UNIT 13 STATISTICAL TESTING OF HYPOTHESIS

Structure

13.1 INTRODUCTION

In Unit 12, we focused on descriptive statistics including the various measures of central tendency, variability, relative positions and relationships. These measures are used to describe the properties of particular samples. In this unit, we shall introduce inferential or sampling statistics. The knowledge of these statistics is useful for testing the hypothesis(es) related to your research problems, and to make generalizations about populations on the basis of data analysis. This requires you to be familiar with certain statistical tests – parametric and non-parametric.

This unit aims to provide you with detailed information about the nature and use of parametric and non-parametric tests in general and the application of some of these tests for drawing inferences and generalizations.

Objectives

After studying this unit, you will be able to :

- classify various statistical tests,
- describe the nature of parametric tests alongwith the assumptions on which they are based,
- work out sampling distribution of means in the context of (i) large samples, and (ii) small samples,
- define and illustrate the concept of confidence intervals and levels of significance,
- define and illustrate the concept of degrees of freedom,
- use Z-test and t-test in testing significance of the difference between means,

- define and illustrate the concept of one-tailed and two-tailed tests of significance,
- describe the nature and uses of analysis of variance,
- describe the nature of the non-parametric tests alongwith their assumptions,
- apply the chi-square test, and
- describe the use of median test and its application.

13.2 CLASSIFICTION OF STATISTICAL TESTS

The descriptive statistics already discussed in Unit 12 are used to explain the properties of samples drawn from a population. The researcher computes certain 'statistics' (sample values) as the basis for inferring the corresponding 'parameters' (population values). Ordinarily, a single sample is drawn from a given population so as to determine how well a researcher can infer or estimate the 'parameter' from a computed sample 'statistics'. For making the inferences about the various parameters, the researcher makes use of parametric and non-parametric tests. There tests are described herewith.

13.3 PARAMETRIC TESTS

Under this section we discuss two sub-themes: sampling distribution of means and application of parametric tests. Sampling distribution of means covers a) large samples, b) confidence intervals and levels of significance, c) small samples, and d) degree of freedom. Application of parametric tests covers three tests, namely Z-test, t-test and F-test.

Parametric tests are the most powerful statistical tests for testing the significance of the computed sampling statistics. These tests are based on the following assumptions:

1) the variables described are expressed in interval or ratio scales and not in nominal or ordinal scales of measurement,
2) the population values are normally distributed,
3) the samples have equal or nearly equal variances – this condition is known as 'equality or homogeneity of variances' and is particularly important to determine for small samples,
4) the selection of one case in the sample is not dependent upon the selection of any other.

Z-test, t-test and F-test are the most commonly used parametric tests. Before discussing the application of these tests, it is necessary to describe certain concepts relating to 'sampling distribution of means', 'confidence intervals', 'levels of confidence of significance' and 'degrees of freedom'. The concepts are discussed next.

13.3.1 Sampling Distribution of Means

As already mentioned earlier under this sub-section we shall cover the sampling distribution of means for large/small samples, concept of confidence interval and level of significance and degree of freedom.

13.3.1.1 *Large Samples*

A important principle, known as the 'central limit theorem', describes the characteristics of sample means. If a large number of equal-sized samples (greater than 30) are selected at random from an infinite population, then:

- the distribution of 'sample means' is normal and it possesses all the characteristics of a normal distribution,

- the average value of 'sample means' will be the same as the mean of the population,
- the distribution of the sample means around the population mean will have its own standard deviation, known as 'standard error of mean' which is denoted as SE_M or σ_M. It is computed by the formula:

$$SE_M = \sigma_M = \frac{\overline{\sigma}}{\sqrt{N}} \quad (1)$$

in which

$\overline{\sigma}$ = standard deviation of the population and

N = the number of cases in the sample

Since the value of $\overline{\sigma}$ (i.e. standard deviation of population) is usually not known, we make an estimate of this standard error of mean by the formula:

$$\sigma_M = \frac{\sigma}{\sqrt{N}} \quad (2)$$

in which

σ = standard deviation of the sample

N = the number of cases in the sample

To illustrate the use of formula (2), we assume that the means of the attitude scores of a sample of 100 mothers towards infant feeding practical is 25 and the standard deviation is 5. The standard error of mean can be calculated accordingly:

$$SE_M = \sigma_M = \frac{5.0}{\sqrt{100}} = 0.50$$

This is the measure of likely variation around mean when different samples of same size are drawn from the same population.

This standard 'error of mean' may be assumed as the standard deviation of a distribution of sample mean, around the fixed population mean of all distance mothers. In the case of large randomly selected samples, the sampling distribution of sample means is assumed to be normal as shown in Figure 13.1.

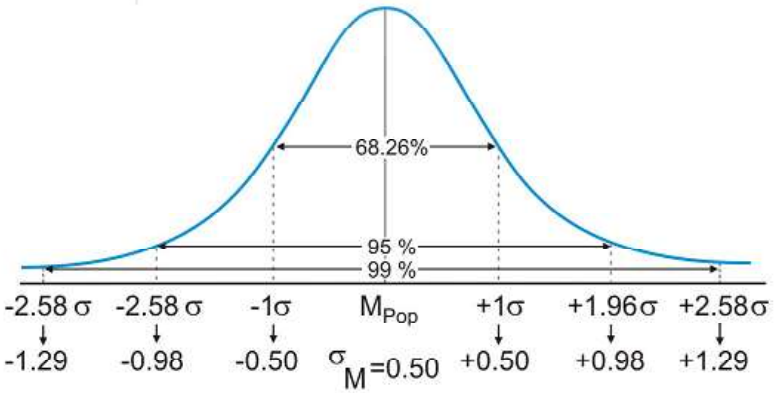

Figure 13.1 : Sampling distribution of means showing variability of obtained means around the population mean in terms of σ_M

The normal curve in Figure 13.1 shows that this sampling distribution is centered around the unknown population mean with standard deviation 0.05. The sample means often fall between the positive and the negative side of the population mean. About 2/3 of our sample means (exactly 68.26 per cent) will lie within ± 1.00 σ_M of the population mean, i.e. within a range of ± 1×0.50 = ± 0.50. Furthermore, 95 of our 100 sample means will lie within ± 2.00 σ_M (more exactly ± 1.96 σ_M) of the population

means i.e. 95 of 100 sample means will lie within ± 1.96×0.50 or ± 0.98 of the population mean. In other words, the probability that our sample mean of 25 does not miss the population mean (M pop.) by more than ± 0.98 is 0.95. Also, 99 of our sample will be within ± 3.00 σ_M (more exactly ± 2.68 σ_M) of the population mean. This indicated that 99 out of 100 sample means will fall within ± 2.58 × 0.50 or ± 1.29 of the population mean. The probability (P) that our sample mean of 25 does not miss the M pop. By more that ± 1.29 is 0.99.

Thus, the value of a population mean, to be inferred from a randomly selected sample mean, can be estimated on a probability basis. In case of proportion (π) the SE(p) can be estimated using the formula:

$$SE_{(p)} = \sqrt{\frac{p(1-p)}{n}}$$

Check Your Progress Exercise 1

1) Describe the assumptions on which the use of parametric tests are based.

.....

.....

.....

2) Given a sample of 100 children 1 - 3 year of age with mean (SD) intake of calcium = 175 (5.82). Compute the standard error of mean.

.....

.....

.....

.....

Now that we are familiar with the concept of standard error of mean; we move on to confidence level and level of significance.

13.3.1.2 *Confidence Intervals and Levels of Significance*

When we draw a large random sample from the population to obtain measures of a variable and compute the mean for the sample, we can use the 'central limit theorem' and 'normal probability curve' to have an estimate of the population mean (M pop.). We can say that M has a 95 per cent chance of being within 1.96 standard error units of M pop. In other words, a mean for a random sample has a chance of 95 per cent of being within 1.96 σ_M units from M pop. It may also be said that there is a 99 per cent chance that the sample mean lies within 2.58 σ_M units of M pop. To be more specific, it may be stated that there is a 95 per cent probability that the limits M $\pm$ 2.58 σ_M enclose the population mean with 99 per cent probability. Such limits M $\pm$ 1.96 σ_M enclose the population mean, and the limits enclosing the population mean are known as the 'confidence intervals'.

These limits help us to adopt particularly two levels of confidence. One is known as 5 per cent level or 0.05 level, and the other is known as 1 per cent level or .01 level. The .05 level of confidence indicates that the probability M pop that lies within the interval M ± 1.96 σ_M is 0.95 and that it falls outside of these limits is .05. By saying that probability is 0.99, it is meant that M pop. lies within the interval M ± 2.58 σ_M and that the probability of its falling outside of these limits is .01.

To illustrate, let us apply the concept to the previous problem. Taking as our limits M ± 1.96 σ_M, we have 25 ± 1.96×0.50 or a confidence interval marked off by the limit 24.02 and 25.98. Our confidence that this interval contains M pop, is expressed by a probability of 0.95. If we want a higher degree of confidence, we can take the 0.99 level of confidence for which the limits are M ± 2.58 σ_M or a confidence interval given by the limits 23.71 and 26.29. We may be quite confident that M pop. is not lower than 23.71 nor higher than 26.96, i.e. the chances are 99 in 100 that the M pop. lies between 23.71 and 26.19.

Next, let us get to know about sampling distribution of mean for small sample.

13.3.1.3 *Small Samples*

When the number of cases in the sample is less than 30, we may estimate the value of σ_M by the formula:

$$SE_M = \frac{S}{\sqrt{N}} \tag{3}$$

in which

S = standard deviation of the small sample

N = the number of cases in the sample

The formula for computing S is

$$S = \sqrt{\frac{\sum x^2}{N-1}} \tag{4}$$

in which

$\sum x^2$ = sum of the squares of deviations of individual scores from the sample mean

N = the number of cases in the sample

The concept of small size was developed by *William Seely Gooset*, a consulting statistician for Guinness Breweries of Dublin (Ireland) around 1915. The principle is that we should not assume that the sampling distribution of means of small samples is normally distributed. He found that the distribution curves of small sample means were somewhat different from the normal curve. When the size of the sample is small, the t-distribution lies under the normal curve, but the tails or ends of the curve are higher than the corresponding parts of the normal curve. Figure 13.2 shows that the t-distribution does not differ significantly from the normal distribution unless the sample size is quite small. Further, as the sample size increases in size, the t-distribution approaches more and more closely to the normal curve.

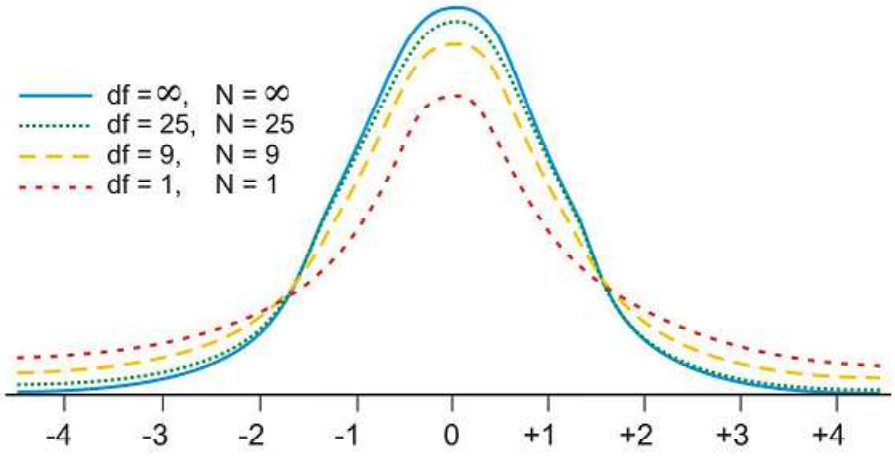

Figure 13.2 : Distribution of t-values for Degrees of Freedom from ∝ to 2 (when df is very large, the distribution of t approaches the normal)

For small samples, it is necessary to make use of selected points in the table of *Gosset's* t-critical values or *student's t-values*, given in Appendix (Table II). As the

sample size increases, the student's t-values approach the Z-values of Normal probability Table. When small samples are used, the use of t-values involves an important concept known as 'degrees of freedom', which we shall now discuss separately.

13.3.1.4 *Degrees of Freeedom*

While finding the standard deviation of small samples we use N – 1 in the denominator instead of N in the basic formula for standard deviation. The difference in the two formulae may seem very little, if N is sufficiently large. But there is a very important difference in the 'meaning' in the case of small samples. N – 1 is known as the 'number of degrees of freedom', denoted by df. The 'number of degrees of freedom' *in a distribution is the number of observations or values that are independent of each other and cannot be deduced from each other. In other words, we may say that the 'degrees of freedom' connote freedom to vary*.

To illustrate as to why the df used here is N – 1, we take 5 scores, i.e., 5,6,7,8 and 9, the mean of which is 7. This mean score is to be used as an estimate of the population mean. The deviations of the scores from the mean 7 are –2, –1, 0, +1 and +2. A mathematical requirement of the mean is that the sum of these deviations should be zero. Of the five deviations only 4, i.e., N – 1 can be chosen freely (independently) as the condition that the sum is equal to zero restricts the value of the 5th deviate. With this condition, we can arbitrarily change any four of the five deviates and thereby fix the fifth. We could take the first four deviates as –2, –1, 0 and +1, which would mean that for the sum of deviates to be zero, the fifth deviate has to be +2. Similarly, we can try other changes and if the sum is to remain zero, one of the five deviates is automatically determined. Hence, only 4, i.e., (5 – 1)'s are free to vary within the restrictions imposed.

When a statistic is to be used to estimate a parameter, the number of degrees of freedom depends upon the restrictions imposed. One df is lost for each of the restrictions imposed. Therefore, the number of df varies from one statistics to another. For example, in estimating and computing the population mean (M pop) from the sample Mean (M), we lose 1 df. So, the number of degrees of freedom is (N – 1).

$\frac{x}{\sigma}$

Let us determine the 0.95 and 0.99 confidence intervals for the population mean (M pop) of the scores 10, 15, 10, 25, 30, 20, 25, 30, 20 and 15, obtained by 10 mothers on an attitude scale to measure infant feeding practices. The mean of the scores is

$$= \frac{10+15+10+25+30+20+25+30+20+15}{10}$$

$$= \frac{200}{10} = 20.00$$

Using formula (4) we compute the standard deviation as :

X	$x = X - M$	x^2
10	–10	100
15	–5	25
10	–10	100
25	5	25
30	10	100
20	0	0
25	5	25
30	10	100
20	0	0
15	–5	25
	$\Sigma x = 0$	$\Sigma x^2 = 500$

$$S = \sqrt{\frac{\sum x^2}{N-1}}$$

$$= \sqrt{\frac{500}{10-1}}$$

$$= 7.45$$

from formula (3) we compute

$$= SE_M = \frac{7.45}{\sqrt{10}}$$

$$= 2.36$$

For estimating the M pop from the sample mean of 20.00, we determine the value of t at the selected points using appropriate number of degrees of freedom. The available df for determining t is N – 1 or 9. Entering Table II (See Appendix) with 9 df. We read that t = 2.26 at .05 level and 3.25 at .01 level. From the first t-value we know that 95 of our 100 sample means will lie within ± 2.26 SE_M or ± 2.26 × 2.36 of the population mean and 5 out of 100 fall outside of these limits. The probability (P) that our sample mean 20.00 does not miss the M pop by more than ± 2.26 × 2.36 or ± 5.33 is 0.95. From the second t-value, we know that 99 per cent of our sample mean will lie between M pop and 3.25 SE_M or ± 3.25 × 2.36, and that 1 per cent fall will beyond these limits. So, the probability (P) that our sample mean of 20.00 does not miss the M pop by more than ± 3.25 × 2.36 or ± 7.67 is 0.99.

Taking our limits as M ± 2.26 SE_M, we have 20.00 ± 2.26 × 2.36 or 14.67 and 25.33 as the limits of the 0.95 confidence interval. The probability (P) that M pop is not less than 14.67 nor greater than 25.33 is 0.95. Taking the limits M ± 3.25 SE_M, we have 20.00 ± 3.25 × 2.36, or 12.33 and 27.67 as the limits of the 0.99 confidence interval, and the probability (P) so that M pop is not less than 12.33 and not greater than 27.67 is 0.99.

The use of small samples to build generalizations in epidemiological research should be made cautiously as it is difficult to ensure that a small sample adequately represents the population from which the sample is drawn. Furthermore, conclusions drawn from small samples are usually unsatisfactory because of the great variability from sample to sample. In other words, large samples drawn randomly from the population will provide a more accurate basis than will small samples for inferring population parameters.

Next, let us learn about the application of the parametric tests.

13.3.2 Application of Parametric Tests

In this subsection, we shall discuss the application of three parametric tests, namely Z-test, t-test and F-test.

13.3.2.1 *Application of Z-test for Testing the Significance of Difference between Means of two Independent Large Samples*

It has already been explained in earlier sections that the frequency distribution of large sample means drawn from the same population fall into a normal distribution around M pop as their measure of central tendency. It is also reasonable to expect that the frequency distribution of the difference between the means computed from the two samples will also tend to be normal with a mean of zero and standard deviation of 1. It is termed the 'Standard error of the difference between two means' and is denoted by d_M. It is computed by the formula:

$$\sigma_{d_M} = \sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2} \quad (5)$$

σ_{M_1} = the SE of the mean of the first sample

σ_{M_2} = the SE of the mean of the second sample

To illustrate, we apply formula (5) to a problem. Suppose a nutrition knowledge test was administered on two groups, one of 120 males and the other of 75 females enrolled in the Masters Programme in Dietetics and Food Service Management M.Sc. (DFSM) programme.

The results are summarized in Table 13.1 below.

Table 13.1: Means and standard deviation of two independent large samples

Statistics	Boys	Girls
N	120	75
Mean (M)	57.50	55.75
Standard Deviation σ	8.42	8.13

Assuming that our samples are random, it is to be ascertained whether the difference between the means 57.50 and 55.75 is significant.

Using formula (5) we compute the 'standard error of the difference between means'.

$$\sigma_{d_M} = \sqrt{\frac{(8.42)^2}{120} + \frac{(8.13)^2}{75}}$$

$$= 1.21$$

The obtained difference between the means of males and females is 1.75 (i.e., 57.50 – 55.75); and the SE of this difference (σ_{d_M}) = 1.21.

To determine whether two groups of males and females actually differ in nutrition knowledge, we set-up a null hypothesis, i.e. the difference between the population means of males and females is zero and that, except for sampling errors, means of males and females is zero and that, except for sampling errors, mean differences from sample to sample will also be zero. In accordance with a null hypothesis, we assume a sampling distribution of differences to be normal with the mean at zero, (or at M pop (males) – M pop (females) = 0). The deviation of each sample difference, $[M_{males} - M_{females}] - [M\ pop_{(males)} - M\ pop_{(females)}]$ or $[M_{males} - M_{females}] = 0$. The deviation of each sampled difference between two means, given in terms of standard measure, would be the deviation divided by the standard error, which gives us a $\bar{z}$ value in terms of a general formula:

$$\bar{z} = \frac{|M_1 - M_2|}{d_M} \quad (6)$$

Using formula (6)

$$\bar{z} = \frac{1.75}{1.21} = 1.45$$

For the sake of convenience, we use .05 and .01 levels of significance as two arbitrary standards for accepting or rejecting a hypothesis. From the normal distribution Table

I (See Appendix) we read that ± 1.96 σ (godown the $\frac{x}{\sigma}$ column till 1.9, then across horizontally to the column headed 0.5 and read only 47.44) mark off points along the base line of the normal curve to the left and right of which lie 5 per cent of the cases (i.e., 2.5 percent at each end of the curve). When a $\bar{z}$-value is 1.96 or more, we reject a null hypothesis at .05 level of significance. The computed $\bar{z}$-value of 1.45 in our problem falls short of 1.96, i.e., it does not reach the .05 level. Accordingly, we retain the null hypothesis and conclude that two groups of males and females actually do not differ in their mean performance on nutrition knowledge test.

Furthermore, from Table I (See Appendix) we know that ± 2.58 σ mark off points to the left and right of which lie 1 per cent (0.5 per cent at each end of the curve) *of* the cases in the normal distribution. If the $\bar{z}$ value is 2.58 or more, we reject the null hypothesis at .01 level and the probability (P) is that not more than once in 100 trails would a difference of this size arise if the true difference (M pop_1 – M pop_2) was zero.

We hope the application of $\bar{z}$-value for testing the significance of difference between means of two independent sample is clear. Next, we shall focus on two-tailed and one-tailed tests of significance.

13.3.2.2 *Two-tailed and one-tailed Tests of Significance*

Suppose a null-hypothesis were set up that there was no difference, other than a sampling error difference, between the mean height of two groups, A and B. We would be concerned only with the difference and not with the superiority or inferiority in height of either group. To test this hypothesis, we apply two-tailed test as the difference between the obtained means of height of two groups may be as often in one direction (plus) as in the other (minus) from the true difference of zero. Moreover, for determining probability, we take both tails of sampling distribution.

For a large sample two-tailed test, we make use of a normal distribution curve. The 5 per cent area of rejection is divided equally between the upper and the lower tails of this curve and we have to go out to ± 1.96 on the base line of the curve to reach the area of rejection as shown in the Figure 13.3.

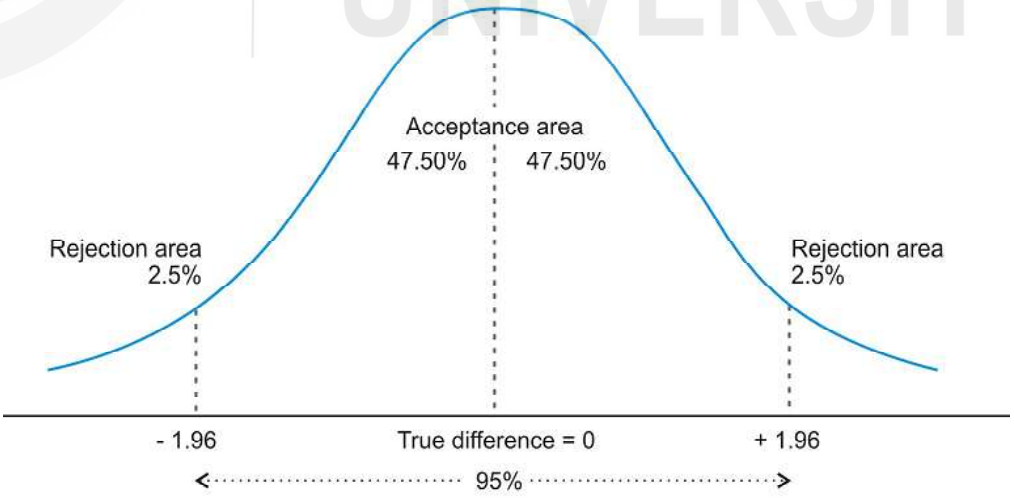

Figure 13.3: A Two-tailed test at .05 level (2.5 per cent at each level)

Similarly, if we have 0.5 per cent area at each end of the normal curve where 1 per cent area of rejection is to be divided equally between its upper and lower tails, it is necessary to go out to ± 2.58 on the base line to reach the area of rejection as shown in the Figure 13.4.

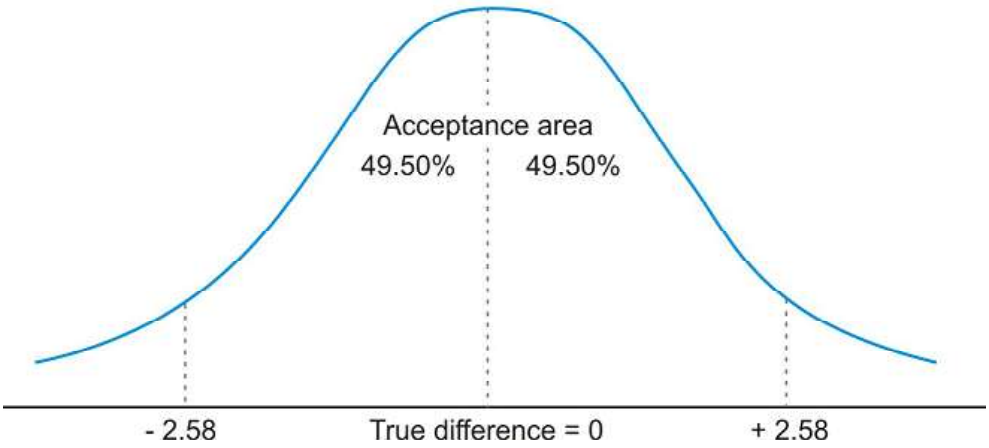

Figure 13.4: A Two-tailed Test at .01 level (0.5 per cent at each level)

In the case of the above example a null hypothesis was set up that there was no difference other than a sampling error difference between the mean nutrition knowledge score of males and females of M.Sc. (DFSM) programme Thus, we were concerned with a difference, and not in superiority or inferiority of either group in the nutrition knowledge test. To test this hypothesis, we applied 'two-tailed test' as the difference between the two means might have been in one direction (plus) or in the other (minus) from the true difference of zero; and we took both tails of sampling distribution in determining probabilities.

As is evident from the above example, we make use of a normal distribution curve in the case of a large sample 'two-tailed test'. The 5 per cent area of rejection is equally divided between the upper and lower tails of the curve and we have to go out to ± 1.96 on the base line of the curve to reach the area of rejection.

Similarly, we have 0.5 per cent area at each end of the normal curve when 1 per cent of rejection is to be divided equally between its upper and lower tails and it is necessary to go out to ± 2.58 on the base line to reach the area of rejection.

In the above problem, if we change the null hypothesis as : male group of M.Sc. (DFSM) have significantly higher nutrition knowledge than that of the female group; or male group have significantly lower nutrition knowledge than the female group of M.Sc. (DFSM) course, then each of these hypotheses indicates a direction of difference. In such situations, the use of 'one-tailed test' is made. For such a test, the 5 per cent area or 1 per cent area of rejection is either at the upper tail or at the lower tail of the curve, to be read from 0.10 column (instead of .05) and .02 column (instead of .01).

13.3.2.3 *Application of t-test for testing the Significance of Difference between two Independent Small Samples*

We have already discussed that the frequency distribution of small sample means drawn from the same population forms a t-distribution and it is reasonable to expect that the sampling distribution of the difference between the means computed from two different populations will also fall under the category of t-distribution. *Fisher* provided the formula for testing the difference between the means computed from independent small samples. The formulae is as follows:

$$t = \frac{|M_1 - M_2|}{\sqrt{\left(\frac{\sum x_1^2 + \sum x_2^2}{N_1 + N_2 - 2}\right)\left(\frac{N_1 + N_2}{N_1 \times N_2}\right)}} \qquad (7)$$

in which

M_1 and M_2 = means of two samples

$\sum x_1^2$ and $\sum x_2^2$ = sums of squares of the deviations from the means in the two samples

N_1 and N_2 = number of cases in the two samples

df = degrees of freedom = $N_1 + N_2 - 2$

To illustrate the use of the formula, let us test the significance of the difference between mean weight scores of 7 boys and 10 girls of 1 - 5 years of age.

Table 13.2: Weight measurements of 7 boys and 10 girls 1 - 5 years of age

Boys	**(N_1 = 7)**				**(N_2 = 10)**	
X_1	x_1		x^2_1	**Girls**	x_2	x^2_2
13	0	0	10	–4	16	
14	1	1	16	2	4	
11	–2	4	12	–2	4	
12	–1	1	13	–1	1	
15	2	4	18	4	16	
13	0	0	13	–1	1	
13	0	0	19	5	25	
			14	0	0	
			13	–1	1	
$\sum X_1 = 91$		$\sum X_2 = 10$	12	–2	4	
			$\sum X^2 = 140$		$\sum x_2^2 = 72$	

$$M_1 = \frac{91}{13} = 13 \qquad M_2 = \frac{140}{10} = 14$$

$$df = N_1 + N_2 - 2 = 7 + 10 - 2 = 15$$

Using formula (7)

$$t = \frac{|13-13|}{\sqrt{\left(\frac{10+72}{7+10-2}\right)\left(\frac{7+10}{7\times 10}\right)}}$$

$$= \frac{1}{\sqrt{\left(\frac{82}{15}\right)\left(\frac{17}{70}\right)}}$$

$$= \frac{1}{\sqrt{9.29}}$$

$$= 0.33$$

To test the significance of difference between the two means by making use of two-tailed test (null hypothesis, i.e., no differences between the two groups), we look for the t-critical values for rejection of null hypothesis in Table II (Appendix) for (7 + 10 – 2) or 15 df. These t-values are 2.13 at .05 and 2.95 at .01 levels of significance. Since the obtained t-value 0.33 is less than the table value necessary for the rejection of the null hypothesis at .05 level for df 15, the null hypothesis is accepted and it may be concluded that there is no significant difference in the mean weight of males and females.

If we change the null hypothesis as: *boys will have higher weight than girls, or males will have lower weight than females*, then each of these hypotheses indicates a direction of difference rather than simply the existence of the difference. So, we make use of one-tailed test. For given degrees of freedom, i.e., 15, the .05 level is read from the 0.10 Column (p/2=.05) and the .01 level from 0.02 column (p/2=.01) of the t-table. In the one-tailed test, for 15 df t-critical values at .05 and .01 levels, as read from the 0.10 and the 0.02 columns of Table II are 1.75 and 2.60, respectively. Since the computed t-value of 0.33 does not reach the table value at .05 level (i.e., 1.75 for .10), we may conclude that the difference in two groups is there merely because of chance factors.

13.3.2.4 *Application of F-test for Testing the Significance of Difference between Independent Means (in two or more groups)*

The use of $\bar{Z}$ and t-test is made by a researcher to determine whether there is any significant difference between the means of two random samples. Suppose we have seven randomly drawn samples from a population and we want to determine whether there are any significant differences among their means. This will require computation of $\frac{7(7-1)}{2} = 21$ t-tests to determine the significance of difference between the seven means by taking two means at a time. This procedure is time consuming, as well as, cumbersome. The technique of analyzing of variance is applied to determine if any two of the seven means differ significantly from each other by a single test, known as *F-test*, rather than 21 t-tests. The *F-test* makes it possible to determine whether the sample means differ from one another (between group variance) to a greater extent than the test scores differ from their own sample means (within group variance). Using the ratio:

$$F = \frac{\text{Variance between the groups}}{\text{Variance within groups}} \qquad (8)$$

The values of F-ratio are given in the Appendix (Table III). This table indicates the F-critical values necessary for rejecting the null hypothesis at selected levels of significance, usually .05 and .01 levels.

The generic method used for comparing means in three or more groups is called *analysis of variance* (ANOVA). The name comes from the fact that the total variance in all the groups combined is broken down into components such as within-groups variance and between-groups variance. Between-groups variance is the systematic variation occurring due to group differentials. If genuine group differentials are present then the between-groups variance should be large relative to the within-groups variance. Thus the ratio of these two components of variance can be used as a criterion to find that the group means are different or not. The common setups for such comparisons are one-way ANOVA and two-way ANOVA.

Let us study the basic assumptions for ANOVA next.

Basic Assumptions for the Analysis of Variance

Certain basic assumptions underlying the technique of analyzing of variance are:

1) The population distribution should be normal. This assumption is, however, not so important. The study of Norton (Guilford, 1965; pp. 300-301) also points out that F is rather insensitive to variations in the shape of population distribution.

2) All groups of a certain criterion or of the combination of more than one criterion should be randomly chosen from the sub-population having the same criterion or having the same combination of more than one criterion. For example, if we wish to select two groups from a study centre, one belonging to a rural area and the other to the urban area, we must, choose the groups randomly from the respective sub-populations.

3) The sub-groups under investigation must have the same variability. In other words, there should be homogeneity of variance. It is tested either by applying Bartlett's test of homogeneity or by applying Hartley's test.

13.3.2.5 *Factor Analysis*

Factor analysis is a statistical approach that can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factors). Unlike other techniques like Regression analysis or ANOVA, factor analysis does not require that the predictor and criterion variables be defined. Factor analysis attempts to identify the relationship between all variables included in the analysis set. The variables included in the analysis have a portion of their variance explained by certain underlying common dimensions, called the factors. The main applications of factor analytic techniques are: (1) to *reduce* the number of variables and (2) to *detect structure* in the relationships between variables, that is to *classify variables.*

Factor Analysis is a technique mainly used in research in psychology and education. We will not examine factor analysis in detail here, but very briefly describe about this technique.

Factor Analysis is a procedure for determining the number and nature of constructs that underlie a set of measures (Wiersman 1986). Construct as you are already aware, is a trait or an attribute that explains some phenomena, e.g., anxiety intelligence, motivation, attitude etc. in factor analysis artificial variables are generated and called factors. Factors analysis is initiated from the correlation matrix of the variables. The variable which are highly correlated are grouped together.

Suppose we have scores of 100 students on 10 different tests. Question here is – How many different traits or constructs these 10 tests measure. The possibility can be that three or four tests measure the same trait or one test may measure two or more traits. The researcher can determine the correlation co-efficient among the 10 different tests. High correlation's between test scores indicate that common constructs are measured. Low or zero correlation's indicate the absence of common constructs.

There are few terms used in factor analysis. If a test measures only one construct, it is labeled as *factorially pure. A factorially complex* test is one that measures two or more factors.

Factor loading is the extent to which a test measures a factor. Factor loading is very important in factor analysis because factors which are artificial variables generated from the data must be described and integrated. It is a correlation coefficient between a test and a factor.

Uses of Factor Analysis in Research

In conducting research, the aim of using factor analysis is to identify the nature and number of constructs that underlie a set of variables. Factor analysis is associated with construct related evidence when establishing validity (construct validity is explained in Unit 7). Factor analysis is used in confirmatory analysis and exploratory analysis.

Confirmatory factor analysis is used in studies where hypothesized constructs measured by a set of variables are either confirmed or refuted. It is also used to analyze a single test by factor analyzing the item scores.

Suppose 50 item test measures three taits, a confirmatory analysis of the item scores would support or refute this proposition.

On the other side, in the exploratory analysis the number of variables are reduced to a manageable number of explanatory purposes. A set of measures can be factor analyzed to enhance the explanation of what is measured in a more persimomons manner.

Eg. A group of teachers in an open university were observed and measured on two different competencies. A factor analysis of the competency scores undoubtedly would generate a smaller number of factors, say three or four, that represents the constructs underlying the performance of teacher.

Thus factor analysis in any research analysis provide valuable insights into the nature of phenomena.

Check Your Progress Exercise 2

1) The following scores were obtained on an interest test for 5 males and 8 females of PGDDE enrolled with IGNOU.

   Male : 20, 22, 30, 32, 26

   Female : 34, 25, 16, 30, 22, 27, 20, 26.

   Is the difference between the Mean Interest Scores of the Males and the Females significant?

   .....

   .....

   .....

   .....

   .....

To illustrate the use of F-test, let us consider an example of twenty children who have been randomly assigned to 4 groups – [Control (C), Supplementation group (S), Nutrition Education Group (NE)] and Supp + Nutrition Education Group (NES). The nutrition knowledge scores test, administered after the completion of experiment are given in Table 13.3.

Table 13.7: Nutrition test score of the four groups of children knowledge

	Methods of Groups				
	C (X_1)	S (X_2)	NE (X_3)	NES (X_4)	
	14	19	12	17	
	15	20	16	17	
	11	19	16	14	
	10	16	15	12	
	12	16	12	17	
ΣX	62	90	71	77	300
ΣX^2	786	1634	1025	1207	4652

we may compute the analysis (ANOVA) of variance using the following steps:

1) Correction = $\frac{\left(\sum X\right)^2}{N} = \frac{(300)^2}{20} = 4500$

2) Total sum of squares (Total SS)

   = ΣX^2 – Correction

$= 14^2 + 15^2 + ...12^2 + 17^2 - 4500$

$= 4652 - 4500$

$= 152$

3) Sum of squares between means of treatments (Methods) A, B, C, and D (between means):

$$= \frac{\left(\sum X_1\right)^2}{N_1} + \frac{\left(\sum X_2\right)^2}{N_2} + \frac{\left(\sum X_3\right)^2}{N_3} + \frac{\left(\sum X_4\right)^2}{N_4} - \text{Correction}$$

$$= \frac{(62)^2}{5} + \frac{(90)^2}{5} + \frac{(71)^2}{5} + \frac{(77)^2}{5} - 4500$$

$= 4582.8 - 4500$

$= 82.8$

4) Sum of squares within treatments (Methods) A, B, C, and D (SS within means):

$=$ Total SS – SS between means

$= 152 - 82.8$

$= 69.2$

5) Calculation of variances from each SS and analysis of the total variance into its components.

Each SS becomes a variance when divided by the degrees of freedom (df) alloted to it. There are 20 scores in all in Table 13.3 and hence there are (N – l) or (20 – 1) = 19 df in all. These 19 df are allocated in the following ways:

If N = number of scores in all and K = number of treatments or groups, we have df for total SS = N – l = 20 – 1 = 19, df for within treatments = N – K = 20 – 4 = 16; and df for between the means of treatments = K – l = 4 – 1 = 3:

The variance among means of treatments is 82.8/3 or 27.60; and the variance within means is 69.2/16 or 4.33.

The summary of the analysis of variance may be presented in tabular form as follows:

Table 13.4 : Summary of analysis of variance

Source of Variance	df	Sum of Squares (SS)	Mean Square (Variance)
Between the Means of treatment	3	82.8	27.60
Within treatment	16	69.2	4.33
Total	**19**	**152.0**	

Using formula (8)

$$F = \frac{27.60}{4.33} = 6.374$$

In the present problem, the null hypothesis asserts that four sets of scores are in reality the scores of four random samples drawn from the same normally distributed population, and that the means of the four groups A, B, C, and D will differ only through

fluctuations of sampling. For testing this hypothesis, we divided the 'between means' variance by the 'within treatments' variance and compared the resulting variance ratio, called F, with the F-values in Table III. The F value of 6.374 in the present case is to be checked for table value for df 3 and 16 (the degrees of freedom for numerator and denominator). The table values for .05 and .01 levels of significance are 3.24 and 5.29. Since the computed F-value of 6.374 is greater than the table values, we reject the null hypothesis and conclude that the means of the four groups differ significantly.

13.4 NON-PARAMETRIC TESTS AND APPLICATION OF CHI-SQUARE TEST

In the preceding section, we described/discussed some important parametric tests involving the assumptions based upon the nature of the population distribution. There are some tests which do not make numerous or stringent assumptions about the nature of the population distribution. These tests are known as distribution-free or non-parametric tests. The non-parametric tests are based upon the following assumptions:

1) The nature of the population, from which samples are drawn, is not known to be normal.
2) The variables are expressed in nominal form, that is, classified in categories and represented by frequency counts. .
3) The variables are expressed in ordinal form, that is, ranked in order or expressed in numerical scores which have the strength of ranks.
4) The sample sizes are small.

The most frequently used non-parametric tests are: Chi-square test, the median test, the sign test, the: Mann-Whitney U test, the Kolmogorov-Smirnov Two Sample Test, theWilcQxon-Matched-Pairs Signed-Ranks Test, the McNemar Test for Significance of changes, contingency co-efficient, etc. For the present unit, we will discuss the applications of Chi-square test and median test only.

13.4.1 Application of Chi-Square Test

The Chi-square (pronounced as Ki-square) test is used with discrete data in the form of frequencies. It is a test of independence and is used to estimate the likelihood that some factor other than chance accounts for the observed relationship. Since the null hypothesis states that there is no relationship between the variables under study, the Chi-square test merely evaluates the probability that the observed relationship results from chance. The formula for Chi-square (χ^2) is

$$\chi^2 = \sum\left[\frac{(fo - fe)^2}{fe}\right] \qquad (9)$$

in which

fo = frequency of the occurrence of observed or experimentally determined facts

fe = expected frequency of occurrence

To test the significance of Chi-square, we enter Table IV of the Appendix with the computed value of Chi-square for the appropriate number of degrees of freedom. The number of degrees of freedom df = (r – l) (c – l), in which r is the number or rows and c is the number of columns in which the data are tabulated.

For example consider the following data (in Table 13.5) of 500 subject who have been categorized into three groups, elder, middle-aged and younger on the basis of age and their preference or four colours, red, blue, yellow and green.

Table 13.5 : The Chi-square test of independence in contingency Table

Colour → Age Group↓	**Red**	**Blue**	**Yellow**	**Green**	**Total**
Elder	40 (38.42)	50 (45.22)	35 (39.10)	45 (47.26)	170
Middle	35 (36.16)	42 (42.56)	44 (36.80)	39 (44.48)	160
Younger	38 (38.42)	41 (45.22)	36 (39.10)	55 (47.26)	170
Total	**113**	**133**	**115**	**139**	**500**

Across the first row of the table, we find that out of 170 subjects in the older age-group, 40 have given their preference for red colour, 50 for blue, 35 for yellow, and 45 for green. Reading down the first column, we find that out of 113 subjects giving preference for red colour, 40 belong to the older age-group, 35 to middle and 38 to younger age-group. The other columns and rows are interpreted in the same way.

The hypothesis to be tested is the null hypothesis, that is, age and colour preferences are essentially unrelated or independent. To compute Chi-square we must calculate an independent value, i.e., expected frequency for each cell in the contingency table. Independent values are represented by the figures in parentheses within the different cells. They give the number of subjects whom we should expect to fall in a particular age-group, showing their preference for a particular colour in the absence of any real association.

The calculation of expected frequencies (fe) and Chi-square (χ^2) are shown as under:

1) *Calculation of expected frequencies (fe)*

Row I : $\frac{113\times170}{500}=38.42;$ $\frac{133\times170}{500}=45.22;$

$\frac{115\times170}{500}=39.10;$ $\frac{139\times170}{500}=47.26;$

Row II : $\frac{113\times160}{500}=36.16;$ $\frac{133\times160}{500}=42.56;$

$\frac{115\times160}{500}=36.80;$ $\frac{139\times160}{500}=44.48;$

Row III : $\frac{113\times170}{500}=38.42;$ $\frac{133\times170}{500}=45.22;$

$\frac{115\times170}{500}=39.10;$ $\frac{139\times170}{500}=47.26;$

2) Computation of the Chi-square value, using formula (9)

$$\chi^2 = \sum\left[\frac{(fo-fe)^2}{fe}\right]$$

$$= \frac{(40-38.42)^2}{38.42}+\frac{(50-45.22)^2}{45.22}+\frac{(35-39.10)^2}{39.10}+\frac{(45-47.26)^2}{47.26}+$$

$$\frac{(35-36.16)^2}{36.16}+\frac{(42-42.56)^2}{42.56}+\frac{(44-36.80)^2}{36.80}+\frac{(39-44.48)^2}{44.48}+$$

$$\frac{(38-38.42)^2}{38.42}+\frac{(41-45.22)^2}{45.22}+\frac{(36-39.10)^2}{39.10}+\frac{(55-47.26)^2}{47.26}+$$

$\chi^2 = 5.182$

3. df = (r – 1) (c – 1)
= (3 – 1) (4 – 1)
= 8

The χ^2 critical values for 8 df as given in Table IV (See Appendix) are 15.507 and 20.090 respectively for .05 and .01 levels of significance and, the obtained value, 5.182, is less than the table value even at .05 level. This indicates that there is no relationship between the age and the colour preference and thus the hypothesis that age and colour preference are essentially independent may be accepted at .05 level of significance.

In the case of 2 × 2 contingency table, with (r – l) (c – l) = 1 df, there is no need of computing the expected frequencies (independence values) for each cell. The formula is:

$$\chi^2 = \frac{N(|AD-BC|)^2}{(A+B)(C+D)(A+C)(B+D)} \quad (10)$$

In the above formula A, B, C and D are the frequencies in the first, second, third and fourth cells respectively and the vertical lines in |AD – BC| mean that the difference is to be taken as positive.

To illustrate the use of formula (10), let us determine whether item 5 of an achievement test differentiates between high and low achievers. The responses to items are given in the 2 × 2 contingency table given in:

Table 13.6 : The Chi-square test in 2×2 fold contingency table

	Passed item 5 (A)	**Failed item 5 (B)**	**Total (C)**
High Achiever	115 (C)	35 (D)	150 (C+D)
Low Achiever	40	90	130
Total	**(A+C) 155**	**(B+D) 125**	**280**

Using formula (10)

$$\chi^2 = \frac{280(|115\times90 - 35\times40|)^2}{(115+35)(40+90)(115+40)(35+90)}$$

$$= \frac{280(|10350 - 1400|)^2}{(150)(130)(155)(125)} = 59.36$$

Since the computed χ^2 value of 59.36 exceeds the critical χ^2 value of 6.635 to be significant at. 01 level, we reject the hypothesis that item 5 of the test, does not discriminate significantly between high and low achievers. In other words, it may be concluded that item 5 of the achievement test discriminates significantly between the two groups, namely high and low achieving students.

Further, when entries in 2×2 table are less than 10, Yate's correction for continuity is applied to formula (10). The corrected formula reads:

$$\chi_c^2 = \frac{N\left(|AD - BC| - N/2\right)^2}{(A+B)(C+D)(A+C)(B+D)} \quad (11)$$

The following example illustrates the use of formula (11).

Fifteen male and twelve female subject in a health clinic were asked to express their attitude towards population education. Both the groups of subject were administered the attitude scale and were classified as having either positive or negative attitude towards population education. The distribution of the sample is shown Table 13.7.

Table 13.7: Distribution of male and female subjects in terms of their positive or negative attitude towards population education

	Positive Attitude	**Negative Attitude**	**Total**
	(A)	(B)	(A+B)
Female	7	5	12
	(C)	(D)	(C+D)
Male	9	6	15
	(A+C)	(B+D)	
Total	16	11	27

$$\chi_c^2 = \frac{27\left(|7\times6 - 5\times9| - 27/2\right)^2}{(7+5)(9+6)(7+9)(5+6)}$$

$$= \frac{27(|42 - 45| - 13.5)^2}{12\times15\times16\times11} = 0.23$$

Since the obtained value of χ^2, 0.23, is less than the table value of 3.842 to be significant at .05 level of significance, it may be inferred that there is no true difference in the attitude of male and female counselors towards population education.

We hope the discussion above was quite useful in helping you understand the concept of chi-square and perhaps enable you to apply the test for inference of your ordinal or nominal data. Next, let us review the application of median tests.

13.4.2 Application of Median Test

The median test is used for testing whether two independent samples differ in central tendencies. It gives information as to whether it is likely that two independent samples have been drawn from populations with the same median. It is particularly useful whenever the measurements for two samples are expressed in an ordinal scale.

In the median test, we first compute the combined median for all rank measures in both samples. Then both sets of rank measures at the combined median are dichotomized and the data are set in a 2×2 Table as shown in Table 13.8.

Table 13.8: 2×2 Table for use of the median test

	Group I	Group II	Total
No. of measures above the combined median	(A)	(B)	(A+B)
No. of measures below the combined median	(C)	(D)	(C+D)
Total	(A+C)	(B+D)	

Under the null-hypothesis, we would expect about half of each group's measures to be above the combined median and about half to be below it, that is, we would expect frequencies A and C to be equal, and frequencies B and D to be nearly equal. In order to test this null-hypothesis, we calculate, f^2 using the formula (12):

$$f^2 = \frac{N(|AD - BC| - N/2)^2}{(A+B)\,(C+D)\,(A+C)\,(B+D)} \qquad (12)$$

Let us illustrate the use of the formula (12) in the following example.

Eighteen male and female distance learners of a study centre of an open university were asked to express their attitude towards the functioning of a study centre. Both the groups were administered an attitude scale and a common median attitude measure was worked out. The number of cases from both the groups falling above and below the median point is shown in Table 13.9.

Table 13.9: Distribution of distance learners (male and female) below and above the common median

	Below Median	Above Median	Total
Female Distance Learners	10	4	14
Male Distance Learners	6	12	18
Total	**16**	**16**	**32**

Using formula (12),

$$f^2 = \frac{32\left(|(10)(12) - (6)(4)| - \frac{32}{4}\right)^2}{16 \times 16 \times 18 \times 14}$$

$$= 3.17$$

Since the obtained value 3.17 of f^2 for 1 df does not exceed the f^2 critical value of 3.84 for a two-tailed test at .05 level, the null hypothesis is retained and we may conclude that there is no difference in the attitude of male and female distance learners towards the functioning of the centre.

Check Your Progress Exercise 4

1) Describe the assumptrons on which non-parametric tests are based.

.....

.....

.....

.....

2) The following table shows the number of the male and the female distance learners who have passed or failed an item of an achievement test. Test whether the item of the test differentiates between the two groups of males and females.

	Number Passed the Test Item	Number Failed the Test Item
Males	30	20
Females	25	15

.....

.....

.....

13.5 LET US SUM UP

In this unit we described the nature of parametric and non-parametric tests alongwith the assumptions they are based on. The applications of some parametric tests (Z, t and F) and non-parametric tests (Chi-square) in the analysis of data have also been discussed.

1) Parametric tests are used in the analysis of data available in interval or ratio scales of measurement. .

2) Parametric tests assume that the data are normally or nearly normally distributed.

3) Z-test, t-test and F-test are the most commonly used parametric tests.

4) If large-sized samples, i.e., those greater than 30 in size, are selected at random from an infinite population, the distribution of sample means is called the 'sampling distribution of means'. This distribution is normal and it possesses all the following characteristics of a normal distribution:

   i) The average value of sample means will be the same as the mean of population.

   ii) The distribution of the sample means around the population mean has its own standard deviation which is known as the 'standard error of the mean' (SE_M or σ_M).

   iii) The sampling distribution is centered at the unknown population mean with its standard deviation 0.50.

   iv) The sample means often fall equally on the positive and negative sides of the population mean.

   v) About 2/3 of the sample means (exactly 68.26 per cent) will lie within ± 1.00 σ_M of the population mean, i.e., within a range of ± 1 × 0.50 or ± 0.50 of the population mean.

vi) 95 of the 100 sample means will lie within ± 1.96 σ_M of the population mean, i.e., 95 of 100 sample means will lie within ± 1.96 × 0.50 or ± 0.98 of the population mean.

vii) 99 of the 100 sample means will be within ± 2.580 σ_M of the population mean, i.e., 99 of our sample means will lie within ± 2.58 × 0.50 or ± 1.29 of the population mean.

5) If we draw a large sample randomly from a population and compute its mean, the mean has 95 per cent chance of being within 1.96 σ_M units from the population mean. Also, there is a 99 per cent chance that the sample mean lies within 2.58 σ_M units from the population mean. To be more specific, there is 95 per cent probability that the limits M ± 1.96 σ_M enclose the population mean and 99 per cent probability that the limits M ± 2.58 σ_M enclose the population mean.

6) The limits M ± 1.96 σ_M and M ± 2.58 σ_M are called 'confidence intervals' for .05 and .01 1eve1s of confidence respectively.

7) When a 'statistic' is used to estimate, a 'parameter', the number of 'degrees of freedom' depends upon the restrictions placed. Therefore, the number of degrees of freedom (df) will vary from one statistic to another. In estimating the population mean from the sample mean, for example, 1 df is lost and so the number of degrees of freedom is N – l.

8) The Z-test is used for testing the significance of the difference between the means of two large samples.

9) The t-test is used for testing the significance of the difference between the means of two small samples.

10) Under the null hypothesis, the difference between the sample means may be either plus or minus and as often in one direction as in the other from the true (population) difference of zero, so that in determining probabilities we take both tails of the normal sampling distribution and make use of two-tailed test. But in many situations our primary concern is with the direction of the difference rather than with its existence in absolute terms. In such situations, we make use of one-tailed test.

11) We use Z-test and t-test to determine whether there is any significant difference between means of two random samples. But when the number of samples is more than two, F-test, based on the technique of analysis of variance, is used for testing the significance of the sample means.

12) Non-parametric tests are used in the analysis of non-parametric data, i.e., when the data are available in nominal or ordinal scales of measurement.

13) Non-parametric tests are distribution-free tests and do not rest upon the more stringent assumptions of normally distributed population.

14) Chi-square test, median test, Sign-test, Mann-Whitney U test, Wilcoxon Matched Pairs Signed Ranks test and Kolmogorov Smirnov two Sample test are examples of some non-parametric tests.

15) Chi-square is used with discrete data in the form of frequencies. It is a test of independence, and is used to estimate the likelihood that some factor other than 'chance' accounts for the observed relationship between the variables.

16) Median test is used for testing whether two independent samples differ in central tendencies. It is particularly useful whenever the measurements for the two samples are expressed in an ordinal scale.

Now, use the following check list and see whether you have learnt to:

- classify various statistical tests,
- describe the nature of parametric tests along with the assumptions on which

they are based,

- define sampling distribution of means,
- define the standard error of mean,
- define the confidence intervals and levels of confidence,
- compute 0.95 and 0.99 confidence intervals for the true mean from a large sample mean,
- define and illustrate the concept of degrees of freedom,
- compute 0.95 and 0.99 confidence intervals for the true mean from the sample mean,
- apply Z-test for testing the significance of the difference between means of two independent large samples involving: (i) one-tailed and (ii) two-tailed tests,
- apply t-test for testing the significance of the difference between means of two independent small samples involving: (i) one-tailed and (ii) two-tailed tests,
- describe the nature and uses of the analysis of variance,
- state the basic assumptions of the technique of analysis of variance,
- apply F-test for testing the significance of the difference between means,
- describe the nature of the non-parametric tests alongwith their assumptions,
- name various non-parametric tests,
- describe the use of Chi-square test,
- illustrate the application of Chi-square test, and
- describe the use of median test.

13.6 GLOSSARY

Parametric Tests : these are statistical tests which are used for analyzing parametric data and making inferences about the parameters from the statistics. These tests are based upon certain assumptions about the nature of data distributions and the types of measure used.

Non-parametric Tests : these are statistical tests which are used for analyzing non-parametric data and making possible useful inferences without any assumptions about the nature of data distributions.

Standard Error of Mean : it is the standard deviation of a distribution of sample means.

Degrees of Freedom : the number of degrees of freedom in a distribution is the number of observations or values that are independent of each other, and cannot be deduced from each other.

13.7 ANSWERS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress Exercise 1

1) Parametric data should be used if the following basic assumptions are met. These assumptions are based on the nature of the population distribution and on the way the scale is used to quantify the data observations.

i) The observations are independent. The selection of one case is not dependent upon the selection of any other case.

ii) The population values are normally distributed.

iii) The samples have equal or near variances. This condition is known as equality or homogeneity of variances and is particularly important to determine when the samples are small.

iv) The variables described are expressed in interval or ratio scale and not in nominal or ordinal scales of measurement.

2) The standard error of Mean is

$$SE_M = \sigma M = \frac{\sigma}{\sqrt{N}}$$

$$= \frac{5.82}{\sqrt{100}}$$

$$= \frac{5.82}{10}$$

$$= 0.582.$$

Check Your Progress Exercise 2

Boys (N_1 = 5)			Girls (N_2 = 10)		
X_1	x_1	x^2_1	X_2	x_2	x^2_2
20	–6	36	34	9	81
22	–4	16	25	0	0
30	4	16	16	–9	81
32	6	36	30	5	25
26	0	0	22	–3	9
$\Sigma X_1 = 130$		$\Sigma X_1^2 = 104$	27	2	4
			20	2	25
			26	–5	1
			$\Sigma X_1 = 200$	1	$\Sigma x_2^2 = 226$
Mean = $M_1 = \frac{\sum X_1}{N_1} = \frac{130}{5}$			Mean = $M_2 = \frac{\sum X_2}{N_2} = \frac{200}{8}$		
= 26.00			= 25.00		

ii) df = $N_1 + N_2 - 2 = 5 + 8 - 2 = 11$

iii) Using formula (7)

$$t = \frac{|M_1 - M_2|}{\sqrt{\left(\frac{\sum x_1^2 + \sum x_2^2}{N_1 + N_2 - 2}\right)\left(\frac{N_1 + N_2}{N_1 \times N_2}\right)}}$$

$$= \frac{|26 - 25|}{\sqrt{\left(\frac{104 + 226}{5 + 8 - 2}\right)\left(\frac{5 + 8}{5 \times 8}\right)}}$$

$$= \frac{1}{\sqrt{9.75}} = 0.32$$

iv) We used a two-tailed test as we are not hypothesizing a direction of the difference between the means. The t-values as given in Table II in the Appendix for 11 df for .05 and .01 columns are 2.20 and 3.1 respectively. Since the obtained value of 0.32 is less than these table values, the difference between the mean interest scores of boys and girls is not significant.

Check Your Progress Exercise 3

1) Non-parametric tests are distribution-free tests and are based on the following assumptions:

   i) The nature of the population distribution, from which samples are drawn is not known to be normal.

   ii) The variables are expressed in nominal form (classified in categories and represented by frequency counts).

   iii) The variables are expressed in ordinal form (ranked in order).

2) i) The number of the males and the females who have passed or failed the test item is given in the following 2 × 2 table.

	Number Passed	**Number Failed**	**Total**
Female	30 (A)	20 (B)	50 (A+B)
Male	25 (C)	15 (D)	40 (C+D)
Total	55 (A+C)	35 (B+D)	90

ii) Using formula (10)

$$\chi^2 = \frac{N(|AD-BC|)^2}{(A+B)(C+D)(A+C)(B+D)}$$

$$= \frac{90(|(30)(15)-(20)(25)|)^2}{(30+20)(25+15)(30+25)(20+15)}$$

$$= 0.58$$

iii) Since the obtained value 0.058 of χ^2 does not exceed the Table value 3.841 of χ^2 at. 05 level of significance, we may conclude that the test item does not differentiate between the two groups of males and females.