

$$MSS_A = \frac{SSA}{k-1} = \frac{0.40}{3-1} = \frac{0.40}{2}$$

Or $MSS_A = 0.20$

Step 6 : MSS_W (Mean sum of squares within the groups)

$$= \frac{SS_W}{N-K} = \frac{48}{15-3} = \frac{48}{12}$$

$MSS_W = 4.00$

Step 7 : F Ratio $= \frac{MSS_A}{MSS_W} = \frac{0.20}{4.00} = 0.05$

Step 8 : Summary of ANOVA

Table 3.2.5 : Summary of ANOVA

Source of variance	df	SS	MSS	F Ratio
Among the Groups	(k-1) 3-1 = 2	0.40	0.20	0.05
Within the Groups	(N-k) 15-3 = 12	48.00	4.00	
Total	14			

From F table (refer to statistics book) for 2 and 12 df at .05 level, the F value is 3.59. Our calculated F value is .05, which is very low than the F value given in the table. Therefore the obtained F ratio is not significant at .05 level of significance for 2 and 12 df. Thus the null hypothesis (H_0) is accepted.

Interpretation of Results

Because null hypothesis is rejected at .05 and .01 level of significance therefore with 99% confidence it can be said that the students studying in Arts, Commerce and Science stream do not differ significantly in their level of intelligence.

Example 2

An experimenter wanted to study the relative effects of four drugs on the physical growth of rats. The experimenter took a group of 20 rats of same age group, from same species and randomly divided them into four groups, having five rats in each group. The experimenter then gave 4 drops of corresponding drug as a one doze to each rat of the concerned group. The physical growth was measured in terms of weight. After one month treatment, the gain in weight is given below. Determine if the drugs are effective for physical growth? Find out if the drugs are equally effective and determine, which drug is more effective in comparison to other one.

Table 3.2.6 : Observations (Gain in weight in ounce)

Group A (Drug P)	Group B (Drug Q)	Group C (Drug R)	Group D (Drug S)
4	9	2	7
5	10	6	7
1	9	6	4
0	6	5	2
2	6	2	7

Solution: Given k = 4, n = 5, N = 20 and Scores of 20 rats in terms of weight

Null hypothesis $H_0 = \mu_1 = \mu_2 = \mu_3$

i.e. All the four drugs are equally effective for the physical growth of the rats.

Therefore:

Table 3.2.7

	Group A		Group B		Group C		Croup D	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
	4	16	9	81	2	4	7	49
	5	25	10	100	6	36	7	49
	1	1	9	81	6	36	4	36
	0	0	6	36	5	25	2	4
	2	4	6	36	2	4	7	49
Sum	12	46	40	334	21	105	27	167
n	5		5		5		5	
Mean	2.40		8.0		4.20		5.40	

Step 1 : Correction Term Cx = $\frac{(\sum x)^2}{N} = \frac{(12+40+21+27)^2}{20} = \frac{(100)^2}{20}$

= 500.00

Step 2 : Sum of Squares of total SS_T = $\sum x^2 - Cx$

= (46+334+105+167) – 500.00

= 152

Step 3 : Sum of Squares Among groups SS_A = $\sum \frac{(\sum x)^2}{n} - Cx$

$$= \left(\frac{(12)^2}{5} + \frac{(40)^2}{5} + \frac{(21)^2}{5} + \frac{(27)^2}{5}\right) - 500.00$$

= 82.80

Step 4 : Sum of Squares Within groups $SS_W = SS_T - SS_A$

= 152 – 82.80

= 69.20

Step 5 : Summary of ANOVA

Table 3.2.8: Summary of ANOVA

Source of variance	df	SS	MSS	F Ratio
Among Groups	4-1 = 3	82.80	$\frac{82.80}{3} = 27.60$	$\frac{27.60}{4.32} = 6.39$
Within Groups (Error variance)	40-4 = 16	69.20	$\frac{69.20}{16} = 4.32$	
Total	19			

In F table $F_{.05}$ for 3 and 16 df = 3.24

and

$F_{.01}$ for 3 and 16 df = 5.24

Our obtained F ratio (6.39) is greater than the F value at .01 level of significance for 3 and 16 df. Thus the obtained F ratio is significant at .01 level of confidence. Therefore the null hypothesis is rejected at .01 level of confidence. i.e. the drugs P, Q, R, S are not equally effective for physical growth.

In the given problem it is also to be determined which drug is comparatively more effective. Thus we have to make post-hoc comparisons.

For post-hoc comparisons, we apply 't' test of significance. The common formula of 't' test is –

$$t = \frac{M_1 \sim M_2}{S.E_{DM}}$$

Where :

M_1 = Mean of first group

M_2 = Mean of second group, and

SE_{DM} = Standard Error of Difference of Means.

Here $SE_{DM} = SDW\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Where SD_W or $\sigma_w = \sqrt{MSS_W}$

i.e. $S.D_W$ is the within groups S.D. and n_1 and n_2 are the size of the samples or groups being compared.

In the given example the means of four groups A, B, C and D are ranging from 2.40 ounce to 8.00 ounce, and the mean difference from 5.60 to 1.20. To determine the significance of the difference between any two selected means we must compute 't' ratio by dividing the given mean difference by its $S.E._{DM}$. The resulting t is then compared with the 't' value given in 't' table (Table no 2.5.1 of Unit 2) keeping in view the df of within the groups i.e. df_W. Thus in this way for four groups we have to calculate 6, 't' values as given below:

Step 6 : Standard deviation of within the groups

$SD_W = \sqrt{MSS_W} = \sqrt{4.32}$

= 2.08

Step 7 : Standard Error of Difference of Mean ($S.E_{DM}$)

$S.E._{DM} = SDW\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

= 1.31

(All the groups have same size therefore the value of SE_{DM} for the two groups will remain same)

Step 8 : Comparison of the means of the various pairs of groups.

Group A vs B

$$t = \frac{M_A - M_B}{S.E_{DM}} = \frac{8.0 - 2.40}{1.31} = \frac{5.60}{1.31} = 4.28$$ (Significant at .01 level for 16 df).

Group A vs C

$$t = \frac{4.20 - 2.40}{1.31} = \frac{1.80}{1.31} = 1.37$$ (Insignificant at .05 level for 16 df).

Group A vs D

$$t = \frac{5.40 - 2.40}{1.31} = \frac{3.0}{1.31} \quad t = 2.29$$ (Significant at .05 level for 16 df).

Group B vs C

$$t = \frac{8.0 - 4.90}{1.31} = \frac{3.80}{1.31} \quad t = 2.90$$ (Significant at .05 level for 16 df).

Group B vs D

$$t = \frac{8.0 - 5.40}{1.31} = \frac{2.60}{1.31} = 1.98$$ (Insignificant at .05 level for 16 df).

Group C vs D

$$t = \frac{5.40 - 4.20}{1.31} = \frac{1.20}{1.31} = 0.92$$ (Insignificant at .05 level for 16 df).

Results :

Out of 6 ‘t’ values, only 3 t values are found statistically significant. Among these three, one value is found significant at .01 level, while the two values are found significant at .05 level of significance. From these ‘t’ values, it is quite clear that the group B is better in physical growth in comparison to the group A and C, similarly group D is found better in comparison to Group A. The group B & D and groups C & D are found almost equal in their physical growth.

Interpretation of the Results

Since the group B is found better in physical growth of the rats in comparison to group A at 99% confidence level and at 95% confidence level it is found better in case of group C. But the group B and D are found approximately equally good in physical growth. Therefore the drug Q and S are effective for physical growth in comparison to the drugs P and R. Further the drug Q is comparatively more effective than the other drugs P, R and S respectively.

From the forgoing illustrations you have noted that if obtained F value is not significant, it means, there is no difference either of the pairs of groups. There is no need to follow ‘t’ test. If F is found significant, then the complete procedure of analysis of variance is to specify the findings by using ‘t’ test. Therefore you have noticed that only the F value is not sufficient when it is found significant. It is to be completed when supplemented by using the ‘t’ test.

3.2.4 Steps of One Way Analysis of Variance

From the foregoing two illustration, it is clear that following steps are to be followed when we use analysis of variance.

Step 1 : Set up null hypothesis.

Step 2 : Set the raw scores in table form as shown in the two illustrations.

Step 3 : Square the individual scores of all the sets and write the same in front of the corresponding raw score.

Step 4 : Obtain all the sum of raw scores and the squares of raw scores. Write them at the end of each column.

Step 5 : Obtain grand sums of raw scores as and square of raw square

as $\sum x^2$

Step 6 : Calculate correction term by using the formula

$$Cx = \frac{\sum x^2}{N} \quad \text{Or} \quad Cx = \frac{\left(\sum x_1 + \sum x_2 + \sum x_3 \ldots\ldots\ldots\ldots + \sum x_k\right)^2}{n_1 + n_2 + n_3 + \ldots\ldots\ldots n_k}$$

Step 7 : Calculate sum of squares i.e. SS_T by using the formula-

$$SS_T = \sum x^2 - Cx$$

Step 8 : Calculate sum of squares among the groups i.e. SS_A by using the formula-

$$SS_A = \frac{\sum x^2}{n} - Cx$$

$$\text{Or } SS_A = \frac{\left(\sum x_1^2\right)^2}{n_1} + \frac{\left(\sum x_2^2\right)^2}{n_2} + \frac{\left(\sum x_3^2\right)^2}{n_3} + \ldots\ldots\ldots\ldots + \frac{\left(\sum x_k^2\right)^2}{n_k} - Cx$$

Step 9 : Calculate sum of squares within the groups i.e. SSw by using the formula

$SS_W = SS_T - SS_A$

Step 10 : Calculate the degrees of freedom as

greater degree of freedom $df_1 = k - 1$ (where k is number of groups)

Smaller degree of freedom $df_2 = N-k$ (where N is the total number in the group)

Step 11 : Find the value of Mean sum of squares of two variances as-

Mean sum of squares between the group $MSS_A = \frac{SS_A}{k-1}$

Mean sum of squares within the groups $MSS_W = \frac{SS_W}{N-K}$

Step 12 : Prepare summary table of analysis of variance as shown in 3.2.5 or 3.2.8.

Step 13 : Evaluate obtained F Ratio with the F ratio value given in F table (Table no. 3.3.1) keeping in mind df_1 and df_2.

Step 14 : Retain or Reject the Null Hypothesis framed as in step no-I.

Step 15 : If F ratio is found insignificant and null hypothesis is retained, stop further calculation, and interpret the results accordingly. If F ratio is found significant and null hypothesis is rejected, go for further calculations and use post-hoc comparison, find the t values and interpret the results accordingly.

3.2.5 Assumptions Underlying the Analysis of Variance

The method of analysis of variance has a number of assumption. The failure of the observations or data to satisfy these assumptions, leads to the invalid inferences. The following are the main assumptions of analysis of variance.

The distribution of the dependent variable in the population under study is normal.

There exists homogeneity of variance i.e. the variance in the different sets of scores do not differ beyond chance, in other words $\sigma_1 = \sigma_2 = \sigma_3 = ... = \sigma_k$.

The samples of different groups are selected from the population by using random method of sample selection.

There is no significant difference in the means of various samples or groups taken from a population.

3.2.6 Relationship between 'F' test and 't' test

The F test and t test are complementary to each other, because-

't' is followed when 'F' value is significant for the specification of inferences.

'F' test is followed, when 't' value is not found significant. Because within groups variance is not evaluated by 't' test. It evaluate only the difference between variance.

There is a fixed relation between 't' and 'F'. the F is the square of 't', while 't' is a square root of F.

F = t^2 $or\ t = \sqrt{F}$

3.2.7 Merits or Advantages of Analysis of Variance

The analysis of variance technique has the following advantages:

- It is the improved technique over the 't' test or 'z' test, it evaluates both types of variance 'between' and 'within'.
- This technique is used for ascertaining the difference among several groups or treatments at a time. It is an economical device.
- It can involve more than one variable in studying their main effects and interaction effects.
- In some of the experimental design e.g. simple random design and levels X treatment designs are based on one way analysis of variance.
- If 't' is not significant, F test must be followed, to analyse the difference between two means.

3.2.8 Demerits or Limitations of Analysis of Variance

The analysis of variance technique has following limitations also:

- We have seen that analysis of variance techniques is based on certain assumptions e.g. normality and homogeneity of the variances among the groups. The departure of the data from these assumptions may effect adversely on the inferences.
- The F value provides global findings of difference among groups, but it can not specify the inference. Therefore, for complete analysis of variance, the 't' test is followed for specifying the statistical inference.

- It is time consuming process and requires the knowledge and skills of arithmetical operations as well the high vision for interpretations of the results.
- For the use of 'F' test, the statistical table of 'F' value is essential without it results can not we interpreted.

3.3 F RATIO TABLE AND THE PROCEDURE TO USE

The significance of difference between two means is analysed by using 't' test or 'z' test as it has been discussed in earlier unit 2. The calculated or obtained t value is evaluated with the help of 't' distribution table. With df at .05 and .01 levels of significance. The df is the main base to locate the 't' values at different levels of significance given in the table.

In a similar way the calculated F ratio value is evaluated by using F table (refer to statistics book) by considering the degree of freedom between the groups and within the groups.

You observe the F table, carefully, you will find there are rows and columns. In the first row there are the degrees of freedom for larges variance i.e. for greater mean squares or between the variance. The first column of the table has also degree of freedom of smaller mean squares or the variance within the groups. Along with these two degree of freedom the F ratio values are given at .05 and .01 level of significance. The normal print values is the values at .05 level and the bold or dark print values are at .01 level of significance.

In the first illustrated example in the summary table the F ratio value is 0.05, df_1, is 2 and df_2 is 12. For evaluating the obtained F value with F value given in table df_1 = 2 is for row and df2 is for column. In the column you will find 12, proceed horizontally or row wise and stop in column 2, you will find F values 3.88, which is for .05 level of confidence and 6.93 (in dark bold print) which is meant for .01 level of confidence. Our calculated F value .05 is much less than these two values. Hence the F ratio is not significant also at .05 level. Thus the null hypothesis is retained.

Self Assessment Questions

1) State the assumptions of ANOVA.

....

....

....

....

2) What happens when these assumptions are violated?

....

....

....

....

3) Compare the 'F Ratio' test and 't Ratio' test in terms of their relative merits and demerits.

....

....

....

....

4) What is the mathematical relationship between F and t.

....

....

....

....

5) What is the relationship between S.D. and Mean sum of squares within the groups?

....

....

....

....

6) When the post ANOVA test of difference is applied?

....

....

....

....

7) How many degree of freedom are associated with the variation in the data for-A comparison of four means for independent samples each containing 10 cases?

....

....

....

....

8) A comparison of three groups selected independently each containing 15 units.

....

....

....

....

3.4 LET US SUM UP

Analysis of variance is used to test the significance of the difference between the means of a number of different populations say two or more than two.

Analysis of variance deals with variance rather to deal with means and their standard error of the difference exist between the means.

The variance is the most important measure of variability of a group. It is simply the square of S.D. of the group i.e. v = σ^2

The problem of testing the significance of the differences between the number of means results from experiments designed to study the variation in a dependent variable with variation in independent variable.

Analysis of variance is used when difference in the means of two or more groups is found insignificant.

There is a fixed relationship between 't' ratio and 'F' ratio. The relationship can be expressed as $F = t^2$ or $t = \sqrt{F}$.

While determining the significance of calculated or obtained ratio, we consider two types of degrees of freedom. One greater i.e. degree of freedom between the groups and second smaller i.e. degree of freedom within the groups.

3.5 UNIT END QUESTIONS

1) The four groups are given four treatments, each group consists of 5 subjects. At the end of treatment a test is administered, the obtained scores are given in the following table. Test significance of difference among four treatments.

Scores of the treatment

Group – A	Group – B	Group – C	Group – D
X1	X2	X3	X4
14	19	12	17
15	20	16	17
11	19	16	14
10	16	15	12
12	16	12	17

2) A Test Anxiety test was given to three groups of students of X class, classified as high achievers, average achievers and low achievers. The scores obtained on the test are shown below. Are the three groups differ in their test anxiety.

High-achievers	Average achievers	Low achievers
15	19	12
14	20	14
11	16	12
12	19	12

3) Apply ANOVA on the following sets of scores. Interpret your results.

Set-I	Set-II	Set-III
10	3	10
7	3	11
6	3	10
10	3	5
4	3	6
3	3	8
2	3	9
1	3	12
8	3	9
9	3	10

4) Summary of analysis of variance is given below:

Source of variance	Df	SS	MSS	F
Between sets	2	180	90.00	17.11
Within sets	27	142	5.26	
Total	29			

Interpret the result obtained.

Note Table F values are

$F_{.05}$ for 2 and 27 df = 3.35

$F_{.01}$ for 2 and 27 df = 5.49

5) Given the following statistics of the two groups obtained on a verbal reasoning test:

Group	N	M	σ
Boys	95	29.20	11.60
Girls	83	30.90	7.80

Calculate:

‘t’ ratio for the two groups.

‘F’ ratio for the two groups.

What should be the degree of freedom for ‘t’ ratio.

What should be the degrees of freedom for ‘F’ ratio.

Interpret the results obtained on ‘t’ ratio and ‘F’ ratio.

6) Why it is necessary to fulfill the assumptions of ‘F’ test, before to apply analysis of variance.

7) Why the 'F' ratio test and 't' ratio tests are complementary to each other.

8) What should be the various problems of psychology and education. Where the ANOVA can be used successfully.

3.6 SUGGESTED READINGS

Aggarwal, Y.P. (1990). *Statistical Methods – Concept, Applications, and Computation*. New Delhi : Sterling Publishers Pvt. Ltd.

Ferguson, G.A. (1974). *Statistical Analysis in Psychology and Education.* New York : McGraw Hill Book Co.

Garret, H.E. & Woodwarth, R.S. (1969). *Statistics in Psychology and Education.* Bombay : Vakils, Feffer & Simons Pvt. Ltd.

Guilford, J.P. & Benjamin, F. (1973). *Fundamental Statistics in Psychology and Education.* New York : McGraw Hill Book Co.

Srivastava, A.B.L. & Sharma, K.K. (1974). *Elementary Statistics in Psychology and Education.* New Delhi : Sterling Publishers Pvt. Ltd.

Walker, H.M. & Lev. J. (1965). *Statistical Inference.* Calcutta : Oxford & I.B.H. Publishing Co.