

# **POSTGRADUATE DIPLOMA IN EDUCATION (PDE)**

## **MODULE 2**

### **PDE 110: STATISTICAL METHODS IN EDUCATION**

#### **UNIT 1: TESTING STATISTICAL SIGNIFICANCE**

##### **INTRODUCTION**

Statistics is used in decision making in everyday life. Decisions are made about a population from a given sample. In order to do this, certain assumptions are made upon which hypotheses are based.

In this Unit, we shall look at these hypotheses, how they are tested, errors associated with them and how decision rules are made prior to experimentation.

##### **OBJECTIVES**

At the end of this unit, you should be able to:

1. Define hypothesis.
2. Distinguish between Null and alternative hypotheses.
3. Give the assumptions of hypothesis testing.
4. Distinguish between Type I and Type II Errors.
5. Explain the concepts of levels of significance, one-tailed and two-tailed tests and degree of freedom.

##### **TESTING HYPOTHESIS**

The idea behind statistical inference has two components:

- (a) The Population which consists every one of interest. This means everything you are interested in measuring or observing. The mean value of the attribute of interest (average height, intelligence etc) of every members in this population is  $\bar{N}$  and generally is not known. The attributes of a population are called **parameters**.
- (b) The Sample, which is randomly drawn from this population. The elements of the sample are part of the population but not all of the population. Note that in census,

the sample represents the entire population. The mean of a sample is represented by  $\bar{X}$ . Attributes of a sample such as  $\bar{X}$  are called **statistics**.

Thus, the parameter of a population can be estimated using the corresponding sample statistics. When claims are made about the value of population mean,  $\mu$ , there may be the need to test this. It is only a sample of the population we can test not all members of the population. This implies that one can never know the absolute or exact value of  $\mu$  unless we test all members. The sample can only allow us to reject or not to reject the claim about the population mean using the sample mean. However, our decision may or may not be correct. This is the nature of hypothesis testing. Before hypothesis is tested, some assumptions are made as these determine the pertinent statistic and its sample distribution. These assumptions are:

- That the sample is randomly selected.
- That the population distribution is normal.
- There is independence of two or more variables.
- Equality of the different population variances.

A statistical hypothesis is an assertion about the distribution of one or more random variables or populations. It is frequently denoted by symbols such as  $H_0$  or  $H_1$ . A test of statistical hypothesis is a procedure, based on the observed values of the random variables, that may lead to the acceptance or rejection of the null hypothesis. If a null hypothesis is found to be true then it is accepted (retained), but if not, it is rejected and an alternative form is then opted for. Thus, there are usually two types of hypothesis in a test – the Null hypothesis ( $H_0$ ) and alternative or research hypothesis ( $H_1$ ).

### THE NULL HYPOTHESIS ( $H_0$ )

A null hypothesis states that *there is no significance difference or relationship between two or more parameters*. It involves a judgement as to whether observed differences or relationships are true differences or relationships or whether they are due to sampling error. The chance explanation for the difference or relationship is called the null hypothesis.

You can hypothesize that any apparent difference between the mean achievement of the experimental and control sample groups at the end of an experimental cycle is simply the result of sampling error, in line with the explanation of the central limit theorem. The use of null hypothesis is not limited to experimental studies alone; it may be used also when inferring generalizations about populations from sample data in descriptive research studies.

Though, the null hypothesis is needed for statistical purposes, more often than not, actual hypotheses are usually in the alternatives to the null. **Alternative hypothesis** ( $H_1$ ). states that differences will exist. If you find out that the difference between the mean achievement of the experimental and the control groups is too great to attribute to the normal fluctuations that result from sampling error, then you may reject the null hypothesis. You are saying in effect

that it is probably not true that the difference is merely the result of sampling error. The means no longer behave as random sample means from the same population after the experiment.

In essence, something must have happened to, or affected the experimental group in such a way that it now behaves like a random sample means from a different or changed population. You can therefore conclude that the experimental variable or the treatment probably accounted for the observed difference in performance as measured by the mean scores.

However, if the difference between means was not great enough to reject the null hypothesis, then you do not reject it. You will conclude that “there was No significant difference and that chance or sampling error probably accounted for any observed difference:

In the two situations, what you have done is to use statistical test to either confirm or discount chance or sampling error as an explanation for the difference observed in mean scores.

### **LEVEL OF SIGNIFICANCE**

You will recall that scientific conclusions are statements with high probability of being correct and not statements of absolute truth. Before you carryout the experiment, you must determine how high the probability will be in order to accept or reject a null hypothesis. You need to carry out a test of significance before you can decide whether to reject or accept a null hypothesis. Based on sample data, the test of significance determines whether the difference is significantly greater than a chance difference resulting from sampling error.

This test of significance is usually made at a pre-determined or pre-selected probability level. You may decide to say that the possibility of the observed difference between the mean is due to chance only if the probability is one chance in a thousand or less (0.001), one chance in a hundred or less, (0.01), 5 chances in 100 or less (0.05). If the level of significance is set at .0001, the null hypothesis will be rejected only if the estimated probability of the observed difference being a function of mere chance is one in 10,000 or less.

The lower your probability level, the lower is your risk of rejecting the null hypothesis when it is true. However, this decision increases the probability of retaining the null hypothesis when in fact it should be rejected. Note that you are more confident in your decision when you say you would expect such a difference by chance only 1 time in 1000 (0.001) studies or 1 time in 100 (0.01) studies than if you say you would expect it 5 times in 100 (0.05).

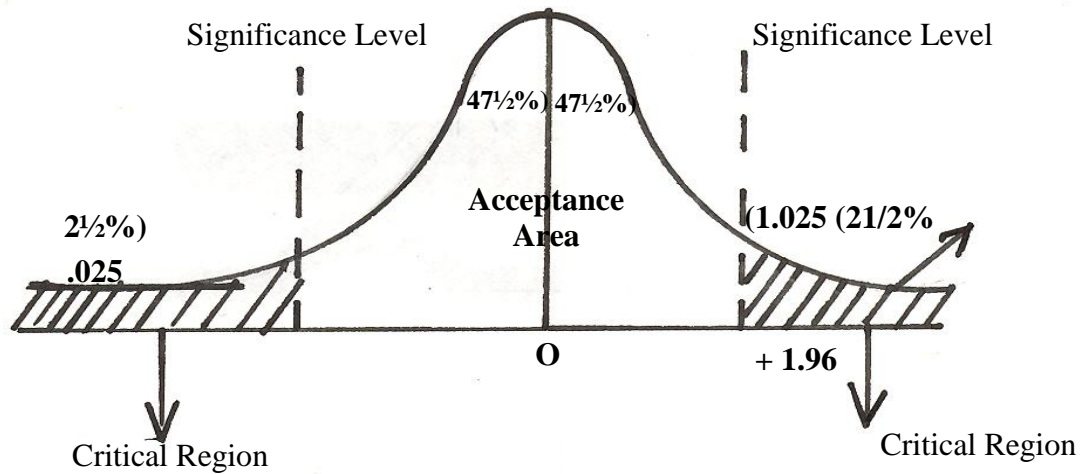
Thus, how confident you are about your decision to reject a null hypothesis or to retain it depends solely upon the level of significance or probability level at which you perform the test of significance. The most commonly used level of significance in the field of education are the 0.05 and 0.01 levels.

### **DECISION MAKING**

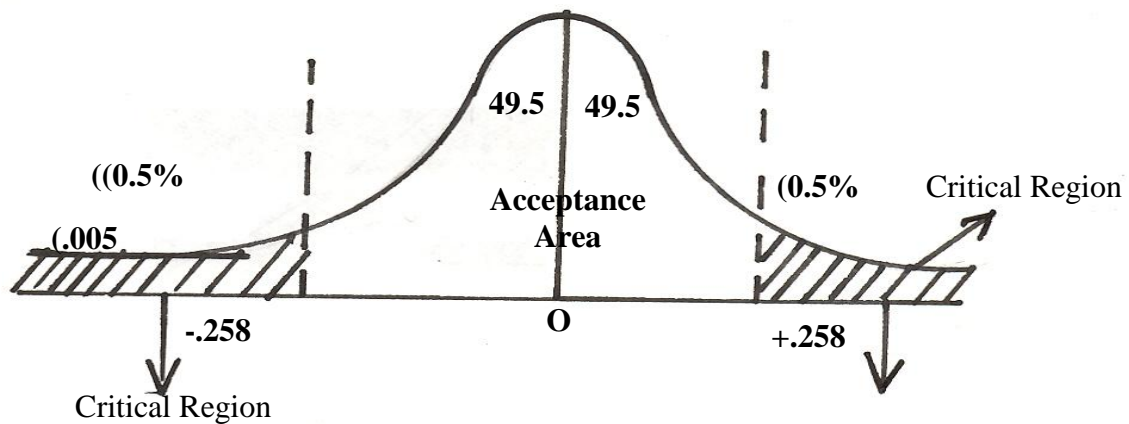
In interpreting the observed difference between the groups, you must choose between the chance explanation, null hypothesis and the explanation that states that there is a relationship between the variables (research or alternatives hypothesis) and you have to do so without

knowing the ultimate truth concerning the populations of interest. This choice is based on incomplete information and therefore subject to possible decision error.

The significance level (or critical level) usually symbolized by  $\alpha$  (alpha) is the probability value that forms the boundary between rejecting and not rejecting the null hypothesis. In education, the precise value is usually taken to be .05. This may be written as " $P < .05$ " in which  $P$  stands for the probability of the null hypothesis. The alpha value, .05 thus represents the probability value that separates a decision to reject null hypothesis from the decision not to reject it.



*Figure - Normal curve showing Critical Region and Significance Level at 95% (.05)*



*Figure - Normal curve showing Critical Region and Significance Level at 95% (.05)*

However, this decision is a guess and it may be wrong. You either reject or retain a null hypothesis. If the null hypothesis is true and you agree that it is true and retain the null hypothesis, then you are correct. But if the null hypothesis is true but you reject it, then you are wrong. You are saying there is a difference between the means whereas there is really no difference. This is an incorrect decision.

The null hypothesis may also be false and you rightly reject it by saying a difference exists between the means. This is also a correct decision. However, it is also possible that that null hypothesis is false, there really is a significant difference between the mean scores, but you retain the null hypothesis that there is no difference between the means; this will be an incorrect decision too.

**TYPE 1 ERROR**

A type 1 error occurs when the decision is to reject the null hypothesis when it is actually true. The probability that a null hypothesis is erroneously rejected by these procedures equals  $\alpha$  - the significance level.

**TYPE II ERROR**

A type II error occurs when the decision is made not to reject the null hypothesis when it is actually false. The probability of this type of error occurring is symbolized by  $\beta$  (beta).

The above represented in a tabular form thus:

(The natural situation which is unknown to research is that the null hypothesis is

		TRUE		FALSE	
		True	Research is correct	Research makes Type II error	
After a test of significance, the researcher concludes that the null hypothesis is	False	Research makes Type I error	Research is correct		

Representation of Types I and Type II errors

The lower your  $\alpha$  is say 0.001, the lower is your rise of committing a Type I error. However, this increases your chances of committing Type II error. In the same vein, as you select a more stringent Beta, you increase the chances of a Type I error. In selecting the probability level,  $P$ , you should consider the relative seriousness of committing a Type I versus a Type II error in your study.

**CONSEQUENCES OF TYPE I AND TYPE II ERRORS**

Consider a new teaching method B that a research or concluded to be superior to method A for teaching statistical concepts. Owing to this, new course books were written, course Tutors were trained on the ‘use of method B. However, after some three or four sessions of further research, no result was found to support this claim. In fact, there were no gains in students achievement. It would appear that the null hypothesis ought not to have been rejected. After this huge expenditure, many course tutors returned to method A.

The researcher will not only be embarrassed, but he would be humiliated. In the extreme, he may loose his or her job as a result of Type I error committed. On the other hand, if he has committed Type II error he would have said that method A be maintained as method B did not indicate any superiority over A. There is no better method. Status quo would be maintained. However, an adventurous Principal had gone ahead to implement method B.

The achievement of the pupils improved tremendously. Method B appeared better. The null hypothesis ought to have been rightly rejected but retained. The harm here is not as much as in committing Type I error. Type I error is more serious.

In view of this, you should do well to guide against Type I error. Your alpha must be appropriate for your study.

## DECISION RULES

Decision rules are statements phrased in terms of the statistics to be calculated, that dictate precisely when the null hypothesis will be rejected and when it will not. The level of significance selected is used to formulate the decision rule.

For example, if the  $\alpha$  level of .05 is chosen, it implies that only 5% or less of the cases will be attributable to chance for the null hypothesis to be retained. This will be P.025 and P.975 of sampling distribution of the mean. These are the end points that would include 95% of the means within the limits but excludes 5% of such means. Of the 5%, 2½ will be extremely low and 2½ extremely high. Note that the curve is perfectly symmetrical. Also note that the sampling distribution of the means is normal in form and so represented as Z - scores (standard normal distribution of the mean, z). The table for Z - values will normally be consulted to read the percentile (P.975) that corresponds to a Z - value such that .025 of the distribution falls to its right (above it) and .025 falls to the left (below it) since the standard normal (Z-th) is a symmetrical distribution.

These points will be 1.96 and - 1.96 to the right and left respectively. Thus, P.025 and P.975 correspond to the points  $Z = -1.96$  and  $Z = 1.96$ . The population mean,  $\mu$  is usually represented by  $Z = 0$ . In your own work, you should also translate your observed sample mean,  $\bar{X}$  to a Z - score. Thus, the decision rule can now be stated as:

If Z is between -1.96 and + 1.96, do not reject  $H_0$ .

If Z is less than or equal to -1.96 or greater than or equal to + 1.96, reject  $H_0$ .

The symbols <, less than and >, greater than, ≤, less than or equal to, and ≥ greater than or equal to can be substituted in the above statement i.e.

If  $-1.96 < z < + 1.96$ , do not reject  $H_0$ .

If  $z \leq -1.96$  or  $z \geq + 1.96$ , reject  $H_0$ .

The z values  $\pm 1.96$  used here to illustrate are called the **Critical Values**. This is because they separate the z values that will result in a decision to reject or not to reject the null hypothesis.

You should note that the sampling distribution of the mean reflects the extent to which the mean of one sample will differ from the mean of another sample taken from the same population, even if these samples are treated in the same manner. If the observed means is not a likely value in this sampling distribution, we reject the null hypothesis. Then, you can say that the observed means was so discrepant from what would have been expected on the basis of random variation from sample to sample (i.e. the basis of sampling error). What you

are saying in essence is that it is unlikely that a mean of this magnitude would be obtained merely because of chance variation from sample to sample. You then go ahead to reject the null hypothesis.

Thus, statistical tests simply assess how different the observed mean was from the population mean, relative to the sampling error one might expect. As this ratio increases, we begin to suspect that the difference between the observed and population means is probably not simply a function of sampling error or random variation between samples or chance.

## POWER OF STATISTICAL TESTS

**The power of a statistical test is the probability that the test will lead to a decision to reject  $H_0$  when  $H_0$  is indeed false.** A good statistical test should have a high power. Though, the power of any statistical test can not be determined, there are some ways of increasing its probability. These are:

1. The power of a statistical test increases as the significance level,  $\alpha$  increases. A test of .05  $\alpha$  is more powerful than that at .01. The less stringent the significance level, the more likely the null hypothesis will be rejected. A larger value of  $\alpha$  will increase the power but will also increase the probability of a Type I error.
2. The power of a test increases with the sample size. The larger the sample the more representative of the population it is.
3. Some statistical tests are more powerful than the others. This has to do with the assumptions upon which the statistic rests.

Parametric tests require that the dependent variable be normally distributed, homogeneously variable across groups, and measured with interval or ratio scales. When any of these assumptions are violated, parametric tests should not be used. Non-parametric tests can therefore be used but then the power may be reduced.

## TWO-TAILED AND ONE-TAILED TESTS OF SIGNIFICANCE

When a null hypothesis is made, we are only interested in whether there is a difference between the means of two samples. We would not say that the mean of A will be greater than that of B or vice versa e.g.

$H_0$ : There is no significant difference between the mean scores of females and males in the statistical method test.

The result of this may give a negative  $-Z$  or  $+Z$  scores. The interpretation remains the same. There is a difference or there is none. Reject or retain the null hypothesis on a stated  $\alpha$  level. This type of hypothesis is said *to be non-directional or that it is two-tailed*.



In a one tailed or directional test, a prediction is made as to what to expect i.e.

*Students who participated in micro-teaching will score higher in practical teaching than those who did not.*

Here, we are concerned with the superiority/inferiority of two groups. Only one end of the curve is required. At 5% level, from the normal probability table, we find that a Z-score of 1.64 cuts off 5% of the mean. At .01  $\alpha$  level, the Z – is 2.33.

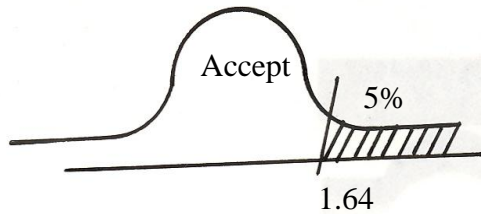


Fig. – Directional Hypothesis at 5% level

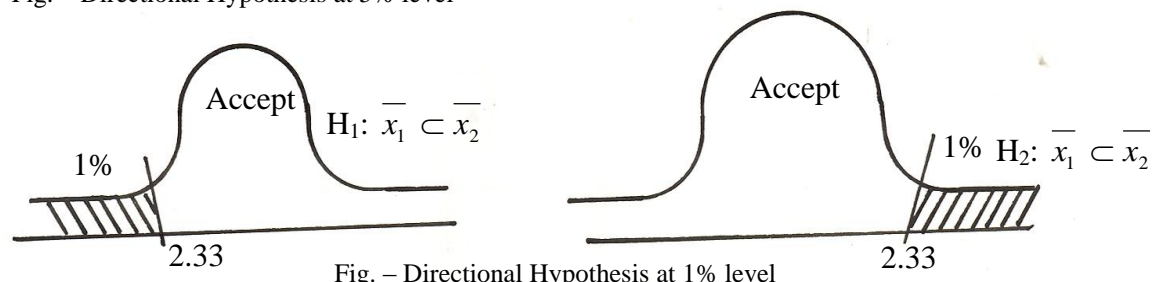


Fig. – Directional Hypothesis at 1% level

Note that the Z-values and invariably the t-values of one-tailed tests are usually smaller than the two tailed values. It is the type of hypothesis that dictates the type of test whether one tail or two tailed test. However, in education, two tailed tests are preferable unless it becomes absolutely necessary to use a one tailed test.

**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 that is that cannot be deduced from each other.** If a coin is thrown up, you can predict that it is either head or tail. Once you know one, you can deduce the other. The degree of freedom here is one. Though, there are two events, you can deduce one from the occurrence of the other.

If you ask a student to name any five odd numbers. He has an infinite degrees of freedom. He can mention any. However, if you ask him to name any five odd numbers between 1 and 10, then you have restricted him. He has just five degrees of freedom i.e. 1, 3, 5, 7 and 9. If you further asked him to list 3 odd numbers between 1 and 10 that will add up to 21 then you have further restricted his chances. Of the five odd numbers, only 5, 7 and 9 will add up to 21. He thus, has three numbers to choose. Of these 3, once he knows any two, the third can be deduced. He has therefore just two degrees of freedom. Each test of significance has its own formula for determining degree of freedom.

The strength of a prediction is increased as the number of independent observations (sample size) or degrees of freedom is increased. Population values can be estimated or inferred more accurately from unbiased large samples rather than small samples. After you have determined whether the test will be two-tailed one-tailed or, selected a probability level and computed the test of significance, you are ready to construct the appropriate table in order to determine the significance of your results. The appropriate table is usually entered at the intersection of your probability level and your degrees of freedom, df.

## **SUMMARY**

In this unit, we have discussed that:

- Inferential statistics dealt with inferences about populations based on the behaviour of samples.
- Some values such as the means are referred to as statistics in samples. The corresponding values in the population are referred to as parameters. Inferential statistics are used to make inferences concerning parameters based on statistics.
- Expected chance variation among the means is referred to as sampling error. Sampling error is normally distributed. If a difference is found between sample means, the question is whether the difference is as a result of sampling error or a reflection of a true difference resulting from the treatment. Differences between two or more sample means are normally distributed around the mean difference in the population.
- The null hypothesis says there is no true difference or relationship between parameters in the population and that any differences or relationships found for the samples is the result of sampling error.
- The null hypothesis is often different from the research hypothesis.
- In order to test a null hypothesis, we need a test of significance, and we need to select a probability or significance level which indicates how much risk we are willing to take that the decision we make is wrong.
- We can make a Type I error by rejecting a true null hypothesis or Type II error by retaining a false null hypothesis.
- The test of significance helps us to decide whether we can reject the null hypothesis.
- There are a number of tests of significance which can be applied – parametric or non-parametric.
- Tests of significance are always two tailed i.e. null hypothesis.
- A one tailed test is directional and assumes that a difference can only occur in one direction.

- Degrees of freedom are a function of such factors as the number of subjects and number of groups.
- The power of a test lies in its ability to reject a false null hypothesis.
- Parametric tests are more powerful and are to be preferred.
- Parametric tests require that certain assumptions be met in order for them to be valid:.
  - The variable measured is normally distributed or that the form of the distribution be known.
  - The data represent an interval or a ratio scale of measurement.
  - Subjects are selected independently (selection of one subject does not affect the other)
  - Variances of the population are homogeneous or at least they are known.
- If one or more of these assumptions is violated, parametric tests should not be used, but a non-parametric one.
- Non-parametric tests make no assumptions about the shape of the distribution.
- Non-parametric tests are used when the data represent an ordinal or nominal scale, when a parametric assumption has been greatly violated, or when the nature of the distribution is not known.
- Parametric tests have the advantage that they permit tests of a number of hypotheses which cannot be tested with a non-parametric test.
- There are a number of parametric statistics which have no counterpart among non-parametric statistics.

### **ASSIGNMENTS**

1. (i) Define the null hypothesis  
(ii) Explain the concept of standard error.
2. Explain the concept of significance level and distinguish clearly between Type I and Type II errors.
3. With good examples, distinguish between one-tailed and two-tailed tests.
4. (i) Explain the concept of degree of freedom.  
(ii) Distinguish between parametric and non-parametric tests.

**REFERENCES**

- Avy, Donal et al (1979): *Introduction to Research in Education* U. S. A. Holt, Rineheat and Winston, Inc.
- Best, J. W and Kahn, J. V. (1986): *Research in Education*. London: Practice Hall Inter.
- Boyinbode I. R. (^1984): *Fundamental Statistical Methods in Education and Research*, Ile-Ife, DC SS Books.
- Gay L. G. (1970) - *Education Research: Competencies for Analysis and Application*. Ohio. Charles E. Merrill.
- Guilford, J. P. and Fruchter, B. (1973): *Fundamental Statistics in Psychology and Education*.
- McCall, R. B (1980): *Fundamental Statistics for Psychology*, U. A. A. Harcourt B. Jovanovich Inc.

## UNIT 2: SAMPLING

### INTRODUCTION

We have discussed how to organize, summarize, describe and present research data using descriptive statistics. In research, the investigator will formulate a problem, propose a set of hypotheses and the procedure to test these hypotheses. Often, it is not possible to study the whole population of concern so, a small portion of it is chosen from which inferences and generalizations are made. For this to be valid, proper and acceptable procedures must be used to select that little portion to be observed. The focus of this unit is to introduce you to sampling. What it is and the available techniques for valid and reliable samples.

### OBJECTIVES

At the end of this unit, you should be able to:

- (1) explain what sampling is.
- (2) describe at least four sampling techniques.
- (3) list the procedures for using each technique.
- (4) apply each of the techniques in research.

### POPULATION OF STUDY

The group you will study is referred to as the population. It is defined as **all members of any well-defined class of people, events or objects**. It is the group about which you want generalization to be made. Each member or item of the group is called a **subject** or **an element**. Population may be type of plants, animals, people, students or any other quantifiable measure. You should note the following about a population:

- ☞ It has at least one characteristic that distinguishes it from another.
- ☞ It may be of any size and may cover almost a given geographical area.
- ☞ It is not often possible to exhaust all its elements i.e. it is infinite. At times, we may target a finite population.
- ☞ The whole group of study may not be available. Only a portion, *the accessible population* is sampled.

The population of interest in a particular investigation is often referred to as **the target population**. However, since it is not possible to deal with the entire target population, the portion of it that one can have access to is called the **Accessible Population**. It is usually from the accessible population that representative samples are taken.

### SAMPLING

Sampling implied the selection of a number of individuals for a study in such a way that the individuals represent the larger group from which they were selected. The individual selected

is called the **Sample** and the larger group from which samples are selected is called the **Population**. A sample is thus a portion of the population of interest.

The purpose of sampling is to gain information about the population because often, owing to the largeness of the population or its geographical spread, it is not economical or feasible to study the entire population. Statistical inference is a procedure by means of which one estimates parameters, characteristics of populations from statistics – characteristics of samples.

If a sample is well selected, research results based on it will be generalizable to the population. However, the estimates from sample remain estimates and are no absolute facts and in such inferences made, a certain degree of error is involved.

### **RATIONALE FOR SAMPLING**

There are some situations where it is possible to study an entire population. This is called a *Census*. The advantages of census include:

- ➔ bias of coverage is reduced.
- ➔ the parameters are those of the population. Conclusions are safely based on observations.
- ➔ public acceptance is easier to secure
- ➔ public compliance and responses may be better secured.

However, in most cases, the whole population may not be available. Sampling is then resorted to. The advantages of sampling include:

- ⊗ economy in terms of time, money, effort personnel and space.
- ⊗ inaccessibility of the population
- ⊗ infinite population
- ⊗ quality and accuracy
- ⊗ speed and timeliness

The general principle of sampling is that it is representative of the accessible population, therefore, findings from the sample can be generalized to that population. The confidence one can have here depends on how representative the sample taken from the accessible population is to the target population. The sample may be biased and the findings of a biased sample cannot legitimately be generalized to the target population.

The closeness of the estimate from the sample of evidence available to us, to the actual target population value (parameter) we would have obtained if a census had been taken is referred to as accuracy. However, accuracy is always difficult to get because the absolute values of population parameters are normally unknown,. If the mean of the sampling distribution of a statistic equals the corresponding parameter, the statistic is called an **Unbiased Estimator** of

the parameter. Otherwise, it is called **Biased Estimator**. These two yield Unbiased and Biased estimates respectively.

**Bias** refers to the difference between the average of the estimates of a population value that would be obtained from a very large number of samples selected by a given procedure and the actual population value, assuming identical measurement processes. Biases are often directional, systematic and usually non-compensatory. **Errors**, unlike biases are variable, random are usually compensatory.

Biases are removed from investigations when representative samples are taken. Various sampling techniques are available to researchers for use in the selection of a sample that will be representative. However, regardless of the specific technique chosen, the basic steps taken to sample a population include:

- ↳ identification of the population
- ↳ determination of the required sample size
- ↳ selection of the sample.

There are four basic sampling techniques. These are *Random Sampling*, *Stratified Sampling*, *Cluster Sampling* and *Systematic Sampling*.

### Random Sampling

Random as used here does not connote haphazard, accidental or without aim or direction. Random sampling is purposeful and methodical. Random sampling is **the process of selecting a sample in such a way that all individuals in the defined population have an equal and independent chance of being selected for the sample**. Every individual has the same probability of being selected and the selection of one individual in no way affects selection of another individual. Simple random sampling as used here means sampling without replacement. This is different from unrestricted random sampling which implies "Sampling with replacement". This means that the individual selected at each draw is replaced into the population before the next draw is made. There are two major ways of carrying out simple random sampling:

(i) *Take – a- pick lottery method*

In this method, the individuals are assigned numbers which are written on piece of paper and then folded properly. These would be put in a container and mixed properly. The required number of individuals are then picked. This method is laborious and the mixing may not be thorough.

(ii) *Use of Random Numbers*

Researchers have produced specially constructed tables of random numbers. The individuals in the target population are enumerated. You will then assign a number to each individual in the population for identification purposes. You will use the number of digits according to the population of study. For 60, use the last two digits, for 500,

the last three digits, and for 1,200 the last four digits.

You need to determine by chance, the point at which the table is entered. You can do this by closing your eyes and touching the page blindly. You, therefore begin wherever the page is touched. The steps involved in this are as follows:

- ≡ identify and define the population
- ≡ determine the desired sample size.
- ≡ list all members of the population
- ≡ assign all individuals on the list a consecutive number from zero to the required number.
- ≡ select an arbitrary number in the table of random numbers by closing your eyes.
- ≡ for the selected number, look at the appropriate number of digits.
- ≡ any number that corresponds to the number assigned to any of the individuals in the population is inclusive in the sample i.e. if the population is 1000 and the number selected is 700, then 700 is in the sample but if the population is 500, then the number assigned 700 is ignored.
- ≡ go to the next number in the column and repeat the step above. Repeat this until the desired number of individual has been selected for the sample.

You can then assign these randomly to two or more treatment groups using the flipping of coin method. Random samplings with small samples do not always guarantee absolute representativeness of the population well. However, random selection with large samples guarantees that differences between the sample and the parent population are only a function of chance and not a result of the researcher's bias.

### **Stratified Random sampling**

When the target population consists of a number of **subgroups** or **strata** that may differ in the characteristics being studied, it is desirable to use stratified random sampling. Stratified random sampling is **the process of selecting a sample in such a way that identified subgroups called strata in the population are represented in the sample in the same proportion that they exist in the population.**

Stratified random sampling is also used in selecting equalized samples from each of a number of subgroups if subgroup comparisons are desired. **Proportional stratified sampling involves the selection of individuals in a sample from the subgroups in proportion to the size of each subgroup in the population.**

Whether one selects equal number of individuals from the strata in the sample or used proportional sampling depends on the nature of research. If the emphasis is on the types of differences among the strata, equal number of cases will be selected from each. However, if the characteristics of the entire population are the main concern, then proportional sampling



is more appropriate. The major advantage of stratified sampling is that of guaranteed representation of defined groups in the population.

When separate estimates are needed for each stratum of the population, then “**Stratified systematic sampling**” is employed. Here the population is stratified using some criteria, then a separate sample is drawn from each stratum with starting points independently determined.

### **Steps in Stratified Random Sampling**

The steps in selecting samples using stratified random sampling include:

- identify and define population
- determine desired sample size
- identify the variable and subgroups for which you want to guarantee appropriate representation.
- classify all members of the population as members of one of the identified strata
- using the table of random numbers, randomly select an appropriate number of individuals from each stratum.

### **CLUSTER SAMPLING**

It is impracticable to obtain or compile a list of all the members of a population say primary school children in Nigeria. The cost of studying subjects of a population that spans a very wide geographical area is prohibitive. Cluster sampling is the kind of sampling in which groups, not individuals are randomly selected. All the members of the selected groups have the same characteristics.

Any intact group of similar characteristics is a cluster e.g. classrooms, schools, city blocks, etc. The procedural requirements for cluster sampling are that:

- = the clusters included in a study must be randomly selected from the population of cluster.
- = once a cluster is selected, then, all the members of the cluster must be included in the sample.

The sampling error in cluster sampling is much higher than in true simple random sampling.

### **Steps in Cluster Sampling**

Stratified random sampling involves the following steps:

- ▶ identify and define the population
- ▶ determine desired sample size

- ▶ identify and define logical cluster
- ▶ list all clusters (or obtain a list) which comprise the population.
- ▶ estimate the average number of population members per cluster.
- ▶ determine the number of clusters needed by dividing the sample size by the estimated size of a cluster.
- ▶ randomly select the needed number of clusters (using a table of random numbers).
- ▶ include in your study all population members in each selected cluster.

The above procedure is often referred to as *single – stage cluster sampling*. *Multi-stage cluster sampling* involves the selection of the elements from the primary sampling units of the clusters. For example, a survey may involve teachers of mathematics in SSI to SS3 in secondary schools. These are the primary units. From these three classes, you can randomly select one class, say SS2. If the class has five arms each, you may use about 3 arms for the study. This is the second stage of the sampling. Any of the methods earlier discussed can then be applied.

### Systematic Sampling

This procedure involves drawing a sample by taking every Kth case from a list of the population. Thus, the major difference between the systematic sampling and the other types of sampling is the fact that all members of the population do not have an independent chance of being selected for the sample. Once K is determined and the first kth element decided then, the rest of the individuals to be included in the sample are automatically determined.

To use this procedure, you need to decide the number of subjects in your sample (n). The already known total number of individuals in the population is N. In order to determine K,

the sampling interval, you simply divide N with n i.e.  $K = \frac{N}{n}$ . The first member is

randomly selected from the first K members of the list and then every Kth member of the population is selected for the sample. Assuming a population of 600 subjects and a sample size (n) of 60. K is given by  $\frac{600}{60} = 10$ .

You can then select any of the first 10 cases, on top of the list and then pick every 10th case from that number. If 5 is chosen, then you add 10 to 5 to pick 15th then, add 10 and pick 25th case etc until you arrive at the required sample size of 60. If the list of the population is randomly ordered, then a systematic sampling can be considered a random sample. The alternative to this is that the selection process be random. However, in practice, randomly ordered lists are rarely available. Thus, systematic sampling is rarely as good as random sampling.

If the list is alphabetical, it is possible that every Kth member of the population might have some unique characteristic that would affect the dependent variable of the study and thus

yield a biased sample. Certain subgroups of the population can be systematically excluded from the sample. Certain national groups tend to cluster under certain letters and the sampling interval could omit them entirely or at least not included them to an adequate extent.

### **Steps in Systematic Sampling**

- ffl identify and define the population
- ffl determine the desired sample size
- ffl obtain a list of the population
- ffl determine what  $K$  is equal to by dividing the size of the population ( $N$ ) with the sample size ( $n$ ).
- ffl pick any number from the list as starting point, then add  $K$  to each number to get the successive  $K$ th term.

The various types of sampling discussed above are the main ones. They are not mutually exclusive. They may be combined in a study. It all depends on the nature of the investigation. However, there are other methods of sampling that are not discussed in this book. Some of these are:

**Area Sampling:** This is a form of multi-stage sampling in which maps rather than lists, serve as the sampling frame. The map is divided into smaller units or areas. An area is randomly selected and then a complete enumeration is carried out in that area.

**Replicated Sampling:** Here, a number of sub - samples rather than one full sample are selected from the population. Each sub-sample is self-contained and adequate sample of the population.

### **Double Sampling**

Double sampling or two-phase sampling is a procedure that is often used in survey studies. The investigator probes the subjects twice. This method is used to gather missing data or information that are absent because the subjects failed to submit the questionnaire. Double Sampling is also used for an intensive follow-up with part of a larger primary sample.

### **NON-PROBABILITY (BIASED SAMPLING)**

This is the method of sampling in which the chances of an element or person being included in the sample is not the same. Some of these methods are:

#### **Quota Sampling**

This is a form of sampling to ensure that all the subsets of the population are represented in the same proportion especially in market and research surveys. The requirement in quota sampling is not that the various population strata be sampled in their correct proportions, but

rather than there being enough cases from each stratum to make possible an estimate of the population stratum value.

### PURPOSIVE SAMPLING

This is the sampling method whereby the researcher just picks those considered to possess the required attributes or information. It is also called *expert choice or judgement sampling*. The common practice is to pick cases that you feel are typical or representative of the population of interest. The assumption here is that errors of judgement in the selection will tend to counter balance each other.

### ACCIDENTAL/HAPHAZARD SAMPLING

This implies choosing any member of the population that is available at the given time until the desired number is achieved. Participation here is based on availability and willingness. Pressmen, Journalists, Astronomers, Pharmacists, Medical Practitioners are some of the fields where representativeness is not the driving force. The precision of this method is not guaranteed. The result may not be generalized into future considerations.

### SAMPLE SIZE DETERMINATION

The size of the sample depends upon the precision the researcher desires in estimating the population parameter at a particular confidence level. The precision of the sample does not depend on the proportion of the population included in the sample  $\left(\frac{n}{N}\right)$  but on the sample size  $n$  of the sample itself. However, the sample should be as large as possible. A larger sample is much more likely to be representative of the population. Larger samples are also likely to be more accurate and precise. This implies that the larger the sample, the smaller the standard error.

The standard error of a sample mean is inversely proportional to the square root of  $n$ , the sample size. This is obvious from the standard error formula for a simple random sample:

$$S . E (\bar{X}) = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$$

In general, the minimum number of subjects believed to be acceptable for a study depends upon the type of research involved. For descriptive research, a sample of 10% of the population is considered minimum. 20% may be required for small populations. For correlational studies, at least 30 subjects. For causal – comparative, 15 subjects per group. However, 30 subjects are considered minimum in general.

Although, larger samples are in general better than smaller samples, even very large samples can lead to erroneous conclusions. There are many sources of sampling bias which can affect a study regardless of sample size. Representativeness is the most important consideration in selecting a sample. Often, the constraint of money and time are limiting factors in selecting adequate sample. However, the formula for estimating the required sample size is given as:

$$S = X^2 NP (1 - P) \div d^2 (N - 1) + X^2 P (1 - P)$$

Where:

- $S$  = sample size being sought
- $X^2$  = table value for chi-square at 1 degree of freedom at the desired alpha level (.05 = 3.84; .01 = 6.64).
- $N$  = Population size
- $P$  = the population proportion (usually .05 as this provides the maximum sample size).
- $d$  = degree of accuracy desired, expressed as a proportion (usually .05).

The table below shows population size and the desired sample size at a chi-square alpha level of .05 and a population proportion also of .05. When the formula is used:

Minimum Sample Size Estimates of a given Population Size

<i>Population</i>	<i>Sample</i>	<i>Population</i>	<i>Sample</i>
10	10	750	254
50	44	1,000	278
100	80	1,500	306
150	108	2,000	322
200	132	3,000	341
250	152	5,000	357
300	169	10,000	370
400	196	30,000	379
500	217	1,000,000	384

You will observe that as the population increases, the proportion of the population represented by the sample tends to decrease. However, in practice, you may not be able to follow this strictly owing to various constraints.

Note that if few subjects are used in a study (N = 6 to 8), then non-parametric analysis are more appropriate. This also depends on the nature of the data. Non-parametric statistics use data on ordinal and nominal scales. Parametric statistics use data on interval and ratio scales. Data on interval and ratio scales can be used with non-parametric statistics with some modifications, though.

**CONCEPT OF SAMPLING ERROR**

Inferential statistics is a procedure by means of which one estimates parameters (characteristics of populations) from statistics (characteristics of sampling). Such estimations are based on law of probability and are at best estimations and not absolute facts. When an inference is made from a sample to a population, a certain amount of error is involved because even random samples can be expected to vary from one to another.

Sampling error is **defined as the difference between a population parameter and a sample statistic**. If the mean of a population is known i.e.  $\mu$  and the mean of random sample is  $\bar{X}$  from that population. Sampling error then is  $SE = \mu - \bar{X}$

**THE CENTRAL LIMIT THEOREM**

The central limit theorem describes the characteristics of sample means if a large number of equal – sized samples (greater than 30 subjects) is selected at random from an infinite population:

- (a) The means of the samples will be normally distributed.
- (b) The mean value of the sample means will be the same as the mean of the population.
- (c) The distribution of sample means will have its own standard deviation. This is the distribution of the expected sampling error. It is known as the **standard error of the mean**. The standard error of a mean is a standard deviation of the distribution of sample means. It is different from the dispersions of individual observations. It is computed from the formula:

$$S_{\bar{x}} = \frac{S}{\sqrt{N}}$$

Where  $S$  = The standard deviation of individual scores.  
 $N$  = The size of the sample.  
 $S_{\bar{x}}$  = The standard error of the mean.

It can be seen that as the size of the sample increases, the standard error of the mean decreases. Thus, as the sample  $N$  approaches infinity, the mean approaches the population mean and the standard error of the mean approaches zero.

$$S_{\bar{x}} = \frac{S}{\sqrt{N}} = \frac{S}{X} = 0$$

As the sample is reduced in size and approaches one, the standard error of the mean approaches the standard deviation of the individual scores.

$$S_{\bar{x}} = \frac{S}{\sqrt{1}} = \frac{S}{1} = S$$

As sample size increases, the magnitude of the error decreases. Sample size and sampling error are negatively correlated.



Sample size.

**Fig: - The relationship between sample size and the magnitude of sampling error.**

It can be safely generalized that as the number of independent observations increases, the error involved in generalizing from sample values to population values decreases and accuracy of prediction increases.

The formula for calculating the standard deviation from a sample is given by:

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

Where  $\sum X^2$  = sum of squares of deviations in the sample.

N = number of cases in the sample.

## SUMMARY

- In this module, we have discussed the rationale for sampling. Distinctions were made between samples and population. Various methods of sampling were discussed. The concept of sampling error was introduced. It was concluded that the larger the sample size the lower the error.

## ASSIGNMENT

- 1 Explain the following terms:
  - (a) population
  - (b) sample
  - (c) parameter
  - (d) statistic
  - (e) Non probability sampling.
2. Discuss four methods of sampling.

**REFERENCES**

- Avy, Donal et al (1979): *Introduction to Research in Education* U. S. A. Holt, Rineheat and Winston, Inc.
- Best, J. W and Kahn, J. V. (1986): *Research in Education*. London: Practice Hall Inter.
- Boyinbode I. R. (^1984): *Fundamental Statistical Methods in Education and Research*, Ile-Ife, DC SS Books.
- Gay L. G. (1970) - *Education Research: Competencies for Analysis and Application*. Ohio. Charles E. Merrill.
- Guilford, J. P. and Fruchter, B. (1973): *Fundamental Statistics in Psychology and Education*.
- McCall, R. B (1980): *Fundamental Statistics for Psychology*, U. A. A. Harcourt B. Jovanovich Inc.



## **UNIT 3: PARAMETRIC STATISTICS**

### **INTRODUCTION**

In Unit one, we discussed various sampling techniques. Remember that “goodness” of the various sampling techniques discussed depend on their effectiveness in producing representative samples. In this unit, we shall discuss methods/techniques of drawing inferences about populations based on the samples, which are called inferential statistics. Parametric Inferential Statistics Methods are concerned with gaining knowledge of a population characteristics from information collected from a sample of the population, which assume that the distributions of the variables being assessed belong to known parameterized families of probability distributions.

### **OBJECTIVES**

At the end of this unit, you should be able to:

1. Explain what a parametric inferential statistics is and states its assumptions;
2. Calculate the t-Test and the Z-Test;
3. Calculate Analysis of variance (ANOVA)

### **PARAMETRIC INFERENTIAL STATISTICS**

Inferential Statistics deals with inferences or generalization about the characteristics of a population based on the behaviour of the sample. Therefore, with inferential statistics, one can draw conclusion about a population of which the data collected represent only a sample. In other words, it is concerned with determining how likely it is that results based on a Sample are the same results which would have been obtained for the entire population.

In making inferences about population based on the behaviour of samples and in testing a hypothesis, there are a number of different methods or tests of significance which can be applied in research studies. Different tests of significance are appropriate for different sets of data. Factors such as the scale of measurement unit represented by the data, method of subject selection, the number of groups, and the number of independent variables determine which test of significance should be selected for a given research design.

The first thing to be considered in selecting an appropriate test of significance is whether parametric test may be used or a nonparametric test. Parametric methods or tests which are our concerned in this unit are always preferred because they are ‘robust’ – that is they are more likely to reject a null hypothesis that is false. In other words, the researcher is less likely to not reject a null hypothesis which should be rejected. However, parametric methods in tests required that certain assumptions be met in order for them to be valid.

Therefore, parametric inferential statistical methods are procedures for statistical hypothesis testing which major assumption is that the distribution of the variable measured is normally distributed or belong to a parameterized families of probability distributions.

## THE ASSUMPTIONS OF PARAMETRIC STATISTICAL ANALYSIS

Parametric statistics require that certain assumptions be met in order for them to be valid. The followings are three very important assumptions that are made when using parametric statistics to test a hypotheses:

1. The variable measured is normally distributed in the population.
2. The normal distributions have the same standard deviation.
3. The data is taken from an interval or ratio scale.

Note that if one or more of these assumptions are violated, a nonparametric test (which will be discussed in Unit 4) should be used. We are now going to discuss some of the commonly used parametric inferential statistical tests by researchers.

### ACTIVITY 1

1. State assumptions that are made when using parametric statistics to test a hypothesis.

## THE t-TEST

The t-test is used to determine whether two means are significantly different when the sample size is small (i.e.  $n < 30$ ). When we use most of the information as contained in the original data in comparison of two groups, there must always, exist a difference no matter how identical the groups are. There are two different types of t-tests, the t-test for independent samples and the t-test for non independent samples.

Independent samples are samples which are randomly formed. The members of one group are not related to members of the other group in any systematic way other than that they are selected from the same population. In a nutshell, the t-test of independent samples is used to determine whether there is probably a significant difference between the means of two independent samples. The formula of t-test for independent samples is:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\frac{\sqrt{S_1^2 + S_2^2}}{n_1 + n_2}}$$

Where  $\overline{X}_1$  &  $\overline{X}_2$  = The means of group 1 and group 2 respectively

$S^2$  = Standard deviation

$n$  = Number of subjects in each group

**Example:** Using the following sets of scores for two randomly formed groups.

Group A = 9, 17, 16, 15, 14, 15, 10, 18, 18, 20, 26, 11, 12

Group B = 6, 10, 12, 18, 13, 1, 11, 9, 19, 5, 15, 10

First, let us label the score and then calculate the mean  $\bar{X}$

$X_1$	$X_2$
9	6
17	10
16	12
15	18
14	13
15	16
10	11
18	9
18	19
20	5
26	15
11	10
<hr style="width: 50px; margin: 0 auto;"/>	<hr style="width: 50px; margin: 0 auto;"/>
$\Sigma X^1$ 210	$\Sigma X^2$ 144

$$\bar{X} = \frac{144}{12} = 12$$

$$\bar{X} = \frac{210}{14} = 15.$$

The next thing is to calculate the  $S^2$ , standard deviation by subtracting the mean ( $\bar{X}$ ) from the scores ( $X - \bar{X}$ ), square each deviation ( $X - \bar{X}$ )<sup>2</sup> and then add the square deviations  $\Sigma(X - \bar{X})^2$ . Divide  $\Sigma(X - \bar{X})^2$  by n, find the square root of  $\Sigma(X - \bar{X})^2$  which is the S standard deviation.

$X_1$	$X_1 - \bar{X}$	$(X_1 - \bar{X})^2$	$X_2$	$X^1 - \bar{X}$	$(X_2 - \bar{X})^2$
9	-6	36	6	-6	36
17	21	4	10	-7	4
16	1	1	12	0	0
15	0	0	18	6	36
14	-1	1	13	1	1
15	0	0	16	4	16
10	-5	25	11	-1	1

18	3	9	9	-3	9
18	3	9	19	7	49
20	5	25	5	-7	49
26	11	121	15	3	9
11	-4	16	10	-2	4
9	-6	39			214
12	-3	9			
		$\Sigma(X - \bar{X})^2 = 295$			$\Sigma(X - \bar{X})^2 = 214$

$$S^1 = \frac{295}{14} = 21.07$$

$$S = \sqrt{21.07} = 5.79$$

$$S^2 = \frac{214}{12} = 17.83$$

$$S = \sqrt{17.83} = 4.50$$

Therefore,

$$t = \frac{15.00 - 12.08}{\sqrt{\frac{5.79^2}{14} + \frac{4.50^2}{12}}} = \frac{2.92}{\sqrt{2.4 + 2.66}}$$

$$= \frac{2.92}{\sqrt{5.1}} = \frac{2.92}{2.25} = 1.3$$

$$\therefore t = 1.3$$

Assuming we selected  $p = .05$ , the only thing we need before we go the t-table is the appropriate degrees of freedom. For the t-test for independent samples, the formula for the degrees of freedom is  $P_1 + P_2 - 2$ . Using the above formula  $\therefore df = 14 + 12 - 2 = 24$

Therefore,  $t = 1.3$

$$p = .05$$

$$df = 24$$

With these, reference is made to the table of  $t$ -distribution. At 0.05 level of significance and 24  $df$ , critical or table value of  $t = 2.064$ . The value 2.064 is the  $t$  value required for rejection or upholding of the null hypothesis. Since the calculated  $t$ -value is less than the critical  $t$ -value (2.064), we do not reject the null hypothesis. Suppose our  $t$ -value was 2.5 we would have rejected the null hypothesis.

**ACTIVITY II**

Using two randomly formed groups of research means,

Group I	Group II
13	12
14	13
15	13
16	13
17	14

Find out whether these scores are significantly different or not. Note that the scores are independent, therefore, the appropriate formula for independent use is that of independent samples.

**T-TEST FOR NON-INDEPENDENT SAMPLES**

Non independent Samples are samples formed by some type of matching. When samples are not independent, the members of one group are systematically related to the members of a second group. Therefore, the t-test for the non independent samples is used to determine whether there is probably a significant difference between the means of two matched or non independent samples or between the means for one sample at two different times. The formula for t-test of non independent samples is:

$$t = \frac{\sum d}{\sqrt{\frac{N \sum d^2 - (\sum d)^2}{N - 1}}}$$

Where  $d$  = difference between each matched samples

$\sum d$  = Sum of the differences between the matched samples.

$d^2$  = the square of the difference between each matched samples.

$N$  = total number of matched samples

$N-1$  = number of degree of freedom.

**Example:** suppose we have the following sets of scores for two matched groups.

$X_1$	$X_2$	$d$	$d^2$
19	29	-10	100
23	63	-40	1600
30	46	-16	256
17	37	-20	400

50	58	-8	64
25	80	-55	3025
10	43	-33	1089
		$\Sigma d + 182$	$\Sigma d^2 \ 6534$

Now let us calculate  $d$  which is the difference between the matched pair ( $X_2 - X_1$ ) then we square the  $d$ , and then sum up  $d$  and  $d^2$  obtained respectively.

Now we have everything we need and all we have to do is substitute the correct number for the correct symbol in the formula.

$$\begin{aligned}
 t &= \frac{-182}{\sqrt{\frac{7 \times 6534 - 182^2}{7 - 1}}} \\
 &= \frac{-182}{\sqrt{\frac{45738 - 33124}{6}}} \\
 &= \frac{-182}{\sqrt{\frac{12614}{6}}} \\
 &= \frac{-182}{\sqrt{2102.33}} \\
 &= \frac{-182}{45.85} \\
 &= -3.97
 \end{aligned}$$

Thus,  $t = -3.97$ . Assuming  $P = 0.05$ , the only thing we need before we go to the  $t$  – table is the appropriate degree of freedom. For the  $t$ -test of non independent samples, the formula for degree of freedom is  $N - 1$ . i.e The number of pairs minus 1. For our example,

$$N - 1 = 7 - 1 = 6. \text{ Therefore, } t = -3.97,$$

$$P = .05$$

$$df = 6$$

With these, reference is made to the table of  $t$  distribution. At 0.05 level of significance and 6df, critical or table value of  $t = 2.365$  This value is the  $\alpha$  value required for rejection or upholding the null hypothesis. This means that with a true null hypothesis and 6 degrees of freedom,  $t$ -ratio of  $+ 2.365$  or more or  $-2.365$  or less will occur by chance five times in a

hundred. Since the calculated t-value of -3.97 is less than the critical value of -2.365, we then reject the null hypothesis..

**THE Z-TEST**

The Z test also is used to determine whether two means are significantly different. It is usually adopted when the sample size is large i.e. when it is equal to or greater than 30 ( $\geq 30$ ). The formula for calculating z test is:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{SD}$$

Where  $\bar{X}$  = The mean of a group

$SD_x$  = Standard error of difference between means

$$SD_x = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$\therefore Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Where  $\bar{X}_1$  = Mean of group 1

$\bar{X}_2$  = Mean of group II

$SD_x$  = Standard error of difference between means.

**Example:** Suppose we have the following sets of scores; let us find the mean, then the standard deviation.

$X_1$	$X_1 - \bar{X}$	$(X_1 - \bar{X})^2$	$X_2$	$X_2 - \bar{X}$	$(X_2 - \bar{X})^2$
3	-2	4	2	-1	1
4	-1	1	3	0	0
5	0	0	3	0	0
6	1	1	3	0	0
7	2	4	4	1	1
<hr/>		10	<hr/>		2
$\Sigma X_1 = 25$		$\Sigma(X_1 - \bar{X})^2 = 10$	$\Sigma X_2 = 15$		$\Sigma(X_2 - \bar{X})^2 = 2$

$$\bar{X} = \frac{25}{5} = 5$$

$$\bar{X} = \frac{15}{5} = 3$$

$$\begin{aligned} SD &= \sqrt{\frac{10}{5}} \\ &= \sqrt{2} \\ SD^1 &= 1.4 \end{aligned}$$

$$\begin{aligned} SD &= \sqrt{\frac{2}{3}} \\ &= \sqrt{0.67} \\ S.D^2 &= \\ &0.82 \end{aligned}$$

Now we have everything we need and all we have to do is substitute the correct number for each symbol.

$$\begin{aligned} \therefore Z &= \frac{5 - 3}{\sqrt{\frac{(1.4)^2}{5} + \frac{(0.82)^2}{5}}} \\ &= \frac{2}{\sqrt{\frac{1.96}{5} + \frac{0.67}{5}}} \\ &= \frac{2}{\sqrt{0.39 + 0.13}} \\ &= \frac{2}{\sqrt{0.52}} \\ &= \frac{2}{.72} \\ &= 2.8 \end{aligned}$$

Assuming we selected  $P = 0.05$ , what we need to do is to make reference to the table of  $z$ -distribution. At 0.05 level of significance, the critical or table value of  $Z = \pm 1.96$ . Since the calculated  $z$ -value of (2.8) is greater than the  $z$  critical value, we reject the null hypothesis; otherwise we do not reject it.



**ACTIVITY IV**

1. What is Z test?
2. Distinguish between t-test and Z-test.
3. Suppose we have the following sets of post tests scores for two groups.
 

<b>Group 1</b>	9, 8, 6, 7, 2, 1, 4, 6, 4, 7, 8, 3, 6, 2, 6, 7, 3, 7, 8, 10, 9, 13, 16, 17, 6, 10, 3, 6, 9, 14, 11.
<b>Group II:</b>	8, 7, 5, 6, 3, 2, 3, 2, 3, 7, 3, 8, 7, 2, 2, 1, 3, 6, 2, 4, 0, 9, 9, 8, 12, 10, 25, 8, 5, 2, 5, 10, 25, 8,

Find out whether these two sets of scores are significantly different or not.

**ANALYSIS OF VARIANCE (ANOVA)**

The analysis of variance commonly referred to by the acronym ANOVA, is used to determine whether there is a significant difference between means of three or more groups, concurrently at a selected probability level. **Anova** not only detects the differences but also brings out the cause or causes of such differences.

As its name suggests, the analysis of variance focuses on variability. It involves the calculation of several measures of variability. The concept underlying analysis of variance is that the total variation or variance of scores can be attributed to two sources – variance between groups (caused by the treatment), and variance within the groups (error variance). As with t-test, a ratio is formed (the F ratio) with group differences as the numerator (variance between groups) and an error term as the denominator (variance within groups). If at the end of a study, the treatment variance is enough larger than the error variance, a significant F ratio results, the null hypothesis is rejected.

If on the other hand, the treatment variance and error variance are essentially the same, the resulting F ratio is not significant and the null hypothesis is not rejected. The greater the difference, the larger the F ratio.

There are two main types of ANOVA. One way ANOVA and two way ANOVA that all analysis of variance end up with the calculation of F- ratio.

**ONE-WAY ANOVA**

This version of ANOVA applies to the case where you have one independent variable and three or more samples of subjects, each sample measured at a different level of variable.

In single classification, or one-way analysis of variance, the relationship between one independent and one independent variable is examined. The analysis of variance consists of the following operations:

- (a) The variance of the scores for three groups are combined into one composite group, known as the **total groups variance** ( $V_t$ ).
- (b) The mean value of the variances of each of the three groups, computed separately, is known as the **within-groups Variance** ( $V_w$ ).
- (c) The difference between the total groups variance and the within-groups variance is known as the between-groups variance ( $V_t - V_w = V_b$ )
- (d) The F ratio is computed as 
$$\frac{V_b}{V_w} = \frac{\text{between - groups variance}}{\text{within - groups variance}}$$

The within-groups variance represents the sampling error in the distributions and is also referred to as the error variance or residual. The between-groups variance represents the influence of the variable of interest or the experimental variable. If the between-groups variance is not substantially greater than the within-groups variance, the researcher would conclude that the difference between the means is probably only a reflection of sampling error. However, if the F ratio were substantially greater than one, it would seem that the ratio of the between-groups variance and the within-groups variance was probably too great to attribute to sampling error.

F ratio is named after Sir Ronald Fisher. The significance of F ratio is found in the F table, which indicates the critical values necessary to test the null hypothesis at selected levels of significance. In the F table, there are usually two different degrees of freedom, one for  $V_b$  (the numerator) and one for  $V_w$  (the denominator).

The degrees of freedom for the between groups variance ( $V_b$ ) is determined by the number of groups minus one i.e. if  $k$  represents the groups, it is  $K - 1$ . The degrees of freedom for the within-groups variance ( $V_w$ ) is determined in the same manner as the t-test-which is, the sum of the subjects for all of the groups minus the number of groups i.e.  $N_1 + N_2 + \dots$  minus  $K$ .

The calculation of F involves finding the mean of the deviations from the mean, squared. Thus, the between-groups variance,  $V_b$  is more commonly referred to as the mean squared between,  $MS_b$  and the within-groups variance,  $V_w$  is more commonly referred to as the mean squared within,  $MS_w$ . The formula for F ratio is then:

$$F = \frac{MS_b}{MS_w}$$

Example, using the following sets of scores of three (3) randomly formed groups.

$X_1$	$X_2$	$X_3$
2	2	2
2	3	3
2	7	2
3	7	2
2	4	1

Now, we need to find out whether these sets of data are significantly different or not. Recall that the total variation or variance is a combination of between (treatment) variance and within (error) variance. In other words:-

$$\text{Total Sum of Square(C)} = \text{between sum of squares(A)} + \text{within sum of square(B) or } SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$$

In order to compute an ANOVA we need each term, but since  $C = A + B$ , or  $C(SS_{\text{total}}) = A(SS_{\text{between}}) + B(SS_{\text{within}})$ , we only have to compute any two terms and we can easily get the third. Since we only have to calculate two, we might as well do the two easiest.  $SS_{\text{total}}$  and  $SS_{\text{between}}$ . Once we have these we can get  $SS_{\text{within}}$  by subtraction,  $SS_{\text{within}}$  will equal  $SS_{\text{total}} - SS_{\text{between}}$ . The formula for  $SS_{\text{between}}$  is as follows:

$$SS_{\text{between}} = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} - \frac{(\sum X)^2}{N} \quad (\text{depending on the number of groups})$$

$$SS_t = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$SS_w = \sum X_1^2 - \frac{(\sum X_1)^2}{n_1} + \sum X_2^2 - \frac{(\sum X_2)^2}{n_2} + \sum X_3^2 - \frac{(\sum X_3)^2}{n_3}$$

Next step is to add the scores in each group for the first three terms in the formula, square the total, and divide by the number of scores in the group. For the fourth term, all we have to do is add up all the scores ( $X_1 + X_2 + X_3$ ), square the total and divide by the total  $N$  i.e ( $n_1 + n_2 + n_3$ )

$X_1$	$X_1^2$	$X_2$	$X_2^2$	$X_3$	$X_3^2$
2	4	2	4	2	4
2	4	3	9	3	9
2	4	7	49	2	4
3	9	7	49	2	4
2	4	4	16	1	1
<hr style="width: 50px; margin: 0;"/>	<hr style="width: 50px; margin: 0;"/>	<hr style="width: 50px; margin: 0;"/>	<hr style="width: 50px; margin: 0;"/>	<hr style="width: 50px; margin: 0;"/>	<hr style="width: 50px; margin: 0;"/>
$\sum X_1 = 11$	$\sum X^2 = 25$	$\sum X_2 = 23$	$\sum X_2^2 = 127$	$\sum X_3 = 10$	$\sum X_3^2 = 22$

Now

$$\begin{aligned} \sum X &= X_1 + X_2 + X_3 &= 11 + 23 + 10 &= 44 \\ \sum X^2 &= X_1^2 + X_2^2 + X_3^2 &= 25 + 127 + 22 &= 174 \\ N &= n_1 + n_2 + n_3 &= 5 + 15 + 5 &= 25 \end{aligned}$$

Let us compute  $SS_{\text{total}}$

$$\begin{aligned}
 SS_{\text{total}} &= \sum X^2 = \frac{(\sum X)^2}{N} \\
 &= 174 - \frac{(44)^2}{15} \\
 &= 174 - \frac{1936}{15} \\
 &= 174 - 129.07 \\
 &= 44.1
 \end{aligned}$$

Let us compute  $SS_{\text{between}}$

$$\begin{aligned}
 SS_{\text{between}} &= \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \frac{(\sum X_3)^2}{n_3} - \frac{(\sum X)^2}{N} \\
 &= \frac{(11)^2}{5} + \frac{(23)^2}{5} + \frac{(10)^2}{5} - \frac{(44)^2}{15} \\
 &= \frac{121}{5} + \frac{529}{5} + \frac{100}{5} - \frac{1936}{15} \\
 &= 24.2 + 105.8 + 20 - 129.07 \\
 &= 20.1
 \end{aligned}$$

Now, we are going to get  $SS_{\text{within}}$  by subtracting  $SS_{\text{between}}$  from  $SS_{\text{total}}$ .

$$\begin{aligned}
 SS_{\text{within}} &= SS_{\text{total}} - SS_{\text{between}} \\
 &= 44.1 - 20.1 \\
 &= 24
 \end{aligned}$$

The next step is to determine the degrees of freedom. The formula for the degree of freedom ( $df$ ) is as follows:

$$df_{\text{between}} = K - 1, \text{ where } K \text{ is the number of groups}$$

$$\therefore df_{\text{between}} = 3 - 1 = 2$$

$df_{\text{within}} = N - K$ , where  $N$  is the total sample size and  $K$  is still the number of groups.

$$\therefore df_{\text{within}} = 15 - 3 = 12$$

$$\begin{aligned} df_{\text{total}} &= N - 1 \\ &= 15 - 1 = 14 \end{aligned}$$

The next step is to determine the mean square.

$$\text{Mean Square} = \frac{\text{sum of degrees } K}{\text{degrees of freedom}}$$

$$MS = \frac{SS}{df}$$

$$\text{for between, } MS_{\text{between}} = \frac{SS_{\text{between}}}{df}$$

$$= \frac{20.1}{2}$$

$$= 10.05$$

$$\text{for within, } MS_{\text{within}} = \frac{SS_{\text{within}}}{df}$$

$$= \frac{24}{12}$$

$$= 2$$

Now all we need is  $F$  ratio. The  $F$  ratio is a ratio of  $MS_{\text{between}}$  and  $MS_{\text{within}}$

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

$$F = \frac{10.05}{2}$$

$$= 5.03$$

Now, we have everything we need, let us fill the summary table of Analysis of variance. When reporting the results of an analysis of variance, it is good to present a summary table such as the below clearly identifying each component of  $SS$ ,  $df$ , and  $MS$ , it allows the reader to take in the main details of the analysis at a glance.

Summary of ONE-WAY ANOVA

Source	SS	df	MS	F	P
Between groups	20.1	2	10.05	5.03	.05
Within groups	24	12	2		
Total	44.1	14			

Now that we have determined the  $F$  ratio, we are ready to go to  $F$  table to see if the calculated values is greater than the table value. Referring to  $F$  – table with 2 and 12  $df$ , and at 0.05 level of significance. The table is entered now where these two values intersect, we find 3.88, the value of  $F$  required for significance. The question is whether an  $F$  value 5.03 is greater than 3.88. If it is, then we reject the null hypothesis and conclude that there are significant differences among the group means.

**ACTIVITY V**

Suppose we have the following sets of posttest scores for three randomly formed groups:

$X_1$	$X_2$	$X_3$
1	1	2
3	2	2
2	3	2
3	4	4
2	5	6

Find out whether these sets of scores are significantly different or not

**TWO – WAY ANALYSIS OF VARIANCE**

The two-way Anova is a procedure that examines the effect of two independent variables concurrently. This not only provides a kind of economy in allowing you to look at two things for the price of one. It also and often much more importantly, allows you to determine whether the two independent variables interact with respect to the effect of the dependent variables.

For example, if we are interested in investigating how sex and course of study (ENG/SOS, PES/SOS and SOS/IRS) affect the attitudes towards general study course in NCE by DLS, the most appropriate Anova to be used is the two-way Anova. This is because two independent variables are involved - Sex and course of study for these different groups (ENG/SOS, PES/SOS, SOS/IRS). This two-way Anova would not only allow you to test the effects of two independent variable (Sex and course of Study) concurrently; but it would also be able to tell you whether the two interact.

Based on our example, the following hypotheses would be stated in the null form:

- Ho1: There is no significant difference between the three groups (course of study) on their attitudes towards General Studies Education (GSE)
- Ho2: There is a significant difference between the sexes on their attitudes towards General Studies Education (GSE)
- Ho3: There is no significance difference between the two independent variables or attitudes towards General Studies Education.

The independent variables are referred to as **factors**. In our example, there are two factors. Factor one (Course of Study) has three levels (ENG/SOS, PES/SOS, IRS/SOS) and factor two (Sex) has two levels (Males/Females).

Based on our example, suppose we have the following scores of the groups according to sex and course of study.

**Scores of the Groups**

		<u><b>Factor A: Courses of Study</b></u>		
		ENG/SOS	PES/SOS	IRS/SOS
<b>FACTOR b: SEX</b>	<b>MALES</b>	9	5	2
		8	5	4
		9	4	4
		6	3	1
		6	3	2
		<b>38</b>	<b>20</b>	<b>13</b>
	<b>FEMALES</b>	7	4	1
		5	6	3
		5	2	2
		6	2	3
		4	1	3
<b>27</b>		<b>15</b>	<b>12</b>	

In considering the effects of the two independent variables separately, what the two-way Anova actually looks at are the differences among the means of the row variables and the differences among the means of the column variables. By convention, these two sets of differences are known as the “*Main Effects*” of the analysis as distinguished from the “*interactive effect*” which is something above and beyond the two main effects. In two-way version of ANOVA, you should always strive for equal group sizes wherever possible.

Note that the number of respondents in group I are subscripted as  $n_{11}$ , which is read as first row, first column and  $n_{23}$  is read as second row and third column group. Now, we shall calculate the sum of squares for total, for two main effects (sex and course of study) for interaction and for within group using our example.

**TWO-WAY ANOVA**

	<i>ENG/SOS</i>		<i>PES/SOS</i>		<i>IRS/SOS</i>		<i>TOTAL</i>		
	$X_1$	$x_1^2$	$X_2$	$x_2^2$	$X_3$	$x_3^2$	$\Sigma X$	$\bar{X}$	$\Sigma X^2$
MALES	9	81	5	25	4	16			
	8	64	5	25	4	16			
	9	81	4	16	2	4			
	6	36	3	9	1	1			
	6	36	3	9	2	4			
							71	4.67	423
$\Sigma$	38	298	20	84	13	41			
	$n_{11} = 5$		$n_{12} = 5$		$n_{13} = 5$				
	$\bar{X}_{11} = 7.6$		$\bar{X}_{12} = 4$		$\bar{X}_{13} = 2.6$				
FEMALES	5	25	4	16	1	1			
	4	16	6	36	3	9			
	5	25	2	4	3	9			
	7	49	2	4	3	9			
	6	36	1	1	2	4			
								3.6	
$\Sigma$	27	151	15	61	12	32	54		244
	$n_{21} = 5$		$n_{22} = 5$		$n_{23} = 5$				
	$\bar{X}_{21} = 5.4$		$\bar{X}_{22} = 3$		$\bar{X}_{23} = 2$				
Grand Total	65	449	35	145	25	73	125		667
Mean Total	6.5		3.5		2.5		7.5		

Now, we are going to calculate total sum of squares ( $SS_{total}$ ). This is determined by subtracting the squared row scores total divided by N from the sum of the square of row scores for the rows and columns. The formula is defined thus:



$$SS_{\text{total}} = \sum X^2 = \frac{(\sum X)^2}{N}$$

$$SS_{\text{total}} = 667 - \frac{(125)^2}{30} = 146.2$$

This is the total variance of all the scores.

1. Let us calculate the SS for Sex row. The formula is:

$$SS_{\text{row}} = \sum \frac{(\sum X)^2}{n} - \left( \frac{\sum X}{N} \right)^2$$

$(\sum X)^2$  = Sum of the row totals divided by the number of subjects in each row add for all the rows.

$$SS_{\text{row}} = \frac{(71)^2}{15} + \frac{(54)^2}{(15)} - \frac{(125)^2}{30}$$

$$= \frac{5,041}{15} + \frac{2,916}{15} - \frac{15,625}{30}$$

$$336.04 + 194.4 - 520.8$$

$$530.16 - 520.8 = 9.66$$

This is the amount of the total sum of squares (variance) that is due to differences between means of the sexes (males and females).

2. Let us calculate the SS for course of study column. The formula is

$$SS_{\text{Column}} = \sum \frac{(\sum X)^2}{\text{Column}} - \frac{(\sum X)^2}{N}$$

$$SS_{\text{Column}} = \frac{(65)^2}{10} + \frac{(35)^2}{10} + \frac{625}{10} - \frac{15625}{30}$$

$$422.5 + 122.5 + 62.5 - 520.8$$

$$607.5 - 520.8 = 86.7$$

This is the amount of the total sum of squares (variance) that is due to the difference between means of the course of study (ENG/SOS, PES/SOS, IRS/SOS).

3. Let us calculate the sum of square for interaction.

$$SS_{\text{interaction}} = SS_{\text{groups}} - SS_{\text{column}} + SS_{\text{rows}}$$

The formula is  $SS_{group} = \sum \frac{(\sum X)^2}{ngroup} - \frac{(\sum X)^2}{Ntotal}$

$$\begin{aligned}
 SS_{group} &= \frac{(38)^2}{5} + \frac{(20)^2}{5} + \frac{(13)^2}{5} + \frac{(27)^2}{5} + \frac{(15)^2}{5} + \frac{(12)^2}{5} - \frac{(125)^2}{30} \\
 &= \frac{1444}{5} + \frac{400}{5} + \frac{169}{5} + \frac{729}{5} + \frac{225}{5} + \frac{144}{5} - 15,625 \\
 &= 288.8 + 80 + 33.8 + 145.8 + 45 + 28.8 - 520.8 \\
 &= 622.2 - 520.8 = 101.4
 \end{aligned}$$

Therefore:  $SS_{interaction} = 101.4 - (9.66 + 86.7) = 5.04$

This is interpreted as part of the total sum square (variances) due to the relationship between the dependent variable (scores or attitude towards GSE and combination of the independent variables).

Let us calculate sum of squares within  $SS_{within}$ , the formula is:

$$SS_{within} = SS_{total} - (SS_{row} + SS_{column} + SS_{interaction})$$

$$\begin{aligned}
 \therefore SS_{within} &= 1462 - (9.66 + 86.7 + 5.04) \\
 &= 146.2 - 101.4 \\
 &= 44.8
 \end{aligned}$$

Now what we need towards calculating F ratio is to determine the number of degrees of freedom for each of the sources of variance. In a two-way analysis of variance, the degrees of freedoms are determined this way:

<b>Source of Variance</b>	-	<b>Degree of freedom</b>
Rows	-	Number of rows - 1
Column	-	Number of Columns - 1
Interaction	-	(column - 1) (rows - 1)
Within Columns	-	Number of subject - Rows X
Total	-	Number of subject - 1

$$df_{row} = 2 - 1 = 1 = r$$

$$df_{column} = 3 - 1 = 2$$

$$df_{interaction} = (3 - 1)(2 - 1) = 2 \times 1 = 2$$

$$\begin{aligned} \text{df within} &= 30 - (2 \times 3) \\ &= 30 - 6 = 24 \\ \text{Total} &= 30 - 1 = 29 \end{aligned}$$

What we need now is to calculate the mean squares by dividing the sum of squares by number of corresponding degrees of freedom.

1. MS row =  $\frac{9 \bullet 66}{1} = 9 \bullet 66$
2. MS column =  $\frac{86 \bullet 7}{2} = 43 \bullet 35$
3. MS interaction =  $\frac{5 \bullet 04}{2} = 2 \bullet 52$
4. MS within =  $\frac{44 \bullet 8}{24} = 1 \bullet 87$

Now, we have all we need to calculate the three F values to test each of the three hypotheses earlier stated. To calculate the F values, we shall divide each of the mean squares by the mean square for within (error term common to all) not accounted for in the sum of variance.

1. F - row (Sex) =  $\frac{9 \bullet 66}{1 \bullet 87} = 5 \bullet 16$
2. F - Column (Course study) =  $\frac{43 \bullet 35}{1 \bullet 87} = 23 \bullet 18$
3. F - Interaction (Sex X Course of study) =  $\frac{2 \bullet 52}{1 \bullet 87} = 1 \bullet 35$

Now, we have everything we need, let us fill the summary table of ANOVA.

**Summary of Two-Way Anova**

<i>Source of Variance</i>	<i>SS Squares</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Row (Sex)	9.66	1	9.66	5.16	
Column (course of study)	86.7	2	43.35	23.18	
Interaction (Sex V CS)	5.04	2	2.52	1.35	
Within	44.8	24	1.87		
Total	164.2	29			

With these we can test the hypothesis and see the decisions to be made. Let us go to F table to see if the calculated value is greater than the table value.

Referring to the F – table to test the first hypothesis which states that “there is no significant difference in attitudes of NCE (DLS) students towards GSE among these groups (ENG/SOS, PES/SOS, IRS/SOS), with 2 and 24 degrees of freedom at 0.05 level of significance. The table is entered where these two values intersect, we find 3.40, the value of F required for significance. The question is whether an F – value of (3.18) is greater than 3.42. If it is, then we reject the null hypothesis and conclude that there is significant difference among the three groups’ (ENG/SOS, PES/SOS, IRS/SOS), means.

Regarding the second hypothesis which states that there is no significant difference in attitudes towards GSE due to sex, with 1 and 24 degrees of freedom at 0.05 level of significance. The table is entered where these two values intersect, we find 4.26. This value is lower than the calculated F – value (9.66), again we reject the null hypothesis.

The third hypothesis which states that there is no significant interaction effect between the variables is tested thus, with 2 degrees 24 degrees of freedom at 0.05 level of significance. The table is entered where the two intersect, we find 3.40. This value is greater than the calculated F – values, therefore we accept the null hypothesis of no interaction effect between the two independent variable and attitudes towards GSE.

**ACTIVITY VI**

Suppose you are interested in investigating how mode of entry into University (UME or DE) and Parents economic status (Rich, Middle class, Poor) affects student’s accommodation in the University. The following were the data obtained from the study:

*Scores of the Group*

*Factor A: Parents Economic Status*

FACTOR b: MODE OF ENTRY	DIRECT ENTRY	RICH	MIDDLE CLASS	POOR
		9	8	7
8	7	6		
7	6	5		
6	5	4		
5	4	3		
		<b>35</b>	<b>30</b>	<b>25</b>
FACTOR b: MODE OF ENTRY	UME	7	9	8
		6	8	7
		5	7	6
		4	6	5
		3	5	4
				<b>25</b>

Based on the above data, test the following null hypotheses;

Ho<sub>1</sub> There is no significant difference in the achievement of the groups based on their parent economic status.

Ho<sub>2</sub>: There is no significant difference in the achievement of the group based on their mode of entry.

## **SUMMARY**

In this unit, you have learnt that:

- Inferential statistics deal with inferences about populations based on the samples
- Parametric inferential statistical methods are procedures for statistical hypothesis testing which major assumption is that the distribution of the variables measured is normally distributed or belongs to a parameterized families of probability distributions.
- The t-test is used to determine whether two means are significantly different at a selected probability level.
- There are two different types of t-tests, the t-test for independent samples and the t-test for non-independent samples.
- Independent samples are samples which are randomly formed, that is, formed without any form of matching.
- Non-independent samples are samples formed by some type of matching.
- Matching, of course, is when the two samples are really taken from the same group at two different times, such as one group which receives two different treatments at two different times or which is presented before a treatment and then post-tested.
- The Z-test is also used to determine whether two means are significantly different. It is usually adopted when the sample size is large ( $\geq 30$ ).
- The analysis of variance is used to determine whether there is significant difference between means of three or more groups concurrently at a selected probability level.
- There are two main types of ANOVA – One-way ANOVA and two-way ANOVA.
- The One-way ANOVA applies to the case where you have one independent variable and three or more samples of subjects, each samples measured at a different level of variable.

- The two-way ANOVA is a procedure that examines the effect of two independent variables concurrently. This not only provides a kind of economy in allowing you to have two things for the price of one.

**REFERENCES**

Anderson, G. (1990) *Fundamentals of Educational Research*: London, the Falmer Press.

Borg, W. R. (1981) *Applying Educational Research: A Practical Guide for Teachers*. New York: Laymic Inc.

Gay, L. G. (1976) *Educational Research: Competencies for analysis and appreciation*: A Bell and Hovel Co. Columbus.

Williams, L. H. (1981) *Statistics for the social Sciences*. New York: The Dryden Press.

## **UNIT 4: NON-PARAMETRIC STATISTICS: COMPUTATION PROCEDURES**

### **INTRODUCTION**

The main distinction between parametric and non-parametric techniques hinges upon the assumption that must be made with regard to the population parameters. In the case of most parametric procedures, such as the t-test and analysis of variance in general, certain assumptions are made regarding the nature of the population data from which the sample data under analysis were drawn. For example it assumed in t-test that the parent population is distributed normally and with equal variability.

Non-parametric statistical methods are applicable to researches that seek to make inferences about a population parameter without rigid mathematical assumptions about the shape or distribution of the population being sampled. As a result, these techniques have been referred to as “distribution free” procedures. However, it is misleading to think of non-parametric test as though they could be legitimately conducted without considering population data. Though never requiring population normality, in some instances nonparametric statistical tests do involve limited assumptions regarding the nature of population distributions. It is certainly true, however, that the assumptions associated with nonparametric tests are much weaker than those demanded by many parametric tests. Statisticians generally agree that when the researcher is working with data, which seriously violate the parameter assumptions, required by appropriate parametric tests, nonparametric procedures are suitable alternatives.

Computation procedures are outlined in this unit for some nonparametric tests, namely, the chi-square test, Mann-whitney U test, the Wilcoxon matched-pairs signed-ranks test, Kruskal-Wallis H test and Friedman two-way analysis of variances.

### **OBJECTIVES**

The objectives of this unit therefore are to enable the students to be able to

1. Determine when to use each of the non-parametric tests in their research work.
2. Compute each of the non-parametric tests.
3. Interpret the result of non-parametric computations

### **THE NEED FOR NON-PARAMETRIC STATISTICS**

Apart from the fact that these tests are used because certain assumptions cannot be made about the populations involved, researchers can use these statistics because:

- (i) they are easier to learn
- (ii) the mathematical derivations are more readily understood.
- (iii) they can be applied to data of a classificatory nature that are common in behavioural research (Nominal or Ordinal scale)

- (iv) they handle ranked data better.
- (v) the computations are relatively simple

The above notwithstanding, parametric tests possess the advantages of being fairly robust with respect to violations of assumptions, having more power – efficiency (the power of a test relative to the sample size which permits one to compare the power of two different statistical tests. The power of a statistical test is the probability that the test will correctly reject the null hypothesis when that hypothesis is false) and sometimes providing more information about a phenomenon (i.e. interactions in the analysis of variance)

### CHI-SQUARE ( $\chi^2$ ) CONTINGENCY TEST

The chi-square test is a measure of relationship, association or independence. Introduced by Karl Pearson in 1900, the chi-square test is probably the best known and the most important of all nonparametric methods. It involves a measure of reliability by comparing observed frequency distributions with theoretical or expected distributions.

There are five basic conditions that must be met for chi-square analysis to be validly applied. These are:

1. The sample observations are independent of each other
2. Sample data are drawn at random from the population.
3. Sample data are expressed in original units.
4. The sample should contain at least 50 observations.
5. There should be no less than five observation in any one cell
6. Not more than 20% of the expected frequency should be less than 5

The  $\chi^2$  can be used to treat data which are classified into nominal, non-ordered categories; it can also be employed with numerical data. The researcher may wish, however, to analyze such data with more powerful parametric tests. But for nominal data, few alternatives to  $\chi^2$  analysis exist.

The basic computation equation for  $\chi^2$  is given below:

$$\chi^2 = \sum \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}} \text{ i.e. } \sum \frac{(O - E)^2}{E}$$

It should be noted that whenever  $\chi^2$  is calculated from 1x 2 (1 by 2) or 2 x 2 (2 by 2) cell tables (instances in which the degree of freedom is one) an adjustment known as Yates correction for continuity must be employed. To use this correction, a value of 0.5 is subtracted from the absolute value (irrespective of algebraic sign) of the numerator contribution of each cell.



$$\frac{(\text{Observed frequency} - \text{Expected frequency})}{\text{Expected frequency}} - 0.5$$

**One Sample Case**

When a researcher intends to determine whether a sample departs significantly from some known or hypothetical distribution, for example, a school proprietor wishes to know whether his 60 recently employed first-year teachers are being drawn from public owned colleges of education or private owned colleges of education in a proportion significantly different from the pattern seen in the previous decade.

In this instance, the basic analytic task is to contrast the proportion of state and private college graduates between (1) the recently employed teachers, and (2) an established proportion drawn from records over the past ten years. The recently employed teachers in this case represent the observed frequencies, while the previously established proportion serves as the basis for calculating the expected frequencies. The data to be used in this example is as presented in the table below

**To calculate the expected frequency:** Step one: determine the previous proportion between state and private college graduates, if this is 40 percent private college graduate to 60 percent state college graduate, this proportion would be applied to the number of the newly employed teachers.

	State College graduate	Private College graduates
Number of teachers employed last year	35	25
Percentage of the teachers employed in preceding 10 years	60%	40%

It is expected that 40%, or 25 of the 60 new teachers would be private college graduates and 60%, or 35 of the 60 teachers would be state college graduate. The observed frequencies and expected frequencies is thus set as the table below.

<i>Frequencies</i>	<i>State College graduate</i>	<i>Private College graduate</i>
Observed	25	35
Expected	36	24

Recall that when calculating  $\chi^2$  from 1X2 or 2X2 tables, Yates correction should be used. Applying Yates correction therefore, we have

$$\begin{aligned}\chi^2 &= \frac{(25-36) - 0.5]^2}{36} + \frac{[(35 - 24) - 0.5]^2}{24} \\ &= 3.67 + 4.59 \\ &= 8.26\end{aligned}$$

The  $\chi^2$  value 8.26 is interpreted from the  $\chi^2$  table of probability values.

For a one-sampled case, in which expected proportions need not be calculated from the sample data, df is always the number of cells less than one thus,  $df = 2 - 1 = 1$ . Consulting the table with one degree of freedom, a  $\chi^2$  value greater than 3.84 is needed to reject the null hypothesis at the 0.05 level. Since the obtained  $\chi^2$  value of 8.26 exceeds 3.84, the null hypothesis that recently employed teachers are being drawn from state and private colleges in the same proportion as in previous years is rejected.

## TWO OR MORE SAMPLE CASES

### Difference between Two Samples

In contrasting differences between two or more samples by a chi-square test, the same principles apply as in the one sample application. Only the methods of determining expected frequencies and the number of degrees of freedom vary.

For example, an investigator wishes to see if boys and girls respond differently to an attitudinal question regarding the educational value of extracurricular student activities. The data collected were as presented in the table below

<b>Responses of 200 boys and girls to question regarding the value of extracurricular activities</b>				
<i>Sex</i>	<i>Very Valuable</i>	<i>Uncertain</i>	<i>Of little Value</i>	<i>Row Total</i>
Boys	40	40	20	100
Girls	60	20	20	100
Colum Total	100	60	40	200 Grand Total

In determining the expected frequencies, since there are equal numbers of boys and girls, as seen in the row subtotals, equal proportion of the column subtotals should be divided between the boys and girls thus, 50 percent of the 100 (40 + 60) frequencies in the “very valuable column would be expected to go to the girls and 50 percent to the boys. Also, the 60 (40 + 20) Uncertain” and 40 (20 + 20) “of little value” frequencies should be divided equally between the boys and girls. Following these procedure we would have the values in the parentheses.

<b>Responses of 200 boys and girls to question regarding the value of extracurricular activities</b>				
<i>Sex</i>	<i>Very Valuable</i>	<i>Uncertain</i>	<i>Of little Value</i>	<i>Row Subtotal</i>
Boys	40 (50)	40 (30)	20 (20)	100
Girls	60 (50)	20 (30)	20 (20)	100
Colum Subtotal	100	60	40	200

*Note that the expected frequencies are in parentheses*

The observed and expected frequencies are then inserted into the  $\chi^2$ -calculation formula as follows and solved accordingly:

$$\begin{aligned} \chi^2 &= \frac{(40 - 50)^2}{50} + \frac{(40 - 30)^2}{30} + \frac{(20 - 20)^2}{20} + \frac{(60 - 50)^2}{50} + \frac{(20 - 30)^2}{30} + \frac{(20 - 20)^2}{20} \\ &= 2.0 + 3.3 + 0 + 2.0 + 3.3 + 0 \\ &= 10.6 \end{aligned}$$

For two or more samples, the degrees of freedom necessary to intercept  $\chi^2$  values such as the above, are always determined from the frequency table by the number of rows minus one times the number of columns minus one.  $(r - 1)(c - 1)$  in the case of the present example  $(2 - 1)(3 - 1) = 2$

The  $\chi^2$  value of 10.6 is significant beyond the 0.01 level. If this satisfies the previously set rejection level, the result indicates that the null hypothesis should be considered untenable and that there is a relationship between sex and response to the question.

**Differences Between More than Two Samples.** The technique for computing a chi-square analysis in which three or more samples are involved is quite similar to the two-sample case. In the table below, a set of data is represented for three secondary schools in the city of Ife in which the number of yearly truancy reports for their student samples is cited. Here the categories for truancy reports are not perfectly equal but are derived in a fairly logical fashion.

<i>School</i>	<i>Number of Truancy Reports by Individual</i>				<i>Row Total</i>
	None	1 - 3	4 - 6	More than 6	
Oduduwa College	400 (400)	100 (87.5)	50 (50)	50 (41.7)	600
St. David's Gram Sch.	300 (266.6)	50 (58.3)	25 (16.7)	25 (20.8)	400

<i>School</i>	<i>Number of Truancy Reports by Individual</i>				<i>Row Total</i>
	None	1 – 3	4 - 6	More than 6	
Ife City Coll.	100 (133.3)	25 (29.2)	25 (16.7)	50 (20.8)	200
Column Total	800	175	100	125	1200 Grand Total

To compute the expected frequencies remember that  $\chi^2 = \frac{(O - E)^2}{E}$

Where  $E = \frac{(\text{Row Total} \times \text{Column Total})}{\text{Grand Total}}$

$$E_{1,1} = \frac{600 \times 800}{1200} \quad E_{1,2} = \frac{600 \times 175}{1200} \quad E_{1,3} = \frac{600 \times 100}{1200} \quad E_{1,4} = \frac{600 \times 125}{1200}$$

$$E_{2,1} = \frac{400 \times 800}{1200} \quad E_{2,2} = \frac{400 \times 175}{1200} \quad E_{2,3} = \frac{400 \times 100}{1200} \quad E_{2,4} = \frac{400 \times 125}{1200}$$

$$E_{3,1} = \frac{200 \times 800}{1200} \quad E_{3,2} = \frac{200 \times 175}{1200} \quad E_{3,3} = \frac{200 \times 100}{1200} \quad E_{3,4} = \frac{200 \times 125}{1200}$$

$E_{1,1}$  implies 1<sup>st</sup> row 1<sup>st</sup> column,  $E_{1,2}$  implies 2<sup>nd</sup> row 2<sup>nd</sup> column etc.

The obtained expected frequencies are in the parentheses

$$\begin{aligned} \chi^2 = & \frac{(400 - 400)^2}{400} + \frac{(100 - 87.5)^2}{87.5} + \frac{(50 - 50)^2}{50} + \frac{(50 - 62.5)^2}{62.5} \\ & + \frac{(300 - 266.6)^2}{266.6} + \frac{(50 - 58.3)^2}{58.3} + \frac{(25 - 33.3)^2}{33.3} + \frac{(25 - 41.7)^2}{41.7} \\ & + \frac{(100 - 133.3)^2}{133.3} + \frac{(25 - 29.2)^2}{29.2} + \frac{(25 - 16.7)^2}{16.7} + \frac{(50 - 20.6)^2}{20.6} \end{aligned}$$

$$\begin{aligned} \chi^2 = & 0 + 1.79 + 0 + 2.50 + 4.18 + 1.18 + 2.07 + 6.69 + 8.32 + 0.60 + 4.13 + 40.99 \\ = & 72.45 \end{aligned}$$

This value of the  $\chi^2$  is intercepted for statistical significance with the degree of freedom equaling  $(r - 1)(c - 1) = (3 - 1)(4 - 1) = 6$ . The  $\chi^2$  value of 72.45 is significant at all levels beyond 0.001.

## WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST

The Wilcoxon matched-pairs signed-ranks test is used to assess the significance of difference between two samples consisting of matched pairs of subjects. Such matched pairs of subject would include two measures taken on the same subject as in a pre- and post comparison of individuals. Wilcoxon is the nonparametric counterpart of the t test for correlated data. The null hypothesis tested is that there is no difference between the two populations from which the matched comparisons are drawn. The implementation of this involved the following steps:

Step 1: the absolute differences between each pair of observations are found, eliminating pairs with an absolute difference of zero.

Step 2: The obtained absolute differences are ranked, assigning the rank of 1 to the smallest difference. Tied values of the absolute differences are given the same rank by averaging the ranks associated with tied positions.

Step 3: The absolute differences are assigned the signs (positive or negative) that they ought to have had if the differences found were not absolute.

Step 4: The sum of the positive or plus differences and that of the negative or minus differences are found. If the sum of the positive differences equals that of the negative differences, It means the null hypothesis is true. But if there is a difference, let the smaller sum be T.

**Example:** A sample of 10 pairs of husbands and wives in a faculty were asked to indicate on a 5-point opinion scale whether under sixteen should be admitted for degree programs. The opinionnaire was administered on husbands and wives separately. A value of 5 in the opinionnaire indicates complete disapproval. The following results are obtained:

<i>Wives</i>	<i>Husbands</i>	<i>Differences</i>	<i>Rank of Difference</i>	<i>Sum of ranks with positive sign</i>	<i>Sum of ranks with negative sign</i>
2	2	0			
2	3	1	1.5		-1.5
1	4	-3	-6		-6
2	2	0			
3	1	2	4	4	
4	4	0			
1	5	-4	-7		-7
3	2	1	1.5	1.5	1.5
3	5	-2	-4		-4
2	4	-2	-4	Total = 5.5	Total = 22.5

Now T is obtained merely by summing the smaller number of the like-signed ranks. In this example the negative ranks are higher and the positive ranks are fewer, so T equals 5.5.

The significance of T, if the number of pairs (n) is 25 or less, can be determined by consulting table of critical values of T in the Wilcoxon Matched-pairs signed-ranks test. In this example, n is 10 so we find that a T value of 3 is needed to yield significance at the 0.05 level. It should be noted that critical values of T in the Wilcoxon Matched-pairs signed-ranks test covers cases involving from six to 25 pairs. When n exceed 25, the following formula is used and the significance of T is determined from the table of the normal curve.

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

To indicate how satisfactory this formula is, even for small samples, we may use the value of T and n from the previous example and compute the value of z. Substituting T = 5.5 and n = 10 in the formula, z is found as follows:

$$Z = \frac{5.5 - [10(10-1)/4]}{\sqrt{[10(10+1)(2 \times 10+1)]/24}}$$

$$z = -0.080$$

## SIGN TEST

The sign test is based on the differences between two pairs of observations without regard to magnitudes. It gets its name from the fact that plus and minus signs are its raw data. The only necessary *assumption for this test is that the variable being studied has a continuous distribution.*

Let there be a pair of observation  $X_1$  and  $X_2$ . The difference  $X_1 - X_2$  will either be a plus, a minus, or a zero. Ignoring pairs for which  $X_1 = X_2$ , the sum of the pairs of observations for which  $X_1 \neq X_2$  constitutes the sample, n. The null hypothesis is that there will be equal number of pluses and minuses. Since zero differences are ignored, the probability of a plus equal the probability of a minus equals 0.5. Employing the binomial, the expected values of pluses, E(P), equals n times the proportion of pluses in the population. The z statistics for the sign test is given by

$$Z = \frac{(P \pm 0.5) - \frac{1}{2}(n)}{1/2 \sqrt{n}}$$

In this formula P +.0.5 is used when P is less than  $\frac{1}{2}(n)$ , and P - 0.5 is used when P is greater than  $\frac{1}{2}(n)$ .

Example: Employing the previous data. Is there any significant difference in the ratings of husbands and wives at the 0.05 level?

Wives	Husbands	Differences
2	2	0
2	3	+
1	4	-
2	2	0
3	1	-
4	4	0
1	5	-
3	2	+
3	5	-
2	4	-

There are 5 minuses and 2 pluses and 3 zeros. Our n is therefore 5 + 2 = 7, ignoring zero differences. Hence,

P = 2, n = 7. Therefore

$$Z = \frac{(2 + 0.5) - \frac{1}{2}(7)}{\frac{1}{2}\sqrt{7}} = \frac{2.5 - 3.5}{0.5(2.65)} = \frac{-1}{1.325}$$

z = 0.755. Since the z-critical ratio for a two-tailed test at the 0.05 significance level is 1.96, and z<sub>c</sub> is 0.755 < 1.96, we conclude that there is no significant difference in the opinion of husbands and wives

**Mann-Whitney U Test**

When the difference between two independent or nonmatched groups is to be assessed, the Mann-Whitney U test represents a powerful alternative to the parametric t-test for uncorrelated samples. The U test needs data which can be ranked. With small samples, U can be calculated in an extremely brief manner. The statistic on which this test is based, that is U, is interpreted for statistical significance in one of three different ways, depending upon the size of the sample involved. This will be demonstrated shortly. Before proceeding, however, it will be necessary to define  $n_1$  as the number of cases in the smaller of the two groups and  $n_2$  as the number of cases in the larger of the two groups.

**Very Small Samples**

When n<sub>1</sub> and n<sub>2</sub> are very small, U is computed in the following order:

Step 1: The scores of the two groups are ranked together but retaining each score's identity as to the group from which it was drawn. Using the following scores as presented below:

Boys Scores: 4, 6, 7, 12, and 15

Girls Scores: 3, 8, and 9.

Combining the ranking we would have

Rank	3	4	6	7	8	9	12	15
Group	Girls	Boys	Boys	Boys	Girls	Girls	Boys	Boys

Step 2: Count the number of boys scores which are ranked below or precede each score in the girls group. The girls score 3 had no Boys score ranked lower. The girls score of 8 and 9 were preceded by three Boys scores each. Hence,  $U = 0+3+3 = 6$ , That is, the number of times a Boy's score preceded a girl's score. Since neither of  $n_1$  and  $n_2$  is larger than 8 the Mann-Whitney subtable that corresponds to the  $n_2$  is to be used to determine the probability of the occurrence of U.  $n_2 = 5$ . Therefore the subtable for  $n_2$  is consulted. In the present example U of 6, when  $n_2 = 5$  and  $n_1 = 3$ , has a 0.393 probability of occurrence.

It some times happens that the value of U is so large that it is not cited in the appropriate subtable. This situation occurs when the researcher has focused on the "wrong" group to determine U. A value of U too large for the table is called  $U^1$ . In the present example  $U^1$  would have been obtained had we used the values of Girls scores that were ranked below the values of boys scores. That is U would have been  $1+1+1+3+3 = 9$ . The subtable  $n_2 = 5$  and  $n_1 = 3$  has no probability value for U of 9. Hence, we know that this U is actually  $U^1$ . We can convert U by the use of the formula:

$$U = n_1n_2 - U^1$$

Thus,  $U = (3)(5) - 9$   
 $= 6$

**Moderately Large Samples**

When  $n_2$  is between 9 and 20, the Mann-Whitney subtable will no longer be suitable for use. Instead critical values of U for a one tailed test or two tailed test is used. This table gives critical values for U for one-and two-tailed tests at commonly used significance levels. For example, if  $n_1 = 8$  and  $n_2 = 16$  in a two-tailed null hypothesis test where the rejection level has been preset at 0.05, a U value of 31 is needed to reject the null hypothesis.

**Large Samples**

When  $n_2$  is larger than 20, a special formula has been developed for interpreting the probability of U by a z value and the normal curve. The formula is stated below:

$$z = \frac{U - \frac{n_1n_2}{2}}{\frac{\sqrt{n_1n_2(n_1 + n_2 + 1)}}{\sqrt{12}}}$$

With large or even moderately large samples, the computation of U by the counting method described earlier is far too time consuming. A procedure which yields identical results is carried out by considering the scores in both groups as though they were in one group and then ranking all scores, giving a 1 to the lowest score, a 2 to the next lowest, and so forth. Then if  $R_1$  equals the sum of the ranks assigned to  $n_1$  group,

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



U for large or moderately large groups can be readily computed using the formula. The obtained value should be inserted in the transformation formula ( $U = n_1n_2 - U^1$ ) to see if  $U^1$  has been obtained noting that the smaller value is U

The formula can then be used to illustrate the hypothetical data presented below, where behavioural rigidity scores and ranks of an experimental and control group students were presented.

Behaviour Rigidity Scores of Nine Experimental and Six Control Students			
<i>Experimental group</i>	<i>Rank</i>	<i>Control group</i>	
47	13	57	15
45	11	54	14
39	9	46	12
37	9	44	10
32	6	35	7
28	5	24	2
27	4		
26	3		$R_1 = 60$
22	1		

Rank all the scores as if they belong to this group. The rankings of the smaller group are summed, to give  $R_1 = 60$ . The  $R_1$  along with  $n_1$  and  $n_2$  are then inserted into the formula  $U = n_1n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$  as follows:

$$U = (6)(9) + \frac{(6)(6 + 1)}{2} - 60$$

$$= 15$$

To ascertain that this value is U and not  $U^1$  the formula  $U^1 = n_1n_2 - U$  is put into use

$$\begin{aligned} U^1 &= (6)(9) - 15 \\ &= 39 \end{aligned}$$

The smaller of these values is U

The next step in the analysis is to insert U,  $n_1$   $n_2$  into the formula

$$U - \frac{n_1n_2}{2}$$

$$z = \frac{-\sqrt{n_1 n_2 (n_1 + n_2 + 1)}}{\sqrt{12}}$$

$$z = 15 - \frac{(6)(9)}{2} = \frac{-12}{8.49} = -1.41$$

$$\frac{\sqrt{(6)(9)(6+9+1)}}{\sqrt{12}}$$

Since a z value of  $-1.41$ , as interpreted from the table of the normal curve, is not significant at the 0.05 level, the null hypothesis would not be rejected.

### TIES

Occasionally, the values of scores are tied. Consider the set of scores (13, 15, 15, 18). The scores of 13 receives rank 1. The two scores 15 are given the average of the next two ranks i.e. 2 and 3 with average 2.5. Each of the two scores of 15 receives a rank of 2.5 score 18 is then assigned a rank of 4.

Occurrence of ties alters the variability of the set of ranks. ties within the same sample cause no theoretical problem in applying the test but ties from different samples suggest the use of the correction formula.

Formula. The Formula is

$$\chi^2_N \left( \frac{n_A n_B}{n_A + n_B (n_A + n_B - 1)} \right) \left( \frac{(n_A + n_B) - (n_A + n_B)}{12} - \sum \left( \frac{t^3 - t}{12} \right) \right)$$

The correction to  $\chi^2_N$  can be ignored for a small number of ties and large  $n_A + n_B$ .

### FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE

The Friedman two-way analysis of variance by ranks is designed to test the null hypothesis that several matched samples have been drawn from the same population. The matched samples may be the same group of subjects who have been exposed to several conditions, or they may be set of individuals who have been matched on relevant variables and are randomly assigned to different conditions. In either situation the condition represents the independent variable while the criterion data, which must be amenable to ranking, represent the dependent variable. The test statistics for the Friedman test for large samples is

$$X^2_r = \frac{12}{NK(K-1)} \sum (R_j)^2 - 3N(K+1)$$

N = the number of rows

K = the number of column

$(R_j)^2$  = direct one to first square, then sum all column rank totals

Example In an experiment, 12 groups of matched subjects were randomly assigned to 5 experimental treatments, A, B, C, D, and E. The scores are represented as in the table below.

Groups	A	RANK	B	RANK	C	RANK	D	RANK	E	RANK
1	3	3	5	4	6	5	2	2	1	1
2	1	1	4	3	3	2	7	5	5	4
3	4	3	6	4	7	5	2	2	0	1
4	1	1	2	2	4	3	5	4	6	5
5	0	1	4	3	10	5	8	4	3	2
6	1	1	4	4	3	3	2	2	5	5
7	6	3	7	4	3	1	4	2	9	5
8	8	6	5	3	6	5	2	1	4	2
9	2	3	4	5	3	4	0	1	1	2
10	1	1	3	2	4	3	8	5	7	4
11	4	3	5	4	3	2	2	1	6	5
12	0	1	3	3	2	2	4	4	9	5
		27		41		40		33		41

In this instance, the most interesting treatment was ranked 1, the next most interesting treatment 2, the ranking is done in that order until when it gets to 5 which indicates the least most interesting treatment for all the groups. In the Friedman test, it makes no difference whether rankings are from lowest to highest or highest to lowest. Tied ranks are assigned the average rank of the ties. The next step is to sum the ranks in each of the columns as has been done in the table above . The values can then be inserted in the formula

$$X_r^2 = \frac{12}{NK(K+1)} \sum (R_j)^2 - 3N(K+1)$$

$$X_r^2 = \frac{12}{12(5)(6)} (27^2 + 41^2 + 40^2 + 33^2 + 41^2) - 3(12)(6)$$

$$= 14.79$$

The probability of  $X_r^2$  is generally determined by using the standard  $\chi^2$  table with df = the number of columns less one. However, when the number of columns is 3 and the number of rows 2 – 9, Or when the number of columns is 4 and the number of rows 2 – 4, the Friedman table should be used. Since in this example there are five columns and 12 rows, we shall consult the standard  $X^2$  table. With df = 4, the obtained value is significant at 0.05 level.

**KRUSKAL WALLIS H-TEST (INDEPENDENT SAMPLES)**

The Kruskal-Wallis test is somewhat an extension of Mann-Whitney test in the same way as the one-way analysis of variance F-test is to the test. The H-test provides an alternative (to the one-way (ANOVA) of finding the significance of the difference between more than two means. However, it makes use of ranks instead of raw scores. When the null hypothesis is true and  $n_j \geq 5$  for all j, the H statistic is defined by

$$H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)$$

Where

$R_j$  = the rank sum associate with jth sample

$n_j$  = the number of observations comprising the jth sample

$$n = n_1 + n_2 + \dots + n_k$$

= total number of observations over all samples.

For large samples, H is approximately chi-square distributed with  $K - 1$  degrees of freedom.

The procedure for computing H is relatively simple. First, all the k samples are pooled to form one singular sample of size n; the array is arranged from smallest to largest and then ranked. Next the rank – sum associated with each independent group is obtained. This value is  $R_j$ ; the corresponding sample size for each group is also found ( $n_j$ ). When the number of cases in each sample exceeds five, H may be interpreted for level of statistical significance from a chi-square table with  $df =$  the number of columns, or samples, less one. However, when there are only three samples and the number of cases per sample is five or less, then the kruskal-Wallis one-way analysis of variance by ranks table is used to interpret the probability of the occurrence of H.

**Example** In a part two final examination taken by students who entered the University by (i) prelim, (ii) direct and (iii) Transfer, the distribution of scores is as follows.

<i>Prelim</i>	<i>Direct</i>	<i>Transfer</i>
69	66	49
67	64	57
60	70	55
80	73	75
74	71	62
65	79	
68	61	
	59	

The null hypothesis to be tested is that there is no difference in the distribution of examination scores over the three groups.

**The computation**

<i>Exam Scores</i>	<i>Group</i>	<i>Rank</i>	<i>Ranks for Prelim</i>	<i>Ranks for Direct</i>	<i>Ranks for Transfer</i>
49	T	1			1
55	T	2			2
57	T	3			3
59	D	4		4	
60	P	5	5		
61	D	6		6	
62	T	7			7
64	D	8		8	
65	P	9	9		
66	D	10		10	
67	P	11	11		
68	P	12	12		
69	P	13	13		
70	D	14		14	
71	D	15		15	
73	D	16		16	
74	P	17	17		
75	T	18			18
79	D	19		19	
80	P	20	20		
Sum of ranks for the groups $R_i$		$R = 20$	$R_1 = 87$	$R_2 = 92$	$R_3 = 31$
Size of groups $n_i$		$n = 20$	$n_1 = 7$	$n_2 = 8$	$n_3 = 5$

$$\begin{aligned}
 H &= \frac{12}{(20)(21)} (87^2/7 + 92^2/8 + 31^2/5) - 3(20 - 1) \\
 &= 66.61 - 57 \\
 &= 9.61
 \end{aligned}$$

Since the number of cases in each sample exceeds five, the chi-square table will be used to interpret the level of statistical significance. Hence comparing the  $H = 9.61$  with the chi-square ( $\chi^2 = 5.99$   $df = 2$ ) at 0.05 level of significant, the null hypothesis is rejected.

**ACTIVITIES**

1. In the last one decade the average percentage of students registering local languages at the private School certificate Examination has been the following: Yoruba, 35%; Igbo, 40; Hausa, 15%; Efik, 10%. This year an apparent marked divergence from previous enrollments was noticed. The 500 students that registered for local languages in the recent year are distributed as follows: Yoruba, 200; Igbo, 150; Hausa, 100; Efik, 50. Using a  $\chi^2$  analysis , determine whether this year registration for local languages is significantly different (at the 0.05 level) from the average distribution pattern during the last decade.
2. Below are responses of two groups of students to a questionnaire item. Using a chi-square two-sample tests determine whether the null hypothesis regarding group differences should be rejected.

	A	B	C	D
Male	60	50	40	45
Female	20	10	30	32

3. Two samples consisting of matched pairs of elementary pupils have been given different sets of spelling materials. The two students in each pair were matched in IQ and scores on an extensive spelling pretest. Below are their scores in a post training spelling examination. Using Wilcoxon matched pairs signed ranks tests, determine whether the performance of the groups using spelling materials X is significantly different beyond the 0.05 level from the group using spelling materials Z

*Spelling Examination Score*

Pair Number	Materials X	Materials Z
1	28	31
2	33	35
3	40	39
4	36	39
5	40	43
6	14	21
7	16	19

8	21	26
9	32	30
10	43	56

4. Using Mann-Whitney tests determine the value of U and decide whether the married significantly out-performed the singles at the 0.05 level. Assume that you have made a prediction in advance of the data gathering that the married will perform better than the singles.

<i>Married scores</i>	2	3	5	6	9	9	10
<i>Singles scores</i>							

5. The following are personality test scores for three groups of secondary teachers, whose college majors were English, Psychology, and Sociology. Using the Kruskal-Wallis one-way analysis of variance, determine the value of H and decide whether the three samples should be considered part of the same population

<b>Psychology</b>	<b>English</b>	<b>Sociology</b>
64	65	59
63	42	42
58	39	29
43	30	
40		

**SUMMARY OF STATISTICAL TESTS AND THEIR DISTINGUISHING CHARACTERISTICS**

**A. TESTS OF THE DIFFERENCES BETWEEN GROUPS**

Types of Sample			Distribution	Variance	Minimum Scale Requirement
Two Independent Samples	(i)	t - Test	Normality	Homogeneity	Interval
Two	(ii)	Mannwhitney U	-	-	Ordinal
Correlated Samples	(iii)	Wvilcoxon test	-	-	Ordinal
Several Independent Samples	(i)	ANOVA	Normality	Homogeneity	Interval
	(ii)	Two Factor ANOVA	Normality	Homogeneity	Interval
	(iii)	Kruska Wallis Test	-	-	Ordinal

**B. TESTS OF ASSOCIATION**

One Sample	(i)	<b>Pearson product-moment correlation</b>	<b>Normality</b>	-	<b>Interval</b>
	(ii)	Spearman Rank-order correlation	-	-	Ordinal
	(iii)	Chi-square test for r x c tables	-	-	Nominal
Two Independent Samples	(i)	Difference between Two Pearson correlations	Normality	-	Interval

**REFERENCES**

Agresti, Alan (1996). *Introduction to categorical data analysis*. NY: John Wiley and Sons. Agresti discusses Mantel-Haenszel chi-square stratified analysis on pp. 231-236.

Conover, W. J. (1998). *Practical Nonparametric Statistics* (3rd Ed.)



Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods* (2nd Ed.).

Lehmann, E. L. (1975). *NONPARAMETRICS: Statistical Methods Based On Ranks*

Lieberman, Bernhardt, ed. (1971). *Contemporary problems in statistics*. NY: Oxford. Section 5 deals with assumptions of chi-square procedures.

Wilcoxon, F. (1945). "Individual comparisons by ranking methods". *Biometrics Bulletin*, 1, 80-83.