Chapter 18

ANALYSIS OF VARIANCE

18.1. Introduction

We have already discussed in Chapter-16 how to test the significance of difference between two population means. But in business decision making, very often, we might be interested to take decision whether there is significant difference among various population means. For example, if a car maker company wants to compare the average fuel consumptions of three models of cars, the company may select a number of drivers for testing purpose, say, 7 drivers are randomly assigned to car of model A, 8 to model B and 5 to model C, then kilometer per liter of fuel consumption for all these three independent random samples will be collected, and mean fuel consumption of different models of cars can be compared using ANOVA technique. Some other applications of ANOVA may be as follows :

i) a company may be concerned to compare the average sales of different salesmen,

ii) a mobile phone company may be thought of checking whether there is any significant difference in the average number of daily calls made by users in different areas of a big city,

iii) an economist may be interested to find out if there is significant difference among average monthly expenditures of families of same number of members in different localities,

iv) a production manager may be interested to compare effectiveness of different promotional devices in terms of sales, etc.

The statistical technique used for taking decision in all such cases is known as Analysis of Variance (ANOVA), developed by R.A. Fisher and popularly known as F-test.

It may seem odd that the technique is called "Analysis of Variance" rather than "Analysis of Means." As it is discussed above, the name is appropriate because inferences about means are made by analyzing variances. The analysis of variance is based on development of two independent estimates of common population variance. One estimate is based on variability among the sample means themselves, called between variation, and another is based on variability of data within each sample, called within variation. By comparing these two estimates of population variance, we will be able to determine whether the population means are equal or not. Since the methodology uses comparison of variances, it is also referred to as analysis of variance (ANOVA).

Definition. ANOVA. Analysis of Variance (ANOVA) is a statistical procedure used to test significance of difference among three or more population means.

A table is to be constructed to summarize the necessary calculations for conducting the testing procedure of more than two population means where variations due to different factors or components are shown. Hence, ANOVA is also recognized as the systematic procedure of partitioning the total variations present in a set of observations into variations due to various components. For example, if the company is concerned to compare the average sales of different salesmen, all collected sales figures will be considered as sales observations, and in ANOVA table, total variations in the sales will be partitioned into two components, viz. the variations due to salesmen (between variation) and variations due to other non-assignable factors (within variation). Actually, the total sum of squares is partitioned into between sum of squares and within or error sum of squares as follows :

$$SS\text{ (Total)} = SS\text{ (Between)} + SS\text{ (within)} \quad \text{or,} \quad SST = SSB + SSW$$

SSW is also sometimes written as sum of squares due to error or SSE.

Assumptions. The methodology of ANOVA is based on the following two assumptions.

i) Each sample is drawn randomly from a normal population and each sample is independent of other samples,

ii) The populations from which the samples are drawn are normally distributed, and

iii) All the population have equal variance, that means, if there are k independent samples, then, the corresponding population variances are equal, i.e.

$$\sigma_1^2 = \sigma_1^2 = \ldots\ldots\ldots\ldots = \sigma_k^2.$$

The observations in the sample may be classified according to one factor or two factors. If the observations are classified according to one factor, then it is called a one-way ANOVA, while if the observations are classified according to two factors, then it is called a two-way ANOVA.

18.2. One-way ANOVA

Most of the business applications involve experiments in which different population are classified with respect to only one attribute of interest, such as gasoline consumption of different *brand* of cars, *flavor* preference of consumers, sales of *salesmen, area* wise daily phone calls, effectiveness of *devices* in terms of sales, etc. In all such cases observations in the sample are classified into two or more groups or levels based on a single attribute (showed in italics) and it is called one-way classification of data or one-way ANOVA.

18.2.1. Factor, Level and Treatment. In case of ANOVA for gasoline consumption of different model of cars, we may be interested to study the effect (say, in terms of gasoline consumption per kilometer) of a single factor brand at different groups or levels (say, types of model A, B, C, etc). These groups or levels are termed as treatments in the context of ANOVA. Again, if it is desired to study the area wise daily phone calls, then the area wise number of daily phone calls in different area will be collected, hence area is a factor and different area such as Dhaka north, Dhaka south, Gazipur, etc are groups or levels. In former example, model A, B, C etc and in later example area Dhaka north, Dhaka south, Gazipur, etc are also known as treatments.

Definition. One-way ANOVA. If the data are classified into two or more groups on the basis of only one criterion and it is required to test the equality of means of observations between the groups, then it is called one-way ANOVA.

Definition. Factor. The qualitative characteristic significance of whose deferent levels are tested is termed a factor.

Definition. Treatment. Different groups or levels or categories of the attribute or factor considered in an ANOVA are termed as treatments.

Definition. Experimental Unit. The unit or object on which experiment is undertaken, or treatments are applied.

18.2.2. Principles of ANOVA. In general, ANOVA experiments need to satisfy three principles - replication, randomization and local control, however, of these three, only replication and randomization have to be satisfied while designing and implementing any one-way ANOVA experiment.

Replication refers to the repeated application of each individual level of the factor to multiple subjects. In the example of brand of cars, in order to apply the principle of replication, a number of drivers are selected to drive each brand of car and gasoline consumption data from all the driven cars are collected. Here, number of cars of each brand is the number of replications of that brand.

Randomization refers to the random allocation of the experimental units. In our example, each car is an experimental unit, and these were selected randomly and independently for each of the brand.

18.2.3. Layout of One-way Classification. Thus the observations obtained for k independent random samples based on one-factor classification can be arranged as shown in following table.

Table 18.1. Sample observations for one-way analysis of variance

Samples No. (Treatment)			
1 (x_1)	2 (x_2)	....	k(x_k)
x_{11}	x_{21}	....	x_{k1}
x_{12}	x_{22}	....	x_{k2}
......	......	....	.....
x_{1n_1}	x_{2n_2}	....	x_{kn_k}

where, x_{ij} = jth observation of ith treatment or level $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n_i)$,

n_i = number of observations or sample size for ith treatment or level,

n = total number of observations in all samples combined (i.e. $n = n_1 + n_2 + \ldots + n_k$).

18.2.4. Decomposition of Total Variation and Degrees of Freedom. As it is mentioned in the introduction part of ANOVA that total sum of squares is partitioned into two components namely, sum of squares between group and sum of squares within group. We can view this as the partition of total variation. This decomposition provides the basis for the analysis of variance test of equality of group population means. Thus, for the above layout, different sum of squares are defined as

Sum of squares between group: $\text{SSB} = \sum_{i=1}^{k} n_i(\bar{T}_i - \bar{X})^2$.

Sum of squares within group: $SSE = SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_j}(x_{ij} - \bar{T}_i)^2$

Total Sum of squares: $SST = \sum_{i=1}^{k}\sum_{j=1}^{n_j}(x_{ij} - \bar{X})^2$

Where, $\bar{T}_i$ is the ith sample mean = $\frac{\sum_{j=1}^{n_j} x_{ij}}{n_i}$, T_i is the sum of observations of ith treatment

$\bar{X}$ is the sample mean of all observations called grand mean $= \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_j} x_{ij}}{n} = \frac{\sum_{i=1}^{k} T_i}{n}$.

Thus, decomposition of total sum of squares implies that,

$$SST = SSB + SSW \quad \text{or} \quad \sum\sum(x_{ij} - \bar{X})^2 = \sum n_j(\bar{T}i - \bar{X})^2 + \sum\sum(x_{ij} - \bar{T}i)^2$$

In case of one-way ANOVA, the degrees of freedom for different sources of variations are computed as follows :

i) if there are k independent samples, then degrees of freedom associated with the between or treatment variation are $k-1$,

ii) if total number of observations are $n = n_1 + n_2 + ...+ n_k$). where n_i is the number of observations in the ith sample, then degrees of freedom for total variation are $n-1$.

iii) then degree of freedom for within variation would be $n-1-(k-1) = n-k$.

The sum of the degrees of freedom for between variation and within variation will be equal to total degrees of freedom.

18.2.5. Hypothesis to be tested. Suppose our aim is to make inference about k sample or treatment means based on sample data. Let $\mu_1, \mu_2, ..., \mu_k$, be the mean of the population of observations for 1st, 2nd, ..., kth sample respectively, then one-way analysis of variance framework is designed to test the null hypothesis that there is no significant difference among the treatment means, symbolically,

$$H_0: \mu_1 = \mu_2 = ... = \mu_k \text{ against}$$
$$H_1: \mu_i \neq \mu_j \text{ for at least one pair of : } \mu_i, \mu_j$$

We have already discussed the some features of F distribution (Section 15.9.3) developed by R.A. Fisher which is simply the ratio of two variances. He determined that the difference of the between variation and within variation values could be expressed as a ratio to be defined as the F-value, hence,

$$F = \frac{\text{Between variation}}{\text{Within variation}} = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{MS(\text{Between})}{MS(\text{within})} = \frac{SSB/(k-1)}{SSW/(n-k)} \sim F_{(k-1),(n-k)}$$

Where MS stands for mean sum of squares or variance. It is seen that, the sum of squares obtained for between and within variations are to be divided by corresponding degrees of freedom in order to compute the mean sum of squares for F-ratio. This is done to obtain unbiased estimates of variances due to different components. If population means are exactly the same, the between mean sum of squares will be equal to the within mean sum of squares, and the value of F will be equal to unity. It is logical to conclude that if the population means are not equal then their samples means will also vary greatly from one another, resulting in a larger value of between mean sum of squares as well as value of F, because within mean sum of squares is based only on the sample variances and not sample means, and hence not affected by the differences in sample means. Accordingly, the larger the value of F, the more likely the decision to reject the null hypothesis. However, like other tests of hypothesis, discussed in chapter 15, the decision of rejecting or accepting the null hypothesis depends on whether the calculated value of F_{Cal} is larger or smaller than the critical value of F_{Tab} at a given level of significance with computed number of degrees of freedom, i.e., the decision rule for an ANOVA test is as follows :

Reject H_0 at $100\alpha\%$ level of significance if $F_{Cal} > F_{k-1,(n-k)}$, otherwise do not reject H_0.

18.2.6. Advantages. The advantages of one-way classification are as follows.

i) One of the principal advantages of this technique is that the number of observations need not be the same in each group,

ii) Layout of the design and statistical analysis is simple.

18.2.7. Computational procedure. For computations of between sum of squares and within sum of squares, the following steps can be followed.

Steps for between sum of squares (SSB)

To compute between sum of squares, we have to compute square of the deviations of each sample mean from the grand mean, and multiply each squared deviation by the respective number of observations, then add up the results obtained, thus steps involved in computation of SSB are as follows :

i) Calculate the sum and mean of the observations for each sample or each treatment, say $T_1 = \Sigma x_1$ is sum of observation of first sample, and $\bar{T}_1 = \frac{\Sigma x_1}{n_1}$, first sample mean, similarly, $T_2 = \Sigma x_2$ and $\bar{T}_2 = \frac{\Sigma x_2}{n_2}$ and so on,

ii) Calculate sum of all observations as $T = T_1 + T_2 + ... + T_k = \Sigma x_1 + \Sigma x_2 + ... + \Sigma x_k$,

iii) Calculate the grand mean $\bar{X}$ as $\bar{X} = \frac{\Sigma T_i}{n} = \frac{T}{N}$ where $n = \Sigma n_i$ = total number of observations,

iv) Compute the deviations of means of the various samples from the grand mean, i.e., $(\bar{T}_1 - \bar{X}), (\bar{T}_2 - \bar{X}), ..., (\bar{T}_k - \bar{X})$,

v) Square these deviations or differences individually and multiply each of these squared deviations by its respective sample size or replications (n_i) and sum up all these products to obtain SSB, i.e., $SSB = \sum_{i=1}^{k} n_i (\bar{T}_i - \bar{X})^2$.

- **Steps for within sum of squares (SSW)**

To compute within sum of squares (SSW), we have to square each deviation between the individual value of each sample and its mean, for all samples and then compute sum of squared deviations, the steps are as follows:

i) Calculate the mean of each observation given for each sample or each treatment, say $\overline{T}_1$, $\overline{T}_2$, ... $\overline{T}_k$,

ii) Consider one sample or treatment at a time and take deviation of each observation of this sample from the respective mean, for example,

For first sample or treatment, compute $(x_{11}-\overline{T}_1)$, $(x_{12}-\overline{T}_1)$, ..., $(x_{1n_{11}}-\overline{T}_1)$,

For second sample compute $(x_{21}-\overline{T}_2)$, $(x_{22}-\overline{T}_2)$, ..., $(x_{2n_2}-\overline{T}_2)$ and so on,

iii) Square these differences and add up all these squared differences which will give SSW,

iv) Divide the SSW by corresponding degrees of freedom to obtain MS (within) or MSW, the degrees of freedom in this case are obtained by subtracting the total number samples or treatments from the total number of observations i.e., df(Within) = N − k, thus,

$$MSW = \frac{\sum_{j=1}^{n_1}(x_{1j}-\overline{T}_1)^2 + \sum_{j=1}^{n_2}(x_{2j}-\overline{T}_2)^2 + \ldots\ldots + \sum_{j=1}^{n_k}(x_{kj}-\overline{T}_k)^2}{(n-k)} = \frac{SSW}{n-k}.$$

- **Steps for total sum of squares SST**

i) Take deviations of all observations from grand mean as $(x_{11}-\overline{X})$, $(x_{12}-\overline{X})$, ..., $(x_{1n_1}-\overline{X})$, $(x_{21}-\overline{X})$, $(x_{22}-\overline{X})$, ..., $(x_{2n_2}-\overline{X})$, and so on,

ii) Square these deviations and add up, which will give SST, i.e.,

$$SST = (x_{11}-\overline{X})^2 + (x_{12}-\overline{X})^2 + \ldots + (x_{1n_1}-\overline{X})^2 + (x_{21}-\overline{X})^2 + (x_{22}-\overline{X})^2 + \ldots$$

$$+ (x_{2n_2}-\overline{X})^2 + \ldots + (x_{k1}-\overline{X})^2 + \ldots + (x_{kn_k}-\overline{X})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{X})^2.$$

Alternatively, more practical or convenient way of computation of sum of squares is as follows:

- **For between sum of squares SSB**

Let T_j for $i = 1, 2, \ldots, k$, be the sum of observations or sample values for jth treatment or sample, then steps to be followed for the computation of SSB are as follows :

i) Compute sum of observations T_i, for each sample for $i = 1, 2, \ldots, k$, for example, $T_1 = \sum x_1$ = first column total, $T_2 = \sum x_2$ = second column total, and so on, $T_k = \sum x_k$ = kth column total,

ii) Compute T (grand total of all observations) as sum of T_j values i.e., $T = \sum_{i=1}^{k} T_i$ and compute the correction term (CT) as $CT = \frac{T^2}{N}$, where $N = \sum_{j=1}^{k} n_j$.

iii) Compute squares of all T_i value, divide each square by corresponding sample size i.e., compute $\frac{(T_1)^2}{n_1}, \frac{(T_2)^2}{n_2}, \ldots\ldots, \frac{(T_k)^2}{n_k}$.

iv) Add up the values obtain in step (iii) to obtain $\sum_{i=1}^{k} \frac{T_i^2}{n_i}$,

v) Subtract the value obtained in step (ii) from the value obtained in step (iv), which will provide the required value of SSB, as $SSB = \sum_{i=1}^{k} \frac{T_j^2}{n_i} - \frac{T^2}{n} = \sum_{i=1}^{k} \frac{T_j^2}{n_i} - CT$.

For within sum of squares SSW

Since the computational procedure for total sum of squares (SST) is easier than within sum of squares and we know that total sum of squares is partitioned into between sum of squares and within sum of squares, hence, within sum of squares is popularly computed using the relationship SSW=SST−SSB, the degrees of freedom for within variation is also computed using the same type of adjustment. Hence easier way of computation of SSW involves following steps.

i) Square all of the observations to obtain

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 = (x_{11}^2 + \ldots + x_{1n_1}^2 + x_{21}^2 + \ldots + x_{2n_2}^2 + \ldots + x_{k1}^2 + \ldots + x_{kn_k}^2)$$

Subtract correction term $CT = \frac{T^2}{N}$ from the value obtained in step (i) to get total sum of squares, SST, i.e.,

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{X})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{T^2}{N} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - CT.$$

ii) Subtract SSB from the computed SST to obtain SSW, i.e., SSW = SST − SSB.

Mean sum of squares are usually computed directly in ANOVA table. These are obtained by dividing sum of squares by corresponding degrees of freedom, thus,

$$\text{MS (between)} = MSB = \frac{SSB}{k-1} \quad \text{and} \quad \text{MS (within)} = MSW = \frac{SSW}{n-k}.$$

Finally, compute the value of F-ratio and compare this value with critical value of F to take decision about the validity of null hypothesis.

18.2.7. ANOVA Table. The table which is constructed to summarize or present the various computations made in analysis of variance is known as analysis of variance table or simply, ANOVA table. This table consists of five columns, namely, (i) Sources of variation, (ii) Sum of squares, (iii) Degrees of freedom, (iv) Mean sum of squares and (v) F-ratio.

Definition. ANOVA Table. A table used to summarize the calculations and results needed for analysis of variance is termed as ANOVA Table.

The format of ANOVA table for a one-way classification is as follows :

ANOVA Table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Treatment	SSB	$k-1$	$SSB/k-1$	MSB/MSW
Within (Error)	SSW	$n-k$	$SSW/n-k$	--
Total	SST	$n-1$	--	--

Decision. Reject null hypothesis in favour of alternative if calculated F-ratio is greater than the tabulated value of F at $100\alpha\%$ level of significance with $(k-1)$ and n-k degrees of freedom.

Example 18.2.1. (Comparison of Fuel consumptions of automobiles) An automobile company is interested to test whether there are any significant differences among the average fuel consumption of three makes of cars A, B and C. For this, 7 drivers are randomly assigned to A-cars, 7 to B-cars and 6 to C-cars. The following data refer to the fuel consumption (km/liter) of three model of cars A, B and C found for respective number of drivers.

Fuel consumption (km/liter) figures for three makes of cars

Car or driver No	A-car (x_1)	B-car (x_2)	C-car (x_3)
1	22.2	24.6	22.7
2	19.9	23.1	21.9
3	20.3	22.0	23.2
4	21.4	23.5	24.1
5	21.2	23.6	22.1
6	21.0	22.1	23.4
7	20.3	23.5	--

Solution. Let μ_1, μ_2, μ_3 be the population mean fuel consumptions for three types of cars A, B, C respectively, then the null hypothesis to be tested here is that the population mean consumption of fuel is the same for all three types of cars, i.e.,

$H_0: \mu_1 = \mu_2 = \mu_3$, against.

H_1 : Average fuel consumption are not the same for at least two cars.

The following table is constructed for the convenience of computations of sum of squares of different sources of variations,

Table 18.2.1a. Computation table

Driver	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2
1	22.2	24.6	22.7	492.84	605.16	515.29
2	19.9	23.1	21.9	396.01	533.61	479.61
3	20.3	22.0	23.2	412.09	484.00	538.24
4	21.4	23.5	24.1	457.96	552.25	580.81
5	21.2	23.6	22.1	449.44	556.96	488.41
6	21.0	22.1	23.4	441.00	488.41	547.56
7	20.3	23.5	--	412.09	552.25	--
Total	$T_1 = 146.3$	$T_2 = 1624$	$T_3 = 137.4$	$\sum x_1^2 = 306143$	$\sum x_2^2 = 377264$	$\sum x_3^2 = 314962$

Here, $T = \Sigma T_j = (146.3 + 162.4 + 137.4) = 446.1.$

$$CT = \frac{T^2}{n} = \frac{(4461)^2}{20} = 9950.26.$$

Total sum of squares (SST) $= (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2) - CT$

$= (3061.64 + 3772.64 + 3149.62) - 9950.26$

$= 9983.99 - 9950.26 = 33.73.$

Between sum of squares (SSB) $= \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CT$ (here, $n_i = 7, 7, 6;\ i = 1, 2, 3$)

$$= \left[\frac{(1463)^2}{7} + \frac{(1624)^2}{7} + \frac{(137.4)^2}{6}\right] - 9950.26$$

$= (3057.67 + 3767.68 + 3146.46) - 9950.2$

$= 9971.81 - 9950.26 = 21.55.$

Within sum of squares (SSW) = SST – SSB = 33.73 – 21.55 = 12.18.

Table 18.2.1b. ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between	SSB = 21.55	$(k-1) = 2$	$SSB/(k-1) = 10.78$	MSB/MSW = 15.04
Within (Error)	SSW = 12.18	$(n-k) = 17$	$SSW/(n-k) = 0.7165$	
Total	SST = 33.73	$(n-1) = 19$	--	

Here the degrees of freedom for F are 2 and 17, and from Table-7 in Appendix, we have the tabulated value of F-ratio at 5% level of significance with 2 and 17 degrees of freedom is 3.59, or $F_{2,17;0.05} = 3.59$, hence these data allow us to reject null hypothesis at 5% level of significance. That means, the mean fuel consumptions for all three types of cars are not the same.

Example 18.2.2. (Effectiveness of drugs) The time required to lessen the temperature of fever patients were recorded to test the differences in the effectiveness of drugs. Three types of drugs, say, A, B, C produced by three different companies were selected. Each of the drugs were applied to six randomly selected patients suffering from high fever. The time needed to reduce the temperature are reported in following table.

Drug A	Drug B	Drug C
15.75	12.63	9.27
11.55	11.46	8.28
11.16	10.77	8.15
9.92	9.93	6.37
9.23	9.87	6.37
8.20	9.12	5.66

Test at 1% level of significance whether population mean time needed to reduce the temperature are the same for all three drugs.

Solution. Let μ_1, μ_2, and μ_3 be the population mean time required to reduce temperature by three types of drugs A, B, C respectively, then, the hypothesis to be tested here is that three population mean time are equal, i.e.,

$H_0: \mu_1 = \mu_2 = \mu_3$, against the alternative.

H_1: Population mean time for at least two drugs is not the same.

The following table is constructed for necessary computation.

Table 18.2.2a. Table for necessary computations

Paper No.	Drug A (x_1)	Drug B (x_2)	Drug C (x_3)	x_1^2	x_2^2	x_3^2
1	15.75	12.63	9.27	248.06	159.52	85.93
2	11.55	11.46	8.28	133.40	131.33	68.56
3	11.16	10.77	8.15	124.55	115.99	66.42
4	9.92	9.93	6.37	98.41	98.60	40.58
5	9.23	9.87	6.37	85.19	97.42	40.58
6	8.20	9.12	5.66	67.24	83.17	32.04
Total	$T_1 = 65.81$	$T_2 = 63.78$	$T_3 = 44.10$	756.85	686.04	334.10

Here, $T = \sum T_i = (65.81 + 63.78 + 44.10) = 173.69.$

$$CT = \frac{T^2}{n} = \frac{(173.69)^2}{18} = 1676.01.$$

$$\text{Total sum of squares (SST)} = (\sum x_1^2 + \sum x_2^2 + \sum x_3^2) - CT$$
$$= (765.85 + 686.04 + 334.10) - 1676.01$$
$$= 1776.99 - 1676.01 = 100.98.$$

$$\text{Between sum of squares (SSB)} = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CT \quad (\text{here, } n_i = n = 6)$$

$$= \left[\frac{(65.81)^2}{6} + \frac{(63.78)^2}{6} + \frac{(44.10)^2}{6}\right] - 1676.01$$
$$= (721.83 + 677.98 + 324.14) - 1676.01$$
$$= 1723.95 - 1676.01 = 47.94.$$

Within sum of squares SSW = SST – SSB = 100.98 – 47.94 = 53.04.

Table 18.2.2b. ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between	SSB = 47.94	(k–1) = (3 – 1) = 2	SSB/(k–1) = 23.97	MSB/MSW = 6.77
Within (Error)	SSW = 53.04	(n– k) = (18 – 3) = 15	SSW/(n–k) = 3.54	
Total	SST = 100.98	(n–1) = (18 – 1) = 17		

Here, the degrees of freedom for F are 2 and 15, and from Table-7 given in Appendix, we have the tabulated value of F-ratio at 1% level of significance with 2 and 15 degrees of freedom is 6.36, or $F_{2,15;0.01} = 6.36$, while the observed value of F is 6.77, which is higher than critical or tabulated value of F, hence, null hypothesis may be rejected. It may be concluded that there is significant differences in the mean time required to reduce temperature by three types of drugs which means drugs are not equally effective.

Example 18.2.3. (Comparison of Sales prediction) Samples of four sales managers of a product from each of four major regions in Bangladesh were asked to predict percentage increase in sales volume for their territories in the next 12-months. The predictions are shown in following table.

Regions			
Dhaka	Chittagong	Rajshahi	Khulna
9.0	7.1	6.8	4.2
8..0	6.6	4.2	4.8
7.2	5.8	5.4	5.8
7.6	7.0	5.0	4.6

a) Set out the analysis of variance table;

b) Test the null hypothesis that the four population mean sales increase predictions are equal.

Solution. (a) For setting out the analysis of variance, we have to compute total sum of squares of sales figures, mean sum of squares of sales percentages between and within the regions. The following table is constructed for necessary computations.

Table 18.2.3a. Table for necessary computations

	Dhaka (x_1)	Ctg (x_2)	Rajshahi (x_3)	Khulna (x_4)	(x_1^2)	(x_2^2)	(x_3^2)	(x_4^2)
	9.0	7.1	6.8	4.2	81.00	50.41	46.24	17.64
	8..0	6.6	4.2	4.8	64.00	43.56	17.64	23.04
	7.2	5.8	5.4	5.8	51.84	33.64	29.16	33.64
	7.6	7.0	5.0	4.6	57.76	49.00	25.00	21.16
Total	31.8	26.5	21.4	19.4	254.6	176.61	118.04	95.48

Here, $T_1 = 31.8,\ T_2 = 26.5,\ T_3 = 21.4,\ T_4 = 19.4.$

So, $T = \sum T_i = (31.8 + 26.5 + 21.4 + 19.4) = 99.1$

$$CT = \frac{T^2}{n} = \frac{(99.1)^2}{16} = 613.80.$$

$$\text{Total sum of squares (SST)} = (\sum x_1^2 + \sum x_2^2 + \sum x_3^2) - CT$$
$$= (254.6 + 176.61 + 118.04 + 95.48) - 613.80$$
$$= 644.73 - 613.80 = 30.93.$$

$$\text{Between sum of squares (SSB)} = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CT$$
$$= \left[\frac{(31.8)^2}{4} + \frac{(26.5)^2}{4} + \frac{(21.4)^2}{4} + \frac{(19.4)^2}{4}\right] - 613.80$$

$= (721.83 + 677.98 + 324.14) - 613.80$

$= 636.95 - 613.80 = 23.15.$

Within sum of squares (SSW) = SST – SSB = 30.93 – 23.15 = 7.78.

Table 18.2.3b. ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between	SSB = 23.15	$(k-1)=(4-1)=3$	SSB/(k-1) = 7.72	MSB/MSW = 11.88
Within (Error)	SSW = 7.78	$(n-k)=(16-4)=12$	SSW/(n-k) = 0.65	
Total	SST = 30.93	$(n-1)=(16-1)=15$	--	

(b) The null hypothesis to be tested here is that four population mean sales increase predictions are equal, against the alternative that they are not equal, i.e.,

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4, \text{ against the alternative.}$$

Here, the degrees of freedom for F are 3 and 12, and from Table-7 given in Appendix, we have the tabulated value of F-ratio at 5% level of significance with 2 and 15 degrees of freedom is 3.49, or $F_{3,12;0.05} = 3.49$, while from ANOVA table we find that the observed value of F is 11.88, which is much higher than critical or tabulated value of F, hence, null hypothesis may be rejected. It may be concluded that there is significant differences in the four population means, that means mean percentage sales predictions are not the same for all four selected regions.

Example 18.2.4. (Performance of Machines) A certain kind of cotton fabrics are produced by of Four machines A, B, C and D are used. The number of flaws found in each of 4 samples of producing 200 square meters of cloths are recorded for each of the machines. The recorded number of flaws are given in following table.

Machines			
A	B	C	D
6	8	20	14
8	9	22	12
10	11	25	18
4	12	23	9

a) Set out the analysis of variance table
b) Do you think that there is a significant difference in the performance of four Machines?

Solution. (a) For setting out the analysis of variance, we have to compute total sum of squares of sales figures, mean sum of squares of number of faults between and within the machines. The following table is constructed for necessary computations.

Table 18.2.4a. Table for computations

	A (x_1)	B (x_2)	C (x_3)	D (x_4)	(x_1^2)	(x_2^2)	(x_3^2)	(x_4^2)
	6	8	20	14	36	64	400	196
	8	9	22	12	64	81	484	144
	10	11	25	18	100	121	625	324
	4	12	23	9	16	144	529	81
Total	28	40	90	53	216	410	2038	745

Here, total number of observations are 16,

$$T_1 = 28, \quad T_2 = 40, \quad T_3 = 90, \quad T_4 = 53,$$

So, $T = \Sigma T_j = (28 + 40 + 90 + 53) = 211.$

$$CT = \frac{T^2}{N} = \frac{(211)^2}{16} = 2782.56.$$

Total sum of squares SST $= (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2) - CT$

$= (216 + 410 + 2038 + 745) - 2782.56$

$= 3409 - 2782.56 = 626.44.$

Between sum of squares SSB $= \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CT$ (here $n_j = n = 4$)

$$= \left[\frac{(28)^2}{4} + \frac{(40)^2}{4} + \frac{(90)^2}{4} + \frac{(53)^2}{4}\right] - 2782.56$$

$= (196 + 400 + 2025 + 702.25) - 2782.56$

$= 3323.25 - 2782.56 = 540.69.$

Within sum of squares SSW = SST − SSB = 626.44 − 540.69 = 85.75.

Table 18.2.4b. ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between	SSB = 540.69	$(k-1)=(4-1)=3$	SSB/(k−1) = 180.23	MSB/MSW = 25.21
Within (Error)	SSW = 85.75	$(n-k)=(16-4)=12$	SSW/(n−k) = 7.15	
Total	SST = 626.44	$(n-1)=(16-1)=15$	--	

b) Let $\mu_A, \mu_B, \mu_C, \mu_D$ be the population mean number of faults incurred by four types of machines A, B, C, D respectively, then the null hypothesis to be tested here is that population mean number of faults for four machines are equal , against the alternative that they are not equal, i.e.,

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D, \text{ against the alternative.}$$

Here, the degrees of freedom for F are 3 and 12, and from Table-7 given in Appendix, we have the tabulated value of F-ratio at 5% level of significance with 3 and 12 degrees of freedom is 3.49, or $F_{3,12;0.05} = 3.49$, while from ANOVA table we find that the observed value of F is 25.21, which is much higher than critical or tabulated value of F, hence, null hypothesis may be rejected. It may be concluded that there is significant differences in the population means flaws, that means mean number of flaws are not the same for all four selected machines.

Example 18.2.5. (Lifetimes of tires) Four different model of tires used by a car rental agency in the process of deciding the brand of tires to purchase as standard equipment for their fleets, the number of kilometers (in '000) run by five tires of each brand are recorded and shown in following table.

Tire type				
A	B	C	D	E
38	48	38	32	47
39	39	42	36	37
37	35	37	39	42
35	37	43	35	48
41	46	35	45	36

Test the hypothesis that the five different model of tires have identical average life.

Solution. Let $\mu_A, \mu_B, \mu_C, \mu_D, \mu_E$ be the population mean lifetimes of five types of tires A, B, C, D, E respectively, then the null hypothesis to be tested here is that population mean lifetime of all the tires are equal, against the alternative that they are not equal, i.e.,

$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$ against the alternative.

H_1: Population mean lifetime of at least two tires are different.

At first we have to set out the ANOVA table, for necessary computations of ANOVA, the following table is constructed.

Table 18.2.5a. Table for necessary computations

	A (x_1)	B (x_2)	C (x_3)	D (x_4)	E (x_5)	(x_1^2)	(x_2^2)	(x_3^2)	(x_4^2)	(x_5^2)
	38	48	38	32	47	1444	2304	1444	1024	2209
	39	39	42	36	37	1521	1521	1764	1296	1369
	37	35	37	39	42	1369	1225	1369	1521	1764
	35	37	43	35	48	1225	1369	1849	1225	2304
	41	46	35	45	36	1681	2116	1225	2025	1296
Total	190	205	195	187	210	7240	8535	7651	7091	8942

Here, total number of observations are 25, and

$$T_1 = 190, \quad T_2 = 205, \quad T_3 = 195, \quad T_4 = 187, \quad T_5 = 210$$

so, $\quad T = \Sigma T = (190 + 205 + 195 + 137 + 210) = 987$

$$CT = \frac{T^2}{N} = \frac{(987)^2}{25} = 38966.76.$$

$$\text{Total sum of squares (SST)} = (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2 + \Sigma x_5^2) - CT$$

$$= (7240 + 8535 + 7651 + 7091 + 8942) - 38966.76$$

$$= 39459 - 38966.76 = 492.24.$$

$$\text{Between sum of squares (SSB)} = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CT \qquad (\text{here, } n_j = 5)$$

$$= \left[\frac{(190)^2}{5} + \frac{(205)^2}{5} + \frac{(195)^2}{5} + \frac{(187)^2}{5} + \frac{(210)^2}{5}\right] - 38966.76$$

$$= (7220 + 8405 + 7605 + 6993.8 + 8820) - 38966.76$$

$$= 39043.80 - 38966.76 = 77.04.$$

Within sum of squares (SSW) = SST – SSB = 492.24- 77.04 = 415.20.

Table 18.2.5b. ANOVA table

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between	SSB = 77.04	$(k-1)=(5-1)=4$	SSB/(k-1) = 19.26	MSB/MSW = 0.9277
Within (Error)	SSW = 415.20	$(n-k)=(25-5)=20$	SSW/(n-k) = 20.76	
Total	SST = 492.24	$(n-1)=(25-1)=24$	--	

The degrees freedom in favour of between variation and within variation are 4 and 20 respectively. From Table-7 in the Appendix, we have the critical value of F at 5% level of significance with 4 and 20 degrees of freedom is 2.87. Thus, since the observed value of F is smaller than the critical value, we fail to reject the null hypothesis of equality of mean lifetime, hence there is no significant difference among the mean lifetime of the five types of tires.

Example 18.2.6. (Waiting time in Banks) A Government Officer just joined in a locality where there are three banks I, II, III. The officer has to open an account in a bank, and he thinks to open account in the bank where the service is better. For this purpose, he made his mind to compare the mean waiting time of the customers required to receive a given service in three banks. He selected customers at random and their waiting time in minutes are recorded as follows :

Bank - I	Bank - II	Bank - III
7.8	9.9	7.3
10.2	12.6	12.6
12.9	11.3	10.7
11.8	12.2	--
14.6	--	--

Test if there is any significant difference in the average waiting time in minutes required to receive the service, so that the officer can select the bank properly.

Solution. Let μ_I, μ_{II}, μ_{III}, be the population mean waiting time for three banks I, II, III respectively, then the null hypothesis to be tested that there is no significant difference in the mean waiting time, against the alternative that there is significant difference, i.e.,

$H_0: \mu_I = \mu_{II} = \mu_{III}$ against the alternative

H_1: Population mean waiting time for at least two banks are different.

At first we have to set out the ANOVA table, for necessary computations of ANOVA, the following table is constructed.

Table 18.2.6a. Table for necessary computations

	I (x_1)	II (x_2)	II (x_3)	(x_1^2)	(x_2^2)	(x_3^2)
	7.8	9.9	7.3	60.84	98.01	53.29
	10.2	12.6	12.6	104.04	158.76	158.76
	12.9	11.3	10.7	166.41	127.69	114.49
	11.8	12.2	--	139.24	148.84	--
	14.6	--	--	213.16	--	--
Total	57.3	46.0	30.6	683.69	533.3	326.54

In this problem the sample size or the replications of all treatments are not the same.

Total number of observations are: (5 + 4 + 3) = 12, and

$$T_1 = 57.3, \quad T_2 = 46.0, \quad T_3 = 30.6$$

so, $$T = \sum T_i = (57.3 + 46.0 + 30.6) = 133.9$$

$$CT = \frac{T^2}{N} = \frac{(133.9)^2}{12} = 1494.10.$$

Total sum of squares (SST) $= \sum\sum x_{ij}^2 = (\sum x_1^2 + \sum x_2^2 + \sum x_3^2) - CT$

$= (683.69 + 533.3 + 326.54) - 1494.10$

$= 1543.54 - 1494.10 = 49.44.$

Between sum of squares (SSB) $= \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CT$ (here, $n_j = 5, 4, 3$)

$$= \left[\frac{(57.3)^2}{5} + \frac{(46.0)^2}{4} + \frac{(30.6)^2}{3}\right] - 1494.10$$

$= (656.66 + 529.00 + 312.12) - 1494.10$

$= 1497.78 - 1494.10 = 3.68.$

Within sum of squares (SSW) = SST – SSB = 49.44 – 3.68 = 45.76.

Table 18.2.6 b. ANOVA table

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between	SSB = 3.68	$(k-1)=(3-1)=2$	SSB/(k–1) = 1.84	MSB/MSW = 0.3622
Within (Error)	SSW = 45.76	$(n-k)=(12-3)=9$	SSW/(n–k) = 5.08	
Total	SST = 49.44	$(n-1)=(12-1)=11$	--	

We see in the table that the degrees freedom in favour of between variation and within variation are 2 and 9 respectively. From Table-7 in the Appendix, we have the critical value of F at 5% level of significance with 3 and 9 degrees of freedom is 4.26 i.e. $F_{2,9;0.05} = 4.26$. Since the observed value of F is smaller than the critical value, the null hypothesis of equality of mean waiting time may be rejected at 5% level of significance, hence there is no significant difference in the average waiting time of three bank, hence the officer may choose any bank convenient to him with respect to other criteria (may be distance from his office, better personal contact, etc).

18.3. Two-Way ANOVA

In case of one-way classification, the observations are assumed to be influenced by different levels of only one assignable factor, but in practice it may happen that the same observations might be affected by a second important assignable factor. In the example given in one-way classification, our objective, was to compare the fuel consumption of three types of automobiles. Data were collected from three independent random samples of three model and analyzed through a one-way ANOVA. In that case, it was assumed that variability in the sample data was due to two causes – a) genuine difference between the performance characteristics of these three types of car and b) a random variation. In fact, it might be suspected that observations might vary not only due to differences in model, but also due to another factor 'drivers' ability'. Hence, a part of observed random variability could be explained by differences in drivers' ability or simply driver too. If the variability of observations due to this last factor could be isolated, the amount of random variability in the experiment would be reduced accordingly, which in turn, makes it easier to detect actual differences in the performance of the cars. In other words, by designing an experiment to account for differences in drivers' ability, we can expect to achieve a more powerful test of the null hypothesis that population mean fuel consumption is the same for all model of cars. Hence, a Two-Way ANOVA is useful when we desire to compare the effect of multiple levels of two factors with multiple observations at each level.

Unlike one way classification, in this case, we consider the experiment where all drivers are to be assigned with a car of each brand randomly. Since every car is run by every driver, it will be possible to extract information about drivers' variability as well as information about differences among different types of cars. The additional variable, drivers' ability, is sometimes called blocking variable and the experiment is needed to be arranged in block (according to drivers). Consideration of variation due to this addition factor would reduce the error variation (Illustrated in Example 19.3.1). If we consider six drivers to test the gasoline consumption, there would be six blocks. If six drivers are assigned to test the gasoline consumption of three models of car, then there would be six blocks and three treatments, and there would be $(3\times6)=18$ observations in total, arranged in six rows and three columns.

Definition. Block. The second factor which is used in an ANOVA in addition to the factor of interest to reduce the variation due to random error.

Definition. Two-way ANOVA. Analysis of variance in which two factors are used to analyze the difference between more than two population means.

Some applications of two-way ANOVA

i) a company may be concerned to see whether there is significant difference in the sales of different salesmen made during different seasons,
ii) a detergent powder company may be interested to test the difference in the performance of different types of detergent powder in different temperature of water,
iii) a production manager may be interested to compare effectiveness of different promotional devices in terms of sales made by different salesmen,
iv) an economist may be involved in comparing the earnings growth predictions for different quarters analyzed by different analysts.

All the three principles of ANOVA design such as replication, randomization and local control are satisfied by a two-way ANOVA. Replications and randomization principles are implemented in the same as in one-way ANOVA. Again, although the existence of a random or error variation in every experiment is unavoidable, it is not desirable to have a large error variance. The methods which are adopted to reduce the error variance are usually known as error control or local control. One of such methods is to make the experimental units homogeneous, and another method is to divide experimental units into different homogenous groups, known as blocks. In case of two-way ANOVA, blocking of experimental units is undertaken as a method of local control to reduce the error variance.

The principle which is followed in the construction of blocks is that 'the experimental units are homogenous within block, and the units are heterogeneous between blocks. Each treatment must be applied in each block only once'.

18.3.1. Assumptions. The following assumptions are to be considered during application of a two-way analysis of variance

i) Each sample is drawn randomly from a normal population and sample statistics tend to reflect the characteristics of the population,

ii) The population from which the samples are derived have identical means and variances,

iii) The blocks are to be made so that the individuals in a particular block are as alike as possible, there may be wide differences between individuals in different blocks,

iv) Each treatment is applied exactly once in each block and number of replications of each treatment is exactly equal to the number of blocks.

18.3.1. Layout of a Two-way ANOVA. Let there are c columns or treatments and each of the treatment is replicated r time in r rows or blocks, the observed values are arranged in c columns and r rows. It is to be noted that in this case, the number of replications of each treatment must be the same, i.e., sample size of all samples is equal. The treatments are replicated randomly in each row or block. Let x_{ij} be the observation corresponding to jth row or block of ith column or treatment, thus, observations obtained for c independent random samples, each of size r based on two-factor classification can be arranged as shown in following table.

Table 18.3. Sample observations for Two-way analysis of variance

Row (Block)	Column (Treatment)				
	C_1	C_2	......	.....	C_c
R_1	x_{11}	x_{12}	......	.....	x_{1c}
R_2	x_{21}	x_{22}	......	.....	x_{2c}
...			......	.....	
R_r	x_{r1}	x_{r2}	......	.....	x_{rc}

For example, for testing the effectiveness of drugs on fever, 3 doses of drugs A, B, C are assigned at random to the 3 categories of patients (according to age group 10-20 years, 20-40 years, 40 years and above), 6 patients of each category are randomly selected as experimental units, then each drug is to be assigned exactly once to each category of patients. In this case, the layout of design is a two-way.

18.3.2. Decomposition of Total Variation and Degrees of Freedom. In a two-way ANOVA that total sum of squares (SST) is partitioned into three components, namely, sum of squares between columns or treatments (SSC), sum of squares between rows or blocks (SSR) and sum of squares within group or error (SSE). That means

$$SST = SSC + SSR + SSE$$

where, Total Sum of squares: $SST = \sum\sum x_{ij}^2 - CT$

Sum of squares between column due to treatment: $SSC = \frac{1}{r}\sum_{j=1}^{c} T_j^2 - CT$

Sum of squares between rows due to blocking: $SSR = \frac{1}{c}\sum_{i=1}^{r} B_i^2 - CT$

and $SSE = SST - SSC - SSR$.

where, $T_j = \sum_{i=1}^{r} x_{ij}$, jth column or treatment total, for example, $T_1 = x_{11} + x_{21} + ... + x_{r1}$,

$B_i = \sum_{j=1}^{c} x_{ij}$, ith row or block total, for example, $B_1 = x_{11} + x_{12} + ... + x_{1c}$,

$T = \sum_{j=1}^{c} T_j = \sum_{i=1}^{r} B_i$, Grand total of all observations, and $CT = \frac{T^2}{N}$

Computations of Degrees of freedom :

In case of two-way ANOVA, the degrees of freedom due to different sources of variations are computed as follows.

i) number of columns or treatments are c, so degrees of freedom for column or treatment variations are $(c-1)$,

ii) number of rows or blocks are r, so degrees of freedom in favour of row or block variation are $(r-1)$,

iii) total number of observations are $N = r \times c = rc$, hence, total degrees of freedom are $(n-1) = (rc-1)$.

iv) Finally, by subtraction, the degrees of freedom associated with error variation become $\{(n-1)-(c-1)-(r-1) = (rc-1)-(c-1)-(r-1) = (rc-c-r+1) = (r-1)\times(c-1)$.

18.3.3. Hypothesis to be Tested. The tests associated with two-way ANOVA proceed in similar fashion to the one-way ANOVA. Only the difference is that in case of two-way ANOVA, tests of significance of difference between column or treatment means and significance of difference between row or block means are undertaken separately. Let α_1, α_2, ... α_c, be the population mean for 1st, 2nd, ..., cth column respectively, and β_1, β_2, ... β_r, be the population mean for 1st, 2nd, ..., rth row respectively, then two-way analysis of variance framework is designed to test the null hypotheses that

i) there is no significant difference between the column or treatment means, symbolically,

$H_{01} : \alpha_1 = \alpha_2 = ... = \alpha_c$,

ii) there is no significant difference between the row or block means, symbolically,

$H_{02} : \beta_1 = \beta_2 = ... = \beta_r$.

Under the first null hypothesis H_{01}, F-ratio is computed using the formula

$$F = \frac{MS(Column)}{MS(Error)} = \frac{SSC/(c-1)}{SSE/(c-1)(r-1)} \sim F_{(c-1), (c-1)(r-1)}.$$

Similarly, under the second null hypothesis H_{02}, F-ratio is computed using the formula

$$F = \frac{MS(Row)}{MS(Error)} = \frac{SSR/(r-1)}{SSE/(c-1)(r-1)} \sim F_{(r-1), (c-1)(r-1)}.$$

18.3.4. Advantages. The advantages of two-way ANOVA are as follows :

i) An important advantage of two-way ANOVA is that it is more efficient than its one-way counterpart. There are two assignable sources of variation, namely, column or treatment and row or block, and this helps to reduce error variation thereby making this design more efficient.

ii) Unlike One-Way ANOVA, it enables us to test the effect of two factors at the same time.

18.3.5. Computational Procedure. For computations of different sum of squares namely, SSC (column or treatment sum of squares), SSR (row or block sum of squares) and SSE (error sum of squares), the following steps can be followed :

- ***Steps for SSC (Column sum of squares) :***

  The steps involved in computation of SSC are as follows :

  i) Calculate the sum of observations for each column or treatment, say, $T_1, T_2, \ldots T_c$, where $T_1 = \sum_{i=1}^{r} x_{i1}$, $T_2 = \sum_{i=1}^{r} x_{i2}$ and so on,

  ii) Calculate sum of all observations to obtain grand total, this can also be done by adding up all the sums obtained in step (i) that means, obtain $T = \sum_{j=1}^{c}\sum_{i=1}^{r} x_{ij} = \sum_{j=1}^{c} T_j = \sum_{i=1}^{r} B_i$ and compute correction factor as $CT = \frac{T^2}{N}$, where $N = r \times c$, total number of observations,

  iii) Calculate the squares of each total obtained in step (i),

  iv) Add up the squares obtained in step (iii) and divide the sum by r (number of rows), that means obtain $\frac{1}{r}\sum_{j=1}^{c} T_j^2$,

  v) Subtract the value obtained in (ii) from that of (iv), which will produce column sum of squares given by $SSC = \frac{1}{r}\sum_{j=1}^{c} T_j^2 - CT$.

- ***Steps for SSR (Row sum of squares) :***

  The steps involved in computation of SSR are as follows.

  i) Calculate the sum of observations for each row or block, say, $B_1, B_2, \ldots, B_r$, where $B_1 = \sum_{j=1}^{c} x_{1j}$ and so on,

  ii) Calculate the squares of each total obtained in step (i)

  iii) Add up the squares obtained in step (ii) and divide the sum by c (number of columns), that means obtain $\frac{1}{c}\sum_{i=1}^{r} B_i^2$,

  iv) Subtract CT obtained in step (ii) during calculation of SSC from the value obtained in above step (iii)), to obtain row sum of squares given by $SSR = \frac{1}{c}\sum_{i=1}^{r} B_i^2 - CT$.

- ***Steps for SST (total sum of squares) :***

For the calculation of total sum of squares, we have to add up the squares of all $r \times c$ observations and subtract the correction factor from total, the steps are as follows :

i) Find the squares of all observations, such as $x_{11}^2, x_{12}^2, \ldots\ldots, x_{rc}^2$,

ii) Add up the squares computed in step (i) to obtain $\sum_{j=1}^{c}\sum_{i=1}^{r} x_{ij}^2$,

iii) Subtract CT from the sum of squares computed in step (ii) to obtain total sum of squares, algebraically, $SST = \sum_{j=1}^{c}\sum_{i=1}^{r} x_{ij}^2 - CT$.

Steps for SSE (Error sum of squares) :

We have already discussed that, in case of two-way ANOVA, the total variation is decomposed into variations due to three sources, such as row, column and random or error, that means, SST=SSC+SSR+SSE, so the error sum of squares is computed by using the relation

$$SSE = SST - SSC - SSR$$

The mean sum of squares of different types of variations are computed dividing sum of squares by corresponding degrees of freedom as shown in following ANOVA table.

18.3.6. ANOVA Table. Unlike one-way ANOVA, there are three sources of variations in two-way ANOVA and we have to compute two F-ratios, one for testing the hypothesis that there is no significant difference among the columns or treatments means, and another for testing the hypothesis that there is no significant difference among rows or block means. The ANOVA table takes form as shown below.

Two-way ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between columns	SSC	$(c-1)$	$MSC = SSC/(c-1)$	$F_1 = MSC/MSE$
Between rows	SSR	$(r-1)$	$MSR = SSR/(r-1)$	$F_2 = MSR/MSE$
Error	SSE	$(r-1)(c-1)$	$MSE = SSE/(c-1)(r-1)$	--
Total	SST	$(rc-1)$	--	--

Decision.

i) Reject null hypothesis of no significant difference between the column means in favour of alternative that there is significant difference between the column means - if calculated F_1-ratio is greater than the tabulated value of F at $100\alpha\%$ level of significance with $(c-1)$ and $(r-1)(c-1)$ degrees of freedom, that means,

reject H_{01} if $F_1 > F_{0.05;(c-1),(c-1)(r-1)}$, otherwise, accept it.

ii) Reject null hypothesis of no significant difference between the row means in favour of alternative that there is significant difference between the row means - if calculated F_2-ratio is greater than the tabulated value of F at $100\alpha\%$ level of significance with $(c-1)$ and $(r-1)(c-1)$ degrees of freedom, that means,

reject H_{02} if $F_2 > F_{0.05;(r-1),(c-1)(r-1)}$, otherwise, accept it.

Example 18.3.1. (Comparison of Fuel consumptions of automobiles using one-way and two-way ANOVA) In Example 19.2.2, we have tested the difference between average fuel consumption of three makes of car and found the significant difference. In this example, how the error variation is reduced due to consideration of a second factor, will be illustrated performing a one-way ANOVA and a two-way ANOVA with almost the same type of information.

Suppose that automobile company is interested to test whether there is any significant differences among the average fuel consumption of three makes of cars A, B and C. For this, 7 classes of drivers are randomly assigned to A-cars, 7 to B-cars and 7 to C-cars. The cars are assigned to the drivers in such a way that each of the car is driven by each of the driver. The following table represents the fuel consumption (km/liter) of three makes of cars A, B and C according to the class of drivers.

Fuel consumption (km/liter) figures for three makes of cars

Class of driver	A-car	B-car	C-car
1	22.2	24.6	22.7
2	19.9	23.1	21.9
3	20.3	22.0	23.2
4	21.4	23.5	24.1
5	21.2	23.6	22.1
6	21.0	22.1	23.4
7	20.3	23.5	22.0

Solution. Under a one-way ANOVA, we have to test the hypothesis

H_0 : Average fuel consumption are the same for all cars.

The following table is constructed for the convenience of computations of sum of squares of different sources of variation.

Table 18.3.1.1a. Computation table

Driver	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2
1	22.2	24.6	22.7	492.84	605.16	515.29
2	19.9	23.1	21.9	396.01	533.61	479.61
3	20.3	22.0	23.2	412.09	484	538.24
4	21.4	23.5	24.1	457.96	552.25	580.81
5	21.2	23.6	22.1	449.44	556.96	488.41
6	21.0	22.1	23.4	441	488.41	547.56
7	20.3	23.5	22.0	412.09	552.25	484.00
Total	$T_1 = 1463$	$T_2 = 1624$	$T_3 = 149.4$	$\Sigma x_1^2 = 306143$	$\Sigma x_2^2 = 377264$	$\Sigma x_3^2 = 363392$

Here, $T = \Sigma T_j = (146.3 + 162.4 + 137.4) = 468.1$

$$CT = \frac{T^2}{N} = \frac{(4681)^2}{21} = 10434.17.$$

Total sum of squares (SST) = $(\Sigma x_1{}^2 + \Sigma x_2{}^2 + \Sigma x_3{}^2) - CT$

$$= (3061.64 + 3772.64 + 3633.92) - 10434.17 = 33.82.$$

Between sum of squares (SSB) = $\sum_{j=1}^{k} \frac{T_j^2}{n_j} - CT$

$$= \left[\frac{(146.3)^2}{7} + \frac{(162.4)^2}{7} + \frac{(149.4)^2}{7}\right] - 10434.17$$

$$= 20.94.$$

Within sum of squares SSW = SST – SSB = 33.82 – 20.94 = 12.88.

Table 18.3.1.1b. ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between cars	SSB = 20.94	$(k-1)=2$	SSB/(k-1) = 10.47	MSB/MSW = 14.63
(Error	SSW = 12.88	$(N-k)=18$	SSW/(N-k) = 0.7156	
Total	SST = 33.82	$(N-1)=20$		

Here the degrees of freedom for F are 2 and 18, and from Table-7 in Appendix, we have the tabulated value of F-ratio at 5% level of significance with 2 and 17 degrees of freedom is 3.55, or $F_{2,18;0.05} = 3.55$, hence like Example 3.2.2, these new data set allow us to reject null hypothesis at 5% level of significance. That means, the mean fuel consumptions for all three types of cars are not the same.

Again, in order to illustrate how the additional factor 'drivers' class' reduces the error variation, let us conduct a two-way ANOVA, in this case we have to formulate two hypotheses as follows.

H_{01}: Average fuel consumption of three types of cars are the same.

H_{02}: Average fuel consumed by all class of drivers are the same.

The following table is constructed for a two-way analysis of variance.

Table 18.3.1.2a. Table for computations

Class of Driver	Car type A(x_1)	B(x_2)	C(x_3)	Row total (B_i)	x_1^2	x_2^2	x_3^2	Row total	B_i^2
1	22.2	24.6	22.7	69.5	492.8	605.2	515.3	1613.3	4830.3
2	19.9	23.1	21.9	64.9	396.0	533.6	479.6	1409.2	4212.0
3	20.3	22	23.2	65.5	412.1	484.0	538.2	1434.3	4290.3
4	21.4	23.5	24.1	69.0	458.0	552.3	580.8	1591.0	4761.0
5	21.2	23.6	22.1	66.9	449.4	557.0	488.4	1494.8	4475.6
6	21	22.1	23.4	66.5	441.0	488.4	547.6	1477.0	4422.3
7	20.3	23.5	22	65.8	412.1	552.3	484.0	1448.3	4329.6
Column Total	T_1 = 146.3	T_2 = 162.4	T_3 = 149.4	T = 468.1	3061.4	3772.6	3633.9	10467.9	31321.1
T_j^2	21403.	26373.7	25408.36	73185.81					

There are $c=3$ columns and $r=7$ rows, so the total number of observations 21,

From the above table, we have,

Sum of all observations (in four samples of salesmen as well as in three samples of months)

$$T=\sum_{i=1}^{3}\sum_{j=1}^{3}x_{ij}=\sum_{j=1}^{3}T_j=\sum_{i=1}^{7}B_i=468.1 \quad \text{and} \quad \sum_{i=1}^{7}\sum_{j=1}^{3}x_{ij}^2=10467.9.$$

$$\text{So,}\quad CT=\frac{T^2}{N}=\frac{4681^2}{21}=10434.17.$$

$$\text{Thus, SST = Total sum of squares} = \sum_{i=1}^{7}\sum_{j=1}^{3}x_{ij}^2-CT=10468.0-10434.2=33.8.$$

SSC = Sum of squares between Texts (column)

$$=\frac{\sum_{j=1}^{3}T_j^2}{r}-CT=73185.81/7-10434.17=20.94\,.$$

SSR = Sum of squares between exam types (rows)

$$=\frac{\sum_{i=1}^{3}B_i^2}{c}-CT=31321.1/3-10434.2=6.2\,.$$

Thus, SSE = within or error sum of squares = SST – (SSC – SSR) = 33.8 – (20.9 + 6.2) = 6.7.

Total degrees of freedom = $N-1=rc-1=21-1=20$,

Column degrees of freedom = $c-1=3-1=2$,

Row degrees of freedom = $r-1=7-1=6$,

And finally, error degrees of freedom = $20-(2+6)=(r-1)(c-1)=12$.

Thus, the required ANOVA table is as presented below.

Table 18.3.1b. Two-way ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between cars (Column)	20.9	$c-1=2$	$MSC = SSC/(c-1) = 20.9/2 = 10.1$	$F_1 = 18.76$
Between drivers (Row)	6.2	$r-1=6$	$MSR = SSR/(r-1) = 6.2/6 = 1.03$	$F_2 = 1.84$
Error	6.7	12	$MSE = SSE/(c-1)(r-1) = 6.7/12 = 0.56$	--
Total	33.8	$N-1=20$	--	--

F_1 is computed using formula: $F_1 = \dfrac{MSC}{MSE} = 10.1/0.56 = 18.76$, and

F_2 is computed using formula: $F_2 = \dfrac{MSR}{MSE} = 1.03/0.56 = 1.84$.

i) From table of percentage points of F-distribution, we have, the critical value of F at 5% with 2 and 12 degrees of freedom is 3.89 while the observed value of F for between cars is 18.76 which is much higher the critical value, so hypothesis is highly significant, that mean we can draw the same inference as in one-way variation, that there is significant difference between the mean fuel consumption of different types of cars.

ii) For the hypothesis related to the driver's class, we have the critical value of F at 5% level of significance with 6 and 12 degrees of freedom 3.00 which is smaller than the observed value of $F=1.84$, hence it may be concluded that there is no significant different between the average fuel consumptions are the same for all class of drivers.

Note. In the first part of this example, we have undertaken a one-way ANOVA and sum of squares due to error (within) is found as 12.88 with 18 degrees of freedom, while in second part, we have considered a second factor 'drivers class' and conducted a two-way ANOVA, we see that due to this additional factor, sum of squares due to error reduces from 12.88 to 6.7 along with its degrees of freedom from 18 to 12. There is no impact of this additional factor on the conclusion regarding equality of fuel consumption by different types of cars. Rather, by conducting a two-way ANOVA, we could make inference about the difference of fuel consumption of due to another factor 'class of driver'.

Example 19.3.2. (Performance of Workers with different types of machines) The following data represent the production per day turned out by 5 different workers using 4 different types of machines.

Workers	Machine type			
	A	B	C	D
1	44	38	47	36
2	46	40	52	43
3	34	36	44	32
4	43	38	46	33
5	38	42	49	39

i) Test whether the mean productivity is the same for different machine types,

ii) Test whether 5 workers differ with respect to mean productivity.

Solution. We have to test the following hypotheses

i) H_{01} : The population mean productivity is the same for different machine types, against the alternative that

H_{11} : The population mean productivity is not the same for different machine types

ii) H_{02} : The workers do not differ with respect to mean productivity, against the alternative that

H_{12} : The workers differ with respect to mean productivity.

For testing the above mentioned two hypotheses, at first, we have to undertake an ANOVA. Necessary computations for ANOVA are done in the following table.

Table 18.3.2a. Table for computations

Workers	Machine types A(x_1)	B(x_2)	C(x_3)	D(x_4)	Row total	x_1^2	x_2^2	x_3^2	x_4^2	Row total	B_i^2
1	44	38	47	36	$B_1 = 165$	1936	1444	2209	1296	6885	27225
2	46	40	52	43	$B_2 = 181$	2116	1600	2704	1849	8269	32761
3	34	36	44	32	$B_3 = 146$	1156	1296	1936	1024	5412	21316
4	43	38	46	33	$B_4 = 160$	1849	1444	2116	1089	6498	25600
5	38	42	49	39	$B_5 = 168$	1444	1764	2401	1521	7130	28224
Column Total	$T_1 =$ 205	$T_2 =$ 194	$T_3 =$ 238	$T_4 =$ 183	$T=820$	--	--	--	--	34194	135126
T_j^2	42025	37636	56644	33489	169794						

Here, number of columns $c=4$ and number of rows $r=5$, so the total number of observations $N=5\times4=20$.

From the above table, we have,

Sum of all observations (in four samples of salesmen as well as in three samples of months)

$$T = \sum_{j=1}^{4}\sum_{i=1}^{5} x_{ij} = \sum_{j=1}^{4} T_j = \sum_{i=1}^{5} B_i = 820 \quad \text{and} \quad \sum_{i=1}^{5}\sum_{j=1}^{4} x_{ij}^2 = 34194.$$

So, $$CT = \frac{T^2}{N} = \frac{820^2}{20} = 33620.$$

Thus, SST = Total sum of squares = $\sum_{i=1}^{5}\sum_{j=1}^{4} x_{ij}^2 - CT = 34194 - 33620 = 574.$

SSC = Sum of squares between machine types (column)

$$= \frac{\sum_{j=1}^{4} T_j^2}{r} - CT = 169794/5 - 33620 = 338.8$$

SSR = Sum of squares between workers (rows)

$$= \frac{\sum_{i=1}^{5} B_i^2}{c} - CT = 135126/4 - 33620 = 161.5\,.$$

Thus, SSE = within or error sum of squares = SST-- (SSC + SSR) = 574 – (338.8 + 161.5) = 73.7.

Total degrees of freedom = $N-1 = rc-1 = 20-1 = 19$

Column degrees of freedom = $c-1=4-1=3$, Row degrees of freedom = $r-1=5-1=4$.

And finally, error degrees of freedom $=19-(3+4)=(r-1)(c-1)=12$.

Thus, the following ANOVA table is constructed.

Table 18.3.2b. Two-way ANOVA table

Source of variation	Sum of Squares SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between machines	338.8	$c-1=3$	$MSC = SSC/(c-1) = 338.8/3 = 112.93$	$F_1 = 18.39$
Between workers	161.5	$r-1=4$	$MSR = SSR/(r-1) = 161.5/4 = 40.375$	$F_2 = 6.574$
Error	73.7	$(c-1)(r-1)=12$	$MSE = SSE/(c-1)(r-1) = 73.7/12 = 6.14$	--
Total	574.0	$N-1=19$	--	--

F_1 is computed using formula $F_1 = \dfrac{MSC}{MSE} = 112.93/6.14 = 18.39$, and

F_2 is computed using formula $F_2 = \dfrac{MSR}{MSE} = 40.375/6.14 = 6.574$.

Decision

i) In case of H_{01}, the tabulated F value at 5% level of significance with 3 and 12 degrees of freedom is 3.49, while the computed F value is $F_1 = 18.39$, which is greater than the tabulated value, so H_{01} may be rejected at 5% level of significance. It may be concluded that mean productivity is not the same for different machines.

ii) In case of H_{02}, the tabulated F value at 5% level of significance with 4 and 12 degrees of freedom is 3.26, while the computed F value is $F_2 = 6.574$, which is also greater than the tabulated value, so H_{02} may also be rejected at 5% level of significance. It may be concluded that different workers differ with respect to mean productivity.

Example 18.3.3. (Performance of salesmen in different months) The following table gives the number of refrigerators sold by 4 salesmen in three months. *(AIS- 2010, 2011, CU)*

Months	Salesmen			
	A	B	C	D
May	50	40	48	39
June	46	48	50	45
July	39	44	40	39

a) Is there a significant difference in the sales made salesmen?

b) Is there a significant difference in the sales made during different months? (Consider 5% level of significance).

Solution. We are interested to test the following two hypotheses.

a) H_{01} : There is no significant difference between the average sales made in four salesmen,

b) H_{02} : There is no significant difference between the average sales made in different months.

For testing the above-mentioned hypotheses, we are to conduct an ANOVA. For necessary computations of this ANOVA, let us construct the following table:

Table 18.3.3a. Table for necessary computations

Month	Showroom				Row Total	x_1^2	x_2^2	x_3^2	x_4^2	Row Total	B_j^2
	$A(x_1)$	$B(x_2)$	$C(x_3)$	$D(x_4)$							
May	50	40	48	39	$B_1 = 177$	2500	1600	2304	1521	7925	31329
June	46	48	50	45	$B_2 = 189$	2116	2304	2500	2025	8945	35721
July	39	44	40	39	$B_3 = 162$	1521	1936	1600	1521	6578	26244
Column Total	$T_1 =$ 135	$T_2 =$ 132	$T_3 =$ 138	$T_4 =$ 123	$T = 528$	6137	5840	6404	5067	23448	93294
T_i^2	18225	17424	19044	15129	69822						

There are $c=4$ columns and $r=3$ rows, so the total number of observations $n=12$.

From the above table, we have,

Sum of all observations (in four samples of salesmen as well as in three samples of months)

$$T = \sum_{i=1}^{3}\sum_{j=1}^{4} x_{ij} = \sum_{j=1}^{4} T_j = \sum_{i=1}^{3} B_i = 528 \quad \text{and} \quad \sum_{i=1}^{3}\sum_{j=1}^{4} x_{ij}^2 = 23448$$

$$\text{So, } CT = \frac{528^2}{12} = 23232.$$

$$\text{Thus, SST = Total sum of squares} = \sum_{i=1}^{3}\sum_{j=1}^{4} x_{ij}^2 - CT = 23448 - 23232 = 216.$$

$$\text{SSC = Sum of squares between salesmen (columns)} = \frac{\sum_{j=1}^{4} T_j^2}{r} - CT = 69822/3 - 23232 = 42.$$

$$\text{SSR = Sum of squares between months (rows)} = \frac{\sum_{i=1}^{3} B_i^2}{c} - CT = 93294/4 - 23232 = 91.5.$$

Thus, SSE = within or error sum of squares = SST – SSC – SSR = 216 – 42 – 91.5 = 82.5.

Total degrees of freedom = $N-1=rc-1=12-1=11$.

Column degrees of freedom = $c-1=4-1=3$, Row degrees of freedom = $r-1=3-1=2$.

And finally, error degrees of freedom = 11 – (3 + 2) = $(r-1)(c-1)=6$.

Computations of mean sum of squares for different sources along with F-ratio are shown in e following ANOVA table.

Table 18.3.3b. Two-way ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between salesmen	42.0	$c-1=3$	MSC = SSC/(c–1) = 42/3 = 14	1.018
Between months	91.5	$r-1=2$	MSR = SSR/(r–1) = 91.5/2 = 45.75	3.327
Error	82.5	11 – (3 + 2) = 6	MSE = SSE/(c–1)(r–1) = 82.5/6 = 13.75	--
Total	216	$n-1=11$	--	--

F_1 is computed using formula $F_1 = \dfrac{MSC}{MSE} = 14/13.75 = 1.018$ and

F_2 is computed using formula $F_2 = \dfrac{MSR}{MSE} = 45.75/13.75 = 3.327$.

ecision

a) The tabulated value of F at 5% level of significance with 3 and 6 degrees of freedom is 4.75, since the observed value $F_1 = 1.018$ does not exceed the tabulated value, we fail to reject null hypothesis. Hence, it may be concluded that there is no significant difference in the sales made by four salesmen.

b) Again, the tabulated value of F at 5% level of significance with 2 and 6 degrees of freedom is 5.14, since the observed value $F_2 = 3.327$ for second hypothesis does not exceed the tabulated value, the second null hypothesis may not also be rejected. That means, there is no significant difference in the sales made during different months.

xample 18.3.4. (Sales in different cities) The following table gives the monthly sales (in 1000 k) of a certain firm in three cities by its four salesmen. *(AIS – 2012, CU)*

Cities	Salesmen				
	A	B	C	D	Total
X	5	4	4	7	20
Y	7	8	5	4	24
Z	9	6	6	7	28
Total	21	18	15	18	72

State whether the difference between sales made by the four salesmen and difference between sales made in three cities are significant.

Solution. We are interested to test the following two hypotheses.

i) H_{01} : There is no significant difference between the average sales made by four salesmen,

ii) H_{02} : There is no significant difference between the average sales made in different months,

For testing the above-mentioned hypotheses, we are to conduct an ANOVA. For necessary computations of this ANOVA, let us construct the following table.

Table 18.3.4a. Table for necessary computations

Cities	Salesmen				Row Total	x_1^2	x_2^2	x_3^2	x_4^2	Row Total	B_I^2
	$A(x_1)$	$B(x_2)$	$C(x_3)$	$D(x_4)$							
X	5	4	4	7	$B_1 = 20$	25	16	16	49	106	400
Y	7	8	5	4	$B_2 = 24$	49	64	25	16	154	576
Z	9	6	6	7	$B_3 = 28$	81	36	36	49	202	784
Column Total	$T_1 =$ 21	$T_2 =$ 18	$T_3 =$ 15	$T_4 =$ 18	$T = 72$	155	116	77	114	462	1760
T_J^2	441	324	225	324	1314						

Here, number of columns $c=4$, number of rows $r=3$,

So the total number of observations $n=12$.

From the above table, we have,

Sum of all observations (in four samples of salesmen as well as in three samples of months)

$$T = \sum_{i=1}^{3}\sum_{j=1}^{4} x_{ij} = \sum_{j=1}^{4} T_j = \sum_{i=1}^{3} B_i = 72 \quad \text{and} \quad \sum_{i=1}^{3}\sum_{j=1}^{4} x_{ij}^2 = 462.$$

So, $$CT = \frac{72^2}{12} = 432.$$

Thus, $$SST = \text{Total sum of squares} = \sum_{i=1}^{3}\sum_{j=1}^{4} x_{ij}^2 - CT = 462 - 432 = 30.$$

$$SSC = \text{Sum of squares between salesmen (columns)} = \frac{\sum_{j=1}^{4} T_j^2}{r} - CT = 1314/3 - 432 = 6,$$

$$SSR = \text{Sum of squares between cities (rows)} = \frac{\sum_{i=1}^{3} B_i^2}{c} - CT = 760/4 - 432 = 5.$$

Thus, SSE = within or error sum of squares = SST – SSC – SSR = 30 – 6 – 5 = 19.

Total degrees of freedom = $N-1 = rc-1 = 12-1 = 11$.

Column degrees of freedom = $c-1=4-1=3$, Row degrees of freedom = $r-1=3-1=2$.

And finally, error degrees of freedom = $11-(3+2)=(r-1)(c-1)=6$.

Computations of mean sum of squares for different sources along with F-ratio are shown in the following ANOVA table.

Table 18.3.4b. Two-way ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between salesmen	6	$c-1=3$	MSC = SSC/(c–1) = 6/3 = 2	$F_1 = 2.0/3.17 = 0.53$
Between cities	5	$r-1=2$	MSR = SSR/(r–1) = 5/2 = 2.5	$F_2 = 2.5/3.17 = 0.66$
Error	19	$11-(3+2)=6$	MSE = SSE/(c–1)(r–1) = 19/6 = 3.17	--
Total	30	$N-1=11$	--	--

Decision

i) From Table-7 in Appendix, we have the tabulated value of F at 5% level of significance with 3 and 6 degrees of freedom is 4.75, i.e., $F_{0.05;3,6} = 4.75$, since $F_1 < F_{0.05;3,6}$, we fail to reject null hypothesis. Hence, it may be concluded that there is no significant difference in the sales made by four salesmen.

ii) Again, we have the tabulated value of F at 5% level of significance with 2 and 6 degrees of freedom is 4.75, i.e., $F_{0.05;2,6} = 5.14$, since the observed value $F_2 < F_{0.05;2,6}$, the second null hypothesis may not be rejected too. That means, there is no significant difference in the sales made during different months.

Example 18.3.5. In an industry, production can be accomplished by four different workers on five different machines. A two-way sample design is being made with two fold objectives of examining whether the four workers differ with respect to mean productivity and whether the mean productivity is the same for the five different machines. The researcher involved in this study reports the following results :

i) Sum of squares for variation between machines = 35.2,

ii) Sum of squares for variation between workmen = 53.8,

iii) Sum of square for total variation = 174.2.

Set up ANOVA table for the given information and draw the inference about variance at 5% level of significance.

Solution. According to the objective of the study, we have to test the following two hypotheses.

i) H_{01} : The workers do not differ with respect to mean productivity, against the alternative that

H_{11} : The workers differ with respect to mean productivity.

ii) H_{02} : Mean productivity is the same for five different machines, against the alternative that

H_{12} : Mean productivity is not the same for five different machines.

Let the analysis of data has been undertaken with workmen in the column and machines in the row, then given the information that,

SSC (for workmen) = 53.8, SSR (for machines) = 35.2 and SST = 174.2.

So the sum of squares due to random variation is obtained as :

$$SSE = SST - SSC - SSR = 174.2 - 53.8 - 35.2 = 85.20.$$

Again, since the design has been made with four workmen on five different machines, so, degrees of freedom for column variation $= c-1 = 4-1 = 3$, for row variation $= r-1 = 5-1 = 4$, for total variation $= 4\times5-1 = 19$, and finally, degrees of freedom for random variation $= (19-4-3) = 12$.Thus, the required ANOVA table is as set up below.

Table 18.3.5a. Two-way ANOVA table

Source of variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	Calculated F-ratio
Between workmen	53.8	$c-1=3$	MSC = SSC/(c−1) = 53.8/3 = 17.93	$F_1 = 2.53$
Between machines	35.2	$r-1=4$	MSR = SSR/(r−1) = 35.2/4 = 8.80	$F_2 = 1.24$
Error	85.2	$(c-1)(r-1)=12$	MSE = SSE/(c−1)(r−1) = 85.2/12 = 7.10	--
Total	174.2	$n-1=19$	--	--

F_1 is computed using formula : $F_1 = \dfrac{MSC}{MSE} = 17.93/7.10 = 2.53$, and

F_2 is computed using formula : $F_2 = \dfrac{MSR}{MSE} = 8.80/7.10 = 1.24$.

Decision

i) In case of H_{01}, the tabulated F value at 5% level of significance with 3 and 12 degrees of freedom is 3.49, while the computed F value is $F_1 = 2.53$, which is smaller than the tabulated value, so H_{01} may not be rejected at 5% level of significance. It may be concluded that the workmen do not differ with respect to mean productivity.

ii) In case of H_{02}, the tabulated F value at 5% level of significance with 4 and 12 degrees of freedom is 3.26, while the computed F value is $F_2 = 1.24$, which is also smaller than the tabulated value, so H_{02} may not also be rejected at 5% level of significance. Hence, like the case of workmen, it may be concluded that mean productivity is the same for five machines.

Group-A : Short questions and answers

1. **What is analysis of variance?**

   Ans. A statistical technique used to test whether the means of three or more populations are all equal.

2. **What is one analysis of variance?**

   Ans. The analysis of variance that analysis one variable only'

3. **What is between sum of squares (SSB)?**

   Ans. The sum of squares between samples is called SSB.

4. **What is within sum of squares (SSW)?**

   Ans. The sum of squares within the samples is called SSW.

5. **What is total sum of squares (SST)?**

   Ans. The total sum of squares is the sum of SSB and SSW.

Group-B : Broad Questions and Problems

- What do you mean by ANOVA? Why is it so called?
- Mention some examples suitable for applying one way ANOVA.
- State the assumptions on which ANOVA is performed? Also state the principles of analysis of variance.
- Define factor, treatment and experimental unit with a suitable example in context to the analysis of variance.

10. What do you mean by one way analysis of variance? State the conditions under which a one-way analysis of variance is preferred.
11. Write down the hypothesis considered in a one-way ANOVA. Also discuss the testing procedure of this hypothesis.
12. State the advantages of one-way ANOVA.
13. How can you undertake a one-way analysis of variance?
14. Define block stating its necessity in analysis of variance. Also state the principles of blocking.
15. Define a two-way analysis of variance along with assumption involved.
16. Mention some examples where two- way ANOVA is to be performed.
17. How does a one-way analysis of variance differ from a two-way analysis of variance?
18. How can you undertake a two-way analysis of variance?

Exercises

19. A ... ds to test whether there is any significant difference between the sales ... C. For this purpose, weekly sales of the salesman A are

taken for 5 weeks, salesman B for 4 weeks and salesman C for 5 weeks. The following results are obtained for a one-way analysis of variance :

Total sum of squares = 40, sum of squares between samples (salesmen) = 10.

Write down the hypothesis to be tested and construct an ANOVA table.

(Ans. F = 1.83)

20. In order to test whether breaking strength (in pounds) of three types of ropes I, II and III differ significantly or not, a random sample of 18 ropes was selected of which 5 ropes belong to type I, 7 ropes belong to type II and the rest belong to type III. The summary of results obtained for a one-way analysis of variance is as follows :

Sum of squares between types of ropes = 5838.44, sum of squares within types of ropes = 1126.

Construct an ANOVA table and comment.

(Ans. F = 38.89)

21. The following results are obtained for a two-way analysis of variance for four salesmen with their sales in three seasons:

Sum of squares between salesmen = 42, sum of squares between seasons = 32, total sum of squares = 210.

Set up ANOVA table and comment on the hypotheses that (i) there is no significant difference between the sales of different salesmen and (ii) there is no significant difference between the sales in different seasons.

(Ans. F_1 = 0.308, F_2 = 0.235)

22. In order to test whether there is any significant difference between the performance of 4 machines which produce the same type of products, five operators are randomly selected and each of the machines is run by each of the operators. The following result is obtained for the number of units produced per day by each machine :

Sum of squares between machines = 247.75, sum of squares between operators = 161.39, sum of squares due to error = 124, set up an ANOVA and comment.

(Ans. F_1 = 5.99, F_2 = 2.92)

23. A medicine company appointed four representatives for four different regions. Each of representatives is asked to sell the product in each area within a given quarter. The number of items sold by each of the representatives in that particular quarter are recorded and the following results are obtained :

Sum of squares between representatives = 32.8, sum of squares between regions = 21.6, total sum of squares due to error = 82.5, set up an ANOVA and comment.

(Ans. F_1 = 3.50, F_2 = 2.31

Applications

24. Four salesmen were posted in different areas by a company. The number of units of the product Y sold by them in four randomly selected weeks are as follows :

Salesman			
A	B	C	D
30	35	33	25
33	42	38	31
38	40	45	29
39	31	28	35

Based on the information given above, can it be concluded at 5% level of significance that there is no significant difference in the performance of these four salesmen.

(Ans. F = 1.37)

The fuel consumption km/liter of three brands of vehicles for a number of replications are provided below :

Brand I	Brand II	Brand III
18	13	13
20	12	14
19	14	15
16	13	14

i) State the null hypothesis and the alternative hypothesis,
ii) Set up ANOVA table,
State your decision regarding null hypothesis.
(Ans. F = 21.94)

The following data refer to the share price (in hundred) of three products for certain randomly selected days of a month.

Product I	Product II	Product III
19	23	20
17	30	19
21	24	25
19	23	24
22	--	25
20	--	--

i) State the hypothesis to be tested,
ii) Perform analysis of variance and set up an ANOVA table,
iii) Comment on the acceptance or rejection of hypothesis. (Ans. F = 5.12, given $F_{0.05;2,12} = 3.89$)

A soft-drink manufacturing company want to compare the effects of types of their products (Coke, Sprite, Fanta) on sales. Four major regions are selected for the test and ten retail stores are randomly selected from each region. The following table shows the sales (in hundred 1-litre bottles) at the end of one-week experimental period.

Region	Drink type		
	Coke	Sprite	Fanta
Dhaka	60	52	47
Chittagong	52	54	56
Rajshahi	44	63	49
Khulna	48	44	41

Set out the appropriate analysis of variance for testing the hypothesis that population mean sales are the same for each type of drink.

8. The manager of a computer software company wishes to study the number of hours senior executives spend per week with computer by type of industry. The manager selected a sample of five executives from each of three industries. The number of hours spent five working days in a particular week are collected and shown in following table.

Security house	Banking	Insurance
32	28	30
30	28	28
30	26	26
32	28	28
30	30	30

Can the manager conclude that there is a difference in the mean number of hours spent per week by the industry at 5% level of significance?

(Ans. F = 5.7325)

29. A big company wishes to test whether its three salesmen A, B and C tend to make sales of the same size or whether they differ in their selling ability as measured by the average size of their sales. During the last week there have been 14 sales calls – A made 5 calls, B made 4 calls and C made 5 calls. Following are the weekly sales record of three salesmen.

A	B	C
300	600	700
400	300	300
300	300	400
500	400	600
100	--	500

i) Set out the one-way analysis of variance table,

ii) Test the null hypothesis that mean selling ability of three salesmen do not differ significantly at 5% level of significance.

(Ans. F = 1.80)

. It is desired to test whether there is any difference in the performance of 3 brands of mobile sets. Samples of 5 sets of each brand are randomly selected and their frequency of repair during first two years of purchase are recorded. The observations are given below.

Brand of Mobile set		
A	B	C
4	7	4
6	4	6
7	3	6
5	6	3
8	5	1

Test whether there is any significant difference between performance of three brands of mobile sets?

(Ans. F = 1.58)

A popular department store is planning to establish a new store at any of the three locations of a city. The important factor to be considered in this case is the household income of people living in these areas. If the average income per household is similar then the authority can select any one of these three locations. Random samples of various number of households from each location are taken and their monthly income are recorded as follows:

Monthly household income (Tk. '000)		
Area I	Area II	Area III
70	100	60
72	110	65
75	108	57
80	112	84
83	113	84
--	120	70
--	100	--

Test if the average income per household in all these three locations can be considered as the same at 5% level of significance.

(Ans. F = 38.96)

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

	Salesman		
Month	A	B	C
January	8	6	14
February	9	8	12
March	11	10	18
April	12	4	8

Do you think at 5% level of significance that there is no significant difference in the average number of appointments made per month by three salesmen?

(Ans. F = 3.95)

3. A manufacturing company has purchased three new machines of different brands and wishes to determine whether one of these is faster than the others in producing a certain product. Five hourly production figures are observed at random from each machine and the results are given below :

Machines		
M_1	M_2	M_3
36	39	25
38	38	28
31	42	30
25	35	28
30	31	24

Use analysis of variance to take decision at 5% level of significance whether the machines are significantly different in their mean speed.

(Ans. F = 7.50)

4. A student with a bachelor's degree in business administration applied to several universities for admission into MBA program in Marketing. He got admission in three private universities. Before deciding on which university to admit, he wants to find out if there is any difference in the average starting salary of MBA graduates from these three universities receive when they get job after graduation. For this purpose, sample of 21 graduates 7 from each university has been selected and their starting salaries (in Taka thousand) have been recorded as follows :

University A	University B	University C
45	41	35
40	42	37
39	43	35
41	45	37
37	40	34
39	45	37
42	43	30

i) State the null and alternative hypothesis,
ii) Compute the values of SSB, SSW and SST,
iii) Complete an ANOVA,
iv) Comment whether there is any significant difference in the average salaries of MBA graduates from the selected three universities.

35. Three different lap top manufacturing companies claim that their battery packs will last for three hours before recharging. A consumer agency selected five computers at random from each of the three companies and tested their duration before recharging. The time (in hours) that these batteries worked before recharging is given in the following table.

Company		
A	B	C
2.86	3.15	3.10
3.00	2.95	2.64
2.55	3.40	2.89
2.90	3.20	2.90
2.60	3.75	3.27

Is there evidence to suggest that there is no difference in the average life batteries in the laptop computers manufactured by three companies? Use 5% level of significance.
(Ans. F = 5.33)

36. A company sells three brands of shampoos, viz. dry, normal and silky. Sales in million of Taka for the past five months are given in the following table.

	Sales in million Taka		
Month	Dry	Normal	Silky
June	7	9	12
July	9	12	14
August	13	11	8
September	8	9	7
October	9	10	13

Using 5% level of significance, test whether the mean sales differ (i) for three types of shampoos, (ii) in different months.
(Ans. $F_1 = 0.39$, $F_2 = 1.71$)

37. There are three main brands of certain liquid milk. A sample of 120 packets sold is examined and found to be allocated among four groups of consumers A, B, C, D. The group and brand wise sales of milk are reported below.

Brand	Group			
	A	B	C	D
I	0	4	8	15
II	5	8	13	6
III	18	19	11	13

i) Is there any significant difference in brand preference by different groups consumers?
ii) Is there any significant difference in average sales of different brands of milk?
(Ans. $F_1 = 0.53$, $F_2 = 0.66$)

The following table represent the number of units of production per day turned out by 4 different workers using 5 different types of machine.

Worker	Machine type					
	A	B	C	D	E	Total
1	14	15	13	17	16	75
2	16	18	16	15	14	79
3	17	16	17	18	18	86
4	13	15	14	18	12	72
Total	60	64	60	68	60	312

On the basis of given information, can it be concluded that (i) the mean productivity of different machines is the same, (ii) the workers do not differ with regard to productivity.
(Ans. $F_1 = 1.28$, $F_2 = 2.93$)

. The following table represents the amount (in lakh Taka) sold by 5 sales agents of a big company in 3 different areas.

Agent	Area		
	A	B	C
I	10	6	6
II	10	10	8
III	10	8	8
IV	12	8	10
V	12	8	10

i) State the hypotheses that can be formulated from the given information
ii) Set out ANOVA table and comment on the mean sales in different area and by different agents.
(Ans. $F_1 = 10.04$, $F_2 = 3.27$)

). A soft-drink manufacturing company want to compare the effects of types of their products (Coke, Sprite, Fanta) on sales. Four major regions are selected for the test and ten retail stores are randomly selected from each region. The following table shows the sales (in hundred 1-litre bottles) at the end of one-week experimental period.

Region	Drink type		
	Coke	Sprite	Fanta
Dhaka	63	49	44
Chittagong	44	41	48
Rajshahi	52	47	60
Khulna	54	56	52

Set out the appropriate analysis of variance to test the following hypotheses.

i) Population mean sales are the same for each type of drink,

ii) Population mean sales are the same in each region.

41. A big manufacturing company operates 24 hours a day, five days a week. The workers rotate shifts each week. Management in interested to check whether there is a difference in the number of units produced when the workers work on various shifts. A sample of five workers is selected and the output recorded on each shift are as follows.

Worker	Shift		
	Morning	Afternoon	Night
A	38	31	38
B	33	31	32
C	36	30	40
D	33	29	35
E	35	34	33

Can it be concluded at 5% level of significance that there is a difference in the mean production rate by shift or by worker?

(Ans. $F_1 = 5.75$, $F_2 = 1.55$)

42. Four different types of drugs have been developed for the cure of certain disease. These drugs are applied on patients in three different hospitals. The information given below show the number of cases of recovery from the disease per 100 patients who have given the drugs.

Hospital	Drug type			
	A	B	C	D
H_1	26	25	37	23
H_2	18	31	20	28
H_3	22	22	31	20

Carry out analysis of variance to test if (i) there is any significant difference in the effectiveness of drugs with respect to the number of recovery, (ii) there is any significant difference in the number of recovery in different hospitals.

(Ans. $F_1 = 0.84$, $F_2 = 0.52$)

43. A manufacturer has just introduced a new product that will be sold in sizes: small, medium and large. Five salesmen are randomly selected from the sales force and given each of these products to sell. The sales figure for a randomly selected month are used to test whether there is a difference in sales volume of different sizes or there is a difference in sales volume sold by different salesmen. The amount sold (in hundred Taka) by the salesmen are given below :

Salesman	Size of product		
	Small	Medium	Large
A	88	90	85
B	76	88	72
C	93	97	88
D	67	89	90
E	88	96	75

i) Formulate the necessary hypotheses
ii) Set out ANOVA table and comment on the hypothesis.
(Ans. $F_1 = 2.95$, $F_2 = 1.60$)

4. Mr. X wants to build a CNG service station on one of he three locations. He measures the traffic passing through each location for six days. The following are the average number of traffics per hour passing through each location for each of six days.

Day	Location		
	L_1	L_2	L_3
1	110	85	95
2	128	88	104
3	135	95	100
4	89	86	78
5	91	98	84
6	120	108	97

(i) Is there any significant difference in the amount of traffic passing through three locations?
(ii) Is there any significant difference in the amount of traffic passing during different days?
(Ans. $F_1 = 6.81$, $F_2 = 3.11$)

Write T for true and F for false of the following:

i The one-way ANOVA test analyzes only one variable
ii. The one –way ANOVA test is always two-tailed
iii. The ANOVA test can be applied to compare three or more population means.
iv. For a one –way ANOVA with k treatments and n observations in all samples taken together the degrees of freedom for the numerator are n-k.
v. For a one –way ANOVA with k treatments and n observations in all samples taken together the degrees of freedom for the denominator are k-1
vi. The t- distribution is in ANOVA test
Ans. (i) T (ii) F (iii) T (iv) F (v) F (vi) F

Multiple Choices

1. **The distribution used in ANOVA is**
 a. χ^2 b. F c. t d. none of these
2. **The one-way ANOVA test is always**
 a. left-tailed b. right-tailed c. two-tailed d. none of these
3. **For a one –way ANOVA with k treatments and n observations in all samples taken together the degrees of freedom for the numerator are**
 a. n-k b. k-1 c. n-1 d. none of these

4. **For a one -way ANOVA with k treatments and n observations in all samples taken together the degrees of.freedom for the denominator are**
   a. k-1 b. n-k c.n-1 d. one of these
5. **The ANOVA test can be applied to compare**
   a .more than two population means b. more than three population means
   c. More than four population means only d. None of these

Answers:

1.	2.	3.	4.	5.
b	b	b	b	a