10.3 ANALYSIS OF VARIANCE (ANOVA)

Analysis of Variance (ANOVA) is a technique for comparing three or more population means. (*Note:* Normal distribution or *t* distribution is sufficient for comparing two population means).

ANOVA can be used for comparing:

(a) the efficiency of workers in 3 plants using the same process.
(b) the salaries of MBAs coming out from 3 different business schools.
(c) the effectiveness of a drug on patients belonging to different age groups.
(d) the effectiveness of a drug on patients belonging to different age groups.

Let us start with an example.

EXAMPLE 10.3 A chain of restaurants in a city wants to compare 3 of its restaurants regarding the service time per customer. One of the owners visited the 3 restaurants during the peak hours and noted the service time for 5 customers in each of the three restaurants. Table 10.3 gives the details.

Table 10.3 Service Time in Minutes in 3 Restaurants

Restaurant 1	Restaurant 2	Restaurant 3
3	3	2
4	4	3.5
5.5	5.5	5
3.5	2.5	6.5
4	3	6

The problem is to test whether the average service time in 3 restaurants are significantly different.

We come back to this problem after discussing the method of ANOVA in the general case.

Notation: Let us use the following notation in the general case.

Let us have k samples from k normal populations with means $\mu_1, \mu_2, \ldots, \mu_k$ and the same variance σ^2. Let the samples have $n_1, n_2, \ldots, n_k$ observations respectively. We list the sample observations as columns. Let N denote the total number of observations.

Let $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k$ be the means of k samples and $s_1^2, s_2^2, \ldots, s_k^2$ their sample variances. Let $\overline{\overline{x}}$ denote the mean of all N observations.

We want to test whether the population means are the same. So, we frame H_0 and H_1 as follows:

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

H_1: $\mu_1, \mu_2, \ldots\ \mu_k$ are not all equal.

Let α be the significance level.

The method of ANOVA is based on the following three steps:

Step 1 (Finding estimate of Between–Column variance)

When the populations have the same mean (by the assumption of H_0) and variance, we can treat the observations from k samples as coming from a single sample of N observations from a normal population of mean μ and variance σ^2.

We estimate σ^2 by considering the variability of sample means. This is called *Between–Column variance*. It is denoted by $\hat{\sigma}_b^2 \cdot \hat{\sigma}_b^2$ and is given by:

$$\boxed{\hat{\sigma}_b^2 = \frac{\Sigma n_j\,(\overline{x}_j - \overline{\overline{x}})^2}{k-1}} \tag{10.5}$$

We get the estimate [RHS of Eq. (10.5)] as follows:

We know that $\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}$ in the case of a single sample of size n. So, $\sigma^2 = n\sigma_{\overline{x}}^2$. As we have k sample we take $s_{\overline{x}}^2 = \dfrac{\Sigma(\overline{x}_j - \overline{\overline{x}})^2}{k-1}$ in place of $\overline{x}^2$. In place of the factor n appearing before $\sigma_{\overline{x}}^2$ in $\sigma^2 = n\sigma_{\overline{x}}^2$ we take n_j (the respective sample size) before $(\overline{x}_j - \overline{\overline{x}})^2$.

Hence,
$$\hat{\sigma}_b^2 = \frac{\Sigma n_j (\bar{x}_j - \bar{\bar{x}})^2}{k-1}$$

Step 2 (Finding the estimate of Within–Column variance).

In this estimate, we consider variations within each sample. The variations within the samples are $s_1^2, s_2^2, ..., s_k^2$. We take the weighted average of $s_1^2, s_2^2, ..., s_k^2$ with $n_1 - 1$, $n_2 - 1$, …, $n_k - 1$ as weights. It is

$$\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1)}$$

i.e., $\dfrac{\Sigma(n_j - 1)s_j^2}{N-k}$, since $n_1 + n_2 + \ldots + n_k = N$

So, the estimate of Within–Column variance is:

$$\boxed{\hat{\sigma}_w^2 = \frac{\Sigma(n_j - 1)s_j^2}{N-k}} \qquad (10.6)$$

Step 3 (Use of F distribution) When H_0 is true, i.e. $\mu_1 = \mu_2 = \cdots = \mu_k$, the two estimates given by Eq. (10.5) and Eq. (10.6) are not significantly different. When H_0 is false, Between–Column variance (derived from sample means) will be significantly different from Within–Column variance (variance within the samples).

We define F by:

$$\boxed{F = \frac{\text{Estimate of Between–Column variance}}{\text{Estimate of Within–Column variance}}} \qquad (10.7)$$

F given by Eq. (10.7) follows on F distribution with $k - 1$ degrees of freedom in the numerator and $N - k$ degrees of freedom in the denominator. [If F defined by Eq. (10.7) is less than 1, we take the ratio of Within–Column variance to Between–Column variance.]

By finding the critical value $F_{(k-1,\, N-k,\, \alpha)}$ we can form the rejection rule.

10.3.1 Method of ANOVA

Let us summarise the method of ANOVA.

We are given k samples drawn from k normal populations of means $\mu_1 = \mu_2 = \cdots = \mu_k$ and the same variance σ^2. Let $n_1, n_2, \ldots, n_k$ be the sizes of the samples,

$\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ be their means, and

$s_1^2, s_2^2, ..., s_k^2$ be their sample variances.

Let $N = n_1 + n_2 + \cdots + n_k$ (Total number of observations)

Let $\bar{\bar{x}}$ be the means of all observations.

Step 1 $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$
$H_1 : \mu_1, \mu_2, \dots \mu_k$ are not all equal.

Step 2 Let the significance level be α.

Step 3 Calculate the estimate of Between–Column variance using Eq. (10.5)

$$\hat{\sigma}_b^2 = \frac{\Sigma n_j (\bar{x}_j - \bar{\bar{x}})^2}{k-1}$$

Step 4 Calculate estimate of Within–Column variance using Eq. (10.6)

$$\hat{\sigma}_w^2 = \frac{\Sigma (n_j - 1) s_j^2}{N-k}$$

Step 5 F follows an F distribution with $k - 1$ degrees of freedom in the numerator and $N - k$ degrees of freedom in the denominator using F distribution defined by Eq. (10.7).

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2}$$

Step 6 Use right-tailed test with significance level α.
If $F > F_{(k-1,\, N-k,\, \alpha)}$ reject H_0.

Note: The main calculation in the method of ANOVA is the calculations of the numerators of Eq. (10.5) and Eq. (10.6). The identity given by Eq. (10.8) makes our calculations easier.

$$\boxed{\Sigma(x_{ij} - \bar{\bar{x}})^2 = \Sigma n_j (\bar{x}_j - \bar{\bar{x}})^2 + \Sigma (n_j - 1) s_j^2} \tag{10.8}$$

As the two terms on RHS of Eq. (10.8) are equal to the numerators of Eqs. (10.5) and (10.6) we call it *sum of squares between columns* (SSC) and *sum of squares within columns* (SSW). LHS of Eq. (10.8) is simply the sum of the squares of the deviations of all observations from $\bar{\bar{x}}$. We call it *total sum of squares* (SST). Equation (10.8) can be stated as:

$$\boxed{\text{SST} = \text{SSC} + \text{SSW}} \tag{10.9}$$

So, $\hat{\sigma}_b^2 = \dfrac{\text{SSC}}{k-1}$ and $\hat{\sigma}_w^2 = \dfrac{\text{SSW}}{N-k}$

We use the identity $\sum_{j=1}^{n} (x_j - \bar{x})^2 = \sum_{j=1}^{n} x_j^2 - n\bar{x}^2$

while calculating SST and SSC. We obtain SSW from Eq. (10.9) (SSW = SST − SSC)

So, the entire method of ANOVA boils down to calculating SST and SSC in a simpler way and these are given in a table called ANOVA table (Table 10.4).

Table 10.4 General ANOVA Table

Source of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between Columns	SSC	$k - 1$	$MSC = \hat{\sigma}_b^2$	$F = \frac{MSC}{MSE}$ or $\frac{MSE}{MSC}$
Within Columns	SSW = SST − SSC	$N - k$	$MSE = \hat{\sigma}_w^2$	
Total	SST	$N - 1$		

Note: The source of variation 'Within Column' is also called *Error*. So, we use MSE for mean sum of squares. SST and SSC are calculated as follows:

The sample observations are given as k columns.

$$\boxed{\text{Correction factor} = \frac{(\Sigma X_1 + \Sigma X_2 + \cdots)^2}{N}} \tag{10.10}$$

$$\boxed{SST = \left(\Sigma X_1^2 + \Sigma X_2^2 + \cdots\right) - \text{Correction factor}} \tag{10.11}$$

$$\boxed{SSC = \left[\frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \cdots\right] - \text{Correction factor}} \tag{10.12}$$

$$\boxed{SSW = SST - SSC} \tag{10.13}$$

Assumptions for using ANOVA

Following are the assumptions for using ANOVA:

(a) The samples are drawn from normal populations.
(b) The samples are random.
(c) The variances of the populations are equal.
(d) The other characteristics of the populations (except those under study) are effectively controlled.

EXAMPLE 10.4 Test whether the average service time in three restaurants given in Example 10.3 are significantly different at a significance level of 0.05.

Solution: Let us arrange the service time (in minutes) in the 3 restaurants as columns under X_1, X_2, X_3 in Table 10.5. The given data are $n_1 = n_2 = n_3 = 5$ $N = 15$ $\alpha = 0.05$. Table 10.6 presents the ANOVA Table.

Table 10.5 Calculation of SST, SSC, SSW for Example 10.4

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
3	9	3	9	2	4
4	16	4	16	3.5	12.25
5.5	30.25	5.5	30.25	5	25
3.5	12.25	2.5	6.25	6.5	42.25
4	16	3	9	6	36
$\Sigma X_1 = 20$	$\Sigma X_1^2 = 83.5$	$\Sigma X_2 = 18$	$\Sigma X_2^2 = 70.5$	$\Sigma X_3 = 23$	$\Sigma X_3^2 = 119.5$

$$\text{Correction factor} = (20 + 18 + 23)^2/15 = 248.07$$

$$\text{SST} = (83.5 + 70.5 + 119.5) - 248.07 = 273.5 - 248.07 = 25.43$$

$$\text{SSC} = \left(\frac{20^2}{5} + \frac{18^2}{5} + \frac{23^2}{5}\right) - 248.07$$

$$= (400 + 324 + 529)/5 - 248.07 = 250.6 - 248.07 = 2.53$$

$$\text{SSW} = \text{SST} - \text{SSC} = 25.43 - 2.53 = 22.9$$

Table 10.6 ANOVA Table for Example 10.4

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between Columns	2.53	2	1.265	1.908/1.265 = 1.508
Within Columns or Error	22.9	12	1.908	
Total	25.43	14		

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

H_1: μ_1, μ_2, μ_3 are not all equal.

$$\alpha = 0.05$$

$$F = 1.508$$

F follows an F distribution with 12 degrees of freedom in the numerator and 2 degrees of freedom in the denominator.

$$F_{(12,\ 2,\ 0.05)} = 19.4$$

As $1.508 < 19.4$, we do not reject H_0.

So, the average service time in the three restaurants are not different.

EXAMPLE 10.5 A company is trying three different training methods to its new employees for enabling them to get familiarised with the company environment and learn the ways the various departments of the company are working. It collected the data given in Table 10.7 regarding the time taken by employees to complete the training methods.

Table 10.7 Time Taken for Training in Hours for Example 10.5

Method 1	Method 2	Method 3
16	23	19
19	28	25
20	19	20
23	22	17
12	18	16

Test whether the three methods are equally effective at a significance level of 0.05.

Solution: Let μ_1, μ_2, μ_3 be the average time taken by the three methods.
The given data are:

$$n_1 = 5 \quad n_2 = 5 \quad n_3 = 6 \quad k = 3 \quad N = 16$$

We frame H_0 and H_1 as follows:

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

H_1: μ_1, μ_2, μ_3 are not all equal.

$$\alpha = 0.05$$

The calculations for SST, SSE, etc. are given in Table 10.8

Table 10.8 Calculation of SST, SSC, SSW for Example 10.5

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
16	256	23	529	19	361
19	361	28	784	25	625
20	400	19	361	20	400
23	529	22	484	17	289
12	144	18	324	23	529
				16	256
$\Sigma X_1 = 20$	$\Sigma X_1^2 = 1690$	$\Sigma X_2 = 18$	$\Sigma X_2^2 = 2482$	$\Sigma X_3 = 23$	$\Sigma X_3^2 = 2460$

$$\text{Correction factor} = \frac{(90 + 110 + 120)^2}{16} = 6400$$

$$\text{SST} = (1690 + 2482 + 2460) - 6400 = 6632 - 6400 = 232$$

$$\text{SSC} = \left[\frac{(90)^2}{5} + \frac{(110)^2}{5} + \frac{(120)^2}{6}\right] - 6400$$

$$= (1620 + 2420 + 2400) - 6400 = 6440 - 6400 = 40$$

$$\text{SSW} = \text{SST} - \text{SSC} = 232 - 40 = 192$$

Table 10.9 presents the ANOVA table for example 10.5.

Table 10.9 ANOVA Table for Example 10.5

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between Columns	40	2	20	20/14.77 = 1.35
Within Columns or Error	192	13	14.77	
Total	232	15		

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

H_1: μ_1, μ_2, μ_3 are not all equal.

$$\alpha = 0.05$$

$$F = 1.35$$

F follows an F distribution with 2 degrees of freedom in the numerator and 13 degrees of freedom in the denominator.

$$F_{(2,\ 13,\ 0.05)} = 3.81$$

As $1.35 < 3.81$, we do not reject H_0.

So, the three methods are equally effective.

10.4 TWO-WAY ANOVA

In Section 10.2, we tested the difference between the means of three or more populations. In Example 10.4, we were testing the difference in mean service time in three restaurants and the variation in service time was due to the restaurants. In Example 10.5, we tested the difference between the time taken for training methods and the variation was due to the nature of the training method. In both cases, the variation we considered was due to one factor alone. (The variation may be due to the nature of the customers and their orders in Example 10.4. In Example 10.5, the variation may be due to the capability and enthusiasm of the workers. But we ignored them).

In this section, we will consider examples where the variation is caused by two factors and conduct ANOVA test for each factor. Let us start with an example.

EXAMPLE 10.6 Let us consider the data regarding the number of five brands of cars in three months given in Table 10.10.

Table 10.10 Number of Five Brands Sold in 3 Months for Example 10.6

Brands	Number of cards sold in '000s (rounded off figures)		
	September 2013	October 2013	November 2013
Alto	24	23	23
Swift	17	19	16
Dzire	17	17	15
Wagon R	15	14	13
Bolero	9	11	8

The variation in the number of cards sold is due to the brand and also the month during which the sales are considered.

We can have some more examples involving two factors.

(a) Prices of bank stocks in stock market under different market conditions (rise in index, fall in index, flat market) and announcement by RBI (increase in SLR rate, decrease in SLR rate, no change). SLR (Statutory Liquidity Ratio)—Minimum percentage of deposits a bank should maintain in the form of cash, gold or other securities.

(b) Number of units produced by different processes in different shifts.

(c) Effectiveness of different brands of detergents under different types of water in removing dirt.

10.4.1 Uses of Two-way ANOVA

The two-way ANOVA is used to test:

(i) the difference between means due to various level of factor 1 (represented as columns)

(ii) the difference between means due to various levels of factor 2 (represented by rows) simultaneously using a single ANOVA table.

Note: The two-way ANOVA eliminates the variation due to the second factor when we test for variation due to first factor.

The key steps of two-way ANOVA method are as follows:

Let r be the number of rows and c be the number of columns in cross tabulation of data.

Step 1 Calculate SSC and SST as in one way ANOVA (see Section 10.3)

$$\text{MSC} = \text{SSC}/c-1$$

Step 2 Calculate SSR (sum of squares due to variation between rows) by considering row sums denoted by ΣY_1, ΣY_2, etc.

$$\text{MSR} = \text{SSR}/r-1$$

Step 3 Calculate SSE (called Error sum of squares) using SSE = SST – SSC – SSR

$$\text{MSE} = \text{SSE}/(C-1)(r-1)$$

Step 4 Use F distribution to test SSC/SSE and SSR/SSE

10.4.2 Construction of Two-way ANOVA Table

Let ΣX_1, ΣX_2,... denote the sum of the first column, second column, etc. (as in one way ANOVA)

Let ΣY_1, ΣY_2, ... denote the sum of the first row, second row, etc.

$$\text{Correction factor} = (\Sigma X_1 + \Sigma X_2 + \cdots)^2/N \text{ [same as Eq. (10.10)]}$$

$$\text{SST} = \left(\Sigma X_1^2 + \Sigma X_2^2 + \cdots\right) - \text{Correction factor [same as Eq. (10.11)]}$$

$$\boxed{SSC = \frac{1}{r}\left[(\Sigma X_1)^2 + (\Sigma X_2)^2 + \cdots\right] - \text{Correction factor } (r\text{-number of rows})} \quad (10.14)$$

$$\boxed{SSR = \frac{1}{c}\left[(\Sigma Y_1)^2 + (\Sigma Y_2)^2 + \cdots\right] - \text{Correction factor } (c\text{-number of columns})} \quad (10.15)$$

$$\boxed{SSE = SST - SSC - SSR} \quad (10.16)$$

Table 10.11 shows two-way ANOVA table.

Table 10.11 Two-way ANOVA Table

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between columns	SSC	$c-1$	MSC	F_1 = MSC/MSE
Between rows	SSR	$r-1$	MSR	F_2 = MSR/MSE
Error	SSE	$(c-1)(r-1)$	MSE	
Total	SST	$N-1$		

EXAMPLE 10.7 Using the data given in Example 10.6, test whether (i) the mean sales of cars are different in different months (ii) the mean sales of cars are different for different brands at $\alpha = 0.05$.

Solution: For (i) H_0 : The sales are equal in different months

H_1: The sales are different in different months

For (ii) H_0: The sales of different brands are equal

H_1: The sales of different brands are different

If we subtract the same number from all entries of the matrix given in Table 10.10, SSC, SSR, SST would not change. So, we subtract 16 from each entry in the matrix given in Table 10.10 (This makes the calculations easier. This technique can be applied for one way ANOVA also). We calculate SST, SSC, SSR etc. in Table 10.12. The only extra column in Table 10.12 is the last column giving ΣY_1, ΣY_2, ΣY_3, ΣY_4 and ΣY_5 being the sum of entries in rows 1, 2, 3, 4, 5. Two-way ANOVA Table is shown in Table 10.13.

Table 10.12 Calculation of SST, SSC, SSR etc. for Example 10.7

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	ΣY_i
8	64	7	49	7	49	22
1	1	3	9	0	0	4
1	1	1	1	−1	1	1
−1	1	−2	4	−3	9	−6
−7	49	−5	25	−8	64	−20
$\Sigma X_1 = 2$	$\Sigma X_1^2 = 116$	$\Sigma X_2 = 4$	$\Sigma X_2^2 = 88$	$\Sigma X_3 = -5$	$\Sigma X_3^2 = 123$	

$$\text{Correction factor} = \frac{(2+4-5)^2}{15} = \frac{1}{15} = 0.067 \text{ [by Eq. (10.9)]}$$

$$\text{SST} = (116 + 88 + 123) - 0.067 = 327 - 0.067 = 326.933$$

$$\text{SSC} = \frac{1}{5}[2^2 + 4^2 + (-5)^2] - 0.067 = 9 - 0.067 = 8.933$$

$$\text{SSR} = \frac{1}{3}[22^2 + 4^2 + 1^2 + (-6)^2 + (-20)^2] - 0.067$$

$$= \frac{1}{3}[484 + 16 + 1 + 36 + 400] - 0.067$$

$$= \frac{1}{3}(937) - 0.067 = 312.263$$

$$\text{SSE} = 326.933 - 8.933 - 312.263$$

Table 10.13 Two-way ANOVA Table for Example 10.7

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between columns	8.933	2	4.467	$F_1 = \frac{4.467}{0.709} = 6.3$
Between rows	312.263	4	78.07	$F_2 = \frac{78.07}{0.709} = 110.11$
Error	5.67	8	0.709	
	326.933	14		

From F tables for 0.05 we get:

$$F_{(2,\,8,\,0.05)} = 4.46$$
$$F_{(4,\,8,\,0.05)} = 3.84$$

As F_1 is greater than 4.46 and F_2 is greater than 3.84, we reject both the null hypotheses.

So, the sales in different months are different (represented by columns) and the sales of different brands (represented by rows) are different.

Supplementary Examples

EXAMPLE 10.8 Which distribution/test is used in testing the following:

(a) Difference between two population means
(b) Difference between three or more population means
(c) Difference between two population proportions
(d) Difference between three or more population proportions.

(e) Difference between two population means using two dependent small samples

Solution:

(a) Normal distribution for two large samples and t distribution for two samples, one sample being small
(b) ANOVA
(c) Normal distribution
(d) χ^2 distribution (test of goodness of fit)
(e) t distribution (using paired difference test)

EXAMPLE 10.9 A steel company is manufacturing the same product in two of its plants A and B having the same capacity. But plant B seems to have variability in output than plant A. The company collected the following data for 31 days in plant A and 41 days in plant B

$$s_A^2 = 960 \text{ tonnes} \quad s_B^2 = 1180 \text{ tonnes}$$

Test whether plant B has more variability in production than plant A at $\alpha = 0.01$

Solution: Let σ_A^2 and σ_B^2 denote the variance of production of plants A and B respectively. As $s_B^2 > s_A^2$, we take

$$F = \frac{s_B^2}{s_A^2}$$

Define H_0 and H_1 as follows:

$$H_0\text{: } \frac{\sigma_B^2}{\sigma_A^2} = 1$$

$$H_1\text{: } \frac{\sigma_B^2}{\sigma_A^2} > 1$$

$$\alpha = 0.02 \quad n = n_1 - 1 = 31 - 1 = 30 \quad d = n_2 - 1 = 41 - 1 = 40$$

As the test is right-tailed, find $F_{(30,\ 40,\ 0.01)} = 2.20$

$$F = \frac{s_B^2}{s_A^2} = \frac{1180}{960} = 1.23$$

As $1.23 < 2.20$, we do not reject H_0.

So, there is no statistical evidence to conclude that plant B has more variability in production than plant A.

EXAMPLE 10.10 A HR Consultant wanted to know whether job satisfaction depends on the individual or profession. He collected job satisfaction score (0–10 scale) of 5 professors, 5 lawyers, 5 chartered accountants and 5 doctors and arrived at the data given in Table 10.14.

Table 10.14 Job Satisfaction in Various Professions for Example 10.10

Professor	Lawyer	Chartered accountant	Doctor
8	7	7	8
6	3	7	7
4	6	3	4
7	5	5	7
5	7	7	6

Test whether there is any difference in the job satisfaction among the four professions at $\alpha = 0.10$.

Solution: Let μ_1, μ_2, μ_3, μ_4 be the average job satisfaction scale of professors, lawyers, chartered accountants and doctors.

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

H_1: μ_1, μ_2, μ_3, μ_4 are not all equal.

$$\alpha = 0.10$$

The calculation of F is given in ANOVA table in Table 10.15.

Table 10.15 Calculation of SSC, SST etc. for Example 10.10

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
8	64	7	49	7	49	8	64
6	36	3	9	7	49	7	49
4	16	6	36	3	9	4	16
7	49	5	25	5	25	7	49
5	25	7	49	7	49	6	36
$\Sigma X_1 = 30$	$\Sigma X_1^2 = 190$	$\Sigma X_2 = 28$	$\Sigma X_2^2 = 168$	$\Sigma X_3 = 29$	$\Sigma X_3^2 = 181$	$\Sigma X_4 = 32$	$\Sigma X_4^2 = 214$

$$\text{Correction factor} = \frac{(30+28+29+32)^2}{20} = 708.05$$

$$\text{SST} = (190 + 168 + 181 + 214) - 708.05 = 44.95$$

$$\text{SSC} = \left(\frac{30^2}{5} + \frac{28^2}{5} + \frac{29^2}{5} + \frac{32^2}{5}\right) - 708.05 = 1.75$$

$$\begin{aligned} \text{SSW} &= \text{SST} - \text{SSC} \\ &= 44.95 - 1.75 \\ &= 43.2 \end{aligned}$$

The ANOVA table is given in Table 10.16.

Table 10.16 ANOVA Table for Example 10.10

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between Columns	1.75	3	0.58	2.7/0.58 = 4.66
Within Columns	43.2	16	2.7	

As the test is two-tailed and $\alpha = 0.10$ find
F_{value} $F_{(3,\ 16,\ 0.05)}$. It is 3.26.
As $4.66 > 3.26$, we reject H_0.

So, the job satisfaction levels are different for different professions.

EXAMPLE 10.11 A dealer in automobiles wants to know whether 3 different brands of tyres give different mileage. He gathered the data presented in Table 10.17 regarding the life of three brands of tyres.

Table 10.17 Different Brands of Tyres

Brand A	Brand B	Brand C
32,000	26,000	29,000
28,500	25,000	32,000
30,000	27,000	31,500
26,000	28,500	29,000
	27,500	

Test whether the average mileage provided by the brands differ significantly at $\alpha = 0.05$.

Solution: Let μ_1, μ_2, μ_3 be the average life of tyres of brands A, B and C (in thousands of kms).

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: μ_1, μ_2, μ_3 are not all equal.

$$\alpha = 0.05$$

The SST, SSC and SSW are calculated in Table 10.18.

Table 10.18 Calculation of SST, SSC, SSW for Example 10.11

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
32	1024	26	676	29	841
28.5	812.25	25	625	32	1024
30	900	27	729	31.5	992.25
26	676	28.5	812.25	29	841
		27.5	756.25		
$\Sigma X_1 = 116.5$	$\Sigma X_1^2 = 3412.25$	$\Sigma X_2 = 134$	$\Sigma X_2^2 = 3598.5$	$\Sigma X_3 = 121.5$	$\Sigma X_3^2 = 3698.25$

We are given the following data:

$$k = 3 \quad n_1 = 4 \quad n_2 = 5 \quad n_3 = 4 \quad N = 13$$

$$\text{Correction factor} = \frac{(116.5 + 134 + 121.5)^2}{13} = \frac{372^2}{13} = 10644.92$$

$$\text{SST} = (3412.25 + 3598.5 + 3698.25) - 10644.92$$
$$= 10709 - 10644.92$$
$$= 64.08$$

$$\text{SSC} = \frac{116.5^2}{4} + \frac{134^2}{5} + \frac{121.5^2}{4} - 10644.92$$
$$= 3393.06 + 3591.2 + 3690.56 - 10644.92$$
$$= 10674.82 - 10644.92$$
$$= 29.9$$

ANOVA table for Example 10.11 is presented in Table 10.19.

Table 10.19 ANOVA Table for Example 10.11

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between Columns	29.9	2	14.95	14.95/3.418 = 4.374
Within Columns	34.18	10	3.418	
	64.18	12		

$$F_{(2,\ 10,\ 0.05)} = 4.10$$

As calculated test statistic 4.374 > 4.10, we reject H_0.

Hence, the average life of tyres of 3 brands are different.

EXAMPLE 10.12 A modified incentive scheme is announced by an IT company having three branches in Karamangala, Electronic City and ITPL (all in Bangalore). The company wants to study the effect of its incentive schemes on professionals in three branches and the effect on three job types—financial services, of dealing with overseas customers and BPO. The data given in Table 10.20 provides the scores of professionals regarding their level of satisfaction in 0–10 scale (10 denoting the maximum satisfaction).

Table 10.20 Effect of Incentive Schemes for Example 10.12

	Score		
Job type	Karamangala	Electronic City	ITPL
Financial services	4	7	9
Dealing with overseas customers	6	5	4
BPO	8	6	7

(a) Test whether the incentive scheme is equally satisfying to professionals of different branches at a significance level of 0.05.

(b) Test whether the scheme is equally satisfying across job types at $\alpha = 0.05$.

Solution: If we subtract the same number from all the entries of Table 10.19, SST, SSC, SSR would not change. So, we subtract 4 from all the entries.

We calculate SST, SSC, SSR etc. in Table 10.21. The last column under ΣY_i gives ΣY_1, ΣY_2 etc. being the row sums. Table 10.22 presents the two-way ANOVA table.

Table 10.21 Calculation of SST, SSC, SSR etc. for Example 10.12

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	ΣY_i
0	0	3	9	5	25	8
2	4	1	1	0	0	3
4	16	2	4	3	9	9
$\Sigma X_1 = 6$	$\Sigma X_1^2 = 20$	$\Sigma X_2 = 6$	$\Sigma X_2^2 = 14$	$\Sigma X_3 = 8$	$\Sigma X_3^2 = 34$	$\Sigma Y_i = 20$

Let H_0: The satisfaction scores are the same in different branches
H_0: The satisfaction scores are the same for different job types
H_1: The satisfaction scores are different in different branches
H_1: The satisfaction scores are different for different job types

$$\alpha = 0.05 \text{ for both tests.}$$

$$\text{Correction factor} = \frac{(6+6+8)^2}{9} = \frac{400}{9} = 44.44$$

$$\text{SST} = (20 + 14 + 34) - 44.44 = 23.56$$

$$\text{SSC} = \frac{1}{3}(6^2 + 6^2 + 8^2) = 44.44 = 45.33 - 44.44 = 0.89$$

$$\text{SSR} = \frac{1}{3}(8^2 + 3^2 + 9^2) - 44.44 = 6.89$$

$$\text{SSE} = \text{SST} - \text{SSC} - \text{SSR} = 23.56 - 0.89 - 6.89$$
$$= 15.78$$

Table 10.22 Two-way ANOVA Table for Example 10.12

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between columns	0.89	2	0.445	3.945/0.445 = 8.87
Within rows	6.89	2	3.445	3.945/3.445 = 1.15
Error	15.78	4	3.945	
	23.56	8		

From F tables for 0.05, we get:

$$F_{2,\,4,\,0.05} = 6.94$$

As 8.87 > 6.94, we reject H_0 (for branches). So, the incentive scheme are not equally satisfying to professional of different branches. As 1.15 < 6.94, we accept H_0 (for job types). So, the incentive schemes are equally satisfying across job types.

EXAMPLE 10.13 The data given in Table 10.23 represents the number of units produced per shift by 5 workers using 3 different machines.

Table 10.23 Units Produced per Shift by 5 Workers Using 3 Machines

Worker	Machine		
	A	*B*	*C*
1	48	44	36
2	50	48	40
3	40	37	38
4	45	45	34
5	50	40	44

(a) Test whether the average productivity is the same for different machines.
(b) Test whether the workers differ in their efficiency.
Take $\alpha = 0.05$ for (a) and (b)

Solution: For (a)

H_0: The average productivity is the same for different machines.
H_1: The average productivity is different for different machines.

For (b)

H_0: The average productivity is the same for different workers.
H_1: The average productivity is different for different workers.

$$\alpha = 0.05 \text{ [for both (a) and (b)]}$$

Subtract 40 from each entry of Table 10.23

$$\text{Correction factor} = \frac{(33+14-8)^2}{15} = \frac{39^2}{15} = 101.4$$

$$\text{SST} = (289 + 114 + 72) - 101.4 = 475 - 101.4 = 373.6$$

$$\text{SSC} = \frac{1}{5}\,[33^2 + 14^2 + (-8)^2] - 101.4 = \frac{1349}{5} - 101.4 = 168.4$$

Calculation of SST, SSC, SSR for Example 10.13 is presented in Table 10.24.

As 8.87 > 6.94, we reject H_0 (for branches). So, the incentive scheme are not equally satisfying to professional of different branches. As 1.15 < 6.94, we accept H_0 (for job types). So, the incentive schemes are equally satisfying across job types.

EXAMPLE 10.13 The data given in Table 10.23 represents the number of units produced per shift by 5 workers using 3 different machines.

Table 10.23 Units Produced per Shift by 5 Workers Using 3 Machines

Worker	Machine		
	A	*B*	*C*
1	48	44	36
2	50	48	40
3	40	37	38
4	45	45	34
5	50	40	44

(a) Test whether the average productivity is the same for different machines.
(b) Test whether the workers differ in their efficiency.
Take $\alpha = 0.05$ for (a) and (b)

Solution: For (a)

H_0: The average productivity is the same for different machines.
H_1: The average productivity is different for different machines.

For (b)

H_0: The average productivity is the same for different workers.
H_1: The average productivity is different for different workers.

$$\alpha = 0.05 \text{ [for both (a) and (b)]}$$

Subtract 40 from each entry of Table 10.23

$$\text{Correction factor} = \frac{(33+14-8)^2}{15} = \frac{39^2}{15} = 101.4$$

$$\text{SST} = (289 + 114 + 72) - 101.4 = 475 - 101.4 = 373.6$$

$$\text{SSC} = \frac{1}{5}\,[33^2 + 14^2 + (-8)^2] - 101.4 = \frac{1349}{5} - 101.4 = 168.4$$

Calculation of SST, SSC, SSR for Example 10.13 is presented in Table 10.24.

Table 10.24 Calculation of SST, SSC, SSR for Example 10.13

X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	Y_i
8	64	4	16	−4	16	8
10	100	8	64	0	0	18
0	0	−3	9	−2	4	−5
5	25	5	25	−6	36	4
10	100	0	0	4	16	14
$\Sigma X_1 = 33$	$\Sigma X_1^2 = 289$	$\Sigma X_2 = 14$	$\Sigma X_2^2 = 114$	$\Sigma X_3 = -8$	$\Sigma X_3^2 = 72$	$\Sigma Y_i = 39$

$$SSR = \frac{1}{3}\,[8^2 + 18^2 + (-5)^2 + 4^2 + 14^2] - 101.4$$

$$= \frac{1}{3}\,(625) - 101.4 = 208.33 - 101.4 = 106.93$$

Table 10.25 presents the ANOVA table for Example 10.13.

Table 10.25 ANOVA Table for Example 10.13

Sources of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between columns	168.40	2	84.2	$F_1 = \dfrac{84.2}{12.28} = 6.86$
Within rows	106.93	2	26.7325	$F_2 = \dfrac{26.7325}{12.28} = 2.1769$
Error	98.27	8	12.28	
	376.6	14		

From F tables for 0.05

$$F_{(2,\ 8,\ 0.05)} = 4.46 \quad F_{(4,\ 8,\ 0.05)} = 3.84$$

As $F_1 > 4.46$ we reject H_0.
As $F_2 < 3.84$, we accept H_0.

So, the average production capacities are different for different machines, but the averag production capacity is the same for different workers.

CASE 10.1

The effect of different therapy methods on mobility among aged patients was evaluated usi a survey of 18 patients. 18 patients were selected to one of the three treatment groups:
Control (No therapy)

Treatment 1 (Physical therapy only)