UNIT 15 ANALYSIS OF QUANTITATIVE DATA: INFERENTIAL STATISTICS BASED ON PARAMETRIC TESTS

Structure

15.1 INTRODUCTION

In the previous unit, you studied the nature of quantitative data and various descriptive statistical measures which are used in the analysis of such data. These include measures of central tendency, variability, relative position and relationship. The concept, characteristics and applications of normal probability curve were also explained.

The computed values of various statistics are used to describe the properties of particular samples. In this unit we shall discuss inferential or sampling statistics, which are useful to a researcher in making generalisations or inferences about the populations from the observations of the characteristics of samples. For making inferences about various population values (parameters), we generally make use of parametric and non-parametric tests. The concept and assumptions of parametric tests will be explained to you in this section along with the inference regarding the means and correlations of large and small samples, and significance of the difference between the means and correlations in large and small independent samples. The assumptions and applications of analysis of variance and co-variance for testing the significance of the difference between the means of three or more samples will also be discussed.

15.2 OBJECTIVES

After studying this unit, you will be able to:

- define the concept of statistical inference;
- explain the nature of parametric tests;
- state the assumptions on which the use of parametric tests are based;
- define the sampling distribution of means and product moment correlation;
- state the characteristics of central limit theorem;
- define the standard error of mean;
- define the confidence intervals and levels of confidence;
- compute the .95 and .99 confidence intervals for the true mean from the large sample mean.
- define and illustrate the concept of degrees of freedom;
- compute the .95 and .99 confidence intervals for the true mean from the small sample mean.
- test the significance of the difference between the means of: (i) two independent large samples, and (ii) two independent small samples involving: (a) one tailed test and (b) two tailed test;
- test the significance of the difference between the means of two dependent samples;
- test the significance of the difference between the means of three or more samples using the technique of: (i) analysis of variance and (ii) analysis of covariance;
- compute the .95 and .99 confidence intervals for the true co-efficient of correlation from the sample Pearson's co-efficient of correlation;
- test the significance of Pearson's co-efficient correlation;
- test the significance of the difference between Pearson's co-efficient of correlations of the independent samples.

15.3 INFERENTIAL STATISTICS

It was explained to you in Unit 14 that the values of descriptive statistics, namely, mean, median, mode, standard deviation, correlation etc. are used to describe properties of particular samples. It is not possible to infer or make generalizations about the populations from the measures of the samples. The sampling or inferential statistics enable the researcher to make generalisations about a population characteristic (parameter) from a probability sample (statistics). The researcher computes certain statistics (sample values) as the basis for inferring what the corresponding parameters (population values) might be as it is not possible to measure all of the members/units of a given population. It may be noted that the values of the parameters for a given population are unknown.

In research situations, a researcher may draw a single sample using any of the probability sampling methods studied by you earlier. The choice of the sampling method depends upon a given set of conditions. His problem is to determine how well he can infer or estimate the parameter from the one sample statistics. Under specified conditions, it is possible for the researcher to forecast the population values (parameters) from the sample values (statistics) with a known degree of accuracy. The degree to which a sample statistics represents its parameter is an index of the significance of the computed sample statistics. While it is possible to make inferences about various parameters, in this unit we shall limit our discussion to mean and product moment correlation using parametric tests.

15.4 PARAMETRIC TESTS : USES AND ASSUMPTIONS

Parametric tests are useful as these tests are most powerful for testing the significance or trustworthiness of the computed sample statistics. However, their use is based upon certain assumptions. These assumptions are based on the nature of the population distribution and on the way the type of scale is used to quantify the data measures. It may be mentioned that there are some parametric tests, namely, t-test and F-test, which are quite robust and are appropriate even when some assumptions are not met.

The assumptions for most parametric tests are the following:

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

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

3. The population values are normally distributed.

4. The observations are independent. The selection of one case in the sample is not dependent upon the selection of any other case.

15.5 STATISTICAL INFERENCE BASED ON PARAMETRIC TESTS

Suppose a large number of researchers selected 50 random samples each of 200 students from the population of all B.Ed. students enrolled with IGNOU and obtained their scores on a scale measuring their attitude towards teaching. The mean attitude scores of the samples would not be identical. A few would be relatively high, a few relatively low, but most of the means would tend to cluster around the population mean. The variation of sample means is due to an error which is known as 'sampling error'. This type of error explains the chance variations which are inevitable when a number of randomly selected sample means are computed.

Parametric tests are useful in making the inferences about the population mean with the help of the sample means on a probability basis. In the following discussion, you will learn how to draw statistical inference about the means of large and small samples.

15.5.1 Statistical Inference Regarding Means of Large Samples

Suppose we wish to measure the teaching aptitude of the B.Ed. distance trainees enrolled with IGNOU using a verbal aptitude teaching test. It is not possible and convenient to measure the teaching aptitude of all the enrolled B.Ed. trainees and hence we must usually be satisfied with a sample drawn from this population. However, this sample should be as large and as randomly drawn as possible to represent adequately all the B.Ed. trainees of IGNOU. If we select a large number of random samples of 100 trainees each from the population of all trainees, the mean values of teaching aptitude scores for all samples would not be identical. A few would be relatively high, a few relatively low, but most of them would tend to cluster around the population mean. The sample means due to 'sampling error' will not vary from sample to sample but will also usually deviate from the population mean. Each of these sample means can be treated as a single observation and these means can be put in a frequency distribution which is known as sampling distribution of the means.

An important principle, known as the 'Central Limit Theorem', describes the characteristics of sample means. According to this theorem, if a large number of equal-sized samples, greater than 30 in size, are selected at random from an infinite population:

1. The means of the samples will be normally distributed.
2. The average value of the sample means will be the same as the mean of the population.
3. The distribution of sample means will have its own standard deviation. This standard deviation is known as the 'standard error of the mean' which is denoted as SE_M or σ_M. It gives us a clue as to how far such sample means may be expected to deviate from the population mean. The standard error of a mean tells us how large the errors of estimation are in any particular sampling situation.

   The formula for the standard error of the mean in a large sample is:

   $$SE_M \text{ or } \sigma_M = \frac{\sigma}{\sqrt{N}}$$

   where

   σ = the standard deviation of the population

   N = the size of the sample

To illustrate the application of the central limit theorem, let us assume that the mean of the teaching aptitude scores of a sample of 64 B.Ed. trainees is 16 and the standard deviation 5. The standard error of the mean is:

$$SE_M = \sigma_M = \frac{5}{\sqrt{64}} = 5/8 = 0.625 \text{ or } 0.63$$

The value 0.63 of SE_M can be thought of as the standard deviation of a distribution of sample mean around the fixed population mean of all the B.Ed. trainees. Since the sampling distribution of sample means is assumed to be normal, it possesses all the characteristics of a normal distribution.

The normal curve in the Figure 15.1 illustrates that this sampling distribution of means is centred at the unknown population mean with its standard deviation 0.63. The sample means falls equally often on the positive and negative sides of the population mean. About 2/3 of sample means (exactly 68.26%) will lie within ± 1.00 σ_M of the population mean i.e., within a range of ± 1 × 0.63 or ± 0.63.

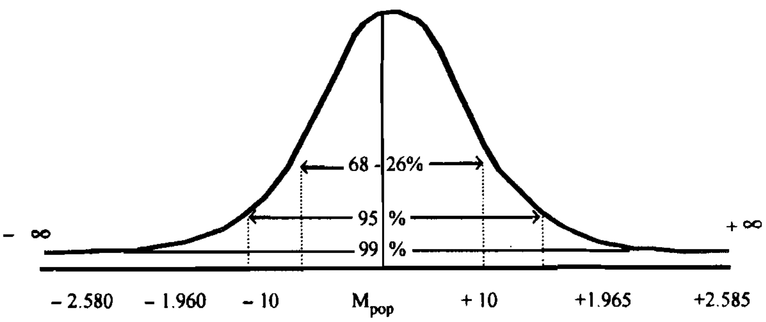

Fig. 15.1: Sampling Distribution of Means Showing Variability of Obtained Means Around Population (M_{pop}) in terms of σ_M.

Furthermore, 95 out of 100 sample means will lie within ± 2.00 σ_M (more exactly ± 1.96 σ_M) of the population mean i.e. 95 out of 100 sample means will lie within ± 1.96 × 0.63 or ± 1.23 of the population mean. The probability (P) is .95, that the sample mean of 64 does not miss the population mean (M_{pop}) by more than ± 1.23. In other words the probability is .05 that the sample mean does miss the M_{pop} by more than ± 1.23. Also, 99 of 100 sample means will lie within ± 3.00 σ_M (more exactly ± 2.58 σ_M) of the population mean i.e. 99 of sample means will lie within ± 2.58 × 0.63 or ± 1.63 of the population mean. The probability (P) is .99 that sample mean of 64 does not miss the M_{pop} by more than ± 1.63. In other words, the probability is .01 that the sample mean of 64 does miss the M_{pop} by more than ± 1.63. The magnitude of probable deviation of a sample mean from its population mean gives us a measure of the probability with which we are able to estimate the population mean (M_{pop}) from the sample mean.

Using the central limit theorem and the normal-curve concept, we may infer that the sample mean of 64 has a 95 per cent chance of being within 1.96 standard error units from M_{pop}. In other words, the mean of 64 for the random sample has a 95 per cent chance of being within 1.96 units from M_{pop}. We may also infer that there is a 99 per cent chance that the sample mean of 64 lies within 2.58 σ_M units from M_{pop}. More specifically, we can say that there is a 95 per cent probability that the limits M ± 1.96 $ó_M$ spans the population mean M_{pop}. Also, there is a 99 per cent probability that the limits M ± 2.58σ_M encloses M_{pop}.

The limits M ± 1.96 σ_M and M ± 2.58 σ_M are called '**confidence intervals**' or '**fiduciary limits**'. These limits help us to adopt particularly two **levels of confidence**. One is known as 5 per cent level or .05 level, and the other as the 1 per cent or .01 level. The .05 level of confidence indicates that the probability is .95 that M_{pop} lies within the interval M ± 1.96 σ_M and .05 that it falls outside these limits. Similarly, .01 level of confidence indicates that the probability is .99 that M_{pop} lies within the interval M ± 2.58 σ_M, and .01 that it falls outside of these limits.

In our example, the limits of M ± 1.96 σ_M will be 64 ± 1.96 × 0.63 i.e. a confidence interval marked off by the limits 62.77 and 65.23. Our confidence that this interval contains M_{pop} is expressed by a probability of .95. For a higher degree of confidence, we can take .99 level of confidence, for which the limits are M ± 2.58 σ_M i.e. a confidence interval given by the limits 62.37 and 65.63. This indicates that M_{pop} is not lower than 62.37 nor higher than 65.63 i.e., the chances are 99 in 100 that the M_{pop} lies between 62.37 and 65.63.

15.5.2 Statistical Inference Regarding Means of Small Samples

In case of small samples, the sampling distribution of means is not normal. It was in about 1815 when William Seely Gosset developed the concept of small sample size.

He found that the distribution curves of small sample means were some what different from the normal curve. This distribution was named as t-distribution. When the size of the sample is small, the t-distribution lies under the normal curve. Figure 2 shows that the t-distribution does not differ greatly from the normal curve unless the sample size is quite small. As the sample increases in size, the t-distribution approaches more and more closely to the normal curve.

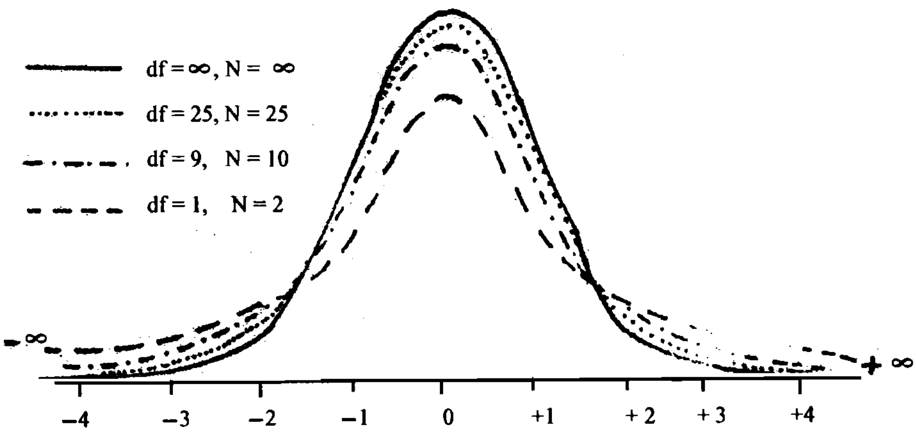

Fig. 15.2: Distribution of t for Degrees of Freedom from I to ∞ (when df is very large, the distribution of t approaches normal).

In case of small samples, we make use of t-table, given in the Appendix, using respective 'degrees of freedom' depending upon the size of the sample for getting t-values instead of 1.96 for .05 and 2.58 for .01 levels of confidence. You can observe from the t-table that the t-values approach z-values (1.96 or 2.58) of normal probability table as the sample size increases. For the application of t-table it is necessary for you to understand the concept of 'degree 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 which cannot be deduced from each other. When a mean is computed for data involving a number of observations, the sum of the measures is calculated and divided by N.

$$\text{Mean} = \frac{\sum x}{N}$$

But computing a mean, 1 degree of freedom is used up or lost, and subsequent calculations of variance and standard deviation will be based on N–1 independent observations or N–1 degrees of freedom.

Let us understand the concept with the help of the following example, using original distribution (0) and altered distribution (A):

O Original Distribution	A Altered Distribution	
+8	12	These four terms can be altered in any way.
+7	9	
+6	14	
+5	8	
+4	−13	This term is dependent on, or determined by, the other four terms.
$\sum x = 30$	$\sum x = +30$	
N = 5	N = 5	
Mean = +6	Mean = +6	

When the deviations are taken from the mean, the sum of the deviations is zero. Hence, in the altered distribution (A), the fifth term must have a value of −13 for the sum to equal +30, the mean to be +6, and the sum of the deviations from the mean is equal to zero. Thus the four terms, 12, 9, 14, 8 in the altered distribution are independent and can be altered. But one term is fixed which is deduced from the other four. There are N-1 i.e. (5-1) or 4 degrees of freedom.

When a sample statistics is used to estimate a population parameter, the number of degrees of freedom depends upon the restrictions placed. One df is lost for each restriction imposed. Thus the number of degrees of freedom (df) will vary from one statistics to another.

In estimating population mean (M_{pop}) from the sample mean (M), we lose 1 df and so the number of degrees of freedom is (N−1). By way of illustration, let us determine the .95 and .99 confidence intervals for the population mean (M_{pop}) of the scores 8, 8, 10, 7 and 9 achieved by 5 students in a test of mathematics. The mean of scores is:

$$= \frac{8+8+10+7+9}{5} = 8.40$$

When the number of cases in the sample is small (less than 30), the formula for computing standard deviation (S) is:

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

in which

$\sum x^2$ = Sum of the squares of deviations from the mean.

N = Number of cases in the sample.

Using this formula, we compute the value of standard deviation as:

X	x = X − M	x^2
8	−0.40	0.16
8	−0.40	0.16
10	+1.60	2.56
7	−1.40	1.96
9	+0.60	0.36
M = 8.40	$\sum x = 0$	$\sum x^2 = 5.20$

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

$$S = \sqrt{\frac{5.20}{5-1}} = 1.14$$

The standard error of the mean (SE_M) is:

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

$$= \frac{1.14}{\sqrt{5}} = 0.51$$

For estimating the M_{pop} from the sample mean of 5.40, we determine the value of t at .05 and .01 points using appropriate number of degrees of freedom. The df for determining t are N–1 or 4. Entering t-table with 4 df, the value of t at 0.05 point is 2.78 and 4.60 at 0.01 point.

For t = 2.78, 95 of our 100 sample means will lie within ± 2.78 SE_M or ± 2.78 × 0.51 of the population mean and that 5 out of 100 fall outside these limits. The probability (P) is .95 that the sample mean 8.40 does not miss the M_{pop} by more than ± 2.78 × 0.51 or ± 1.92. The limits of the .95 confidence interval are 8.40 ± 2.78 × 0.51 or 6.48 and 10.32. The probability (P) is .95 that M_{pop} is not less than 6.48 and greater than 10.32.

For t=4.60, we know that 99 of our 100 sample means will lie between M_{pop} and ± 4.60 SE_M or ± 4.60 × 0.51 and that 1 out of 99 fall beyond these limits. The probability (P) is .99 that our sample mean of 8.40 does not miss the M_{pop} by more than ± 4.60 × 0.51 or ± 2.35. The limits of .99 confidence interval are 8.40 ± 4.60 × 0.51 or 6.05 and 10.75. The probability (P) is .99 that M_{pop} is not less than 6.05 and greater than 10.75.

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

1. i) State the assumptions upon which parametric tests are based.

....

....

ii) Compute: (i) .95 and (ii) .99 confidence intervals for the population mean for a sample with mean = 30, SD = 5 and N = 100.

....

....

iii) Find the .99 confidence interval for the population mean for a sample of 5 students whose scores on a test are 6, 5, 7, 4, 5.

....

15.6 TESTING THE STATISTICAL SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS

In some research situations we require the use of a statistical technique to determine whether a true difference exists between the population parameters of two samples. The parameters may be means. standard deviations, correlations etc. For example, we wish to determine whether the population of male B.Ed. trainees enrolled with IGNOU differ from their female counterparts in their attitude towards teaching. In this case we would first draw samples of male and female B.Ed. trainees. Next, we would administer an attitude scale measuring attitude towards teaching on the selected samples, compute the means of the two samples, and find the difference between them. Let the mean of the male sample be 55 and that of the females 59. Then it has to be ascertained if the difference of 4 between the sample means is large enough to be taken as real and not due only to sampling error or chance. In order to test the significance of the obtained difference of 4, we need to first find out the standard error of the difference of the two means because it is reasonable to expect that the difference between two means will be subject to sampling errors. Then from the difference between the sample means and its standard error we can determine whether a difference probably exists between the population means. In the following sections we will discuss the procedure of testing the significance of the difference between the means and correlations of the samples.

15.6.1 Significance of the Difference Between the Means of Two Independent Large and Small Samples

Means are said to be independent or uncorrelated when computed from samples drawn at random from totally different and unrelated groups.

Large Samples

You have learnt that the frequency distribution of large sample means, drawn from the same population, fall into a normal distribution around the population mean (M_{pop}) as their measure of central tendency. It is reasonable to expect that the frequency distribution of the difference between the means computed from samples drawn from two different populations will also tend to be normal with a mean of zero and standard deviation which is called the standard error of the difference of means. The standard error is denoted by σ_{dM} which is estimated from the standard errors of the two sample means, σ_{M1} and σ_{M2}. The formula is:

$$\sigma_{dM} = \sqrt{\sigma_{M1^2} + \sigma_{M2^2}}$$

in which

σ_{M1} = SE of the mean of the first sample

σ_{M2} = SE of the mean of the second sample

N_1 = Number of cases in the first sample

N_2 = Number of cases in the second sample

To illustrate the application of this formula, let us consider a problem in which a teaching aptitude test was administered on two randomly selected groups of teachers, one of 32 males and the other 34 females. The data are summarized in the following table:

Statistics	Male Teachers	Female Teachers
N	32	34
Mean	87.43	82.58
Standard Deviation	6.27	6.39

For testing the significance of the difference between the means, 87.43 and 82.58, of the two groups, we first compute the standard error of the difference of these means. Using the formula of σ_{dM}, we have:

$$\sigma_{dm} = \sqrt{\frac{(6.27)^2}{32} + \frac{(6.39)^2}{34}}$$

$$= \sqrt{\frac{39.40}{32} + \frac{40.80}{34}}$$

$$= \sqrt{1.23 + 1.20}$$

$$= \sqrt{2.43} = 1.56$$

In order to find out whether the male and female teachers actually differ in teaching aptitude, it is needed to set up a 'Null-Hypothesis'. This type of hypothesis asserts that the difference between the population means of male and female teachers is zero and that, except for sampling errors, mean differences from sample to sample will all be zero. In accordance with the null hypothesis, it is assumed that the sampling distribution of differences is normal with the mean at zero, or at [M_{pop} (males) – M_{pop} (females)] = 0. The deviation of each sample difference, (M_{males} – $M_{females}$), from this central reference point is equal to [M_{males} – $M_{females}$] – [M_{pop} (males) – M_{pop} (females)], or [M_{males} – $M_{females}$] = 0. The deviation of each sampled difference given in terms of standard measure would be deviation divided by the standard error, which gives us a C.R. value (Z-value) in terms of a general formula:

$$CR = Z = \frac{M_1 - M_2}{\sigma dm}$$

$$CR = Z = \frac{87.43 - 82.58}{1.56} = \frac{4.85}{1.56} = 3.11$$

Since the CR value of 3.11 exceeds 2.58, the null hypothesis may be rejected at the .01 level of significance i.e. the mean teaching aptitude of male and female teachers differs significantly at .01 level.

For the sake of convenience, the researchers have .05 and .01 levels of significance as two arbitrary standards for accepting or rejecting a null hypothesis. When the value of CR or Z is 1.96 or more, we reject a null hypothesis at .05 level of significance. If the 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 trials would a difference of this size arise if the true difference (M_{pop1} – M_{pop2}) were zero.

In testing a null-hypothesis, we make use of a two tailed test or a one-tailed test. If we set up a hypothesis that there is no difference, other than a sampling error difference between the means, we would be concerned only with a difference, and not in superiority or inferiority of either group. For testing this type of hypothesis, we apply a two tailed test since the difference between the obtained means may be as often in one direction (plus) as in the other (minus) from the true difference of zero; and in determining probabilities we take both tails of sampling distribution.

When we are hypothesizing a direction of difference, rather than the mere existence of a difference, we make use of a one tailed test. In determining probabilities, we take either the upper tail or the lower tail of the sampling distribution.

Small Samples

It has already been explained to you that the frequency distribution of small sample means drawn from the same population forms a t-distribution. It is reasonably expected that the sampling distribution of the difference between the means computed from small samples drawn from two different populations will also fall under the category of t-distribution.

The formula for testing a difference between means computed from independent or uncorrelated samples is:

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

in which:

M_1 and M_2 = means of the two samples

$\Sigma x_1^2 + \Sigma x_2^2$ = Sums of squares of the deviations from the means in the two samples

N1 and N2 = Number of cases in the two samples

In the use of t-test, the following conditions need to be fulfilled:

1. The population distribution should be normal. If the population distribution is badly skewed, t-test should not be used.
2. The samples under investigation should have the same variability. This condition is known as 'homogeneity of variances.'

To illustrate the use of t-test, let us consider the following data obtained from the two samples of elementary school in-service teachers enrolled in a distance teacher training programme, one of 10 males and the other of 12 females. The data in respect of means and sum of the square of deviations from means of the intelligence test scores of the samples are as under:

Teachers	Mean	Σx^2	N
Males	9	20.44	10
Females	14	19.60	12

Using t-test:

$$t = \frac{|9-14|}{\sqrt{\left(\frac{19.60+20.44}{12+10-2}\right)\left(\frac{1+1}{12\times10}\right)}}$$

$$= \frac{5}{\sqrt{3.67}} = 8.26$$

Using a two-tailed test, the t critical values for rejection of the null hypothesis at .05 and .01 level of significance for (12 + 10 –2) or 20 degrees of freedom are 2.09 and 2.84 respectively. Since the obtained t value 8.26 is greater than 2.09, the null hypothesis is rejected at the .05 level for 20 degrees of freedom. We may conclude that the difference between the means of intelligence test scores of the male and female teachers is significant. The value 8.26 of t is also greater than 2.84, the null hypothesis is rejected at .01 level for 20 degrees of freedom concluding that there is a significant difference in the means of the two samples at the higher level of confidence.

15.6.2 Significance of the Differences Between the Means of Two Dependent Samples

Means are said to be dependent or correlated when obtained from the scores of the same test administered to the same sample upon two occasions, or when the same test is administered to equivalent samples in which the members of the group have been matched person for person, by one or more attributes.

When the same test is administered to the same sample on two occasions, the formula used for testing the significance of the difference between means obtained in the initial and final testing is:

$$t = \frac{|M_1 + M_2|}{\sigma M_1^2 + \sigma M_2^2 - 2\, r_{12}\, \sigma M_1\, \sigma M_2}$$

in which

M_1 and M_2 = Means of the scores of the initial and final testing.

σ_{M1} = Standard error of the initial test mean.

σ_{M2} = Standard error of the final test mean.

r_{12} = Correlation between the scores on initial and final testing

To illustrate the use of this formula, let us consider the scores of 12 sixth grade students on trial 1. After providing remedial instruction to the students the scores on trial 2 on the mastery learning test in mathematics were also obtained.

Trial 1	Trial 2
53	65
77	87
70	78
85	99
56	66
75	83
57	67
50	45
41	50
66	76
57	55
65	77

$M_1 = 62.66$ $\qquad$ $M_2 = 70.66$

$\sigma_1 = 9.44$ $\qquad$ $\sigma_2 = 15.15$

$\sigma_{M1} = 3.46$ $\qquad$ $\sigma_{M2} = 4.37$

$r_{12} = 0.94$

Using formula:

$$t = \frac{|62.66 - 70.66|}{\sqrt{(3.46)^2 + (4.37)^2 - (2)(0.94)(3.46)(4.37)}}$$

$$= \frac{8.00}{\sqrt{2.6425}}$$

$$= \frac{8}{1.6255}$$

$$= 4.92$$

Since there are 12 students in the sample, we have 12 pairs of scores, the degrees of freedom (df) becomes 12 – 1 or 11. To test the hypothesis that remedial instruction has enhanced the achievement of students in mathematics, we would make use of a one tailed test. In the one-tailed test, for 11 df the .05 level is read from .10 column $\left(\frac{P}{2}=.05\right)$ of the t-table to be 1.80 and .01 level from .02 column $\left(\frac{P}{2}=.01\right)$ is 2.72.

Since our t of 4.92 is greater than 2.72, we may conclude that the gain from Trail 1 to Trial 2 is significant and the hypothesis that the remedial instruction in mathematics has increased the achievement is accepted.

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

2. i) The means and standard deviations of two independent large samples for 95 rural and 64 urban seventh grade students of intelligence test scores are given below:

	Rural	Urban
N	95	64
Mean	24.5	23.4
Standard Deviation	5.12	4.07

Assuming that the samples are random, test the difference between means at .05 level of confidence.

....

....

....

ii) The means and Σx^2 of achievement scores in respect of two dependent small groups matched for intelligence are given below. Compute t-value for testing the significance between the means.

Group 1	Group 2
$N_1 = 20$	$N_2 = 20$
$\sigma_1^2 = 54.76$	$\sigma_2^2 = 42.25$
$M_1 = 53.20$	$M_2 = 49.80$

$r_{12} = 0.60$

....

....

....

....

15.6.3 Significance of the Difference Between the Means of Three or More Samples

We compute CR and t-values to determine whether there is any significant difference between the means of two random samples. Suppose we have N (N > 2) random samples and we want to determine whether there are any significant differences among their means. For this we have to compute $\frac{N(N-1)}{2}$ values of t to determine the significance of the difference between N means by taking two means at a time. This procedure is time consuming and hence, it is advisable to use the technique of 'Analysis of Variance' which would make it possible to determine if any of the two of the N means differ significantly from each other by a single test, called F test, rather than computing $\frac{N(N-1)}{2}$ values of t.

Analysis of variance has the following basic assumptions underlying it which should be fulfilled in the use of this technique.

1. The population distribution should be normal. This assumption, however, is not especially important. Eden and Yates showed that even with a population departing considerably from normality, the effectiveness of the normal distribution still held.

2. All the 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 instance, if we wish to select two groups in a population of B.Ed. trainees enrolled with IGNOU, one of males and the other of females, we must choose randomly from the respective sub-populations. The assumption of randomness is the key stone of the analysis of variance technique. There is no substitute for randomisation.

3. The sub-groups under investigation should have the same variability. This assumption is tested by applying F_{max} test.

$$F_{max} = \frac{\text{Largest Variance}}{\text{Smallest Variance}}$$

In analysis of variance, we have usually three or more groups i.e. there will be three or more variances. Unless the computed value of F_{max} equals or exceeds the appropriate F critical value at .05 level in the Table N of the Appendix, it is assumed that the variances are homogeneous and the difference is not significant.

Suppose we want to apply the technique of analysis of variance to test the significance of the difference between the means of the scores of a test of five groups, one group of greater variance (28.40) with 10 degrees of freedom and the other group of smaller variance (18.06) with 12 degree of freedom. The value of F_{max} is:

$$F_{max} = \frac{28.40}{18.60} = 1.57$$

Since the value of F_{max} 1.57 is less than the value 3.28 of F_{max} at .05 level in the Table N, it may be concluded that the variances fulfilled the condition of homogeneity. It may be noted the table of F_{max} used for testing the homogeneity of variances is different from the F table which is used for testing the significance of the difference between the means of the samples.

Analysis of Variance for Randomized Group Design (One-Way Analysis of Variance)

In an analysis of variance for a randomized group design' (one-way analysis of variance), we examine the relationship between one independent and one dependent variable. The analysis consists of following operations:

i) The variance of the measures (scores) for the randomized groups are combined into one composite group, known as the 'total groups variance' (V_t).

ii) The mean value of the variances of each of the three groups computed separately, is known as the 'within-groups variances' (V_w).

iii) The 'difference between the total groups variance and within-group variance' is known as the 'between-groups variance' ($V_t - V_w = V_b$).

iv) The F-ratio is computed as:

$$F = \frac{V_b}{V_w} = \frac{\text{Between Groups Variance}}{\text{Within Groups Variance}}$$

The within-groups variance represents the sampling error in the distributions. It is also referred to as the error variance or residual. The between-groups variance represents the influence of the variable under study or the experimental variable. If the 'between-groups variance' is not substantially greater than the 'within-groups variance', we may conclude that the difference between the means is probably due to sampling error. If the F-ratio is substantially greater than one, it means that the ratio of the 'between-groups variance' and 'within-groups variance' is probably too large to attribute to sampling error.

The critical values of the F-ratio for .05 and .01 levels of significance are read from the F-table, given in the Appendix. The table indicates two different degrees of freedom, one for V_b (the numerator) and one for V_w (the denominator). The degrees of freedom for the within a groups variance (V_w) is determined in the same way as it is for the t-test i.e., the sum of the subjects (individuals) for all the sample groups minus the number of the groups. If there are K number of groups with number of subjects as N_1, N_2, N_3, N_4 ..... in each group, then ($N_1 + N_2 + N_3 + N_4 + \ldots\ldots - K$) represents the degrees of freedom for within groups variance.

Let us illustrate the computation of F ratio with help of following example. In a remedial instruction programme, 5 subjects were randomly assigned to each of the three groups. Each group was administered the same mastery or criterion test but under slightly different remedial treatments. The scores of the three groups are presented as under:

Group 1		Group 2		Group 3	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
14	196	16	256	12	144
11	121	16	256	19	361
17	289	12	144	20	400
20	400	21	441	24	576
18	324	16	256	14	196
$\sum X_1 = 80$	$\sum X_1^2 = 1330$	$\sum X_2 = 81$	$\sum X_2^2 = 1353$	$\sum X_3 = 89$	$\sum X_3^2 = 1677$

To test the hypothesis that the groups do not differ in mean performance, let us apply the technique of analysis of variance.

The basic assumptions underlying the technique are tested as under:

1. Assumption of Normality

In the light of the finding of Eden and Yates discussed earlier, the assumption of normality may not be considered important for the data of the study.

2. Assumption of Randomness

The requirement of randomness has been amply fulfilled in this study as the subjects were randomly assigned to three treatment groups from the same pool sample.

3. Assumption of Homogeneity of Variance

The F critical value exceeds F max value for the data, it may be assumed that the variances are homogeneous.

After testing the data for the basic assumptions we may proceed with the computations for the analysis of variance using the following steps:

Step 1 : Correction

$$= \frac{\left(\sum x\right)^2}{N} = \frac{\left(\sum x_1 + \sum x_2 + \sum x_3\right)^2}{N}$$

$$= \frac{(80+81+89)^2}{15}$$

$$= \frac{(250)^2}{15}$$

$$= 4166.67$$

Step 2 : Total Sum of Squares (Total SS_t)

$$SS_t = \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 - \text{Correction}$$

$$= 1330 + 1353 + 1677 - 4166.67$$

$$= 4360 - 4166.67$$

$$= 193.33$$

Step 3 : Sum of Squares Between Means of Treatment (SS_b)

$$SS_b = \frac{\left(\sum x_1\right)^2}{N_1} + \frac{\left(\sum x_2\right)^2}{N_2} + \frac{\left(\sum x_3\right)^2}{N_3} - \text{Correction}$$

$$= \frac{(80)^2}{5} + \frac{(81)^2}{5} + \frac{(89)^2}{5} - 4166.67$$

$$= 1280 + 1312.2 + 1584.2 - 4166.67$$

$$= 4176.40 - 4166.67$$

$$= 9.73$$

Step 4: Sum of Squares Within Treatments (SS_w)

$$SS_w = SS_t - SS_b$$

= 193.33 – 9.73

= 183.60

Step 5 : Calculation of Variances for Each SS and Analysis of the Total Variance into Components

Each SS becomes a variance when divided by the degrees of freedom (df) allotted to it. There are 15 scores in all and hence there are (N–1) or (15–1) or 14 df in all. These 14 df are allotted in the following way:

If N = number of scores in all and K = number of groups, we have df for total sum of squares (SS_t) = 15 – 1 = 14; df_w for within treatments (SS_w) = 15 – 3 = 12 and df_3 for between means of treatments (SS_b) = K – 1 = 3 – 1 = 2.

To find the mean square between (MS_b) and mean square within (MS_w), we divide the sum of squares between (SS_b) and the sum of squares within (SS_w) by their respective degrees of freedom (df).

$$F = \frac{MS_b}{MS_w} = \frac{SS_b}{df_b} \Big/ \frac{SS_w}{df_w}$$

$$= \frac{\frac{9.73}{2}}{\frac{183.60}{12}}$$

$$= \frac{4.865}{15.30}$$

$$= 0.318$$

Summary : Analysis of Variance of Three Groups

Source of Variance	df	Sum of Squares (SS)	Mean Square (Variance)	F
Between Groups	2	9.73	4.865	0.318
Within Groups	12	183.60	15.30	
Total	14	193.33		

In our problem the null hypothesis asserts that three sets of scores are in reality the score of three random samples drawn from the same population, and that the means of the treatments of the three groups 1, 2 and 3 differ only through the fluctuations of sampling. To test this hypothesis we divided the 'between means' variance by the 'within treatments' variance. The resulting variance ratio, called F, is to be compared with the F values in the F table presented in the Appendix. The F value in our problem is 0.318 and the df are 2 for the numerator (df_1) and 12 for the denomination (df_2). Entering the F table we read from column 2 and row 12 that an F of 3.88 is significant at .05 level and F of 6.93 is significant at the .01 level. Since the obtained F value of 0.318 is less than the table values, we accept the null hypothesis and conclude that the means of three groups do not differ significantly.

Analysis of Variance for Factorial Design

In many experimental studies, we are concerned with the effect of two or more independent variables, usually called **factors,** on a dependent variable. The number of ways in which a factor is varied is called the number of levels of the factor. A factor that is varied in two ways is said to have two levels and a factor that is varied in three ways is said to have three levels. With two or more factors each with two or more levels; a treatment in some experiments consists of a combination of one level for each factor.

In such multiple classification or factorial designs, both the independent and interactive effects of two or more independent variables on one dependent variable are analysed. The total variance is divided into more than two parts: one part for each independent variable (main effect), one part for each interaction of two or more independent variables, and one part for the residual, or within-group variance. Thus, in a design with two independent variables, the variance is divided into four parts. For example, in case of a three group randomized design we could also divide each of the three treatment groups into males and females. We then have a factorial design with two independent variables, treatments and sex. Since there are three treatment conditions and two conditions of sex, this is a 3×2 factorial design. In the analysis of variance of this design, the variance is divided into four parts: (i) the main effect of treatments; (ii) the main effect of sex; (iii) the interaction effect of treatments with sex; and (iv) residual. Thus we are able to compute three sepearate values for F; one to test the difference among the treatments, one to test the difference between males and females, and one to test the interaction of sex and treatments.

With the availability of computers, analysis of variance can be used with any number of independent variables. However, the analysis does not take into account the correlation between the dependent variables and the pertinent intervening variables. For example, in the studies of memory and learning, for certain reasons it is not possible to equate experimental groups on some pertinent intervening variables at the start of the experiment. In such situations, we may use the technique of **'analysis of co-variance'**.

Analysis of co-variance is an extension of analysis of variance that tests the significance of the difference between means of the final experimental data by taking into account the correlation between the dependent variable and one or more co-variates or pertinent control variables, and by adjusting initial mean differences in the groups. The technique is particularly appropriate when the subjects in two or more groups are found to differ on a pretest of other initial variable. In such a situation, the effects of the pretest and other relevant variables are partialled out and the resulting adjusted means of the posttest scores are compared. The initial status of the experimental groups may be determined by the pretest scores in a pretest – post-test study, or in post-test only studies, by such measures as previous knowledge of subject matter, intelligence, academic achievement etc. The differences in the initial status of groups in terms of such pertinent variables can be removed statistically so that they can be compared as though their intial status had been equated.

To illustrate, suppose in an experiment the dependent variable was the performance in mathematics of three groups of seventh grade distance learners enrolled with National Institute of Open Schooling. One group was taught through computer assisted instruction, second through PSI and the third through conventional method. The three groups were framed by randomly assigning 10 subjects to each of the groups and subjects were equated on the pre-test scores on the criterion (mastery) test in mathematics. Since intelligence was also considered to be a significant factor related to the scores on the criterion test, the scores on the intelligence test were also taken into account so as to provide a statistical control for testing the significance of the difference between the means of criterion test of the three groups at the end of the experiment.

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

3. i) List the assumptions of analysis of variance.

....

....

ii) Thirty subjects were randomly assigned to three groups. Each group performed the same task under three experimental treatments. The scores of the subjects on the task are as under:

Group I: 26, 27, 18, 22, 23, 19, 27, 26, 24, 26

Group II: 18, 22, 18, 23, 19, 24, 20, 21, 19, 25

Group III: 18, 14, 15, 14, 19, 21, 17, 17, 18, 19

Apply analysis of variance and test the significance of the difference between the means of the three groups.

....

....

15.7 STATISTICAL INFERENCE REGARDING PEARSON'S CO-EFFICIENT OF CORRELATION

The mathematical basis for standard error of a Pearson's co-efficient of correlation (r) is rather complicated because of the difficulty in its nature of sampling distribution. The sampling distribution of r is not normal except when population r is near zero and size of the sample is large (N = 30 or greater). When r is high (0.80 or more) and N is small, the sampling distribution of r is skewed. It is also true when r is low (0.20 or less).

In view of this, a sound method for making the inference regarding Pearson's r, especially when its magnitude is very high or very low, is to convert r into Fisher's Z coefficient using conversion table provided in the Appendix and find the standard error (SE) of Z. The sampling distribution of Z co-efficient is normal regardless of the size of sample N and the size of the population r. Furthermore, the SE of Z depends only upon the size of sample N.

The formula for standard error of Z (σ_z) is:

$$SE_z = \sigma_z = \frac{1}{\sqrt{N-3}}$$

To illustrate, let us consider the correlation of 0.55 between scores of an achievement test in science obtained from a random sample of 39 female students of sixth grade.

Using the formula of standard error of Z, we have:

$$SE_z = \sigma_z = \frac{1}{\sqrt{39-3}} = \frac{1}{6} = 0.17$$

For the r = 0.55, the corresponding value of Z is 0.62. Since the sampling distribution of Z is normal, the confidence interval at .95 level for the population of true Z is Z $\pm \sigma_Z \times 1.96$ i.e. $0.62 \pm 0.17 \times 1.96$ or 0.29 and 0.95. The corresponding r's are 0.28 and 0.74, which give a well estimated interval within which we expect the population r with .95 confidence. The chances are 95 in 100 that population r lies between 0.28 and 0.74. For a higher degree of confidence we can take .99 level, for which limits are $Z \pm \sigma_Z \times 2.58$ i.e. $0.62 \pm 0.17 \times 2.58$ or 0.18 and 1.06. The corresponding r's from the conversion table are 0.18 and 0.78. The chances are 99 in 100 that population r lies within 0.18 and 0.78.

The significance of r is also tested by using the formula:

$$t_r = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

With N – 2 degrees of freedom, a co-efficient of correlation is judged as statistically significant when t_r value equals or exceeds the t critical value in the t table.

To illustrate, for r = 0.60 and N = 27:

$$t_r = \frac{(0.60)\left(\sqrt{27-2}\right)}{\sqrt{1-(0.60)^2}}$$

$$= \frac{(0.60)(5)}{\sqrt{1-36}} = 3.75$$

The obtained t_r value of 3.75 exceeds the t critical values of 2.06 and 2.79 at .05 and .01 levels for 25 degrees of freedom which indicates that the correlation of 0.60 is significant.

We can also test the significance of r directly by using the Table K presented in the Appendix. This table presents critical values of r which can be read directly without computing the t value using N – 2 degrees of freedom.

15.8 SIGNIFICANCE OF THE DIFFERENCE BETWEEN PEARSON'S CO-EFFICIENTS OF CORRELATION OF TWO INDEPENDENT SAMPLES

The method of determining the standard error of the difference between Pearson's co-efficients of correlation of two samples is first to convert the r's into Fisher's Z co-efficients and then to determine the significance of the difference between the two Z's.

When we have two correlations between the same two variables, X and Y, computed from two totally different and unmatched samples, the standard error of a difference between two corresponding Z's is computed by the formula:

$$SE_{DZ} = \sigma_{Z_1 - Z_2} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

in which

N_1 and N_2 = Sizes of the two samples

The significance of the difference two Z's is tested with the help of formula:

$$CR = \frac{Z_1 - Z_2}{SE_{DZ}}$$

Illustrating the use of the formula, let us consider the correlations of 0.80 and 0.85 computed between teaching aptitude and teaching attitude scores on the basis of the data obtained from two groups, one males and other females, of B.Ed. trainees enrolled with IGNOU. The data are presented as under:

Male B.Ed. Trainees	Female B.Ed. Trainees
$N_1 = 208$	$N_2 = 220$
$r_1 = 0.80$	$r_2 = 0.85$

To test the significance of the difference between the two r's, we have to first convert these r's into Z-co-efficients. The corresponding Z-coefficients from the Table F given in the Appendix are 1.10 and 1.26 respectively.

Using formula of SE_{DZ}:

$$SE_{DZ} = \sigma_{Z_1 - Z_2} = \sqrt{\frac{1}{208-3} + \frac{1}{220-3}}$$

$$= \sqrt{\frac{1}{205} + \frac{1}{217}}$$

$$= \sqrt{.0048 + .0046}$$

$$= \sqrt{.0094}$$

$$= .097$$

$$CR = \frac{|1.10 - 1.26|}{.097}$$

$$= 1.65$$

The obtained CR value is less than 1.96, the difference between $Z_1 = 1.10$ and 1.26 is not significant at .05 level which indicates that the difference between the corresponding correlations 0.80 and 0.85 is also not significant. We may infer that the correlation between teaching aptitude and teaching attitude does not differ in the two groups drawn from two populations of male and female B.Ed. trainees.

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

4. i) List the advantages of converting Pearson's r into Fisher's Z.

....

....

....

....

ii) Find the .95 confidence interval for population Pearson's correlation co-efficient of 0.78 obtained from a random sample of 84 cases.

....

....

iii) Test the significance of r for the following :

$r = 0.40$

$N = 25$

....

....

15.9 LET US SUM UP

1. In this unit we discussed the statistical inference based on parametric tests. It included the assumptions on which the use of parametric tests are based; inferences regarding means of large and small samples; significance of the difference between the means of two large and small independent samples; significance of the difference between means of the two dependent samples; significance of the difference between means of three or more samples; significance of Pearson's coefficients of correlation; and significance of the difference between Pearson's coefficients of correlation of two independent samples.

2. Parametric statistical tests are based on certain assumptions about the nature of distribution and types of measurement scales used in the collection of data. These tests should be used: (i) when the data are expressed in interval or ratio scales of measurement; (ii) the population values are normally distributed; (iii) the samples have equal or nearly equal variances (homogeneity of variances); and (iv) selection of cases is random.

3. Statistical inferences about sample means are not made with certainty but are based upon probability estimates making use of 'Central Limit Theorem', standard error of the mean, null hypothesis, levels of significance, and one-tailed and two-tailed tests.

4. In testing the significance of the difference between two means, we make use of Z or CR value or t-test depending upon the size of samples.

5. F test is used for testing the significance between the means of three or more sampling. It involves the use of analysis of variance or analysis of co-variance.

6. Analysis of variance is based on the assumptions of: (i) randomness; (ii) homogeneity of variances and (iii) normality. However, the assumption of randomness is the key stone of this technique; failure to fulfil this assumption gives biased results.

7. For testing the significance of Pearson's r, we make use of Fisher's Z transformation or t-test.

8. The inference regarding the significance of the difference between Pearson's r based on two independent sample is made by computing the value of Z using Fisher's Z transformation.

15.10 UNIT-END ACTIVITIES

1. List the assumptions on which the use of Parametric Tests is based.
2. Describe the characteristics of Central Limit Theorem.
3. Define the standard error of mean.
4. Define the confidence intervals and levels of confidence.
5. Given a sample with mean = 35.80, SD = 6.35 and N = 100 Compute the .95 and .99 confidence intervals for the true mean.
6. The mean of 12 independent observations of a test is 80 and SD is 14. Compute the .99 confidence intervals for the true mean.
7. In a study of attitude measurement, a sample of 120 rural school teachers and a sample of 130 school urban teachers scored as below:

Number	Rural Teachers	Urban Teachers
112	115	14.52
108	119	9.81

   Assuming that our samples are random, is the difference between means significant at .01 level.

8. Ten subjects were given 4 successive trials upon a mastery test in physics, of which only the scores for trails 1 and 4 one shown below. Test the hypothesis that remedial programmes between the trials have increased the test scores from initial to final trials significantly:

   Trial 1 : 80, 72, 81,56, 65, 72, 80, 71, 60, 85, 82, 68

   Trail 4 : 91, 85, 79, 62, 71, 84, 91, 82, 71, 87, 91,72

9. Apply Analysis of Variance and test the significance of the difference between the means of scores of the following five groups.

Group 1	Group 2	Group 3	Group 4	Group 5
27	41	34	19	41
37	29	37	20	42
24	35	51	25	38
20	62	52	31	39
18	36	39	38	37

10. State the Assumptions of Analysis of Variance.
11. The Pearson's coefficient of correlation between the height and weight of 100 tenth grade boys is 0.85 and that of 175 tenth grade girls is 0.73. Is the difference between the correlations significant at .01 level?

15.11 SUGGESTED READINGS

Best, John and James V. Kahn (1992): *Research in Education.* New Delhi: Prentice Hall of India Pvt. Ltd.

Edwards, A.L. (1971): *Experimental Design in Psychological Research.* New Delhi: Amerind Publishing Co. Pvt. Ltd. Indian Edition.

Garrett, H.E. (1962): *Statistics in Psychology and Education.* Bombay: Allied Pacific Pvt. Ltd.

Guilford, J.P. (1965): *Fundamental Statistics in Psychology and Education.* New York: McGraw Hill Book Company.

Koul, Lokesh (1997): *Methodology of Educational Research.* Third Revised Edition, New Delhi: Vikas Publishing House Pvt. Ltd.

15.12 ANSWERS TO CHECK YOUR PROGRESS

1. i) Parametric tests should be used when the:

   a) Variables described are expressed in interval or ratio scales;

   b) Samples have equal or nearly equal variances; and

   c) Observations are independent.

   ii) a) 29.02 and 30.98

   b) 28.71 and 31.29

   iii) 3.05 and 7.70

2. i) CR = 1.51, not significant at .05 level.

   ii) t = 2.43, significant at .05 level for 19 df.

3 i) a) The population distribution should be normal.

   b) The sub-groups under investigation should have the same variability.

   c) The groups should be selected at random from the same population.

   ii) F = 14.77, which is significant at .01 level.

4. i) a) The sampling distribution of Z - coefficient is normal regardless of the size of sample N.

   b) The standard error of Z depends only upon the size of sample N.

   ii) 0.68 and 0.85.

   iii) t_r = 2.08, which is not significant at .05 level. Hence r = 0.40 is not significant.