

TEST DEVELOPMENT AND CORRELATION - I

Unit Structure :

- 1.0 Objectives
- 1.1 Introduction
- 1.2 Test Conceptualization
- 1.3 Test Construction;
 - 1.3.1 Scaling
 - 1.3.2 Types of Scales
 - 1.3.3 Method of Scaling
 - 1.3.4 An item on the scale is
- 1.4 Test Tryout
- 1.5 Item Analysis
 - 1.5.1 An index of the item's difficulty.
 - 1.5.2 Analysis of item alternatives
- 1.6 Test Revision
 - 1.6.1 Test Revision as a stage in New Test Development.
 - 1.6.2 Standardization
- 1.7 Summary
- 1.8 Questions
- 1.9 References

1.0 OBJECTIVES

The objective of this unit is –

- To impart knowledge and understanding about test development.
- To create awareness about the technical process of construction of a good test.
- To explore number of techniques designed for construction and selection of good items.
- To compare the custom-made test with those of newly constructed tests.

1.1 INTRODUCTION

Developing a test is not an easy task and a good test is not developed by chance, it requires a great deal of thoughtfulness and sound application of standardized principles based on statistical techniques.

Test construction occurs in five stages:

- a. Test conceptualization
- b. Test construction
- c. Test tryout
- d. Item analysis
- e. Test revision

Test conceptualization refers to a novel idea for a test to be conceived. Items for the test drafted refers to as construction. The first draft of the test is then applied on a group of testtakers in the sample (test tryout). Once the data from tryout are collected, performance of testtaker on each item is analyzed.

Then, item analysis in terms of statistical procedures is applied to know which items are good and which items need to be revised and which of them needs to be dropped or discarded. And finally, the results are analyzed and further revised if necessary- so it goes.

1.2 TEST CONCEPTUALISATION

In behaviour terms, the starting point to develop a new test begins with self-talk and test developer asks himself something like "which type of test should be designed that measures a 'construct' for which the test is being developed. For example, once a new disease comes to the attention of medical researchers, they try to develop diagnostic tests to assess its causes, symptoms, presence or absence, its severity of manifestations in the body. Thus, development of a new test may be a response to a need to face severe situation and to develop mastery in the fields of occupation or profession, telecommunications and computer networking, etc. However, a test developer is confronted with a number of questions when he develops a new test. Some of them are as under:

- *What is the test designed to measure?* Thus, a simple deceptive question which is related to how the test developer defines the construct being measured? How this definition is different from another test measuring the same construct?
- *What is the objective of the test?* This question is directly related to aim / goal and purpose of the test.
- *Is there a need for this test?* Are other tests available to measure the same attribute or trait? Are these tests reliable and valid? In what ways the newly constructed test will be better than other existing tests?

- *Who will use this test?* Clinicians? Educators? Others? What purpose this test would serve?
- *Who will take this test?* This question is related with age range and qualification of test takers.
- *What content will the test cover?* This question covers the content of existing test and the content of culture-specific.
- *How will the test be administered?* This question is linked with the application of test on individual or group and on both.
- *What is ideal format of the test?* is it in true/false form, essay type, multiple - choice, or in some other forms? which is the best form selected for administration?
- *Should more than one form of the test be developed?*
- *What special training will be required for test users for administering or interpreting the test?* This question requires background and qualifications of test users.
- *Who benefits from an administration of this test?*
- *How will meaning be attributed to score on this test?* This question is related with scores of one test user to another taking the test at the same time and others in a criterion-group.

The last question needs attention relating to test development with regard to norm-versus criterion- referenced tests.

Norm- referenced versus criterion- referenced tests. '(Item development issues).

There are two main important approaches to test development and individual item analysis. They are norm referenced or criterion-referenced approaches.

A good item on a norm-referenced achievement test is an item that scores high on the test items answered correctly. Whereas, low score indicates on the same item answered incorrectly. In other words, High scorers on the test get a particular item right whereas low scorers on the test get that same item wrong. But in a criterion oriented test, high scores are referred to as right and low scores as wrong.

However, criterion-oriented test is used in licensing contexts, a license to practice medicine or a license to drive a car. This approach is also employed in educational contexts, to strengthen the knowledge, skills or both of students in class room teaching. The test developer may attempt to sample criterion- related knowledge relevant to the criterion being assessed. For the targeted skills or knowledge, they may conduct experimentation with different items, tests, formats or measurement procedure which may help them to discover the best measure of mastery for required cognitive or master skills.

Whereas, norm-referenced approach is insufficient and inappropriate when knowledge of mastery is required for test user.

However, the best items are those items that discriminate between these two groups.

Pilot Work.

In behaviour sciences, pilot work, pilot study and pilot research generally refer to the preliminary administration of test on selected sample before final administration.

Commonly, pilot study is conducted to evaluate the reliability and validity of newly constructed test and to find out whether it may be included in the final form of the instrument. For this purpose, a structured interview is conducted. In addition, interviews with parents, teachers, and others may be arranged.

In pilot study, a test developer attempts to find out to measure a targeted construct. This process may help him in creation, revision, deletion of many test items, literature reviews, and related activities. The need for additional pilot study is always knocking at door.

Check your progress

1. Define conceptualization.

2. What are the processes of developing a test?

3. Explain any four preliminary questions which a test developer is immediately confronted with?

4. What is pilot work, pilot study or pilot research? Explain.

1.3 TEST CONSTRUCTION

After finalization of pilot study, test developer pays attention to formal aspect of professional test construction. The process of test construction begins with scaling.

1.3.1 Scaling

Scaling may be defined as "the process of setting rules for assigning numbers in measurement". In other words, scaling is the process by which a measuring device is designed and scale values are assigned to different amount of the trait, attribute or characteristic being measured.

Credit goes to L.L. Thurston who developed methodologically sound scaling methods (1929,1932). His technique of scaling method is known as Equal- appearing interval.

However, Thurston and his students have developed a series of scales, each consisting of statements. These scales are developed to measure attitudes of individuals towards Negroes, Chinese, war, Censorship, the Bible, patriotism, and freedom of speech. An important scale of Thurston (1932) was developed to measure attitudes. This scale consisting of several statements or items which are followed by five responses along with scoring weights.

These five responses are:	Points
Strongly Agree (SA) having	(5)
Agree (A)	(4)
Undecided (U)	(3)
Disagree (D)	(2)
Strongly Disagree (SD)	(1)

This scale is also known as five-point-scale.

1.3.2 Types of Scales

It is generally agreed that there are four types of scales of measurement. They are : nominal scales; ordinal scales, interval scales and ratio scales.

1. Nominal scales

Nominal scales are the simplest forms of measurement. They are based on one or more distinct characteristics in which all things measured are classified or categorized and placed into mutually and exhaustive categories. For example, clinical psychologists often use nominal scale taking the help of (DSM-IV) to find out the nature, causes, symptoms and therapeutic method to know about mental disorder. This DSM-IV has assigned its own number for each disorder. For example, the number 303.00 identifies alcohol intoxication, and number 307.00 identifies stuttering. But these numbers are used exclusively for classification purpose.

The following statement can explain the test item on Nominal scale:
Instruction : Answer either Yes or No.

Are you actively contemplating suicide? -----

2. Ordinal scales

Ordinal scale is a system of measurement in which all things measured can be rank-ordered. It permits classification. In business and organizational settings, job applicants may be rank-ordered according to their desirability for a position. For individual subject ordinal form of measurement is also used. The Rokeach Value Survey which consists of a list of personal values, such as freedom, happiness, and wisdom is one of the best examples of ordinal sale (1973). A set of values ca be ranked in order which may assign a value of Vto the most important and '10' to the least important.

3. Interval scale

Interval scale is a system of measurement that contains equal intervals between numbers. Each unit on the scale is exactly equal to any other unit on the scale. Each unit in interval scales gets a meaningful result. Intelligence test is the best example of this type of scale. The IQs of 80 and 100, for example, is thought to be similar to that existing between IQs of 100 and 120. Like nominal and ordinal scales, interval scales also does not have zero ability or I.Q.

4. Ratio scales

A ratio scale has a true zero point. It is a system of measurement in which all things measured can be put in rank order and equal intervals exist between each number on the scale. There are few scales in psychology or education that can come closer to ratio scales. The best examples of ratio scales are a test of hand grip and a timed test of perceptual-motor ability (Cohen- Swerdlik, 2005)

In addition, we can also characterize scales in other ways. There are various other types of scales also which test takers use in different forms. When the test taker tests the performance of individuals as a function of age, it is referred to. as age-based scale. If he tests

performance as a function of grade, it is known as grade based scale. Similarly, when all raw scores of a test are transformed into scores, ranging from 1 to 9, is referred to as a stanine scale.

However, there are many different methods of scaling but there is no best type of scale. We shall discuss here some important scaling methods :

1.3.3 Method of Scaling

It is recorded in the history of test construction that L.L.Thurston (1929,1932) has developed a scaling method known as Equal appearing interval.

In this regard, another important scaling method has been developed by Katz et al (1999). It is known as ***Morally Debatable Behavioral Scale-Revised*** (MDBS-R). This scale assesses people's belief, the strength of their convictions and moral tolerance. It is 10 - point scale that ranges from justified to always justified. Here is a sample.

Cheating on taxes if you have a chance is

1	2	3	9	5	6	7	8	9	10
Never justified					Always justified				

The MDBS-R is an example of a rating scale which explains grouping of words, statements or symbols on which judgment about trait, attitude or opinion and emotion are indicated by test takers. This type of rating scale is used to obtain judgment of oneself, others, experiences, or objects. This scale consists of 30 items or statements. So, 30 scores indicate a low score and 300 indicates a high score (always justified). The final score is obtained by summing the rating scores of all items. This is termed as a summative scale.

Another important scaling method has been developed by Likert (1932). This scale is widely used to measure the attitudes of individual. Likert scales are very easy to construct. It is 5 - point scale because each item has five alternative responses, usually on an agree / disagree or approve / disapprove type of continuum.

Another scale method is *Guttman scale* (1999,1997), based on ordinal-level of measurement. This scale measures attitude, belief or feeling. This scale's items range from weaker to stronger expression. In other words, all respondents who agree with the stronger statements of the attitude will also agree with milder statements. However, this scale is not widely used, because other useful scales available measure various constructs.

Another scaling method is used by Katz et al, known as *paired comparisons*. In this method, the test taker selects one stimulus out of two (two photographs, two objects, two statements) according to his interest.

1.3.4 An Item on the Scale

Q. Select the behaviour that you think would be more justified.

- a. cheating on taxes if one has a chance.
- b. accepting a bribe in the course of one's duties.

In these items, option a is more justified on which most of judges reflect their agreement.

Another scaling methods are *comparative scaling and categorical scaling*. For comparative scaling, the judgments of a stimulus are compared with every other stimulus on the scale. For example, a test taker is given a list of 30 items on a sheet of paper and asked then to rank the items from 1 to 30, as most justifiable to least justifiable.

In categorical, scaling stimuli are placed into one of two or more alternative categories that differ quantitatively and then test takers may be asked to sort the cards into three piles, never justified, sometimes justified and always justified.

We have discussed above the nature, types and methods of scaling which are important components of test construction. No single scaling type or method is perfect. The selection of scaling or method depends upon the choice of test developer and the behavioral constructs being measured.

Writing Items.

The selection of scaling methods and actual writing of the test's items go together. While writing items test developer has to pay attention to:

- The Range of Content the test's items cover.
- The different types of formats a test developer should employ.
- The number of items the test should contain (written items in No.)

For construction of new test, the test developer has to write item according to the format of the test. To prepare the initial items pool, a test developer has to write a large number of items from personal experience or can obtain from other sources including experts. For writing items of psychological test, a test developer may conduct interviews in clinical setting, on clinicians, patients, parents, family members, and professionals who can assist test developer. It is useful to have at least 1200 sample items in an item pool. Item pool is a type of reservoir of items from which final version of items will be drawn or discarded. A comprehensive sampling gives a good basis for content validity of the final version of the test.

Item format

Variables such as the form, plan, structure, arrangement and layout of individual test items are collectively referred to as item format. We shall discuss here two important types of item format, namely the *selected response format* and the *constructed - response format*.

In a *selected response format*, there are a set of alternative responses. The test takers have to select only one response. But in a constructed - response format, the test takers are required to create the correct answer, not merely to select it.

However, there are three types of selected - response format available. They are multiple - choice, matching, and true/false format.

A good example of *multiple - choice item* in achievement test is -

1. Of the following, which is the best answer or measure of public interest in a particular election?

- A. The number of offices to be filled.
- B. The size: of the popular vote.
- C. The amount of campaigning preceding the election.
- D. The amount of money spent by the opposing parties for campaign purposes.
- E. The importance of the issues at stake.

Matching test:

Directions: After each name of topic which is intimately associated with that person.

	Column A	Column B
	Name of Persons	Topics
1.	Conditional Reflex	Titchener
2.	Age scale for testing intelligence.	Stanley Hall
3.	Reaction time experiments.	Pavlov
9.	Psychoanalysis.	Cattle, J. M.
5.	Psychology of Adolescence.	Sigmund Freud
6.	Existential Psychology.	Alfred Binet
7.	Factorial analysis.	

However, a multiple - choice item that contains only two possible responses is called a *binary-choice item*. The most familiar binary choice item is the true/false, agree/disagree, Yes/No, right/wrong or fact/opinion.

On the other hand, the constructed-response items are also of three types. They are the *completion item*, *the short answer*, and *the essay type*.

A completion item requires the test taker to provide a word, or phrase that completes a sentence. Following is the example of completion item.

The mean is the most stable and useful measure of the correct answer is central tendency.

If we write this sentence in short-answer item, we can write as: what descriptive statistics is generally considered the most useful measure of central tendency.

Whereas, essay item is concerned, it requires the test taker to respond to a question by writing a composition that is related to recall of facts, understanding, analysis or interpretation.

Here is an example of an essay item.

Distinguish between classical and operant conditioning in terms of definitions, principles and techniques.

An essay item is useful for test developer to demonstrate a depth of knowledge about a single topic. One can communicate his ideas in writing very well.

Writing items for computer administration.

Various computer programmes are designed to help the construction of tests and their administration, scoring and interpretation. These programmes have two main advantages

1. to store items in an " item bank "
2. test individual's ability through the technique called "item branching

An item bank collects a large number of questions which an instructor has to teach, sometimes useful for examination. These item bank of questions are compiled by subject areas, item statistics, or other variables. These items may be added to, withdrawn from, and even modified in an item bank.

However, computer-adaptive testing (CAT) is an important test device which can be administered on test taker's performance on previous items. The great advantage of CAT is that it records total number of items in terms of item pool for administering on test taker. CAT is very useful to reduce the number of test items by 50% and also reducing error of measurement by 50%.

The ability of the computer to present test items taken from an item bank based on previous responses of test taker is called item branching. Thus, CAT presents programme's items according to rule. For example, a test taker cannot take third items unless last two previous items are answered correctly. These items are presented on the basis of difficulty level.

Item branding technique is applied as construction of test of achievement and in test of personality. For example, if a person answers an item in such a way that seems for us that he/she is anxious about nothing, then computer may automatically provide the anxiety-related symptoms and behaviour.

Scoring Items

There are various scoring models available to score test items. Among them, *Cumulative Model* is commonly used for ease and simplicity. It is said that

the higher the score on the test, the higher is the ability of test taker on trait, attribute or other characteristics the test is applied to measure.

Another important model is *Class or Category scoring*. This test is used to assess diagnostic symptoms of individuals that are exhibited in specific diagnosis.

A third scoring model is *ipsative scoring*, used to compare test taker's score on one scale within a test with another scale within that test. For example, Edwards Personal Preference Schedule (EPPS), is designed to measure the relative strength of different psychological needs. The EPPS ipsative scoring system yields information on the strength of various needs in relation to the strength of other needs of the test taker. The test does not yield information on the strength of a test taker's need relative to the presumed strength of that need in the general population.

Edwards constructed his test of 210 pairs of statements in such a way that respondents were “forced” to answer *true* or *false* or *yes* or *no* to only one of two statements. A sample of an EPPS-like forced-choice item, to which the respondents would indicate which is “more true” of themselves is:

I feel depressed when I fail at something.

I feel nervous when giving a talk before a group.

On the basis of such an ‘I’ positively scored personality test, it would be possible to draw only intra-individual conclusions about the test taker

After making decision about scoring models, the first draft is ready for administration, the next step is test tryout.

Check your progress.

Q1. Define scaling and explain Nominal scales and Ordinal scales. Q2. Explain L.L. Thurston's scaling method in brief.

Q3. Write short note on Likert's scale. Q4. Define item pool and its format.

Q5. Explain any two types of selected- response format.

Q6. Explain in brief:

- a. Example of multiple-choice items
- b. Example of matching items
- c. Item bank
- d. Computerized Adaptive Testing (CAT)
- e. Scoring Items.

1.4 TEST TRYOUT

After writing a pool of items from the final version of the test, the test developer will try out the test on selected sample of people for whom the test is designed. The number of people to be taken in tryout is equally important. The rule of the thumb is that there should be minimum 5 or maximum 10 subjects for each item in tryout test.

Another point to keep in mind is that test must be administered under conditions as similar as possible to the conditions under which the standardized test will be administered. For example, instructions, time limit allotted for completing the test, atmosphere at the test site, etc.

What is a good item?

Just as a good test should be reliable and valid, so also, a good test item should be reliable and valid. A good test item helps to discriminate testtakers. Thus, a good item is the one that is answered correctly by the high scorers on the test as a whole. In academic context, an item that is answered incorrectly by high scorers on the test as a whole is not a good item. In other words, a good test item is one on which maximum scorers can do right on that item. However, different types of statistical techniques are used to analyze the test data, one such technique is item analysis.

Item analysis is the fourth step in test construction. It is an integral part of the reliability and validity. The quality and merit of a test depend upon the individual items of which it is composed. Thus, it is essential to analyze each item in the standardized manner in order to retain only those that suit the purpose and rationale of the device being measured.

However, item analysis is a general term which is related with various procedures designed to explore individual items of test that work as compared to other items of the whole test. Item analysis is also conducted to find out the level of difficulty of individual items on an achievement or another test.

There are many approaches to study item analysis. We shall discuss here the approaches particularly dependent upon statistical methods:

1.5.1 An index of the item's difficulty.

The difficulty of an item is determined by so many ways. The following three ways are worth mentioning:

- a. by the judgment of competent expert who, rank the items in order of difficulty;
- b. by how quickly the item can be solved, and
- c. by the number of people in the group who get the item right.

An index of an item's difficulty is obtained by calculating the proportion of the total number of test takers who answered the item correctly. The value of an item-difficulty index can theoretically range from 0 (if no one got the item right) to 1 (if everyone got the item right).

Thus, if everyone gets the item right, the item is too easy. If everyone gets the item wrong, the item is too difficult. There is no formula for determining the exact distribution of item difficulties. Thus, a common practice is to retain such items whose level of difficulty is 50 percent in terms of passing.

The item - Reliability Index

The item – reliability index is a statistical technique designed to provide an indication of a test's internal consistency. The higher the item - reliability index, the greater the test's internal consistency.

This index is equal to the item-score standard deviation (s) and the correlation (r) between the item score and the total test score.

Factor analysis and inter-item consistency

Factor analysis is a useful mathematical procedures used to reduce data. It is designed to find out variables on which people differ. There are two types of factor analysis. Exploratory factor analysis is a class of mathematical

procedure applied to estimate factors, extract factors, or decide how many factors should be retained for final revision of test.

Whereas confirmatory factor analysis (CFA) is also a mathematical procedure used when a factor structure is tested for its fit with the observed relationship between the variables. Factor analysis can be useful in the test interpretation process, especially when comparing the constellation of responses to the items from two or more groups

The item-validity Index

Item-validity Index is also a statistical technique which indicates the degree to which a test measures what it purports to measure. So higher the item-validity index, greater the test's criterion-related validity.

The item-validity Index can be calculated by either of the two following methods :

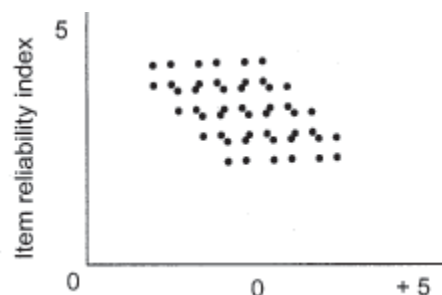
The item-score standard deviation. The correlation between the item-score and the criterion-score.

The item-score standard deviation of item 1 (denoted by s_i) can be calculated using the index of the item difficulty (P) in the following formula:

$$S_i = P_i(1 - P_i)$$

However, the type of validation employed in a test construction may be content, construct, predictive, or a combination of these. The techniques used in item analysis for external validation criteria are correlation coefficients, expectancy tables, and standard errors of measurement. The standardizations sample of the population includes age range, sex, socioeconomic distribution, range of ability or trait variation, educational level, type of school. For separate validity findings for different age groups, grade groups, ability groups, clinical groups, culture and subculture groups and occupational groups are also necessary.

In short, the best items on a test can be achieved by plotting each item's item-validity index and item-reliability index shown below:



The Item - Discrimination Index

It is also a statistical method which is designed to indicate how adequately a test item separates or discriminates between high and low scores. The Item

Discrimination Index is a measure of item discrimination. It is symbolized by letter d (d). Kelley (1939) has pointed out that marked and significant discrimination between extreme groups is obtained when item analysis is based upon the highest 27% and lowest 27% of the group. The higher the value of d , the greater the number of high scorers answering the item correctly.

We can use this method to find out what percentage of the highest percent and what percentage of the lowest 27% passed each item, then by statistical calculation, to determine if the difference between the two percentages is significant. There is another method which can be used by applying with reference to high average, low average, and low groups, classification being based upon total test score or upon external validation criteria.

However, successes and failures on each item may be correlated with the total scores for whole test. When this is done, biserial correlation can be calculated.

1.5.2 Analysis of item alternatives

There is no formula or statistics for item alternative. For this purpose, two groups are selected known as upper level (U) and lower level (L) of the distribution. We shall analyze the responses to five items, 1 and 2, a good item and poor item. Examples are given below:

Item 1	Alternatives				
	a	b	c	d	e
U	29	3	2	0	3
L	10	5	6	6	5

Response pattern to item 1 indicates that the item is a good one. More U group members than L group members answered item correctly.

Item 2	Alternatives				
	a	b	c	d	e
U	19	0	0	5	13
L	7	0	0	16	9

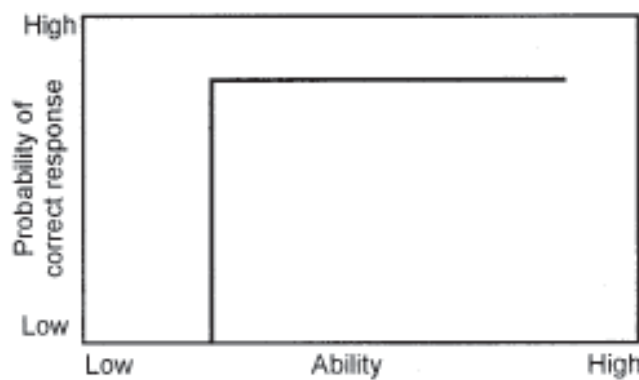
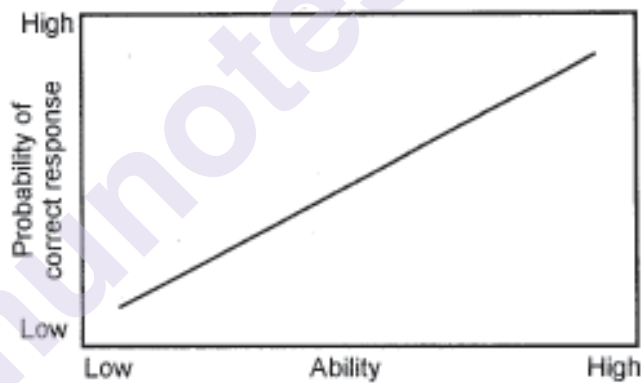
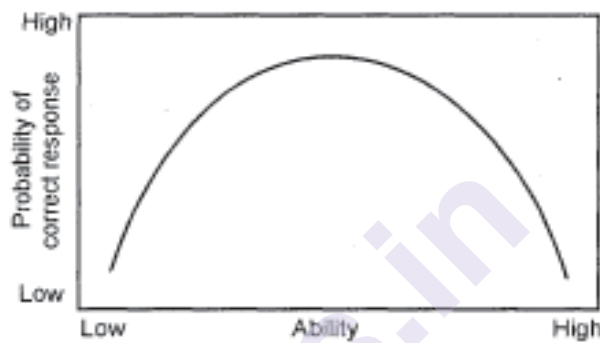
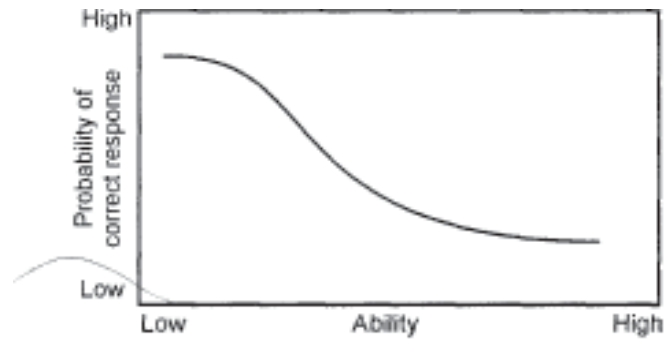
Item 2 is a poor item because more L group members than U group members answered the item correctly.

Item - Characteristics Curves (ICC)

ICC also represents item difficulty and discrimination.

ICC is a graph on which ability is plotted on the horizontal axis and probability of correct response is plotted on the vertical axis. The slope of the curve shows the item discriminating high- from- low scoring. The slope is seemed to be positive, more high scorers are getting the item correct than low scorers.

Item C is a good item. Item D has excellent discriminative ability and will be useful in a test designed to select individuals for cutting off scores.



Item response theory (IRT)

Item Response theory is also referred to as latent - theory or the latent trait model. It is a system of assumptions about measurement of a trait and the

extent to which each test item measures that trait. It is a true score model (Lord, 1980). It is widely used by commercial test developers and large-scale test publishers in test development.

An important model in this regard is offered by Rasch (2001) which explains that a person with x ability will be able to perform at a level of y. In other words, the probability of a person exhibiting a personality trait in x ability may exhibit the same trait of personality in y. According to Mitchell (1999), the Rasch model is more sophisticated model than classical test theory.

Latent - Trait Theory (LTT) is not widely used because of technical and complex issues. This theory provides us an estimate of the amount of knowledge, ability or strength of a test. Since this theory is unidimensional, the test measures all the traits.

Latent - Trait Model (LTM) can be found in the Illness Causality Scale (ICS), a measure of children's understanding of illness (Sayer et al. 1993). This scale reveals three important latent traits which are labeled as verbal intelligence, level of cognitive development, and understanding of illness. This scale has been correlated with other scales and the results are quite convincing.

However, this model is criticized on the ground that it is not widely used when others are available and it is more technically complex.

Despite these objections, LTM plays an increasing dominant role in the development of new tests and testing programmes. But large testing firms, state agencies and district's school rely on IRT method to construct, analyze, and score major achievement, entrance, and professional licensure examinations ((Rais and Henson, 2003).

Other Considerations in Item Analysis. Guessing:

Guessing an item correctly by chance. is not always true. The guessing of a person on true/false choice item alone may be 5 out of ten items. The probability of guessing correctly on two such items is equal; to .52 or .25%. The probability of guessing correctly on ten such items is equal; to .510 /or .001. Therefore, one in thousand items can be found correct on ten true/false on the basis of chance alone.

However, three criteria have been published with regard to the problem of guessing answer correctly.

1. A respondent while guessing a correct answer does not make on random basis, rather he applies his knowledge of the subject matter and ability to rule out unfit alternative.
2. A correction for guessing also depends on the problem of omitted items. This problem is related with questions : Should the omitted item be scored " wrong "? or such item be handled in different ways.

3. The rule of guessing also depends upon "Luck by chance," but any correction for guessing may be underestimated or overestimated the effect for lucky and unlucky test users.

However, no criterion is found to be satisfactory for guessing items correctly but two reasons are seemed to be very reasonable

- a. Clear instructions given by examiner to examinees and
- b. Specific instructions for scoring and interpreting omitted items.

Thus, guessing responses correctly is not a complex problem, it depends on risk - taking behavior of test taker.

Items fairness

An item is considered fair when the different group members pass on any given item that measures that ability, regardless of race, social class, sex or any other background characteristics. In other words, the same proportion of persons from each group should pass any given item of the test, provided that persons earn the same total score on the test (Jenson, 1980, P.44).

Speed tests

Speed tests are tests in which the time limit imposed is so short that all examinees cannot attempt all of the items in test. They are of low difficulty level. The reliability of speed test can be calculated by split-half-technique or the basis of odd-even split. The reliability coefficient of correlation may be closed to 1.00. parallel forms or test-retest are also methods to estimate reliability of speed tests.

For a good speed test, all items should be of uniform, or nearly uniform, degree of difficulty. The best practice is to correlate subtest scores and total

- Test scores with scores of criteria, under various time limits.

Qualitative Item Analysis (QIA)

QIA is general term for various non-statistical procedures designed to explore how individual test items work. The analysis of this procedure is compared with individual test items to each other and to the test as a whole.

There are some important topics which a researcher can explore for qualitative analysis. They are: cultural sensitivity, face validity, test administration, test fairness, test language, test length, test taker's preparation and so on.

Qualitative methods are techniques of data analysis through verbal means such as interview and group discussion. It is better to provide an opportunity to test takers and students to describe their instructors. If students fail to respond adequately for test items, they may be given chance to evaluate their performance. The other related aspect with qualitative item analysis is Think Aloud Test Administration.

Think Aloud Administration

Different researchers use different procedures to respondents to verbalize thoughts as they occur (Davison, 1997; Hurlburt, 1997; Klinger 1978). This approach is employed by them for adjustment, problem solving, educational remediation and clinical intervention.

Cohen et al. (1988) have pointed out that "think aloud" test administration is a tool of QIA focuses on the thought process of test taker during the administration of a test.

For achievement test, verbalizations may be useful in assessing low or high scores and also why and how they are misinterpreting the items. And for personality test "Think aloud" device may provide insight to perceive, interpret, and respond to the items.

Expert panels

Expert panels also provide qualitative analysis of test items. These panels try to obtain an understanding of the history and philosophy of the test battery and to discuss and define the problem of bias (Stanford Special Report, 1992). Some of the content bias that have been identified are given as under:

- a. Status - the situations that do not involve authority.
- b. Stereotype - members of particular group show aptitude, interest, occupation and personality characteristics.
- c. Familiarity - groups know vocabulary and experiences about items.
- d. Offensive Choice of words - using of correct wording for items.
- e. Other - Panel members should be asked another indication of bias they detect.

On the basis of qualitative information from an expert panel, test users or developers may elect to modify or revise the test. However, rewording items, deleting items or creating new items is known as test revision, a final stage in the development of a new test. The process of revision is very expensive work. It requires a lot of efforts, time and expense.

Check your progress:

Q1. What are the tools a test developer employs to analyze and select items ?

Q2. Explain in brief:

I. The item - difficulty Index

II. The item- reliability Index

III. The item- Discrimination Index

Q3. Explain any two theories of item response theory.

Q4. Define: -

a) Item fairness

b) speed test

c) Qualitative method

Q5. Write brief note on expert panels and Think Aloud Test Administration.

1.6 TEST REVISION

Test revision is the last (fifth) stage of test construction. This is final form of a new test in which some items are eliminated and others will be rewritten.

1.6.1. Test Revision as a stage in New Test Development.

On the basis of information obtained from the item - difficulty, item analysis, item reliability, validity, discrimination and bias of test items, a test developer tries his best to revise the test if he feels adequate and necessary.

There are many ways or approaches which help developers in revision of final test newly constructed. One such approach is to characterize each item according to its strength and weakness. Some items may be very easy or hard, highly reliable but lack criterion validity, some items may lack reliability and validity because of their restricted range. The same may be true of very easy items.

So if the test developer may find that he can't maintain balance between strength and weakness across items, then good items need to include more difficult items. Thus, the purpose of test revision will be affected. Furthermore, if highly skilled individuals are being tested, then the best possible test discrimination will be made a priority.

As revision proceeds, writing large items pool becomes clear. Thus, poor items can be eliminated and good items should be retained. By doing so, a test developer will come out of the revision stage with a better test. Now, the next step is to administer the revised test under standardized conditions to obtain a second appropriate sample of examinees. After administration of second draft of the test, the test developer may give a final touch. Once the test is in finished form, the test's norm may be developed from the data. Now the test will be said to have been "standardized".

1.6.2 Standardization

Standardization refers to "the process employed to introduce objectivity and uniformity into test administration, scoring, and interpretation". (Robertson, 1990). A standardizations sample represents the group (s) individual with whom examinee's performance will be compared.

This sample must be representative of the population on those variables that may affect the performance. For example, in a ability test the standardizations group must represent the norms such as age, gender, geographical region, type -of community, ethnic group and educational level.

However, process of revision continues until the test is satisfactory and standardizations can occur. So once the test is ready, its validity requires for a cross-validation of findings. Before we discuss cross-validation, it is reasonable to briefly consider some issues related to the development of a new edition of an existing test.

Test Revision in the Life cycle of an existing Test.

There is no hard-and-fast rule that exists for revision of a test (APA, 1996. 3.18). It give us suggestion that a test can be kept in its present form as long

as it remains "useful" and be revised when significant changes make the test useless for application.

However, a test can be revised if it meets any of the following conditions:

- When the current test takers can't correlate stimulus materials.
- When current test takers are not able to understand the verbal content of the test, including test administration, instructions and test items containing dated vocabulary.
- When a popular culture changes and the words take the new meanings and certain test items or directions seem to be inappropriate, thus the test needs to be revised.
- As the group membership changes, the test norms are also seemed to be inadequate.
- Through revision, the reliability and validity can be improved significantly.
- As time passes, an age extension of norms may take a shift upward, downward, or in both directions, a change in test is necessary.
- The original theory on which test is based also need revision that reflect the design and content of the test.

As we have noted that revision takes place in all stages of test construction starting from conceptualization phase, test tryout, item analysis and test revision. There are variety of tests and developed scales which has under gone revisions. For example, Strong Vocational Interest Bank, MMPI, Binet Tests of Intelligence, etc.

However, a key step in the development of all tests is crossed validation and co-validation.

Cross-validation and Co-validation.

The term cross-validation refers to the revalidation of a test on a sample of test takers other than those on whom test performance has originally been found to be a valid predictor of some criterion. For example, a test's items are selected for a newly constructed test may be administered on first and second groups of sample. Suppose, an Indian test developer constructs a test in Indian 'population, he can find out its reliability and validity very easily. But when the same test is applied on foreign sample and obtained validity, this is called cross-validation.

However, the decrease in item validities that occurs by chance after cross-validation of findings is referred to as validity shrinkage.

Co-validation is defined as a test validation process conducted on two or more tests using the same sample of subjects. However, a test is used to

revise existing norms, it is referred to as co-norming. Co-validation is useful for:

- Publishers because of its economical purpose;
- Collecting data by means of face-to-face or telephone interview; and multiple testing

Co-validation requires qualified examiners to administer the test, to assist in scoring, interpretation and statistical analysis. There are some tests which are used by publishers and test users. For example, Wechsler Adult Intelligence Scale (WAIS-111) and Wechsler Memory Scale (WMS-111) are used together in the clinical evaluation of an Adult. Since the two tests are normed on the same population, sampling error is greatly minimized, if not eliminated completely.

Quality assurance during test revision.

There is no mechanism of quality assurance that a test publisher can adopt in course of standardising a new test or restandardising an existing test. But for quality control, it is essential to- recruit examiners with extensive experience of testing the children and adults. They must be well off about their educational and professional qualifications, administration experience with various intellectual measures, certification, and licensing status. The selected examiners must be very familiar with childhood assessment practices (Wechsler, 2003).

In nut-shell, every examiner must be having a doctoral degree. Regardless of educational qualifications or experience, all examiners must be trained enough to handle the process of test construction and statistical procedures of test administration, scoring, interpretation, and process of revision. They may be involved to greater degrees in the final scoring of protocol.

For quality assurance of WISC-IV two trained and qualified scorers were appointed during national tryout and standardisation stage of WISC-IV test development.

Another mechanism for ensuring quality assurance is an anchor protocol. It is test protocol scored by a highly authoritative scorer, who resolves scoring discrepancies, if any existing.

However, when discrepancy exists between scoring in an anchor protocol and for another protocol is referred to as scoring drift. Anchor protocols were used for quality assurance in the developments of the WISC-IV. For quality assurance we have computer programmes to find out and identify any error in score reporting.

Check your progress

Q1. Explain any one approach of test revision.

Q2. Explain the nature and uses of standard isation.

Q3 What are the conditions under which a test can be revised ?

Q4 Define the following terms

- a. Cross-validation.
- b. Co-validation.
- c. An anchor protocol.
- d. Scoring drift.

1.7 SUMMARY

In this chapter we have highlighted on creation of a good item, the basics of test development and the. process by which tests are constructed. We have also discussed a number of techniques designed for construction and selection of good items. We have also focused on five stages of the process of development. These stages are test conceptualization, test construction, test tryout, item analysis and test revision.

While discussing the process of test conceptualization, we have explained some preliminary questions confronted by test developers. Under this topic we have also pointed out item development issues, concept of pilot work or pilot study.

The second stage, test construction, is also dealt with various important instruments relating to test construct. These instruments are scaling, methods of scaling, types of scaling, writing items, items formats, writing items for computer administration and different test scoring models.

In test tryout stage we have focused on the nature of a good test, followed by item analysis.

In fourth stage we have revealed the nature and uses of various test tools which a test developer has to adopt. These tools are an index of item's difficulty, item's reliability, item's validity and item discrimination. To provide a detailed explanation of all these indexes, we have discussed item response theories too, including speed test, test administration and appointment of expert panels.

And the final stage is discussed through highlighting on various ways of revision of new test development, standardizations of test, test revision in the life cycle of an existing test, the nature and uses of cross-validation and co-validation, quality assurance during test revision and so on.

1.8 QUESTIONS

Q1. Explain any two stages of Test development.

Q2. Define scaling and describe the various types of scaling method.

Q3. Explain in brief

- a. Multiple-choice format and matching test
- b. Speed test and expert panel
- c. Test revision
- d. Likert scale
- e. Thurstone scale
- f. Tools of Test development

Q4. Define the following terms:

- i. CAT
 - ii. Rasch model
 - iii. Think aloud test administration
 - iv. Item Pool
 - v. Item branching
 - vi. Anchor protocol
 - vii. Item - Discrimination Index
 - viii. Item - difficulty Index
 - ix. Cross –validation
-
-
-
-
-

1.9 REFERENCES

Anastasi, L. R. (1937) - A Hand Book of Psychological Testing (7th edi)
Indian Reprint, 2002

Campbell, D.P. (1972) - The practical problems of revising on established
psychological test. In J.N. Butcher (Ed) Objective Personality Assessment
: Changing Perspective (pp. 117-130) Newyork Academic Press.

Freeman, F. S. (1962) - A Hand Book on Theory and Practice of
Psychological Testing (6th Ed), Oxford and IBH Publishing Co. Bombay.

Cohen, J.R. and-Swerdlik, M.E. (2010) - Psychological Testing and
Assessment ; An Introduction to Test and Measurement. (7th ed), Newyork,
McGraw- Hill international edition.

Guttman, L. A. (1999) A Basis for Scaling Qualitative Data. American
Sociological Review, 9,179-190

Guttman, L. A. (1997) - The Cornell Technique for Scale and Intensity
Analysis. Education and Psychological Measurement, 7, 297-280.

- Jensen, A. R. (1980) Bias in Mental Testing, Newyork Free press.
- Katz, R. C. and Lonero, P. (1999) Findings on the Revised Morally Debatable Behavioral scale. Journal of Psychol. 128,15-21.
- Likert, R. (1932) A Technique for Measurement of Attitude. Archives of Psychol. Number 190.
- Mitchell, J. (1999) Measurement in Psychol : Critical History of a Methodological concept. Newyork : Cambridge university press.
- Rasch, G. (2001) Applying Fundamental Measurement in Human Science chapter - 2 Mahwah, N.J. Erlbaum.
- Reise, S. P. and Henson, J. M. (2003) A Discussion of Modern versus Traditional Psychometrics as Applied to Personality Assessment Scales Journal of Personality Assessment, 81. 93-103.
- Robertson, G. J. (1990) A practical model for test development Hand Book psychological and educational Assessment of children. pp, 62-85, New York Guilford.
- Rokeach,, M. (1973) The nature of human values. Newyork. Free Press.
- Sayer, A. G. and Perrin, E.C. (1993) Measuring understanding of illness causality in healthy children and in children with chronic illness. A construct validation. Journal of applied developmental Psychol. 19, 11-36.
- Thurstone, L. L. (1929) Theory of attitude measurement, Psychological Bulletin, 36, 222-291.
- Thurstone, L. L. (1932) - Multiple Factor analysis. Chicago : university of Chicago Press Wechsler, D. (2003)
- WISC - iv, Technical and interpretive manual, 9th (Ed) San Antonio, Tx. Psychological corporation.

TEST DEVELOPMENT AND CORRELATION - II

Unit Structure :

- 2.0 Objectives
- 2.1 Introduction
- 2.2 Meaning and types of correlation.
- 2.3 Graphic representation of correlation - Scatterplots.
- 2.4 The steps involved in calculation of Pearson's product moment correlation coefficient.
- 2.5 Calculation of rho by Spearman's rank-difference method.
- 2.6 Uses and Limitations of correlation coefficient.
- 2.7 Simple regression and Multiple regression.
- 2.8 Summary
- 2.9 Questions
- 2.10 References

2.0 OBJECTIVES

- To impart knowledge and understanding of the meaning, types and methods of Calculation of Correlation.
- To create awareness about the various steps involved in calculation of Pearson's product-moment coefficient of correlation.
- To provide basic knowledge about calculation of rho by Spearman's rank-order method.
- To make the foundation of statistical techniques strong about the knowledge of correlation coefficient and its applications.

2.1 INTRODUCTION

The measures of variability or dispersion are defined on univariable i.e., the observations based on single characteristic. We will study the concept of variability or dispersion along with its various measures in detail in Unit 7. In some situations, we may observe two or more characteristics simultaneously for each unit in a population - for example - height and weight of an individual, income and expenditure, supply and demand of a commodity, imports and exports of a country, etc. The statistical procedure that measures the relationship between two or more variables is called correlation. In this chapter we will study the meaning, and types of

correlation. We shall also study Graphic representation of correlation, specifically Scatterplots. Attention is also given to calculation, correlation by Pearson's product- moment correlation coefficient and rho by Spearman's rank difference method.

Uses and limitations of coefficient of correlation will be also studied. Toward the end of the unit we shall discuss the nature and uses of simple regression and multiple Regression.

2.2 MEANING AND TYPES OF CORRELATION

Correlation is defined as an expression of the degree and direction of relationship between two things or two sets of variables, when each of them is continuous in nature, Thus, the relationship between two variables is called simple or bivariate correlation. For example, height and weight, supply and demand of a commodity.

Types of correlation

There are three types of correlation- these are discussed below.

Positive Correlation: -

When the values of two variables move in the same direction, so that an increase in the value of one variable tends to increase the value with other variable, the correlation is called positive. Similarly, if two variables simultaneously decrease then, two variables are also said to be positively correlated. In our observation such as height and weight, profit and investment, income and expenditure of a family, etc., are positively correlated.

In Figure 2.1, direction of plotted points are from lower left corner to upper right corner. This is a positive correlation. The slope of the line is also positive.

Figure 2.1 Positive Correlation

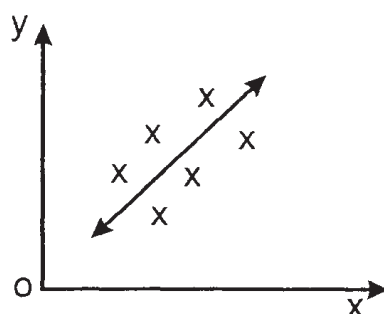


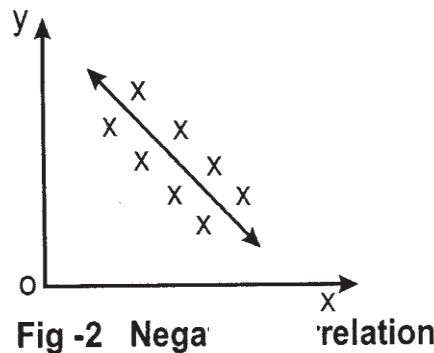
Fig -1 Positive correlation

Negative Correlation:

A negative correlation occurs when one variable increases and other variable decreases. For example, when supply increases and demand decreases, or price of the commodity increases but consumption decreases.

In Figure 2.2, Most of points lie near the line or on the line. Slope of the line is also negative.

Figure 2.2 Negative Correlation

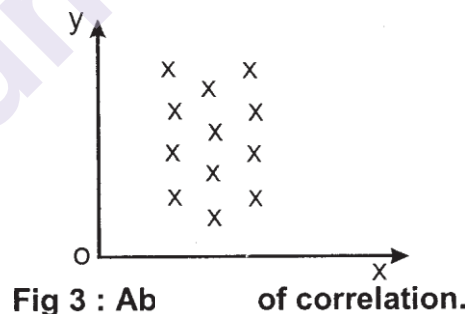


Zero correlation:-

When there is absolutely no relationship existing between the two variables, the correlation is said to be zero (0). In any case, perfect positive correlation (+1), perfect negative correlation (-1) and zero correlation are hard to identify. Most of the time, two variables are fractionally correlated.

Figure 2.3 shows a zero correlation. It shows no direction. Therefore, there is no correlation between the values of two variables. No line can be drawn passing through most of the points.

Figure 2.3 Absence of Correlation/ Zero Correlation



2.3 GRAPHIC REPRESENTATION OF CORRELATION - SCATTER PLOTS

An important type of representing correlation is a graphic description known as the Scatterplot or Scatter diagram. It is very simple method to study correlation by the use of n - pairs of observations (x_i, y_i) , (X_2, Y_2) and so on. Values of two variables of each observation are taken as co-ordinates of the point, where the values of one variable are placed on X-axis (horizontal line) and the values of another variable are placed on y axis (vertical line). Paired observations are plotted as points on the graph. The graph of these points shows how far the observations are scattered, Hence,

it is a scattered plot or scatter diagram. A scatterplot of the data helps us in having a visual idea about the nature of association between two variables. The relationship shown by the points plotted on the scatterplot involves two aspects, first the direction of the relationship i.e., positive or negative and the closeness of the points to some line.

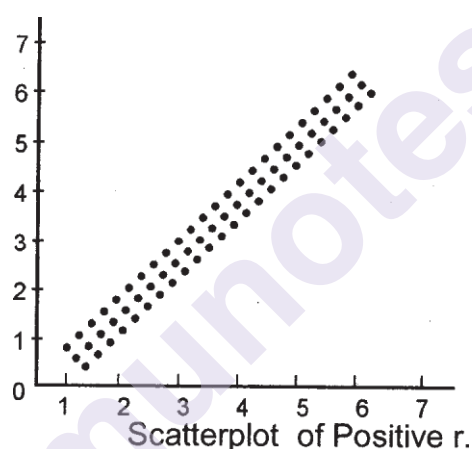
However, scatterplot reveals the direction and strength of magnitude of the relationship between two variables.

Scatterplot or graph also provides data to know the range of scores and types of relationship exist between two variables, scores, group of scores, etc. It is relatively very simple technique that provides a hint of some of deficiency in the testing or scoring procedures. Scatterplots are useful in revealing the presence of *curvilinearity* in a relationship

Graphic representation of Scatterplot.

The following Figure 2.4 shows the Scatterplot for positive correlation (r) correlation coefficient = 0.60 (Moderate degree of positive correlation) correlation coefficient = 0.95 (very high degree of positive correlation).

Figure 2.4 Scatterplot for Positive Correlation



However, the Figures 2.5 and 2.6 shows the negative correlation in the forms of graph; whereas Figure 2.7 shows a zero correlation, that is the absence of any kind of correlation.

Figure 2.5 Correlation coefficient (r) = -0.50 (moderate degree of a negative correlation)

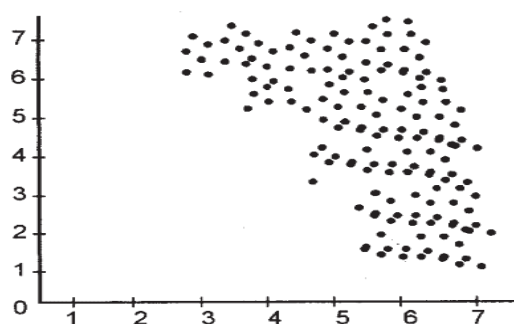
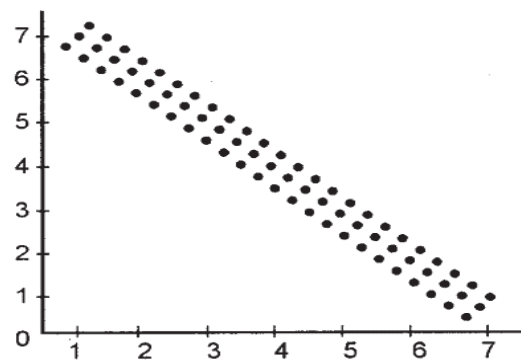
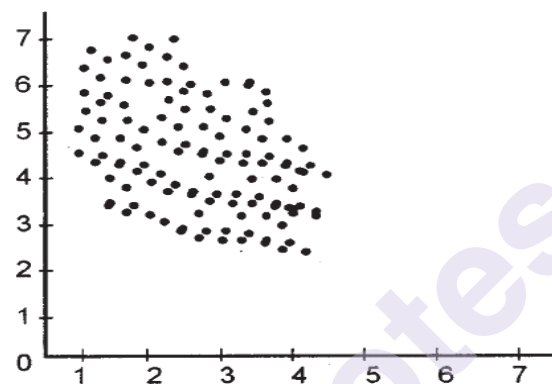


Figure 2.6 Correlation coefficient (r) = - 0.90 (very high degree of negative r)



Scatterplots and Correlations for Negative values of r

Figure 2.7 Correlation coefficient (r) = 0.00 (zero correlation)



Scatterplot and Correlations for 0 values of r

2.4 THE STEPS INVOLVED IN THE CALCULATION OF PEARSON'S PRODUCT- MOMENT CORRELATION COEFFICIENT

There are many techniques which have been developed to measure correlation. The most widely used technique is of Karl Pearson's product - moment coefficient correlation. Pearson's technique is the standard index of the amount of correlation between two variables. When the relationship between the variables is linear and when the two variables being correlated are continuous, Pearson's r is used.

There are a number of Pearson's r formula. We shall study the following short-cut and modified formula for our purpose step by. step.

$$\frac{\sum xy}{(\sum x^2) (\sum y^2)}$$

When we see this formula, it seems to be more complex than other formulas. But it is easier to use when deviations are taken from the means of the two distributions.

To find r ;

1. The first step to find out of number of paired scores.
2. Σxy is the sum of the product the paired X and Y scores.
3. ΣX is the sum of the x scores and Σy is the sum of the Y scores.
4. (ΣX^2) is the sum of the squared X scores and (Σy^2) is the sum of the squared Y scores.

However, similar results are obtained with the use of each formula. The sign of the resulting r would be a function of the sign and the magnitude of the standard scores used.

Check your progress.

Q1. Define correlation and explain its types with examples.

Q2. Define/Explain scatterplot and explain the graphic representation of its calculation.

Q3. What is product-moment correlation coefficient? Explain the various steps involved in its calculation.

2.5 CALCULATION OF RHO BY SPEARMAN'S RANK-DIFFERENCE METHOD

When complex behaviour such as honesty, athletic ability, social adjustment is hard to measure, we should put these behaviours in order of merit. In computing the correlation between two sets of ranks, special methods are used.

Charles Spearman (1927) has developed a method known as Spearman's rho rank-difference method. It is conveniently applied as a quick substitute when the number of pairs is less than 30.

It is also conveniently used when data are already in terms of rank-order. Therefore, it is also called rank-order correlation coefficient, a rank-difference correlation coefficient, or simply Spearman's rho. However, Rho method has only one formula, but it has different names such as rank-order correlation of co-efficiency, a rank-difference correlation coefficient, or simply Spearman rho. It has both sets which are in ordinal form, (rank-order) However, the formula of Rank order is:

$$1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

or

$$1 - \frac{6 \sum d^2}{N^3 - N}$$

Example Calculate rank correlation coefficient for the following data given marks in two tests in Psychology, for a group of 10 students.

X	Y	R ₁	R ₂	d	d ²
67	78	2	2	0	0
42	80	8	1	7	49
53	77	7	3	4	16
66	73	3	6	-3	9
62	75	4	4	0	0
60	68	5	7	-2	4
54	63	6	8	-2	4
68	74	1	5	-4	16

$$\sum d^2 = 98$$

$$1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$r = 1 - \frac{6 \times 98}{8 \times (8^2 - 1)}$$

$$r = 1 - \frac{588}{504}$$

$$r = -0.166$$

This is poor negative correlation.

Example 2.2

X	Y	R ₁	R ₂	d	d ²
50	48	5.5	5	0.5	0.25
63	30	3	8.5	5.5	30.25
48	35	9	7	2	4
70	60	1	1	0	0
45	55	8	2	6	36
65	30	2	8.5	-6.5	42.25
38	25	10	10	0	0
40	45	7	6	1	1
52	50	4	4	0	0
50	52	5.5	3	2.5	6.25

$$\sum d^2 = 120$$

$$r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$r = 1 - \frac{6 \times 120}{10(10^2 - 1)}$$

$$r = 1 - \frac{720}{990}$$

$$r = 1 - 0.73$$

$$r = 0.27$$

Simple low positive correlation.

2.6 USES AND LIMITATIONS OF CORRELATION COEFFICIENT

A correlation coefficient is the numerical index that expresses the relationship between two sets or variable which is indicated by the letter small Y. This coefficient may be of any size from zero to +1.00 or - 1.00. The sign of the coefficient does not determine its significance, there are high, moderate, or low coefficients that may be either positive or negative. Thus + 1.00 indicates a perfect positive and -1.00 a perfect negative correlation.

Uses

1. The techniques of correlation of coefficient are used in physical sciences and social sciences, especially psychological research where we postulate doctrine and theories with principles and hypotheses.
2. This technique is also applied in calculation of reliability and validity of coefficient of correlation.

3. It is also used to find out simple r , multiple r and rank-order correlation.
4. Through scatterplot we can identify outliers.
5. Correlation coefficient indicates not only the direction of relationship but also the magnitude of the relationship between two variables.

Limitations of correlation coefficient

Since there are several techniques of calculating r , so we can't apply all of these to solve one type of problem. For example, Pearson's r is applied on group data and ungroup data, but it does not apply on ordinal form a Rank-order method is used. So, both techniques have their own limitations and advantages.

Similarly, by any means, a perfect positive correlation coefficient (+1.00) or a perfect negative coefficient (-1.00) is impossible to obtain. It is challenging to try to think of.

It is also very difficult to use all techniques of Pearson r because they are very complicated and time consuming except one.

2.7 SIMPLE REGRESSION AND MULTIPLE REGRESSION

Regression is commonly explained as retreat back or reversion to some previous state. In statistics, regression also describes a kind of reversion- to the mean overtime or generation. The term "regression" was first used by Francis Galton with reference to the inheritance of status. Galton found that children of tall parents tend to be less tall, and children of short parents less short, than their parents. In other words, the heights of the offspring tend to "move back" toward the mean height of the general population. This tendency toward maintaining the mean height is called the principle of regression and the line describing the relationship of height, in parent and offspring is called a "regression line". The term is still employed but in other meaning. Today, regression is defined as the relationship among variables for the purpose of understanding how far one variable predicts other one. In other words, regression is the measure of the average relationship between two or more variables in terms of original units of data.

Although, there are various types of regression, but we shall explain simple and multiple regression.

Simple regression involves the analysis of only two variables; one is independent variable (X), typically known as "the predictor" variable and another is dependent variable (Y), typically referred to as to the "outcome variable." Simple regression results in an equation for regression line - that line of best fit, the strait line that comes to the closest to the greatest number of points on the scatterplot of X and Y.

However, the main use of regression equation is to predict the effect of one value on other and to make interpretation of Y.

Multiple regression is another type of regression. It is also used as a predictor. Its analysis requires the use of more than one variable. To predict Y requires the use of a multiple regression equation. This type of equation explains the interrelations among all the variables involved. If many predictors are used, and one is not correlated with any predictor, but is correlated with the predicted score, it gives more weight and provides unique information.

Method of Regression Analysis

We have already seen scatter plot that represents the correlation between two variables for a bivariate data. In this method the points are plotted on a graph paper representing pairs of values of the concerned variables where values of independent variables are taken on X -axis and the values of dependent variables are taken on Y-axis. A regression line may be drawn in between these points by free hand or by a scale rule in such a way that squares of the vertical distances or the horizontal distances between the points and the line of regression so drawn is the least. It should be drawn carefully as the line of best fit leaving equal number of the points on either side of the line.

Example 2.3

X	2	3	4	6	7	8	10
Y	5	6	10	11	15	17	21

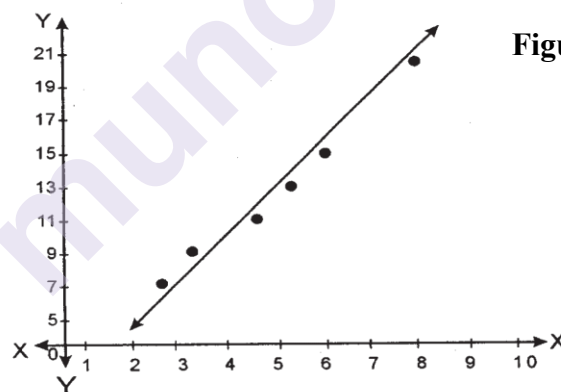


Figure 2.8 Regression Line

This method provides a rough estimate of the dependent variable because the line drawn is subjective to the person drawing it.

Check your progress

Q1 When rho by Spearman's rank-difference method is used?

Q2 Find the Spearman's rank correlation coefficient from the following

X	50	70	65	38	90	50	40	75
Y	35	90	70	48	95	85	40	80

X	15	12	16	25	15	17	14	18	26	20
Y	17	14	16	16	25	20	24	22	26	23

Q3 Explain the uses and limitations of correlation coefficient.

Q4 Write short notes on simple regression and multiple regression.

2.8 SUMMARY

In this chapter we have discussed the meaning and types of correlation with special attention given on positive, negative and zero correlation with examples. Correlation is also graphically represented in the form of scatterplot. We have also studied the nature of Pearson's product- moment correlation coefficient and its various steps involved in its calculation of Y.

Calculation of rho by Spearman's rank-difference method is also highlighted with questions and answers of the problems.

Finally, we have focused on the uses, limitations of correlation coefficient and its two types - simple regression and multiple regression followed by method of regression analysis shown by a graph.

2.9 QUESTIONS

Q1. a) Define correlation and explain its types.

b) Calculate the rank - order correlation coefficient from the following distribution.

X	22	35	70	80	70	65	50	55	40	50
Y	78	68	60	65	60	55	45	52	75	76

c) Interpret your answer.

Q2. Explain the uses and limitations of correlation coefficient.

Q3. Explain the various steps involved in calculation of Pearson's product-moment correlation coefficient.

Q4. Write short notes on Scatterplot, simple regression and multiple regression.

2.10 REFERENCES

Annapornaa R. et al (2008) A hand book of Mathematics and Statistics, Chetana Publications Pvt. LTD. 263, Khatauwadi, Girgaon, Mumbai-400004

Anastasi, A. and Urbina, S. (1997) Psychological testing (7th ed) Pearson Education, India Reprint -2002

Cohen, J.R. and Swerdik, m.E. (2010) Psychological testing and Assessments, An introduction to tests and measurements (7th ed) Newyork McGraw-Hill international edition.

Garrett, Henry, E. (1973) Statistics in psychology and Education (6th edition) Vakills, Feffer and Simon Pvt. Ltd. Ballard Estate, Mumbai 400001

Guilford, J.P. (1956) Fundamental Statistics in Psychology and Education (3rd ed) Newyork McGraw-Hill book, Co.

Spearman, C. (1927) The ability of man: their nature and measurement, Newyork, MaCmillan.

Walker, H. M. (1943) Elementary statistical method (Newyork) Henry Stoll and C. pp. 308-310

munotes.in

MEASUREMENT OF INTELLIGENCE, INTELLIGENCE SCALES, PROBABILITY, NORMAL PROBABILITY CURVE AND STANDARD SCORES - I

Unit Structure :

- 3.0 Objective
- 3.1 Introduction
- 3.2 What is Intelligence? - Definitions and Theories
 - 3.2.1 Definition of Intelligence
 - 3.2.2 Theories of Intelligence
- 3.3 Measuring Intelligence
- 3.4 The Stanford-Binet Intelligence Scales
- 3.5 The Wechsler Tests
 - 3.5.1 Wechsler Intelligence Scale for Children - Fourth Edition (WISC -IV)
 - 3.5.2 Wechsler Preschool and Primary Scale of Intelligence III edition (abbreviated as WPPSI-III)
- 3.6 Summary
- 3.7 Questions
- 3.8 References

3.0 OBJECTIVES

After studying this unit, you should be able to:

- Understand the various definitions of intelligence given by the layman as well as scholars and test professionals.
- Comprehend the various theories of intelligence.
- Know the process of measurement of intelligence and types of tasks involved in intelligence tests as well as theory in intelligence test development and interpretation.
- Understand the Stanford-Binet Intelligence scales as well as Wechsler tests.

3.1 INTRODUCTION

In this unit we will discuss about the definition of intelligence and the various theories of intelligence. Among the definitions of intelligence, we would examine the views of the layman as well as the views of scholars and test professionals. A few theories of intelligence would also be examined. The most important theories of intelligence include the Factor-analytic theories and the information processing views. Following this we would study about measuring intelligence and issues related to it.

The Stanford- Binet Tests and the Wechsler tests are an important group of tests. Among the Stanford Binet Tests we would discuss the 5th edition of Stanford-Binet Intelligence Scales in brief. We would also discuss few Wechsler tests and briefly discuss the short forms of few of these tests. Towards the end of the unit we provide a summary of the unit, followed by questions and a list of references for further study.

3.2 WHAT IS INTELLIGENCE? - DEFINITIONS AND THEORIES

Intelligence is a multifaceted capacity that is expressed in many ways. Intelligence means many things to different people including layman, scholars and psychological test professionals. The following various abilities are said to constitute intelligence.

- Ability to acquire and apply a specific knowledge.
- Ability to use the right words and to generate quick thoughts in given event.
- Ability to reason well, judge well and to be self-critical
- Ability to plan and foresee things and events
- Ability to grasp, visualize concepts.
- Ability to make good judgments to solve problems efficiently and economically
- Ability to infer perceptively
- Ability to comprehend people, events and situations.
- Ability to pay attention to minute details and to be intuitive and innovative
- Ability to cope, adjust, adapt to new situations and culture.
- Ability to be practical and street smart and get one's work done.

3.2.1 Definition of Intelligence

Intelligence Defined: Views of the Layman

Sternberg and his associates have done considerable work with respect to how lay people conceptualize intelligence. According to Sternberg, non-psychologists conceptualize intelligence as follows:

- Reasons logically and well
- Reads widely
- Displays Commonsense
- Keeps an open mind and
- Reads with high comprehension

Non-psychologists view about unintelligence is reflected in the following statements:

- Does not tolerate diversity of views
- Does not display curiosity and
- Behaves with insufficient consideration of others.

According to Sternberg, non-psychologists and experts consider intelligence, in general, as follows:

- a. Practical problem solving ability
- b. Verbal ability and
- c. Social competence

Sternberg found that there was considerable degree of similarity between the experts and lay person's conceptions of intelligence. The following table lists as to what both meant by academic intelligence and everyday intelligence.

Academic Intelligence	Everyday Intelligence
<ul style="list-style-type: none">• Verbal Ability• Problem Solving Ability• Social Competence• Specific behaviours associated with acquiring academic skills, such as studying hard.	<ul style="list-style-type: none">• Practical problem solving ability• Social Competence• Character• Interest in learning and Culture

The major difference between the layperson's definition and expert's conceptualization of intelligence was with respect to academic intelligence. Experts emphasized on the motivation whereas lay persons stressed on interpersonal and social aspects of intelligence.

Some researchers have also found that intelligence is a function of stages of development as can be seen from brief description below:

- **Infancy** - During this period of development, intelligence was associated with physical coordination, awareness of people, verbal output and attachment.
- **Childhood** - During this period of development, intelligence was associated with verbal facility, understanding and characteristics of learning
- **Adulthood** - During this period of development, intelligence was associated with verbal facility, use of logic and problem solving.

It has been observed that children develop notions of intelligence when they reach the first grade. Young children's concept of intelligence emphasize interpersonal skills, such as being polite, acting nice, being helpful and good to others. Older children conceive of intelligence as involving academic skills, such as reading well.

Intelligence Defined: Views of Scholars and Test Professionals:

The definition of intelligence is one area where psychologists most disagree with each other. There is no unanimity among them. Spearman (1927) remarked "In truth, intelligence has become a word with so many meanings that finally it has none". Similarly, decades later Wesman (1968) concluded that "There appears to be no more general agreement as to the nature of intelligence or the most valid means of measuring intelligence today than was the case 50 years ago." It is also relevant to note the statement of Edwin G. Boring (1923) that he made decades ago with respect to what intelligence is. According to him "intelligence is what the tests test". Some important definitions of intelligence given by experts are as follows:

Francis Galton: Galton is the pioneer of intelligence movement. He was a contemporary of Alfred Binet and his work on heredity and intelligence is remarkable. He wrote extensively on heritability of intelligence.

According to Galton, intelligent people had excellent sensory abilities. According to him intelligence tests were nothing but measures of sensory abilities, Hence, experts influenced by his views developed tests of visual acuity and hearing ability. Galton measured intelligence through various sensory motor and other perception related tests.

Alfred Binet: He did not give an explicit definition of intelligence, though he believed that there are certain components of intelligence. According to Alfred Binet, intelligence is made up of the following components: Reasoning, Judgement, Memory and Abstraction.

According to Alfred Binet, "intelligence is the capacity of an individual to reason well to judge well and to be self-critical."

Binet and his colleague, Henri attempted to assess complex measures of intellectual ability. Binet was interested in measuring intelligence because

he was faced with the practical problem of identifying intellectually limited school children in schools of Paris who could be benefited from regular instructional programme and may require special educational experiences.

David Wechsler: David Wechsler is another, pioneer in the measurement of intelligence whose work in the 1950s and 1960s led to development of several tests of intelligence.

According to him "Intelligence can be operationally defined as the aggregate or global capacity of an individual to act purposefully, think rationally and deal effectively with the environment".

According to Wechsler, intelligence is composed of abilities which though not independent, but are qualitatively different.

Intelligence, according to him is not the mere sum of different abilities. According to Wechsler, in the assessment of intelligence, even the non-intellectual aspects must be taken into consideration. Some important non-intellectual factors that influence intelligence and its assessment are as follows:

- Drive
- Persistence
- Goal awareness
- Individual's potential to perceive and respond to social, moral and aesthetic values and
- General personality of an individual.

He viewed intelligence as constituting of four factors: Verbal Comprehension, Working Memory, Perceptual Organization- and Processing Speed.

Jean Piaget: Jean Piaget was a Swiss developmental psychologist whose work on intelligence among children has been very influential. Jean Piaget studies development of cognition in children. He specifically studied as to how children think? How they understand themselves and the world around them and how they reason and solve problems.

Jean Piaget conceived of intelligence as a kind of evolving biological adaptation to the outside world. Piaget viewed that intelligence is a result of joint interaction of biology and environmental forces. He viewed intelligence as a cognitive ability consisting of four stages. Individuals move through these four stages at different rates and ages. According to Piaget, biological aspects of mental development is governed by inherent maturational mechanisms. Experiences at each stage of cognitive development helps to organize and reorganize the mental structures (also called as schemas). Piaget used the word schema to refer to an organized action or mental structure, that, when applied to the world, leads to knowing or understanding. The Plural of schema is schemata. In initial years schema is tied to simple behaviour such as sucking and grasping. As they grow

older, schemata become more complicated and are tied less to overt actions than to mental transformations.

Piaget conceived of learning as occurring through two basic mental operations: assimilation and accommodation. Assimilation is defined as actively organizing new information so that it fits in with what is already perceived and thought. Accommodation is defined as changing what is already perceived or thought so that it fits with new information. The following table lists the Piaget stages of cognitive development.

Sensori Motor	Birth to 24 months
Preoperational	2 to 7 years

Period of concrete operations	7 to 11 years
Period of formal operations	7 to 11 years

A common theme among all the above mentioned researchers on intelligence is their focus on interactionism. According to interactionism viewpoint intelligence is influenced by and is a result of joint interaction of heredity and environment.

3.2.2 Theories of Intelligence:

1. Factor-Analytic Theories of Intelligence:

The focus of factor analytic theories is on identifying the ability or groups of abilities deemed to constitute intelligence.

Factor analysis is one statistical method. It consists of a group of statistical techniques that is designed to determine the existence of underlying relationships between sets of variables including test scores. The method of factor analysis has been used to study correlations between tests measuring varied abilities presumed to reflect intelligence.

Charles Spearman: It was Charles Spearman who pioneered the methods of determining Inter-correlations between tests. Charles Spearman developed the theory of general intelligence which is also called as the two - factor theory of intelligence. He found that measures of intelligence tended to correlate to various degrees with each other.

According to Spearman, all intellectual activity is dependent primarily upon and is an expression of a general factor common to all mental activities. It is designated by the symbol 'g'. This factor enters into various degrees in various proportions and different individuals possess this factor in different degrees. All activities do not make an equal demand for this factor. Thus, a task, which requires more 'g' factor, will be done poorly by an individual who has less of such factor. Spearman also pointed out that this factor can only be indirectly observed by constructing psychological tests. Hence, according to him the aim of psychological testing should be to measure the

amount of an individual's 'g' factor, because it is this factor that provides the only basis for prediction of subject's performance from one situation to another.

A test that is highly loaded with 'g' factor requires insight into relationship e.g., reasoning test. On the other hand, a test that is not highly loaded with 'g' factor does not require insight into relationship. E.g., of such a test is a test involving mechanical or rote type of learning.

According to Spearman, a test that exhibited high positive correlations with other intelligence tests were thought to be highly saturated with 'g'. On the other hand, tests with low or moderate correlations with other intelligence tests were thought as a means of possible measure of 's' factors. (e.g. visual acuity, motor ability, etc.)

A test that is highly loaded with 'g' factor is better able to measure and predict intelligence. According to Spearman, it is the 'g' factor rather than the 's' factor that is the better predictor of overall intelligence.

The best measure of 'g' factor in test of intelligence were items dealing with abstract-reasoning.

Spearman also noted that between the 'g' factor and the 's' factor were intermediate class of factors common to a group of activity which were called as group factors. These group factors include the following:

- Linguistic abilities
- Mechanical abilities
- Arithmetical abilities

Some researchers have found highly specific factors. One such group of g factors is listed below:

Factors	Description of Factors
R	Ability to repeat a chain of verbally presented materials
R1	Ability to process sound
R2	Ability to retain verbally presented stimuli
R3	Speed of processing verbally presented stimuli

Many Multiple factor models of intelligence has been proposed. These include the work of Thurstone, Guilford, Gardner, Raymond Cattell, Carroll, etc. We would discuss about each of them.

Thurstone's work: Louis Thurstone found that intelligence is not composed of one or two factors such as "g" or "s" factor as proposed by Spearman, but it is composed of many factors, which he called it as Primary Mental Abilities. Thurstone identified seven factors, which he used to construct Test of Primary Mental Abilities. Though revised versions of this

test are frequently used, the predictive power of this test has been questioned on a number of grounds.

Guilford's work: Guilford was of the view that intelligence is composed of 150 different abilities. He de-emphasized the importance of g factor.

Gardner's work: Gardner also believed that there are different types of intelligence requiring different abilities and different areas of the brain controlled different types of ability. The seven different and independent types of intelligence identified by Gardner are as follows:

1. Interpersonal (social skills).
2. Intrapersonal (personal adjustment).
3. Spatial (artistic).
4. Logical mathematical.
5. Linguistic (verbal).
6. Musical.
7. Kinesthetic (athletic).

Gardner's descriptions of interpersonal and intrapersonal intelligence have found expressions in the concept emotional intelligence.

Raymond Cattell: Raymond Cattell (1941) and later on Horn (1966) postulated the two major types of cognitive abilities which they called as:

- a. Crystallized
- b. Fluid Intelligence

Crystallized Intelligence (GC) consists of acquired skills and knowledge that are dependent on exposure to a particular culture as well as formal and informal education. Crystallized intelligence consists of the following:

- Retrieval of information
- Application of general knowledge

Fluid Intelligence (GF): It is made up of abilities that are:

- Nonverbal
- Culture free
- Independent of specific instructions

Over the years Horn has proposed several additional factors which include as follows:

- Visual Processing (Gv)
- Auditory Processing (Ga)
- Quantitative Processing (Gq)
- Speed of Processing (Gs)

- Facility with Reading and Writing (Grw)
- Short Term Memory (Gsm)
- Long Term Storage and Retrieval (Glr)

He divided these abilities in to two broad groups:

Vulnerable abilities: These abilities, such as Visual Processing (Gv), decline with age and tend not to return to, pre-injury levels following brain damage.

Maintained abilities: These are other abilities such as, Quantitative Processing (Gq). They are called maintained abilities as they tend not to decline with age and may return to pre-injury levels following brain damage

Three stratum theory of cognitive abilities:

This theory was proposed by Carroll (1997) and is based on factor analytic approach. It is a hierarchical model, meaning that all the abilities listed in the stratum are subsumed by or incorporated in the strata above. According to Carroll the three strata are as follows:

1. The top stratum, consists of "g" factor or general intelligence.
2. The second stratum, consists of eight abilities and processes as follows:
 - i. Fluid intelligence (Gf)
 - ii. Crystallized intelligence (Gc)
 - iii. General memory and learning (Y)
 - iv. Broad visual perception (V)
 - v. Broad auditory perception (U)
 - vi. Broad retrieval capacity (R)
 - vii. Broad cognitive speediness (S)
 - viii. Processing / decision speed (T)

At the third stratum are the "level factors" and / or "speed factors". Each of these is different and based on their linkage to second stratum. The three factors linked to Fluid intelligence (Gf) include, general reasoning, quantitative reasoning and Piagetian reasoning. A speed factor linked to Gf include speed of processing. Similarly, four factors linked to Crystallized intelligence (Gc) include: language development, comprehension, spelling ability and communication ability. Two speed factors linked to Gc are oral fluency and writing ability. The three-stratum theory is a **hierarchical model**, meaning that all of the abilities listed in a stratum are subsumed by or incorporated in the strata above.

The CHC Model: Some researchers, using factor analysis and other statistical methods have attempted to extract, blend and combine the various factors creating more complex models. One such model is called as the

Cattell-Horn-Carroll (CHC) model. An integration of Cattell-Horn and Carroll models was proposed by Kevin S McGrew (1997) as well as by McGrew and Flanagan. the McGrew- Flanagan CHC model features ten “broad-stratum” abilities and over seventy “narrow-stratum” abilities, with each broad-stratum ability subsuming two or more narrow-stratum abilities. The ten broad-stratum abilities, with their “code names” in parentheses, are labeled as follows: fluid intelligence (*Gf*), crystallized intelligence (*Gc*), quantitative knowledge (*Gq*), reading/ writing ability (*Grw*), short-term memory (*Gsm*), visual processing (*Gv*), auditory processing (*Ga*), long-term storage and retrieval (*Glr*), processing speed (*Gs*), and decision/reaction time or speed (*Gt*). These models were developed to improve the practice of psychological assessment in education. The Cattell-Horn and Carroll models are similar in many ways. For example, the broad abilities (second-stratum level in Carroll’s theory) envelop many narrow abilities (first-stratum level in Carroll’s theory). However, there are some differences too. For example, for Carroll, *g* is the third-stratum factor, subsuming *Gf*, *Gc*, and the remaining six other broad, second-stratum abilities. By contrast, *g* has no place in the Cattell-Horn model.

2. Information Processing Theories of Intelligence:

The focus of information processing theories is on identifying the specific mental processes that constitute intelligence. Many Information Processing view of intelligence has been developed.

Luria's approach: Russian neuropsychologist Aleksander Luria (1966) emphasized on the information processing approach to intelligence. His approach focuses on the mechanisms by which information is processed. He focuses on how information is processed rather than what is processed. The two basic types of information processing style that he emphasized includes

- a. Simultaneous (parallel) processing
- b. Successive (sequential) processing

In simultaneous processing, information is integrated all at one time, whereas in successive processing, each bit of information is individually processed in sequence. Successive (sequential) processing is logical and analytic in nature. According to Luria, memorizing a telephone number or learning the spelling of a new word is typical of the type of tasks that involve acquisition of information through successive processing.

Simultaneous processing is information processed, integrated and synthesized at once as a whole. For e.g., when one is observing and appreciating a painting in an art museum, the information conveyed by the painting is processed in a manner that, at least for most of us, could reasonably be described as simultaneous.

Information processing perspective is evident in the intelligence test developed by Kaufman Assessment Battery for Children. Information processing view is also evident in the PASS model of intellectual functioning developed by Das (1972) and Naglieri (1990). The word PASS

is an acronym for Planning, Attention (Arousal), Simultaneous and Successive. In the PASS model the term:

- Planning refers to strategy development for problem solving.
- Attention, also called as arousal refers to receptivity of information.
- Simultaneous and successive refers to the type of information processing employed.

One important test that has been developed to assess PASS is called as the cognitive Assessment System.

Robert Sternberg: Robert Sternberg has developed another information processing approach to intelligence. According to Sternberg the essence of intelligence is that it provides a means to, govern ourselves so that our thoughts and actions are organized, coherent and responsive to both our internally driven needs as well as the needs of our environment. According to Robert Sternberg intelligence is influenced by three main factors. They are: Context, Experience and basic information processing mechanism.

Sternberg is well known for his Triarchic theory of intelligence which can be divided into three parts as follows:

- a. The componential part, which is basically concerned with, processing or cognitive processing.
- b. The experiential part, which is basically concerned with the processes by which experience influences intelligence. It deals with the effect of experience on one's intelligence.
- c. The contextual part which is basically concerned with the effect of one's culture and environment on one's intelligence. The componential part of this theory is highly developed. According to it, there are three types of components.
 - i. Knowledge acquisition component: This component of intelligence is concerned with an individual's ability to learn new information.
 - ii. Performance component: This component of intelligence is concerned with knowing how to solve specific problems.
 - iii. Metacomponent: This component of intelligence is concerned with solving problem in general or learning general ways to approach problem solving. It is the metacomponent, which will help us to distinguish between the more intelligent and the less intelligence individual.

Each of these three components overlaps to a great extent and operate in a collective and integrative manner, rather than each operating independently. During the process of problem solving, each of these three operates together.

3.3 MEASURING INTELLIGENCE

Measurement of intelligence involves sampling an examinee's performance on different types of tests and tasks as a function of developmental level. Two important topics related to measuring intelligence include:

- a. Types of Tasks used in Intelligence Tests
 - b. Theory in Intelligence Test Development and Interpretation We would discuss both these topics briefly.
- a. Types of Tasks used in Intelligence Tests:** An important issue related to types of tasks used in intelligence tests are as to whose intelligence are we measuring; Intelligence of infant, children or adult. Different types of tasks are used in the measurement of intelligence of these three groups of people

Measuring Infant's Intelligence: In infancy (i.e. from birth to 18 months) measuring intelligence consists of measuring sensorimotor development. The measurement of sensori-motor development consists of the following activities.

- Turning over
- Lifting one's head
- Sitting up
- Following a moving object with the eyes
- Imitating gestures
- Reaching for a group of objects.

In measuring and assessing intellectual ability of infants, examiner must be highly skillful in establishing and maintaining rapport.

Parents, guardians and caretakers are an important source of information about activities of children. Structured interviews taken with them will help us to accurately assess the intelligence of infants.

Measuring Intelligence of Children: Measuring intelligence of children is also a highly skillful task. Children's intelligence is measured by assessing their verbal and performance abilities. Their intelligence is assessed by evaluating the following:

- a. General fund of information
- b. Vocabulary
- c. Social judgment
- d. Language
- e. Reasoning
- f. Numerical concepts

- g. Auditory and visual memory
- h. Attention
- i. Concentration and
- j. Space visualization

In the earlier period, intelligence tests were scored and interpreted with reference to mental age. Mental age can be defined as an index that refers to the chronological age equivalent to one's performance on tests or a subtest. The index of mental age was typically derived by reference to norms that indicate the age at which most test takers are able to pass or otherwise meet some criterion performance.

Besides performance on various measures, examiners also note the nonverbal and other behaviour of the children who are taking intelligence test. Verbal and nonverbal behaviour of the children who are taking the test of intelligence can yield considerable information about their performance and this can help, the examiner to interpret the obtained results on the test.

Measuring Intelligence of Adult: Wechsler pioneered in measurement of adult intelligence. According to him, adult intelligence scales should be able to measure the following abilities:

1. Retention of general information
2. Quantitative reasoning
3. Expressive language
4. Memory and
5. Social judgment

It should be remembered that tests of intelligence are seldom administered to adults for purposes of educational placement. They are generally given to obtain clinically relevant information or for the measurement of learning potential and skill acquisition. Adult intelligence helps us to assess: faculties of an impaired individual (i.e., whether an individual is senile, traumatized or otherwise impaired) for the purpose of judging that person's competency to make important decisions.

- b) **Theory in Intelligence Test Development and Interpretation:** Measurement of intelligence is also influenced by as to what we mean by intelligence. Galton viewed intelligence to be made up of sensorimotor and perceptual abilities and hence he devised tests to measure sensorimotor and perceptual differences between individuals. On the other hand Binet as well as Spearman developed formal theories to assess intelligence. Spearman emphasized the universal unity of the intellectual function with g as its centerpiece. David Wechsler wrote extensively on the intelligence and viewed intelligence as multifaced and conceived intelligence as not only made up of cognitive abilities but also factors related to personality.

Thorndike conceived of intelligence in terms of three clusters of ability: Social intelligence (dealing with people), Concrete intelligence (dealing with objects) and abstract intelligence (dealing with verbal and mathematical symbols). Factor analytic theories and various other theories have been used in the process of test development and interpretation.

3.4 THE STANFORD-BINET INTELLIGENCE SCALES

After Binet's death in 1911, Binet tests underwent more and more revision, especially in America. The most widely used revision of this test came to be called as Stanford-Binet revision or test, because these tests were revised under the direction of Prof. L.M. Terman at Stanford University.

The following are the important revisions of this test.

1. 1916: Stanford Revision and extension of the Binet-Simon scale by Lewis M. Terman.
2. 1937: Revised Stanford-Binet tests of intelligence (Forms L and M) by Lewis M. Terman and Maud A. Merrill.
3. 1960: Stanford-Binet, third edition, Form L-M, by Lewis M. Terman and Maud A. Merrill.
4. 1972: Stanford-Binet, Form L-M (renaming) by Lewis M. Terman, Maud A. Merrill and Robert L. Thorndike.
5. 1986: Stanford-Binet Intelligence Scale, 4th edition by Robert C. Thorndike, Elazabeth P. Hagen, and Jerome M. Sattler.
6. 2003: Stanford-Binet Intelligence 5th edition by Gale H. Roid.

The first edition of the Stanford Binet Scales (S-B Scales) was not without major flaws. There was lack of representativeness of the standardization sample. The first edition of the S-B Scales was the first published test to provide organized and detailed administration and scoring instructions. It was also the first American test to employ the concept of IQ and the first test to introduce the concept of alternate form.

The 1937 revision of this test began in 1926 and it took 11 years to complete this revision. This revision had two equivalent forms form L (for Lewis) and M (for Maude Merrill who was one of the coauthors of the 02nd edition of this test). New types of tasks for use with preschool-level and adult-level test takers were developed. The manual contained many examples to aid the examiner in scoring. One important criticism of this edition was the lack of representation of minority groups during the test's development.

The test was again revised in the year 1960. This test had only one form and included the items considered to be the best from the two forms of the 1937 test. A major innovation of the 1960 revision was the use of concept of Deviation IQ. The second innovation of this scale was that the IQ tables have been expanded to include chronological ages 17 and 18 because latest

finding indicated that mental development, as measured by Stanford-Binet, continues at least up to that age.

Another revision of the S-B Scales was published in the year 1972. This scale was criticized for the quality of its standardization sample. Its manual was vague with respect to the number of minority individuals in the standardization sample. It is also said that this scale over represented the western urban communities in the standardization sample.

The fourth revision of S-B Scales appeared in 1986 and constituted a major departure from the previous versions with respect to theoretical organization, test organization, test administration, test scoring and test interpretation. As opposed to the earlier three revisions of the scale which were age scales. The fourth edition was a point scale. A point scale is a test organized in to subtests by category of items, not by age at which most test takers are presumed capable of responding in the way that is keyed as correct. The manual of the fourth edition contains an explicit exposition of the theoretical model of intelligence that guided the revision. The model was based on the Cattell-Horn model of intelligence.

The Stanford-Binet intelligence was once again revised in 2003. It is the fifth revision and is also called as the SB5. It is an individually administered assessment of intelligence and cognitive abilities and is suitable for people in the age range of 2 years to 85 + years. The SB-5 blends many of the important features of the earlier editions with significant improvements in psychometric design. It provides comprehensive coverage of five factors of cognitive ability:

- Fluid reasoning
- Knowledge
- Quantitative processing
- Visual-spatial processing
- Working memory

The new features of the SB-5 are as follows:

- Wide variety of items requiring nonverbal performance by examinee ideal for assessing subjects with limited English, deafness, or communication disorders
- Ability to compare verbal and nonverbal performance useful in evaluating learning disabilities
- Greater diagnostic and clinical relevance of tasks, such as verbal and nonverbal assessment of working memory
- Includes Full Scale IQ, Verbal and Non-verbal IQ, and Composite Indices spanning 5 dimensions -with a standard score mean of 100, SD 15
- Includes subtest scores with a mean of 10, SD 3

- Extensive high-end items, many adapted from previous Stanford-Binet editions and designed to measure the highest level of gifted performance
- Improved low-end items for better measurement of young children, low functioning older children, or adults with mental retardation
- Enhanced memory tasks provide a comprehensive assessment for adults and the elderly.
- Co-normed with measures of visual-motor perception and test-taking behaviour
- Scoreable by hand or with computer software.
- Enhanced artwork and manipulatives those are both colorful and child- friendly.

Uses :

The SB-5 may be used to diagnose a wide variety of developmental disabilities and exceptional abilities and may also be useful in:

- Clinical and neuropsychological assessment.
- Early childhood assessment.
- Psycho-educational evaluations for special education placements.
- Adult workers compensation evaluations.
- Providing information for interventions such as IFPs, IEPs, career assessment, industrial selection, and adult neuropsychological treatment.
- A variety of forensic contexts.
- Research on abilities and aptitudes.

Some important points worth noting about SB-5 are as follows:

- i. The SB-5 is based on the Cattell-Horn-Carroll (CHC) theory of intellectual abilities.
- ii. SB-5 has good reliability data, interscorer reliability ranges between 0.74 to 0.97 with a median of 0.90.
- iii. Only a few subtest items of SB-5 are timed. Most of the SB-5 items are not timed.
- iv. Normative data for the SB-5 were gathered from 44,800 individuals between the ages of 2.0 and 85 + years. The normative sample closely matches the 2000 U.S. Census (education level based on 1999 data).
- v. The SB-5 was co-normed with the Bender (R) Visual-Motor Gestalt Test, Second Edition. Reliability for the SB-5 are very high. For the FSIQ, NVIQ, and VIQ, reliabilities range from .95 to .98
- vi. Reliabilities for the Factor Indexes range from .90 to .92. For the ten subtests, reliabilities range from .84 to .89.

3.5 THE WECHSLER TESTS

David Wechsler initially developed test for measuring adult intelligence and later on he developed a series of tests to measure intelligence of children and infants. Some general types of items used in Wechsler tests of intelligence are as follows:

WAIS IV - The Wechsler Adult Intelligence Scale - Fourth Edition is the current available test in the Wechsler series, Some of the earlier tests include:

- Wechsler-Bellevue - I (WB-I)
- Wechsler Bellevue - II (WB-II)
- WAIS
- WAIS-R
- WAIS - III

WAIS-IV is the most recent edition to the family of Wechsler Adult Scales. The fourth edition of the test (WAIS-IV) was released in 2008 by Pearson. It consists of 10 core subtests and five supplemental subtests. A core subtest is one that is administered to obtain a composite score. Under usual circumstances, a supplemental subtest (also sometimes referred to as an optional subtest) is used for purposes such as providing additional clinical information or extending the number of abilities or processes sampled. In certain situations, supplemental subtest is used in place of a core subtest, under following conditions:

- when examiner incorrectly administered a core subtest
- or the assessee had been inappropriately exposed to the subtest items prior to the administration of the test
- the assessee has the physical limitation that affected the assessee's ability to effectively respond to the items of a particular test.

Some important features of WAIS-IV are as follows:

- The WAIS-IV was standardized on a sample of 2,200 people in the United States ranging in age from 16 to 90 years and 11 months.
- More explicit administration instructions.
- Expanded use of demonstration and sample items
- Floor and Ceiling limits are extended. WAIS IV has a full scale IQ ceiling of 160 and Full Scale IQ floor of 40.
- It is sensitive to the needs of the older adults
- The images in the picture completion, symbol search and coding subtests have been enlarged.
- An average reduction in the overall test administration time from 80 to 67 minutes.

- In WAIS IV we do not measure three different types of IQ such as Full - Scale IQ, Verbal IQ and Performance IQ as we used to do in the earlier versions.

3.5.1 Wechsler Intelligence Scale for Children - Fourth Edition (WISC-IV)

Wechsler scale for children was first published in 1949. It represented the downward extension of Wechsler Bellvue scale. The original Wechsler Intelligence Scale for children had many flaws in it. The standardization sample contained only white children and some test items were viewed as perpetuating gender and cultural stereotypes. WISC was revised in 1974 and came to be called as WISC R. This test has been adapted in India by Prof. Malin and is called as Malin's Indian adaptation of WISC.

WISC-IV is the latest revision and improved version of WISCIII and was published in the year 2003. WISC-IV yields the following measures:

- i. General intelligence functioning (Full Scale IQ also called as FSIQ)
- ii. Four index scores viz.:
 - a. Verbal comprehension index
 - b. Perceptual reasoning index
 - c. Working memory index
 - d. Processing speed index

Each of these indices is based on scores on three to five subtests.

In WISC-IV, following subtests have been eliminated

- a. Picture arrangement
- b. Object assembly and
- c. Mazes

The following subtests are supplementary tests

- Information
- Arithmetic
- Picture completion

WISC-IV contains 10 core subtests and 5 supplemental tests.

3.5.2 Wechsler Preschool and Primary Scale of Intelligence Third edition (abbreviated as WPPSI - III)

The origin of Wechsler Preschool and Primary Scale Intelligence (WPPSI) can be traced to the year 1967 when Wechsler for the first time decided that a new scale should be developed and restandardised especially for children who were under the age of 6 years.

The WPPSI was the first major intelligence test that adequately sampled the total population of the United States, including racial minorities. The WPPSI was revised in the year 1989 and came to be called as WPPSI-R. This test was designed to assess the intelligence of children from ages 3 years through 7 years and 3 months. In this revision new items were developed to extend the range of the test both upward and downward.

WPPSI-III was published in the year 2002. This test further extended to the age range of children who could be tested with this instrument downward to 2 years and 6 months.

The WPPSI-III had many changes incorporated in it as compared to earlier editions. Some important changes were as follows:

1. The following 5 subtests which were present in the earlier editions were dropped from WPPSI. These five subtests are as follows:
 - Arithmetic
 - Animal pegs
 - Geometric designs
 - Mazes and
 - Sentences
2. The following seven new subtests were added in WPPSI-III.
 - Matrix reasoning
 - Picture concepts
 - Word reasoning
 - Coding
 - Symbol search
 - Receptive vocabulary
 - Picture naming
3. WPPSI-III has different labels for certain subtest. Some of these are as follows:
 - Core subtests
 - Supplemental subtests
 - Optional subtests

Core subtests are those that are required for the calculation of composite score.

Supplemental subtests are used to provide broader sampling of intellectual functioning. These subtests may also substitute for a core subtest if a core subtest cannot be administered due to some reasons or was administered but its score cannot be used or has become unusable.

Optional subtests are those that may not be used to substitute for core subtests but may be used in the derivation of optional scores.

The WPPSI was aimed at measuring two variables:

- Fluid reasoning
- Processing speed

Wechsler, Binet and the Short Form: Short forms of intelligence tests including that of Wechsler tests have been developed. Short form refers to test that has been abbreviated in length, typically to reduce the time needed for test administration, scoring and interpretation. Short forms of the test are used for two purposes: convenience of the test administrator and the practical necessities with the client that mandates the use of short forms. Short forms of the test are not new. Doll (1917) used the short form of the Binet Simon test. In 1958 David Wechsler endorsed the use of short forms, but only for screening purposes. Wechsler Abbreviated Scale of Intelligence (WAIS) was developed in 1999. Watkins (1986) concluded that short forms may be used for screening purposes only, but not to make placement and educational decisions. Smith McCarthy and Anderson (2000) held the view that short form must be used with caution. Silverstein (1990) has pointed out as to how short forms can be used with caution.

3.6 SUMMARY

In this unit we had discussed about the definition of intelligence and the various theories of intelligence. Among the definitions of intelligence, we had examined the views of the lay public as well as the views of scholars and test professionals. We studied the views of Francis Galton, Alfred Binet, David Wechsler, Jean Piaget. Two most important theories of intelligence that we discussed were the factor-analytic theories and the information processing views. The topic of measuring intelligence was also discussed.

The Stanford- Binet Tests and the Weschsler tests are an important group of tests.. Among the Stanford Binet Tests we discussed the 05th edition of Stanford-Binet Intelligence Scales in brief. We also discussed Weschsler tests, such as the Weschsler Adult Intelligence Scale, 04 th Edition (WAIS - IV), Wechsler Intelligence Scale for the Children, 04 1h Edition (WISC - IV) as well as the Wechsler Preschool and Primary Scale of Intelligence, 03rd Edition (WPPSI-III). A brief mention of the short forms of intelligence tests was also discussed.

3.7 QUESTIONS

1. Define Intelligence and discuss the views of the lay public as well as Scholars and Test Professionals with respect to intelligence.

2. Discuss the Factor-Analytic and Information Processing Theories of Intelligence.

3. Write short notes on:

a. Measuring Intelligence

b. The Stanford-Binet Scales of Intelligence

c. WAIS IV - The Wechsler Adult Intelligence Scale Fourth Edition

d. Wechsler Intelligence Scale for Children - Fourth Edition (WISC -IV)

e. Wechsler Preschool and Primary Scale of Intelligence Illrd edition (abbreviated as WPPSI-III).

3.8 REFERENCES

Cohen, J.R., & Swerdlik, M.E. (2010). Psychological Testing and Assessment: An introduction to Tests and Measurement. (7 th ed.). New York. McGraw-Hill International edition.

Anastasi, A. & Urbina, S. (1997). Psychological Testing. (7th ed.). Pearson Education, Indian reprint 2002.

MEASUREMENT OF INTELLIGENCE, INTELLIGENCE SCALES, PROBABILITY, NORMAL PROBABILITY CURVE AND STANDARD SCORES - II

Unit Structure :

- 4.0 Objectives
- 4.1 Introduction
- 4.2 The concept of Probability
- 4.3 Characteristics, importance and applications of the Normal Probability Curve
 - 4.3.1 Characteristics of Normal Probability Curve:
 - 4.3.2 Importance and Applications of The Normal Probability Curve
- 4.4 Areas under the Normal Curve
- 4.5 Skewness- positive and negative, causes of Skewness, formula for calculation
 - 4.5.1 Measurement of skewness
 - 4.5.2 Causes of Skewness
- 4.6 Kurtosis - meaning and formula for calculation
- 4.7 Standard scores - z, T -Score, stanine; linear and non-linear transformation; Normalized Standard scores
 - 4.7.1 Common Types of Standard Scores
- 4.8 Summary
- 4.9 Questions
- 4.10 References

4.0 OBJECTIVES

After studying this unit, you should be able to:

- Understand the concept of Probability.
- Explain the characteristics, importance and applications of the Normal Probability Curve.
- Know the Areas under the Normal Curve.
- Understand the concept of skewness and discuss the causes of skewness
- Explain the concept of Kurtosis
- Explain the concept of standard scores and know its various types.

4.1 INTRODUCTION

In this unit, we will first discuss the concept of probability. We will define probability and understand its meaning through some examples. The concept of Normal Probability Curve is very important in psychological measurement. We would discuss its characteristics as well as importance and the various applications of the said curve. Towards the end of the unit we will discuss the areas under normal curve.

We will discuss the concept of skewness and kurtosis in this unit and understand the causes for the same. We will also attempt to understand how divergence from normality is measured. Types of skewness and kurtosis would also be studied in this unit. Also, we will discuss the concept of standard scores and its various types.

4.2 THE CONCEPT OF PROBABILITY

The word probability or chance as it is sometimes called is very commonly used in day-to-day conversation. The theory of probability has its origin in the games of chances related to gambling such as throwing a die, tossing a coin, drawing cards from pack of cards, etc. Cardano, an Italian mathematician was the first to write a book on the subject titled "Book on Games of Chances" which was published in 1663, after his death.

Galileo, an Italian mathematician was the first man to attempt quantitative measure of probability while dealing with some problems related to theory of dice in gambling. However, systematic and scientific foundation of the mathematics theory of probability was laid in mid-seventeenth century by two French mathematician B Pascal and Pierre de Fermat. Swiss mathematician James Bernoulli came out extensive work in the following decades.

The probability of a given event is an expression of likelihood or chance of occurrence of event. A probability is a number which ranges from zero to one, zero for an event which cannot occur and 1 for an event certain to occur.

Probability can be defined as the ratio of number of favorable results to the total number of results for e.g. say a coin is tossed and we want head at the top. Since there is only one head out of two possibilities (Head and Tail) the required probability of getting Head or Tail is $1/2$.

Whenever there is more than one result, the probability comes into existence and it is possible to use a procedure for calculating the probability in such cases. For certainty probability does not exist. But for every uncertainty probability should exist.

There are certain examples in which the question of probability does not arise at all. For example, when kerosene is poured into a fire, the result is certain and hence there is no question of thinking any other probability or when a stone is thrown in the air, there is absolutely no doubt that it will fall

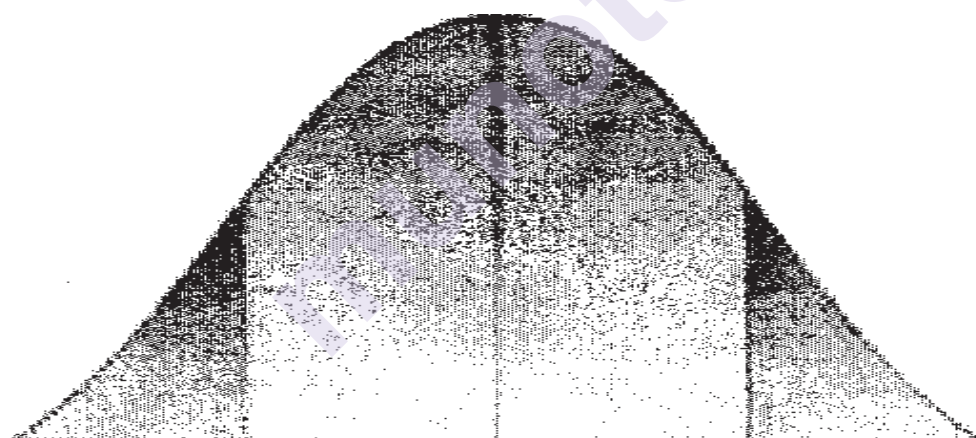
down. Hence there is certainty and no probability. Thus, we observe that when a result is certain the question of probability does not exist at all.

If you roll a six-sided die, there are six possible outcomes, and each of these outcomes is equally likely. A six is as likely to come up as a three, and likewise for the other four sides of the die. What, then, is the probability that a one will come up? Since there are six possible outcomes, the probability is $1/6$. What is the probability that either a one or a six will come up? The two outcomes about which we are concerned (a one or a six coming up) are called favorable outcomes. Given that all outcomes are equally likely, we can compute the probability of a one or a six using the formula.

4.3 CHARACTERISTICS, IMPORTANCE AND APPLICATIONS OF THE NORMAL PROBABILITY CURVE

Normal probability curve is one type of theoretical distribution that is of immense use in statistics. Normal probability curve is also called as the normal curve, the Gaussian curve (after a great German mathematician who investigated its properties and wrote the equation for it). It is also called as the Bell-Shaped curve or Mesokurtic curve (Mesos means middle or medium). The following figure depicts the Normal probability curve:

Figure 4.1 Normal Probability Curve



2.15% 13.6% 34.1% 34.1% 13.6% 2.15%

168.2%

95.4%

99.7%

Much evidence has accumulated to show that the normal distribution serves to describe the frequency of occurrence of many variable facts with a relatively high degree of accuracy. Phenomenon which follow the normal probability curve (at least approximately) may be found in Biological statistics, Anthropological data, Social and Economic data, Psychological measurements and Errors of observation.

4.3.1 Characteristics of Normal Probability Curve:

Following are the major characteristics of Normal Probability Curve:

1. The normal curve is symmetrical at about the mean. The number of cases below the mean in a normal distribution is equal to number of cases above the mean. Hence, the mean and median coincide.
2. The height of the curve is maximum at its mean. Hence, the mean and mode of normal distribution coincide. Thus, Mean, Median and Mode are equal in normal probability curve.
3. There is one maximum point of the normal curve which occurs at the mean. The height of the curve declines as we go in higher direction from the mean. The dropping off is slow at first, then rapid, and then slow again. Theoretically, the curve never touches the base line. Its tail approaches, but never reach the base line. Hence, the range is unlimited.
4. Since there is only one point in the curve which has maximum frequency the normal probability curve is unimodal, i.e. it has only one mode.
5. The point of inflection i.e. the points where the curvature changes in direction are each plus or /and minus one standard deviation from the mean ordinate.
6. The height of the curve at a distance of one standard deviation from the mean is 60.7% of the height at the mean. The height of the curve at 2 and 3 standard deviation distances from the mean is 13.5% and 1.1 % of the height at mean respectively.
7. The total interval from plus one standard deviation to minus one standard deviation contains 68.26% of the cases. Similarly, 95.44% of the total area will be included between the mean ordinate and an ordinate 2 standard deviation from the mean. Similarly, 99.74% of the total area will be included between the mean ordinate and a point 3 standard deviation away from the mean.
8. For the normal curve the value of $Ku = 0.263$.
9. In normal probability curve $Q1$ and $Q3$ are equidistant from the median. When there is any Skewness in the distribution, the two distances will be unequal.
10. In the normal probability curve the height declines symmetrically in either direction from the maximum point.
11. The Normal Curve is a Mathematical Model in Behavioral Sciences. The curve is used as a measurement scale. The measurement unit of this scale is plus or minus 1, 2, 3, etc., standard deviation

4.3.2 Importance and Applications of The Normal Probability Curve

Normal probability curve has wide significance and applications in the field of measurement concerning education, psychology and sociology. The Normal Distribution is by far the most used distribution for drawing inferences from statistical data. Number of evidences are accumulated to show that normal distribution provides a good fit or describe the frequencies of occurrence of many variables and facts in (i) biological statistics e.g., sex ratio in births in a country over a number of years, (ii) the anthropometrical data e.g., height, weight, (iii) wages and output of large numbers of workers in the same occupation under comparable conditions, (iv) psychological measurements e.g., intelligence, reaction time, adjustment, anxiety and (v) errors of observations in Physics, Chemistry and other Physical Sciences.

The Normal distribution is of great value in educational evaluation and educational research, when we make use of mental measurement. It may be noted that normal distribution is not an actual distribution of scores on any test of ability or academic achievement, but is instead, a mathematical model. The distribution of test scores approach the theoretical normal distribution as a limit, but the fit is rarely ideal and perfect.

Some of its main applications are as follows:

1. **Use as Model:** Normal curve represents a model distribution. For this reason, it may be used as a model.
 - a. To compare various distributions with it, to say, whether the distribution is normal or not, if not what way it diverges from the normal.
 - b. To compare two or more distributions in terms of overlapping, and
 - c. To distribute short marks and categorical rating.

Often, phenomena in the real world follow a normal (or near normal) distribution. This allows researchers to use the normal distribution as a model for assessing probabilities associated with real-world phenomena. Typically, the analysis involves two steps.

- **Transform raw data.** Usually, the raw data are not in the form of z- scores. They need to be transformed into z-scores, using the transformation equation such as :

$$z = (X - p) / a.$$

- Find probability. Once the data have been transformed into scores, you can use standard normal distribution tables, online calculators (e.g., Stat Trek's free normal distribution calculator), or handheld graphing calculators to find probabilities associated with the z-scores.

2. **To compute Percentile and Percentile Ranks:** Normal probability curve may be conveniently used for computing percentile and percentile Ranks in a given normal distribution.
3. **To understand and apply the concept of Standard Error of Measurement:** The normal curve is also known as the normal curve of error, or simply the curve of error on the grounds that it helps in understanding the concept of standard errors of measurement. For e.g., if we compute mean for the distributions of the various samples taken from a single universe (population), then, these means will be found to be distributed normally around the mean or the center of population. The sigma distance of a particular sample mean may help us to determine standard error of measurement for the mean of that sample.
4. **For ability grouping :** A group of individuals may be conveniently grouped into certain categories like good, average, poor, etc. in terms of some trait (assumed to be normally distributed) with the help of the normal curve.
5. **To convert Raw Scores into Comparable Standard Normalized Scores :** Sometimes, we have records of an individual performance on two or more different kinds of measurement and we wish to compare his score on one measurement with the score on the other measurement unless the scales of these two tests are the same, we cannot make a direct comparison. With the help of the normal curve, we can convert the raw scores belonging different scales of measurement into standard normalized scores like Sigma (or Z scores) and T scores.
6. **To determine the relative difficulty of test items:** Normal curve provides the simplest rationale method of scaling test items for difficulty and therefore, may be conveniently employed for determining the relative difficulty of test questions problems and other test items.

Thus, from the above discussion we see that there are number of applications of normal curve in the field of educational measurement and evaluation. These are:

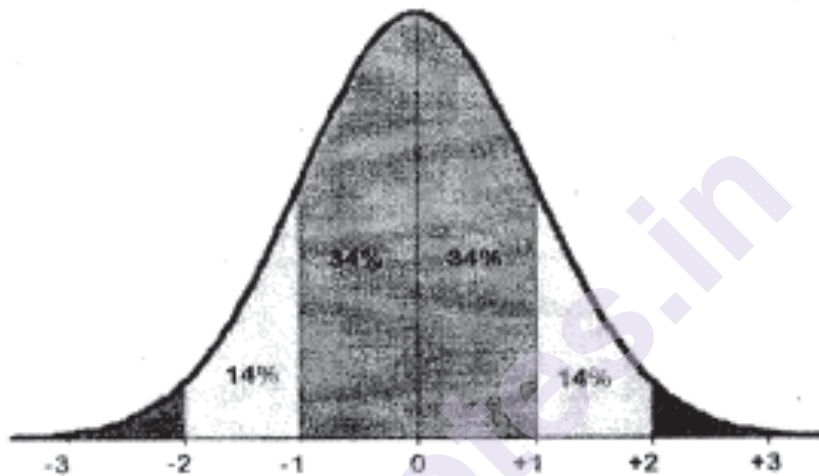
- i. To determine the percentage of cases (in a normal distribution) within given limits or scores
- ii. To determine the percentage of cases that are above or below a given score or reference point
- iii. To determine the limits of scores which include a given percentage of cases
- iv. To determine the percentile rank of a student in his own group
- v) To find out the percentile value of a student's percentile rank
- v. To compare the two distributions in terms of overlapping

- vi. To determine the relative difficulty of test items, and
- vii. Dividing a group into sub-groups according to certain ability and as- signing the grades.

4.4 AREAS UNDER THE NORMAL CURVE

We are often interested in finding the areas under the normal curve which are associated with some given z score. The normal curve can be divided into sections by each standard deviation, beginning with a z score of zero in the center.

Figure 4.2 Areas under the Normal Curve



Notice that the total area under the curve is 1.000 by definition (that would correspond to 100% in Figure 4.2 above). The area under the curve is depicted in Figure 4.3 below.

Figure 4.3

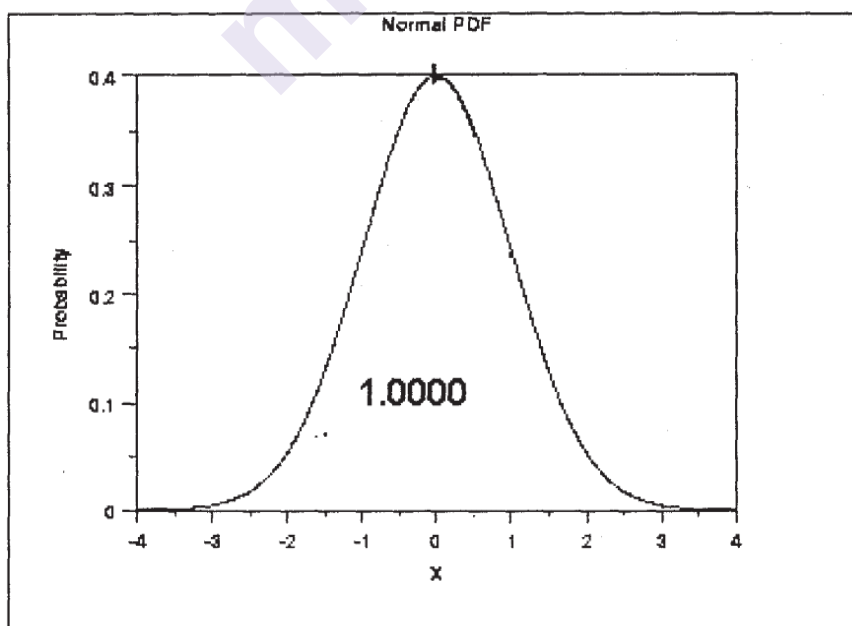
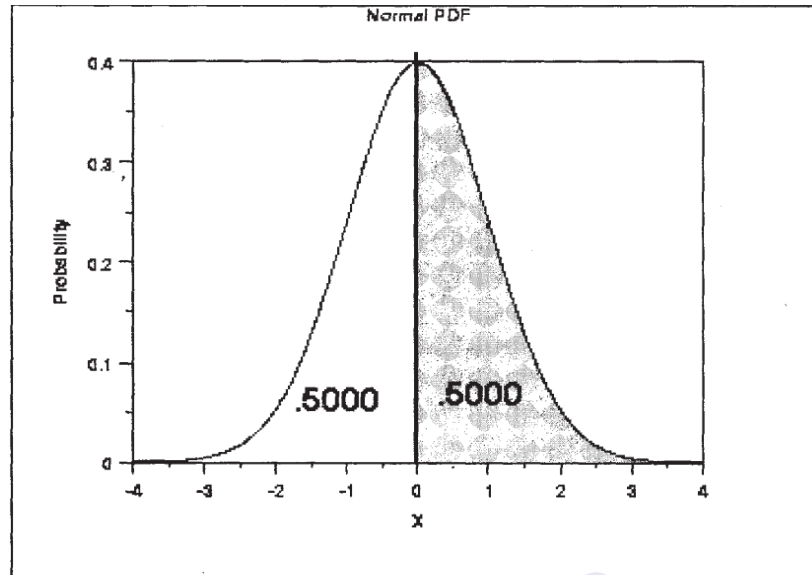


Figure 4.4



Now try to see (Figure 4.4 below) that the curve can theoretically be split in half at $z = 0$, and there will be .5000 area above $z = 0$ and .5000 are below $z = 0$.

Frequently we ask questions in statistics which require us to know the area above or below a given z value, or the area between two z values. Referring to Figure 3 above, note that one can say that the area above $z=0$ is .50, and the area below $z=0$ is .50. Notice from Figure 1 we can make all kinds of statements about integer z values. For instance, the area between $z = -1$ and $z = +1$ is .68 (because $.34 + .34 = .68$). The area between $z = -2$ and $z = +1$ is $.14 + .34 + .34$, or .82.

Frequently, though, our z values of interest are no perfect integers. For instance, we might need to know the area above $z = +1.74$. That is why we have z tables, and now is the time to learn about them.

There are basically three types of z tables. However, this being a basic introductory course on Psychological Testing and Statistics we need not go in to its details.

The table of areas of normal probability curve is referred to find out the proportion of area between the mean and the z value. One determines the probability of occurrence of a random event in a normal distribution by consulting tables of areas under a normal curve. Tables of the normal curve have a mean of 0 and a standard deviation of 1. To use the table, you must convert your data to have a mean of 0 and standard deviation of 1. This is done by transforming your raw values into z -scores, according to this formula:

z -score

$$z_1 = \frac{x_1 - \bar{X}}{s_1}$$

4.5 SKEWNESS - POSITIVE AND NEGATIVE, CAUSES OF SKEWNESS, FORMULA FOR CALCULATION

Measurement of Intelligence,
Intelligence Scales,
Probability, Normal
Probability Curve and
Standard Scores - II

Whenever a curve lacks symmetry, we call it skewness. Divergence from Normal Probability Curve gives rise to skewness or Kurtosis. Not all distributions in the real life display normal curve phenomenon. Some distributions deviate from normality.

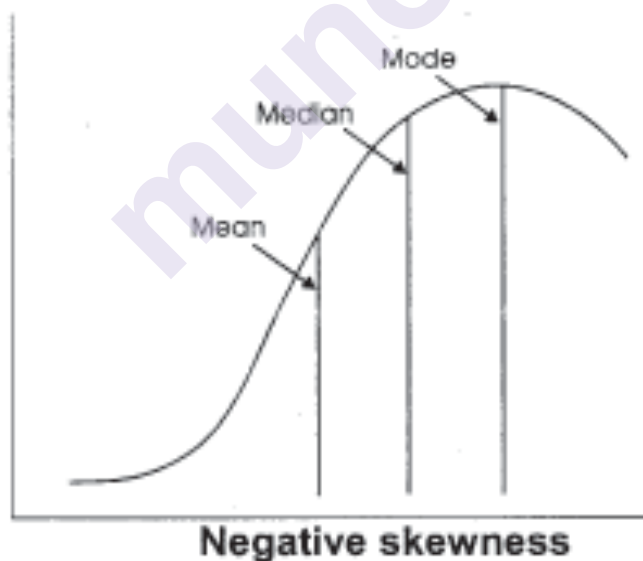
The word Skewed means lacking symmetry or distorted. Skewness shows the direction of symmetry. Skewness of the distribution tells how lopsided the distribution is. A distribution is said to be skewed when the mean and median fall at different points in the distribution and the balance or center of gravity is shifted to one side or the other.

Skewness depends upon the manner in which the scores in a series scatter about the average value. When the scatter is greater on one side of the point of central tendency than on the other, the distribution is said to be skewed.

In a normal distribution the mean equals the median and the skewness is of course zero. The more nearly the distribution approaches the normal form, the closer together are the mean and the median and the less the skewness.

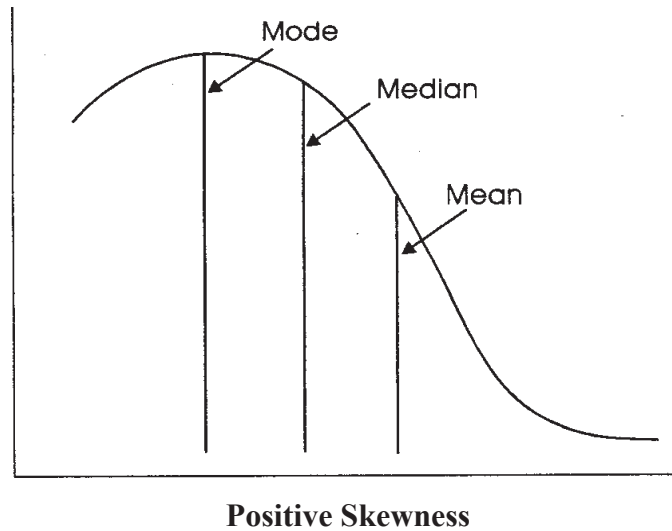
Distributions are said to be skewed negatively or to the left when scores are massed at the high end of the scale (the right end) and are spread out more gradually towards the low end or left.

Figure 4.5 Negative Skewness



Distributions are said to be skewed positively when there is piling of scores at the low end and a long tail running up in high scores.

Figure 4.6 Positive Skewness



In skewness the mean is pulled more toward the skewed end of the distribution, than the median. In fact, greater the gap between the mean and the median the greater the skewness. Moreover, when the skewness is negative the mean lies to the left of the median and when skewness is positive the mean lies to the right of the median.

4.5.1 Measurement of skewness:

There are three different measures of skewness which are as follows:

1. One method used for measuring skewness is by graphic analysis. Whenever Mean Median and Mode are not equal, the curve is skewed. But graphic method fails to give exact numerical value of skewness.
2. Second method used for measuring skewness is by the following formula. This is also called as the Karl Pearson coefficient of Skewness

$$Sk = 3(\text{mean} - \text{median}) / \text{Standard Deviation.}$$

Interpretation:

- If $Sk = 0$, then the frequency distribution is normal and symmetrical.
 - If Sk is greater than 0, then the frequency distribution is positively skewed.
 - If Sk is less than 0, then the frequency distribution is negatively skewed. It will take value between +1 and -1.
3. Third method of measuring skewness is in terms of percentiles.

$$Sk = P_{90} + P_{10} - 2(P_{50}P_{90}) - P_{10}$$

4.5.2 Causes of Skewness:

Skewed data distributions are a result of extreme values, also known as outliers. These can be due to many causes which are discussed below:

1. **Selection:** Selection is a potent cause of skewness. If the sample you choose is a biased one, the distribution of the scores will not exhibit the bell-shaped form.
2. **Unsuitable or poorly made tests:** Normality or lack of normality is dependent upon the number of items and their difficulty. If a test is too easy, scores will pile up at the high scores end of the scale and will give negative skewness, whereas if the test is too hard, scores will pile up at the low score end of the scale giving a positively skewed curve.
3. **Non-normal Distributions:** Skewness or Kurtosis or both will appear when there is a real lack of normality in the trait being measured e.g. if a loaded die is tossed for a number of times, the resulting distribution will certainly be skewed and probably be peaked. This is because the loaded side is a dominant factor in determining the result of the tosses. Non-normal curves often occur in the medical statistics. In the case of childhood disease, for example, death rate would be maximum in the early ages and would decrease with increase in age. The distribution would be positively skewed.
4. **Errors in the use of Test:** Errors in timing or in giving instructions, errors in scoring, differences in motivation, all of these factors, if they cause some students to score higher and others to score lower than they normally would, tend to make for skewness in the distribution.

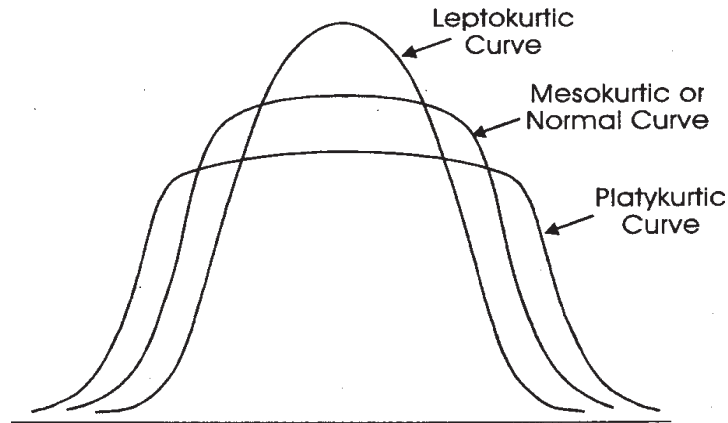
4.6 KURTOSIS - MEANING AND FORMULA FOR CALCULATION

The term kurtosis refers to the peakedness or flatness of a frequency distribution as compared with the normal. Kurtosis describes how peaked or flat the distribution is.

When there is high concentration of scores in the neighborhood of the point of central tendency, the distribution is relatively narrow across the shoulders. Relatively high and narrow distributions are described as leptokurtic. When there is low concentration of scores in the neighborhood of the central tendency the distribution is relatively broad across the shoulders. Such relatively flat-topped distributions are described as platykurtic.

A normal distribution is called mesokurtic. The Figure 5.3 depicts roughly the kurtosis of different types.

Figure 4.7 Different Types of Kurtosis



Kurtosis can be measured by the graphic method. Besides the graphic method we can measure kurtosis, based on quartile or percentile method by the following formula:

A measure of kurtosis based on quartiles and percentiles is as follows:

$$K = \frac{Q.D.}{P_{90} - P_{10}}$$

This is known as the Percentile Coefficient of Kurtosis. It has been shown that K for a normal distribution is 0.263 and that it lies between 0 and 0.50.

We can also measure kurtosis from normal probability curve. For the normal curve the value of $K_u = 0.263$, if the value of K_u is less than 0.263 we infer that distribution is Leptokurtic. If the value of K_u is greater than 0.263, the distribution is platykurtic. If kurtosis values tend toward 0, then the distribution approximates a normal distribution.

4.7 STANDARD SCORES - Z, T - SCORE, STANINE; LINEAR AND NON-LINEAR TRANSFORMATION; NORMALISED STANDARD SCORES

In the areas of measurement of personality individuals scores are often compared with reference to standard scores to obtain an assessment of an individual. Standard scores are ways to measure positions on the normal curve. They are standard because the size of the distribution or type of measurement always fall in the same place. They include standard deviations, percentile ranks, z- scores, T-scores, etc. A **standard score** is a raw score that has been converted from one scale to another scale, where the latter scale has some arbitrarily set mean and standard deviation.

The percentages of the area under the normal curve are the same for each standard deviation. Once you calculate the standard deviation for any set of

scores, you can find any standard score for the data set. You can also figure out where any score falls relative to others.

Standard scores assume a normal distribution. They provide a method of expressing any score in a distribution in terms of its distance from the mean in standard deviation units. A Standard Score indicates how far a particular score is from a test's average. With a standard score, the position of a test-taker's performance relative to other test-takers is readily known.

In standard scores the unit that tells the distance from the average is the standard deviation (sd) for that test. For WAIS-III, the average is 100 and the sd is 15. The standard deviation (sd) is always given for a standard score. Standard Scores between -1 sd (85) and +1 sd (115) fall in the normal range on the ability being tested.

Above + 1 sd (115+) a learner is in the top 15% of performance. Below -1 sd (-85), she / he is in the lowest 15% of performances.

Standard scores are used in norm-referenced assessment to compare one student's performance on a test to the performance of other students her age. Standard scores estimate whether a student's scores are above average, average or below average compared to peers. They also enable comparison of a student's scores on different types of tests, as in diagnosing learning disabilities.

Standard Scores are used generally for the following purposes:

- i. To tell the exact location of a score in a distribution. For e.g. Raju is 10 years old and has a weight of 50 kg. How does his weight compare to other 10 years old boys.
- ii. Standard Scores also help us to compare scores across different distributions. For e.g. Geeta scored 65 in her chemistry paper, 75 in maths and 60 in English. On which test she performed better.

Standard Scores can be obtained by either linear or non-linear transformations of the original raw scores. When founded by linear transformation, they retain the exact numerical relations of the original raw scores, because they are computed by subtracting a constant from each raw score and then dividing the result by another constant. Linearly, derived standard scores are often designated simply as "Standard Scores" or "z-scores". To compute a z-score we find the difference between the individual's raw score and the mean of the normative group and then divide the difference by the S D of the normative group.

4.7.1 Common Types of Standard Scores:

1. Z-Scores:

These scores are scaled on a number line ranging from -4 to + 4 with zero being in the middle. On this scale, zero is average. Positive scores are above average, and negative scores are below average.

One type of standard score is a z-score, in which the mean is 0 and the standard deviation is 1. This means that a z-score tells us directly how many standard deviations the score is above or below the mean. For example, if a student receives a z score of 2 her score is two standard deviations above the mean or the 84th percentile. A student receiving a z score of -1.5 scored one- and one-half deviations below the mean. Any score from a normal distribution can be converted to a z score if the mean and standard deviation is known. The formula is

$$Z \text{ score} = (\text{Score} - \text{mean score}) / (\text{Standard deviation})$$

So, if the score is 130, the mean is 100, and the standard deviation is 15, then the formula leads to this calculation:

$$Z = \frac{(130 - 100)}{15} = 2$$

2. **T-Scores:** These scores range from 10 - 90 in intervals of 10 points. Fifty is average

on this scale. A T-score, by definition, has a mean of 50 and a standard deviation of 10. This means that a T score of 70 is two standard deviations above the mean and so is equivalent to a z score of 2.

3. **Stanines:** Stanines (pronounced stay-nines) are often used for reporting students'. The stanine scale is also called the standard nine scale. These scores range from 1 - 9 with five being average. Scores below five are below average. Scores above five are above average.

4.8 SUMMARY

In this unit, we have discussed the concept of Probability, which can be defined as the ratio of number of favorable results to the total number of results. The normal probability curve is based upon the law of probability was also defined and discussed. Its characteristic features were discussed. The importance and applications of the Normal Probability Curve and the areas under the Normal Curve were also discussed in brief with illustrations.

We have discussed the concepts of skewness and kurtosis in this unit along with their measurement and types, and the causes of skewness. Standard scores are frequently used in psychological measurements to compare one student's performance with another or score obtained by one student on one test with scores obtained by him on another test. There are many different types of standard scores. The three most common types include: Z-Scores, T-Scores and Stanines.

4.9 QUESTIONS

Measurement of Intelligence,
Intelligence Scales,
Probability, Normal
Probability Curve and
Standard Scores - II

1. Discuss the concept of Probability.

2. Explain the characteristics features and importance of Normal Probability Curve.

3. Discuss the applications of the Normal Probability Curve.

4. Write a note with illustrations on the 'Areas under the Normal Curve'.

5. Define Skewness and Kurtosis and explain the causes of skewness.

6. Discuss how Skewness and Kurtosis are measured.

7. Explain the different types of Skewness and Kurtosis.

4.10 REFERENCES

Cohen, JR, & Swerdlik, M.E. (2010). Psychological Testing and Assessment: An introduction to Tests and Measurement. (7 th ed.). New York. McGraw-Hill International edition.

Anastasi, A. & Urbina, S. (1997). Psychological Testing. (7th ed.). Pearson Education, Indian reprint 2002.

ASSESSMENT OF PERSONALITY - I

Unit Structure :

- 5.0 Objectives
- 5.1 Introduction
- 5.2 Definitions of Personality and Personality Assessment
 - 5.2.1 Traits, Types and States
- 5.3 Personality Assessment - some basic questions
- 5.4 Developing instruments to assess Personality - logic and reason, theory, data reduction methods, criterion groups.
- 5.5 Personality Assessment and Culture
- 5.6 Summary
- 5.7 Questions
- 5.8 References

5.0 OBJECTIVES

After studying this unit you should be able to:

- Define Personality, Personality Assessment and related terms such as Traits, Types and States.
- Understand some basic questions related to Personality Assessment.
- Comprehend the tools or instruments used in the process of assessment of personality, such as the use of logic and reason, theory, data reduction methods and criterion groups
- Know the relationship between Personality and Culture.

5.1 INTRODUCTION

In this unit, we will first define the concepts of Personality, Personality Assessment and related terms such as Traits, Types and States. Following this we will discuss some basic questions related to personality assessment such as who is being actually assessed, what is assessed when a personality assessment is conducted, where are personality assessments conducted, how are personality assessments structured and conducted.

In the process of personality assessment various tools or instruments are used which includes logic, theory, data reduction methods, such as factor analysis and the criterion groups. These constitute the technical aspects involved in developing instruments to assess personality. Assessment of personality is intimately tied to one's culture as well as language. Issues related to acculturation and other considerations would also be discussed.

5.2 DEFINITIONS OF PERSONALITY AND PERSONALITY ASSESSMENT

Personality: Personality is an important aspect of human individuality and psychologists have defined and measured it in scientific ways. Psychologists do not agree with the definitions of personality. Hence, different scholars have defined it in different ways leading to varied definitions of personality.

McClelland defined personality as "the most adequate conceptualization of a person's behaviour in all its details". According to Menninger, personality can be defined as "the individual as a whole, his height, and weight and love and hates, blood pressure and reflexes, his smiles and hopes and bowed legs and enlarged tonsils. It means all that anyone is and that he is trying to become". Some scholars such as Goldstein (1963) has defined personality very narrowly and focuses on a particular aspect of an individual whereas Sullivan (1953) has defined personality in the context of society. Some Psychologists avoid defining the term personality (Byrne (1974). Byrne characterized the entire area of personality as "Psychology's garbage bin in that any research which does not fit other existing categories can be labeled personality".

Hall and Lindzey (1970) in their classic text "Theories of Personality" state that "personality is defined by the particular empirical concepts which are a part of the theory of personality employed by the observer". According to Cohen and Swerdlik (7th Edition), personality can be defined as an individual's unique constellation of psychological traits and states.

Personality Assessment:

Personality assessment is also sometimes incorrectly referred to as psychological testing. Personality assessment can be defined as the measurement and evaluation of psychological traits, states, values, interests,

attitudes, worldview, acculturation, personal identity, sense of humor, cognitive and behavioral styles, and/or related individual characteristics. The term assessment as used in personality assessment is different from psychological testing.

5.2.1 Traits, Types and States

Personality Traits:

There is no consensus among psychologists as to the meaning of the term traits. In general, it refers to enduring characteristics or aspects of personality that are stable over time and across situations. Gordon Allport viewed personality traits as real physical entities that are bona fide mental structures in each personality. Psychological traits can be viewed as attributions made in an effort to identify threads of consistency in behavioral patterns.

According to Guilford trait refers to "any distinguishable, relatively enduring way in which one individual varies from another"

Personality psychologists have developed formal methods for describing and measuring personality. Their use of the term trait (defined as stable and enduring characteristic of an individual) is different from the everyday use of this term that we make in three ways.

- Personality psychologists have used statistical methods to reduce the number of traits on the basis of similarity.
- They have used reliable and valid instruments to measure traits.
- They have used carefully designed method to conduct empirical research to demonstrate the relationship between certain specific behaviour and traits.

Personality theorists and assessors have assumed for years that personality traits are relatively enduring over the course of one's life. Roberts and DelVecchio (2000) explored the endurance of traits by means of a meta-analysis of 152 longitudinal studies. These researchers concluded that trait consistency increases in a step like pattern until one is 50-59 years old, at which time such consistency peaks.

Personality Types:

Personality type can be defined as a unique constellation of traits and states that is similar in pattern to one identified category of personality within a taxonomy of personalities. Personality types are actually descriptions of people. A type is a class of individuals who share a common collection of traits together in an individual.

Some important personality types as identified and discussed by different scholars and psychologists in their works are as follows:

One of the earliest personality type theorists was Greek physician Hippocrates around 400 BC. He is also called as the father of modern medicine.

Working on the assumption that the human body contains four fluids, or humours (blood, phlegm, black bile, and yellow bile), he categorized people into four corresponding personality types as follows:

- i. Phlegmatic (a calm, apathetic temperament caused by too much phlegm).
- ii. Choleric (a hot headed, irritable temperament due to an excess of yellow bile).
- iii. Sanguine (an optimistic, hopeful temperament attributed to a predominance of blood).
- iv. Melancholic (a sad, depressed temperament based on black bile). This types of theory of personality, today, is only of historical importance.

In 1925, a German psychiatrist Kretschmer, in his book, physique and character classified individuals into certain biological types according to their physical structure. Kretschmer classified personality into three types :Pycnic (having fat bodies), Athletic (balanced bodies) and Leptosomatic (lean and thin bodies).

William Sheldon, Carl G. Jung , John Holland, etc., has also given different type theories of personality. The Myers-Briggs Type Indicator (MBTI), test is based on Carl Jung's personality types. John Holland categorized people into following six personality types

- i. Artistic
- ii. Enterprising
- iii. Investigative
- iv. Social
- v. Realistic
- vi. Conventional

Meyer Friedman and Ray Rosenman developed two personality types called the Type A and Type B personality types. Type A personality is characterized by competitiveness, haste, restlessness, impatience, feeling of being time- pressured and strong needs for achievement and dominance. Type B personality is opposite of Type A's traits. Type B people are relaxed, easy going, laid-back or mellowed.

The personality type that has attracted the most attention from researchers of clinicians is the one associated with scores on Minnesota Multiphasic Personality Inventory.

Data from administration of MMPI tests are frequently discussed in terms of patterns of scores that emerge on the subtests. This pattern is referred to as a profile. In general, a profile is a narrative description, graph, table or other representation of the extent to which a person has demonstrated certain targeted characteristics as a result of the administration or application of the tools of assessments.

Personality States:

The terms personality state in psychological assessment literature is used in the following two different contexts:

- a. It is used to refer to an inferred psychodynamic disposition designed to convey the dynamic quality of Id, Ego and Superego in perceptual conflict.
- b. It is a term used to refer to the transitory exhibition of some personality trait. In other words, the term state is indicative of a relatively temporary predisposition.

Measuring personality state amounts, in essence, to a search for and an assessment of the strength of traits that are relatively transitory of fairly

situation specific. Charles D. Spielberger and his associates have developed a number of personality inventories designed to distinguish various states from traits. They distinguish between Trait and State anxiety. Trait anxiety or anxiety proneness refers to relatively stable or enduring personality characteristics. State anxiety on the other hand refers to a transitory experience of tension because of a particular situation. One test developed by Spielberger and his associates to measure states from traits is called as State-Trait Anxiety Inventory (STAI).

5.3 PERSONALITY ASSESSMENT - SOME BASIC QUESTIONS

Personality assessment is aimed to find practical solutions or answers to issues that confront us either related to our health, work, career, life decisions, etc. Questions related to personality assessment either in basic research or practical problems seek to explore answers that can help us to handle a given problem or issue at hand. Four important basic questions related to personality assessment are as follows:

1. Who is actually being assessed? Can the test taker be someone other than the subject of the assessment?
2. What is being assessed? What is assessed when personality assessment is conducted?
3. Where is personality assessment conducted?

How is personality assessment structured and conducted, we would discuss each of these briefly.

Who : One of the important questions related to personality assessment is that while assessing an individual's personality, who needs to be assessed. The concerned individual, his/her spouse, children, parents, friends or other informants. Who can be a test taker? An individual herself or someone other than the subject of assessment.

Many tests make use of self-report or in which one assess the individual herself whose personality assessment is required. The individual answers in wide variety of ways. Some of which are as follows:

- They respond to interview questions.
- Answer questionnaires in writing.
- Blacken squares on computer answer forms.
- Sort cards with various terms on them, etc.

In certain types of assessment, we rely on informants other than the person being assessed to acquire personality related information. For e.g., while assessing children we ask parents and/or teachers to participate in personality assessment of children by asking them their judgments, opinion and impressions about a particular child being assessed. Thus, while

assessing an individual, following three things are important, each of which we will discuss in brief:

a. The self as the primary referent:

In many testing or assessment situations an individual who is being assessed is an important source of information. Many different types of assessment require self-report. It is a process whereby information about assesseees is supplied by the assesseees themselves. Self report information is generally obtained in the following ways:

- From diaries report by the assesseees.
- Response to oral or written questions or test items.

Self-report methods are very commonly used to assess an individual's self- concept. A number of self-concept measures for children have been developed. Some representative tests include:

- Tennessee Self-Concept Scale
- Piers-Harris Self-Concept Scale

self-concept differentiation:

Callero, 1992 believed that states and traits related to self-concept are to a large degree context-dependent—that is, ever-changing as a result of the particular situation. The term **self-concept differentiation** refers to the degree to which a person has different self-concepts in different roles. People characterized as highly differentiated are likely to perceive themselves quite differently in various roles. For example, a highly differentiated college professor in his 40s may perceive himself as motivated and hard driving in his role at work, conforming and people-pleasing in his role as son and emotional and passionate in his role as a husband. By contrast, people whose concept of self is not very differentiated tend to perceive themselves similarly across their social roles. Self-report measures are highly valuable source of information provided by an individual. The person who gives the said information is assumed to have reasonably accurate insight into their own thinking and behavior, assumed to be highly motivated and answers the questions in an honest manner. The greatest limitation of self-report method includes:

Faking on the part of the assessee to impress the examiner or to conceal some vital or embarrassing or highly personal information, such as that related to certain behaviour.

Some assesseees may lack insight into their own behaviour and hence may not be able to reveal the accurate picture of their own self. In that case, another person is used as a primary referent.

b. Another person as the primary referent:

In many situations, we can acquire reliable and best information from a third party. The third party can be a parent, a spouse, a teacher, peer, supervisor, etc. For example, in the assessment of an emotionally disturbed child parents and/ or teachers can be a best source of information.

While acquiring information from others the raters is likely to make many errors. Raters can make biased judgments, consciously or unconsciously as it is in their own self-interest to do so. Some of the most common types of errors on the part of the raters are as follows:

- i. ***Error of leniency or generosity:*** Tendency to score or rate leniently.
- ii. ***Error of severity or stringency:*** Tendency to score or rate strictly.
- iii. ***Halo effect:*** A type of rating error wherein the rater views the object of rating with extreme favor and tends to bestow ratings inflated in apposite direction
- iv. ***Error of central tendency:*** It is the general tendency to rate everyone near the midpoint of a rating scale.

Many other factors may also influence how raters rate a given individual. Some important factors that can bias raters are as follows:

- The rater may feel competitive with, physically attracted to or physically repelled by many extraneous or similar factors between the ratee and rater.
- The rater may also not have the proper background, experience or training in rating a given individual.
- The rater's judgment may also be limited by his/her general level of conscientiousness and willingness to devote the time and effort required to do the job properly.
- Different raters may also have different perspectives on the individual they are rating by virtue of their context in which they typically view that person. For e.g., a parent may indicate on a rating scale that a child is hyperactive, whereas the class teacher, on some rating scale may indicate that the child's activity level is within normal limits.

c. The cultural background of the assessee:

Many researchers and test administrators take into account the cultural variables in the use of assessment. While administering scoring and interpreting a given assessment instrument or method due

regard should be given to cultural, factors and/or variables which are likely to impact on the assessment process.

What: One of the important questions with regard to assessment is as to what is assessed when a personality assessment is conducted. Two important aspects of what is being test in personality assessment are as follows:

- a. Primary content area sampled
- b. Test taker response style

With respect to primary content area it should be remembered that personality test may measure one area or aspect of personality such as anxiety or extroversion or shyness or it may measure many aspects of personality as in MMPI.

Many tests, today measure test takers' response style also. **Response style** refers to a tendency to respond to a test item or interview question in some characteristic manner regardless of the content of the item or question. A particular response style such as responding to a personality test in an inconsistent, contrary or random way or attempting to take good or bad, etc., may invalidate a given test. Some testtakers engage in impression management, i.e., they try to give socially desirable answers. Now, many personality tests contain items designed to detect different types of response styles.

Where: Another basic. question related to personality assessment is as to where are personality assessments conducted. Traditionally personality assessments have been conducted in different places such as in the:

- Schools
- Clinics and hospitals
- Academic research laboratories
- Employment counseling e.g., vocational selection centers
- Office of psychologists and counselors

Today, assessment is also conducted in natural settings as well as online.

How: Another important question related to personality assessment is with respect to how are personality assessments structured and conducted. How a personality assessment is conducted generally depends upon the scope. The scope of personality assessment can be very wide or very narrow. For e.g., MMPI and California Psychological Inventory are used when an assessment is aimed at a broader scope. On the other hand, when we measure a single trait such as loss of control, our scope is very narrow. Certain tests are based on particular theory of personality whereas there are tests whose development is atheoretical in nature. For example,

Blacky Pictures Test is a theory-based instrument while the Minnesota Multiphasic Personality Inventory (MMPI) is atheoretical instrument.

Procedures and item formats: With respect to "how" of personality assessment, it is also important to note the procedures or item formats used in assessing one's personality. Some important methods of assessing an individual's personality are as follows:

- Face-to-face interviews
- Computer administered tests
- Behavioral observations
- Paper-and-pencil tests
- Evaluation of case history data
- Evaluation of portfolio data
- Recording of physiological responses

Certain methods of personality assessment are highly structured whereas other methods are highly unstructured. The same personality trait or construct may be measured with different instruments in different ways. For example, aggression can be measured by using a paper-and-pencil test; a computerized test; an interview with the assessee; an interview with family, friends, and associates of the assessee; analysis of official records and other case history data; behavioural observation; and laboratory experimentation.

Frame of reference: Another important aspect of how personality measurement is carried out has to do with frame of reference of assessment. We can define frame of reference as aspects of the focus of exploration such as the time frame (the past, the present, or the future) as well as other contextual issues dealing with people, places and events.

One important method that can be applied in the exploration of varied frames of reference is the Q-sort technique. It was a method developed by Stephenson and used extensively by Carl Rogers. Besides Q-sort method two other item presentation formats readily adoptable to different frames of reference are as follows:

- Adjective checklist
- Sentence completion format

Scoring and interpretation: Personality assessments also differ with respect to how tests are scored and interpreted. Two important approaches to scoring and interpretation of personality assessment are as follows:

- a. **Nomothetic approach:** This approach to assessment is characterized by efforts to learn how a limited number of personality traits can be applied to all people. Nomothetic approach assumes that certain personality traits exist in all people to varying degrees. The assessor's

task is to determine what the strength of each of these traits are in the assessee.

- b. **Idiographic approach:** It is characterized by efforts to learn about each individual's unique constellation of personality traits with no attempt to characterize each person according to any particular set of traits.

"How" of test also involves examining the issues in personality test development and use. Many issues in personality test development and use also determine as to how the test will be used. Will the test be a self-report inventory or will it be a projective test. Should personality inventories be used or some other tests.

5.4 DEVELOPING INSTRUMENTS TO ASSESS PERSONALITY - LOGIC AND REASON, THEORY, DATA REDUCTION METHODS, CRITERION GROUPS

Most personality tests employ two or more of the following tools in the development of personality assessment instruments. We would briefly discuss each of these.

1. **Logic and Reason:** While preparing test items we make use of logic and reason, which dictate as to what contents is covered by the items. Use of logic and reason in the development of test items is sometimes referred to as the content or content-oriented approach to test development. In the process of development of test items, we see to it that they are based on American Psychiatric Association's Diagnostic and Statistical Manual Criteria for diagnosis of a particular disorder. Attempts to develop content oriented, face valid items began during First World War in an attempt to develop instruments to assess recruit's personality and adjustment problems. One of the well-known personality tests was Woodworth's Personal Data Sheet (1917) that later on came to be called as Woodworth Psychoneurotic Inventory. This test was designed to elicit self-report fears, sleep disorders and other problems deemed symptomatic of psych neuroticism. The greater the number of problems reported, the more psychoneurotic the test taker was presumed to be.

Self-report instruments of this type can help us to collect a great deal of clinically actionable information in a relatively little time. In order to administer such types of test a highly trained professional is not required. Such instruments are more suited in clinical settings where emphasis is on cost cutting.

Logic reason and intuition are often used in item development. A sound research knowledge and clinical experience is also needed

2. **Theory:** Personality measures differ in the extent to which they rely on a particular theory of personality in their development and interpretation. When a psychological theory is the guiding force

behind the development of a psychological test, rather than reason and logic, then the items are quite different. One theory based test that is predominantly used today is the Self Directed Search (SDS) which is a measure of one's interest and perceived abilities. It was developed by John Holland and his associates. The test is based on Holland's theory of vocational personality. The central idea of the theory is the view that occupational choice has a great deal to do with one's personality and self-perception of abilities. The SDS is self-administered, self-scored and self-interpreted. Test scores direct test takers towards specific occupational themes. From there, test takers follow instructions to learn about various occupations that are consistent with their expressed pattern of interests and abilities.

3. **Data Reduction Methods:** Another category of widely used tool in test development is data reduction methods. Such methods make use of wide variety of statistical techniques collectively -known as factor analysis or cluster analysis. One use of data reduction methods is to, aid in the identification of minimum number of variables or factors that account for the intercorrelations in observed phenomenon. Psychologists using data reduction methods have identified certain primary factors of personality. Considerable research using data reduction methods was carried out by Raymond Cattell who developed, 16 PF Questionnaire. Cattell identified 36 surface traits and 16 source traits.

Eysenck (1991) have argued that primary factors can be narrowed down to three. Some researchers have identified five factor model. One well known such model is that developed by Costa and McCrae called the Big Five Model.

Big Five Model: One test developed to measure big five factors include revised NEO Personality Inventory (NEO PI - R). The five factors of big five include the following:

- **Neuroticism** domain taps aspects of adjustment and emotional stability.
- **Extraversion** domain taps aspects of sociability and assertiveness.
- **Openness** encompasses openness to experience as well as active imagination, aesthetics sensitivity, attentiveness to inner feelings, preference for variety, intellectual curiosity and independence of judgment.
- **Agreeableness** is primarily a dimension of interpersonal tendencies that include altruism, sympathy towards others and belief that others are similarly inclined.
- **Conscientiousness** is a dimension of. personality that involves the active process of planning, organizing and following through.

NEO PI - R is generally used with individuals who are 17 years and above. It is a self-administered test. It can be computer scored and interpreted.

Criterion Groups:

Criterion is defined as a standard on which a judgment or decision can be made. A criterion group is a reference group of test takers who share specific characteristics and whose responses to test items serve as a standard according to which items will be included in or discarded from the final version of the scale. The process of using the criterion groups to develop test items is referred to as empirical criterion keying because the scoring or keying of items has been demonstrated empirically to differentiate among group of test takers.

One test developed using criterion group is the MMPI (Minnesota Multiphasic Personality, Inventory) originally called the Medical and Psychiatric Inventory (Dahlstrom and Dahlstrom, 1980). MMPI consists of various types. Some of the most commonly included types of MMPI are:

- a. MMPI
- b. MMPI - 2
- c. MMPI - 2 - RF
- d. MMPI - A

We would briefly discuss the MMPI as it is the original and one of the classic tests in the history of personality assessment.

MMPI was originally developed by Psychologists Starke K Hathaway and Psychiatrist - Neurologist John Charnley McKinley (1940). It consists of 566 True-False items and was devised as an aid to psychiatric diagnosis with adolescents and adults of 14 years and above. MMPI consists of the following 10 clinical Scales:

1. Hs: Hypochondriasis
2. Hy: Hysteria
3. Mf: Masculinity-femininity
4. Pt: Psychasthenia
5. Ma: Hypomania
6. D: Depression
7. Pd: Psychopathic deviate
8. Pa: Paranoia
9. Sc: Schizophrenia
10. Si: Social Introversion

These above mentioned diagnostic categories were very popular in 1930s. The clinical criterion group for the MMPI was made up of psychiatric inpatients at the University of Minnesota Hospital. The validity scales were also built in the MMPI. These scales include:

- i. Validity Score (F)
- ii. Lie Score (L)
- iii. Question Score
- iv. Correction Score (K)

These scales were not designed to measure validity in the technical, psychometric sense. Out of the 566 items, 16 items are repeated. Scores of MMPI are reported in the form of T-Scores which is one type of standard score with a mean set at 50 and standard deviation set at 10. In addition to the above mentioned clinical and validity scales, there are MMPI content scales, such as Wiggins Content Scales. Content Scales are composed of group of test items of similar content. Supplementary scales are a catch all phrase for the hundreds of different MMPI scales that have been developed since the test's publication. These scales have been developed by different researchers using a variety of methods and statistical procedures, mostly factor analysis. MMPI is today administered by many different methods:

- Online
- Offline on disc
- Index cards
- Audio version for semiliterate test takers is also available with instructions recorded on audiocassette.

MMPI can be scored by hand. Computer scoring is also available. Different psychologists have described different ways of scoring this test. For example, Paul Meehl (1951) proposed a two-point code derived from a number of clinical scales on which the test taker achieved (most pathological) the highest scores. Another popular approach to scoring and interpretation was developed by Welsh called as Welsh Codes.

The other scales such as MMPI - 2, MMPI - 2 - RF and MMPI - A are different developments in MMPI.

5.5 PERSONALITY ASSESSMENT AND CULTURE

Psychologists have become increasingly concerned with the relationship between assessment of an individual personality and the various aspects of an individual's culture. Psychologists are often required to assess individual's personality and other related variables of people who belong to varied and diverse cultures. It should be remembered that with members of culturally and linguistically diverse populations, a routine and business-as-usual approach to psychological testing and assessment is inappropriate.

Research studies, especially in the area of cross cultural psychology has emphasized the sensitivity on the part of psychologists as to how culture relates to behaviours of cognitions that are being measured. Some important aspects of the culture that impacts assessment enterprise are as follows:

- a. Acculturation
- b. Values
- c. Identity
- d. World view and
- e. Language of the assessee

We would briefly discuss each of these.

Acculturation is an process by which an individual's thoughts, behaviours, values, world view and identity develops in relation to general thinking, behaviour, customs, values of a particular group. It is at birth, that the process of acculturation begins and it is through the process of acculturation that develop culturally accepted ways of thinking, feeling and behaving. Few tests have been developed to assess an individual's levels of acculturation to their native culture or dominant culture.

Values are closely associated with acculturation. One's values considerably influence the process of assessment. Different culture emphasizes different values. Indian culture emphasizes group, family and spiritual values as compared to individuality and material culture emphasized by western cultures. Similarly, some cultures emphasize and value "future". Where are other cultures emphasizes on "here and now" the present. Assessment instruments should reflect one's cultural values. One of the earliest works on values was the book titled Types of Men (Spranger 1928), which listed different types of people based on whether they valued things like truth, practicality and power. Rokeach (1973) distinguished between two types of values called as Instrumental and Terminal Values. Instrumental Values are guiding principles to help one attain some objective. Honesty, imagination, ambition and cheerfulness are examples of instrumental values. Terminal values are guiding principles and a mode of behaviour that is an endpoint objective. Some examples of terminal values are a comfortable life, an exciting life, a sense of accomplishment and self-respect. According to Kluckhohn (1960) values provide answers to key questions with which civilizations must grapple.

Personal identity is another important aspect of one's culture that should be kept in mind in the process of assessment. We can define identity as a set of cognitive or behavioral characteristics by which an individual define themselves as members or a particular group. The term identity is closely associated with the term self.

World view is another important aspect of one's culture. It is a unique way in which people interpret and make sense of their perceptions as a consequence of their learning experiences, cultural background and related variables.

One's **language** also influences the ways in which one conceptualizes a given item or statement. Language also influences the way in which one responds to test item.

Any assessment instrument must take into consideration above mentioned cultural factors if meaningful information has to be obtained from assessment data.

In Unit 6, we will discuss various methods of personality assessment, which include objective methods and projective methods.

5.6 SUMMARY

In this unit we have discussed the definitions of personality and personality assessment. We also distinguished between the concepts of traits, types and states. Some important basic questions related to personality assessment were discussed. The questions pertained to Who, what, where and how of personality assessment. Some important techniques used in personality assessment such as logic and reason, theory, data reduction methods, criterion groups were briefly discussed. Following this, we discussed the relationship between personality assessment and culture.

5.7 QUESTIONS

1. Define the concept of Personality and Personality Assessment.

2. Explain the terms Personality Traits, Personality Types and Personality States.

3. Explain some basic questions with regards to personality assessment such as Who, What Where and How of assessment.

4. Write a note on personality assessment and culture.

5.8 REFERENCES

Cohen, JR., & Swerdlik, M.E. (2010). Psychological Testing and Assessment: An introduction to Tests and Measurement. (7 th ed.). New York. McGraw-Hill International edition.

Anastasi, A. & Urbina, S. (1997). Psychological Testing. (7th ed.). Pearson Education, Indian reprint 2002.

munotes.in

ASSESSMENT OF PERSONALITY - II

Unit Structure :

- 6.0 Objectives
- 6.1 Introduction
- 6.2 Objective Methods of Personality Assessment
- 6.3 Projective Methods of Personality Assessment
 - 6.3.1 Inkblot as Projective Stimuli:
 - 6.3.2 Evaluation of the Rorschach Test:
 - 6.3.3 Pictures as Projective Stimuli:
- 6.4 Projective Methods in Perspective
- 6.5 Summary
- 6.6 Questions
- 6.7 References

6.0 OBJECTIVES

After studying this unit you should be able to:

- Understand the various objective methods of Personality Assessment.
- Know the various Projective Methods of Personality Assessment.

6.1 INTRODUCTION

In last unit, we defined the concepts of personality, personality assessment and related terms, such as traits, types and states. We also discussed some basic questions related to personality assessment, as well as various tools or instruments used in the process of personality assessment. We also learned about the technical aspects involved in developing instruments to assess personality. Apart from this, issues related to acculturation and other considerations were discussed.

In this unit, we will discuss the objective and projective methods of personality assessment briefly. Among the projective techniques the Rorschach ink blot test is the most common followed by the Thematic Apperception Test, the Word Association Tests and the Sentence Completion Tests. Sounds and Figure Drawing are also used as projective techniques in the assessment of personality.

6.2 OBJECTIVE METHODS OF PERSONALITY ASSESSMENT

Objective methods of personality assessment consist of paper and pencil tests having multiple objectives and the assessee has to select one of the correct items. The objective methods of personality assessment contain short answer items for which the assessee's task is to select one response from the two or more provided, and the scoring is done according to set procedures involving little or no judgment on the part of the scorer.

Objective methods of personality assessment may include items written in a multiple choice, true/false or matching format.

Response on an objective ability test may be scored correct or incorrect. Objective personality tests share many advantages with objective tests of ability.

Objective items can usually be scored quickly and reliably by varied means, from hand scoring to computer scoring.

In objective, multiple-choice tests of ability there is little room for emotion, bias or favoritism on the part of the test scorer.

Objective personality tests are objective in the sense that they employ a short-answer, typically multiple-choice format, one that provide little, if any, room for discretion in terms of scoring.

Many personality inventories such as MMPI, EPPS or CPI are objective in nature.

6.3 PROJECTIVE METHODS OF PERSONALITY ASSESSMENT

Projective methods of personality assessment are based on the projective hypothesis which holds that individuals introduces structure to unstructured stimuli and thereby reveals his/her personality. It is an ideal that when an individual is presented with an unstructured and ambiguous stimuli and individual will introduce structure and clarity in to it thus revealing his/her unconscious desires, wishes, needs and unconscious impulses. Projective methods refer to a group of related and unrelated techniques used for studying both intellectual and non-intellectual aspects of personality. In these tests, an individual is presented with a relatively unstructured or ambiguous task like a picture, inkblot or incomplete sentences, which permits a wide variety of interpretations by the subjects. The basic assumption underlying projective tests is that individual's interpretation of the task will project his characteristic mode of responses, his personal motives, emotions and desires and thus enable the examiner to understand more subtle aspects of his personality.

One of the well-known projective techniques is the Rorschach Inkblot Test which was developed by Hermann Rorschach in 1921. He called it as the "Form Interpretation Test" as inkblots were the forms to be interpreted. In 1921 published his work in a Monograph titled 'Psychodiagnostic'. Many consider this year as the date for the development of this test. Unfortunately, his first published work was also his last work because he died in the following year. In this monograph he provided 28 case studies, which included normal as well as individuals with psychiatric diagnosis such as neurosis, psychosis and manic-depressive psychosis. The Rorschach test consists of 10 cards having bilateral systematical inkblots. Half of the cards are achromatic and other five share one or more colors. The cards are presented to the subjects in a definite sequence. Rorschach did not impose any time limit on this test nor do present users. There is also freedom to give any type of responses and as many type of responses that are possible.

Though even before Rorschach, researchers have used inkblots to study imagination and other functions, Rorschach was the first to apply inkblots to the diagnostic investigation of the personality as a whole. Today, Rorschach is one of the most frequently used, very popular, widely criticized and extensively researched test. Various innovations in this test have led to the development of multiple systems of test administration scoring and interpretation. Some of the commonly used Rorschach systems are by Klopfer, Beck, Hertz, Piotrowski, Rapaport Schaffer and the last one developed by Exner, called as Comprehensive System.

The cards are presented to the test taker in a definite sequence. The cards are numbered from I to X. While administering the test, the examiner notes the various aspects of subject's behaviour. He keeps a verbatim record of the responses, notes the time elapsed between the presentation of each card and the first response to it (called as initial reaction time), length of the pause between responses, total time required for each card response and the subject's extraneous movements, spontaneous remarks, emotional expressions and any other specific behaviours which appears significant to the examiner.

After all the 10 cards have been presented in a sequential order another phase called as the inquiry phase starts in which the examiner questions the individual systematically regarding the parts and aspects of each blot to which the associations were given. The two important purposes of the inquiry are :

- First, to determine which aspects of blot initiated and sustained the association process,
- Second the inquiry gives the subject an opportunity to add, elaborate or to clarify his original responses, but if this is done it must be completely spontaneous on the part of the subject and without any suggestions from the examiners.

Another important component of test administration is what is called as Testing of Limits, In this procedure the examiner asks specific questions to seek additional information concerning subject's personality functioning. In testing of the limits, the examiner:

- Might ask if the test taker has used the entire blot or a part of it or the white space within the blot to form the -percept (i.e., perception of an image)
- Any confusion or misunderstanding that the subject might have with regard to the task that is the blot.
- Finds out if the test taker is able to refocus percepts given a new frame of reference.
- Assesses if the test taker becomes anxious by the ambiguous nature of the task or is able to perform better, etc.

Several scoring categories for Rorschach test have been developed but the most commonly scored categories are as follows:

- i. Location responses refer to the area of the blot, which has been perceived by the subject as the basis of his response. The subject may respond to entire blot, to a larger or small portion of it, to small minute detail and sometimes even to the white background. The location responses can be very well-defined or only vaguely defined. The location responses to the test are categorized into following classes :

- W (responses to the whole blot),
- 'D' (responses to large usual details),
- 'Dd'(responses to small unusual details),
- 'S' (responses to white space on the card).

The location responses can be given by combining some of the above categories for e.g. D and S can also be perceived together.

The location responses and the subject's ability to delineate them are regarded as indicative of subject's organizing process and the ability to analyze and articulate the part of everyday experience. The analysis of the subject's location response is made in the light of norms prepared by the test author.

- ii. Determinants refer to the characteristics of the inkblot as perceived by the subject. They are those qualities of the blot that have produced the response to it. The four most important determinants are as follows :

- Form,
- Shading,
- Colour, and
- Movement.

Although there is of course no movement in the blot itself, the respondent's perception of the blot as a representation of moving object is scored in this category. Further differentiation is made within these categories. Movement responses are Human Movement (M), Animal Movement (FM), or inanimate movement (m). Similarly, form may be perceived with ordinary accuracy (F), with unusual accuracy (F+) or in a very poor accuracy (F-). Shading may be perceived as representing depth (V) or texture (T), which can be perceived in its pure form (V or T) or it can be combined with form (FV, VF, FT, and TF), colour can be perceived along with form being dominant (FC) or colour being dominant (CF). Colour can be perceived all by itself also (C). These are some of the commonly scored categories of determinant developed by Rorschach and various other systematizers.

Determinants reveal a lot of information about a wide variety of personality characteristics. They tell us about one's emotionality, imagination, fantasy life, an individual's ability to indulge in creative and conceptual thinking, about ego strength, conflicts, opposition tendency, etc.

- iii. The treatment of content varies from one scoring system to another, some emphasizing it heavily whereas others ignoring it altogether. Some of the commonly scored content categories are as follows: i) Human (H), ii) Animal (A), iii) Human detail (Hd), iv) Animal detail (Ad), v) Clouds (CI), vi) Botany (Bt), vii) Blood (BI), viii) Sexual materials (Sx), ix) X-Rays, x) Anatomy (At), etc.

Content categories help us to discover many important aspects of an individual's behaviour. Content analysis is a source of ascertaining the subject's personal meaning, attitudes, interest and even complexes. Content responses are supposed to possess psychiatric and psychoanalytic interpretations often pointing towards pathological tendencies present in the subject.

- iv. Original and popular scoring category tells us whether subject's responses are common or original. A popularity score is often found on the basis of the relative frequency of different responses among people in general. However, there is often difference of opinions - among specialists as to which responses should be regarded as original and which as popular. Popular responses help us to know about an individual's interpersonal behaviour. It also tells us whether an individual is socially confirming or conventional in his approach or not.

6.3.2 Evaluation of the Rorschach Test:

Rorschach test has received mixed reception, some have regarded it as an X-ray of personality, an indispensable tool for diagnostic purposes, whereas, others have regarded its use as unethical. Many have judged its worth with contempt and have advocated its abandonment because it has,

proved baffling to the researcher and irritating to those with strong allegiance to stringent measurement theory.

1950s were a decade of great controversy of Rorschach and even today this controversy is fresh. In 1956, Cronbach remarked "the test has repeatedly failed as predictor of practical criteria" and in 1958, Jensen made a similar statement by pointing out that "the Rorschach has been worthless as a research instrument. The Rorschach methodology has nothing to show for its application in the personality field." Contrary to this, Sundberg (1961) noted that "Rorschach techniques is the most commonly used assessment instrument in clinical psychology". In 1969, Exner offered the opinion that "the Rorschach despite its inherent weakness and the strain of academic disparagement is providing to be surprisingly handy. It not only survives perennially but is still a mainstay of psycho-diagnostic methodology." However, it is strange that Exner (1974) recently remarked that "the survival of Rorschach is no longer as certain as once was the case clearly, it is no longer synonymous with clinical psychology." Nowhere has there been the discrepancy and controversy between the researchers and clinicians so great and bitter as it has been in the area of Rorschach. Researchers have consistently presented a poor picture, whereas on the other hand, clinicians have been using this test with increasing frequency.

The nature of the Rorschach test makes it an important tool, which can be used with any age group, any cultural group and which does not make any demand on the part of the subject. It can, very easily, be used with illiterate and culturally backward groups.

6.3.3 Pictures as Projective Stimuli:

Wide varieties of pictures have been used as projective stimuli ranging from real people to animals, objects or anything such as paintings, drawings etc. The use of pictures in psychometric assessment dates back to 1907 when Brittain (1907) reported sex differences in response to pictures. Similarly, Libby (1908) as well as Schwartz (1932) developed the projective tests using pictures. The most famous test is the Thematic Apperception Test (TAT) developed in 1935 at Harvard Psychological Clinic by Christiana D Morgan and Henry Murray.

The only other projective technique that has approached the Rorschach method in amount of use and volume of research is the TAT which was introduced by C.D. Morgan and Henry A. Murray as a method to explore the unconscious thoughts and fantasies.

Murray found that TAT enabled the trained examiner or interpreters to reconstruct on the basis of subject's stories, his dominant drives, emotions, sentiments, complexes and conflicts. Although the TAT was at first slow in gaining wide acceptance, it is now a test that approximates the Rorschach in popularity and in amount of research it has stimulated.

The TAT was originally designed as an aid to eliciting fantasy material from patients in psychoanalysis. At first, it gained popularity only among clinical psychologists, but gradually, it became a research tool in development,

social and personality psychology and in the cross-cultural studies in Anthropology. It is also used for personality assessment in the fields of counselling and industrial psychology.

In contrast to the inkblot techniques, the Thematic Apperception Test (TAT) presents more highly structured stimuli and requires more complex and meaningfully organized verbal responses. Interpretations of the responses by the examiners is usually based on content analysis of a rather qualitative nature. Murray's system for scoring TAT is largely quantitative whereas McClelland and Eron has developed a highly quantitative system for scoring and interpreting TAT responses.

Administration of TAT:

The third revision of TAT consists of 30 pictures and blank card. The pictures have been selected and marked in such a way that there are four sets of 20 cards each, one for boys, one for girls, one for males and one for females, over 14 years. The testing process is divided into two sessions and for each of these, it is suggested that no more than 10 TAT cards be administered with at least one day intervening between the two sessions. More recently, practical considerations have led to reduction in the number of cards administered. Most testers now present the subject with 8 to 12 cards and use only a single session. The cards are presented individually and the respondent is instructed to provide a story about the picture that described the depicted scene, what led up to it, what the characters in the picture are thinking and what the outcome will be.

Although typically administered as an oral test in clinical situation, the TAT may also be administered in writing and as a group test.

Scoring of the TAT:

Like Rorschach, the TAT also has multiple scoring system. Murray's lack of detailed scoring instructions in his manual and the relative ease with which the TAT can be administered have been cited as factors contributing to the multiplicity of the scoring system (Murstein, 1963). Furthermore, the * non- technical nature of the TAT and the simple verbal contents of stories have encouraged clinicians to invent their own system of analysis. Some important TAT scoring systems are as follows:

- a. Murray's scoring system (non-quantitative)
- b. McClelland's system (quantitative) and
- c. Eron's system (quantitative).

Here we will examine Murray's scoring system in brief. Although amenable to qualifications, Murray's recommended system of analysis is highly content- oriented and relies heavily on the qualitative characteristics of the stories. Murray emphasized three important concepts called as Need (determinants of behaviour arising from within the individual), Press (determinants of behaviour arising from within the environment) and themes (a unit of interaction between needs and press). The following points are noted in the analysis of the story.

1. **The Hero :** The first step in the analysis of the story is to distinguish the hero or the character with whom the subject seemed to have identified in principle. This would be the character in whom the story teller is most interested and the individual who mostly resembles him. The tester must be aware of the fact that the hero of the story may be no hero in the story. The interpreter should direct his attention to the following aspects of the hero's personality. His intelligence, achievement ability, conflicts, leadership qualities, feelings, etc.
2. **Needs of the Hero:** After the identification of the hero, the interpreter must formulate the reactions of the hero to various forces. These formulations are usually influenced by the theoretical orientations of the test interpreter. However, Murray recommends that this may be accomplished within a classification of the needs of the hero. The needs can be either primary or secondary.
3. **Environmental Forces:** These are categories according to their effect on the hero. Murray's system consists of a comprehensive list of environmental forces or presses. These presses could be real or imaginary and include aggression in which the hero's property and/or possessions are destroyed. Dominance, where the hero is exposed to commands, order or forceful arguments and rejection in which persons reject, repudiate are indifferent or leave the hero.
4. **Outcomes:** Outcome refers to the results of the story. It refers to the relative strengths of the forces emitting from the hero and the strengths of these. The amount of frustration and hardships experienced and the relative degree of success and failure of the hero must be assessed.

Themes or Themas: Themas(a unit of interaction between needs and press) refer to the interplay within the story of the hero's needs, presses (determinants of behavior arising from within the environment), and successful or unsuccessful resolutions of his conflicts. Themes represent need- press combination, it can be simple or complex.

Interests, Sentiments and Relationship: These are the last category to be scored. In this, a note is made of the various interests, sentiments and inter-personal relationship as expressed in the stories by the subject.

Research studies have suggested that many situational factors such as who the examiner is, how the test is administered, test taker's experiences prior to the test and during the test administration process considerably influences as to how the test taker responds to the test. Test takers responses are also influenced by transient internal needs such as hunger, thirst, fatigue, and higher than ordinary levels of sexual tension. This test is considerably used in many competitive examinations including armed forces as well as UPSC (Union Public Service Commission) examinations for selecting candidates for employment in armed forces and government jobs.

Other tests using pictures as projective stimuli: Many different varieties of tests have been developed using pictures as projective stimuli. Some of the well known tests are as follows:

- a. **Hand Test:** This test was developed by Wagner (1983) and consists of 09 cards with pictures of hand on them and the tenth card is a blank card. The test taker is asked what the hands on each card might be doing. When presented with the blank card, the test taker is instructed to imagine a pair of hands on the card and then describe what they might be doing. Test takers may make several responses to each card and all responses are recorded. Responses on this test are interpreted according to 24 categories such as aggression, dependence and affection.
- b. **Rosenzweig Picture-Frustration Study:** This test was first developed in 1945. It employs cartoons depicting frustrating situations. The test taker's task is to fill in the response of the cartoon figure being frustrated. The Rosenzweig Picture-Frustration Study (P-FS), is a semi-projective technique that has been widely used to assess patterns of aggressive responding to everyday stress. When shown a picture, the client provides a reply for the anonymous frustration person depicted. The instrument contains 24 cartoon-like pictures, each depicting 2 people in a mildly frustrating situation that commonly occurs. Three versions of this instrument exist: Child, Adolescent and Adult.

Some other versions of picture tests include:

- Children's Apperception Test (CAT)
- Thompson Modification of TAT (T-TAT)
- The Blacky Picture
- Make Picture Story (MAPS)
- Michigan Picture Test (MPT)

Word as Projective Stimuli: There are following two types of such tests, each of which We will discuss briefly:

- a. **Word Association Test** can be defined as a semi structured, individually administered, projective technique of personality assessment that entails the presentation of a list of stimulus words, to each of which an assessee responds verbally or in writing with whatever comes to mind first upon hearing the word. Responses are then analyzed on the basis of content and other variables.

This test was originally known as the free association test and was first systematically described by Francis Galton, half cousin of Charles Darwin, in 1879, Wilhelm Wundt, the father of experimental psychology, subsequently introduced it into psychoanalytic researchers. Carl G. Jung (1910) used it as a psychiatric screening instrument by providing objective scoring and statistical norms.

Forensic psychology made use of it as a 'lie detector'. It was Jung (1910) who pointed out how word association test can be used as a lie detector. Burt in 1931 and Lindsley in 1955 carried out extensive research demonstrating its utility and reliability as a 'lie detector'

There have been many versions of the word association test. The three most important are as follows:

- i. Jung (1910) used a list of 100 words to represent 'common emotional complexes'. The subject is told that the examiner will speak a series of words, one at a time, after each word, the subject is to reply as quickly as possible with the first word that comes to mind. There are no right or wrong responses. The examiner records the reply to each stimulus word, the reaction time and any unusual speech or behavioral manifestations accompanying a given response. Analysis of the content of responses, reaction times and other aspects of overt behaviour aid in discovering certain emotional problems and in drawing certain inferences which are then used for further psychological exploration by interview.
- ii. In 1968, Rapaport and his associates at the Menninger clinic developed another version of word association test which was very similar to Jung's word association test. The test developed by Rapaport and his associates consists of 60 words list. The approach used by them reflects strong psychoanalytic orientation and many of the words are associated with psychosexual conflicts. This test aids us in two ways :
 - One in detecting the impairment of thought process and
 - Second to suggest areas of significant internal conflicts.Analysis of the results is done on the basis of popular responses, reaction time, associative disturbances and impaired reproduction on retest.
- iii. Still another version of the word association test was developed by Kent- Rosanoff called as the Kent-Rosanoff Free Association Test to differentiate between mentally ill and the normal. It consists of 100 stimulus words. They have provided a percentage in each category that was expected to differentiate the normal from abnormal. However, the diagnostic use of this test declined with the gradual realization that response frequency varies with age, socio- economic status, educational level, regional and cultural background, creativity, etc. Hence, proper interpretation of scores requires norms of many subgroups as well as supplementary information about the examinee, which no doubt is a very difficult task.

- b. **Sentence Completion Test** is another verbal projective technique that has been extremely used in research and clinical practice. A wide

variety of sentence completion test are presently available. The content of a particular test and the nature of the sentences will depend upon the group of persons and the purpose for which they are intended.

In this test, any individual is presented with a series of incomplete sentences, generally open at the end, to be completed by him in one or more words. They resemble the word association test. However, sentence completion test is regarded as superior to word association test because the subject may respond with more than one-word, greater flexibility and variety of responses are possible and more areas of personality and experiences may be tapped.

Some of the most commonly used completion tests are as follows:

- i. Sack's Sentence Completion, which is commonly used in clinical practice, consists of 60 sentences stems which can tell us about an individual's adjustment in family area, sex area, about his interpersonal relationship, self-concept and goals. Sample items are as follows :

- Some day, I ...
- My sex life ...
- If I were in charge ... etc.

The subject has to complete these sentences by writing the first word or few words that comes to mind.

- ii. Another important and widely used such test was developed by Rotter who called it as **Incomplete Sentence Blank** which consists of 40 sentences stems and is similar to Sack's completion test except that it is scored more rigidly and precisely.
- iii. A novel approach to sentence completion technique is to be seen in **Washington University Sentence Completion Test (WUSCT)** which is largely based on Loevinger's broadly defined construct of ego development. This test classified responses with reference to a seven stage scale of ego development as follows: pre-social and symbiotic, impulsive, self-protective, conformist, conscientious, autonomous and integrated. Although most of the research with this test has been done with adult women, forms are also available for use with men and with younger persons of either sex.

In general, a sentence completion test may be useful for obtaining diverse information about an individual's interests, educational aspirations, future goals, fears, conflicts, needs, etc. The sentence completion tests have high degree of face validity. However, they are vulnerable for "faking good" or "faking bad" on the part of the examinee.

Sound as Projective Stimuli: Though not much popular, it was the behaviorist B.F. Skinner who developed this test. Skinner created a series of recorded sounds to which people were told to respond. Saul Rosenweig as well as David Shakow also did some pioneering work with audition as a projective techniques. Behaviorist B. F. Skinner is not typically associated with the fields of personality assessment or projective testing. However, early in his career Skinner developed an instrument he named the verbal summator, which, at one point, he referred to as a device for snaring out complexes," much like an auditory analogue of Rorschach inkblots. Skinner's interest in the projective potential of his technique was relatively short lived, but whereas he used the verbal summator to generate experimental data for his theory of verbal behavior, several other clinicians and researchers exploited this potential and adapted the verbal summator technique for both research and applied purposes. The idea of an auditory inkblot struck many as a useful innovation, and the verbal summator spawned the tautophone test, the auditory apperception test, and the Azzageddi test, among others.

The Production of Figure Drawings: Analysis of drawings is another projective method. Drawings, especially in the case of young children provides a wealth of diagnostic information about the clinical aspects of individual's personality functioning. Today, the use of drawings in clinical and research settings have extended beyond the area of personality assessment. Attempts have been made to use artistic productions as a source of information about individual's intelligence, neurological impairment, visual motor coordination, cognitive development and learning disabilities.

Figure drawing tests are projective methods of personality assessment that entails the production of a drawing by the assessee, which is analyzed on the basis of its content and related variables.

Karen Machover did classic work on figure drawing tests. One of the well-known figure drawing test is the one developed by Karen Machover called as Draw-A-Person Test. In this test, the examinee is provided with a paper and pencil and is told simply to 'draw a person'. Upon completion of the first drawing, he or she is asked to draw a person of the opposite sex from that of the first figure.

While the individual draws, the examiner notes his or her comments, the sequence in which different parts are drawn, length of time required to complete the picture, placement of the figures, the size of the figure, pencil pressure used, symmetry, line quality, shading, the presence of erasures, facial expressions, posture, clothing, and overall appearance and other procedural details. The drawing is normally followed by an inquiry in which the examinee is asked to make up a story about each person drawn. A series of questions is also asked to the subject about the person drawn. Emanuel Hammer (1958, 1981, 2014) believed that people project their self-image or self-concept in figure drawings, as well as in other ways (such as in disguised form in dreams and paintings)

Scoring of this test is essentially quantitative and is lacking in validation studies. Clinician and researchers now opine that the draw a person test can serve best not as a psychometric test but as the clinical instrument, in which, the drawings are interpreted in the context of other information about the individual.

Another well-known drawing test is the House-Tree-Person test. Still another test that can help the examiner to understand and examinee in relation to his/her family is the Kinetic Family Drawing (KFD). In this test a child is given a paper 8 1/2 by 11 inch sheet, a pencil and an eraser and is told to "draw a picture of everyone in your family, including you, doing something."

Emphasis is laid on the interaction between various family members.

Figure drawing tests are clinically highly useful, especially while dealing with children. However, these tests have its own disadvantages and reliability and validity data on these tests is lacking.

6.4 PROJECTIVE METHODS IN PERSPECTIVE

Many psychologists question the usefulness of projective methods. They are especially critical of scoring system, the assumptions on which their use is based, the situational variables that influence their use, and their reliability and validity.

Some of the assumptions on which projective methods are based are that every response can be used for personality analysis, there is a link between the strength of need and its manifestation in projective tests, testtakers do not know what they are disclosing about themselves, the more ambiguous the stimuli, the more subjects reveal about their personality, environmental variables, response sets, reactions to the examiner, and related factors all contribute to response patterns, etc.

Murstein dismissed all these assumptions as "cherished beliefs" that have been accepted without validation. It was pointed out that similarities in the response themes of different testtakers to the same stimuli indicates that the stimulus material may not be as ambiguous and amenable to projection as previously assumed.

Projective tests are based on the assumption that projection is greater onto stimulus material that is similar to the subject (in physical appearance, gender, occupation, and so on). There is no support for this assumption too.

Many interpretive systems for the Rorschach and other projective instruments are based on psychodynamic theory, which itself has been criticized.

Situational variables

Research has shown that situational variables such as the examiner's presence or absence have significantly affected the responses of

experimental subjects. For example, Bernstein (1956) showed that TAT stories written in private are likely to be less guarded, less optimistic, and more affectively involved than those written in the presence of the examiner. Similarly, the age of the examiner, the specific instructions given and other subtle reinforcement cues provided by the examiner are likely to affect projective protocols.

Masling (1960) noted that subjects use every available cue in the testing situation, including cues related to the actions or the appearance of the examiner and even examiners also use situational cues such as their own needs and expectations, their own subjective feelings about the person being tested, and their own constructions regarding the total test situation. It has been observed that even in situations involving objective (not projective) tests or simple history taking, the effect of the clinician's training (Chapman & Chapman, 1967; Fitzgibbons & Shearn, 1972) and role perspective (Snyder et al., 1976) as well as the patient's social class (Hollingshead & Redlich, 1958; Lee, 1968; Routh & King, 1972) and motivation to manage a desired impression (Edwards & Walsh, 1964; Wilcox & Krasnoff, 1967) are capable of influencing ratings of pathology (Langer & Abelson, 1974) and related conclusions (Batson, 1975). These and other variables are given wider latitude in the projective test situation, where the examiner may be at liberty to choose not only the test and extra-test data on which interpretation will be focused but also the scoring system that will be used to arrive at that interpretation.

Psychometric considerations

Critics of projective techniques point out that uncontrolled variations in protocol length, inappropriate subject samples, inadequate control groups, and poor external criteria as factors contributing to spuriously increased ratings of validity. There are methodological obstacles in researching projectives because many test-retest or split-half methods are inappropriate. It is very difficult to conduct validity studies that effectively rule out, limit, or statistically take into account all of the unique situational variables that attend the administration of such tests.

Objective Tests vs. Projective Tests

Academics differentiating between objective and projective tests criticize objective tests by saying that scores on objective tests get influenced by response styles, malingering and other sources of test bias. Moreover, they point out that the testtakers may lack sufficient insight or perspective to respond "objectively" to objective test items. Meehl (1945) was of the opinion that so-called objective test items may actually serve as projective stimuli for some testtakers.

On the other hand, looking at the doubtfulness of assumptions on which the projective tests are based, projective tests also may not be as projective as they were once thought to be. Weiner (2005) pointed out that many projective tests feature scoring systems that has "objective" coding. So, having a dichotomy of objective and projective tests is misleading. Weiner (2005) in fact suggested substituting the terms *structured*, in place of

objective, and *unstructured*, in place of projective. The more structured a test is, the more likely it is to tap relatively conscious aspects of personality. By contrast, unstructured or ambiguous tests are more likely to access material beyond immediate, conscious awareness (Stone & Dellis, 1960; Weiner & Kuehnle, 1998).

6.5 SUMMARY

In this unit we discussed various objective methods of personality assessment. These methods mostly use the paper and pencil tests. We also discussed various projective methods of personality assessment at length. Some of the most commonly used projective methods include the Rorschach Inkblot test, The Thematic Apperception Test, Hand Test, Rosenzweig Picture-Frustration Study, Word Association Test, Sentence Completion Tests, Sound as Projective Stimuli and Figure Drawings which includes the Draw-A-Person Test and House-Tree-Person test.

6.6 QUESTIONS

Write notes on:

1. Objective Methods of Personality Assessment

2. Sound as Projective Stimuli

3. Discuss Inkblot and Pictures as Projective Stimuli

4. Write a note on Word Association Test and Sentence Completion Tests

5. Discuss Analysis of Drawings as a projective method

6.7 REFERENCES

Cohen, JR., &Swerdlik, M.E. (2010). Psychological Testing and Assessment: An introduction to Tests and Measurement. (7 th ed.). New York. McGraw-Hill International edition.

Anastasi, A. & Urbina, S. (1997). Psychological Testing. (7th ed.). Pearson Education, Indian reprint 2002.

MEASURES OF VARIABILITY, PERCENTILES AND PERCENTILE RANKS - I

Unit Structure :

- 7.0 Objectives
- 7.1 Introduction
- 7.2 Range and Average Deviation (AD).
- 7.3 Quartile Deviation (QD) and Standard Deviation (SD)
- 7.4 Calculation of the four measures of variability.
- 7.5 Comparison of the 4 measures of variability: Merits, Limitations, Uses of measures of variability
 - 7.5.1 Merits Limitations and uses of Range
 - 7.5.2 Merits, Limitations, uses of Average Deviation Merits
 - 7.5.3 Merits, Limitations and Uses of Quartile Deviation
- 7.6 Summary
- 7.7 Questions
- 7.8 References

7.0 OBJECTIVES

- To understand and to impart knowledge about various measures of variability or dispersion, their uses, limitations and statistical approaches.
- To create an awareness among students and to know the various measures or technique to calculate the measures of variability.

7.1 INTRODUCTION

The concept of variability or dispersion is fundamental in statistics that shows how far the number of the series scattered or spread on either side of central tendency, so it measures the quantities of the distributions. It also indicates relative measures of deviation of a group.

In Unit 8, we will deal with the best method or measure of variability, and comparison of the four measures of variability, along with the concept of percentile, its merits, limitation and uses, along with the methods of calculations of percentiles and percentile rank.

7.2 RANGE AND AVERAGE DEVIATION (AD)

Out of 4 measures of variability, Range and Average Deviation are two measures that are not so important as quartile deviation (QD) and Standard deviation (SD).

Range is the simplest measure of variability. It is based on the value of the two extreme scores, an extreme score can alter the value of the Range.

It is defined as "the interval between the highest and lowest scores" i.e.,
Range = Highest - Lowest score.

For example, if the highest score in a raw score of 50 is 90 and the lowest score is 25, the range will be= 90-25=65

(R = Highest score – Lowest score, i.e., 90-25=65 R)

Whereas, Average Deviation (AD) is concerned, it is rarely used but AD provides a solid foundation for understanding the conceptual basis of another more widely used measure, the SD.

It is the mean of the deviation of all separate scores in a series taken from the mean.

$$AD = \frac{\sum |x|}{N}$$

The formula of AD is:

Σ = Sigma, sum total of

χ = deviation from mean (Raw score minus mean)

N = total Number of scores.

7.3 QUARTILE DEVIATION (QD) AND STANDARD DEVIATION (SD)

Quartile Deviation (QD) is another measure of variability. It is one half of the scale distance between the 75th and 25th percentile in a frequency distribution. The 25th percentile is Q1 and 75th percentile is called Q3, the first and the third quartiles respectively. The formula of QD is:

$$QD = \frac{Q3 - Q1}{2}$$

In a perfectly symmetrical distribution, Q1 and Q2 is exactly the same distance from the median (Q_2).

On the other hand, standard deviation is one of the best methods or measures of variability. It is very useful measure of variation.

It is defined as "the square root of the arithmetic mean of the square of all deviations taken from mean."

$$SD = \sqrt{\frac{\sum fx^2}{N}}$$

The formulas of SD are:

- i. Long method:
 Σ Sum total of
 f frequencies in a frequency distribution
 x_2 = deviation from the mean
 X^2 = deviation of mean square
 N = total number of frequencies
 = under root

ii.

$$SD = \sqrt{\frac{\sum fx^2}{N} - c^2}$$

Here c^2 = is the squared correction of fx^2/N other calculations are as equal like long method.

7.4 CALCULATION OF THE 4 MEASURES OF VARIABILITY

1. Range = Highest - Lowest score. (R = H - L) example, 100 - 25 = 75
2. Average Deviation (AD) Formula
3. $AD = \frac{\sum X - xN}{N}$
4. Whereas x is a deviation from mean score, i.e., $X - M = x$ (small letter of x)

Example: Find out AD from the following scores: 6, 8, 10, 12, 14.

Step - 1 find out the mean of these 5 scores by $X/N = \frac{6+8+10+12+14}{5}$

$$= \frac{50}{5} = 10$$

Therefore,

Mean is = 10

Step - 2 find out x. Deviation of all the scores from the mean score, 10. will be, 6-10 or -4 ; 8-10 or -2; 10-10 or 00; 12-10 or 02 and 14-10 or 4. The sum of these 5 deviations, disregarding signs, is 12; and dividing 12 by 5 (N) will be 2.4 or AD.

We can put these five scores as under and thus can find out AD:

Scores	X
	-4
8	-2
10	0
12	2
14	4
50	12
	Disregarding sign X
N = 5	

$$\text{Mean} = \sum X/N = \frac{50}{5} = 10$$

$$\text{So, AD is } (\sum |X|)/N = \frac{12}{5} = 2.4$$

7.4.1 Quartile Deviation or QD:

To calculate QD, the formula is:

$$\text{QD} = \frac{Q3 - Q1}{2}$$

However, to find QD, we must first calculate Q3 and Q1.

To find Q3, the formula is:

$$Q3 = I + [(3N/4 - F)/f_m] \times ci$$

And the formula of Q1 is:

$$Q1 = I + [(N/4 - F)/f_m] \times ci$$

Calculating Q1 from the following frequency distribution by short-method (Table)

CI	Frequencies
136-139	3
132-135	5
128-131	16
124-127	23
120-123	52
116-119	49
112-115	27
114-111	14
104-107	7
	N = 200

$$Q1 = I + \frac{(N/4 - F)}{Fm} \times ci$$

I = the exact lower limit of the interval in which the quartile falls, i.e., 111.5

i = the length of interval = 4

F= the sum total of all the frequencies below which Q1 falls. in this case 25

fm= the frequency on which the Q1 falls, i.e., 27

N/4= Total N is divided by 4 = 50 (200/4).

Now applying this formula in frequency distribution Table No. 1, we get the following:

$$Q1 = 111.5 + \frac{(50-25)}{27} \times 4$$

$$Q1 = 111.5 + \frac{(25)}{27} \times 4$$

$$Q1 = 111.5 + 0.92 \times 4$$

$$Q1 = 111.5 + 3.68$$

$$Q1 = 115.18$$

$$\mathbf{Q1 = 115.18}$$

$$Q3 = I + \frac{(3N/4 - F)}{fm} \times ci$$

where,

I = the exact lower limit of the interval in which the Q3 falls, i.e., 119.5

i = the length of class interval, 4

F= the sum total of all the frequencies below which Q3 falls, 101

fm= the exact frequency on which the Q3 falls, 52

3N/4= 150 as N/4=50, so 3N/4=3x50=150

Thus Q3:

$$Q3 = 119.5 + \frac{(150-101)}{52} \times 4$$

$$Q3 = 119.5 + \frac{(49)}{52} \times 4$$

$$Q3 = 119.5 + 0.94 \times 4$$

$$Q3 = 119.5 + 3.76$$

$$Q3 = 123.26$$

$$Q3 = 123.26$$

$$Q3 = \frac{Q3 - Q1}{2}$$

$$Q3 = \frac{123.26 - 115.18}{2}$$

$$Q3 = \frac{8.08}{2} = 4.04$$

$$Q3 = 4.04$$

7.5 COMPARISON OF THE FOUR MEASURES OF VARIABILITY: MERITS, LIMITATIONS, USES OF MEASURES OF VARIABILITY

7.5.1 Merits, Limitations and uses of Range

Merits of Range

1. It is simplest measure of variability.
2. It provides gross description of the spread of scores.
3. It is most general measure of spread or scatter.

Limitations of Range

1. It is less used in other measures of statistics and maths.
2. Its values are always changed.
3. Except frequency distribution it is not used in other distribution, therefore its use is limited.
4. It is also not used in algebraic treatment.

Uses of Range

1. It is used when the data are too scattered.
2. It is used when the knowledge of extreme score is wanted.
3. It is also used when the median is the measure of central tendency.
4. It is used when the scores are likely to affect SD.
5. It is used when the 50% of score is of primary interest.

7.5.2 Merits, Limitations, uses of Average Deviation (AD) Merits

Merits

1. It is another measure of variability in which all signs of deviation (+ or -) for calculation is treated as positive.
2. Other merit of this measure is to weigh all deviations from the mean according to their size.

Limitations of AD

1. This method or measure is rarely used because of algebraic signs.
2. It is useless measure for further operations. Uses of AD

Uses

1. It is used in Psychology, Economics and Statistics.
2. It is very accurate measure.
3. It is not affected by extreme values of the distribution.
4. For all practical purposes, AD is replaced by SD.

7.5.3 Merits, Limitations and Uses of Quartile Deviation Merits:

1. Q is a very important measure of variability. This method is applied when 75th and 25th percentiles are required. In a normal distribution Q is called the probable error or (PE) is used interchangeably.
2. It is better than range as it is not affected by extreme values.

Limitations

1. This measure does not calculate Q2, 50% percentile.
2. This method is only used in Psychology and Statistics.
3. This measure is not used for further operations.

Uses

1. It is very easy measure to calculate the percentiles.
2. It is reliable measure because it is not affected by extreme values or scores.
3. It is used when the median is the measure of central tendency.
4. It is also used when extreme scores would influence the SID disproportionately.

7.6 SUMMARY

In this chapter, we have discussed the nature, methods of variability along with its four important measures such as Range, Average deviation, Quartile deviation and Standard deviation. We have highlighted the merits, uses and limitations of these four variables with separate headings. We have also provided the measures or techniques for calculation of these four measures of variability. Mention is also made about the comparison of the 4 measures. We have also answered the question "which is the best measure of variability", but in brief. We discussed the merits, uses and limitations of the three of these important measures of variability or dispersion, that is, range, average deviation and quartile deviation.

7.7 QUESTIONS

Q.1 Explain the uses and limitations of QD.

Q.2 (a) Calculate QD for the following distribution.

(b) Calculate SD for the following distribution.

Scores	F
50-54	3
45-49	6
40-44	9
35-39	15
30-34	19
25-29	12
20-24	14
15-19	6
10-14	2
	N = 140

7.8 REFERENCES

Annapornaa R. et al (20014) - A Handbook of Mathematics and Statistics, Chetana Publications Pvt. LTD. 262, Khatauwadi, Girgaon, Mumbai-400004

Anastasi, A. and Urbina, S. (1997) -Psychological Testing (7th ed) Pearson Education, India Reprint -2002

Cohen, J.R. and Swerdik, M.E. (2010) Psychological Testing and Assessments, An Introduction to tests and Measurements (7th ed) New York McGraw-Hill International Edition.

Garrett, Henry, E. (1973) - Statistics in Psychology and Education, Vakills, Feffer and Simon Pvt. Ltd. Ballard Estate, Mumbai- 400001

MEASURES OF VARIABILITY, PERCENTILES AND PERCENTILE RANKS - II

Unit Structure :

- 8.0 Objectives
- 8.1 Introduction
- 8.2 Comparison of measures of variability: Merits, Limitations, Uses of measures of variability
 - 8.2.1 Merits, Limitations, Uses of SD
 - 8.2.2 Comparison of the Four Measures of Variability
- 8.3 Percentiles - nature, merits, limitations and uses
- 8.4 Calculation of Percentile ranks and Percentile Scores
- 8.5 Summary
- 8.6 Questions
- 8.7 References

8.0 OBJECTIVES

- To understand and to impart knowledge about various measures of variability or dispersion, their uses, limitations and statistical approaches.
- To create an awareness among students and to know the various measures or technique to calculate the measures of variability.
- To provide basic knowledge about statistical procedures for calculating percentiles and percentile ranks.

8.1 INTRODUCTION

In the last unit, we studied the concept of variability or dispersion, which is fundamental in statistics. Variability shows how far the number of the series scatter or spread on either side of central tendency, as well as indicates relative measures of deviation of a group.

Now, in this unit, we will discuss the various measures of variability, their merits, limitation, uses and methods of calculations. We will also deal with the best method or measure of variability. Comparison of the four measures of variability will be also explained in detail. And finally, we shall focus on the concept of percentile, its merits, limitation and uses, along with the methods of calculations of percentiles and percentile rank.

8.2 COMPARISON OF THE FOUR MEASURES OF VARIABILITY: MERITS, LIMITATIONS, USES OF MEASURES OF VARIABILITY

In last unit, we learned about the merits, limitations and uses of the first two measures of variability, that is, range, and average deviation. In this unit, we will discuss the merits, limitations and uses of other two remaining measures, that is, quartile deviation and standard deviation. We will also compare all four measures of variability to know the best among them.

8.2.1 Merits, Limitations, Uses of SD Merits:

1. SD as a matter of fact is the best method of variability because of its reliability and accuracy.
2. This method is used in psychological research, and is also applied to estimate the population, significant differences between means computing coefficient of correlation.
3. It is very good measure of regression equation.

Limitations:

1. It is very complex method.
2. It gives more weightage to extreme values.
3. It is useless for measures of central tendency.
4. Its operations are limited only for psychological research and statistics.

Uses

- a. It is widely used in Psychology, Economics and Statistics.
- b. It is more reliable than other measures of dispersion.
- c. It is also used in the interpretation of curve.
- d. It is used when coefficient of correlation is computed.
- e. It is used for estimating the population mean.

With regard to reliability of SD, an important question is often asked as which is the best measure of variability? why? To answer this question, we can simply say that SD is the best method of variability. Why? Because of the following reasons (focusing on other three measures):

1. Range and AD are rarely used because of their limitations and further operations.
2. QD is used when 75th percentiles and 25th percentiles are required.

But SD is used for its wider scope and importance. It is equally important for its reliability and accuracy.

Keeping Psychological research aside, SD is used in further operations such as estimating the population mean, calculation of significant difference between two means, computing coefficient of correlation, without SD no Y can be calculated, and the like. Therefore, we can reasonably say that SD is the best measure of variability.

Check your progress

Q.1 Define/Explain the Range, AD, Q and SD.

Q.2 Explain the uses of Range and Q.

Q.3 Describe the limitations of AD and SD.

Q.4 Calculate Q and SD from the frequency distribution given below Q5. Find out AD from the following scores

6, 14, 13, 15, 17, 20 = (N=6)

8.2.2 Comparison of the Four Measures of Variability

It is hard to compare these four measures of variability because Range and AD are very simple measures of variability and rarely used in Psychological

research and Statistical treatment. Due to their limitations, they are useless for further operations.

Quartile Deviation (Q), on the other hand, are useful measures of variability. Q is used when 75th and 25th percentiles are required but SD is used in variety of ways such as in calculating coefficient of correlation, estimating significant difference between means, regression equations and so on. However, we shall compare these four measures of variability as under (Table 8.1).

	Range	QD	AD	SD
1. Definition	The interval between the highest and the lowest scores.	It is one of the scale distance between the 75 th and 25 th Percentiles.	It is the mean of the deviation of all separate scores in a series taken from the mean.	The square root of the arithmetic mean of the square of all the deviation taken from mean.
2. Merits	It provides gross description of the spread scores. It is most general measure of scatter.	It is applied to find 75% and 25% of score or value.	In this measure, all signs of deviations (+ -) for computing AD is treated as positive.	This is widely used measures of variability which is adequate and reliable and trustworthy.
3. Uses	It is used when knowledge of extreme scores are wanted and when median is measure of the central tendency.	It is easy measure and reliable because it is not affected by extreme values or scores.	It is very accurate measure and is used in Economics, Psychology, and Statistics.	It is the best measure of variability. It is widely used in estimating correlation, population means and regression equation.
4. Demerits	Its use limited. It is used only in frequency distribution. Its values are changed.	It is always used in Psychology and Statistics, but does not calculate 50% of percentiles.	It is rarely used because of signs of Algebra.	It is very complex measure which gives weightage to extreme values.
5. Formulas	Difference between highest score and the lowest score.	$\frac{Q_3 - Q_1}{2}$	$(\sum X) / N$	$SD = \sqrt{\frac{\sum fx^2 - C^2}{N}}$

	Range	QD	AD	SD
6. Applications	For calculating and preparing a frequency table, used in Psychology, Economics, and Statistics.	For calculating measures of variability used in Psychology and Statistics.	For calculating mean and deviation.	For calculating correlation, coefficient of correlation, regression equation, estimating population mean, etc. It is also used in Psychology, Economics and Statistics.

Check your progress

Q1 Compare Range with AD.

Q2 Define Q and SD and make a comparison between these two measures of variability.

Q3 What are the applications of Range, AD, Q and SD.

8.3 PERCENTILES - NATURE, MERITS, LIMITATIONS AND USES

Nature /Definition of Percentile

If we remember the calculation of median (50%) of score, we can be immediately acquainted with Percentile or percentage. To calculate median, we get 50% of measure or scores. Similarly, Q3 gives us 75th percentile and Q1 gives us 25th percentile.

But percentile gives us any percentage, may be 10%, 20%, 30% and even 90% of scores. However, points below which lie, 10%, 45% and 85% or any percent of scores is called Percentiles. .

However, a percentile may be defined as an expression of the percentage of people whose scores on a test or measure fall below a particular raw score. Thus, a percentile (P) is a converted score that refers to a percent- age of test takers.

Merits of Percentiles

1. The main advantage of percentile is to determine at which 10% or 43% of individual's scores or cases is located.
2. Percentage are based upon the number of scores or cases falling within a certain Range.
3. The distance between any two percentiles show a certain area or number of cases ($N/10$, $N/20$).

Limitations of Percentiles

1. When the number of scores in a distribution is small, percentiles are not used.
2. When there is little or no significance in making distinctions in rank, percentiles are not used.
3. There is a restriction in using percentiles that real differences between raw scores may be minimized but near the end of distribution it increases, and the errors may be even worse with highly skewed data.
4. Except in calculation Percentile Point, they are used for further operations.
5. Percentiles have limited scope in their application.

Uses of Percentiles

1. The percentile technique is very easy in calculation.
2. The percentile has another advantage of being easily understood.
3. The percentile technique does not make any assumption with regard to the characteristics of the total distribution.
4. This technique also answers the question: "Where does an individual's scores rank him in his group"? or, "Where does an individual's scores rank him in another group whose members have taken the same test?"

5. The differences in scores between any two percentile points become greater as we move from the median (P50) toward the extremes.

Check your progress:

Q.1 Define/Explain Percentiles and describe their merits. Q.2 Explain the limitations and uses of Percentiles.

8.4 CALCULATION OF PERCENTILES AND PERCENTILE RANK

In percentiles we calculate 10%, 15%, 40% and so on. This indicates that 10% or 15% cases of individual score falls below in a distribution. In other words, 15th percentile or 10th percentile is the score at or below which 15% or 10% of the scores in the distribution fall.

Whereas Percentile Rank indicates an individual's percentile rank on a test referring to the percentage of cases or scores lying below PR. For ex- ample, a person is having a percentile rank of 20. This means that 20 is situated above twenty percent of the group of which he is a member or twenty percent of the group fall below this person's rank.

Methods of calculating Percentile points from distribution (Table 8.2)

Scores	f
95-99	1
90-94	2
85-89	4
80-84	5
75-79	14
70-74	10
65-69	6
(P _R) 60-64	4
55-59	4
50-54	2
45-49	3
40-44	1
	N=50

The method of calculating percentiles is the same as that of finding median. the formula is

$$P_p = (PN - F) \frac{X_i}{fp}$$

fp

P = Percentage of the distribution wanted, e.g., 10%, 25%, 40%, etc.

I = Exact lower limit of the class interval upon which P_p lies

PN = Part of N to be counted off in order to reach P_p

F = Sum of all scores upon intervals below I

fp = number of scores with the interval upon which P_p falls.

i = length of the class interval

Calculation of Percentile Points, P_{10} , P_{20} , P_{30} , P_{40} , P_{50} , P_{90} (Table No.8.2)

$$10\% \text{ of } 50 = 5 \quad 49.5 + \frac{(5 - 4)}{2} \times 5 = 52.0$$

$$20\% \text{ of } 50 = 10 \quad 59.5 + \frac{(10 - 10)}{4} \times 5 = 59.5$$

$$30\% \text{ of } 50 = 15 \quad 64.5 + \frac{(15 - 14)}{6} \times 5 = 65.3$$

$$40\% \text{ of } 50 = 20 \quad 69.5 + \frac{(20 - 20)}{10} \times 5 = 69.5$$

$$50\% \text{ of } 50 = 25 \quad 69.5 + \frac{(25 - 20)}{10} \times 5 = 72.0 \text{ (Mdn)}$$

$$90\% \text{ of } 50 = 45 \quad 84.5 + \frac{(45 - 43)}{4} \times 5 = 85.0$$

Calculation of P_R from the same above formula and by same method. (Table No.3)

If we want to find the P_{R7} of a man who scores 63, what is the answer? Score 63 falls on interval 60-64. There are 10 scores up to 59.5 exact lower limit of this interval (see table No. 3) and four scores spread over this interval. Dividing 4 by 5 (length of interval) gives us 0.8 score percent of interval.

The score of 63, which we are finding is 3.5 score units from 59.5, exact lower limit of this interval. So multiply 3.5 by 0.8 we get 2.8 as the score-distance of 63 from 59.5, and adding 2.8 to 10 (number below 59.5) we get 12.8 where N lying below 63. Hence the percentile rank of score 63 is 26.

P_R s from ordered data

There are so many instances where individuals and things can be put in 1-2-3-4 order with respect to some trait or characteristics, such trait or quality

cannot be measured directly. Then we apply the following formula to calculate P_R

$$\text{The formula is } P_R = 100 - \frac{(100 R - 50)}{N}$$

Example, if 20 Doctors/ Professors have been ranked from 1 to 20, it is possible to convert this order of merit into P_R or scores on a scale of 100. Applying the formula we get:

Professor gets 1 first rank

$$P_R = 100 - \frac{(100 \times 1 - 50)}{20}$$

$$P_R = 100 - \frac{(100 - 50)}{20}$$

$$P_R = 100 - \frac{50}{20}$$

$$P_R = 100 - 2.5$$

$$P_R = 97.5$$

Doctor gets 10th rank, therefore

$$P_R = 100 - \frac{(100 \times 10 - 50)}{20}$$

$$P_R = 100 - \frac{(1000 - 50)}{20}$$

$$P_R = 100 - \frac{950}{20}$$

$$P_R = 100 - 47.5$$

$$P_R = 52.5$$

Check your progress

Q1 Define percentiles and calculate P_{60}, P_{70}, P_{80} from the distributions (Table No.3)

Q2 Calculate $P_R 70$ and $P_R 75$ from the same distribution.

8.5 SUMMARY

In this chapter, we discussed the merits, uses and limitations of the remaining one important measure of variability or dispersion, that is, standard deviation and also compared the four measures and answered the question "which is the best measure of variability" in brief. Apart from this, we explained the nature, merits, limitations and uses of percentiles along with the formula for calculating percentiles and percentile ranks.

8.6 QUESTIONS

Q1. (a) Explain SD and its uses.

(b) Calculate P_{140} and PR from the following frequency distribution.

Scores	F
50-54	3
45-49	6
40-44	9
35-39	15
30-34	19
25-29	12
20-24	14
15-19	6
10-14	2
	N = 140

Q.2 Define percentiles and describe the uses and limitations of percentiles.

8.7 REFERENCES

Annaporna R. et al (20014) - A Handbook of Mathematics and Statistics, Chetana Publications Pvt. LTD. 262, Khatauwadi, Girgaon, Mumbai-400004

Anastasi, A. and Urbina, S. (1997) -Psychological Testing (7th ed) Pearson Education, India Reprint -2002

Cohen, J.R. and Swerdik, M.E. (2010) Psychological Testing and Assessments, An Introduction to tests and Measurements (7th ed) New York McGraw-Hill International Edition.

Garrett, Henry, E. (1973) - Statistics in Psychology and Education, Vakills, Feffer and Simon Pvt. Ltd. Ballard Estate, Mumbai- 400001
