UNIT 12 ANALYSIS OF VARIANCE TESTS

Structure

12.1 INTRODUCTION

In Unit 4 of this block, we tested the equality of means of two independent groups by using t-test. Sometimes situations may arise where testing of more than two means is required. As an example, in crop-cutting experiments it may be required to test whether under similar conditions the average yield of some crop in a number of fields is same or not. For obvious reasons, in such cases, t-test cannot be applied.Generally, for such situations, the technique of Analysis of Variance (ANOVA) is used, in which the testing of equality of several means is doe by dividing the population variability into different components. The usual F-test is used to test the equality of means of several groups.

As its name suggests, the analysis of variance focuses on variability. It involves the calculation of several measures of variability, when the total variability of the population is divided into many components, like, variability within the smaller sub-groups, variability between the smaller sub-groups, etc. In other words, ANOVA is a technique which split up the total variation of data which may be attributed to various "sources" or "causes" of variation. There may be variation between variables and also within different levels of variables. In this way, Analysis of Variance is used to test the homogeneity of several population means by comparing the variances between the sample and within the sample.

In this unit, we shall discuss the one-way as well as two-way Analysis of Variance. One-way Analysis of Variance is a technique where only one independent variable at different levels is considered which affects the response variable whereas in two-way Analysis of Variance technique, we will consider two variables at different levels which affect the response variables.

12.2 OBJECTIVES

After studying this unit, you should be able to:

- understand the Analysis of Variance technique;
- describe various types of assumptions underlying the Analysis of Variance technique and applications of it;
- define various types of linear models used in Analysis of Variance technique;
- understand how to test the hypothesis under One-way Analysis of Variance; and
- explain the method of performing Two-way ANOVA Test.

12.3 ANALYSIS OF VARIANCE (ANOVA)

The statistical technique known as "Analysis of Variance"(abbreviated as ANOVA), was propounded by Professor R.A. Fisher in 1920's, in which he developed the method of testing equality of means of several sub-populations by dividing the total variability into different components. Variation is inherent in nature, so analysis of variance means examining the variation present in data or parts of data. In other words, analysis of variance means to find out the cause of variation in the data. The total variation in any set of numerical data of an experiment is due to number of causes which may beAssignable causes and Chance causes.

The variation in the data due to assignable causes can be detected, measured and controlled whereas the variation due to chance causes is not in the control of human being and cannot be traced or find out separately.

According to Professor R.A. Fisher, Analysis of Variance (ANOVA) is the "separation of variance ascribable to one group of causes from the variance ascribable to other group". So, by this technique, the total variation present in the data is divided into two components of variation; one due to assignable causes (Between the Groups Variability) and the other due to chance causes (Within Group Variability). Analysis of variance technique can be classified as Parametric ANOVA and Non-Parametric ANOVA.The topic of this unit is related to parametric ANOVA.

Parametric ANOVA can be classified as simply ANOVA if only one response variable is considered. If more than one response variables are under consideration than it is called multivariate analysis of variance (MANOVA).

If we consider, only one independent variable which affects the response/dependent variable then it is called One-way ANOVA. If the independent variables/explanatory variables are more than one, say n, then it is called n-way ANOVA. If n is equal to two than the ANOVA is called Two-way classified ANOVA.

12.3.1 Significance of Analysis of Variance

One obvious question may arise that why do we call it Analysis of Variance, even though we are testing for the equality of means? Why do not we simply call it the Analysis of Means? How do we propose test for means by analysis the variances? As a matter of fact, in order to determine if means of several populations are equal, we do consider the measure of variance, σ^2.

The estimate of population variance σ^2 is computed by two different estimates of it each one by a different method. One approach is to compute an estimator of σ^2 in such a manner that even if the population means are not equal it will have no effect on the value of this estimator. This means that the differences in the values of the population means does not affect the value of σ^2 as calculated by a given method. This estimator of σ^2 is the average of the variances found within each of the samples. For example, if we take 10 samples of size n, then each sample will have a mean and a variance. Then the mean of these 10 variances would be considered as an unbiased estimator of σ^2, the population variance, and its value remains appropriate irrespective of whether the population means are equal or not. This is really done by pooling all the sample variances to estimate a common population variance, which is the average of all sample variances. This common variance is known as variance within sample or σ^2_{within}

The second approach to calculate the estimate of σ^2 is based upon the Central Limit Theorem and is valid only under the null hypothesis assumption that all the population means are equal. This means that in fact, if there are no differences among the population means, then the computed value of σ^2 by the second approach should not differ significantly from the computed value of σ^2 by the first approach. Hence, if these two values of σ^2 are approximately the same, then we can decide to accept the null hypothesis of equality of means.

The second approach results in the following computation:

Based upon the Central Limit Theorem, we have previously found that the standard error of the sample means is calculated by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

or, the variance would be:

$$\sigma_{\bar{x}}^{\ 2} = \frac{\sigma^2}{n} \qquad \text{or,} \qquad \sigma^2 = n\sigma_{\bar{x}}^{\ 2}$$

Thus, by knowing the square of the standard error of the mean $(\sigma_{\bar{x}})^2$, we could multiply it by n and obtain a precise estimate of σ^2. This approach of estimating σ^2, is known as $\sigma^2_{\text{between}}$. Now, if all population means are equal then, $\sigma^2_{\text{between}}$ value should be approximately the same as $\sigma^2_{\text{between}}$ value. A significant difference between these two values would lead us to conclude that this difference is the result of difference between the population means.

But, how do we know that any difference between these two values is significant or not? How do we know whether this difference, if any, is simply due to random sampling error or due to actual differences among the population means?

R. A. Fisher developed a Fisher test or F-test to answer the above question. He determined that the difference between the $\sigma^2_{\text{between}}$ and σ^2_{within} could be expressed as a ratio to be designated as the F-value, so that

$$F = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}}$$

In the above, case, if the population means are exactly the same, then $\sigma^2_{\text{between}}$ will be equal to the σ^2_{within} and the value of F will be equal to 1.

However, because of sampling errors and other variations, some disparity between these two values will be there, even when the null hypothesis is true, meaning that all population means are equal. The extent of disparity between the two variances and consequently, the value of F, will influence our decision on whether to accept or reject the null hypothesis. It is logical to conclude that if the population means are not equal then their sample means will also vary greatly from one another, resulting in a large value of $\sigma^2_{\text{between}}$, and hence a larger value of F (σ^2_{within} is based only on sample variances and not on sample means and hence is not affected by differences in sample means). Accordingly, larger the value of F statistic, the more likely the decision is to reject the null hypothesis. But how large the value of F be so as to reject the null hypothesis? The answer is that the computed value of F must be larger than the critical value of F, given in the table for a given level of significance and calculated number of degrees of freedom. (The F distribution is a family of curves, so that there are different curves for different degrees of freedom.)

12.3.2 Degrees of Freedom

We have talked about the F-distribution being a family of curves, each curve reflecting the degrees of freedom relative to both $\sigma^2_{\text{between}}$ and σ^2_{within}. This means that the degrees of freedom are associated both with the numerator as well as with the denominator of the F-ratio.

Since the variance between samples $\sigma^2_{between}$ comes from many samples and if there are k number of samples, then the degrees of freedom, associated with the numerator would be (k – 1).

The denominator is the mean variance of the variances of k samples and since each variance in each sample is associated with the size of the sample (n), then the degrees of freedom associated with each sample would be (n – 1). Hence, the total degrees of freedom would be the sum of degrees of freedom of k sample or df = k(n – 1), when each sample is of size n.

12.3.3 Uses of ANOVA

The following are some of the uses of ANOVA:

1. **To test the homogeneity of several means (say, k groups) that is, H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$**

   If H_0 is rejected then we can say that there is a significant difference between these k groups or there is a significant effect of these k independent variables.

2. **To test the relationship between two variables**

   This test provides evidence that dependent variable Y and independent variable X are related in their movements. If Y do not relate with Xthen we expect H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$, which is the null hypothesis for testing the absence of relationship.

3. **Test for Linearity of Regression**

   If in 2, some relationship is established, the next step is to find the appropriate regression function. At the first stage we try to find out whether the linear regression fits the observed data, that is the null hypothesis will be

   $$H_0: \mu_i=\alpha+\beta X_i$$

   with the sample model $Y_{ij} = \mu_i + e_{ij}$, when α and β are the parameters of the model.

4. **Test for Polynomial Regression**

   If a linear relationship is not established, we proceed to test the hypothesis

   $$H_0: \mu_i = \alpha+\beta_1 X_i + \beta_2 X_i^2 +\ldots+ \beta_k X_i^k$$

   That is, the relationship between X and Y can be explained by a polynomial of degree k.

5. **Some Other Uses of ANOVA**

   - Test of Homogeneity of a Group of Regression Coefficients.
   - Test for Equality of Regression Equations from p Groups.
   - Test for Multiple Linear Regression Model.

CHECK YOUR PROGRESS 1

Note: i) Use the space given below for your answers.

ii) Check your answers with those given at the end of the unit.

1. Describe the Analysis of Variance and differentiate between One-way and Two-way ANOVA.

…………………………………………………………………………….

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

2. Explain briefly the uses of Analysis of Variance.

…………………………………………………………………………….

……………………………………………………………………………

……………………………………………………………………………

……………………………………………………………………………

12.4 ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

One-factor analysis of variance (or one-way analysis of variance) is a special case of ANOVA.Whereas two sample t-test is used to decide whether two groups (two levels) of a factor have the same mean; one-way ANOVA generalizes this problem to k levels (greater than two) of a factor.

In the following, subscript i refers to the i^{th} level of the factor and subscript j refers to the j^{th} observations within a level of factor. For example, y_{23} refers to third observation in the second level of a factor.

The observations on different levels of a factor can be exhibited as shown below:

Level of a factor	**Observations**			**Totals**	**Means**
1	y_{11}	y_{12}............	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}............	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
.			.	.	.
.			.	.	.
.			.	.	.
k	y_{k1}	y_{k2}............	y_{kn}	$y_{k.}$	$\bar{y}_{k.}$

A linear mathematical model for one-way classified data is used which can be written as

$$y_{ij} = \mu_i + e_{ij} \qquad \text{where } i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n.$$

Here, y_{ij} is continuous dependent or response variable, whereas μ_i is discrete independent variable, also called an explanatory variable. Total number of observations is

$$n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i = N$$

In fact, this model decomposes the responses into a mean for each level of a factor and error term, that is,

Response = A mean for each level of a factor + error term

The analysis of variance provides estimates for each level mean. These estimated level means are the predicted values of the model and the difference between the response variable and the estimated/predicted level means are the residuals.

That is,

$$y_{ij} = \mu_i + e_{ij} \text{ implying that } \quad e_{ij} = y_{ij} - \mu_i$$

The above model can also be re-written as $y_{ij} = \mu + (\mu_i - \mu) + e_{ij}$

or $\quad y_{ij} = \mu + \alpha_i + e_{ij}, \ \forall\, i = 1, 2, ..., k$ and $j = 1, 2, ..., n_i$

where, $\alpha_i = \mu_i - \mu$.

12.4.1 Assumptions

The following are the basic assumptionsof one-way ANOVA:

1. Dependent variable measured on interval scale;
2. k samples are independently and randomly drawn from the population;
3. Population can be assumed reasonably to have a normal distribution;
4. k samples have approximately equal variance;
5. Various effects are additive in nature; and
6. e_{ij} are independently and identically distributed (i.i.d.) normal variables with mean zero and variance σ_e^2.

12.4.2 Procedure of One-Way ANOVA Test

Now, when we discuss the step-by-step computation procedure for one way analysis of variance for k independent samples, the first step of the procedure is to make the null and alternative assumptions.

Null Hypothesis

We want to test the equality of the population means, that is, to test the homogeneity of effect of different levels of a factor. Hence, the null hypothesis is given by

$$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_k$$

against the alternative hypothesis

$$H_1: \quad \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k \qquad \text{(or some } \mu_i\text{'s are not equal)}$$

Determining Level of Significance:

The next is to decide a criterion for acceptance or rejection of null hypothesis i.e., cut-off value and level of significance α, at which we have to test our null hypothesis. As discussed in Unit 3 and 4, level of significance is a fixed assumed value which is decided before starting the test procedure. The most commonly used values of α are 0.05 (5%) and 0.01 (1%). **Let us take it as** α **= 5% (or 1%) if not given** anywhere or follow whatever given for α.

Computation of Ratio of Variations F:

Since here F ratio contains only two elements, which are the variances between the samples and within the samples respectively, as discussed before, let us recapitulate the calculation of these variances.

If all the means of samples were exactly equal and all samples were exactly representative of their respective populations so that all the sample means were exactly equal to each other and to the population mean, then there will be no variance. However, this can never be the case. We always have variation both between samples and within samples, even if we take these samples randomly and from the same population. This variation is known as the total variation.

The total variation designated by $\sum\left(X-\overline{\overline{X}}\right)^2$, where X represents individual observations for all samples and $\overline{\overline{X}}$ is the grand mean of all sample means and equals μ, the population mean, is also known as the total sum of squares or SST, and is simply the sum of squared differences between each observation and the overall mean. This total variation represents the contribution of two elements. These elements are:

a) **Variance between samples.** The variance between samples may be due to the effect of different treatments, meaning that the population means may be affected by the factor under consideration, thus making the population means actually different, and some variance may be due to the inter-sample variability. This variance is also known as the sum of squares between samples. Let this sum of squares be designed as SSB. Then SSB is calculated by the following steps:

Step-I: Take k samples of size n each and calculate the mean of each sample, designated as $\overline{X}_1, \overline{X}_2, \overline{X}_3, ..., \overline{X}_k$.

Step-II: Calculate the grand mean $\overline{\overline{X}}$ of the distribution of these sample means, so that

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k}\overline{X}_i}{k}$$

Step-III: Take the difference between the means of the various samples and the grand mean, which can be denoted as

$$(\overline{X}_1-\overline{\overline{X}}), (X_2-\overline{\overline{X}}), (\overline{X}_3-\overline{\overline{X}}), ..., (\overline{X}_k-\overline{\overline{X}})$$

Step-IV: Square these deviations or differences individually, multiply each of these squared deviations by its respective sample size and sum up all these products, so as to find

$$\sum_{i=1}^{k} n_i \left(\overline{X}_i - \overline{\overline{X}}\right)^2, \text{ where } n_i = \text{size of the } i^{th} \text{ sample.}$$

This will be the value of the SSB.

However, sometimes individual observations in each k samples are not available, but means of the samples are given, then the step 1 can be skipped and we start the computation procedure from step 2. Divide SSB by the degrees of freedom, which are (k – 1), where k is the number of samples and this would give us the value of $\sigma^2_{between}$, so that:

$$\sigma^2_{between} = \frac{SSB}{(k-1)}$$

This is also known as mean square between samples or MSB.

b) **Variance within samples.** Even though each observation in a given sample comes from the same population and is subjected to the same treatment, some chance variation can still occur. The variance may be due to sampling errors or other natural causes. The variance or sum of squares is calculated through the following steps:

Step-I: Calculate the mean value of each sample, that is, $\overline{X}_1, \overline{X}_2, \overline{X}_3, ..., \overline{X}_k$.

Step-II: Take one sample at a time and take the deviation of each item in the sample from its mean. Do this for all the samples, so that we would have a difference between each value in each sample and their respective means for all values in all samples.

Step-III: Square these differences and take a total sum of all these squared differences (or deviations). This sum is also known as SSW or sum of squares within samples.

Step-IV: Divide this SSW by the corresponding degrees of freedom. The degrees of freedom are obtained by subtracting the total number of samples from the total number of items.

Thus if N is the total number of items or observations, and k is the number of samples, then, df = (N – k) which are the degrees of freedom within samples (If all samples are of equal size n, then df = k(n – 1), since (n – 1) are the degrees of freedom for each sample and there are k samples).

This figure SSW/df is also known as σ^2_{within}, or MSW (mean of sum of squares within samples).

Now the value of F can be computed as:

$$F = \frac{\sigma^2_{between}}{\sigma^2_{within}} = \frac{SSB/df}{SSW/df} = \frac{SSB/(k-1)}{SSW/(N-k)} = \frac{MSB}{MSW}$$

Construction of ANOVA Table

After the various calculations for SSB, SSW and the degrees of freedom have been, made, these figures can be presented in a simple table called Analysis of Variance table or simply ANOVA table, as follows;

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	Variance Ratio F
Treatment	SSB	(k -1)	$MSB = \frac{SSB}{(k-1)}$	$F = \frac{MSB}{MSW}$
Within	SSW	(N – k)	$MSW = \frac{SSW}{(N-k)}$	
Total	SST			

Then, the variance ratio is

$$F = \frac{MSB}{MSW}$$

This calculated value of F is then compared with the critical value of F, obtained from the F-table at respective d.f. and a decision is made about the validity of null hypothesis. The critical value of F from the table for α level of significance and degrees of freedom as follows:

df(numerator) = (k – 1)

df (denominator) = (N – k)

This value of F from the table is compared with calculated value of F and if calculated value of F is less than the critical value of F, we cannot reject the null hypothesis.

Now after discussing the procedure of One-way ANOVA test let us practice to solve some examples.

Example 1: To test whether all professors teach the same material in different sections of the introductory statistics class or not, 4 sections of the same course were selected and a common test was administered to 5 students selected at random from each section. The scores for each student from each section were noted and are given below. We want to test for any differences in learning, as reflected in the average scores for each section.

Student #	Section 1 Scores (X_1)	Section 2 Scores (X_2)	Section 3 Scores (X_3)	Section 4 Scores (X_4)
1	8	12	10	12
2	10	12	13	15
3	12	10	11	13
4	10	8	12	10
5	5	13	14	10
				10
Totals	$\sum X_1 = 45$	$\sum X_2 = 55$	$\sum X_3 = 60$	$\sum X_4 = 60$
Means	$\overline{X}_1 = 9$	$\overline{X}_2 = 11$	$\overline{X}_3 = 12$	$\overline{X}_4 = 12$

Solution: The method is as follows:

1) State the null hypothesis. We are assuming that there is no significant difference among the average scores of students from these 4 sections and hence all professors are teaching the same material with the same effectiveness, i.e.

   $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$.

   H_1 : All means are not equal or at least two means differ from each other

2) Establish a level of significance. Let $\alpha = 0.05$.

3) Calculate the variance between the samples, as follows:

   a) The mean of each sample is,

   $\overline{X}_1 = 9,\ \overline{X}_2 = 11,\ \overline{X}_3 = 12,\ \overline{X}_4 = 12$

   b) The grand mean or $\overline{\overline{X}}$ is:

   $$\overline{\overline{X}} = \frac{\sum \overline{X}}{n} = \frac{9+11+12+12}{4} = 11$$

   c) Calculate the value of SSB.

   $$SSB = \sum n\left(\overline{X} - \overline{\overline{X}}\right)^2$$
   $$= 5(9-11)^2 + 5(11-11)^2 + 5\ (12-11)^2 + 5\ (12-11)^2$$
   $$= 20 + 0 + 5 + 5 = 30$$

   d) The variance between samples $\sigma^2_{between}$ or MSB is given by:

   $$MSB = \frac{SSB}{df} = \frac{(30)}{(k-1)} = \frac{(30)}{3} = 10$$

4) Calculate the variance within samples, as follows:

   The find the sum of squares within samples (SSW) we square each deviation between the individual value of each sample and its mean, for all samples and then sum these squares deviations, as follows:

   $$\sum\left(X_i - \overline{X}_i\right)^2 = \sum\left(X_1 - \overline{X}_1\right)^2 + \sum\left(X_2 - \overline{X}_2\right)^2 + \sum\left(X_3 - \overline{X}_3\right)^2 + \sum\left(X_4 - \overline{X}_4\right)^2$$

We have the mean of each sample as

$\overline{X}_1 = 9,\ \overline{X}_2 = 11,\ \overline{X}_3 = 12,\ \overline{X}_4 = 12$

Thus

$$\sum\left(X_i - \overline{X}_i\right)^2 = (8-9)^2 + (10-9)^2 + (12-9)^2 + (10-9)^2 + (5-9)^2$$
$$+ (12-11)^2 + (12-11)^2 + (10-11)^2 + (8-11)^2 + (13-11)^2$$
$$+ (10-12)^2 + (13-12)^2 + (11-12)^2 + (12-12)^2 + (14-12)^2$$
$$+ (12-12)^2 + (15-12)^2 + (13-12)^2 + (10-12)^2 + (10-12)^2$$
$$= 1 + 1 + 9 + 1 + 16 + 1 + 1 + 1 + 9 + 4 + 4 + 1 + 1 + 0 + 4$$
$$+ 0 + 9 + 1 + 4 + 4$$
$$= 28+16+10+18=72$$

Then SSW = 28 + 16 + 10 + 18 = 72

Now, the variance within samples, σ^2_{within}, or MSW is given by

$$MSW = \frac{SSW}{df} = \frac{SSW}{(N-K)} = \frac{72}{20-4} = \frac{72}{16} = 4.5$$

Then the F-ratio $= \frac{MSB}{MSW} = \frac{10}{4.5} = 2.22$.

We can construct an ANOVA table for the problem solved above as follows:

ANOVA Table

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Variance Ratio F**
Treatment	SSB = 30	(k – 1) = 3	$MSB = \frac{SSB}{(k-1)}$ $MSB = \frac{30}{3} = 10$	$\frac{MSB}{MSW}$ = 2.22
Within (or error)	SSW = 72	(N – k) = 16	$MSW = \frac{SSW}{(N-k)}$ $MSW = \frac{72}{16} = 4.5$	
Total	SST = 102			

Now, we check for the critical value of F from the table for α = 0.05 and degrees of freedom as follows:

df(numerator) = (k – 1) = (4 – 1) = 3

df (denominator) = (N – k) = (20 – 4) = 16

This value of F from the table is given as 3.24. Now, since our calculated value of F = 2.22 is less than the critical value of F = 3.24, we cannot reject the null hypothesis.

Example 2: A department store chain is considering building a new store at one of the three locations. An important factor in making such a decision is the household income in these areas. If the average income per household is similar then they can pick any one of these locations. A random survey of various households in each location is undertaken and their annual combined income is recorded. This data is tabulated as follows:

Annual Household Income ($ 1, 000s)

	Area (1)	Area (2)	Area (3)
	70	100	60
	72	110	65
	75	108	57
	80	112	84
	83	113	84
	-	120	70
	-	100	-
Total	380	763	420

Test if the average income per household in all these localities can be considered as the same at $\alpha = 0.01$.

Solution: If μ_i denotes the mean of the i[th] area (i= 1, 2, 3) then the null hypothesis is:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$\text{against } H_1 : \text{At least two means are not equal.}$$

The null hypothesis can be tested by computing the F-ratio for the data given and then comparing it with the critical ratio of F from the table.

As before, let us first calculate the values of SSB and SSW.

Here: $\overline{X}_1 = \dfrac{380}{5} = 76 \quad \overline{X}_2 = \dfrac{763}{7} = 109 \qquad \overline{X}_3 = \dfrac{420}{6} = 70$

so that, $\overline{\overline{X}} = \dfrac{76 + 109 + 70}{3} = \dfrac{255}{3} = 85.$

Then, $SSB = 5(76 - 85)^2 + 7(109 - 85)^2 + 6(70 - 85)^2$

$$= 405 + 4032 + 1350 = 5787$$

$$SSW = \sum_{i=1}^{n_1}(X_{i1} - \overline{X}_1)^2 + \sum_{i=-}^{n_2}(X_{i2} - X_2)_2 + \sum_{i=1}^{n_{k1}}(X_{ik} - \overline{X}_k)^2$$

where $n_i = 5 \; n_1 = 7 \quad n_3 = 6 \;\; k = 3$

$$SSW = \sum_{i=1}^{5}(X_{i1} - \overline{X}_1)^2 + \sum_{i=1}^{7}(X_{i2} - \overline{X}_2)^2 + \sum_{i=1}^{6}(X_{i3} - \overline{X}_3)^2$$

$$= (70 - 76)^2 + (72 - 76)^2 + (75 - 76)^2 + (80 - 76)^2 + (83 - 76)^2$$

$$+(100-109)^2+(110-109)^2+(108-109)^2+(112-109)^2$$

$$+(113-109)^2+(120-109)^2+(100-109)^2$$

$$+(60-70)^2+(65-70)^2+(57-70)^2+(84-70)^2+(84-70)^2+(70-70)^2$$

= 36 + 16 + 1 + 16 + 49 +81 + 1 + 1 + 9 + 6 + 121 + 81

+100 + 25 +169 + 196 + 196 + 0

= 118+310 + 686

Then, SSW = 118 + 310 + 686 = 1114.

Now, $$\text{MSB} = \frac{\text{SSB}}{(k-1)} = \frac{5787}{2} = 2893.5$$

$$\text{MSW} = \frac{\text{SSW}}{(N-k)} = \frac{1114}{15} = 74.26$$

Then, $$F = \frac{\text{MSB}}{\text{MSW}} = \frac{2893.5}{74.26} = 38.96$$

We can construct an ANOVA table for the problem solved above as follows:

ANOVA Table

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**F**
Treatment	SSB =5787	(k – 1) = 2	$\text{MSB} = \frac{\text{SSB}}{(k-1)}$ $\frac{5787}{2} = 2893.5$	$F = \frac{\text{MSB}}{\text{MSW}}$ = 38.96
Within (or error)	SSW =1114	(N – k) = 15	$\text{MSW} = \frac{\text{SSW}}{(N-k)}$ $\frac{1114}{15} = 74.26$	
Total	SST =6901			

The critical value of F from table for α = 0.01 and df 2 and 15 respectively is 6.36. Since our calculated value of F is higher than the table value of F, we cannot accept the null hypothesis.

CHECK YOUR PROGRESS 2

Note: i) Check your answers with those given at the end of the unit.

3) There are three sections of an introductory course in Statistics. Each section is being taught by a different professor. There are some complaints that at least one of the professors does not cover the necessary material. To make sure that all the students receive the same level of material in a similar manner, the chairperson of the department has prepared a common test to be given to students of the three sections. A random sample of seven students is selected from each class and their test scores out of a total of 20 points are tabulated as follows;

Students	Section (1)	Section (2)	Section (3)
1	20	12	16
2	18	11	15
3	18	10	18
4	16	14	16
5	14	15	16
6	18	12	17
7	15	10	14
Total	119	84	112

Do you think that at 95% confidence level, there is significant difference in the average test scores of students taught by the different professors?

4) Able Insurance Company wants to test whether three of its salesmen, A, B, and C, in a territory make similar number of appointments with prospective customers during a given period of time. A record of previous four months showed the following results for the number of appointments made by each salesman for each month.

	Salesman		
Month	**(A)**	**(B)**	**(C)**
1	8	6	14
2	9	8	12
3	11	10	18
4	12	4	8
Totals	40	28	52

Do you think that at 95% confidence level, there is significant difference in the average number of appointments made by the salesmen per month?

12.5 TWO-WAY ANALYSIS OF VARIANCE (ANOVA)

In the previous Section we considered the case where only one predictor/ independent variable/explanatory was categorized at different levels. In this section, let us consider the case with two categorical predictors, each categorized at different levels and a continuous response variable. Then it is called two-way classification and the analysis is called Two-way ANOVA.

In such an ANOVA, generally we have an experiment in which we simultaneously study the effect of two factors in the same experiment. For each factor, there will be a number of classes/groups or levels. In the fixed effect model, there will be only fixed levels of the two factors. We shall first consider the case of one observation per cell. Let the factors be A and B and the respective levels be $A_1, A_2, \ldots, A_r$ and $B_1, B_2, \ldots, B_{qs}$. Let y_{ij} be the observation/response/dependent variable under the i^{th} level of factor A and j^{th}level of factor B. Further, let $\mu_{1A}, \mu_{2A}, \ldots, \mu_{rA}$be the means of levels $A_1, A_2, \ldots, A_r$ and $\mu_{1B}, \mu_{2B}, \ldots, \mu_{sB}$ be the means of levels $B_1, B_2, \ldots, B_s$in the population. The observations then can be represented in a table as follows:

TWO-WAY CLASSIFIED DATA

A/B	B_1	B_2	…	B_j	…	Bs	TOTAL	MEAN
A_1	y_{11}	y_{12}	…	y_{1j}	…	y_{1s}	$y_{1.}$	□$_{1.}$
A_2	y_{21}	y_{22}	…	y_{2j}	…	y_{2s}	$y_{2.}$	□$_{2.}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{is}	$y_{i.}$	□$_{i.}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_r	Y_{r1}	Y_{r2}	…	Y_{rj}	…	Y_{rs}	$Y_{r.}$	□$_{r.}$
Total	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.s}$	$y_{..} = G$	
Mean	□$_{.1}$	□$_{.2}$	…	□$_{.j}$	…	□$_{.s}$	-	□$_{..}$

Mathematical Model

Here, the mathematical model may be written as

$y_{ij} = \mu_{ij} + e_{ij}$ wheree_{ij}'s are error terms.

12.5.1 Assumptions of Two-Way ANOVA

For the validity of F-test, the following assumptions should be satisfied:

(i) All the observations y_{ij} are independent;

(ii) Different effects (effects of levels of factor A, effects of levels of factor B and error effect) are additive in 'nature';

(iii) e_{ij}'s are independently and identically distributed normal variables with mean zero and variance σ_e^2 that is, $e_{ij} \approx \text{i i d } N(0, \sigma_e^2)$; and

(iv) There is no interaction between different levels of factor A and B.

12.5.2 Test of significance in Two-way ANOVA

As we have discussed above that if we are interested in studying the simultaneous effect of two independent factors on the dependent variable, we use two-way ANOVA. For example, we may wish to study the simultaneous effects of five varieties of wheat (first criterion) and four different types of fertilizers (second criterion) on the yield (dependent variable) or test the stress level of employees in three different organizations in different regions, and so on. In such situations, we can also apply two separate one-way ANOVA for each treatment/factor. However, it is more advantageous to use two-way ANOVA because the variance can be reduced by introducing the second factor.

In two-way ANOVA, the total variation in the data is divided into three components: variation due to the first criterion (factor), variation due to the second criterion (factor) and variation due to error.

Now let us discuss the testing procedure of two-way ANOVA briefly mentioning the main steps and formulae as follows:

Step 1: We first formulate the null hypothesis (H_0) and alternative hypothesis (H_1). In two-way ANOVA, we can test two hypotheses simultaneously: one for different levels of factor A and the other for different levels of factor B. If factor A has r levels, we can set up the null and alternative hypotheses as follows:

$$H_{0A} : \alpha_1 = \alpha_2 = ... = \alpha_r$$

$$H_{1A} = \text{At least one } \alpha_i \neq \alpha_j (i \neq j = 1,2,...,r)$$

Similarly, if factor B has s levels, we can set up the null and alternative hypotheses as follows:

$$H_{0B} : \beta_1 = \beta_2 = ... = \beta_s$$

$$H_{1B} = \text{At least one } \beta_i \neq \beta_j (i \neq j = 1,2,...,s)$$

Step 2: We calculate the Correction factor (CF) and the raw sum of squares (RSS) using the formulae given below:

Correction Factor $(CF) = \dfrac{G^2}{N}$ andRaw Sum of Squares $(RSS) = \sum_{i=1}^{r}\sum_{j=1}^{s} y_{ij}^2$

where G = sum of all observations, N = the total number of observations, and y_{ij} = observation of the i^{th} level of factor A and the j^{th} level of factor B.

Step 3: We calculate the total sum of squares (TSS), the sum of squares between rows or sum of squares due to factor A (SSA), the sum of squares between columns of sum of squares due to factor B (SSB) and sum of squares due to error (SSE) as follows:

$$\text{Total Sum of Squares(TSS)} = \text{RSS} - \text{CorrectionFactor(CF)}$$

$$\text{Sum of Squares due to Factor A (SSA)} = \frac{1}{q}\sum_{i=1}^{r} y_i^2 - \text{Correction Factor (CF)}$$

$$\text{Sum of Squares due to Factor B (SSB)} = \frac{1}{p}\sum_{i=1}^{s} y_i^2 - \text{Correction Factor (CF)}$$

$$\text{Sum of Squares due to Error (SSE)} = \text{TSS} - \text{SSA} - \text{SSB}$$

where y_i = the sum of the observations of the i^{th} level of factor A.

y_j = the sum of the observations of the j^{th} level of factor B.

Step 4: We determine the degrees of freedom (df) as

The degrees of freedom (df) for factor A = r – 1

The degrees of freedom (df) for factor B = s – 1

The degrees of freedom (df) for error = (r – 1) (s – 1)

Step 5: We obtain the various mean sums of squares as follows:

$$\text{Mean sum of squares due to factor A (MSSA)} = \frac{\text{SSA}}{r-1}$$

$$\text{Mean sum of squares due to factor B (MSSB)} = \frac{\text{SSB}}{s-1}$$

$$\text{Mean sum of squares due to error (MSSE)} = \frac{\text{SSE}}{(r-1)(s-1)}$$

Step 6: We calculate the value of the test statistics using the formulae given below:

$$F_A = \frac{\text{MSSA}}{\text{MSSE}}$$

$$F_B = \frac{\text{MSSB}}{\text{MSSE}}$$

ANOVA TABLE

Sources of Variation	**DF**	**SS**	**MSS**	**F-Test**
Variation Due to	(r-1)	SSA	$(\text{MSSA}) = \frac{\text{SSA}}{r-1}$	$F_A = \frac{\text{MSSA}}{\text{MSSE}}$

Factor A				
Variation Due to Factor B	(s-1)	SSB	$(MSSB) = \dfrac{SSB}{s-1}$	$F_B = \dfrac{MSSB}{MSSE}$
Variation Due to Error	(r-1)(s-1)	SSE	$(MSSE) = \dfrac{SSE}{(r-1)(s-1)}$	
Total	rs-1	TSS		

Step 7: We take decisions about the null hypotheses for factor A and factor B as explained below:

(i) Compare the calculated value of F_A with tabulated value of F_A at the respective df's. If calculated value is greater than the tabulated value then reject the hypothesis H_{0A}, otherwise it may be accepted.

(ii) Compare the calculated value of F_B with tabulated value of F_B at the respective df's. If calculated value is greater than the tabulated value then reject the hypothesis H_{0B}, otherwise it may be accepted.

Now after discussing the procedure of Two-way ANOVA test let us practice to solve some examples:

Example 3: Future group wishes to enter the frozen shrimp market. They contract a researcher to investigate various methods of groups of shrimp in large tanks. The researcher suspects that temperature and salinity are important factor influencing shrimp yield and conducts a two-way analysis of variance with their levels of temperature and salinity. That is each combination of yield for each (for identical gallon tanks) is measured. The recorded yields are given in the following chart:

Categorical variable Salinity (in pp)

Temperature	700	1400	2100	Total	Mean
60^0 F	3	5	4	12	4
70^0 F	11	10	12	33	11
80^0 F	16	21	17	54	18
Total	30	36	33	99	11

Compute the ANOVA table for the model.

Solution: Since in each all there is one observation. So we will use the model.

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

where, y_{ij} is the yield corresponding to i^{th} temperature and j^{th}salinity, μ is the general mean, α_i is the effect due to i^{th} temperature, β_j is the effect due to j^{th}salinity and $e_{ij} \sim$ i i d N $(0, \sigma^2)$. The hypotheses to be tested are

H_{0A}: $\alpha_1 = \alpha_2 = \alpha_3$against H_{1A}: $\alpha_1 \neq \alpha_2 \neq \alpha_3 \neq 0$

H_{0B}: $\beta_1 = \beta_2 = \beta_3$against H_{1B}: $\beta_1 \neq \beta_2 \neq \beta_3 \neq 0$.

The computations are as follows:

Grand Total (G) = 99

No. of observations(N) = 9

Correction Factor CF = (99 × 99) / 9 = 1089

Raw Sum of Square (RSS) = 1401

Total Sum of Square (TSS) = RSS - CF= 1401-1089 = 312

Sum of Square due to Temperature (SST) = $(12)^2/3+(33)^2/3+(54)^2/3-1089$ = 294

Sum of Square due to Salinity (SSS)= $(30)^2/3+(36)^2/3+(33)^2/3-1089 = 6$

Sum of Square due to Error= TSS – SST – SSS= 312 -294 -6 = 12

ANOVA TABLE

Sources of Variation	**DF**	**SS**	**MSS**	**F-Test**
Due to Temperature	2	294	147	F_T=MSST/MSSE=147/3 =49
Due to Salinity	2	6	3	F_S = 3/3 = 1
Due to error	4	12	3	
Total	8	312		

Since tabulated value of $F_{2,4}$ at 5% level of significance is 10.649 which is less than the calculated F_T for testing the significant difference in shrimp yield due to differences in levels of temperature (49). So, H_{0A} isrejected. Hence, there aredifferences in shrimp yield due to temperature at 5% level of significance.

Since tabulated value of $F_{2,4}$ at 5% level of significance is 10.649 which is greater than the calculated F_S for testing the significant difference in shrimp yield due to difference in the level of salinity (1). So, H_{0B} is not rejected. Hence there is no any difference in shrimp yield due the salinity level of 5% level of significance.

CHECK YOUR PROGRESS 3

Note: i) Check your answers with those given at the end of the unit.

5) An experiment was conducted to determine the effect of different dates of planting and different methods of planting on the field of sugar-cane. The data below show the fields of sugar cane for four different data and the methods of planting:

Dates of Planting

Method of Planting	October	November	February	March
I	7.10	3.69	4.70	1.90
II	10.29	4.79	4.50	2.64
III	8.30	3.58	4.90	1.80

Carry out an analysis of the above data.

6) A researcher wants to test four diets A, B, C, D on growth rate in mice. These animals are divided into 3 groups according to their weights. Heaviest 4, next 4 and lightest 4 are put in Block I, Block II and Block III respectively. Within each block, one of the diets is given at random to the animals. After 15 days increase in weight is noted and given in the following table:

Blocks	Treatments/Diets			
	A	B	C	D
I	12	8	6	5
II	15	12	9	6
III	14	10	8	5

Perform a two-way ANOVA to test whether the data indicate any significant difference between the four diets due to different blocks.

12.6 LET US SUM UP

In this unit, we have discussed:

1. Basic ideas and concepts related to ~~and~~the technique of Analysis of Variance as applied to test the equality of several population means simultaneously;
2. Basic underlying assumptions of ANOVA, terminologies and notations which are frequently used in Analysis of Variance;
3. Different applications of ANOVA;
4. Types of data which can be analyzed through ANOVA technique and the meaning of one -way and two-way classified data as well as the concept of one-way and two-way ANOVA;
5. Formation of null and alternative hypotheses to be tested and different steps of One-way ANOVAfor testing the hypothesis under consideration;
6. Null and alternative hypotheses to be tested and different steps of two-way ANOVA for testing the hypothesis under consideration;

12.7 KEY WORDS

Analysis of variance (ANOVA): An overall test of the null hypothesis for more than two population means.

Two-factor ANOVA: A more complex type of analysis of variance that tests whether differences exist among population means categorized by two factors or independent variables.

12.8 SUGGESTED FURTHER READING/ REFERENCES

Witte, R., & Witte, J. (2017). *Statistics*. Hoboken, NJ: John Wiley & Sons.

12.9 ANSWERS TO CHECK YOUR PROGRESS

1) Please refer to section 12.6excluding sub-sections 12.3.1 and 12.3.2.

2) Please refer to section 12.3.3.

3) The null hypothesis states that there are no differences among the mean scores of the three sections(denoted by μ_i, i = 1, 2, 3), so that we have:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \text{At least two means differ.}$$

a) As we know, the sum squares between the samples is given by:

$$SSB = \sum_{i=1}^{k} n_i (\overline{X}_i - \overline{\overline{X}})^2$$

In our case, we have 3 samples, therefore,

$$\overline{X}_1 = \frac{119}{7} = 17,\ \overline{X}_2 = \frac{84}{7} = 12,\ \overline{X}_3 = \frac{112}{7} = 16$$

And hence, $\overline{\overline{X}} = \frac{45}{3} = 15$

Then, $SSB = 7(17-15)^2 + 7(12-15)^2 + 7(16-15)^2$ =28+ 63 + 7 = 98.

b) Sum of squares within (SSW) samples is calculated as follows:

$$SSW = \sum_{I=1}^{n_1}(X_{i1} - \overline{X}_1)^2 + \sum_{i=1}^{n_2}(X_{i2} - \overline{X}_2)^2 + \sum_{i=1}^{n_k}(X_{ik} - \overline{X}_k)^2$$

$$\begin{aligned}
&= (20-17)^2 + (18-17)^2 + (18-17)^2 + (16-17)^2 \\
&+ (14-17)^2 + (18-17)^2 + (15-17)^2 \\
&+ (12-12)^2 + (11-12)^2 + (10-12)^2 + (14-12)^2 \\
&+ (15-12)^2 + (12-12)^2 + (10-12)^2 \\
&+ (16-16)^2 + (15-16)^2 + (18-16)^2 + (16-16)^2 \\
&+ (16-16)^2 + (17-16)^2 + (14-16)^2 \\
&= 58
\end{aligned}$$

Further, we have N = 21, k = 3 so that k-1= 2 and N-k= 18. Therefore, calculated

$$F = \frac{SSB/df}{SSW/df} = \frac{98/2}{58/18} = 15.21$$

The tabulated $F_{2,18}$ at 5% level of significance is 3.55. Hence, since $F_{tab}<F_{cal}$; H_0 is rejected implying that mean scores of three sections are not same.

4) In the usual notations,our null hypothesis is

H_0: $\mu_1 = \mu_2 = \mu_3$, and H_1 at least two means are unequal.

Here $\overline{X}_1 = \frac{40}{4} = 10 \quad \overline{X}_2 = \frac{28}{4} = 7 \quad \overline{X}_3 = \frac{52}{4} = 13$

so that, $\overline{\overline{X}} = \frac{10+7+13}{3} = \frac{30}{3} = 10.$

Then, $SSB = 4(10-10)^2 + 4(7-10)^2 + 4(13-10)^2 = 0 + 36 + 36 = 72$

Degrees of freedom = df = (k – 1) = (3 – 1) = 2.

$$SSW = \sum_{i=1}^{4}(X_{i1} - \overline{X}_1)^2 + \sum_{i=1}^{4}(X_{i2} - \overline{X}_2)^2 + \sum_{i=1}^{4}(X_{i3} - \overline{X}_3)^2$$

$$= (8-10)^2 + (9-10)^2 + (10-10)^2 + (12-10)^2$$

$$+ (6-7)^2 + (8-7)^2 + (10-7)^2 + (4-7)^2$$

$$+ (14-13)^2 + (12-13)^2 + (18-13)^2 + (8-13)^2 = 82$$

Degrees of freedom = df = (N – k) = (12 – 3) = 9.

Then, the F-ratio is given as:

$$F = \frac{SSB/df}{SSW/df} = \frac{72/2}{82/9} = \frac{36}{9.1} = 3.95$$

The F-ratio from the table at 95% confidence level and degrees of freedom 2 and 9 respectively is given as 4.26.

Since our calculated value of F is less than the tabulated value of F, we cannot reject the null hypothesis.

5) Here we have H_{0A} : There is no differencein the average yield due to different methods of planting, that is,

$H_{0A} : \alpha_1 = \alpha_2 = \alpha_3$ against, H_{1A}: $\alpha_1 \neq \alpha_2 \neq \alpha_3$

H_{0B} : There is no difference in the average yield due to different dates of planting, that is,

$H_{0B}: \beta_1 = \beta_2 = \beta_3 = \beta_4$

$H_{1B}: \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$

G= $\sum\sum y_{ij}$ = Total of all observations

=7.10+3.69+4.70+1.90+10.29+4.79+4.58+2.64+8.30+ 3.58+4.90+1.80

=58.28

N = No. of observations = 12

Correction Factor = CF = $\frac{G^2}{N} = \frac{58.28 \times 58.28}{12} = 283.0465$

Raw Sum of Square (RSS) =

$$(7.10)^2 + (3.69)^2 + (4.70)^2 + (1.90)^2 + (10.29)^2 + (4.79)^2 + (4.58)^2 + (2.64)^2 + (8.30)^2 + (3.58)^2 + (4.90)^2 + (1.80)^2 = 355.5096$$

Total Sum of Square (TSS) = RSS – CF = 355.5096-283.0465= 72.4631

Sum of Square due to Dates of Planting (SSD)

$$= \frac{D_1^2}{3} + \frac{D_2^2}{3} + \frac{D_3^2}{3} + \frac{D_4^2}{3} - CF$$

$$= \frac{(25.69)^2}{3} + \frac{(12.06)^2}{3} + \frac{(14.18)^2}{3} + \frac{(6.35)^2}{3} - 283.0465$$

Sum of Square to Method of Planting (SSM) $= \frac{M_1^2}{4} + \frac{M_2^2}{4} + \frac{M_3^2}{4} - CF$

$$= \frac{(17.39)^2}{4} + \frac{(22.31)^2}{4} + \frac{(15.58)^2}{4} - 283.0465$$

= 286.3412-283.0465 = 3.2947

Sum of Square due to Error (SSE)= TSS-SSD-SSM

= 72.4631-3.2947-65.8917 = 3.2767

$\text{MSSD} = \frac{65.8917}{3} = 21.9639$, $\text{MSSM} = \frac{3.2947}{2} = 1.6473$,

$\text{MSSE} = \frac{3.2767}{6} = 0.5461$

For testing $H_{0A} : F_M$ is $\frac{\text{MSSM}}{\text{MSSE}} = \frac{1.6473}{0.5461} = 3.02$

For testing $H_{0B} : F_D$ is $\frac{\text{MSSD}}{\text{MSSE}} = \frac{21.9639}{0.5461} = 40.22$

The tabulated value of $F_{2,6}$ at 5% level of significance is 5.14 which is greater than the calculated value of F_M (3.02) so H_{0A} is accepted. So, we conclude that there is no significant difference among the different methods of planting.

The tabulated value of $F_{3,6}$ at 5% level of significance is 4.76 which is less than calculated value of F_D (40.22). So, we reject the null hypothesis H_{0B} . Hence there is a significant difference among the dates of planting.

6) Null Hypotheses are:

H_{01}: There is no significant difference between mean effects of diets.

H_{02}: There is no significant difference between mean effects of different blocks.

Against the alternative hypothesis

H_{11}: There is significant difference between mean effects of diets

H_{12}: There is significant difference between mean effects of different blocks.

Blocks	Treatments/Diets				
	A	B	C	D	Totals
I	12	8	6	5	31
II	15	12	9	6	42
III	14	10	8	5	37
Totals	$T_{1.}$=41	$T_{2.}$=30	$T_{3.}$=23	$T_{4.}$=16	110

Squares of observations

Blocks	**Treatments/Diets**				**Totals**
	A	B	C	D	
I	144	64	36	25	269
II	225	144	81	36	486
III	196	100	64	25	385
Totals	565	308	181	86	1140

Grand Total = G = $\sum\sum y_{ij}$ = 110

Correction Factor (CF) = $\dfrac{G^2}{N} = \dfrac{(110)^2}{12} = 1008.3333$

Raw Sum of Squares (RSS) = $\sum\sum y_{ij}{}^2 = 1140$

Total Sum Squares (TSS) = RSS - CF = 1140 - 1008.3333 = 131.6667

Sum of Squares due to Treatments/ Diets (SST)

$$= \frac{T_{1.}^2}{3} + \frac{T_{2.}^2}{3} + \frac{T_{3.}^2}{3} + \frac{T_{4.}^2}{3} - CF$$

$$= \frac{1}{3}[(41)^2 + (30)^2 + (23)^2 + (16)^2] - 1008.3333$$

$$= \frac{1}{3}[1681 + 900 + 529 + 256] - 1008.3333$$

$$= 1122 - 1008.3333 = 113.6667$$

Sum of squares due to block (SSB) = $\frac{1}{4}\left(T_{.1}^2 + T_{.2}^2 + T_{.3}^2\right) - CF$

$$= \frac{1}{4}[(31)^2 + (42)^2 + (37)^2] - 1008.3333$$

$$= 1023.5 - 1008.3333 = 15.1667$$

Sum of Squares due to Errors (SSE) = TSS – SST − SSB

$$= 131.6667 - 113.6667 - 15.1667 = 2.8333$$

Mean Sum of Squares due to Treatments (MSST)

$$= \frac{SST}{df} = \frac{113.6667}{3} = 37.8889$$

Mean Sum of Squares due to Blocks (MSSB)

$$= \frac{SSB}{df} = \frac{15.1667}{2} = 7.58335$$

Mean Sum of Squares due to Errors (MSSE)

$$= \frac{SSE}{df} = \frac{2.8333}{6} = 0.4722$$

ANOVA TABLE

Source of Variation	**SS**	**df**	**MSS**	**F**
Between Treatments/ Diets	113.6667	3	$\frac{113.6667}{3} = 37.8889$	$F_1 = \frac{37.8889}{0.4722} = 80.2391$
Between Blocks	15.1667	2	$\frac{15.1667}{2} = 7.58335$	$F_2 = \frac{7.58335}{0.4722} = 160.596$
Due to Errors	2.8333	6	$\frac{2.8333}{6} = 0.4722$	
Total	131.6667	11		

Tabulated F at 5% level of significance with (3, 6) degree of freedom is 4.76 & tabulated F at 5% level of significance with (2, 6) degree of freedom is 5.14

Conclusion: Since calculated F_1 > Tabulated F_1, so we reject H_{01} and conclude that there is significant difference between mean effect of diets.

Also calculated F_2 is greater than tabulated F_2, so we reject H_{02} and conclude that there is significant difference between mean effect of different blocks.