The critical value of $\chi^2$ = 12.59 for $df = (r-1)(c-1) = (3-1)(4-1) = 6$ and $\alpha = 0.05$. Since the calculated value of $\chi^2$ is greater than its critical, the null hypothesis is rejected. Hence we conclude that the opinion of all age groups is not the same about the new movie.

## Conceptual Questions 11A

1. Describe the nature of a situation that gives rise to the multinomial distribution.
2. Consider a multinomial experiment in which the outcomes are classified into $n$ categories. Explain why there are $n-1$ degrees of freedom when using the chi-square goodness-of-fit procedure.
3. In the application of the chi-square procedure, describe how the expected frequencies are determined.
4. What is the $\chi^2$-test? Under what conditions is it applicable? Point out its role in business decision-making. [*Kumaon Univ., MBA, 2000*]
5. Describe the $\chi^2$-test of significance and state the various uses to which it can be put.
6. What is the $\chi^2$-test of goodness-of-fit? What cautions are necessary while applying this test? [*Sukhadia Univ., MBA, 1998*]
7. Explain Yate's method of correction for small frequencies in a contingency table.
8. Explain how the $\chi^2$-test is used in the test of homogeneity.
9. Under what conditions should the $\chi^2$-test of independence be used?
10. What are the advantages and limitations of the chi-square test of association? What is the general rule governing the applicability of the chi-square test?
11. What are similarities and differences between the $z$-test the and chi-square test?
12. Write a short note on each of the following:
    (a) Chi-square statistic
    (b) Critical value of chi-square
    (c) Chi-square distribution
    (d) Degrees of freedom

## 11.6 THE SIGN TEST FOR PAIRED DATA

This test is also known as **paired-sample sign test** and based upon the sign of difference in paired observations, say $(x, y)$ where $x$ is the value of an observation from population 1 and $y$ is the value of an observation from population 2. This test assumes that the pairs $(x, y)$ of values are independent and that the measurement scale within each pair is at least ordinal.

For comparing two populations, the sign test is stated in terms of probability that values from population 1 are greater than values from population 2, that are paired. The direction of the difference in values whether plus (+) or minus (–) is also recorded for each pair.

The probability ($p$) that a value $x$ from population 1 will be greater than $y$, a value from population 2, is denoted by $p = P(x > y)$. Every pair of values in a sample is written as $(x, y)$, where $x > y$ is denoted by a plus (+) sign and $x < y$ is denoted by a minus (–) sign. Possibility that $x = y$ is ignored and is denoted by 0 (zero).

Since in the sign test ordering rather than actual measurements are involved, this test is acceptable when distribution is not symmetrical, i.e. when the mean would not be an appropriate measure of centerality.

### Null and Alternative Hypotheses

In sign test, the null hypothesis, $H_0$ is stated that the probability of a plus (+) sign is equal to the probability of a minus (–) sign and both are 0.50, i.e. there is no difference between two populations. This situation is very similar to the fair-coin toss, where we use binomial distribution to describe sampling distribution. The possible hypotheses for the sign test are stated as follows:

| One-tailed Test | Two-tailed Test |
|---|---|
| • Right-tailed Test<br>$H_0 : p \le 0.50$<br>$H_1 : p > 0.5$<br>• Left-tailed Test<br>$H_0 : p \ge 0.5$<br>$H_1 : p < 0.5$ | • $H_0 : p = 0.50 \leftarrow$ No difference between two types of events<br>$H_1 : p \neq 0.50 \leftarrow$ There is a difference between two types of events. |

The test-statistic to test the null hypothesis is defined as:

$$T = \text{Number of plus signs}$$

Since only two signs are considered and the probability ($p$) of getting a plus (+) sign is same as probability ($1 - p$) of getting a minus (–) sign, i.e. 0.50, therefore binomial distribution properties can be used to calculate expected number of plus (+) signs and possible variance. The binomial probability distribution can be approximated to a normal distribution if the conditions:

$$np \ge 5 \quad \text{and} \quad n(1-p) \ge 5$$

are satisfied. We can apply $z$-test statistic to test the null hypothesis $H_0 : p = 0.50$, so that

$$z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}$$

where $x$ is the number of possible signs observed.

**Decission rule:** If calculated value, $z_{cal}$ of $z$-test statistic is less than its critical value, accept the null hypothesis. Otherwise reject $H_0$.

**Remark.** The sign test can also be used to test the hypothesis that the median difference between two populations is zero. The null and alternative hypotheses are stated as:

$$H_0 : \text{Population median} = A, \text{ and } H_1 : \text{Population median} \neq A$$

where $A$ is some number.

To conduct the sign test, observations are paired with the null hypothesis value of the median, and rest of the procedure remains same as discussed before.

**Example 11.15:** The nutritionists and medical doctors have always believed that vitamin C is highly effective in reducing the incidents of cold. To test this belief, a random sample of 13 persons is selected and they are given large daily doses of vitamin C under medical supervision over a period of one year. The number of persons who catch cold during the year is recorded and a comparison is made with the number of cold contacted by each such person during the previous year. This comparison is recorded as follows, along with the sign of the change.

| Observations | : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Without vitamin C | : | 7 | 5 | 2 | 3 | 8 | 2 | 4 | 4 | 3 | 7 | 6 | 2 | 10 |
| With vitamin C | : | 2 | 1 | 0 | 1 | 3 | 2 | 3 | 5 | 1 | 4 | 4 | 3 | 4 |
| Sign | : | – | – | – | – | – | 0 | – | + | – | – | – | + | – |

Using the sign test at $\alpha = 0.05$ level of significance, test whether vitamin C is effective in reducing colds.

**Solution:** Let us take the null hypothesis that there is no difference in the number of cold contacted with or without vitamin C and this probability ($p$) = 0.50, i.e.

$$H_0 : p = 50 \quad \text{and} \quad H_1 : p \neq 0.50$$

Given $n = 12$ (because differnece is zero in observation 6); number of minus signs = 10 and number of plus signs = 2.Thus

$$\mu = np = 12\,(0.5) = 6 \text{ and } \sigma = \sqrt{npq} = \sqrt{12 \times 0.5 \times 0.5} = \sqrt{3} = 1.73$$

Applying the $z$-test statistic; we get

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{9.5 - 6}{1.73} = 2.02$$

where, $\bar{x} = 10 \cong 9.5$ because of approximation from discrete distribution to normal distribution.

Since $z_{cal}$ (= 2.02) is greater than the critical value $z_\alpha$(= 1.96) at $\alpha$ = 0.05 level of significance, $H_0$ is rejected. Hence we conclude that there is a significant difference in the number of cold contacted with or without vitamin C.

**Example 11.16:** The median age of tourists who has come to India is claimed to be 40 years.A random sample of 18 tourists gives the following ages:

24, 18, 37, 51, 56, 38, 45, 45, 29, 48, 39, 26,38, 43, 62, 30, 66, 41

Test the hypothesis using $\alpha$ = 0.05 level of significance.

**Solution:** Let us take the null and alternative hypotheses as stated below:

$$H_0: \mu = 40 \quad \text{and} \quad H_1: \mu \neq 40$$

Arranging data on ages of tourists in an ascending order and making pair of those with median age 40 years to determine plus (+) sign and minus (–) sign as follows:

| Tourist Age : | 18 | 24 | 26 | 29 | 30 | 37 | 38 | 38 | 39 | 41 | 43 | 45 | 45 | 48 | 51 | 56 | 62 | 66 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Median age : | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |
| Sign : | + | + | + | + | + | + | + | + | + | – | – | – | – | – | – | – | – | – |

Given $n$=18, number of plus (+) sign, $x$=9, and $p$=0.5. Thus

$$\mu = np = 18(0.5) = 9, \text{ and } \sigma = \sqrt{npq} = \sqrt{18 \times 0.5 \times 0.5} = 1.060$$

Applying the $z$-test statistics, we get

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{8.5 - 9}{1.060} = -1.533$$

where, $\bar{x} = 9 \cong 8.5$ because of approximation form discrete distribution to normal distribution.

Since $z_{cal}$(= – 1.533) is greater than its critical value $z_\alpha$ = – 1.96 at $\alpha$ = 0.05 level of significance, $H_0$ is accepted. Hence, we conclude that the claim is correct.

## 11.7 RUNS TEST FOR RANDOMNESS

The randomness of the sample drawn from a population is essential for all types of statistical testing because sample results are to be used to draw conclusions regarding the population under study. The *run test* helps to determine whehter the order or sequence of observations (symbols, items or number) in a sample is random. The runs test examines the number of 'runs' of each of two possible characteristics that sample elements may have. *A run is a sequence of indentical occurrences of elements (symbols or numbers) preceded and followed by different occurrences of elements or by no element at all.* For example, in tossing a coin, the outcome of three tails in succession would constitute a run, as would a succession of five heads. To quantify how many runs are acceptable before raising doubt about the randomness of the process, a probability distribution is used that leads to a *statistical test for randomness*.

Suppose that, tossing a coin 20 times produces the following sequence of heads (H) and tails (T).

| HHH | TT | HH | TTT | HHH | TT | HHH | T |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

In this example, the first run of HHH is considered as run 1, the second run of *TT* as run 2 and so on, so that there are 8 runs in all or in other words, $r$ = 8. However, in this exmaple, rather than perfect separation between H and T, it appears to be a perfect clustering together. It is a form of regularity not likely to have arisen by chance.

### Small Sample Run Test

In order to test the randomness, let $n_1$ = number of elements of one kind, and $n_2$ = number of elements of second kind. Total sample size is $n = n_1 + n_2$. In the above example, $n_1$ = 12 heads and $n_2$ = 8 tails. Let one kind of elements be denoted by plus (+) sign and second kind of elements be denoted by minus (–) sign. In the above example, *H* is represented by plus (+) sign and T by minus (–) sign. The concept of plus (+) or minus (–) provides the direction of change from an establish pattern. Accordingly, a plus (+) would be considered a change from an establish pattern value in one direction and a minus (–) would be considered a change in the other direction.

If the sample size is small, so that either $n_1$ or $n_2$ is less than 20, then test is carried out by comparing the deserved number of runs R to critical values of runs for the given values of $n_1$ and $n_2$. The critical values of R are given in Appendix. The null and alternative hypotheses stated as:

$H_0$ : Observations in the samples are randomly generated

$H_1$ : Observations in the samples are not randomly generated

can be tested that the occurrences of plus (+) signs and minus (–) signs are random by comparing $r$ value with its critical value at a particular level of significance. *Decision rule*

**Decision rule:** • Reject $H_0$ at a if $R \leq C_1$ or $R \geq C_2$

• Otherwise accept $H_0$

where $C_1$ and $C_2$ are critical values obtained from standard table with total tail probability $P(R \leq C_1) + P(R \geq C_2) = \alpha$.

**Large Sample Run Test**

If the sample size is large so that either $n_1$ or $n_2$ is more than 20, then the sampling distribution of R statistic (i.e. run) can be closely approximated by the normal distribution. The mean and standard deviation of the number of runs for the normal distribution are given by

$$\text{Mean, } \mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$\text{Standard devication, } \sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

Thus the standard normal test statistic is given by

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{r - \mu_R}{\sigma_R}$$

The critical $z$-value is obtained in the usual manner at a specified level of significance.

**Example 11.17:** A stock broker is interested to know whether the daily movement of a particular share averages in the stock market showed a pattern of movement or whether these movements were purely random.For 14 business days, he noted the value of this average and compared it with the value at the close of the previous day. He noted the increase as plus (+) and decrease as minus (–). The record was as follows:

+, +, – , –, +, +, +, –, +, +, –, +, –, –

Test whether the distribution of these movements is random or not at $\alpha = 0.05$ level of significance.

**Solution:** Let us state the null and alternative hypotheses as follows:

$H_0$ : Movement is random and $H_1$: Movement is not random

There are $r = 8$ runs (plus signs) with $n_1 = 8$ plus (number of increases) and $n_2 = 6$ minus (number of decreases) so that $n = n_1 + n_2 = 14$. The critical vlaue of $r = 8$ for $n_1 = 8$ and $n_2 = 6$, implies that $H_0$ is rejected when $r \leq 3$ and $r \geq 12$ at $\alpha = 0.05$ level of significance. Since $3 \leq r \leq 12$, therefore $H_0$ cannot be rejected.

**Example 11.18:** Some items produced by a machine are defective. If the machine follows some pattern where defective items are not randomly produced throughout the process the machine needs to be adjusted. A quality control engineer wants to determine whether the sequence of defective ($D$) versus good ($G$) items is random. The data are

GGGGG, DDD, GGGGGG, DDD, GGGGGGGGGG, DDDD,

GGGGGGGGGGG, DDD, GGGGGGGGGGG, DDDD

Test whetehr the distribution of defective and good items is random or not at $\alpha = 0.05$ level of significance.

**Solution:** Let us state the null and alternative hypotheses as follows:

$H_0$ : Sequence is random, and $H_1$ : Sequence is not random

There are $r = 10$ runs with $n_1 = 43$ (number of G) and $n_2 = 17$ (number of D), so that $n = n_1 + n_2 = 60$. Since sample size is large, we compute mean and standard deviation as follows:

$$\text{Mean, } \mu_R = \frac{2n_1n_2}{n_1+n_2}+1 = \frac{2\times 43\times 17}{43+17}+1 = 24.8$$

$$\text{Standard deviation, } \sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$

$$= \sqrt{\frac{2(43)(17)[2(43)(17)-43-17]}{(43+17)^2(43+17-1)}}$$

$$= \sqrt{\frac{1462\times 1402}{3600\times 59}} = \sqrt{\frac{20,49,724}{2,12,400}} = 3.106$$

The computed value of z-test statistic is

$$z = \frac{R-\mu_R}{\sigma_R} = \frac{10-24.8}{3.106} = -7.871$$

The critical value of $z_\alpha = 2.58$ for two-tailed test at $\alpha = 0.05$ level of significance. Since $z_\alpha < z_{cal}$, $H_0$ is rejected. Hence, we conclude that the sequence of defective versus good item is not random.

## 11.8 MANN-WHITNEY U-TEST

If sample size is small and we cannot or do not wish to make the assumption that the data is taken from normally distributed population, then Mann-Whitney U-test or simply U-test is used to test the equality of two population means. This test is the substitute for t-test statistic when the stringent assumtions of parent population being normally distributed with equal variance are not met or when the data are only ordinal in measurement.

To apply the U-test, also called the **Wilcoxon rank sum test** or simply the **Rank Sum Test** values from two samples are combined into one group and are arranged in ascending order. The pooled values are then ranked from 1 to $n$ with smallest value being assigned a rank 1. Such a test produces good results when data are on ordinal scale of measurement. The sum of ranks of values from sample 1 is denoted as $R_1$. Similarly the sum of the ranks of values from sample 2 is denoted as $R_2$. If both $n_1$ and $n_2 < 30$, samples are considered small.If either $n_1$ or $n_2$ is greater than 10, the samples are considered large.

### Small Sample U-Test

1. Combine the two random samples of size $n_1$ and $n_2$ and rank values from smallest to largest. If several values are tied, then assign each the average of the ranks that would otherwise have been assigned.

   If the two sample sizes are unequal, then suppose $n_1$ represent smaller-sized smaple and $n_2$ the larger-sized sample. The rank sum test statistic $U_1$ is the sum of the ranks assigned to the $n_1$ observations in the smaller sample. However, for equal-sized samples, either group may be selected for determining $U_1$.

   The test statistic $U_1$ plus the sum of the ranks assigned to the $n_2$ observations in the larger sample:

   $$U_1 + U_2 = \frac{n(n+1)}{2}$$

   represents the sum of first consecutive integers.

2. State the null and alternative hypotheses for U-test as follows:

   $H_0 : u_1 = u_2 \leftarrow$ Two populations distribution have equal mean

   $H_1 : u_1 \neq u_2 \leftarrow$ Two populations distribution have different means

   The test of the null hypothesis can either be two-tailed or one-tailed.

3. The value of U-statistic is the smallest of the following two U-values computed as follows:

   $$U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

The U-statistic is a measure of the difference between the ranked observations of the two samples. Large or small values of statistic provide evidence of a diffference between two populations. If differences between populations are only in location, then large or small values of U-statistic provide evidence of a difference in location (median) of two populations.

Both $U_1$ and $U_2$ need not be caluclated, instead one of $U_1$ or $U_2$ can be calculated and other can be formed by using the equation: $U_1 = n_1 n_2 - U_2$.

**Large Sample U-Test**

For large samples (i.e. when both $n_1$ and $n_2$ are greater than 10) the sampling distribution of the U-statistic can be approximated by the normal distribution so that z-test statistic is given by

$$z = \frac{U - \mu_U}{\sigma_U}$$

where Mean $\mu_U = \frac{n_1 n_2}{2}$ and standard deviation, $\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$

**Decision rules**

- When $n_1$ and $n_2$ are both less than or equal to 10, standard table value can be used to obtain the critical value of the test statistic for both one and two tailed test at a specified level of significance.
- For large sample, at a specified level of significance.
  - Reject $H_0$, if computed vlaue of $z_{cal} \geq$ critical value $z_\alpha$
  - Otherwise accept $H_0$

**Example 11.19:** It is generally believed that as people grow older, they find it harder to go to sleep. To test if there was a difference in time in minutes before people actually went to sleep after lying in the bed, a sample of 10 young persons (ages 21 to 25) and 10 old persons (ages 65 to 70) was randomly selected and this sleeping habits were monitored. The data show the number of minutes these 20 persons were awake in bed before getting to sleep:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Young men : | 58 | 42 | 68 | 20 | 15 | 35 | 26 | 40 | 47 | 28 |
| Old men : | 100 | 152 | 147 | 70 | 40 | 95 | 68 | 90 | 112 | 58 |

Is there evidence that young men are significantly take more time to get to sleep than old men. Use $\alpha = 0.05$ level of significance.

**Solution:** First arrange the data in ascending order for ranking as shown below. When ties occur, we assign to each tied observation the average rank of the ties.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Young men : | 15 | 20 | 26 | 28 | 35 | 40 | 42 | 47 | 58 | 68 |
| Old men : | – | – | – | – | – | 40 | – | – | 58 | 68 |
| | 70 | 90 | 95 | 100 | 112 | 147 | 152 | – | – | – |

List the ranks of all the observations in each of the two groups as follows:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Young men : | 1 | 2 | 3 | 4 | 5 | 6.5 | 8 | 9 | 10.5 | 12.5 | = 64.5 |
| Old mewn : | 6.5 | 10.5 | 12.5 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | = 148.5 |

Let us define young men as population 1 and old men as population 2. The rank sum are $R_1 = 64.5$ and $R_2 = 148.5$. Computing the test statistic U, we get

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 10 \times 10 + \frac{10 \times 11}{2} - 64.5$$

$$= 100 + 55 - 64.5 = 90.5$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 10 \times 10 + \frac{10 \times 11}{2} - 148.5$$

$$= 100 + 55 - 148.5 = 6.5$$

Consider the lower value between $U_1$ and $U_2$, i.e. $U_2 = 6.5$, so that

$$\text{Mean}, \mu_U = \frac{n_1 n_2}{2} = \frac{10\times 10}{2} = 50$$

and $$\text{Standard deviation}, \sigma_U = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{10\times 10(10+10+1)}{12}}$$

$$= \sqrt{\frac{100\times 21}{12}} = \sqrt{\frac{2100}{12}} = 13.23.$$

Since both $n_1$ and $n_2$ are greater than 8, applying z-test statistic so that

$$z = \frac{U-\mu_U}{\sigma_U} = \frac{6.5-50}{13.23} = -3.287$$

Since computed value of $z_{cal}$ (= – 3.287) is less than its critical value $z_\alpha$ = – 1.96 at $\alpha$ = 0.05 significance level, null hypothesis to rejected. Hence we conclude that there is a significant difference in time to get to sleep between young men and old men.

## 11.9 WILCOXON MATCHED PAIRS TEST

This test is also known as **Wilcoxon signed rank test.** When two samples are related, the U-test discussed before is not applicable. Wilcoxon test is a non-parametric test alternative to t-test for two related samples. In U-test the differences between paired observations is not believed to be normally distributed and also ignore the size of magnitude of these differences. Wilcoxon test takes into consideration both the direction and magnitude of differences between paired values.

**Procedure**

1. Compute differences in the same manner as in the case of the U-test. Then assign ranks for 1 to $n$ to the absolute values of the differences starting from the smallest to largest defferences. All pairs of values with zero differences are ignored. If differences are equal in magnitude, a rank equal to the average of ranks that would have been assigned otherwise is given to all the equal differences.
2. Take the sum of the ranks of the positive and of the negative differences. The sum of positive and negative differences is denoted by $s_+$ and $s_-$ respectively.
3. Define Wilcoxon T-statistic as the smallest sum of ranks (either $s_+$ or $s_-$):

   $$T = \min(s_+, s_-) = s$$

   Since values of $s_+$, $s_-$ and $s$ may vary in repeated sampling, these sums for a given sample may be treated as specific values of their respective sample statistic.

   When the number of pairs of values, $n$ is more than 15, the value of $T$ is approximately normally distributed and z-test statistic is used to test the null hypothesis.
4. Define null and alternative hypotheses as follows:

| *One-tailed Test* | *Two-tailed Test* |
|---|---|
| $H_0 : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = d_0$ | $H_0 : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = d_0$ |
| $H_1 : \mu_1 > \mu_2$ or $\mu_1 - \mu_2 > d_0$ | $H_1 : \mu_1 \neq \mu_2$ or $\mu_1 - \mu_2 \neq d_0$ |

5. See table value of $s_\alpha$ for different sample size $n$ and a specified level of significance $\alpha$ for $s_+$, $s_-$ and $s_\alpha$. Decision rules for one-tailed and two-tailed test are as follows:

| *One-tailed Test* | *Two-tailed Test* |
|---|---|
| • Reject $H_0$ when $s_- < s_\alpha$ or $s_+ < s_\alpha$ | • Reject $H_0$ when $s < s_\alpha$ |
| • Otherwise accept $H_0$ | • Otherwise accept $H_0$ |

**Remarks**

1. If the sample size is small, i.e. $n \leq 15$ (number of pairs), then a critical value against which to compare T can be found by using $n$ and $\alpha$. If calculated value of T is less than or equal to the critical value of T at $\alpha$ level of significance and sample size $n$, then $H_0$ is rejected.

2. If the sample size $n$ (> 15), then the sampling distribution of T (i.e. $s_+$ and $s_-$) approaches normal distribution with

$$\text{Mean } \mu_T = \frac{n(n+1)}{4}$$

and Standard deviation, $\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$

Given the level of significance $\alpha$, z-test statistic is computed as:

$$z = \frac{T - \mu_T}{\sigma_T}$$

where $n$ = number of pairs.

**Example 11.20:** Ten workers were given on-the-job training with a view to shorten their assembly time for a certain mechanism. The results of the time (in minutes) and motion studies before and after the training programme are given below:

| Worker : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Before : | 61 | 62 | 55 | 62 | 59 | 74 | 62 | 57 | 64 | 62 |
| After : | 59 | 63 | 52 | 54 | 59 | 70 | 67 | 65 | 59 | 71 |

Is there evidence that the training programme has shortened the average assembly time?

**Solution:** Let us take the null hypothesis that the training programme has helped in reducing the average assembly time. Calculations to compute Wilcoxon T-statistic are shown below:

**Table 11.2: Computation of Wilcoxon T-statistic.**

| Worker | Before the Training ($x_1$) | After the Training ($x_2$) | Difference $d = x_1 - x_2$ | Absolute Rank | Signed Rank Positive ($s_+$) | Negatie ($s_-$) |
|---|---|---|---|---|---|---|
| 1 | 61 | 59 | + 2 | 2 | 2 | – |
| 2 | 62 | 63 | – 1 | 1 | – | 1 |
| 3 | 55 | 52 | + 3 | 3 | 3 | – |
| 4 | 62 | 54 | + 8 | 7 | 7 | – |
| 5 | 59 | 59 | 0 | Ignored | – | – |
| 6 | 74 | 70 | + 4 | 4 | 4 | – |
| 7 | 62 | 67 | – 5 | 5.5 | – | 5.5 |
| 8 | 57 | 45 | + 12 | 9 | 9 | – |
| 9 | 64 | 59 | + 5 | 5.5 | 5.5 | – |
| 10 | 62 | 71 | – 9 | 8 | – | 8 |
| | | | | | $\Sigma s_+ = 30.5$ | $\Sigma s_- = 14.5$ |

Since the smaller sum is associated with the negative ranks, value of the Wilcoxon test statistic is: $T = \Sigma s_- = 14.5$. We compare the computed value of T with its critical value $s = 8$ at $n = 10$ and $\alpha = 0.05$ level of significance (See Appendix). Since computed value of T is more than its critical value, null hypothesis is accepted and we conclude that the training has not helped in raducing the average assembly time.

**Alternative approach**

$$\mu_T = \frac{n(n+1)}{4} = \frac{10 \times 11}{4} = 27.5$$

and $$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{9.81}{24}} = 9.81$$

Applying z-test statistic, we get

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{14.5 - 27.5}{9.81} = -1.325$$

Since computed value of $z$ (= – 1.325) is less than the critical value $z_\alpha$(= ± 1.96) at $\alpha = 0.05$ level of significance, null hypothesis is accepted.

**Example 11.21:** The average score on a vocatinal training test has been known to be 64. Recently, several changes have been carried out in the programme; the effect of these changes on performance on the test is unknown. It is therefore desirable to test the null hypothesis that the average score for all people who will complete the programme will be 64 versus the alternative that it will not be 64. The following random sample of scores is available

87, 91 65, 31, 8, 53, 99, 44, 42, 60, 77, 73, 42, 50
79, 90, 54, 39, 77, 60, 33, 41, 42, 85, 71, 50

Is there evidence that the average score for all people who will complete the programme will be 64?

**Solution:** Let us take the null and alternative hypotheses as:

$$H_0 : \mu = 64 \quad \text{and} \quad H_1 : \mu \neq 64$$

Subtracting the average score of 64 (the null hypothesis mean) from every data point to form pairs. This gives the differences:

$87 - 64 = +23$, $91 - 64 = +27$, $65 - 64 = +1$, $31 - 64 = 33$, $8 - 64 = -56$,
$53 - 64 = -11$, $99 - 64 = +35$, $44 - 64 = -20$, $42 - 64 = -22$, $60 - 64 = -4$,
$77 - 60 = +17$, $73 - 64 = -9$, $42 - 64 = -22$, $50 - 64 = -14$, $79 - 64 = +15$,
$90 - 64 = +26$, $54 - 64 = -10$, $39 - 64 = -25$ $77 - 64 = -13$, $60 - 64 = -4$,
$33 - 64 = -31$, $41 - 64 = -23$, $42 - 64 = -22$, $85 - 64 = +21$ $71 - 64 = +7$,
$50 - 64 = 14$.

Ranking the absolute value of the differences from smallest to largest, we get

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Difference : | +23 | +27 | +1 | −33 | −56 | −11 | +35 | −20 | −22 | −4 | +17 | +9 | −22 |
| Rank : | 18.5 | 22 | 1 | 24 | 26 | 7 | 25 | 13 | 16 | 2.5 | 12 | 5 | 16 |
| Difference : | +4 | +15 | +26 | −10 | −25 | 13 | −4 | −31 | −23 | −22 | +21 | +7 | −14 |
| Rank : | 9.5 | 11 | 21 | 6 | 20 | 8 | 2.5 | 23 | 18.5 | 16 | 14 | 4 | 9.5 |
| $\Sigma s_+$ : | 18.5 + 22 + 1 + 25 + 12 + 5 + 11 + 21 + 8 + 14 + 4 = 141.5 | | | | | | | | | | | | |
| $\Sigma s_-$ : | 24 + 26 + 7 + 13 + 16 + 2.5 + 16 + 9.5 + 6 + 20 + 2.5 + 23 + 18.5 + 16 + 9.5 = 209.5 | | | | | | | | | | | | |

Since the smaller sum is associated with positive ranks, we define T as that sum, i.e. T = 141.5.

Consider the sampling distribution of T-statistic as normal so that

$$\mu_T = \frac{n(n+1)}{4} = \frac{26 \times 27}{4} = 175.50$$

and
$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{26 \times 27 \times 53}{24}} = \sqrt{\frac{37206}{24}} = 39.37$$

Applying $z$-test statistic, we get

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{141.5 - 175.50}{39.37} = -0.863$$

Since computed value of $z_{cal}$ (= − 0.863) is less than the critical value $z_\alpha = \pm 1.96$ at $\alpha = 0.05$ level of significance, null hypothesis is accepted. Hence, we conclude that average score for all people who will complete the programme will be 64.

## 11.10 KRUSKAL-WALLIS TEST

This test is the non-parametric alternative to the one-way analysis of variance to identify differences among populations that does not require any assumption about the shape of the population distribution. This test uses the ranks of the observations rather than the data themselves with the assumption that the observations are on an interval scale.

**Procedure:** This test is used for comparing $k$ different populations having identical distribution. The summary of procedure is as follows:

1. Draw $k$ independent samples $n_1, n_2, ..., n_k$ from each of $k$ different populations. Then combine these samples such that $n = n_1 + n_2 + ... + n_k$, and arrange these $n$ observations in an ascending order.
2. Assign ranks to observations from 1 to $n$ such that smallest value is assigned rank 1. For ties each value is assigned average rank.
3. Identify rank values whether they belong to samples of size $n_1, n_2, ..., n_k$. The ranks corresponding to different samples are totaled and the sum for the respective sample is denoted as $t_1, t_2, ..., t_k$.

   The following formula is used to compute Kruskal Wallis H-statistic:

$$H = \frac{12}{n(n+1)}\left[\sum_{j=1}^{k}\frac{t_j^2}{n_j}\right] - 3(n+1)$$

This $H$ value is approximately Chi-square distributed with $k - 1$ degrees of freedom as long as $n_j$ is not less than 5 observations for any population

However, if one or more samples have two or more equal observations, the value of H is adjusted as : $H' = H/C$, where C is the correction factor defined as:

$$C = 1 - \frac{1}{n^3 - n}\left[\sum_{j=1}^{r}(O_j^3 - O_j)\right]$$

where $O_j$ = the number of equal observations in the $j$th sample.

$r$ = the number of samples which has equal observations

The null and alternative hypotheses are stated as:

$H_0$ : The $k$ different populations have identical distribution

$H_1$ : At least one of the $k$ populations has different distribution.

### Decision Rule

- Reject $H_0$ when computed value of H is greater than $\chi^2$ (Chi-square) at $df = k - 1$ and $\alpha$ level of significance
- Otherwise acept $H_0$.

**Example 11.22:** Use Kruskal-Wallis test to determine whether there is a significant difference in the following populations. Use $\alpha = 0.05$ level of significance

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Populatin 1 : | 17 | 19 | 27 | 20 | 35 | 40 | |
| Pupulatin 2 : | 28 | 36 | 33 | 22 | 27 | | |
| Population 3 : | 37 | 30 | 39 | 42 | 28 | 25 | 31 |

**Solution:** Three populations are considered for study, so $k = 3$ and $n = 18$. The observations in three populations are combined and ranked. The smallest value is given rank 1, as shown below:

| *Populatin 1* | | *Populations 2* | | *Population 3* | |
|---|---|---|---|---|---|
| *Value* | *Rank* | *Value* | *Rank* | *Value* | *Rank* |
| 17 | 1 | 22 | 4 | 25 | 5 |
| 19 | 2 | 27 | 6.5 | 28 | 8.5 |
| 20 | 3 | 28 | 8.5 | 30 | 10 |
| 27 | 6.5 | 33 | 12 | 31 | 11 |
| 35 | 13 | 36 | 14 | 37 | 15 |
| 40 | 17 | $n_2 = 5$ | $t_2 = 45$ | 39 | 16 |
| $n_1 = 6$ | $t_1 = 42.5$ | | | 42 | 18 |
| | | | | $n_3 = 7$ | $t_3 = 83.5$ |

Suppose $H_0$ : Three populations are identical, i.e. $\mu_1 = \mu_2 = \mu_3$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

Kruskal-Wallis test statistic is

$$H' = \frac{H}{C},$$

where $H = \frac{12}{n(n+1)}\left[\sum_{j=1}^{3} (t_j^2/n_j)\right] - 3(n+1)$

$= \frac{12}{18\times 19}\left[\frac{(42.5)^2}{6} + \frac{(45)^2}{5} + \frac{(83.5)^2}{7}\right] - 3\times 19$

$= 0.035[301.4 + 405 + 996.03] - 57 = 2.572$

and $C = 1 - \frac{1}{n^3 - n}\left[\sum_{j=1}^{2} (O_j^3 - O_j)\right]$

$= 1 - \frac{1}{(18)^3 - 18}[\{(2)^3 - 2\} + \{(2)^2 - 2\}]$

$= 1 - \frac{12}{5814} = 1 - 0.002 = 0.998$

Thus $H' = \frac{2.572}{0.998} = 2.577$

Since computed value of H′ (= 2.577) is less than table value of $\chi^2$(= 5.99) at $df = k - 1 = 2$ and $\alpha = 0.05$, the nul hypothesis is accepted and conclude that three populations are identical.

## Self-Practice Problems 11C

**11.18** Pre- and post-test scores after a particular training programme are known to be non-normal in their distribution. A sample of the scores, with the calculated changes, is given below:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Pre-test : | 67 | 71 | 83 | 69 | 68 | 36 | 52 | 72 | 56 |
| | 64 | 76 | 83 | 69 | 68 | 36 | | | |
| | 52 | 72 | 56 | | | | | | |
| Post-test : | 58 | 62 | 84 | 67 | 72 | 38 | 63 | 72 | 55 |
| | 59 | 76 | 84 | 69 | 72 | 38 | | | |
| | 63 | 74 | 66 | | | | | | |

Conduct a sign test for determinig whether any significant change has taken place.

**11.19** The median age of a tourist to India is claimed to be 41 years. A random sample of 18 tourists gives the followign ages:

25, 19, 38, 52, 57, 39, 46, 46, 30 49, 40, 27, 39, 44, 63, 31, 67, 42

Test the hypothesis against a twio-tailed alternative using $\alpha = 0.05$.

**11.20** A courier service employs eight men and nine women. Every day, the assignments of delivery are supposed to be done at random. On a certain day, all the best jobs, in order of desirability, were given to the eight men. Is there evidence of sex discrimination? Discuss this also in the context of a continuing, daily operation. What would heppen if you tested the randomness hypothesis everyday?

**11.21** The owner of a garment shop wants to test whether his two salesman A and B are equally effective. That is, he wants to test whether the number of sales made by each salesman is about the same or whether one salesman is bettter than the other. He gets the following random samples of daily sales made by each salesman.

| | | | | | | |
|---|---|---|---|---|---|---|
| Salesman A : | 35, | 44, | 39, | 50, | 48, | 29, |
| | 60, | 75, | 49, | 66 | | |
| Salesman B : | 17, | 23, | 13, | 24, | 33, | 21, |
| | 18, | 16, | 32 | | | |

Test whether salesperson A and B are equally effective.

**11.22** Suppose 26 cola drinkers are sampled randomly to determine whether they prefer brand A or brand B. The random sample contains 18 drinkers of brand A and 8 drinkers of brand B. Let *C* denotes brand A drinkers and *D* denote brand B drinkers. Suppose the sequence of sampled cola drinkers is DCCCCCDCCDCCCCDCDC CCDDDCCC. Is this sequence of cola drinkers evidence that the sample is not random.

**11.23** A machine produces parts that are occasionally flawed. When the machine is working in adjustment, flaws still occur but seem to happen randomly. A quality control person randomly selects 50 of the parts produced by the machine today and examines them one at a time in the order that they were made. The result is 40 parts with no flaws and 10 parts with flaws. The sequence of no flaws (denoted by N) and flaws (denoted by F) is shown here. Using $\alpha$=0.05 level of significance, the quality controller tests to determine whether the machine is producing randomly (the flaws are occurring randomly).

NNN F NNNNNNN F NN FF NNNNNN F NNNN F NNNNNN FFFF NNNNNNNNNNNN

Is this sequence of flaws and no flaws evidence that the sample is not random

**11.24** A panel of 8 members has been asked about their perception of a product before and after they had an opportunity to try it. Their perceptions, measured on an ordinal scale, gave the results given below:

| Member : | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Before : | 8 | 3 | 6 | 4 | 5 | 7 | 6 | 7 |
| After : | 9 | 4 | 4 | 1 | 6 | 7 | 9 | 2 |

Have the perception scores changed after trying the product? Use $\alpha = 0.05$ level of significance.

**11.25** Samples have been taken from two branches of a chain of stores. The samples relate to the daily turnover of both the branches. Is there any difference in turnover between the two branches.

| | | | | |
|---|---|---|---|---|
| Banrch 1 : | 23500 | 25500 | 35500 | 19500 |
| | 24400 | 24000 | 23600 | 25900 |
| | 26000 | | | |
| Branch 2 : | 24000 | 19800 | 22000 | 21500 |
| | 24500 | | | |

**11.26** The average hourly number of messages transmitted by a communications satellite is believed to be 149. If there is a possibility that demand for this service may be declining, then test the null hypothesis that the average hourly number of relayed messages is 149 (or more) versus the alternative hypothesis that the average hourly number of relayed messages is less than 149. A random sample of 25 operation hours is selected. The data (numbers of messagers relayed per hour) are

151, 144, 123, 178, 105, 112, 140, 167, 177, 185, 129, 160, 110, 170, 198, 165, 109, 118, 155,102, 164, 180, 139, 166, 182

Is there any evidence of declining the use of the satellite?

**11.27** The average life of a 100-watt light bulb is stated on the package to be 750 hours. The quality control manager at the plant making the lightbulbs needs to check whether the statement is correct. The manager is only concerned about a possible reduction in quality and will stop the production process only if statistical evidence exists to conclude that the average life of a lightbulb is less than 750 hours. A random sample of 20 bulbs is collected and left on until they burn out. The lifetime of each bulb is recorded. The data are (in hours of continuous use) 738, 752, 710, 701, 689, 779, 650, 541, 902, 700, 488, 555, 870, 609, 745, 712, 881, 599, 659, 793. Should the process be stopped and corrected? Explain why or why not.

**11.28** An accounting firm wants to find out whether the current ratio for three companies is same. Random samples of eight firms in industry A, six firms in industry B, and six firms in industry C are available. The current ratios are as follows:

| *Company A* | *Company B* | *Company C* |
|---|---|---|
| 1.38 | 2.33 | 1.06 |
| 1.55 | 2.50 | 1.37 |
| 1.90 | 2.79 | 1.09 |
| 2.00 | 3.01 | 1.65 |
| 1.22 | 1.99 | 1.44 |
| 2.11 | 2.45 | 1.11 |
| 1.98 | | |
| 1.61 | | |

Conduct the test at $\alpha = 0.05$ level of significance, and state your conclusion.

**11.29** Results of a survey indicated that people between 55 and 65 years of age contact a physician an average of 9.8 times per year. People of age 66 and older contact doctors on an average of 12.9 times per year. Suppose you want to validate these results by taking your own samples. The following data represent the number of annual contacts people make with a physician.The samples are independent. Apply a suitable test statistic to determine whether the number of contacts with physicians by the people of age 65 years and older is greater than the number by people of age 55 to 65 years

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 55 to 65 : | 12 | 13 | 8 | 11 | 9 | 6 | 11 | | |
| 65 and loder : | 16 | 15 | 10 | 17 | 13 | 12 | 14 | 9 | 13 |

**11.30** Consider the survey that estimated the average annual household spending on healthcare. The metropolitan average was Rs.1800. Suppose six families in Delhi are matched demogrphically with six families in Mumbai and their amounts of household spending on healthcare for last year are obtained. The data are as follow:

| *Family Pair* | *Delhi* | *Mumbai* |
|---|---|---|
| 1 | 1950 | 1760 |
| 2 | 1840 | 1870 |
| 3 | 2015 | 1810 |
| 4 | 1580 | 1660 |
| 5 | 1790 | 1340 |
| 6 | 1925 | 1765 |

Apply a suitable test statistic to determine whether there is a significant difference in annual household healthcare spending between these two cities.

## Hints and Answers

**11.18** $H_0$: No difference between the scores

$H_1$: There is a difference between scores

$s_+ = 10, s_- = 5, s = 5$; Accept $H_0$

**11.19** T = 9, $H_0$ is accepted.

**11.20** $z = -3.756$, $H_0$ is rejected

**11.21** $n_1 = 10$ and $n_2 = 9$. Arrange values in two samples in increasing order and denote them by A or B based on which populatin they come from:

BBBBBBBABBAAAAAAAAA

Total runs R = 4. Salesman A sells more than B.

**11.22** $H_0$ : The observations in the sample are random.

$H_1$ : The observations in the sample are not random.

Tally the number of runs

| D | CCCCC | D | CC | D | CCCC |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| D | C | D | CCC | DDD | CCC |
| 7 | 8 | 9 | 10 | 11 | 12 |

The number of runs, $R = 12$. Since the value of $R$ falls betwen the critical values of 7 and 17, do not reject $H_0$.

**11.23** $H_0$: The observations in the sample and random.

$H_1$: The observations in the sampel are not random.

Apply z-test statistic is

$$\mu_R = R - \left(\frac{2n_1n_2}{n_1+n_2}+1\right) = \frac{2\times 40\times 10}{40+10} + 1 = 17$$

$$\sigma_R = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

$$= \sqrt{\frac{2\times 40\times 10\,(2\times 40\times 10 - 40 - 10)}{(40+10)^2(40+10-1)}} = 2.213$$

$$z = \frac{13-27}{2.213} = -1.81$$

Since $z_{cal}$ (= −1.81) is greater than $z_{\alpha/2} = -1.96$, reject the null hypothesis.

**11.24** $H_0$: There is no difference in perception.

$H_1$: There is a difference in perception.

| *Member* | *Before* | *After* | *Difference* | *Rank* |
|---|---|---|---|---|
| A | 8 | 9 | + 1 | 2 |
| B | 3 | 4 | + 1 | 2 |
| C | 6 | 4 | − 2 | 4 |
| D | 4 | 1 | − 3 | 5.5 |
| E | 5 | 6 | + 1 | 2 |
| F | 7 | 7 | 0 | – |
| G | 6 | 9 | + 3 | 5.5 |
| H | 7 | 2 | − 5 | 7 |

Sum of ranks: $s_+ = 11.5$; $s_- = 16.5$; Consider the smaller test statistic, $s_+ = 11.5$, which is less than its critical value, accept null hypothesis.

**11.25** $H_0$ : Both samples come from the same population.

$H_1$ : The two samples come from different populations.

| *Branch 1* | *Order* | *Branch 2* | *Order* |
|---|---|---|---|
| 23,500 | 5 | 24,000 | 7.5 |
| 25500 | 11 | 19,800 | 2 |
| 35,500 | 14 | 22,000 | 4 |
| 19,500 | 1 | 21,500 | 3 |
| 24,400 | 9 | 24,500 | 10 |
| 24,000 | 7.5 | | |
| 23,600 | 6 | | |
| 25,900 | 12 | | |
| 26,000 | 13 | | |

Rank sum are $R_1 = 78.5$ and $R_2 = 26.5$

$$U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= 9\times 5 + \frac{9\times 6}{2} - 78.5 = -6.5$$

$$U_2 = 45 + 27 - 26.5 = 45.5$$

$$\mu = \frac{n_1n_2}{2} = \frac{54}{2} = 27$$

$$\sigma_u = \sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{45\times 15}{12}} = 7.5$$

$$z = \frac{U-\mu}{\sigma_u} = \frac{-6.5-27.0}{7.5} = -4.466$$

Since $z_{cal} < z_\alpha$ (= 1.645) accept $H_0$.

**11.26**

$$z = \frac{T-\mu_T}{\sigma_T} = \frac{T - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

$$= \frac{163.5 - (25)(26)/4}{\sqrt{(25)(26)(51)/24}} = 0.027$$

Since $z_{cal}$ (= 0.027) < $z_\alpha$( 1.645), $H_0$ is accepted.

**11.30** Let $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$

| *Family Pair* | *Delhi* | *Mumbai* | *d* | *Rank* |
|---|---|---|---|---|
| 1 | 1950 | 1760 | + 190 | + 4 |
| 2 | 1840 | 1870 | − 30 | − 1 |
| 3 | 2015 | 1810 | + 205 | + 5 |
| 4 | 1580 | 1660 | − 80 | − 2 |
| 5 | 1790 | 1340 | + 450 | + 6 |
| 6 | 1925 | 1765 | + 160 | + 3 |

Since $s = \text{Min}(s_+, s_-) = \text{Min}(18, 3) = 3$ is greater than its critical value $s = 1$, $H_0$ is accepted.

# Formulae Used

1. $\chi^2$-test statistic $\chi^2 = \Sigma(O - E)^2/E$
2. Expected frequencies for a contingency table
$$E_{ij} = \frac{\text{Row } i \text{ total} \times \text{Column } j \text{ total}}{\text{Sample size}}$$
3. Degree of freedom for a contingency table
$$df = (r - 1)(c - 1)$$

# Chapter Concepts Quiz

## True or False

1. The critical value of a statistic is the value which cuts off the region for the rejection of null hypothesis. (T/F)
2. The $\chi^2$-test requires that the sampling distribution should be normally distributed. (T/F)
3. If null hypothesis is accepted, then it is concluded that the data analysis did not produce significant results. (T/F)
4. If results of hypothesis testing are statistically significant, the independent variable must have had a very large effect. (T/F)
5. When the results of hypothesis testing are statistically significant, they are unlikely to reflect sampling error. (T/F)
6. The use of non-parametric tests depends on the normal distribution of the underlying population. (T/F)
7. A basic assumption for using $\chi^2$ is that the numbers (frequencies) in each cell of the contingency table are independent. (T/F)
8. When using $\chi^2$, the closer the observed frequency for each cell of the contingency table is to the expected frequency, the higher the probability of rejecting null hypothesis (T/F)
9. In order to reject the null hypothesis $\chi^2_{cal} > \chi^2_{table}$. (T/F)
10. The $\chi^2$ sampling distribution is a family of curves, the distribution of which varies with the degree of freedom. (T/F)
11. The $\chi^2$-test is appropriate for testing hypothesis about proportions. (T/F)
12. If the expected and observed values are the same for each cell in the contingency table, $\chi^2_{cal}$ will not be statistically significant. (T/F)
13. If sample data is used to calculate expected values, then contingency tables are used for calculating $\chi^2_{cal}$. (T/F)
14. The level of significance is generally set at $\alpha = 0.05$ or $\alpha = 0.01$ for $\chi^2$-test. (T/F)
15. A contingency table shows the relationship between two variables. (T/F)

## Multiple Choice

16. Which of the following is a 'non-parametric' test?
(a) $\chi^2$-test (b) $t$-test
(c) $z$-test (d) none of these
17. Which of the following tests is appropriate for analysing data where 3 or more groups were used?
(a) $z$ (b) $\chi^2$
(c) $t$ (d) none of these
18. The larger the difference between the expected and observed frequencies for each cell in a contingency table,
(a) the more likely it is that null hypothesis will be rejected
(b) the more likely it is that the population proportions are different
(c) both (a) and (b)
(d) none of these
19. For any given level of significance, $\chi^2_{table}$ value:
(a) increases as sample size increases
(b) decreases as degree of freedom increases
(c) increases as degree of freedom increases
(d) decreases as sample size increases
20. A contingency table
(a) always involves two degrees of freedom
(b) always involves two dependent variables
(c) always involves two variables
(d) all of these
21. Entries into the cells of a contingency table should be:
(a) frequencies (b) mean values
(c) percentages (d) degrees of freedom
22. The degrees of freedom for a contingency table:
(a) equal $n - 1$ (b) equal $rc - 1$
(c) equal $(r - 1)(c - 1)$ (d) cannot be determined
23. Chi-square should not be used with a $2 \times 2$ contingency table if:
(a) $df > 1$
(b) expected values are below 5 in any cell
(c) observed values are below 5 in any cell
(d) both (b) and (c)
24. The $\chi^2$-test requires that:
(a) data be measured on a nominal scale
(b) data conform to a normal distribution
(c) expected frequencies are equal in all cells
(d) all of these
25. In selecting an appropriate statistical test:
(a) $z$ should be used as it is most powerful

(b) $t$ should be used as it takes the sample size into account
(c) the choice depends on the design of the study
(d) $\chi^2$ should be avoided

**26.** What type of $\chi^2$-test will be conducted to see whether the proportions of two genders at the class is representative of the proportions of the two genders enrolled at the college as a whole?
(a) one-way (b) two-way
(c) contingency table (d) a parametric chi-square

**27.** Statistical tests are used:
(a) only when the investigation involves a true experimental design
(b) to increase the internal velocity of experiments
(c) to establish the probability of the outcome of data analysis being due to chance alone
(d) both (a) and (b)

**28.** If the null hypothesis is accepted, then you may be making:
(a) a correct decision about the data
(b) a Type I error
(c) a Type II error
(d) either (a) or (b)

**29.** When the results of hypothesis testing are statistically significant, this means:
(a) the obtained probability is equal to or less than $\alpha$
(b) rejecting null hypothesis
(c) both (a) and (b)
(d) either (a) or (b)

**30.** A null hypothesis is accepted when:
(a) $\chi^2_{cal} \le \chi^2_{table}$ (b) $\chi^2_{cal} > \chi^2_{table}$
(c) $\chi^2_{cal} < \chi^2_{table}$ (d) none of these

## Concepts Quiz Answers

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **1.** T | **2.** F | **3.** T | **4.** F | **5.** T | **6.** F | **7.** T | **8.** F | **9.** T |
| **10.** T | **11.** T | **12.** T | **13.** T | **14.** T | **15.** T | **16.** (a) | **17.** (b) | **18.** (c) |
| **19.** (c) | **20.** (c) | **21.** (a) | **22.** (c) | **23.** (b) | **24.** (a) | **25.** (c) | **26.** (a) | **27.** (c) |
| **28.** (d) | **29.** (c) | **30.** (c) | | | | | | |

# Review Self-Practice Problems

**11.31** 1000 students at college level were graded according to their IQ and the economic conditions of their homes. Use the $\chi^2$-test to find out whether there is any association between economic condition at home and IQ level.

| *Economic Condition* | *IQ Level* | | |
|---|---|---|---|
| | *High* | *Low* | *Total* |
| Rich | 460 | 140 | 600 |
| Poor | 240 | 160 | 400 |
| Total | 700 | 300 | 1000 |

[*Osmania, MBA, 1998 Kumaon; Univ., MBA, 1999*]

**11.32** An automobile company gives you the following information about age groups and the liking for a particular model of car which it plans to introduce.

| *Opinion* | *Age Groups* | | | | |
|---|---|---|---|---|---|
| | *Below 20* | *20-39* | *40-59* | *60 and above* | *Total* |
| Liked the car | 140 | 80 | 40 | 20 | 280 |
| Disliked the car | 60 | 50 | 30 | 80 | 220 |
| Total | 200 | 130 | 70 | 100 | 500 |

On the basis of this data can it be concluded that the model's appeal is independent of the age group?

[*HP Univ., MCom; MD Univ, MCom, 1996*]

**11.33** The following results were obtained when two sets of items were subjected to two different treatments X and Y, to enhance their tensile strength.

- Treatment X was applied on 400 items and 80 were found to have gained in strength.
- Treatment Y was applied on 400 items and 20 were found to have gained in strength.
- Is treatment Y superior to treatment X?

[*Calcutta Univ., MCom, 1998; Madras Univ., MCom, 1996*]

**11.34** Three samples are taken comprising 120 doctors, 150 advocates, and 130 university teachers. Each person chosen is asked to select one of the three categories that best represents his feeling towards a certain national policy. The three categories are in favour of policy (F), against the policy (A), and indifferent toward to policy (I). The results of the interviews are given below:

| *Occupation* | *Reaction* | | | |
|---|---|---|---|---|
| | *F* | *A* | *I* | *Total* |
| Doctors | 80 | 30 | 10 | 120 |
| Advocates | 70 | 40 | 40 | 150 |
| University teachers | 50 | 50 | 30 | 130 |
| Total | 200 | 120 | 80 | 400 |

On the basis of this data can it be concluded that the views of doctors, advocates, and university teachers are homogeneous insofar as the National Policy under discussion is concerned?

**11.35** Following information is obtained in a sample survey:

| *Condition of Child* | *Condition of Home* | | |
|---|---|---|---|
| | *Clean* | *Dirty* | *Total* |
| Clean | 70 | 50 | 120 |
| Fairly clean | 80 | 20 | 100 |
| Dirty | 35 | 45 | 80 |
| Total | 185 | 115 | 300 |

State whether the two attributes, that is, condition of home and condition of child are independent. Use the $\chi^2$-test for the purpose.

[*Madurai Kamraj Univ., MCom, 1996*]

**11.36** National Healthcare Company samples its hospital employees' attitude towards performance. Respondents are given a choice between the present method of two reviews a year and a proposed new method of quality reviews. The responses are given below:

| | *North* | *South* | *East* | *West* |
|---|---|---|---|---|
| Method I | 68 | 75 | 79 | 57 |
| Method II | 32 | 45 | 31 | 33 |

Test whether there is any significant difference in the attitude of employees in different regions at 5 per cent level of significance. [*Bharthidasam Univ., MCom, 1996*]

**11.37** A milk producers union wishes to test whether the preference pattern of consumers for its products is dependent on income levels. A random sample of 500 individuals gives the following data:

| *Income* | *Product Preferred* | | |
|---|---|---|---|
| | *A* | *B* | *C* |
| Low | 170 | 30 | 80 |
| Medium | 50 | 25 | 60 |
| High | 20 | 10 | 55 |

Can you conclude that the preference patterns are independent of income levels?

[*Calicut Univ., MCom, 1999*]

**11.38** The following data relate to the sales, in a time of trade depression of a certain article in wide demand. Do the data suggest that the sales are significantly affected by depression?

| *Pattern of Sales* | *Districts* | | |
|---|---|---|---|
| | *Not Hit by Depression* | *Hit* | *Total* |
| Satisfactory | 140 | 60 | 200 |
| Not satisfactory | 40 | 60 | 100 |
| Total | 180 | 120 | 300 |

[*GND Univ., MCom, 1997*]

**11.39** A controlled experiment was conducted to test the effectiveness of a new drug. Under this experiment 300 patients were treated with new drug and 200 were not treated with the drug. The results of the experiment are given below:

| *Details* | *Cured* | *Condition Worsened* | *No Effect* | *Total* |
|---|---|---|---|---|
| Treated with the drug | 200 | 40 | 60 | 300 |
| Not treated with the drug | 120 | 30 | 50 | 200 |
| Total | 320 | 70 | 110 | 500 |

Use the $\chi^2$-test and comment on the effectiveness of the drug.

**11.40** A survey of 320 families with 5 children each revealed the following distribution:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| No. of boys | : | 5 | 4 | 3 | 2 | 1 | 0 |
| No. of girls | : | 0 | 1 | 2 | 3 | 4 | 5 |
| No. of families | : | 14 | 56 | 110 | 88 | 40 | 12 |

Is the result consistent with the hypothesis that male and female births are equally probable?

**11.41** The following figures show the distribution of digits in numbers chosen at random from a telephone directory:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Digit | : | 0 | 1 | 2 | 3 | 4 | 5 |
| Frequency | : | 1026 | 1107 | 997 | 966 | 1075 | 933 |
| Digit | : | 6 | 7 | 8 | 9 | | |
| Frequency | : | 1107 | 972 | 964 | 853 | = 10,000 | |

Test whether the digits may be taken to occur equally frequently in the directory. [*Roorkee Univ., MBA, 2000*]

**11.42** The number of customers that arrived in 128, 5-minute time periods at a service window were recorded as:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Customer | : | 0 | 1 | 2 | 3 | 4 | 5 |
| Frequency | : | 2 | 8 | 10 | 12 | 18 | 22 |
| Customer | : | 6 | 7 | 8 | 9 | | |
| Frequency | : | 22 | 16 | 12 | 6 | | |

Is the probability distribution for the customer arrivals a Poisson distribution with a 0.05 level of significance?

**11.43** A production supervisor is interested in knowing if the number of breakdowns of four machines is independent of the shift using the machines. Test this hypothesis based on the following sample information:

| *Shift* | *Machine* | | | | |
|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | *Total* |
| Morning | 15 | 10 | 18 | 12 | 55 |
| Evening | 12 | 8 | 15 | 10 | 45 |
| Total | 27 | 18 | 33 | 22 | 100 |

**11.44** You are given the distribution of the number of defective units produced in a single shift in a factory over 100 shifts.

| Number of defective units : | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Number of shifts: | 4 | 14 | 23 | 23 | 18 | 9 | 9 |

Would you say that the defective units follow a Poisson distribution? [*Delhi Univ., MBA, 1995*]

**11.45** The number of car accidents that occurred during the various days of the week are as follows:

| *Day* | *No. of accidents* |
|---|---|
| Sun. | 14 |
| Mon. | 16 |
| Tues. | 8 |
| Wed. | 12 |
| Thurs. | 11 |
| Fri. | 9 |
| Sat. | 14 |

Find whether the accidents are uniformly distributed over the week. [*MD Univ., MCom, 1999*]

## Hints and Answers

**11.31** Let $H_0$ : No association between economic condition and IQ level.

| *O* | *E* | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 460 | 420 | 1600 | 3.810 |
| 240 | 280 | 1600 | 5.714 |
| 140 | 180 | 1600 | 8.889 |
| 160 | 120 | 1600 | 13.333 |
| | | | 31.746 |

Since $\chi^2_{cal}$ = 31.746 is more than its critical value $\chi^2$ = 3.84 for *df* = (2 – 1) 2 – 1) = 1 and α = 5 per cent, the $H_0$ is rejected.

**11.32** Let $H_0$ : Personal liking is independent of the age group.

| *O* | *E* | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 140 | 112.0 | 784.00 | 7.000 |
| 60 | 88.0 | 784.00 | 8.910 |
| 80 | 72.8 | 51.84 | 0.712 |
| 50 | 57.2 | 51.84 | 0.906 |
| 40 | 39.2 | 0.64 | 0.016 |
| 30 | 30.8 | 0.64 | 0.021 |
| 20 | 56.0 | 1296.00 | 23.143 |
| 80 | 44.0 | 1296.00 | 29.454 |
| | | | 70.162 |

Since $\chi^2_{cal}$ = 7.162 is more than its critical value $\chi^2$ = 7.815 for *df* = (2 – 1) (4 – 1) = 3 and α = 5 per cent, the $H_0$ is rejected.

**11.33** Let $H_0$ : No difference in the types of treatments X and Y.

| *Treatment* | *Gained* | *Not Gained* | *Total* |
|---|---|---|---|
| X | 80 | 320 | 400 |
| Y | 20 | 380 | 400 |
| Total | 100 | 700 | 800 |

| *O* | *E* | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 80 | 50 | 900 | 18.000 |
| 20 | 50 | 900 | 18.000 |
| 320 | 350 | 900 | 2.571 |
| 380 | 350 | 900 | 2.571 |
| | | | 41.142 |

Since $\chi^2_{cal}$ = 41.142 is more than its critical value $\chi^2$ = 3.84 for *df* = (2 – 1) (2 – 1) = 1 and α = 5 per cent, the $H_0$ is rejected.

**11.34** Let $H_0$ : Opinion of three different groups of people about National Policy is same.

$$E_{11} = \frac{120\times200}{400} = 60 \qquad E_{12} = \frac{120\times120}{400} = 36$$

$$E_{21} = \frac{150\times200}{400} = 75 \qquad E_{22} = \frac{150\times120}{400} = 45$$

| *O* | *E* | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 80 | 60 | 400 | 6.667 |
| 70 | 75 | 25 | 0.333 |
| 50 | 65 | 225 | 3.462 |
| 30 | 36 | 36 | 1.000 |
| 40 | 45 | 25 | 0.556 |
| 50 | 39 | 121 | 3.103 |
| 10 | 24 | 196 | 8.167 |
| 40 | 30 | 100 | 3.333 |
| 30 | 26 | 16 | 0.616 |
| | | | 27.237 |

Since $\chi^2_{cal}$ = 27.237 is more than its critical value $\chi^2$ = 9.488 for $df$ = (3 – 1) (3 – 1) = 4 and $\alpha$ = 5 per cent, the $H_0$ is rejected.

**11.35** Let $H_0$ : Two attributes, that is, condition of home and condition of child are independent.

| $O$ | $E$ | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 70 | 74 | 16 | 0.216 |
| 80 | 62 | 324 | 5.226 |
| 35 | 49 | 196 | 4.000 |
| 50 | 46 | 16 | 0.348 |
| 20 | 38 | 324 | 8.526 |
| 45 | 31 | 196 | 6.322 |
| | | | 24.638 |

Since $\chi^2_{cal}$ = 24.638 is more than its critical value $\chi^2$ = 5.99 for $df$ = (3 – 1) (2 – 1) = 2 and $\alpha$ = 5 per cent, the $H_0$ is rejected.

**11.36** Let $H_0$ : No difference in the attitude of employees in different regions

| $O$ | $E$ | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 68 | 66 | 4 | 0.061 |
| 32 | 34 | 4 | 0.118 |
| 75 | 80 | 25 | 0.312 |
| 45 | 40 | 25 | 0.625 |
| 79 | 73 | 36 | 0.493 |
| 31 | 37 | 36 | 0.973 |
| 57 | 60 | 9 | 0.150 |
| 33 | 30 | 9 | 0.300 |
| | | | 3.032 |

Since $\chi^2_{cal}$ = 3.032 is less than its critical value $\chi^2$ = 7.82 at $df$ = (2 – 1) (4 – 1) = 3 and $\alpha$ = 5 per cent, the $H_0$ is accepted.

**11.37** Let $H_0$ : Income level and product preferred are independent.

| $O$ | $E$ | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 170 | 134.40 | 1267.36 | 9.430 |
| 50 | 64.80 | 219.04 | 3.380 |
| 20 | 40.80 | 432.64 | 10.604 |
| 30 | 36.40 | 40.96 | 1.125 |
| 25 | 17.55 | 55.50 | 3.162 |
| 10 | 11.05 | 1.10 | 0.099 |
| 80 | 109.20 | 852.64 | 7.808 |
| 60 | 52.65 | 54.02 | 1.026 |
| 55 | 33.15 | 477.42 | 14.402 |
| | | | 51.036 |

Since $\chi^2_{cal}$ = 51.036 is more than its critical value $\chi^2$ = 14.860 at $df$ = (3 – 1) (3 – 1) = 4 and $\alpha$ = 5 per cent, the $H_0$ is rejected.

**11.38** Let $H_0$ : Sales are not significantly affected by depression.

| $O$ | $E$ | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 140 | 120 | 400 | 3.333 |
| 40 | 60 | 400 | 6.667 |
| 60 | 80 | 400 | 5.000 |
| 60 | 40 | 400 | 10.000 |
| | | | 25.000 |

Since $\chi^2_{cal}$ = 25 is more than its critical value $\chi^2$ = 3.84 at $df$ = (2 – 1) (2 – 1) = 1 and $\alpha$ = 5 per cent, the $H_0$ is rejected.

**11.39** Let $H_0$ : No significant difference in patients, condition between those treated with the new drug and those not treated

| $O$ | $E$ | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|
| 200 | 192 | 64 | 0.333 |
| 120 | 128 | 64 | 0.500 |
| 40 | 42 | 4 | 0.095 |
| 30 | 28 | 4 | 0.143 |
| 60 | 66 | 36 | 0.545 |
| 50 | 44 | 36 | 0.818 |
| | | | 2.434 |

Since $\chi^2_{cal}$ = 2.434 is less than its critical value $\chi^2$ = 5.49 at $df$ = (2 – 1) (3 – 1) = 2 and $\alpha$ = 5 per cent, the $H_0$ is accepted.

**11.40** Let $H_0$ : Male and female births are equally probable.

Given $p = q = 0.5$. Since events are only two and mutually exclusive, the binomial distribution is used to calculate the expected frequencies (number of families). Thus the expected number of families having $x$ male in a family of 5 children is given by

$nP(x = r) = n\,{}^nC_r p^r q^{n-r} = 320\,{}^5C_r(0.5)^5, r = 0, 1, 2, \ldots, 5$

| | | | | | | |
|---|---|---|---|---|---|---|
| O-values: | 14 | 56 | 110 | 88 | 40 | 12 |
| E-values: | 10 | 50 | 100 | 100 | 50 | 10 |

Since $\chi^2_{cal} = \Sigma(O-E)^2/E$ = 7.16 is less than its critical value $\chi^2$ = 11.07 at $df = n - 1$ = 6 – 1 = 5 and $\alpha$ = 5 per cent, the null hypothesis is accepted.

**11.41** Let $H_0$ : Digits are uniformly distributed in the directory.

The expected frequency (E) for each digit 0 to 9 is 10,000/10 = 1000.

Since $\chi^2_{cal}$ = 58.542 is more than its critical value $\chi^2$ = 16.919 at $df$ = 10 – 1 = 9 and $\alpha$ = 5 per cent, the null hypothesis is rejected.

**11.42** Let $H_0$ : The population has a Poisson distribution

Total number of customers who arrived during the sample of 128, 5-minute time periods is given by 0(2) + 1(8) + 2(10) + . . . + 9(6) = 640. Thus 640 customers arrival over a sample of 128 periods provide a mean arrival rate of $\lambda$ = 640/128 = 5 customers per 5-minute period.

The expected frequency (E) of customer arrivals for each of the variable from 0 to 9 is calculated by using the Poisson distribution formula

$$nP(x = r) = n\,\frac{e^{-\lambda}\lambda^r}{r!} = 128\,\frac{e^{-5}(5)^r}{r!}$$

$r = 0, 1, ..., 9; \lambda = 5$

| *Number of customers* | *Observed frequency* | *Expected frequency* |
|---|---|---|
| 0 or 1 | 10 | 5.17 |
| 2 | 10 | 10.77 |
| 3 | 12 | 17.97 |
| 4 | 18 | 22.46 |
| 5 | 22 | 22.46 |
| 6 | 22 | 18.71 |
| 7 | 16 | 13.36 |
| 8 | 12 | 8.35 |
| 9 or more | 6 | 8.71 |

Since $\chi^2_{cal} = 10.97$ is less than its critical value $\chi^2 = 14.07$ at $df = n - 1 - 2 = 10 - 1 - 2 = 7$ and $\alpha = 0.05$, the null hypothesis is accepted.

**11.43** Let $H_0$ : Number of breakdowns is independent of the shift using the machines.

$$\text{Expected frequencies} = \frac{\text{Row } i \text{ total} \times \text{Column } j \text{ total}}{\text{Grand total}}$$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| O-values | : | 15 | 12 | 10 | 8 | 18 | 15 | 12 | 10 |
| E-values | : | 14.85 | 12.15 | 9.90 | 8.10 | 18.15 | 14.85 | 12.10 | 9.90 |

# Chapter 12

*There is no merit in equality unless it be equality with best.*

—John Spalding

# Analysis of Variance

## LEARNING OBJECTIVES

After studying this chapter, you should be able to

- understand how 'analysis of variance'(ANOVA) can be used to test for the equality of three or more population means.
- understand and use the terms like 'response variable', 'a factor' and 'a treatment' in the analysis of variance.
- learn how to summarize F-ratio in the form of an ANOVA table.

## 12.1 INTRODUCTION

In Chapter 10 we introduced hypothesis testing procedures to test the significance of differences between two sample means to understand whether the means of two populations are equal based upon two independent random samples. In all these cases the null hypothesis states that there is no significant difference among population mean, that is, $H_0 : \mu_1 \neq \mu_2$. However, there may be situations where more than two populations are involved and we need to test the significance of differences between three or more sample means. We also need to test the null hypothesis that three or more populations from which independent samples are drawn have equal (or homogeneous) means against the alternative hypothesis that population means are not all equal. Let $\mu_1, \mu_2, \ldots, \mu_r$ be the mean value for population 1, 2, . . ., $r$ respectively. Then from sample data we intend to test the following hypotheses:

$$H_0 : \mu = \mu_2 = \ldots = \mu_r \qquad (12\text{–}1)$$

and $\quad H_1$ : Not all $\mu_j$ are equal ($j = 1, 2, \ldots, r$)

In other words, the null and alternative hypotheses of population means imply that the null hypothesis should be rejected if any of the $r$ sample means is different from the others.

For example, the production level in three shifts in a factory can be compared to answer questions such as: Is the production level higher/lower on any day of the week? Is Wednesday morning shift's production better/worse than any other shift? and so on. Production level can also be analysed using other days and shifts of the week in combination.

The following are a few examples involving more than two populations where it is necessary to conduct a comparative study to arrive at a statistical inference:

- Effectiveness of different promotional devices in term of sales
- Quality of a product produced by different manufacturers in terms of an attribute
- Production volume in different shifts in a factory
- Yield from plots of land due to varieties of seeds, fertilizers, and cultivation methods

Under certain circumstances we may not conduct repeated $t$-tests on pairs of the samples. This is because when many independent tests are carried out pairwise, the probability of the outcome being correct for the combined results is reduced greatly. Table 12.1 shows how the probability of being correct decreases when we intend to compare the average marks of 2, 3, and 10 students at the end of an examination at 95 per cent confidence level or 0.95 probability of being correct in our statistical inferences in this experiment.

**Table 12.1: Calculations of Probability of being Correct**

| *Number of Students, n* | *Number of Pairs* | *Confidence Level* | *Probability of Error* |
|---|---|---|---|
| 2 | 1 | 0.950 | 0.050 |
| 3 | 3 | $(0.95)^3 = 0.857$ | 0.143 |
| 10 | 45 | $(0.95)^{45} = 0.100$ | 0.900 |

**Analysis of variance:** A statistical procedure for determining whether the means of several different populations are equal.

It is clear from the calculations in Table 12.1 that as the size of student population or sample increases, the probability of error in statistical inference of population means increases. Under certain assumptions, a method known as **analysis of variance** *(ANOVA)* developed by R. A. Fisher is used to test the significance of the difference between several population means.

The following are few terms that will be used during discussion on analysis of variance:

- *A sampling plan or experimental design* is the way that a sample is selected from the population under study and determines the amount of information in the sample.
- An *experimental unit* is the object on which a measurement or measurements is taken. Any experimental conditions imposed on an experimental unit provides effect on the response.
- A *factor or criterion* is an indepenent variable whose values are controlled and varied by the researcher.
- A *level* is the intensity setting of a factor.
- A *treatment or population* is a specific combination of factor levels.
- The *response* is the dependent variable being measured by the researcher.

For example,

**Factor:** Another word for independent variable of interest that is controlled in the analysis of variance.

**Factor level:** A value at which the factor is controlled.

1. A tyre manufacturing company plans to conduct a tyre-quality study in which quality is the independent variabnle called *factor or criterion* and the *treatment levels or classifications* are low, medium and high quality. The *dependent* (*or response*) variable might be the number of kilometers driven before the tyre is rejected for use. A study of daily sales volumes may be taken by using a completely randomized design with demographic setting as the independent variable. A treatment levels or classifications would be inner-city stores, stores in metro-cities, stores in state capitals, stores in small towns, etc. The dependent variable would be sales in rupees.
2. For a production volume in three shifts in a factory, there are two variables—days of the week and the volume of production in each shift. If one of the objectives is to determine whether mean production volume is the same during days of the week, then the *dependance (or response) variable* of interest, is the mean production volume. The *variables* that are related to a response variable are called **factors**, that is, a day of the week is the *independent variable* and the value assumed by a factor in an experiment is called a **level**. The combinations of levels of the factors for which the response will be observed are called *treatments*, i.e. days of the week. These treatments define the populations or samples which are differentiated in terms of production volume and we may need to compare them with each other.

## 12.2 ANALYSIS OF VARIANCE APPROACH

The first step in the analysis of variance is to partition the total variation in the sample data into the following two component variations in such a way that it is possible to estimate the contribution of factors that may cause variation.

1. The amount of variation *among the sample means* or the variation attributable to the difference among sample means. This variation is either on account of difference in treatment or due to element of chance. This difference is denoted by SSC or SSTR.
2. The amount of variation *within the sample observations*. This difference is considered due to chance causes or experimental (random) errors. The difference in the values of various elements in a sample due to chance is called an estimate and is denoted by SSE.

The observations in the sample data may be classified according to *one factor* (criterion) or *two factors* (criteria). The classifications according to one factor and two factors are called *one-way classification* and *two-way classification*, respectively. The calculations for total variation and its components may be carried out in each of the two-types of classifications by (*i*) *direct method*, (*ii*) *short-cut method*, and (*iii*) *coding method*.

### Assumptions for Analysis of Variance

The following assumptions are required for analysis of variance:

1. Each population under study is normally distributed with a mean $\mu_r$ that may or may not be equal but with equal variances $\sigma_r^2$.
2. Each sample is drawn randomly and is independent of other samples.

## 12.3 TESTING EQUALITY OF POPULATION (TREATMENT) MEANS: ONE-WAY CLASSIFICATION

Many business applications involve experiments in which different populations (or groups) are classified with respect to only one attribute of interest such a (i) *percentage of marks* secured by students in a course, (ii) *flavour preference* of ice-cream by customers, (iii) *yield of crop* due to varieties of seeds, and so on. In all such cases observations in the sample data are classified into several groups based on a single attribute and is called **one-way classification** of sample data.

**One-way analysis of variance:** Analysis of variance in which only one criterion (variable) is used to analyse the difference between more than two population means.

As mentioned before, for all theoretical purposes we refer populations (i.e., several groups classified based on single factor or criterion in a sample data) as treatments. We will study the effect of a factor (criterion) such as flavour preference on the dependent variable (i.e. sales) at different groups (i.e. variety of ice-creams). These groups are the treatments in this particular example.

Suppose our aim is to make inferences about $r$ population means $\mu_1, \mu_2, \ldots \mu_r$ based on independent random samples of size $n_1, n_2, \ldots, n_r$, from normal populations with a common variances $\sigma^2$. That is, each of the normal population has same shape but their locations might be different. The null hypothesis to be tested is stated as:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_r \quad \leftarrow \text{Null hypothesis}$$

$$H_1: \text{Not all } \mu_j \ (j = 1, 2, \ldots, r) \text{ are equal} \quad \leftarrow \text{Alternative hypothesis}$$

Let $n_j$ = size of the $j$th sample ($j = 1, 2, \ldots, r$)

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

$x_{ij}$ = the $i$th observation value within the sample from $j$th population

The observations values obtained for $r$ independent samples based on one-criterion classification can be arranged as shown in Table 12.2.

**Table 12.2: One-Criterion Classification of Data**

| *Observations* | *Populations (Number of Samples)* | | | |
|---|---|---|---|---|
| | *1* | *2* | ... | *r* |
| 1 | $x_{11}$ | $x_{12}$ | ... | $x_{1r}$ |
| 2 | $x_{21}$ | $x_{22}$ | ... | $x_{2r}$ |
| ⋮ | ⋮ | | | |
| $k$ | $x_{k1}$ | $x_{k2}$ | ... | $x_{kr}$ |
| Sum | $T_1$ | $T_2$ | ... | $T_r = T$ |
| A.M. | $\bar{x}_1$ | $\bar{x}_2$ | ... | $\bar{x}_r = \bar{\bar{x}}$ |

where $T_i = \sum_{i=1}^{k} x_{i1}$ $\quad T = \sum_{j=1}^{r} T_i$

$$\bar{x}_i = \frac{1}{k}\sum_{i=1}^{k} x_{i1} \qquad \bar{\bar{x}} = \frac{1}{rk}\sum_{j=1}^{r} \bar{x}_j = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{r} x_{ij}$$

The values of $\bar{x}_i$ are called sample means and $\bar{\bar{x}}$ is the grand mean of all observations (or measurements) in all the samples.

Since there are $k$ rows and $r$ columns in Table 12.2, therefore total number of observations are $rk = n$, provided each row has equal number of observations. But, if the number of observations in each row varies, then the total number of observations is $n_1 + n_2 + \ldots + n_r = n$.

**Illustration:** Three brands A, B, and C of tyres were tested for durability. A sample of four tyres of each brand is subjected to the same test and the number of kilometres until wearout was noted for each brand of tyres. The data in thousand kilometres is given in Table 12.3.

**Table 12.3: Example of Data in ANOVA**

| *Observations* | *Population (Number of Brands)* | | |
|---|---|---|---|
| | *A* | *B* | *C* |
| 1 | 26 | 18 | 23 |
| 2 | 25 | 16 | 19 |
| 3 | 28 | 17 | 26 |
| 4 | 12 | 18 | 30 |
| Sum | 91 | 69 | 98 |
| Sample size | 4 | 4 | 4 |
| Mean | 22.75 | 17.25 | 24.50 |

Since the same number of observations is obtained from each brand of tyres (population), therefore the number of observations in the table is $n = rk = 3\times 4 = 12$.

- The sample mean of each of three samples is given by

$$\bar{x}_1 = \frac{1}{4}\sum_{i=1}^{4} x_{i1} = \frac{1}{4}(91) = 22.75$$

$$\bar{x}_2 = \frac{1}{4}\sum_{i=1}^{4} x_{i2} = \frac{1}{4}(69) = 17.25 \text{ and } \bar{x}_3 = 24.50$$

- The grand mean for all samples is

$$\bar{\bar{x}} = \frac{1}{3}(\bar{x}_1 + \bar{x}_2 + \bar{x}_3) = \frac{1}{3}(22.75 + 17.25 + 24.50) = 21.50$$

### 12.3.1 Steps for Testing Null Hypothesis

**Step 1:** State the null and alternative hypotheses to test the equality of population means as:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_r$ ← Null hypothesis

$H_1:$ Not all $\mu_j s$ are equal $(j = 1, 2, \ldots, r)$ ← Alternative hypothesis

$\alpha$ = level of significance

**Step 2: Calculate total variation** If a single sample of size $n$ is taken from the population, then estimate of the population variance based on the variance of sampling distribution of mean is given by

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1} = \frac{SS}{df}$$

The numerator in $s^2$ is called *sum of squares* of deviations of sample values about the sample mean $\bar{x}$ and is denoted as SS. Consequently 'sum of squares' is a measure of variation. Thus when SS is divided by $df$, the result is often called the *mean square* which is an alternative term for sample variance.

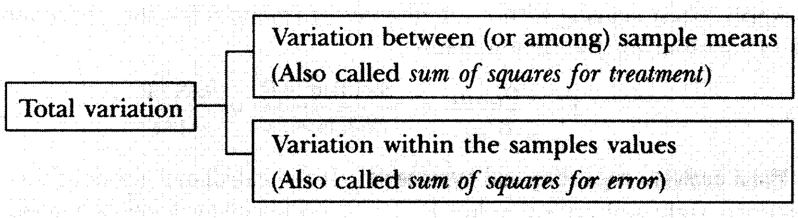
Total variation is represented by the '*sum of squares total*' (SST) and is equal to the sum of the squared differences between each sample value from the grand mean $\bar{\bar{x}}$

$$\text{SST} = \sum_{i=1}^{r} \sum_{j=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2$$

where $r$ = number of samples (or treatment levels)

$n_j$ = size of $j$th sample

The total variation is divided into two parts as shown below:




**Step 3: Calculate variation between sample means** This is usually called the 'sum of squares between' and measures the variation between samples due to treatments. In statistical terms, variation between samples means is also called the *between-column variance*. The procedure is as follows:

(a) Calculate mean values $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_r$ of all $r$ samples

(b) Calculate grand mean $\bar{\bar{x}} = \frac{1}{r}(\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_r) = \frac{T}{n}$

where T = grand total of all observations.

$n$ = number of observations in all $r$ samples.

(c) Calculate difference between the mean of each sample and the grand mean as $\bar{x}_1 - \bar{\bar{x}}$, $\bar{x}_2 - \bar{\bar{x}}$, ..., $\bar{x}_r - \bar{\bar{x}}$. Multiply each of these by the number of observations in the corresponding sample and add. The total gives the sum of the squared differences between the sample means in each group and is denoted by SSC or SSTR.

$$\text{SSTR} = \sum_{j=1}^{r} n_j(\bar{x}_j - \bar{\bar{x}})^2$$

This sum is also called *sum of squares for treatment* (SSTR)

**Step 4: Calculate variation within samples** This is usually called the 'sum of squares within' and measures the difference within samples due to chance error. Such variation is also called *within sample variance*. The procedure is as follows:

(a) Calculate mean values $\bar{x}_1$, $\bar{x}_2$, ..., $\bar{x}_r$ of all $r$ samples.

(b) Calculate difference of each observation in $r$ samples from the mean values of the respective samples.

(c) Square all the differences obtained in Step (b) and find the total of these differences. The total gives the sum of the squares of differences within the samples and is denoted by SSE.

$$\text{SSE} = \sum_{i=1}^{k} \sum_{j=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

This sum is also called the *sum of squares for error*, SSE = SST − SSTR.

**Step 5: Calculate average variation between and within samples—mean squares** Since $r$ independent samples are being compared, therefore $r - 1$ degrees of freedom are associated with the sum of the squares among samples. As each of the $r$ samples contributes $n_j - 1$ degrees of freedom for each independent sample within itself, therefore there are $n - r$ degrees of freedom associated with the sum of the squares within samples. Thus total degrees of freedom equal to the degrees of freedom associated with SSC (or SSTR) and SSE. That is

$$\text{Total } df = \text{Between samples (treatments) } df + \text{Within samples (error) } df$$
$$n - 1 = (r - 1) + (n - r)$$

When these 'sum of squares' are divided by their associated degrees of freedom, we get the following variances or *mean square* terms:

$$\text{MSTR} = \frac{\text{SSTR}}{r-1}; \quad \text{MSE} = \frac{\text{SSE}}{n-r}; \quad \text{and} \quad \text{MST} = \frac{\text{SST}}{n-1}$$

It may be noted that the quantity MSE = SSE/$(n - r)$ is a pooled estimate of $\sigma^2$ (weighted average of all $r$ sample variances whether $H_0$ is true or not)

**Step 6:** Apply F-test statistic with $r - 1$ degrees of freedom for the numerator and $n - r$ degrees of freedom for the denominator

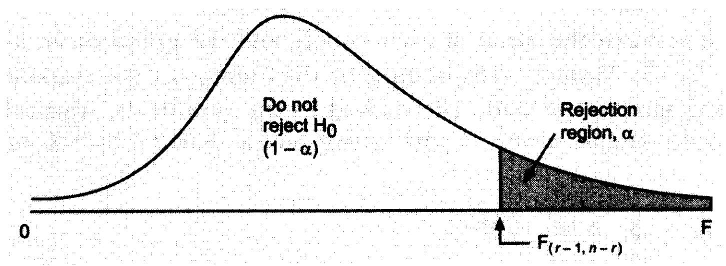
$$F = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{\text{SSTR}/(r-1)}{\text{SSE}/(n-r)} = \frac{\text{MSTR}}{\text{MSE}}$$

**Step 7: Make decision regarding null hypothesis** If the calculated value of F-test statistic is more than its right tail critical value $F_{(r-1,\, n-r)}$ at a given level of significance $\alpha$ and degrees of freedom $r - 1$ and $n - r$, then reject the null hypothesis. In other words, as shown in shown in Fig. 12.1, the decision rule is:

- Reject $H_0$ if the calculated value of F > its critical value $F_{\alpha(r-1,\, n-r)}$
- Otherwise accept $H_0$

The F-distribution is a family of distributions, each identified by a pair of degrees of freedom. The first number refers to the number of degrees of freedom in the numerator of the F ratio, and the second refers to the number of degrees of freedom in the denominator. In the F table, columns represent the degrees of freedom for numerator and the rows represents the degrees of freedom for denominator.

**Figure 12.1**
Rejection Region for Null Hypothesis using ANOVA




If null hypothesis $H_0$ is true, then the variance in the sample means measured by MSTR = SSTR/$(r - 1)$ provides an unbiased estimate of $\sigma^2$. But if $H_0$ is false, and population means are different, then MSTR is large as shown in Fig 12.2

Table 12.4 shows the general arrangement of the ANOVA table for one-factor analysis of variance.

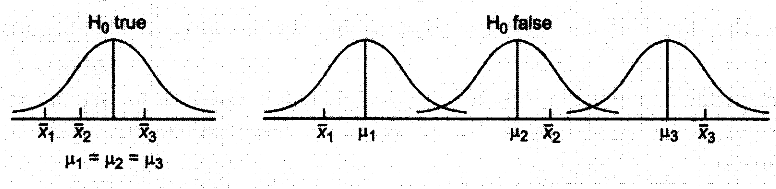



**Figure 12.2**
Sample Means Drawn from Identical Populations

**ANOVA table:** A standard table used to summarize the analysis of variance calculations and results.

**Table 12.4: ANOVA Summary Table**

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Squares* | *Test-Statistic or F-Value* |
|---|---|---|---|---|
| • Between samples (Treatments) | SSTR | $r-1$ | $\text{MSTR} = \dfrac{\text{SSTR}}{r-1}$ | |
| | | | | $F = \dfrac{\text{MSTR}}{\text{MSE}}$ |
| • Within samples error | SSE | $n-r$ | $\text{MSE} = \dfrac{\text{SSE}}{n-r}$ | |
| Total | SST | $n-1$ | | |

**Short-Cut Method** The values of SSTR and SSE can be calculated by applying the following short-cut methods:

- Calculate the grand total of all observations in samples, T
$$T = \Sigma x_1 + \Sigma x_2 + \ldots + \Sigma x_r$$
- Calculate the correction factor $\text{CF} = \dfrac{T^2}{n}$; $n = n_1 + n_2 + \ldots + n_r$
- Find the sum of the squares of all observations in samples from each of $r$ samples and subtract CF form this sum to obtain the total sum of the squares of deviations SST:

$$\text{SST} = (\Sigma x_1^2 + \Sigma x_2^2 + \ldots + \Sigma x_r^2) - \text{CF} = \sum_{i=1}^{k} \sum_{j=1}^{n_j} x_{ij}^2 - \text{CF}$$

$$\text{SSTR} = \frac{(\Sigma x_j)^2}{n_j} - \text{CF}$$

and $$\text{SSE} = \text{SST} - \text{SSTR}$$

**Coding Method** Sometimes the method explained above takes a lot of computational time due to the magnitude of numerical values of observations. The coding method is based on the fact that the F-test statistic used in the analysis of variance is the ratio of variances without unit of measurement. Thus its values does not change if an appropriate constant value is either multiplied, divided, subtracted or added to each of the observations in the sample data. This adjustment reduces the magnitude of numerical values in the sample data and reduces computational time to calculate F value without any change.

**Example 12.1:** To test the significance of variation in the retail prices of a commodity in three principal cities, Mumbai, Kolkata, and Delhi, four shops were chosen at random in each city and the priceswho lack confidence in their mathematical ability observed in rupees were as follows:

| | | | | |
|---|---|---|---|---|
| Mumbai : | 16 | 8 | 12 | 14 |
| Kolkata : | 14 | 10 | 10 | 6 |
| Delhi : | 4 | 10 | 8 | 8 |

Do the data indicate that the price in the three cities are significantly different?

[*Jammu Univ., M.Com, 1997*]

**Solution:** Let us take the null hypothesis that there is no significant difference in the prices of a commodity in the three cities. Calculations for analysis of variance are as under.

| *Sample 1* | | *Sample 2* | | *Sample 3* | |
|---|---|---|---|---|---|
| *Mumbai* | | *Kolkata* | | *Delhi* | |
| $x_1$ | $x_1^2$ | $x_2$ | $x_2^2$ | $x_3$ | $x_3^2$ |
| 16 | 256 | 14 | 196 | 4 | 16 |
| 8 | 64 | 10 | 100 | 10 | 100 |
| 12 | 144 | 10 | 100 | 8 | 64 |
| 14 | 196 | 6 | 36 | 8 | 64 |
| $\Sigma x_1 = 50$ | $\Sigma x_1^2 = 660$ | $\Sigma x_2 = 40$ | $\Sigma x_2^2 = 432$ | $\Sigma x_3 = 30$ | $\Sigma x_3^2 = 244$ |

There are $r$=3 treatments (samples) with $n_1$=4, $n_2$=4, $n_3$=4, and $n$=12.

$$\text{T} = \text{Sum of all the observations in the three samples}$$
$$= \Sigma x_1 + \Sigma x_2 + \Sigma x_3 = 50 + 40 + 30 = 120$$

$$\text{CF} = \text{Correction factor} = \frac{\text{T}^2}{n} = \frac{(120)^2}{12} = 1200$$

$$\text{SST} = \text{Total sum of the squares}$$
$$= (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2) - \text{CF} = (660 + 432 + 244) - 1200$$
$$= 136$$

$$\text{SSTR} = \text{Sum of squares between the samples}$$
$$= \left(\frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3}\right) - \text{CF}$$
$$= \left\{\frac{(50)^2}{4} + \frac{(40)^2}{4} + \frac{(30)^2}{4}\right\} - 1200$$
$$= \left\{\frac{2500}{4} + \frac{1600}{4} + \frac{900}{4}\right\} - 1200 = \frac{5000}{4} - 1200 = 50$$

$$\text{SSE} = \text{SST} - \text{SSTR} = 136 - 50 = 86$$

Degrees of freedom: $df_1 = r - 1 = 3 - 1 = 2$ and $df_2 = n - r = 12 - 3 = 9$

Thus $\text{MSTR} = \frac{\text{SSTR}}{df_1} = \frac{50}{2} = 25$ and $\text{MSE} = \frac{\text{SSE}}{df_2} = \frac{86}{9} = 9.55$

**Table 12.5: ANOVA Table**

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Squares* | *Test-Statistic* |
|---|---|---|---|---|
| • Between samples | 50 | 2 | 25 | $\text{F} = \frac{25}{9.55}$ |
| • Within samples | 86 | 9 | 9.55 | = 2.617 |
| Total | 136 | 11 | | |

The table value of F for $df_1 = 2$, $df_2 = 9$, and $\alpha = 5$ per cent level of significance is 4.26. Since calculated value of F is less than its critical (or table) value, the null hypothesis is accepted. Hence we conclude that the prices of a commodity in three cities have no significant difference.

**Example 12.2:** A study investigated the perception of corporate ethical values among individuals specializing in marketing. Using $\alpha = 0.05$ and the following data (higher scores indicate higher ethical values), test for significant differences in perception among three groups.

| *Marketing Manager* | *Marketing Research* | *Advertising* |
|---|---|---|
| 6 | 5 | 6 |
| 5 | 5 | 7 |
| 4 | 4 | 6 |
| 5 | 4 | 5 |
| 6 | 5 | 6 |
| 4 | 4 | 6 |

**Solution:** Let us assume the null hypothesis that there is no significant difference in ethical values among individuals specializing in marketing. Calculations for analysis of variance are as under:

| *Marketing Manager* | | *Marketing Research* | | *Advertising* | |
|---|---|---|---|---|---|
| $x_1$ | $x_1^2$ | $x_2$ | $x_2^2$ | $x_3$ | $x_3^2$ |
| 6 | 36 | 5 | 25 | 6 | 36 |
| 5 | 25 | 5 | 25 | 7 | 49 |
| 4 | 16 | 4 | 16 | 6 | 36 |
| 5 | 25 | 4 | 16 | 5 | 25 |
| 6 | 36 | 5 | 25 | 6 | 36 |
| 4 | 16 | 4 | 16 | 6 | 36 |
| 30 | 154 | 27 | 123 | 36 | 218 |

There are $r = 3$ treatments (samples) with $n_1 = n_2 = n_3 = 6$ and $n = 18$.

$$\text{T} = \text{Sum of all the observations in three samples}$$
$$= \Sigma x_1 + \Sigma x_2 + \Sigma x_3 = 30 + 27 + 36 = 93$$

$$\text{CF} = \text{Correction factor} = \frac{\text{T}^2}{n} = \frac{(93)^2}{18} = 480.50$$

$$\text{SST} = \text{Total sum of the squares}$$
$$= (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2) - \text{CF} = (154 + 123 + 218) - 480.50$$
$$= 14.50$$

$$\text{SSTR} = \text{Sum of squares between the samples}$$
$$= \left(\frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3}\right) - \text{CF}$$
$$= \left\{\frac{(30)^2}{6} + \frac{(27)^2}{6} + \frac{(36)^2}{6}\right\} - 480.50$$
$$= \left(\frac{900}{6} + \frac{729}{6} + \frac{1296}{6}\right) - 480.50$$
$$= (150 + 121.5 + 216) - 480.50 = 7$$

$$\text{SSE} = \text{SST} - \text{SSTR} = 14.50 - 7 = 7.50$$

Degrees of freedom: $df_1 = r - 1 = 3 - 1 = 2$ and $df_2 = n - r = 18 - 3 = 15$

Thus $$\text{MSTR} = \frac{\text{SSTR}}{df_1} = \frac{7}{2} = 3.5 \text{ and MSE} = \frac{\text{SSE}}{df_2} = \frac{7.50}{15} = 0.5$$

**Table 12.6: ANOVA Table**

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Squares* | *Test-Statistic* |
|---|---|---|---|---|
| • Between samples | 7 | 2 | 3.5 | $F = \frac{3.5}{0.5}$ |
| • Within samples | 7.5 | 15 | 0.5 | = 7 |
| Total | 14.5 | 17 | | |

The table value of F for $df_1 = 2$, $df_2 = 15$, and $\alpha = 0.05$ is 3.68. Since calculated value of F=7 is more than its table value, the null hypothesis is rejected. Hence we conclude that there is significant difference in ethical values among individuals specializing in marketing.

**Example 12.3:** As head of the department of a consumer's research organization, you have the responsibility for testing and comparing lifetimes of four brands of electric bulbs. Suppose you test the life-time of three electric bulbs of each of the four brands. The data are shown below, each entry representing the lifetime of an electric bulb, measured in hundreds of hours:

| *Brand* | | | |
|---|---|---|---|
| *A* | *B* | *C* | *D* |
| 20 | 25 | 24 | 23 |
| 19 | 23 | 20 | 20 |
| 21 | 21 | 22 | 20 |

Can we infer that the mean lifetimes of the four brands of electric bulbs are equal?

*[Roorkee Univ., MBA, 2000]*

**Solution:** Let us take the null hypothesis that the mean lifetime of the four brands of electric bulbs is equal.

Substracting a common figure 20 from each observation. The calculations with code data are as under:

| *Sample 1(A)* | | *Sample 2(B)* | | *Sample 3(C)* | | *Sample 4(D)* | |
|---|---|---|---|---|---|---|---|
| $x_1$ | $x_1^2$ | $x_2$ | $x_2^2$ | $x_3$ | $x_3^2$ | $x_4$ | $x_4^2$ |
| 0 | 0 | 5 | 25 | 4 | 16 | 3 | 9 |
| –1 | 1 | 3 | 9 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 2 | 4 | 0 | 0 |
| 0 | 2 | 9 | 35 | 6 | 20 | 3 | 9 |

$$T = \text{Sum of all the observations in four samples}$$
$$= \Sigma x_1 + \Sigma x_2 + \Sigma x_3 + \Sigma x_4 = 0 + 9 + 6 + 3 = 18$$

$$CF = \text{Correction factor} = \frac{T^2}{n} = \frac{(18)^2}{12} = 27$$

$$SST = \text{Total sum of the squares}$$
$$= (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2) - CF$$
$$= (2 + 35 + 20 + 9) - 27 = 39$$

$$SSTR = \text{Sum of squares between the samples}$$
$$= \left\{\frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} + \frac{(\Sigma x_4)^2}{n_4}\right\} - CF$$
$$= \left\{0 + \frac{(9)^2}{3} + \frac{(6)^2}{3} + \frac{(3)^2}{3}\right\} - 18 = (0 + 27 + 12 + 3) - 27 = 15$$

$$SSE = SST - SSTR = 39 - 15 = 24$$

Degrees of freedom: $df_1 = r - 1 = 4 - 1 = 3$, $df_2 = n - r = 12 - 4 = 8$. Thus

$$\text{MSTR} = \frac{\text{SSTR}}{df_1} = \frac{15}{3} \quad \text{and} \quad \text{MSE} = \frac{\text{SSE}}{df_2} = \frac{24}{8} = 3$$

**Table 12.7: ANOVA Table**

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Squares* | *Test Statistic* |
|---|---|---|---|---|
| • Between samples | 15 | 3 | 5 | $F = \frac{5}{3} = 1.67$ |
| • Within samples | 24 | 8 | 3 | |
| Total | 39 | 11 | | |

The table value of F for $df_1 = 3$, $df_2 = 8$, and $\alpha = 0.05$ is 4.07. Since the calculated value of F = 1.67 is less than its table value, the null hypothesis is accepted. Hence we conclude that the difference in the mean lifetime of four brands of bulbs is not significant and we infer that the average lifetime of the four brands of bulbs is equal.

## 12.4 INFERENCES ABOUT POPULATION (TREATMENT) MEANS

When null hypothesis $H_0$ is rejected, it implies that all population means are not equal. However, we may not be satisfied with this conclusion and may want to know which population means differ. The answer to this question comes from the construction of confidence intervals using the small sample procedures, based on t-distribution.

For a single populatin mean, $\mu$ the confidence interval is given by

$$\bar{x} \pm t_{\alpha/2}(s/\sqrt{n})$$

where $\bar{x}$ is the sample mean from a population. Similarly, confidence interval for the difference between two population means $\mu_1$ and $\mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $\bar{x}_1$ and $\bar{x}_2$ = mean of sample population 1 and 2 respectively;

$n_1$ and $n_2$ = number of observations in sample 1 and 2 respectively.

To use these confidence intervals, we need to know

- How to calculate $s$ or $s^2$, which is the best estimate of the common sample variance?
- How many degrees of freedom are used for the critical value of $t$-test statistic?

To answer these questions, recall that in the analysis of variance, we assume that the population variances are equal for all populations. This common value is the **mean square error** MSE = SSE/$(n - r)$ which provides an unbiased estimate of $\sigma^2$, regardless of test or estimation used. We use, $s^2$ = MSE or $s = \sqrt{\text{MSE}} = \sqrt{\text{SSE}/(n-r)}$ with $df = (n - r)$ and $t_{\alpha/2}$ at specified level of significance $\alpha$ to estimate $\sigma^2$, where $n = n_1 + n_2 + \ldots + n_r$.

**Mean square error (MSE):** The mean of the squared errors used to judge the quality of a set of errors.

**Illustration**

From Example 12.1 we know that, $\bar{x}_1 = 5$; $\bar{x}_3 = 6$, $n = n_1 + n_2 + n_3 = 18$, $s^2 = MSE = 0.5$ and $\alpha = 0.05$ level of signifiance. So the confidence interval is computed as:

$$(\bar{x}_1 - \bar{x}_3) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = (5-6) \pm 2.131\sqrt{0.5\left(\frac{1}{6} + \frac{1}{6}\right)}$$

$$= -1 \pm 1.225 = -2.225 \text{ and } 0.225$$

*Since zero is included in this interval, we may conclude that there is no significant difference in the selected population means*. That is, there is no difference between eithical values of marketing and advertising managers.

**Remark:** If the end points of the confidence interval have the same sign, then we may conclude that there is a significant difference between the selected population means.

## Self-Practice Problems 12A

**12.1** Kerala Traders Co. Ltd., 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 the three salesmen:

| | | | | | |
|---|---|---|---|---|---|
| A : | 300 | 400 | 300 | 500 | 0 |
| B : | 600 | 300 | 300 | 400 | — |
| C : | 700 | 300 | 400 | 600 | 500 |

Perform the analysis of variance and draw you conclusions.

*[Madras Univ., MCom, 1996; Madurai Univ., MCom, 1996]*

**12.2** There are three main brands of a certain powder. A sample of 120 packets sold is examined and found to be allocated among four groups A, B, C, and D, and brands I, II and III, as shown below:

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

Is there any significant difference in brand preferences?

**12.3** An agriculture research organization wants to study the effect of four types of fertilizers on the yield of a crop. It divided the entire field into 24 plots of land and used fertilizer at random in 6 plots of land. Part of the calculations are shown below:

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Squares* | *F-Test Statistic* |
|---|---|---|---|---|
| • Fertilizers | 2940 | 3 | — | 5.99 |
| • Within groups | — | — | — | |
| Total | 6212 | 24 | | |

(a) Fill in the blanks in the ANOVA table.

(b) Test at $\alpha = 0.05$, whether the fertilizers differ significantly

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

| *Observations* | $A_1$ | $A_2$ | $A_3$ |
|---|---|---|---|
| 1 | 25 | 31 | 24 |
| 2 | 30 | 39 | 30 |
| 3 | 36 | 38 | 28 |
| 4 | 38 | 42 | 25 |
| 5 | 31 | 35 | 28 |

Use analysis of variance and determine whether the machines are significantly different in their mean speed.

**12.5** The following figures related to the number of units of a product sold in five different areas by four salesmen:

| *Area* | *Number of units* | | | |
|---|---|---|---|---|
| | *A* | *B* | *C* | *D* |
| 1 | 80 | 100 | 95 | 70 |
| 2 | 82 | 110 | 90 | 75 |
| 3 | 88 | 105 | 100 | 82 |
| 4 | 85 | 115 | 105 | 88 |
| 5 | 75 | 90 | 80 | 65 |

Is there a significant difference in the efficiency of these salesmen? *[Osmania Univ., MBA, 1998]*

**12.6** Four machines A, B, C, and D are used to produce a certain kind of cotton fabric. Samples of size 4 with each unit as 100 square metres are selected from the outputs of the machines at random, and the number of flowers in each 100 square metres are counted, with the following results:

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

Do you think that there is significant difference in the performance of the four machines?

*[Kumaon Univ., MBA, 1998]*

## Hints and Answers

**12.1** Let $H_0$ : No difference in average sales of three salesmen.

Divide each observation by 100 and use the code data for analysis of variance.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Test Statistic |
|---|---|---|---|---|
| • Between Samples | 10 | 2 | 5 | $F = \frac{5}{2.73} = 1.83$ |
| • Within Samples | 30 | 11 | 2.73 | |
| Total | 40 | 13 | | |

Since the calculated value of F = 1.83 is less than its table value F = 3.98 at $df_1 = 2, df_2 = 11$, and $\alpha = 0.05$, the null hypothesis is accepted.

**12.2** Let $H_0$ : There is no significant difference in brand preference.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Test Statistic |
|---|---|---|---|---|
| • Between Samples | 168.5 | 2 | 84.25 | $F = \frac{84.25}{22.83} = 3.69$ |
| • Within Samples | 205.5 | 9 | 22.83 | |
| Total | 374.0 | 11 | | |

Since calculated value of F = 3.69 is less than its table value F = 4.26 at $df_1 = 2, df_2 = 9$, and $\alpha = 0.05$, the null hypothesis is accepted.

**12.3** Given total number of observations, $n = 24$; Number of samples, $r = 4$

$df = n - 1 = 24 - 1 = 23$ (For between the groups–fertilizers)

$df_1 = r - 1 = 4 - 1 = 3$; $df_2 = n - r = 24 - 4 = 20$ (For within the groups)

SSTR = 2940;

SSE = SST – SSB = 6212 – 2940 = 3272

$$\text{MSTR} = \frac{\text{SSTR}}{df_1} = \frac{2940}{3} = 980;$$

$$\text{MSE} = \frac{\text{SSE}}{df_2} = \frac{3272}{20} = 163.6$$

$$\therefore \quad F = \frac{\text{MSTR}}{\text{MSE}} = \frac{980}{163.6} = 5.99$$

Since the calculated value of F = 5.99 is more than its table value F = 3.10 at $df_1 = 3, df_2 = 20$, and $\alpha = 0.05$, the null hypothesis is rejected.

**12.4** Let $H_0$ : Machines are not significantly different in their mean speed.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Test Statistic |
|---|---|---|---|---|
| • Between Samples | 250 | 2 | 125 | $F = \frac{125}{16.66} = 7.50$ |
| • Within Samples | 200 | 12 | 16.66 | |
| Total | 450 | 14 | | |

Since the calculated value of F = 7.50 is more than its table value F = 3.89 at $df_1 = 2, df_2 = 12$, and $\alpha = 0.05$, the null hypothesis is rejected.

**12.5** Let $H_0$ : No significant difference in the performance of four salesmen.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Test Statistic |
|---|---|---|---|---|
| • Between Samples | 2340 | 3 | 780 | $F = \frac{780}{73.5} = 7.50$ |
| • Within Samples | 1176 | 16 | 73.5 | |
| Total | 3516 | 19 | | |

Since the calculated value of F = 10.61 is greater its table value F = 3.24 at $df_1 = 3$, $df_2 = 12$, and $\alpha = 0.05$, the null hypothesis is rejected.

**12.6** Let $H_0$ : Machines do not differ significantly in performance.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Test Statistic |
|---|---|---|---|---|
| • Between Samples | 540.69 | 3 | 180.23 | $F = \frac{180.23}{7.15} = 7.50$ |
| • Within Samples | 85.75 | 12 | 7.15 | |
| Total | 626.44 | 15 | | |

Since the calculated value of F = 25.207 is more than its table value F = 5.95 at $df_1 = 3$, $df_2 = 12$, and $\alpha = 0.05$, the null hypothesis is rejected.

## 12.5 TESTING EQUALITY OF POPULATION (TREATMENT) MEANS: TWO-WAY CLASSIFICATION

In one-way ANOVA, the partitioning of the total variation in the sample data is done into two components: (i) Variation among the samples due to different samples (or treatments) and (ii) Variation within the samples due to random error. However, there might be a possibility that some of the variation left in the random error from one-way analysis of variation was not due to random error or chance but due to some other measurable factor. For instance, in Example 12.1 we might feel that part of the variation in price was due to the inability in data collection or condensation of data. If so, this accountable variation was deliberately included in the sum of squares for error (SSE) and therefore caused the mean sum of squares for error (MSE) to be little large. Consequently, F-Value would then be small and responsible for the rejection of null hypothesis.

**Two-way analysis of variance:** Analysis of variance in which two criteria (or variables) are used to analyse the difference between more than two population means.

The **two-way analysis of variance** can be used to

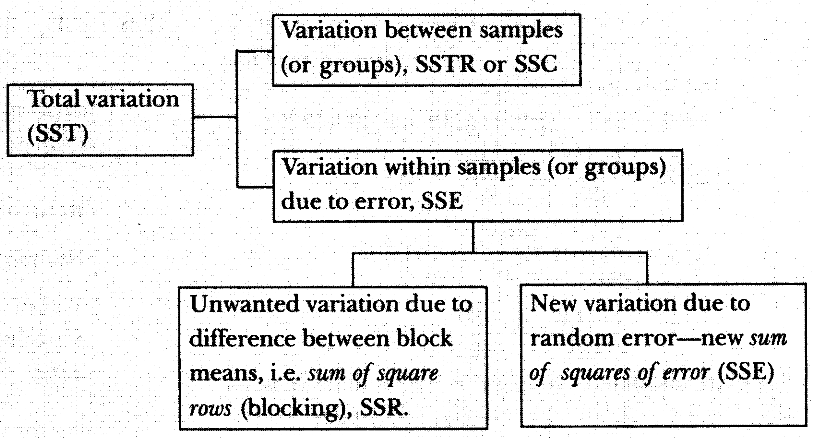
- explore one criterion (or factor) of interest to partition the sample data so as to remove the unaccountable variation, and arriving at a true conclusion.
- investigate two criteria (factors) of interest for testing the difference between sample means.
- consider any interaction between two variables.

**Blocking:** The removal of a source of variation from the error term in the analysis of variance.

In two-way analysis of variance we are introducing another term called *'blocking variable'* to remove the undesirable accountable variation. A block variable is the variable that the researcher wants to control but is not the treatment variable of interest. The term '**blocking**' refers to block of land and comes from agricultural origin. The 'block' of land might make some difference in the study of growth pattern of varieties of seeds for a given type of land. R. A. Fisher designated several different plots of land as blocks, which he controlled as a second variable. Each of the seed varieties were planted on each of the blocks. The main aim of his study was to compare the seed varieties (independent variable). He wanted only to control the difference in plots of land (blocking variable).For instance, in Example 12.1, each set of three prices in three cities under a given condition would constitute a 'block' of sample data. 'Blocking' is an extension of the idea of pairing observations in hypothesis testing. Blocking provides the opportunity for one-to-one comparison of prices, where any observed difference cannot be due to difference among blocking variables.

To ensure a right conclusion to be reached, each sample data (group) should be measured under the same conditions by removing variations due to these conditions by the use of a blocking factor.

The partitioning of total variation in the sample data is shown below:




The general ANOVA table for $c$ samples (columns), $r$ blocks, and number of observations $n$ is shown in Table 12.8.

**Table 12.8: General ANOVA Table for Two-way Classification**

| *Source of Variation* | *Sum of Square* | *Degrees of Freedom* | *Mean Square* | *Test Statistic* |
|---|---|---|---|---|
| • Between columns | SSTR | $c-1$ | MSTR = SSTR/($c-1$) | $F_{treatment}$ = MSTR/MSE |
| • Between rows | SSR | $r-1$ | MSR = SSR/($r-1$) | $F_{blocks}$ = MSR/MSE |
| • Residual error | SSE | $(c-1)(r-1)$ | MSE = SSE/($c-1$) ($r-1$) | |
| Total | SST | $n-1$ | | |

As stated above, total variation consists of three parts: (i) variation between columns, SSTR; (ii) variation between rows, SSR; and (iii) actual variation due to random error, SSE. That is

$$SST = SSTR + (SSR + SSE)$$

The degrees of freedom associated with SST are $cr - 1$, where $c$ and $r$ are the number of columns and rows, respectively

Degrees of freedom between columns $= c - 1$

Degrees of freedom between rows $= r - 1$

Degrees of freedom for residual error $= (c - 1)(r - 1) = N - n - c + 1$

**Randomized block design:** A two-way analysis of variance designed to eliminate any assignable variation from the analysis.

The test-statistic F for analysis of variance is given by

$$F_{treatment} = MSTR/MSE; \quad MSTR > MSE \quad \text{or} \quad MSE/MSTR; \quad MSE > MSTR$$

$$F_{blocks} = MSR/MSE; \quad MSR > MSE \quad \text{or} \quad MSE/MSR; \quad MSE > MSR$$

### Decision rule

- If $F_{cal} < F_{table}$, accept null hypothesis $H_0$
- Otherwise reject $H_0$

**Example 12.5:** The following table gives the number of refrigerators sold by 4 salesmen in three months May, June and July:

| *Month* | *Salesman* | | | |
|---|---|---|---|---|
| | *A* | *B* | *C* | *D* |
| May | 50 | 40 | 48 | 39 |
| June | 46 | 48 | 50 | 45 |
| July | 39 | 44 | 40 | 39 |

Is there a significant difference in the sales made by the four salesmen? Is there a significant difference in the sales made during different months?

[*Delhi Univ., MCom, 1998*]

**Solution:** Let us take the null hypothesis that there is no significant difference between sales made by the four salesmen during different months. The given data are coded by subtracting 40 from each observation. Calculations for a two-criteria—month and salesman—analysis of variance are shown in Table 12.9.

**Table 12.9: Two-way ANOVA Table**

| *Month* | *Salesman* | | | | | | | | *Row* |
|---|---|---|---|---|---|---|---|---|---|
| | $A(x_1)$ | $x_1^2$ | $B(x_2)$ | $x_2^2$ | $C(x_3)$ | $x_3^2$ | $D(x_4)$ | $x_4^2$ | *Sum* |
| May | 10 | 100 | 0 | 0 | 8 | 64 | – 1 | 1 | 17 |
| June | 6 | 36 | 8 | 64 | 10 | 100 | 5 | 25 | 29 |
| July | – 1 | 01 | 4 | 16 | 0 | 0 | – 1 | 1 | 2 |
| Column sum | 15 | 137 | 12 | 80 | 18 | 164 | 3 | 27 | 48 |

T = Sum of all observations in three samples of months = 48

$$CF = \text{Correction factor} = \frac{T^2}{n} = \frac{(48)^2}{12} = 192$$

$$SSTR = \text{Sum of squares between salesmen (columns)}$$

$$= \left\{\frac{(15)^2}{3} + \frac{(12)^2}{3} + \frac{(18)^2}{3} + \frac{(3)^2}{3}\right\} - 192$$

$$= (75 + 48 + 108 + 3) - 192 = 42$$

$$SSR = \text{Sum of squares between months (rows)}$$

$$= \left\{\frac{(17)^2}{4} + \frac{(29)^2}{4} + \frac{(2)^2}{4}\right\} - 192$$

$$= (72.25 + 210.25 + 1) - 192 = 91.5$$

$$\text{SST} = \text{Total sum of squares}$$
$$= (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 + \Sigma x_4^2) - \text{CF}$$
$$= (137 + 80 + 164 + 27) - 192 = 216$$
$$\text{SSE} = \text{SST} - (\text{SSC} + \text{SSR}) = 216 - (42 + 91.5) = 82.5$$

The total degrees of freedom are, $df = n - 1 = 12 - 1 = 11$. So

$df_c = c - 1 = 4 - 1 = 3,\ df_r = r - 1 = 3 - 1 = 2;\ df = (c - 1)(r - 1) = 3 \times 2 = 6$

Thus $\text{MSTR} = \text{SSTR}/(c - 1) = 42/3 = 14,$

$\text{MSR} = \text{SSR}/(r - 1) = 91.5/2 = 45.75$

$\text{MSE} = \text{SSE}/(c - 1)(r - 1) = 82.5/6 = 13.75$

The ANOVA table is shown in Table 12.10.

**Table 12.10: Two-way ANOVA Table**

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Squares* | *Variance Ratio* |
|---|---|---|---|---|
| • Between salesmen | 42.0 | 3 | 14.00 | $F_{treatment} = 14/13.75 = 1.018$ |
| • Between months | 91.5 | 2 | 45.75 | $F_{block} = 45.75/13.75 = 3.327$ |
| • Residual error | 82.5 | 6 | 13.75 | |
| Total | 216 | 11 | | |

(a) The table value of F = 4.75 for $df_1 = 3$, $df_2 = 6$, and $\alpha = 0.05$. Since the calculated value of $F_{treatment} = 1.018$ is less than its table value, the null hypothesis is accepted. Hence we conclude that sales made by the salesmen do not differ significantly

(b) The table value of F = 5.14 for $df_1 = 2$, $df_2 = 6$, and $\alpha = 0.05$. Since the calculated value of $F_{block} = 3.327$ is less than its table value, the null hypothesis is accepted. Hence we conclude that sales made during different months do not differ significantly.

**Example 12.6:** To study the performance of three detergents and three different water temperatures, the following 'whiteness' readings were obtained with specially designed equipment:

| *Water Temperature* | *Detergent A* | *Detergent B* | *Detergent C* |
|---|---|---|---|
| Cold water | 57 | 55 | 67 |
| Warm water | 49 | 52 | 68 |
| Hot water | 54 | 46 | 58 |

Perform a two-way analysis of variance, using 5 per cent level of significance.

[*Osmania Univ., MBA, 1998*]

**Solution:** Let us take the null hypothesis that there is no significant difference in the performance of three detergents due to water temperature and vice-versa. The data are coded by subtracting 50 from each observation. The data in coded form are in Table 12.11:

**Table 12.11: Coded Data**

| *Water Temperature* | *Detergents* $A(x_1)$ | $x_1^2$ | $B(x_2)$ | $x_2^2$ | $C(x_3)$ | $x_3^2$ | *Row Sum* |
|---|---|---|---|---|---|---|---|
| Cold water | + 7 | 49 | + 5 | 25 | + 17 | 289 | 29 |
| Warm water | − 1 | 01 | + 2 | 04 | + 18 | 324 | 19 |
| Hot water | + 4 | 16 | − 4 | 16 | + 8 | 64 | 8 |
| Column sum | 10 | 66 | 3 | 45 | 43 | 677 | 56 |

$$\text{T} = \text{Sum of all observations in three samples of detergents} = 56$$

$$\text{CF} = \text{Correction factor} = \frac{\text{T}^2}{n} = \frac{(56)^2}{9} = 348.44$$

$$\text{SSTR} = \text{Sum of squares between detergents (columns)}$$

$$= \left\{\frac{(10)^2}{3} + \frac{(3)^2}{3} + \frac{(43)^2}{3}\right\} - \text{CF}$$

$$= 33.33 + 3 + 616.33 - 348.44 = 304.22$$

$$\text{SSR} = \text{Sum of squares between water temperature (rows)}$$

$$= \left\{\frac{(29)^2}{3} + \frac{(19)^2}{3} + \frac{(8)^2}{3}\right\} - \text{CF}$$

$$= (280.33 + 120.33 + 21.33) - 348.44 = 73.55$$

$$\text{SST} = \text{Total sum of squares}$$

$$= (\Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2) - \text{CF} = (66 + 45 + 677) - 348.44 = 439.56$$

$$\text{SSE} = \text{SST} - (\text{SSC} + \text{SSR}) = 439.56 - (304.22 + 73.55) = 61.79$$

Thus

$$\text{MSTR} = \text{SSTR}/(c - 1) = 304.22/2 = 152.11;$$

$$\text{MSR} = \text{SSR}/(r - 1) = 73.55/2 = 36.775$$

$$\text{MSE} = \text{SSE}/(c - 1)(r - 1) = 61.79/4 = 15.447$$

**Table 12.12: Two-way ANOVA Table**

| *Source of Variation* | *Sum of Squares* | *Degree of Freedom* | *Mean Squares* | *Variance Ratio* |
|---|---|---|---|---|
| Between detergents (columns) | 304.22 | 2 | 152.110 | $F_{treatment} = 152.11/15.447 = 9.847$ |
| Between temp. (rows) | 73.55 | 2 | 36.775 | $F_{block} = 36.775/15.447 = 2.380$ |
| Residual error | 61.79 | 4 | 15.447 | |
| Total | 439.56 | 8 | | |

(a) Since calculated value of $F_{treatment} = 9.847$ at $df_1 = 2$, $df_2 = 4$, and, $\alpha = 0.05$ is greater than its table value F = 6.94, the null hypothesis is rejected. Hence we conclude that there is significant difference between the performance of the three detergents.

(b) Since the calculated value of $F_{block} = 2.380$ at $df_1 = 2$, $df_2 = 4$, and $\alpha = 0.05$ is less than its table value F = 6.94, the null hypothesis is accepted. Hence we conclude that the water temperature do not make a significant difference in the performance of the detergent.

## Conceptual Questions 12A

1. What are some of the criteria used in the selection of a particular hypothesis testing procedure?
2. What are the major assumptions of ANOVA?
3. Under what conditions should the one-way ANOVA F-test be selected to examine the possible difference in the means of independent populations?
4. How is analysis of variance technique helpful in solving business problems? Illustrate your answer with suitable examples. *[Kumaon Univ., MBA, 2000]*
5. Distinguish between one-way and two-way classifications to test the equality of population means.
6. What is meant by the term analysis of variance? What types of problems are solved using ANOVA? Explain.
7. Describe the procedure for performing the test of hypothesis in the analysis of variance. What is the basic assumption underlying this test?
8. What is meant by the critical value used in the analysis of variance? How is it found?
9. How is the F-distribution related to the student's *t*-distribution and the chi-square distribution? What important hypothesis can be tested by the F-distribution?

10. Discuss the components of total variation when samples are selected in blocks.
11. Define the terms treatment, error – 'with in' 'between' and the context in which these are used.
12. Explain the sum-of-square principle.
13. Explain how the total deviation is partitioned into the treatment deviation and the error deviation.
14. Does the quantity MSTR/MSE follow an F-distribution when the null hypothesis of ANOVA is false? Explain.

## Self-Practice Problems 12B

**12.7** 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:

| *Season* | *Salesman* | | | | *Total* |
|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | |
| Summer | 36 | 36 | 21 | 35 | 128 |
| Winter | 28 | 29 | 31 | 32 | 120 |
| Monsoon | 26 | 28 | 29 | 29 | 112 |
| Totals | 90 | 93 | 81 | 96 | 360 |

(a) Do the salesmen significantly differ in performance?
(b) Is there significant difference between the seasons?
[*Calcutta Univ., MCom, 1996; Calcultta Univ., MCom, 1998*]

**12.8** Perform a two-way ANOVA on the data given below:

| *Plots of Land* | *Treatment* | | | |
|---|---|---|---|---|
| | *A* | *B* | *C* | *D* |
| 1 | 38 | 40 | 41 | 39 |
| 2 | 45 | 42 | 49 | 36 |
| 3 | 40 | 38 | 42 | 42 |

Use the coding method for subtracting 40 from the given numbers. [*CA, May 1996*]

**12.9** 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 |

(a) Test whether the mean productivity is the same for the different machine types.
(b) Test whether the 5 men differ with respect to mean productivity. [*Madras Univ., MCom, 1997*]

**12.10** The following table gives the number of units of production per day turned out by four different types of machines:

| *Employees* | *Type of Machine* | | | |
|---|---|---|---|---|
| | $M_1$ | $M_2$ | $M_3$ | $M_4$ |
| $E_1$ | 40 | 36 | 45 | 30 |
| $E_2$ | 38 | 42 | 50 | 41 |
| $E_3$ | 36 | 30 | 48 | 35 |
| $E_4$ | 46 | 47 | 52 | 44 |

Using analysis of variance (a) test the hypothesis that the mean production is same for four machines and (b) test the hypothesis that the employees do not differ with respect to mean productivity.

[*Osmania Univ., MCom, 1999*]

**12.11** In a certain factory, production can be accomplished by four different workers on five different types of machines. A sample study, in the context of a two-way design without repeated values, 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 while analysing the gathered data as under:

(a) Sum of squares for variance between machines = 35.2
(b) Sum of squares for variance between workmen = 53.8
(c) Sum of squares for total variance = 174.2
Set up ANOVA table for the given information and draw the inference about variance at 5 per cent level of significance.

**12.12** Apply the technique of analysis of variance of the following data showing the yields of 3 varieties of a crop each from 4 blocks, and test whether the average yields of the varieties are equal or not. Also test equality of the block means

| *Varieties* | *Blocks* | | | |
|---|---|---|---|---|
| | *I* | *II* | *III* | *IV* |
| A | 4 | 8 | 6 | 8 |
| B | 5 | 5 | 7 | 8 |
| C | 6 | 7 | 9 | 5 |

**12.13** Three varieties of potato are planted each on four plots of land of the same size and type, each variety is treated with four different fertilizers. The yield in tonnes are as follows:

| Fertilizer | Variety | | |
|---|---|---|---|
| | $V_1$ | $V_2$ | $V_3$ |
| $F_1$ | 164 | 172 | 174 |
| $F_2$ | 155 | 157 | 147 |
| $F_3$ | 159 | 166 | 158 |
| $F_4$ | 158 | 157 | 153 |

Perform an analysis of variance and show whether (a) there is any significant difference between the average yield of potatoes due to different fertilizers being different used, and (b) there is any difference in the average yield of potatoes of different varieties.

## Hints and Answers

**12.7** Let $H_0$ : No significant difference between sales by salesmen and that of seasons.

Decoding the data by subtracting 30 from each figure.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Variance Ratio |
|---|---|---|---|---|
| • Between salemen (column) | 42 | 3 | 14 | $F_1 = \frac{22.67}{14} = 1.619$ |
| • Between seasons (row) | 32 | 2 | 16 | $F_2 = \frac{22.67}{16} = 1.417$ |
| • Residual error | 136 | 6 | 22.67 | |
| Total | 210 | 11 | | |

- Since $F_1 = 1.619 < F_{0.05\,(6,\,3)} = 4.76$, accept null hypothesis.
- Since $F_2 = 1.417 < F_{0.05(6,\,2)} = 5.14$, accept null hypothesis.

**12.8**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Variance Ratio |
|---|---|---|---|---|
| • Between columns | 42 | 3 | 14 | $F_1 = \frac{14}{10.67} = 1.312$ |
| • Between rows | 26 | 2 | 13 | $F_2 = \frac{13}{10.67} = 1.218$ |
| • Residual error | 64 | 6 | 10.67 | |
| Total | 132 | 11 | | |

(a) $F_1 = 1.312 < F_{0.05\,(3,\,6)} = 4.76$, accept null hypothesis.
(b) $F_2 = 1.218 < F_{0.05\,(2,\,6)} = 5.14$, accept null hypothesis.

**12.9** Let $H_0$ :(a) Mean productivity is same for all machines
(b) Men do not differ with respect to mean productivity
Decoding the data by subtracting 40 from each figure.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Variance Ratio |
|---|---|---|---|---|
| • Between machine type | 338.8 | 3 | 112.933 | $F_1 = \frac{112.933}{6.142} = 18.387$ |
| • Between workers | 161.5 | 4 | 40.375 | $F_2 = \frac{40.375}{6.142} = 6.574$ |
| • Residual error | 73.7 | 12 | 6.142 | |
| Total | 574 | 19 | | |

(a) $F_{0.05} = 3.49$ at $df_1 = 3$ and $df_2 = 12$. Since the calculated value $F_1 = 18.387$ is greater than the table value, the null hypothesis is rejected.
(b) $F_{0.05} = 3.26$ at $df_1 = 4$ and $df_2 = 12$. Since the calculated value $F_2 = 6.574$ is greater than the table value, the null hypothesis is rejected.

**12.10** Let $H_0$: (a) Mean production does not differ for all machines
(b) Employees do not differ with respect to mean productivity
Decoding the data by subtracting 40 from each figure.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Variance Ratio |
|---|---|---|---|---|
| • Between machine | 312.5 | 3 | 104.17 | $F_1 = \frac{104.17}{10.72} = 9.72$ |
| • Between employees | 1266.0 | 3 | 88.67 | $F_2 = \frac{88.67}{10.72} = 8.27$ |
| • Residual error | 96.5 | 9 | 10.72 | |
| Total | 675.0 | 15 | | |

(a) $F_{0.05} = 3.86$ at $df_1 = 3$ and $df_2 = 9$. Since the calculated value $F_1 = 9.72$ is more than its table value, reject the null hypothesis.
(b) $F_{0.05} = 3.86$ at $df_1 = 3$ and $df_2 = 9$. Since the calculated value $F_2 = 8.27$ is more than its table value, reject the null hypothesis

**12.11** Let $H_0$ :(a) Workers do not differ with respect to their mean productivity

(b) Mean productivity of all machines is the same

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Square* | *Variance Ratio* |
|---|---|---|---|---|
| • Between machine | 35.2 | 4 | 8.8 | $F_1 = \frac{8.8}{7.1} = 1.24$ |
| • Between workmen | 53.8 | 3 | 17.93 | $F_2 = \frac{17.93}{7.1} = 2.53$ |
| • Residual error | 85.2 | 12 | 7.1 | |
| Total | 174.2 | 19 | | |

(a) The calculated value of $F_1 = 1.24$ is less than its table value $F_{0.05} = 3.25$ at $df_1 = 4$ and $df_2 = 12$, hence the null hypothesis is accepted.

(b) The calculated value of $F_2 = 2.53$ is less than its table value $F_{0.05} = 3.49$ at $df_1 = 3$ and $df_2 = 12$, hence the null hypothesis accepted.

**12.12** Let $H_0$: (a) Mean yields of the varieties are equal

(b) Block means are equal

Decoding the data by subtracting 5 from each figure.

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Square* | *Variance Ratio* |
|---|---|---|---|---|
| • Between blocks | 9.67 | 3 | 3.22 | $F_1 = \frac{3.22}{2.80} = 1.15$ |
| • Between varieties | 0.5 | 2 | 0.25 | $F_2 = \frac{2.80}{0.25} = 11.20$ |
| • Residual error | 16.83 | 6 | 2.80 | |
| Total | 27.00 | 12 | | |

(a) Since the calculated value $F_1 = 1.15$ is less than its table value $F_{0.05} = 4.757$ at $df = (3, 6)$, the null hypothesis is accepted.

(b) Since the calculated value $F_2 = 11.20$ is less than its table value $F_{0.05} = 19.33$ at $df = (6, 2)$, the null hypothesis is accepted.

**12.13** Let $H_0$ :(a) No significant difference in the average yield of potatoes due to different fertilizers

(b) No significant difference in the average yield of the three varieties of potatoes

Decoding the data by subtracting 158 from each figure.

| *Source of Variation* | *Sum of Squares* | *Degrees of Freedom* | *Mean Square* | *Variance Ratio* |
|---|---|---|---|---|
| • Between varieties | 56 | 2 | 28 | $F_1 = \frac{28}{18} = 1.55$ |
| • Between fertilizers | 498 | 3 | 166 | $F_2 = \frac{166}{18} = 9.22$ |
| • Residual error | 108 | 6 | 18 | |
| Total | 662 | 11 | | |

(a) $F_{cal} = 1.55$ is less than its table value $F_{0.05} = 5.14$ at $df = (2, 6)$, the null hypothesis is accepted.

(b) $F_{cal} = 9.22$ is more than its table value $F_{0.05} = 4.67$ at $df = (3, 6)$, the null hypothesis is rejected.

## Formulae Used

1. One-way analysis of variance

• Grand sample mean

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{r} x_{ij}}{n}, n = n_1 + n_2 + \ldots + n_r$$

• Correction factor $CF = \frac{T^2}{n}$

• Total sum of squares

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{r}(x_{ij} - \bar{\bar{x}})^2 = \sum_i\sum_j x_{ij}^2 - CF$$

• Sum of squares of variations between samples due to treatment

$$SSTR = \sum_{j=1}^{r} n_j(\bar{x}_j - \bar{\bar{x}})^2 = \frac{1}{n_j}\sum_{j=1}^{r} x_j^2 - CF$$

- Sum of squares of variations within samples or error sum of squares

$$SSE = \sum_i^k \sum_j^r (x_{ij} - \bar{x}_i)^2 = SST - SSTR$$

- Mean square between samples due to treatments

$$MSTR = \frac{SSTR}{r-1}$$

- Mean square within samples due to error

$$MSE = \frac{SSE}{n-r}$$

- Test statistic for equality of $k$ population means

$$F = \frac{MSTR}{MSE}$$

- Degrees of freedom

  Total $df(n-1)$ = Treatment $df\,(r-1)$ + Random error $df\,(n-r)$

2. Two-way analysis of variance
   - Total sum of squares

$$SST = \sum_{j=1}^{r} \sum_{i=1}^{k} x_{ij}^2 - n\,(\bar{\bar{x}})^2$$

- Sum of squares of variances between columns due to treatments

$$SSTR = \sum_{j=1}^{r} (\bar{x}_j)^2 - n\,(\bar{\bar{x}})^2$$

- Sum of squares between rows due to blocks

$$SSR = \sum_{i=1}^{k} (\bar{x}_i)^2 - n\,(\bar{\bar{x}})^2$$

- Sum of squares due to error

$$SSE = SST - (SSTR + SSR)$$

- Degrees of freedom

$$df_c = c - 1; \quad df_r = (r-1)$$

  $df$ (residual error) = Blocks $df$ + Treatments $df$ $= (r-1)(c-1)$

- Mean squares between columns due to treatment

$$MSTR = \frac{SSTR}{c-1}$$

- Mean square between rows due to blocks

$$MSR = \frac{SSR}{r-1}$$

- Mean square of residual error

$$MSE = \frac{SSE}{(c-1)(r-1)}$$

- Test statistic

$$F_1 = \frac{MSTR}{MSE}; \quad F_2 = \frac{MSR}{MSE}$$

provided numerator is bigger than denominator.

# Chapter Concepts Quiz

## True or False

1. Analysis of variance serves as the basis for a variety of statistical models related to the design of experiments. (T/F)
2. Analysis of variance is used to test the hypothesis that the means of several populations do not differ. (T/F)
3. Analysis of variance is used to test the hypothesis that the variances of several populations do not differ. (T/F)
4. Analysis of variance cannot be used when samples are of unequal size. (T/F)
5. For analysis of variance, samples drawn from populations need not be independent. (T/F)
6. All populations need not be normally distributed to draw samples for analysis of variance. (T/F)
7. If samples of size $n$ each are drawn from $k$ normal populations, the degrees of freedom for variation within the samples is $n\,(k-1)$. (T/F)
8. Analysis of variance is an extension of the tests for differences between two means. (T/F)
9. In the analysis of variance the assumption that population variances are equal is called homogeneity of variance. (T/F)
10. The homogeneity of variance means that the test is concerned with the hypothesis that the several means came from the same population. (T/F)

## Multiple Choice

11. The number of parts in which total variance in a one-way analysis of variance partitioned is:
    (a) 2 (b) 3
    (c) 4 (d) none of these
12. The number of parts in which total variance in a two-way analysis of variance partitioned is:
    (a) 2 (b) 3
    (c) 4 (d) none of these
13. Any difference among the population means in the analysis of variance will inflate the expected value of
    (a) SSE (b) MSTR
    (c) MSE (d) all of these
14. If data in a two-way classification is displayed in $r$ rows and $c$ columns, then the degrees of freedom will be
    (a) $r-1$ (b) $c-1$
    (c) $r\,(c-1)$ (d) $(r-1)\,(c-1)$

15. The degrees of freedom between samples for $k$ samples of size $n$ will be
(a) $r - 1$ (b) $n - 1$
(c) $nr - 1$ (d) none of these
16. The degrees of freedom associated with the denominator of F-test in the analysis of variance are:
(a) $r(n - 1)$ (b) $n(r - 1)$
(c) $nr - 1$ (d) none of these
17. The sum of squares within a one-way analysis of variance is given by:
(a) SST + SSTR (b) SSTR − SST
(c) SST − SSTR (d) none of these
18. The total sum of squares in a two-way analysis of variance is given by
(a) $\Sigma\Sigma (x_{ij} - \bar{x}_j)^2$ (b) $\Sigma\Sigma x_{ij}^2 - n(\bar{\bar{x}})^2$
(c) $\Sigma\Sigma (x_{ij} - \bar{\bar{x}})^2$ (d) none of these
19. The error sum of squares can be obtained from the equation:
(a) SSE = SST + SSR + SSTR
(b) SSE = SSR + SSC − SST
(c) SSE = SST − SSR − SSTR
(d) none of these
20. The degrees of freedom for the error sum of squares are:
(a) $df_e = df_t - df_r - df_c$ (b) $df_e = df_r + df_c - df_t$
(c) $df_e = df_t + df_r + df_c$ (d) none of these
21. Which of these distributions has a pair of degrees of freedom
(a) Binomial (b) Chi-square
(c) Poisson (d) none of these
22. To test equality of proportions of more than two populations which of following techniques is used
(a) interval estimate
(b) analysis of variance
(c) chi-square test
(d) none of these
23. Which of the following assumptions of ANOVA can be discarded in case the sample size is large
(a) Each population has equal variance
(b) samples are drawn from a normal population
(c) both (a) and (b)
(d) none of these
24. Test statistic for equality of $r$ population means is
(a) MSTR/MSE (b) MSTR/MSR
(c) MSR/MSE (d) none of these

**Concepts Quiz Answers**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1. T | 2. T | 3. F | 4. T | 5. F | 6. F | 7. F | 8. T | 9. T |
| 10. T | 11. (a) | 12. (b) | 13. (b) | 14. (d) | 15. (a) | 16. (a) | 17. (c) | 18. (b) |
| 19. (c) | 20. (a) | 21. (d) | 22. (d) | 23. (b) | 24. (a) | | | |

## Review Self-Practice Problems

12.14 Complete the ANOVA table and determine the extent to which this information supports the claim that on an average there are no treatment differences:

| *Source* | *df* | *SS* | *MSS* | *F-value* |
|---|---|---|---|---|
| Between | 2 | — | 5 | |
| Within | — | 14 | — | — |
| Total | 9 | — | | |

12.15 A manager obtained the following data on the time (in days) needed to do a job. Use these data to test whether the mean time needed to complete a job differs for four persons. Use $\alpha = 0.05$.

| *Persons* | | *Time* | |
|---|---|---|---|
| 1 | 8 | 10 | 9 |
| 2 | 12 | 16 | 15 |
| 3 | 15 | 18 | 14 |
| 4 | 6 | 10 | 7 |

12.16 A leading oil company claims that its engine oil improves engine efficiency. To verify this claim, the company's brand A is compared with three other competing brands B, C, and D. The data of the survey consists of the km per litre consumption for a combination of city and highway travel, and are as follows:

| *Size of Container* | *Brand* | | | |
|---|---|---|---|---|
| | *A* | *B* | *C* | *D* |
| 1 | 36 | 34 | 33 | 35 |
| 2 | 29 | 26 | 28 | 27 |
| 3 | 25 | 24 | 25 | 23 |
| 4 | 19 | 20 | 18 | 18 |

(a) Is there any difference in the average mileage for these four brands?
(b) Is there any difference in the average mileage for a combination of city and highway travel?

12.17 A TV manufacturing company claims that the performance of its brand A TV set is better than two other brands. To verify this claim, a sample of 5 TV sets are selected from each brand and the frequency of repair during the first year of purchase is recorded. The results are as under:

| TV Brands | | |
|---|---|---|
| *A* | *B* | *C* |
| 4 | 7 | 4 |
| 6 | 4 | 6 |
| 7 | 3 | 6 |
| 5 | 6 | 3 |
| 8 | 5 | 1 |

In view of this data, can it be concluded that there is a significant difference between the three brands?

**12.18** Three varieties of coal were tested for ash content by five different laboratories. The results are as under:

| *Variety of Coal* | *Laboratory* | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| A | 9 | 7 | 6 | 5 | 8 |
| B | 7 | 4 | 5 | 4 | 5 |
| C | 6 | 5 | 6 | 7 | 6 |

In view of this data, can it be concluded that all three varieties of coal have an equal amount of ash content?

**12.19** An Insurance Company wants to test whether three of its field officers, A, B, and C in a given territory, meet equal number of prospective customers during a given period of time. A record of the previous four months showed the following results for the number of customers contacted by each field officer for each month:

| *Month* | *Field Officer* | | |
|---|---|---|---|
| | *A* | *B* | *C* |
| 1 | 8 | 6 | 14 |
| 2 | 9 | 8 | 12 |
| 3 | 11 | 10 | 18 |
| 4 | 12 | 4 | 8 |

Is there any significant difference in the average number of contacts made by the three field officers per month?

**12.20** A departmental store chain is considering opening a new store at one of three locations. An important factor in making such a decision is the household income in these areas. If the average income per household is similar, then the management can choose any one of these three locations. A random survey of various households in each location is undertaken and their annual combined income is recorded. This data is as under:

| *Annual Household Income (Rs '000s)* | | |
|---|---|---|
| *Area 1* | *Area 2* | *Area 3* |
| 70 | 100 | 60 |
| 72 | 110 | 65 |
| 75 | 108 | 57 |
| 80 | 112 | 84 |
| 83 | 113 | 84 |
| — | 120 | 70 |
| — | 100 | — |

Can the average income per household in these areas be considered to be the same?

**12.21** Four types of advertising displays were set up in twelve retail outlets, with three outlets randomly assigned to each of the displays. The data on product sales according to the advertising displays are as under:

| *Types of Display* | | | |
|---|---|---|---|
| *A* | *B* | *C* | *D* |
| 40 | 53 | 48 | 48 |
| 44 | 54 | 38 | 61 |
| 43 | 59 | 46 | 47 |

Does the type of advertising display used at the point of purchase affect the average level of sales?

## Hints and Answers

**12.14** *df* (within) = total *df* – *df* (between) = 9 – 2 = 7, that is, $k-1=2$ and $n-k=7$

$$\text{MSB} = \frac{\text{SSB}}{k-1} \quad \text{or} \quad 5 = \frac{\text{SSB}}{2}, \text{ SSB} = 10 \text{ and}$$

$$\text{MSW} = \frac{\text{SSW}}{n-k} = \frac{14}{7} = 2$$

$$\text{SST} = \text{SSB} + \text{SSW} = 10 + 14 = 24;$$

$$\text{F} = \frac{\text{MSB}}{\text{MSW}} = \frac{5}{2} = 2.5$$

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between raws | 10 | 2 | 5 | |
| • Within column | 14 | 7 | 2 | 2.5 |
| Total | 24 | 9 | | |

**12.15** Let $H_0$: (a) No significant difference in efficiency of four persons

(b) Mean time needed to complete the job is equal

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between rows | 138.667 | 3 | 46.22 | $F_1 = 47.64$ |
| • Within columns | 22.167 | 2 | 11.08 | $F_2 = 11.42$ |
| • Residual error | 5.833 | 6 | 0.97 | |
| Total | 166.667 | 11 | | |

(a) $F_{cal} = 47.64$ is more than its table value $F_{0.05} = 4.75$ at $df = (3, 6)$, the null hypothesis is rejected.

(b) $F_{cal} = 11.42$ is more than its table value $F_{0.05} = 5.14$ at $df = (2, 6)$, the null hypothesis is rejected.

**12.16** Let $H_0$ : (a) No significant difference in average mileage for four brands of engine oils.
(b) No significant difference in average mileage for a combination of city and highway travel.

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between sizes | 519.50 | 3 | 173.16 | $F_1 = \frac{173.16}{1.11} = 156.00$ |
| • Within brands | 5.50 | 3 | 1.83 | $F_2 = \frac{1.83}{1.11} = 1.64$ |
| • Residual error | 10.00 | 9 | 1.11 | |
| Total | 535.00 | 15 | | |

(a) $F_{cal}$ = 156 is more than its table value $F_{0.05}$ = 3.862, for $df$ = (3, 9), the null hypothesis is rejected.
(b) $F_{cal}$ = 1.64 is less than its table value $F_{0.05}$ = 3.862 for $df$ = (3, 9), the null hypothesis is accepted.

**12.17** Let $H_0$ : No significant difference in the performance of three brands of TV sets.

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between samples | 10 | 2 | 5.00 | $\frac{5.00}{3.17} = 1.58$ |
| • Within samples | 38 | 12 | 3.17 | |
| Total | 48 | 14 | | |

Since $F_{cal}$ = 1.58 is less than its table value $F_{0.05}$ = 3.89 at $df$ = (2, 12), the null hypothesis is accepted.

**12.18** Let $H_0$ : Ash content is equal in all varieties of coal.

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between samples | 8.673 | 2 | 4.337 | $\frac{4.337}{1.611} = 2.69$ |
| • Within samples | 19.336 | 12 | 1.611 | |
| Total | 28.01 | 14 | | |

Since $F_{cal}$ = 2.69 is less than its table value $F_{0.05}$ = 3.89 at $df$ = (2, 12), the null hypothesis is accepted.

**12.19** Let $H_0$ : All field officers met equal number of customers in the previous four months.

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between samples | 72 | 2 | 36 | $\frac{36}{9.1} = 3.95$ |
| • Within samples | 10 | 9 | 9.1 | |
| Total | 82 | 11 | | |

Since $F_{cal}$ = 3.95 is less than its table value $F_{0.05}$ = 4.26 at $df$ = (2, 9), the null hypothesis is accepted.

**12.20** Let $H_0$ : No significant difference in the average income per household in all the three areas.

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between samples | 5787 | 2 | 2893.5 | $\frac{2893.5}{74.26}$ |
| • Within samples | 114 | 15 | 74.26 | = 38.96 |
| Total | 6901 | 17 | | |

Since $F_{cal}$ = 38.96 is more than its table value $F_{0.05}$ = 3.68 at $df$ = (2, 15), the null hypothesis is rejected.

**12.21** Let $H_0$ : Type of advertising display used at the point of purchase does not affect the average level of sales.

| *Source* | *SS* | *df* | *MSS* | *F-value* |
|---|---|---|---|---|
| • Between types of display | 351.5 | 3 | 117.2 | $\frac{117.2}{25.9} = 4.53$ |
| • Within displays | 207.4 | 8 | 25.9 | |
| Total | 558.9 | 11 | | |

Since $F_{cal}$ = 4.53 is more than its table value $F_{0.05}$ = 4.07 at $df$ = (3, 8), the null hypothesis is rejected.

# Case Studies

### Case 12.1: FMCG Company

A FMCG company wished to study the effects of four training programmes on the sales abilities of their sales personnel. Thirty-two people were randomly divided into four groups of equal size, and the groups were then subjected to the different sales training programmes. Because there were some dropouts during the training programmes due to illness, vacations, and so on, the number of trainees completing the programmes varied from group to group. At the end of the training programmes, each salesperson was randomly assigned a sales area from a group of sales areas that were judge to have equivalent sales potentials. The sales made by each of the four groups of salespeople during the first week after completing the training programme are listed in the table:

**Training Programme**

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| 78 | 99 | 74 | 81 |
| 84 | 86 | 87 | 63 |
| 86 | 90 | 80 | 71 |
| 92 | 93 | 83 | 65 |
| 69 | 94 | 78 | 86 |
| 73 | 85 | 73 | 79 |
| 97 | 70 | | |
| 91 | | | |
| 482 | 735 | 402 | 588 |

## Questions for Discussion

1. Analyse the experiment using the appropriate method.
2. Identify the treatments or factors of interest to the researcher and investigate any significant effects.
3. What are the practical implications of this experiment?
4. Write a paragraph explaining the results of your analysis.

*Nothing is good or bad by comparison.*

—Thomas Fuller

# Chapter 13

# Correlation Analysis

## LEARNING OBJECTIVES

After studying this chapter, you should be able to

- express quantitatively the degree and direction of the covariation or association between two variables.
- determine the validity and reliability of the covariation or association between two variables.
- provide a test of hypothesis to determine whether a linear relationship actually exists between the variables.

## 13.1 INTRODUCTION

The statistical methods, discussed so far, are used to analyse the data involving only one variable. Often an analysis of data concerning two or more quantitative variables is needed to look for any statistical relationship or association between them that can describe specific numerical features of the association. The knowledge of such a relationship is important to make inferences from the relationship between variables in a given situation. Few instances where the knowledge of an association or relationship between two variables would prove vital to make decision are:

- Family income and expenditure on luxury items.
- Yield of a crop and quantity of fertilizer used.
- Sales revenue and expenses incurred on advertising.
- Frequency of smoking and lung damage.
- Weight and height of individents.
- Age and sign legibility distance.
- Age and hours of TV viewing per day.

A statistical tichnique that is used to analyse the strength and direction of the relationship between two quantitative variables, is called *correlation analysis*. A few definitions of correlation analysis are:

- An analysis of the relationship of two or more variables is usually called correlation. — A. M. Tuttle
- When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation. — Croxton and Cowden

**Coefficient of correlation:** A statistical measure of the degree of association between two variables.

The **coefficient of correlation**, is a number that indicates the *strength (magnitude)* and *direction* of statistical relationship between two variables.

- The ***strength*** of the relationship is determined by the closeness of the points to a straight line when a pair of values of two variables are plotted on a graph. A straight line is used as the frame of reference for evaluating the relationship.
- The ***direction*** is determined by whether one variable generally increases or decreasis when the other variable increases.

The importance of examining the statistical relationship between two or more variables can be divided into the following questions and accordingly requires the statistical methods to answer these questions:

(i) Is there an association between two or more variables? If yes, what is form and degree of that relationship?

(ii) Is the relationship strong or significant enough to be useful to arrive at a desirable conclusion?

(iii) Can the relationship be used for predictive purposes, that is, to predict the most likely value of a dependent variable corresponding to the given value of independent variable or variables?

In this chapter the first two questions will be answered, while the third question will be answered in Chapter 14.

For correlation analysis, the data on values of two variables must come from sampling in pairs, one for each of the two variables. The pairing relationship should represent some time, place, or condition.

## 13.2 SIGNIFICANCE OF MEASURING CORRELATION

The objective of any scientific and clinical research is to establish relationships between two or more sets of observations or variables to arrive at some conclusion which is also near to reality. Finding such relationships is often an initial step for identifying causal relationships. Few advantages of measuring an association (or correlation) between two or more variables are as under:

1. Correlation analysis contributes to the understanding of economic behaviour, aids in locating the critically important variables on which others depend, may reveal to the economist the connections by which disturbances spread and suggest to him the paths through which stabilizing forces may become effective. —W. A. Neiswanger
2. The effect of correlation is to reduce the range of uncertainty of our prediction. The prediction based on correlation analysis will be more reliable and near to reality. — Tippett
3. In economic theory we come across several types of variables which show some kind of relationship. For example, there exists a relationship between price, supply, and quantity demanded; convenience, amenities, and service standards are related to customer retention; yield of a crop related to quantity of fertilizer applied, type of soil, quality of seeds, rainfall, and so on. Correlation analysis helps in quantifying precisely the degree of association and direction of such relationships.
4. Correlations are useful in the areas of healthcare such as determining the validity and reliability of clinical measures or in expressing how health problems are related to certain biological or environmental factors. For example, correlation coefficient can be used to determine the degree of inter-observer reliability for two doctors who are assessing a patient's disease.

## 13.3 CORRELATION AND CAUSATION

There are at least three criteria for establishing a causal relationship; correlation is one of them. While drawing inferences from the value of correlation coefficient, we overlook the fact that it measures only the strength of a linear relationship and it does not necessarily imply a causal relationship. That is, there are several other explanations for finding a correlation.

The following factors should be examined to interpret the nature and extent of relationship between two or more variables:

1. **Chance coincidence:** A correlation coefficient may not reach any statistical significance, that is, it may represent a nonsense (spurious) or chance association. For example, (i) a positive correlation between growth in population and wheat production in the country has no statistical significance. Because, each of the two events might have entirely different, unrelated causes. (ii) While estimating the correlation in sales revenue and expenditure on advertisements over a period of time, the investigator must be certain that the outcome is not due to biased sampling or sampling error. That is, he needs to show that a correlation coefficient is statistically significant and not just due to random sampling error.
2. **Influence of third variable:** If the correlation coefficient does not establish any relationship, it can be used as a source for testing null and allternative hypotheses about a population. For example, it has been proved that smoking causes lung damage. However, given that there is often multiple reasons of health problems, the reason of stress cannot be ruled out. Similarly, there is a positive correlation between the yield of rice and tea because the crops are influenced by the amount of rainfall. But the yield of any one is not influenced by other.
3. **Mutual influence:** There may be a high degree of relationship between two variables but it is difficult to say as to which variable is influencing the other. For example, variables like price, supply, and demand of a commodity are mutually correlated. According to the principle of economics, as the price of a commodity increases, its demand decreases, so price influences the demand level. But if demand of a commodity increases due to growth in population, then its price also increases. In this case increased demand make an effect on the price. However, the amount of export of a commodity is influenced by an increase or decrease in custom duties but the reverse is normally not true.

## 13.4 TYPES OF CORRELATIONS

There are three broad types of correlations:
1. Positive and negative,
2. Linear and non-linear,
3. Simple, partial, and multiple.

In this chapter we will discuss simple linear positive or negative correlation analysis.

### 13.4.1 Positive and Negative Correlation

A positive (or direct) correlation refers to the same direction of change in the values of variables. In other words, if values of variables are varying (i.e., increasing or decreasing) in the same direction, then such correlation is referred to as **positive correlation.**

A **negative (or inverse) correlation** refers to the change in the values of variables in opposite direction.

The following examples illustrate the concept of positive and negative correlation.

| | | | | | | |
|---|---|---|---|---|---|---|
| *Positive Correlation* | | | | | | |
| Increasing $\rightarrow x$ | : | 5 | 8 | 10 | 15 | 17 |
| Increasing $\rightarrow y$ | : | 10 | 12 | 16 | 18 | 20 |
| Decreasing $\rightarrow x$ | : | 17 | 15 | 10 | 8 | 5 |
| Decreasing $\rightarrow y$ | : | 20 | 18 | 16 | 12 | 10 |
| *Negative Correlation* | | | | | | |
| Increasing $\rightarrow x$ | : | 5 | 8 | 10 | 15 | 17 |
| Decreasing $\rightarrow y$ | : | 20 | 18 | 16 | 12 | 10 |
| Decreasing $\rightarrow x$ | : | 17 | 15 | 12 | 10 | 6 |
| Increasing $\rightarrow y$ | : | 2 | 7 | 9 | 13 | 14 |

It may be noted here that the change (increasing or decreasing) in values of both the variables is not proportional or fixed.

### 13.4.2 Linear and Non-Linear Correlation

A linear correlation implies a constant change in one of the variable values with respect to a change in the corresponding values of another variable. In other words, a correlation is referred to as *linear correlation* when variations in the values of two variables have a constant ratio. The following example illustrates a linear correlation between two variables $x$ and $y$.

| | | | | | |
|---|---|---|---|---|---|
| $x$ : | 10 | 20 | 30 | 40 | 50 |
| $y$ : | 40 | 60 | 80 | 100 | 120 |

When these pairs of values of $x$ and $y$ are plotted on a graph paper, the line joining these points would be a straight line.

A non-linear (or curvi-linear) correlation implies an absolute change in one of the variable values with respect to changes in values of another variable. In other words, a correlation is referred to as a *non-linear correlation* when the amount of change in the values of one variable does not bear a constant ratio to the amount of change in the corresponding values of another variable. The following example illustrates a non-linear correlation between two variables $x$ and $y$.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $x$ : | 8 | 9 | 9 | 10 | 10 | 28 | 29 | 30 |
| $y$ : | 80 | 130 | 170 | 150 | 230 | 560 | 460 | 600 |

When these pair of values of $x$ and $y$ are plotted on a graph paper, the line joining these points would not be a straight line, rather it would be curvi-linear.

### 13.4.3 Simple, Partial, and Multiple Correlation

The distinction between simple, partial, and multiple correlation is based upon the number of variables involved in the correlation analysis.

If only two variables are chosen to study correlation between them, then such a correlation is referred to as ***simple correlation***. A study on the yield of a crop with respect to only amount of fertilizer, or sales revenue with respect to amount of money spent on advertisement, are a few examples of simple correlation.

In ***partial correlation***, two variables are chosen to study the correlation between them, but the effect of other influencing variables is kept constant. For example (i) yield of a crop is influenced by the amount of fertilizer applied, rainfall, quality of seed, type of soil, and pesticides, (ii) sales revenue from a product is influenced by the level of advertising expenditure, quality of the product, price, competitors, distribution, and so on. In such cases an attempt to measure the correlation between yield and seed quality, assuming that the average values of other factors exist, becomes a problem of partial correlation.

In ***multiple correlation***, the relationship between more than three variables is considered simultaneously for study. For example, employer-employee relationship in any organization may be examined with reference to, training and development facilities; medical, housing, and education to children facilities; salary structure; grievances handling system; and so on.

## 13.5 METHODS OF CORRELATION ANALYSIS

The correlation between two ratio-scaled (numeric) variables is represented by the letter $r$ which takes on values between –1 and +1 only. Sometimes this measure is called the '**Pearson product moment correction**' or the **correlation coefficient**. The correlation coefficient is scale free and therefore its interpretation is independent of the units of measurement of two varibles, say $x$ and $y$.