## 1.4 F-distribution [Test for variance] [Snedecor's F-distribution]

### 1.4.1. The F-distribution

Suppose that two independent normal populations are of interest, when the population means and variances, say $\mu_1$, $\mu_2$, $\sigma_1^2$ and $\sigma_2^2$, are unknown. We wish to test hypothesis about the equality of the two variances, say, $H_0 : \sigma_1^2 = \sigma_2^2$. Assume that two random samples of size $n_1$ from population 1 and of size $n_2$ from population 2 are available, and let $S_1^2$ and $S_2^2$ be the sample variances. We wish to test the hypothesis.

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

The development of a test procedure for these hypothesis requires a new probability distribution, the F distribution.

If $s_1^2$ and $s_2^2$ are the variances of two samples of sizes $n_1$ and $n_2$ respectively, the estimates of the population variance based on these samples are respectively $S_1^2 = \dfrac{n_1 s_1^2}{n_1 - 1}$ and $S_2^2 = \dfrac{n_2 s_2^2}{n_2 - 1}$. The quantities $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ are called the degrees of freedom of these estimates. We want to test if these estimates $S_1^2$ and $S_2^2$ are significantly different or if the samples may be regarded as drawn from the same population or from two populations with same variance $\sigma^2$.

$$\text{Let } F = \frac{S_1^2}{S_2^2} = \frac{\dfrac{n_1 s_1^2}{n_1 - 1}}{\dfrac{n_2 s_2^2}{n_2 - 1}} \quad \dots (1)$$

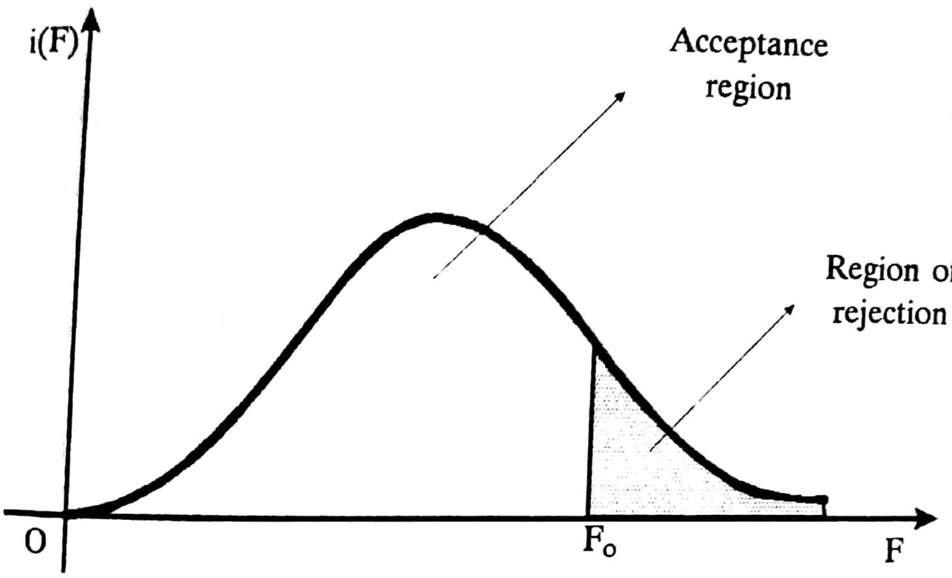
The sampling distibution of F is of the form.

$$y = f(F) = K \frac{F^{\frac{\nu_1 - 2}{2}}}{(\nu_1 F + \nu_2)^{\frac{\nu_1 + \nu_2}{2}}} \quad \cdots (2)$$

where $\nu_1, \nu_2$ are the degrees of freedom of two estimates and K is got by

$$\int_0^\infty f(F)\, dF = 1$$

A rough sketch of this curve is given below :




If $S_1^2 = S_2^2$, then $F = 1$. Hence, our object is to find how far any observed value of F differs from unity, consistent with our assumption of the equality of the population variances.

The area of the curve $y = f(F)$ to the right of $F_0$ gives the probability that $F > F_o$. Here, $F_0$ is the critical value of F. If $F > F_0$ the difference of variances is significant. If $F < F_0$, the difference is not significant. The critical values of F, say $F_0$, is got from F-table corresponding to degrees of freedom $(\nu_1, \nu_2)$ at $\alpha$ level of significance.

In setting $F = \frac{S_1^2}{S_2^2}$, the numerator is greater than denominator i.e., $S_1^2 > S_2^2$, so that $F > 1$.

$$F = \frac{S_1^2}{S_2^2} = \frac{n_1 s_1^2 (n_2 - 1)}{n_2 s_2^2 (n_1 - 1)}$$ where $s_1^2$, $s_2^2$ are variances of two samples

and

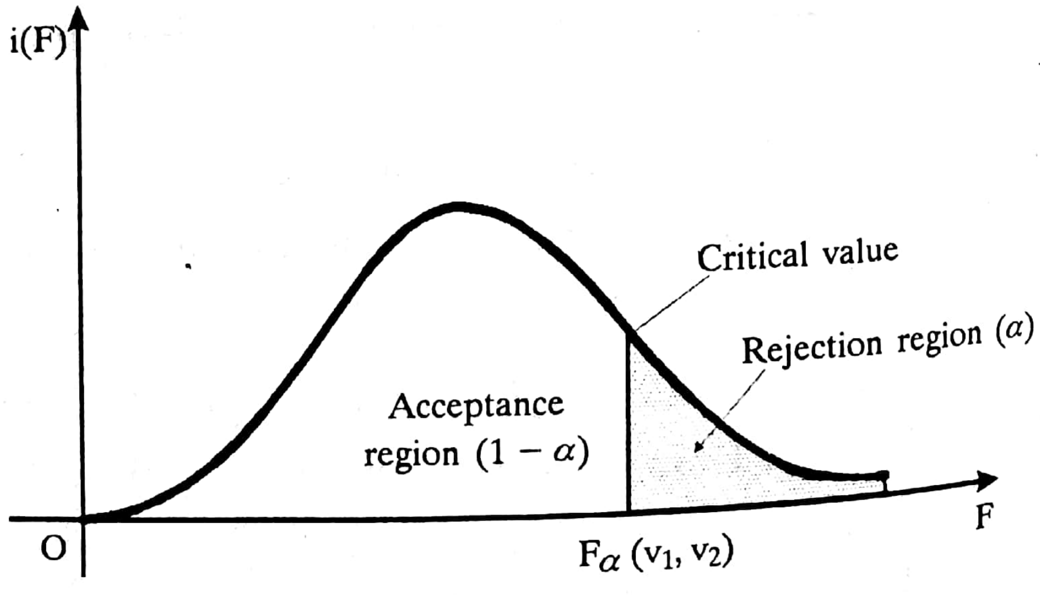
$$S_1^2 = \frac{\Sigma (x - \bar{x})^2}{n_1 - 1}, \quad S_2^2 = \frac{\Sigma (x - \bar{x})^2}{n_2 - 1}, \quad \text{and } S_1^2 > S_2^2$$

If $F < F_0$, critical value, the difference is not significant and if $F > F_0$, the difference is significant.

**Note.** $F > 0$ always.

## Applications :

1. F-test is used to test (i) whether two independent samples have been drawn from the normal populations with the same variance $\sigma^2$, or (ii) whether the two independent estimates of the population variance are homogeneous or not.



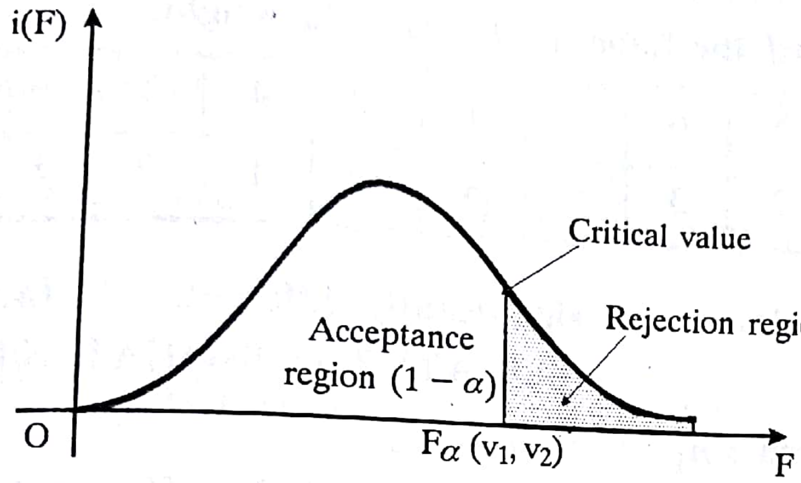

$F_\alpha (v_1, v_2)$

Here, $P(F > F_\alpha (v_1, v_2)) = \alpha.$

Refer to tables for critical $F_\alpha (v_1, v_2)$

### 1.4.2. Properties of the F-distribution

1. The probability curve of the F-distribution is roughly sketched in Fig.




2. The square of the $t$-variate with $n$ degrees of freedom follows a F-distribution with $l$ and $n$ degrees of freedom.

3. The mean of the F-distribution is $\frac{v_2}{v_2-2}\,(v_2 > 2)$

4. The variance of the F-distribution is

$$\frac{2v_2^2\,(v_1+v_2-2)}{v_1\,(v_2-2)^2\,(v_2-4)}\,(v_2 > 4)$$

**Notes :**

1. We should always make F > 1. This is done by taking the larger of the two estimates of $\sigma^2$ as $\sigma_1^2$ and by assuming that the corresponding degree of freedom as $v_1$.

2. To test if two small samples have been drawn from the same normal population, it is not enough to test if their means differ significantly or not, because in this test we assumed that the two samples came from the same population or from populations with equal variance. So, before applying the $t$-test for the significance of the difference of the sample means, we should satisfy ourselves about the equality of the population variances by F-test.

**Example 1.4.1**

A group of 10 rats fed on diet A and another group of 8 rats fed on diet B, recorded the following increase in weight.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Diet A : | 5 | 6 | 8 | 1 | 12 | 4 | 3 | 9 | 6 | 10 |
| Diet B : | 2 | 3 | 6 | 8 | 10 | 1 | 2 | 8 | | |

Find if the variances are significantly different. [A.U N/D 2011]
[A.U A/M 2015 R-13] [A.U N/D 2013 R-13]

**Solution : Given :** $n_1 = 10, n_2 = 8$

| | | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample I | $x_1$ | 5 | 6 | 8 | 1 | 12 | 4 | 3 | 9 | 6 | 10 | 64 |
| | $x_1^2$ | $5^2$ | $6^2$ | $8^2$ | $1^2$ | $12^2$ | $4^2$ | $3^2$ | $9^2$ | $6^2$ | $10^2$ | 512 |
| Sample II | $x_2$ | 2 | 3 | 6 | 8 | 10 | 1 | 2 | 8 | | | 40 |
| | $x_2^2$ | $2^2$ | $3^2$ | $6^2$ | $8^2$ | $10^2$ | $1^2$ | $2^2$ | $8^2$ | | | 282 |

$$\bar{x}_1 = \frac{\Sigma x_1}{10} = \frac{64}{10} = 6.4$$

$$\bar{x}_2 = \frac{\Sigma x_2}{8} = \frac{40}{8} = 5$$

$$s_1^2 = \frac{\Sigma x_1^2}{n_1} - (\bar{x}_1)^2 = \frac{512}{10} - (6.4)^2 = 10.24$$

$$s_2^2 = \frac{\Sigma x_2^2}{n_2} - (\bar{x}_2)^2 = \frac{282}{8} - 25 = 10.25$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{(10)(10.24)}{9} = 11.3777$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{(8)\ (10.25)}{7} = 11.7143$$

$$S_2^2 > S_1^2$$

The parameter of interest is $\sigma_1^2$ and $\sigma_2^2$

1. $H_0 : \sigma_1^2 = \sigma_2^2$ [The difference of variance is not significant]

2. $H_1 : \sigma_1^2 \neq \sigma_2^2$

3. $\alpha = 0.05$, d.f $(v_1) = 9$, d.f $(v_2) = 7$

4. **Table value of F** : 3.29

5. **The test statistic is** $F = \frac{S_2^2}{S_1^2} = \frac{11.7143}{11.3777} = 1.02958$

6. **Conclusion :**

If cal. F < table F, then we accept $H_0$; otherwise, we reject $H_0$

Here, F = 1.02958 < 3.29, we accept $H_0$ at 5% level of significance.

We conclude that the two samples have come from populations with equal variances.

**Example 1.4.2**

**Two independent samples of sizes 9 and 7 from a normal population had the following values of the variables.**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Sample I : | 18 | 13 | 12 | 15 | 12 | 14 | 16 | 14 | 15 |
| Sample II : | 16 | 19 | 13 | 16 | 18 | 13 | 15 | | |

**Do the estimates of the population variance differ significantly at 5% level ?** **[A.U M/J 2014]**

**Solution : Given :** $n_1 = 9$, $n_2 = 7$

| | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample I | $x_1$ | 18 | 13 | 12 | 15 | 12 | 14 | 16 | 14 | 15 | 129 |
| | $x_1^2$ | $18^2$ | $13^2$ | $12^2$ | $15^2$ | $12^2$ | $14^2$ | $16^2$ | $14^2$ | $15^2$ | 1871 |
| Sample II | $x_2$ | 16 | 19 | 13 | 16 | 18 | 13 | 15 | | | 110 |
| | $x_2^2$ | $16^2$ | $19^2$ | $13^2$ | $16^2$ | $18^2$ | $13^2$ | $15^2$ | | | 1760 |

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{129}{9} = 14.3333$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{110}{7} = 15.7143$$

$$s_1^2 = \frac{\Sigma x_1^2}{n_1} - (\bar{x}_1)^2 = \frac{1879}{9} - (14.3333)^2$$

$$= 3.3342$$

$$s_2^2 = \frac{\Sigma x_2^2}{n_2} - (\bar{x}_2)^2 = \frac{1760}{7} - (15.7143)^2$$

$$= 4.4894$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \left(\frac{9}{8}\right)(3.3342) = 3.751$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \left(\frac{7}{6}\right)(4.4894) = 5.2376$$

$$S_2^2 > S_1^2$$

The parameter of interest is $\sigma_1^2$ and $\sigma_2^2$

1. $H_0 : \sigma_1^2 = \sigma_2^2$

2. $H_1 : \sigma_1^2 \neq \sigma_2^2$

3. $\alpha = 0.05$, d.f $(\nu_1) = n_1 - 1 = 9 - 1 = 8$ i.e., d.f (6, 8)

   d.f $(\nu_2) = n_2 - 1 = 7 - 1 = 6$

4. **Table value of F** : 3.58

5. **The test statistic is**

$$\therefore F = \frac{S_2^2}{S_1^2} = \frac{5.2376}{3.7510} = 1.3963$$

6. **Conclusion :**

If cal. F<table F, then we accept $H_0$; otherwise, we reject $H_0$

Here, F = 1.3963 < 3.58, we accept $H_0$ at 5% level of significance.

We conclude that the difference is not significant.

**Example 1.4.3**

**In one sample of 8 observations, the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations was 102.6. Test whether this difference is significant at 5% level, given that the 5 percent point of F for $\nu_1 = 7$ and $\nu_2 = 9$ degrees of freedom is 3.29.**

**Solution : Given :** $n_1 = 8, n_2 = 10$

$$\Sigma (X_1 - \overline{X}_1)^2 = 84.4$$

$$\Sigma (X_2 - \overline{X}_2)^2 = 102.6$$

$$S_1^2 = \frac{\Sigma (X_1 - \overline{X}_1)^2}{n_1 - 1} = \frac{84.4}{7} = 12.057$$

$$S_2^2 = \frac{\Sigma (X_2 - \overline{X}_2)^2}{n_2 - 1} = \frac{102.6}{9} = 11.42$$

Here, $S_1^2 > S_2^2$

The parameter of interest is $\sigma_1^2$ and $\sigma_2^2$

1. $H_0 : \sigma_1^2 = \sigma_2^2$

2. $H_1 : \sigma_1^2 \neq \sigma_2^2$

3. $\alpha = 0.05$, d.f $(v_1) = n_1 - 1 = 7$,
   d.f $(v_2) = n_2 - 1 = 9$

4. **Table value of F** : 3.29

5. **The test statistic is** $F = \dfrac{S_1^2}{S_2^2} = \dfrac{12.057}{11.42} = 1.057$

6. **Conclusion :**

   If cal. F < table F, then we accept $H_0$; otherwise, we reject $H$

   Here, $F = 1.057 < 3.29$, we accept $H_0$ at 5% level of significance.

**Example 1.4.4**

**In one sample of 10 observations, the sum of the squares of th deviations of the sample values from the sample mean was 120 an in another sample of 12 observations it was 314. Test whether thi difference is significant at 5% level of significance.**

**Solution : Given :** $n_1 = 10$, $n_2 = 12$

$$\Sigma (X_1 - \overline{X}_1)^2 = 120$$

$$\Sigma (X_2 - \overline{X}_2)^2 = 314$$

$$S_1^2 = \frac{\Sigma (X_1 - \overline{X}_1)^2}{n_1 - 1} = \frac{120}{9} = 13.33$$

$$S_2^2 = \frac{\Sigma (X_2 - \overline{X}_2)^2}{n_2 - 1} = \frac{314}{11} = 28.55$$

Here, $S_2^2 > S_1^2$

The parameter of interest is $\sigma_1^2$ and $\sigma_2^2$

1. $H_0 : \sigma_1^2 = \sigma_2^2$
2. $H_1 : \sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 0.05$, d.f $(v_1) = n_1 - 1 = 9$,
   d.f $(v_2) = n_2 - 1 = 11$
4. **Table value of F** : 3.11
5. **The test statistic is** $F = \dfrac{S_1^2}{S_2^2} = \dfrac{28.55}{13.33} = 2.14$
6. **Conclusion :**

If cal. F < table F, then we accept $H_0$; otherwise, we reject $H_0$

Here, $F = 2.14 < 3.11$, we accept $H_0$ at 5% level of significance.

We conclude that the samples might have come from two populations having the same variance.

**Example 1.4.5**

**Two random samples give the following results**

| Sample | Size | Sample mean | Sum of squares of deviations from the mean |
|---|---|---|---|
| I | 10 | 15 | 90 |
| II | 12 | 14 | 108 |

**Test whether the samples could have come from the same normal population.** **[A.U M/J 2006, M/J 2012] [A.U N/D 2016 R-13]**

**Solution :** A normal population has two parameters namely the mean $\mu$ and the variance $\sigma^2$. If we want to test the samples from the same normal population, we have to test

(i) the equality of population variances (using F-test)

(ii) the equality of population means (using $t$-test)

Since $t$-test assumes $\sigma_1^2 = \sigma_2^2$ we shall first apply F-test and then $t$-test.

## (i) F-test

**Given :** $n_1 = 10,\ n_2 = 12,\ \bar{x}_1 = 15,\ \bar{x}_2 = 14,$

$$S_1^2 = \frac{\Sigma (x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{90}{9} = 10$$

$$S_2^2 = \frac{\Sigma (x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{108}{11} = 9.8181$$

$$S_1^2 > S_2^2$$

The parameters of interest is $\sigma_1^2$ and $\sigma_2^2$

1. $H_0 : \sigma_1^2 = \sigma_2^2$
2. $H_1 : \sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 0.05$, d.f $(v_1) = n_1 - 1 = 9$
   d.f $(v_1) = n_2 - 1 = 11$
4. **Table value of F** : 2.90
5. **The test statistic is** $F = \dfrac{S_1^2}{S_2^2} = \dfrac{10}{9.8182} = 1.019$
6. **Conclusion :**

If cal. F < table F, then we accept $H_0$ ; otherwise, we reject $H_0$

Here, $F = 1.019 < 2.90$, we accept $H_0$ at 5% level of significance.

[Note : If F-test failed, then $t$-test should not be used]

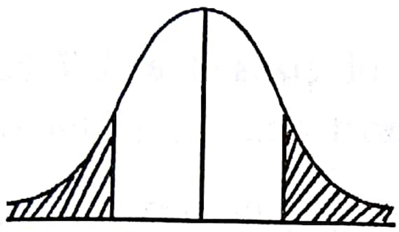
## (ii) $t$-test

**Given :** $n_1 = 10,\ n_2 = 12,\ s_1^2 = 10,\ s_2^2 = 9.8181$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{100 + 117.82}{10 + 12 - 2} = 10.9$$

∴ The parameter of interests is $\mu_1$ and $\mu_2$

1. $H_0 : \mu_1 = \mu_2$
2. $H_1 : \mu_1 \neq \mu_2$
3. $\alpha = 0.05$, d.f $= n_1 + n_2 - 2 = 20$ [Two-tailed test]
4. Critical region



$-t_{\frac{\alpha}{2}} = -2.086$ $\qquad$ $t_{\frac{\alpha}{2}} = 2.086$

5. The test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$= \frac{15 - 14}{\sqrt{(10.9)\left(\frac{1}{10} + \frac{1}{12}\right)}} = 0.707$$

6. Conclusions : Sincem $|t| = 0.707 < 2.086$, we accept $H_0$ at 5% level of significance.

## EXERCISES

1. Two random samples drawn from two normal populations are :

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample I : | 20 | 16 | 26 | 27 | 23 | 22 | 18 | 24 | 25 | 19 | | |
| Sample II : | 27 | 33 | 42 | 35 | 32 | 34 | 38 | 28 | 41 | 43 | 30 | 37 |

Obtain estimates of the variances of the populations and test whether the populations have the same variance.

[**Ans.** They have same variances]

2. **Definition** : Analysis of variance (ANOVA) is the separation of variance ascribable to one group of causes from the variance ascribable to other groups".

It is nothing but an arithmetical procedure, used to express the total variation of data, as the sum of its non-negative components.

3. **Assumptions** : For the validity of the F-test in ANOVA, the following assumptions are made : [A.U. CBT A/M 2011]

(i) The observations are independent

(ii) Parent population from which observations are taken is normal and

(iii) Various treatment and environmental effects are additive, in nature

| | |
|---|---|
| S.S.C | - Sum of Squares between columns |
| T.S.S | - Total Sum of Squares |
| S.S.T | - S.S. due to treatments |
| M.S.S | - Mean sum of squares |
| S.S.E | - Error sum of squares (or) within sum of squares |
| R.S.S | - Row sum of squares |
| C.F | - Correction factor |
| C.D | - Critical Difference |
| S.S.R | - Sum of squares between Rows |
| M.S.C | - Mean sum of squares (between columns) |
| M.S.E | - Mean sum of squares (within columns) |
| M.S.R | - Mean sum of squares (between rows) |
| N | - Number of Observations |
| $N_1$ | - Number of elements in each column |
| $N_2$ | - Number of elements in each row |

## 2.2 ONE WAY CLASSIFICATION

One-way classification observations are classified according to one factor. This is exhibited column-wise.

A set of N = cr observations classified in one direction may be represented as follows.

| | $X_1$ | $X_2$ | ... | $X_c$ |
|---|---|---|---|---|
| | $x_{11}$ | $x_{21}$ | ... | $x_{c1}$ |
| | $x_{12}$ | $x_{22}$ | ... | $x_{c2}$ |
| | $x_{13}$ | $x_{23}$ | ... | $x_{c3}$ |
| | . . . | . . . | . . . | . . . |
| | . . . | . . . | . . . | . . . |
| Mean | $\overline{X}_1$ | $\overline{X}_2$ | ... | $\overline{X}_c$ |

$$\text{Grand mean } \overline{\overline{X}} = \frac{\overline{X}_1 + \overline{X}_2 + \ldots + \overline{X}_c}{c}$$

Then, 'between column' sum of squares : $SSC = \sum_j (\overline{X}_j - \overline{\overline{X}})^2$

'Within column' sum of squares : $SSE = \sum_i \sum_j (x_{ij} - \overline{X}_j)^2$

Total sum of squares : $TSS = \sum_i \sum_j (x_{ij} - \overline{\overline{x}})^2$

Also, it can be shown that TSS = SSC + SSE

If each of these sum of squares is divided by the corresponding number of degrees of freedom, we get the mean sum of squares. These information are presented in the following table called ANOVA table.

**(Analysis of Variance Table)**

| Source of Variation | Sum of squares | Degrees of freedom | Mean sum of squares | Variance ratio |
|---|---|---|---|---|
| Between columns | SSC | C − 1 | $MSC = \frac{SSC}{C-1}$ | $F = \frac{MSC}{MSE}$ (or) |
| Within columns | SSE | N − C | $MSE = \frac{SSE}{N-C}$ | $F = \frac{MSE}{MSC}$ |
| Total | TSS | N − 1 | | |

The F ratio should be calculated in such a way that F > 1.

## ...ndomized Design (C.R.D)

The completely randomized design is the simplest of all the designs, based on principles of randomization and replication. In this design, treatments are allocated at random to the experimental units over the entire experimental material.

### ■ 1. Merits ■

(i) **C.R.D.** results in the maximum use of the experimental units, since all the experimental material can be used.

(ii) **The design is very flexible.** Any number of treatments can be used and different treatments can be used unequal number of times without unduly complicating the statistical analysis in most of the cases.

(iii) The statistical analysis remains simple, if some or all the obsevations for any treatment are rejected or lost or missing for some purely random accidental reasons. We merely carry out the standard analysis on the available data. Moreover, the loss of information due to missing data is smaller in comparison with any other design.

(iv) It provides the maximum number of degrees of freedom for the estimation of the error variance, which increases the sensitivity or the precision of the experiment for small experiments, (i.e.,) for experiments with small number of treatments.

### ■ 2. Demerits ■

(i) In certain circumstances, the design suffers from the disadvantage of being inherently less informative than other more sophisticated layouts. This usually happens, if the experimental material is not homogeneous. Since, randomization is not restricted in any direction to ensure that the units receiving one treatment are similar to those receiving the other treatment, the whole variations among the experimental units is included in the residual variance. This makes the design less efficient and results in less sensitivity in detecting significant effects. As such C.R.D. is seldom used i

## 3. Applications

(i) Completely randomised design is more useful in laboratory technique and methodological studies, (e.g.,) in physics, chemistry or cookery, in chemical and biological experiments, in some green house studies, etc., where either the experimental material is homogeneous or the intrinsic variability between units can be reduced.

(ii) C.R.D is also recommended in situations where an appreciable fraction of units is likely to be destroyed or fail to respond.

### Working Procedure [One-way classification CRD]

1. $H_0$ : There is no significant difference between the treatments.

2. $H_1$ : There is significant difference between the treatments.

Step 1 : Find N i.e., the number of observations

Step 2 : Find T i.e., the total value of all observations

Step 3 : Find $\frac{T^2}{N}$ i.e., the correction factor

Step 4 : Calculate the total sum of squares.

$$TSS = \sum X_1^2 + \sum X_2^2 + \dots - \frac{T^2}{N}$$

Step 5 : Calculate the column sum of squares

$$SSC = \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_1} + \frac{(\Sigma X_3)^2}{N_1} + \dots - \frac{T^2}{N}$$

Here, $N_1$ is the number of elements in each column.

$$SSE = TSS - SSC$$

Step 6 : Prepare the ANOVA table to calculate F-ratio.

Step 7 : Find the table value.

Step 8 : Conclusion :

The following are the numbers of mistakes made in 5 successive days of 4 technicians working for a photographic laboratory :

| Technician I ($X_1$) | Technician II ($X_2$) | Technician III ($X_3$) | Technician IV ($X_4$) |
|---|---|---|---|
| 6 | 14 | 10 | 9 |
| 14 | 9 | 12 | 12 |
| 10 | 12 | 7 | 8 |
| 8 | 10 | 15 | 10 |
| 11 | 14 | 11 | 11 |

Test at the level of significance $\alpha = 0.01$, whether the differences among the 4 sample means, can be attributed to chance.

[A.U. A/M 2004, A.U A/M 2011]

**Solution :** $H_0$ : There is no significant difference between the technicians.

$H_1$ : Significant difference between the technicians.

We shift the origin to 10.

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | Total | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ |
|---|---|---|---|---|---|---|---|---|
| –4 | 4 | 0 | –1 | –1 | 16 | 16 | 0 | 1 |
| 4 | –1 | 2 | 2 | 7 | 16 | 1 | 4 | 4 |
| 0 | 2 | –3 | –2 | –3 | 0 | 4 | 9 | 4 |
| –2 | 0 | 5 | 0 | 3 | 4 | 0 | 25 | 0 |
| 1 | 4 | 1 | 1 | 7 | 1 | 16 | 1 | 1 |
| Total –1 | 9 | 5 | 0 | 13 | 37 | 37 | 39 | 10 |

**Step 1 :** N = 20

**Step 2 :** T = 13

**Step 3 :** $\frac{T^2}{N} = \frac{(13)^2}{20} = 8.45$

**Step 4 : TSS** $= \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}$

$= 37 + 37 + 39 + 10 - 8.45 = 114.55$

**Step 5 : SSC** $= \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_1} + \frac{(\Sigma X_3)^2}{N_1} + \frac{(\Sigma X_4)^2}{N_1} - \frac{T^2}{N}$

$N_1 \rightarrow$ Number of elements in each column

$= \frac{(-1)^2}{5} + \frac{(9)^2}{5} + \frac{(5)^2}{5} - 0 - 8.45$

$= \frac{1}{5} + \frac{81}{5} + 5 - 8.45 = 0.2 + 16.2 + 5 - 8.45 = 12.95$

SSE = TSS − SSC

$= 114.55 - 12.95 = 101.6$

**Step 6 : ANOVA**

| Source of Variation | Sum of squares | d.f. | Mean square | variance ratio | Table value at 1% level |
|---|---|---|---|---|---|
| Between columns | SSC = 12.95 | C − 1 = 4−1 = 3 | $MSC = \frac{SSC}{C-1} = \frac{12.95}{3} = 4.317$ | $F_c = \frac{MSE}{MSC} = \frac{6.35}{4.317} = 1.471 > 1$ | $F_c(16,3) = 26.9$ |
| Error | SSE = 101.6 | N−C = 20−4 = 16 | $MSE = \frac{SSE}{N-C} = \frac{101.6}{16} = 6.35$ | Since $\frac{MSC}{MSE} < 1$ | |
| Total | 114.55 | | | | |

**Step 7 : Conclusion :** Cal $F_c$ < Tab $F_c$

So, we accept $H_0$

It is concluded that there is no significant difference between the technicians at 5% level of significance.

**There are three main brands of a certain powder. A set of 120 sample values is examined and found to be allocated among four groups (A, B, C and D) and three brands (I, II, III) as shown here under :**

**[A.U. A/M. 2004]**

| **Brands** | **Groups** | | | |
|---|---|---|---|---|
| | **A** | **B** | **C** | **D** |
| I | 0 | 4 | 8 | 15 |
| II | 5 | 8 | 13 | 6 |
| III | 8 | 19 | 11 | 13 |

**Is there any significant difference in brands preference ? Answer at 5% level.**

**Solution :** $H_0$ : There is no significant difference in brands.

$H_1$ : There is significant difference in brands.

| **Brands** | **Groups** | | | | **Total** | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ |
|---|---|---|---|---|---|---|---|---|---|
| | **A** ($X_1$) | **B** ($X_2$) | **C** ($X_3$) | **D** ($X_4$) | | | | | |
| I ($Y_1$) | 0 | 4 | 8 | 15 | 27 | 0 | 16 | 64 | 225 |
| II ($Y_2$) | 5 | 8 | 13 | 6 | 32 | 25 | 64 | 169 | 36 |
| III($Y_3$) | 8 | 19 | 11 | 13 | 51 | 64 | 361 | 121 | 169 |
| Total | 13 | 31 | 32 | 34 | 110 | 89 | 441 | 354 | 430 |

**Step 1 :** N = 12

**Step 2 :** T = 110

**Step 3 :** $\frac{T^2}{N} = \frac{(110)^2}{12} = 1008.3$

**Step 4 :** TSS $= \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^4 - \frac{T^2}{N}$

$= 89 + 441 + 354 + 430 - 1008.3$

$= 305.7$

**Step 5 :** SSR $= \frac{(\Sigma Y_1)^2}{N_2} + \frac{(\Sigma Y_2)^2}{N_2} + \frac{(\Sigma Y_3)^2}{N_2} - \frac{T^2}{N}$

[$N_2 \rightarrow$ No. of elements in each row]

$= \frac{(27)^2}{4} + \frac{(32)^2}{4} + \frac{(51)^2}{4} - 1008.3$

$= 182.25 + 256 + 650.25 - 1008.3 = 80.2$

SSE = TSS − SSR

$= 305.7 - 80.2 = 225.50$

**Step 6 :** ANOVA

| Source of variation | Sum of squares | d.f. | Mean square | Variance ratio | Table value at 5% level |
|---|---|---|---|---|---|
| between rows | SSR = 80.2 | $r - 1 = 3 - 1 = 2$ | $MSR = \frac{SSR}{r-1} = \frac{80.2}{2} = 40.1$ | $F_R = \frac{MSR}{MSE} = \frac{40.1}{20.06} = 1.999$ | $F_R$ (2,9) = 4.26 |
| Error | SSE = 225.5 | $N-r = 12-3 = 9$ | $MSE = \frac{SSE}{N-r} = \frac{225.5}{9} = 20.06$ | | |
| Total | 305.7 | | | | |

**Conclusion :** Cal $F_R$ < Table $F_R$

So, we accept $H_0$.

**A completely randomised design experiment with 10 plots and 3 treatments gave the following results :**

[A.U. N/D 2007]

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Plot No : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Treatment : | A | B | C | A | C | C | A | B | A | B |
| Yield : | 5 | 4 | 3 | 7 | 5 | 1 | 3 | 4 | 1 | 7 |

**Analyse the results for treatment effects.**

**Solution :**

| | | | | |
|---|---|---|---|---|
| A | 5 | 7 | 3 | 1 |
| B | 4 | 4 | 7 | |
| C | 3 | 5 | 1 | |

$H_0$ : There is no significant difference

$H_1$ : There is significant difference

| $X_1$ A | $X_2$ B | $X_3$ C | Total | $X_1^2$ | $X_2^2$ | $X_3^2$ |
|---|---|---|---|---|---|---|
| 5 | 4 | 3 | 12 | 25 | 16 | 9 |
| 7 | 4 | 5 | 16 | 49 | 16 | 25 |
| 3 | 7 | 1 | 11 | 9 | 49 | 1 |
| 1 | | | 1 | 1 | | |
| 16 | 15 | 9 | 40 | 84 | 81 | 35 |

Step : 1. $N = 10$

2. $T = 40$

3. $\frac{T^2}{N} = \frac{(40)^2}{10} = 160$

4. TSS $= \sum X_1^2 + \sum X_2^2 + \sum X_3^2 - \frac{T^2}{N}$

$= 84 + 81 + 35 - 160 = 40$

5. SST $= \frac{(\sum X_1)^2}{N_1} + \frac{(\sum X_2)^2}{N_1} + \frac{(\sum X_3)^2}{N_1} - \frac{T^2}{N}$

$N_1 \rightarrow$ Number of elements in each column.

$= \frac{(16)^2}{4} + \frac{(15)^2}{3} + \frac{(9)^2}{3} - 160$

$= 64 + 75 + 27 - 160 = 6$

SSE $=$ TSS $-$ SST $= 40 - 6 = 34$

6. ANOVA Table

| Source of variation | Sum of squares | d.f. | MSS | Variance ratio | Total value 5% level |
|---|---|---|---|---|---|
| Between treat-ments | SST = 6 | $C - 1 = 3 - 1 = 2$ | $MST = \frac{SST}{C-1} = \frac{6}{2} = 3$ | $F_T = \frac{MSE}{MST} = 1.62 > 1$ Since $\frac{MST}{MSE} < 1$ | $F_T(7,2) = 19.35$ |
| Error | SSE = 34 | $N - C = 10 - 3 = 7$ | $MSE = \frac{SSE}{N-C} = \frac{34}{7} = 4.86$ | | |

Step 7 : Conclusion Cal F < Table F

So, we accept $H_0$

The following table shows the lives in hours of four brands of electric lamps.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Brand A : | 1610 | 1610 | 1650 | 1680 | 1700 | 1720 | 1800 | |
| B : | 1580 | 1640 | 1640 | 1700 | 1750 | | | |
| C : | 1460 | 1550 | 1600 | 1620 | 1640 | 1660 | 1740 | 1820 |
| D : | 1510 | 1520 | 1530 | 1570 | 1600 | 1680 | | |

Perform an analysis of variance test the homogneity of the mean lives of the four brands of Lamps. [A.U. A/M. 2008] [A.U N/D 2011] [A.U Tvli M/J 2011] [A.U A/M 2015 R-13]

*Solution :*

$H_0$ : There is no significant difference between the four brands.

$H_1$ : There is a significant difference between the four brands.

Subtract 1600 and then divided by 10, we get

| $X_1$ A | $X_2$ B | $X_3$ C | $X_4$ D | Total | $X_1^2$ | $X_2^2$ | $X_3^3$ | $X_4^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | –2 | –14 | –9 | –24 | 1 | 4 | 196 | 81 |
| 1 | 4 | –5 | –8 | –8 | 1 | 16 | 25 | 64 |
| 5 | 4 | 0 | –7 | 2 | 25 | 16 | 0 | 49 |
| 8 | 10 | 2 | –3 | 17 | 64 | 100 | 4 | 9 |
| 10 | 15 | 4 | 0 | 29 | 100 | 225 | 16 | 0 |
| 12 | – | 6 | 8 | 26 | 144 | – | 36 | 64 |
| 20 | – | 14 | – | 34 | 400 | – | 196 | – |
| – | – | 22 | – | 22 | – | – | 484 | – |
| **57** | **31** | **29** | **–19** | **98** | **735** | **361** | **957** | **267** |

Step 1 : N = 26

Step 2 : T $= 98$

Step 3 : C.F $= \frac{T^2}{N} = \frac{9604}{26} = 369.39$

Step 4 : TSS $= \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}$

$= 735 + 361 + 957 + 267 - 369.39$

$= 1950.61$

Step 5 : SSC $= \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_1} + \frac{(\Sigma X_3)^2}{N_1} + \frac{(\Sigma X_4)^2}{N_1} - \frac{T^2}{N}$

$N_1 \rightarrow$ Number of elements in their respective columns.

$$= \frac{(57)^2}{7} + \frac{(31)^2}{5} + \frac{(29)^2}{8} + \frac{(-19)^2}{6} - 369.39$$

$$= \frac{3249}{7} + \frac{961}{5} + \frac{841}{8} + \frac{361}{6} - 369.39$$

$$= 464.14 + 192.2 + 105.13 + 60.17 - 369.39 = 452.25$$

SSE $=$ TSS $-$ SSC

$= 1950.61 - 452.25 = 1498.36$

Step 6 : ANOVA

| Source of Variation | Sum of squares | d.f. | Mean square | Variance Ratio | Table value 5% level |
|---|---|---|---|---|---|
| Between columns | SSC $= 452.25$ | $C - 1$ $= 4 - 1$ $= 3$ | $MSC = \frac{SSC}{C-1}$ $= \frac{452.25}{3}$ $= 150.75$ | $F_C = \frac{MSC}{MSE}$ $= \frac{150.75}{68.11}$ $= 2.21 > 1$ | $F_C$ (3,22) $= 3.05$ |
| Error | SSE $= 1498.36$ | $N - C$ $= 26 - 4$ $= 22$ | $MSE = \frac{SSE}{N-C}$ $= \frac{1498.36}{22}$ $= 68.11$ | Since $\frac{MSE}{MSC} < 1$ | |

Scanned by CamScanner

**Working Rule**

1. $H_0$ : There is no significant difference.

2. $H_1$ : There is a significant difference.

Test the hypothesis that variation between varieties and between blocks do not differ significantly from the variance due to random errors.

**Arrange calculation of sum of squares.**

**Step 1 :** Find N. **Step 2 :** Find T. **Step 3 :** Find $\frac{T^2}{N}$

**Step 4 :** Find TSS **Step 5 :** Find SSC. **Step 6 :** Find SSR

**Step 7 : SSE = TSS − SSC − SSR**

Prepare the ANOVA Table

**Step 8 :** Find Table $F_c$ and $F_R$

**Step 9 :** Conclusion

**Example 2.3.1**

**An experiment was designed to study the performance of 4 different detergents for cleaning fuel injectors. The following "cleanliness" readings were obtained with specially designed equipment for 12 tanks of gas distributed over 3 different models of engines :**

| | Engine 1 | Engine 2 | Engine 3 | Total |
|---|---|---|---|---|
| Detergent A | 45 | 43 | 51 | 139 |
| Detergent B | 47 | 46 | 52 | 145 |
| Detergent C | 48 | 50 | 55 | 153 |
| Detergent D | 42 | 37 | 49 | 128 |
| **Total** | 182 | 176 | 207 | 565 |

Perform the ANOVA and test at 0.01 level of significance, whether there are differences in the detergents or in the engines.

[A.U. Model] [A.U. N/D. 2004] [A.U CBT A/M 2011] [A.U M/J 2007, N/D 2008] [A.U N/D 2015 R13]

**Solution :** The above data are classified according to criteria (i) Detergent (ii) Engine.

In order to simplify calculations, we code the data by subtracting 50 from each figure.

| Detergent | Engine ($X_1$) | ($X_2$) | ($X_3$) | Total | $X_1^2$ | $X_2^2$ | $X_3^2$ |
|---|---|---|---|---|---|---|---|
| A ($Y_1$) | –5 | –7 | 1 | –11 | 25 | 49 | 1 |
| B ($Y_2$) | –3 | –4 | 2 | –5 | 9 | 16 | 4 |
| C ($Y_3$) | –2 | 0 | 5 | 3 | 4 | 0 | 25 |
| D ($Y_4$) | –8 | –13 | –1 | –22 | 64 | 169 | 1 |
| Total | –18 | –24 | 7 | –35 | 102 | 234 | 31 |

1. $H_0$ : There is no significant difference between columns means as well as row means.

2. $H_1$ : There is significant difference between columns means or the row means.

**Step 1 :** $N = 12$ [Total number of entries]

**Step 2 :** $T = -35$

**Step 3 :** $\frac{T^2}{N} = \frac{(-35)^2}{12} = 102.08$

**Step 4 :** $TSS = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - \frac{T^2}{N}$

$$= (102) + (234) + (31) - 102.08 = 367 - 102.08 = 264.92$$

**Step 5 :** $SSC = \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_1} + \frac{(\Sigma X_3)^2}{N_1} - \frac{T^2}{N}$

[$N_1 \rightarrow$ Number of elements in each column]

**Example 2.3.5**

A tea company appoints four salesmen A, B, C and D and observes their sales in three seasons - summer, winter and monsoon. The figures (in lakhs) are given in the following table.

| Seasons | Salesmen | | | | Season's Total |
|---|---|---|---|---|---|
| | A | B | C | D | |
| Summer | 36 | 36 | 21 | 35 | 128 |
| Winter | 28 | 29 | 31 | 32 | 120 |
| Monsoon | 26 | 28 | 29 | 29 | 112 |
| Salesmen's Total | 90 | 93 | 81 | 96 | 360 |

**(i) Do the salesmen significantly differ in performance ?**

**(ii) Is there significant difference between the seasons ?**

**[A.U N/D 2012] [A.U M/J 2016 R13]**

*Solution* : The above data are classified according to criteria (i) salesmen and (ii) seasons. In order to simplify calculations, we code the data by subtracting 30 from each figure. The data in the coded form are given below :

| Seasons | | Salesmen | | | | Seasons' Total | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | A ($X_1$) | B ($X_2$) | C ($X_3$) | D ($X_4$) | | | | | |
| $Y_1$ | Summer | +6 | +6 | −9 | +5 | +8 | 36 | 36 | 81 | 25 |
| $Y_2$ | Winter | −2 | −1 | +1 | +2 | 0 | 4 | 1 | 1 | 4 |
| $Y_3$ | Monsoon | −4 | −2 | −1 | −1 | −8 | 16 | 4 | 1 | 1 |
| Total | | 0 | 3 | −9 | 6 | Grand Total T = 0 | 56 | 41 | 83 | 30 |

1. $H_0$ : There is no significant difference between column means as well as row means.

Scanned by CamScanner

2. $H_1$ : There is significant difference between column means or the row means.

Step 1 : $N = 12$ Step 2 : $T = 0$

Step 3 : $\frac{T^2}{N} = \frac{(0)^2}{12} = 0$ (number of items or N is 12)

Step 4 : $TSS = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \frac{T^2}{N}$

$= 56 + 41 + 83 + 30 - 0 = 210$

Step 5 : $SSC = \frac{(\Sigma X_1)^2}{N_1} + \frac{(\Sigma X_2)^2}{N_1} + \frac{(\Sigma X_3)^2}{N_1} + \frac{(\Sigma X_4)^2}{N_1} - \frac{T^2}{N}$

[$N_1$ = Number of elements in each column]

$= \frac{(0)^2}{3} + \frac{(3)^2}{3} + \frac{(-9)^2}{3} + \frac{(6)^2}{3} - \frac{T^2}{N} = 0 + 3 + 27 + 12 - 0 = 42$

Step 6 : $SSR = \frac{(\Sigma Y_1)^2}{N_2} + \frac{(\Sigma Y_2)^2}{N_2} + \frac{(\Sigma Y_3)^2}{N_2} - \frac{T^2}{N}$

[$N_2$ = Number of elements in each row]

$= \frac{(8)^2}{4} + \frac{(0)^2}{4} + \frac{(-8)^2}{4} - \frac{T^2}{N} = 16 + 0 + 16 - 0 = 32$

**SSE = TSS − SSC − SSR** = 210 − 42 − 32 = 136

Step 7 : Table of Analysis of Variance :

| Sources of variation | Sum of squares | d.f. | Mean squares | Variance | Table value at 5% level |
|---|---|---|---|---|---|
| Between columns (salesmen) | SSC = 42 | $c-1 = 4-1 = 3$ | $MSC = \frac{SSC}{c-1} = 14$ | $F_C = \frac{MSE}{MSC} = \frac{22.67}{14} = 1.619$ | $F_C(6, 3) = 8.94$ |
| Between rows (seasons) | SSR = 32 | $r-1 = 3-1 = 2$ | $MSR = \frac{SSR}{r-1} = \frac{32}{2} = 16$ | $F_R = \frac{MSE}{MSR} = \frac{22.67}{16} = 1.417$ | $F_R(6,2) = 19.33$ |
| Residual | SSE = 136 | $N-c-r+1 = 6$ | $MSE = \frac{SSE}{N-c-r+1} = \frac{136}{6} = 22.67$ | | |
| | 210 | 11 | | | |

**Step 8 : Conclusion :** Cal F < Table F

Hence, there is no significant difference in the seasons as far as the sales are concerned.

Thus, the test shows that the salesmen and the seasons are alike, so far as the sales are concerned.

## Example 2.3.6

**A set of data involving four "four tropical feed stuffs A, B, C, D" tried on 20 chicks is given below. All the twenty chicks are treated alike in all respects except the feeding treatments and each feeding treatment is given to 5 chicks. Analyze the data.**
**Weight gain of baby chicks fed on different feeding materials composed of tropical feed stuffs.**

| | | | | | | Total $T_i$ |
|---|---|---|---|---|---|---|
| A | 55 | 49 | 42 | 21 | 52 | 219 |
| B | 61 | 112 | 30 | 89 | 63 | 355 |
| C | 42 | 97 | 81 | 95 | 92 | 407 |
| D | 169 | 137 | 169 | 85 | 154 | 714 |
| | | | | Grand total | | G=1695 |

**[A.U A/M 2010]**

**Solution :**

1. $H_0$ : There is no significant difference between column means as well as row means.

2. $H_1$ : There is significant difference between column means or the row means.

Subtract 50 from each value

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | T | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ | $X_5^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y_1=A$ | 5 | −1 | −8 | −29 | 2 | −31 | 25 | 1 | 64 | 841 | 4 |
| $Y_2=B$ | 11 | 62 | −20 | 39 | 13 | 105 | 121 | 3844 | 400 | 1521 | 169 |
| $Y_3=C$ | −8 | 47 | 31 | 45 | 42 | 157 | 64 | 2209 | 961 | 2025 | 1764 |
| $Y_4=D$ | 119 | 87 | 119 | 35 | 104 | 464 | 14161 | 7569 | 14161 | 1225 | 10816 |
| T | 127 | 195 | 122 | 90 | 161 | 695 | 14371 | 13623 | 15586 | 5612 | 12753 |

Step 1 : $N = 20$

Step 2 : $T = 695$

Step 3 : $\frac{T^2}{N} = \frac{(695)^2}{20} = 24151.25$

Step 4 : $TSS = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 + \Sigma X_5^2 - \frac{T^2}{N}$

$= 14371 + 13623 + 15586 + 5612 + 12753 - 24151.25$

$= 37793.75$

Step 5 : $SSC = \frac{(\Sigma X_1^2)}{N_1} + \frac{(\Sigma X_2)^2}{N_1} + \frac{(\Sigma X_3)^2}{N_1} + \frac{(\Sigma X_4)^2}{N_1} + \frac{(\Sigma X_5)^2}{N_1} - \frac{T^2}{N}$

$= \frac{(127)^2}{4} + \frac{(195)^2}{4} + \frac{(122)^2}{4} + \frac{(90)^2}{4} + \frac{(161)^2}{4} - 24151.25$

$= 25764.75 - 24151.25$

$= 1613.50$

Step 6 : $SSR = \frac{(\Sigma Y_1)^2}{N_2} + \frac{(\Sigma Y_2)^2}{N_2} + \frac{(\Sigma Y_3)^2}{N_2} + \frac{(\Sigma Y_4)^2}{N_2} - \frac{T^2}{N}$

$= \frac{(-31)^2}{5} + \frac{(105)^2}{5} + \frac{(157)^2}{5} + \frac{(464)^2}{5} - 24151.25$

$$= 50386.20 - 24151.25$$

$$= 26234.95$$

$$SSE = TSS - SSC - SSR$$

$$= 37793.75 - 1613.50 - 26234.95$$

$$= 9945.3$$

Step 7 : ANOVA Table

| Source of Variation | SS | DF | MSS | VR | Table value at 5% level |
|---|---|---|---|---|---|
| Columns treatment | SSC = 1613.50 | $c-1=5-1=4$ | $MSC=\frac{SSC}{c-1}$ $=403.375$ | $F_C=\frac{MSE}{MSC}$ $=2.055$ | $F_C(12,4)$ $=5.91$ |
| Between rows | SSR= 26234.95 | $r-1=$ $4-1=3$ | $MSR=\frac{SSR}{r-1}$ $=8744.98$ | $F_R=\frac{MSR}{MSE}$ $=10.552$ | $F_R(3,12)$ $=3.49$ |
| Error | SSE=9945.3 | $N-c-r+1$ $=20-5-4+1$ $=12$ | $MSE=\frac{SSE}{12}$ $=828.775$ | | |

Step 8 : Conclusion : Cal $F_C <$ Table $F_C$ , so accept $H_0$

Cal $F_R >$ Table $F_R$ , so reject $H_0$

The following data represent the number of units production per day turned out by different workers, using 4 different types of machines.

| | | Machine type | | | |
|---|---|---|---|---|---|
| | | A | B | C | D |
| Workers | 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 |

Test whether the five men differ with respect to mean productivity and test whether the mean productivity is the same for the four different machine types. [A.U. M/J 2006, N/D 2007, M/J 2013] [N/D 2010, A/M 2011]

**Solution :** The coded data is,

| Workers | Machine Type | | | | Total | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | | $X_1^2$ | $X_2^2$ | $X_3^2$ | $X_4^2$ |
| $Y_1$ | 4 | −2 | 7 | −4 | 5 | 16 | 4 | 49 | 16 |
| $Y_2$ | 6 | 0 | 12 | 3 | 21 | 36 | 0 | 144 | 9 |
| $Y_3$ | −6 | −4 | 4 | −8 | −14 | 36 | 16 | 16 | 64 |
| $Y_4$ | 3 | −2 | 6 | −7 | 0 | 9 | 4 | 36 | 49 |
| $Y_5$ | −2 | 2 | 9 | −1 | 8 | 4 | 4 | 81 | 1 |
| Total | 5 | −6 | 38 | −17 | T = 20 | 101 | 28 | 326 | 139 |

$H_0$ : (i) the mean productivity is the same for four different machines and

(ii) the 5 men do not differ with respect to mean productivity, Code the data by subtracting 40 from each value.

**Step 1 :** $N = 20,$

**Step 2 :** $T = 20$

**Step 3 :** Correction factor $= \dfrac{T^2}{N} = \dfrac{400}{20} = 20$

**Step 4 :** $\text{TSS} = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2 - \dfrac{T^2}{N}$

$$= 101 + 28 + 326 + 139 - 20$$

$$= 574.$$

**Step 5 :** $\text{SSC} = \dfrac{(\Sigma X_1)^2}{N_1} + \dfrac{(\Sigma X_2)^2}{N_1} + \dfrac{(\Sigma X_3)^2}{N_1} + \dfrac{(\Sigma X_4)^2}{N_1} - \dfrac{T^2}{N}$

[$N_1$ = Number of elements in each column]

$$= \frac{5^2}{5} + \frac{(-6)^2}{5} + \frac{(38)^2}{5} + \frac{(-17)^2}{5} - \frac{T^2}{N}$$

$$= 5 + 7.2 + 288.8 + 57.8 - 20 = 338.8$$

**Step 6 :** $SSR = \frac{(\Sigma Y_1)^2}{N_2} + \frac{(\Sigma Y_2)^2}{N_2} + \frac{(\Sigma Y_3)^2}{N_2} + \frac{(\Sigma Y_4)^2}{N_2} + \frac{(\Sigma Y_5)^2}{N_2} - \frac{T^2}{N}$

[$N_2$ = Number of elements in each row]

$$= \frac{5^2}{4} + \frac{(21)^2}{4} + \frac{(-14)^2}{4} + \frac{0^2}{4} + \frac{8^2}{4} - \frac{T^2}{N}$$

$$= \frac{25}{4} + \frac{441}{4} + \frac{196}{4} + \frac{64}{4} - 20$$

$$= 6.25 + 110.25 + 49 + 16 - 20$$

$$= 181.5 - 20 = 161.5$$

$$SSE = TSS - SSC - SSR$$

$$= 574 - 338.8 - 161.5 = 73.7$$

**Step 7 :** ANOVA Table

| Source of Variation | SS | DF | MSS | VR | Table value at 5% level |
|---|---|---|---|---|---|
| Between columns | SSC $= 338.8$ | $c-1$ $= 4-1$ $= 3$ | $MSC = \frac{SSC}{c-1}$ $= \frac{338.8}{3}$ $= 112.933$ | $F_C = \frac{MSC}{MSE}$ $= \frac{112.933}{6.142}$ $= 18.38$ | $F_C(3, 12)$ $= 3.49$ |
| Between rows | SSR $= 161.5$ | $r-1$ $= 5-1$ $= 4$ | $MSR = \frac{SSR}{r-1}$ $\frac{161.5}{4} = 40.375$ | $F_R = \frac{MSR}{MSE}$ $= \frac{40.375}{6.142}$ $= 6.574$ | $F_R(4,12)$ $= 3.26$ |
| Residual | SSE $= 73.7$ | $N-c-r+1$ $= 12$ | $MSE = \frac{SSE}{N-c-r+1}$ $= \frac{73.7}{12} = 6.142$ | | |

**Step 8 : Conclusion :** (i) Table $F_C$ (3, 12) at 5% level = 3.49

Cal $F_C$ > Table $F_C$. Reject $H_0$

(ii) Take $F_R$ (4, 12) = 3.26 at 5% level

Cal $F_R$ > Table $F_R$. Reject $H_0$

∴ The workers differ with respect to mean productivity.

## EXERCISE 2.3 [Two-way classification][RBD]

1. Set up two-way ANOVA table for the data given below without any interpretation.

| Plots of Land | Treatment | | | |
|---|---|---|---|---|
| | A | B | C | D |
| I | 30 | 27 | 31 | 30 |
| II | 35 | 30 | 29 | 31 |
| III | 34 | 32 | 35 | 28 |

Use coding method subtracting 30 from the given numbers.

2. Four different manufacturing processes were tried at three different stations and the average measurements of quality characters of the product by these processes are given in the following table. Perform the analysis of variance of the data and test for the difference between the processes.

| Station | Processor | | | |
|---|---|---|---|---|
| | A | B | C | D |
| I | 7 | 14 | 11 | 11 |
| II | 15 | 16 | 14 | 10 |
| III | 8 | 15 | 10 | 12 |