NETAJI SUBHAS OPEN UNIVERSITY

STUDY MATERIAL

BDP
EDUCATION

(EED-V)

PAPER V

GROUPS : A & B

EDUCATION EVALUATION AND STATISTICS IN EDUCATION

PREFACE

In the curricular structure introduced by the University for students of Bachelor Degree Programme, the opportunity to pursue Graduate course in subjects introduced by this University is equally available to all learners. Instead of being guided by any presumption about ability level, it would perhaps stand to reason if receptivity of a learner is judged in the course of the learning process. That would be entirely in keeping with the objectives of open education which does not believe in contrived differentiation.

Keeping this in view, study materials of the Graduate level in different subjects are being prepared on the basis of a well laid-out syllabus. The course structure combines the best elements in the approved syllabi of Central and State Universities in respective subjects. It has been so designed as to be upgradable with the addition of new information as well as results of fresh thinking and analysis with the changing time.

The accepted methodology of open and distance education has been followed in the preparation of these study materials. Co-operation in every form of experienced scholars is indispensable for a work of this kind. We, therefore, owe an enormous debt of gratitude to everyone whose tireless efforts went into the writing, editing and devising of a proper lay-out of the materials. Practically speaking, their role amounts to an involvement in invisible teaching. For, whoever makes use of these study materials would virtually derive the benefit of learning under their collective care and wisdom without each being seen by the other.

The more learners would seriously pursue these study materials the easier it will be for him/her to reach out to the wider horizons of a subject. Care has also been taken to make the language lucid and presentation attractive so that may be rated as quality self-learning materials. If anything remains still esoteric or difficult to follow, arrangements are there to come to terms with them through the counseling sessions regularly available at the network of study centres set up by the University.

Needless to add, a great part of these efforts is still experimental-in fact, pioneering in certain areas. Naturally, there is every possibility of some omission or inadequacy here and there. However, these do admit of restitution and furtherance in due course. On the whole, therefore, these study materials are expected to evoke wider appreciation the more they receive serious attention of all stakeholders.

Professor (Dr.) Subha Sankar Sarkar
Vice-Chancellor

First Edition : October, 2019

Printed in accordance with the regulations of the Distance Education Bureau of the University Grants Commission.

BACHELOR DEGREE PROGRAMME IN EDUCATION (EED)

Paper-V
Educational Evaluation and Statistics in Education

Chairman

Prof. Subha Sankar Sarkar
Vice-Chancellor,
Netaji Subhas Open University,
Kolkata-64

Convenor

Prof. Atindranath Dey
Director
School of Education
Netaji Subhas Open University,
Kolkata-64

Course Contributors

Group-A: Educational Evaluation

Unit - 1 & 2 : Dr. Chandan Adhikary
Unit - 3 & 4 : Dr. Keka Trivedi

Content Editing

Prof. Nemai Chand Maity

Group-B: Statistics in Education

Unit - 5 & 6: Dr. Kartik Pramanik
Unit - 7 & 8: Dr. Arjun Chandra Das

Former Editing **Dr. Papiya Upadhyay & Dr. Parimal Sarkar**

Course Design & Coordination: Dr. Parimal Sarkar & Dr. Papiya Upadhyay

---

Netaji Subhas Open University

Educational Evaluation and Statistics in Education
EED
Paper-V (Group—A & B)

Group—A : Educational Evaluation

Unit 1	❑	Measurement and Evaluation in Education	7-21
Unit 2	❑	Tools and Techniques of Evaluation	22-54
Unit 3	❑	Characteristics of a Good Tool	55-84
Unit 4	❑	Evaluation Process	85-116

Group—B : Satistics in Education

Unit 5	❑	Educational Statistics	117-140
Unit 6	❑	Descriptive Statistics	141-173
Unit 7	❑	Normal Distribution and Derived Score	174-201
Unit 8	❑	Relationship & Inferential Statistics	202-238

Unit-1 ❑ Measurement and Evaluation in Education

Course Content

1.1 Introduction

1.2 Objectives

1.3 Concept, Scope and Need of Evaluation

- **1.3.1 Definition of Evaluation**
- **1.3.2 Concept of Evaluation**
- **1.3.3 Principles of Evaluation**
- **1.3.4 Scope of Evaluation**
- **1.3.5 Purpose or need of Evaluation in Education**
- **1.3.6 Characteristics of a Good Evaluation Programme**
- **1.3.7 Functions of Evaluation**
- **1.3.8 Evaluation in teaching and Learning**

1.4 Relation between Evaluation and Measurement

1.5 Scales of measurement-Nominal, Ordinal, Interval and Ratio

1.6 Let us sum up

1.7 Check your progress (10 questions: 5 SA type, 5 objective type)

1.8 References

1.1 Introduction

Evaluation is an integral part of our teaching-learning process. Very often we blame our evaluation system as it failed to assess the learners' attainment with high degree of accuracy. There is an urgent demand to establish faith in our existing evaluation system. In order to enhance our human resource, proper evaluation of all aspects of learners' personality is mandatory. In this regard continuous and comprehensive evaluation covering cognitive, affective and psychomotor domain of the learners has to be addressed in formal, informal and non formal education sectors respectively. Due importance should be laid upon to understand the process of evaluation and its application. This would help the learners know their strengths and weaknesses and the abilities of the instructors in the curriculum constriction.

Evaluation is a generic term which involves value judgment based on sound

criteria. In this unit, we will be learning about the concept and meaning of evaluation, general principles of evaluation and characteristics of evaluation programme.

1.2 Objectives

After going through this unit you will be able to:

- Define educational Evaluation
- Explain concept of Evaluation
- State the principles of Evaluation
- Explain scope of Evaluation
- Explain need or purpose of evaluation
- Describe the features of a good evaluation programme
- Justify the features of a good evaluation tool
- Justify the pace and relevanceof evaluation in teaching learning
- State the functions of evaluation
- Correlate among Measurement, Assessment and Evaluation
- State the features of different measurement scale

1.3 Concept, Scope and Need of Evaluation

1.3.1 Definition of Evaluation

Different educationist has defined evaluation as following:

SI No	Educationist	Definition of Evaluation
1.	James M. Bradfield	Evaluation is the assignment of symbols to phenomenon, in order to characterize the worth or value of a phenomenon, usually with reference to some cultural or scientific standards.
2.	Thorndike and Hegan	The term evaluation is closely related to measurement. It is in some respect, inclusive including informal and intuitive judgement of pupil's progress. Evaluation is describing something in term of selected attributes and judging the degree of acceptability or suitability of that which has been described.

SI No	Educationist	Definition of Evaluation
3.	Norman E. Gronlund and Robert L. Linn	Evaluation is a systematic process of collecting, analysing and interpreting information to determine the extent to which pupil's are achievement instructional objectives.
4.	C.V. Good	The process of ascertaining or judging the value or amount of something by use of a standard of standard of appraisal includes judgement in terms of internal evidence and external criteria.
5.	Tyler	Tyler defined evaluation as "a systematic process of determining the extent to which educational objectives are achieved by pupils". This definition indicates that evaluation is a systematic process, and it omits tile casual, informal or uncontrolled observation of the pupils. The definition also implies that objectives of education have to be identified in advance. Without predetermined objectives, it is not possible to judge the progress, growth and development of students.
6.	Cronbach	Cronbach defined evaluation as "the collection and use of information to make decisions about an educational programme".
7.	Wheeler	Wheeler defined evaluation as a more general judgment of the outcome of a programme, which involves the use of observations, various tests, questionnaires, interviews, etc. His emphasis was on the processes of educational evaluation.

1.3.2 Concept of Evaluation

From the above definition Evaluation can be conceptualized in the following manner:

1) Evaluation is an act or a process that allows one to make a judgment about the desirability or value of a measure.

2) Evaluation is a process of delineating, obtaining and providing useful information for judging decision alternatives

3) The word evaluation refers to the act or process of determining the value of something.

Accordingly evaluation is providing information for decision making. Thus evaluation is a systematic process of collecting evidence about students' achievement in both cognitive and non-cognitive areas of learning on the basis of which judgments are formed and decisions are made. Thus Evaluation is a concept that has emerged as a prominent process of assessing, testing and measuring. Its main objective is Qualitative Improvement. Evaluation is a process of making value judgements over a level of performance or achievement. Making value judgements in Evaluation process presupposes the set of objectives. Evaluation is the process of determining the extent to which the objectives are achieved. Concerned not only with the appraisal of achievement, but also with its improvement. Evaluation is continuous and dynamic. In this regard it is to be kept in mind that a test is a set of question measurement is assigning numbers to the results of test according to some specific rules on the other hand evaluation adds value judgment.

1.3.3 Principles of Evaluation:

Evaluation is a systematic process of determining to what extent instructional objectives has been achieved. Therefore evaluation process must be carried out with effective techniques. The principles of evaluation can be categorized in two heads, viz., 1) General Principles and 2) Specific Principles.

1) General principles of evaluation

The general principles of evaluation is depicted in following diagram (Fig. 1.1)

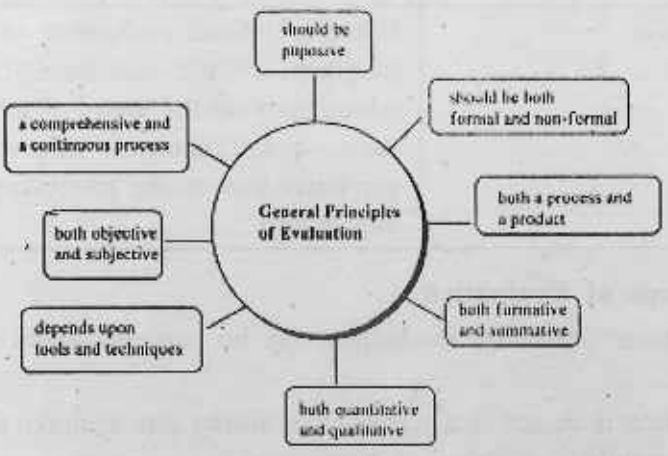


Fig. 1.1 General principles of Evaluation

2) Specific principles of evaluation

The following principles will help to make the evaluation process an effective one

1. It must be clearly stated what is to be evaluated

A teacher must be clear about the purpose of evaluation. He must formulate the instructional objectives and define them clearly in terms of student's observable behaviour. Before selecting the achievement measures the intended learning out comes must be specified clearly.

2. A variety of evaluation techniques should be used for a comprehensive evaluation

It is not possible to evaluate all the aspect of achievement with the help of a single technique. For the better evaluation the techniques like objective tests, essay tests, observational techniques etc. should be used. So that a complete' picture of the pupil achievement and development can be assessed.

3. An evaluator should know the limitations of different evaluation techniques

Evaluation can be done with the help of simple observation or highly developed standardized tests. But whatever the instrument or technique may be it has its own limitation. There may be measurement errors. Sampling error is a common factor in educational and psychological measurements. An achievement test may not include the whole course content. Error in measurement can also be found due to students guessing on objective tests. Error is also found due to incorrect interpretation of test scores.

4. The technique of evaluation must be appropriate for the characteristics or performance to be measured

Every evaluation technique is appropriate for some uses and inappropriate for another. Therefore while selecting an evaluation technique one must be well aware of the strength and limitations of the techniques.

5. Evaluation is a means to an end but not an end in itself

The evaluation technique is used to take decisions about the learner. It is not merely gathering data about the learner. Because blind collection of data is wastage of both time and effort. But the evaluation is meant for ,some useful purpose.

6. Determining and Clarifying' What' aspect of the Evaluation

The classroom teacher or evaluator should always be perfectly clear in his mind about what he is aiming to achieve i.e. what to evaluate and how to evaluate.

7. Selection of Appropriate Evaluation Techniques

There are a number of evaluation techniques. Out of them one technique is appropriate in some cases which may not be so in others. Therefore, the evaluator needs to select the one which serves his/her purpose best.

8. Determining Comprehensiveness of Evaluation Programmes

It means to assess pupils' progress in all areas. Educational evaluation, apart from testing knowledge, should also bring about student's originality and his ability to use the ideas, and his ability to think and apply the knowledge and skills already achieved.

9. For Comprehensive Evaluation combining a variety of Evaluation Techniques be adopted

To make evaluation comprehensive, different types of evaluation procedures should be adopted depending on their suitability. Moreover, use of a variety of techniques provides an evaluator sufficient evidences of different aspects of pupil achievement on different objectives, because more the evidence better the evaluation.

10. Treat Evaluation as a Means to an End, not an End in itself

In the teaching-learning process, evaluation should be done with a purpose, and not for the sake of evaluation only. Administering a test, scoring the scripts and collecting the data without making any use of this information for the pupils is a waste of effort.

1.3.4 Scope of Evaluation:

The term evaluation is a broadest concept with multifarious activities. Therefore, subject matter or scope of evaluation covers a wide area. The scope of evaluation is predicted below.

1. Evaluation and instructional objectives:

Learning outcomes expected from class-room discussion can be fixed by using evaluation results.

What type of knowledge and understanding the student should develop? What skill they should display? What interest and attitude they should develop?

Can only be possible when we shall identify the instructional objectives and state them clearly in terms of intended learning outcomes. Only a good evaluation process helps us to fix up a set of perfect instructional objectives.

2. Evaluation process and assessing the learner's needs

In the teaching learning process it is very much necessary to know the needs of the learners. The instructor must know the knowledge and skills to be mastered by the students. Evaluation helps to know whether the students possess required knowledge and skills to proceed with the instruction.

3. Evaluation and feed back to the students

An evaluation process helps the teacher to know the learning difficulties of the students. It helps to bring about an improvement in different school practices. It also ensures an appropriate follow-up service.

4. Evaluation and preparing programmed materials

Programmed instruction is a continuous series of learning sequences. First the instructional material is presented in a limited amount then a test is given to response the instructional material. Next feedback is provided on the basis of correctness of response made. So that without an effective evaluation process the programmed learning is not possible.

5. Evaluation and curriculum development

Curriculum development is an important aspect of the instructional process. Evaluation data enable the curriculum development, to determine the effectiveness of new procedures, identify areas where revision is needed. Evaluation also helps to determine the degree to what extent an existing curriculum is effective. Thus evaluation data are helpful in constructing the new curriculum and evaluating the existing curriculum.

6. Evaluation and reporting pupil's progress to parents

A systematic evaluation procedure provides an objective and comprehensive picture of each pupil's progress. This comprehensive nature of the evaluation process helps the teacher to report on the total development of the pupil to the parents. This type of objective information about the pupil provides the foundation for the most effective co-operation between the parents and teachers,

7. Evaluation and guidance and counselling

Evaluation procedures are very much necessary for educational, vocational and personal guidance. In order to assist the pupils to solve their problems in the educational, vocational and personal fields the counsellor must have an objective knowledge of the pupils abilities, interests, attitudes and other personal characteristics.

An effective evaluation procedure helps in getting a comprehensive picture of the pupil which leads to effective guidance and of counselling.

8. Evaluation and school administration

Evaluation data helps the administrators to judge the extent to which the objectives of the school are being achieved, to find out strengths and weaknesses of the curriculum and arranging special school programmes. It also helps in decisions concerning admission, grouping and promotion of the students.

9. Evaluation and school research:

In order to make the school programme more effective, researches are necessary. Evaluation data help in research areas like comparative study of different curricula, effectiveness of different methods, effectiveness of different organizational plans, etc.

1.3.5 Purpose or need of Evaluation in Education:

Evaluation serves numerous purposes in education, Some of the important purposes are to grade, rank, classify, compare and promote the students, It is also used for certifying the completion of a course, selection of students for admission or scholarship, and for predicting their future success in different endeavors.

The sole purpose of evaluation has been to bring about quality improvement in education which it does by providing feedback regarding students' learning, classroom teaching, effectiveness of curriculum and course content. It also helps bring about all round development of the students' personality when it is used for developing their non-cognitive capacities.

1.3.6 Characteristics of a Good Evaluation Programme:

The meaning, types and purpose of evaluation lead us to arrive at the following characteristics of a good evaluation programme in educational institutions. Evaluation is a Dynamic Process.

Evaluation is based on learning experiences; it also provides evidence about the effectiveness of that learning experience. Thus, evaluation keeps validating the whole teaching-learning process through regular feedback. Thus evaluation programme brings in dynamism and leads to continuous improvement in the entire educational process.

Evaluation is a Cooperative process

The teacher alone cannot gel all the evidence required about student's growth.

To collect evidence regarding social relationships, emotional behaviour, initiative, scientific attitudes, social attitudes, likes and dislikes, etc. collaboration ofthe student peers, parents, other teachers and all those who watch him/her grow and develop is necessary.

Evaluation is an Objective-oriented Process

It is for the achievement of the instructional objectives that the instruction is given. Evaluation is made to confirm whether the instructional objectives have been achieved and to what extent. The selection of evaluation techniques and tools is also based on the objectives to be evaluated.

Evaluation is a Continuous Process

Continuous evaluation is, therefore, essential for getting reliable evidence about student's growth and development.

Evaluation is a Comprehensive Process.

As a good evaluation programme should evaluate both the cognitive and non-cognitive aspects of learner growth. Apart from evaluating all possible objectives, comprehensive evaluation involves the use of multiple tools and techniques to procure information on different aspects of personality growth.

Evaluation is a Decision Making Process.

At every step of the teaching-learning process evaluation is a must . Before the instruction is started, it is necessary to determine the entering behaviour of students to decide the strategies, learning material and even appropriate objectives of teaching. Evaluation helps the teachers to make judgments and take decisions at different stages in a pupil's educational career.

1.3.7 Functions of Evaluation:

The main functions of evaluation are as follows:

1) Diagnosis:
   (a) To locate and identify the weaknesses and strength in learning on the part of a learner.
   (b) To pinpoint areas where remedial measures may be desirable.
2) Modification:

   To provide a basis for a modification ofthe curriculum, syllabus or courses.

3) Prediction:

To bring out the inherent capabilities of a student, such as proper attitudes, habits, manipulative skills, appreciation and understanding in addition to conventional acquisition of knowledge.

4) Selection:

To select suitable persons of a particular course or career.

5) Motivation:

To motivate pupils towards better attainment and growth.

6) Teaching:

a) To improve instruction.

b) To ascertain how far could learning objective be achieved;

c) To provide the empirical evidences about the effectiveness of teaching strategies, tactics and aids.

7) Guidance :

a) To assist a person in decision making about a course or subjects within a course and careers;

b) To enable a learner to know his pace of learning;

c) To make provision for guiding the growth of individual pupils;

d) To provide a basis for the introduction of experiences to meet the needs of individuals and groups of pupils.

8) Testing:

a) To test the efficiency of teachers in providing learning experience and the effectiveness of instruction and of classroom activates;

b) To help in developing a comprehensive criterion test.

9) Grading:

To assign rank or grade to the learners of a give group. (Example: The unit test)

10) Feedback:

To give reinforcement and feedback to teachers and learners.

1.3.8 Evaluation in teaching and Learning:

Evaluation is an integral part of any teaching and learning programme. Whenever a question is asked and answered evaluation takes place. Thus, both teaching and evaluation overlap and merge into each other. In fact, it is not possible to have teaching and learning without evaluation.

Both teaching and evaluation are based on the instructional objectives which provide direction to them. Instructional objectives are those desirable behaviours which are to be developed in students. It is for achieving the instructional objectives that instruction is provided and it is to see whether the instructional objectives have been achieved and to what extent, that the evaluation is made. Three components of teaching and learning constitute an integrated network in which each component depends on the other. Thus, through evaluation, the teacher not only assesses as to how far the student has achieved the objectives of teaching but also judges the effectiveness of the learning experiences, methodologies, means and the materials used for achieving those objectives.

1.4 Relationship between Measurement and Evaluation

The terms measurement and evaluation assessment are sometimes used interchangeably; The word 'evaluation' is often confused with assessment, testing and measurement. Testing is only a technique to collect evidence regarding pupil behaviour. Measurement on the other hand, is limited to quantitative description of the student behaviour. Evaluation is a more comprehensive term which includes testing and measurement and also qualitative description of the student behaviour. It also includes value judgment regarding the worth or desirability of the behavior measured or assessed.

Considering the importance of this relationship, Gronlund has indicated this relationship in the following equation:

Measurement = quantitative description of pupils (measurement) + value judgment

Evaluation = qualitative description of pupils (non-measurement) + value judgment

Thus, evaluation may not be based on measurement alone but it goes beyond the simple quantitative score.

For example when we say Rohan secured 45 numbers in Arithmetic. It just indicates 'how much' Rohan has successfully answered. It does not include any

qualitative description i.e. 'how good' he is in Arithmetic. Evaluation on the other hand includes both quantitative description (measurement) and qualitative description (Non measurement) along with value judgments. This relationship between measurement, non measurement and evaluation can be illustrated with the help of following diagram (Fig. 1.2).

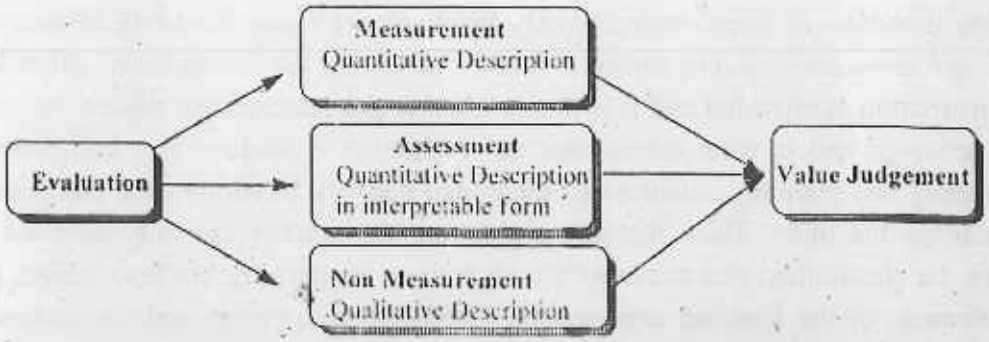


Fig. 1.2 Relationship among Measurement, Assessment and Evaluation

1.5 Scales of Measurement

Traditional Classification Statisticians call an attribute on which observations differ a variable. The type of unit on which a variable is measured is called a scale. Traditionally, statisticians talk of four types of measurement scales: (1) nominal, (2) ordinal, (3) interval, and (4) ratio.

(1) Nominal Scales:

The word nominal is derived from nomen, the Latin word for name. Nominal scales merely name differences and are used most often for qualitative variables in which observations are classified into discrete groups. The key attribute for a nominal scale is that there is no inherent quantitative difference among the categories. Sex, religion, and race are three classic nominal scales used in the behavioral sciences. Taxonomic categories (rodent, primate, canine) are nominal scales in biology. Variables on a nominal scale are often called categorical variables.

(2) Ordinal Scales:

Ordinal scales rank-order observations. Class rank and horse race results are

examples. There are two salient attributes of an ordinal scale. First, there is an underlying quantitative measure on which the observations differ. For class rank, this underlying quantitative attribute might be composite grade point average, and for horse race results it would be time to the finish line. The second attribute is that individual differences individual on the underlying quantitative measure are either unavailable or ignored. As a result, ranking the horses in a race as 1st, 2nd, 3rd, etc. hides the information about whether the first-place horse won by several lengths or by a nose. There are a few occasions in which ordinal scales may be preferred to using a quantitative index of the underlying scale. College admission officers, for example, favor class rank to overcome the problem of the different criteria used by school districts in calculating GPA. In general, however, measurement of the underlying quantitative dimension is preferred to rank-ordering observations because the resulting scale has greater statistical power than the ordinal scale.

(3) Interval Scales:

In ordinal scales, the interval between adjacent values is not constant. For example, the difference in finishing time between the 1st place horse and the 2nd horse need not the same as that between the 2nd and 3rd place horses. An interval scale has a constant interval but lacks a true 0 point. As a result, one can add and subtract values on an interval scale, but one cannot multiply or divide units. Temperature used in day-to-day weather reports is the classic example of an interval scale. The assignment of the number 0 to a particular height in a column of mercury is an arbitrary convenience apparent to everyone anyone familiar with the difference between the Celsius and Fahrenheit scales. As a result, one cannot say that 30°C is twice as warm as 15°C because that statement involved implied multiplication. To convince yourself, translate these two into Fahrenheit and ask whether 86°F is twice as warm as 50°F. Nevertheless, temperature has constant intervals between numbers, permitting one to add and subtract. The difference between 28°C and 21°C is 7 Celsius units as is the difference between 53°C and 46°C. Again, convert these to Fahrenheit and ask whether the difference between 82,4°F and 69.8°F is the same in Fahrenheit units as the difference between 127.4°F and 114.8°F?

(4) Ratio Scales:

A ratio scale has the property of equal intervals but also has a true 0 point. As a result, one can multiply and divide as well as add and subtract using ratio

scales. Units of time (msec, hours), distance and length (cm, kilometers), weight (mg, kilos), and volume (cc) are all ratio scales. Scales involving division of two ratio scales are also themselves ratio scales. Hence, rates (miler per hour) and adjusted volumetric measures (mg/dL) are ratio scales. Note that even though a ratio scale has a true 0 point, it is possible that the nature of the variable is such that a value of 0 will never be observed. Human height is measured on a ratio scale but every human has a height greater than 0. Because of the multiplicative property of ratio scales, it is possible to make statements that 60 mg is three times as great as 20 mg.

1.6 Let us sum up

In this Unit, you studied about the meaning and concept of evaluation, principle of evaluation and features of evaluation. You studied that evaluation is a systematic process of determining the extent to which the instructional objectives are achieved by the pupils.

i. Evaluation is a systematic, comprehensive and continuous process of value judgment covering all aspects of learners' attainment in cognitive, affective and psychomotor domain. .

ii. Evaluation always assumes that instructional objectives have been previously determined and properly formulated. Evaluation includes measurement (quantitative description), assessment (quantitative description in interpretable form) and non measurement (qualitative description), i.e. value judgment. You studied the principles of evaluation. There the purpose should be carefully defined, appropriate tools/techniques should be selected, it should be comprehensive as far as practicable, it should be both quantitative and qualitative etc.; You also studied the qualities of a good evaluation programme. They are reliability, validity, objectivity, comprehensiveness, practicability and discrimination power.

In the last part of this unit, you studied about the different types of evaluation and their importance, measurement scales and their uses.

1.7 Check your progress (10 questions: 5 SA type, 5 objective type)

A) Short Answer type questions:

1) Evaluation is an integral part of teaching-learning process-Explain the statement
2) Bring out the importance of continuous and comprehensive evaluation.
3) Distinguish between evaluation and measurement
4) Write a brief account on different types of measurement scales.
5) Mention the features of a good evaluation programme.

B) Objective Answer type questions:

1) Evaluation is a —— and ——— process, (fill in the blanks)
2) Evaluation is both qualitative and quantitative where as measurement——————. (fill in the blank)
3) Arrange the following in the hierarchical order of comprehensiveness: Measurement, Assessment, Evaluation.
4) Which of the following is not the characteristic feature of formative evaluation? a) Diagnosis b) Remedial instruction c) Immediate feedback d) Gradation.
5) Educational data are collected in ratio scale (say True or False).

1.8 References

1. Aggarwal, J.C. *Essentials of Examination System, Evaluation, Tests and Measurement.* Vikash Publication Pvt. Ltd., New Delhi, 1997.
2. Gronlund, W. E. *Measurement and Evaluation in Teaching.* McMillan Publishing Co, New York, 1976.
3. Bhatia, K. K. *Measurement and Evaluation in Education.* Prakash Brothers, Ludhiana, 1973.
4. Dandekar, W.M. *Evaluation in Schools,* Srividya Prakashana, Poova, 1970.
5. Rawat. D.S. *Measurement, Evaluation and Statistics in Education.* New Raj Book Depot, New Delhi, 1970.
6. NSOU Study Material (B.Ed. ODL Programme, *Teacher Functions,* School of Education, Kolkata, 2013.
7. IGNOU Study Material (MESE-057), *Measurement and Evaluation in Education,* School of Education, New Delhi.

Unit-2 ❑ Tools and Techniques of Evaluation

Course Content

2.1 Introduction

2.2 Objectives

2.3 Tests: Essay type and Objective type, Short answer type and Oral type

2.3.1 Techniques and Tools of Evaluation

2.3.2 Features of a Good Evaluation Tool

2.3.3 Test

2.3.3.1. Essay Type Test

2.3.3.1.1 Meaning of Essay Type Test

2.3.3.1.2 Types of Essay Test

2.3.3.1.3 Advantages of the Essay Tests

2.3.3.1.4 Limitations of Essay Tests

2.3.3.1.5 Suggestions for Improving Essay Tests

2.3.3.2 Short Answer Type Test

2.3.3.2.1 Meaning of Short Answer Type Test

2.3.3.2.2 Classification of Short Answer Type Test

2.3.3.2.3 Merits of Short Answer Type Test

2.3.3.2.4 Demerits of Short Answer Type Test

2.3.3.3 Objective Type Test

2.3.3.3.1 Meaning of Objective Type Test

2.3.3.3.2 Objective-Centered Test/Objective based Test

2.3.3.3.3 Objective-type tests have two characteristics viz

2.3.3.3.4 Merits of Objective Type Test

2.3.3.3.5 Limitations of Objective Type Test

2.3.3.3.6 Guidelines for Constructing Better Objective Type Test Items:

2.3.3.4 Oral Type Test

2.3.3.4.1 Meaning of Oral Type Test

2.3.3.4.2 Merits of Oral Type Tests

2.3.3.4.3 Limitations of oral Type Test

2.4 Personality test-Rorschach Ink Blot Test

2.4.1 Meaning of Personality Test

2.4.2 Projective Technique of Assessing Personality

2.4.3 The Rorschach Ink-blot Test

2.5 Interest Test-Kuder Test

2.5.1 Meaning of the term Interest

2.5.2 Key Aspects of Interest

2.5.3 Interest Inventory and Kuder's Test

2.6 CRC

2.6.1 Meaning of Cumulative Records

2.6.2 Need for School Record

2.6.3 Characteristics of Cumulative Record

2.6.4 Data contained in the cumulative record card (CRC) should be

2.6.5 Types of Information Maintained in the-CRC

2.6.6 Sources of Collection of Information

2.6.7 Maintenance of the Record

2.7 Observation

2.7.1 Meaning of Observation

2.7.2 Advantages of Observation

2.7.3 Limitations of Observation Technique

2.8 Self-reporting (Interview, Questionnaire)

2.8.1 Meaning of Self Reporting

2.8.2 Interview

2.8.2.1 Definition of Interview

2.8.2.2 Types of Interview

2.8.2.3 Advantages of Interview

2.8.2.4 Limitations of the Interview

2.8.3 Questionnaire

2.8.3.1 Meaning of Questionnaire

2.8.3.2 Definition of Questionnaire

2.8.3.3 Types of Questionnaire

2.8.3.4 Advantages of Questionnaires

2.8.3.5 Limitations of Questionnaire

2.9 Let us sum up

2.10 Check your progress (10 questions: 5 SA type, 5 objective type)

2.11 References

2.1 Introduction

Evaluation of learners' learning outcomes is one of the key functions of teaching-learning process. It is the evaluation process, during or after instruction, that reveals whether a set of learning outcomes has been achieved or not. It is this process that helps to measure, test, assess and appraise different learning outcomes or competencies. But all learning outcomes, which fall under different domains, cannot be tested through the written tests alone. There are different kinds of evaluation techniques and tools to measure different competencies and skills. These learning outcomes are generally categorized into cognitive and non-cognitive learning outcomes. Cognitive outcomes are associated with scholastic domain of students' performance; while non-cognitive learning outcomes are related to non-scholastic areas. Evaluation of cognitive outcomes is carried out with the help of certain techniques and tools, whereas some other techniques and tools are used for assessing the learners' performance in non-cognitive sectors. In this unit we will discuss tools and techniques which are conventionally used to evaluate cognitive and non cognitive learning outcomes of the learners. This would cover three basic techniques names examination, self reporting and observation; this is followed by tools testing,

interview, questionnaire and observation schedule and cumulative record card. A detailed discussion on essay type, short answer type, objective type testing, personality test and interest test have also been provided in this unit.

2.2 Objectives

After going through this unit you will be able to:

- Differentiate between techniques and tools of evaluation
- List the needs for a variety of tools and techniques
- Explain various types of testing and their advantages and disadvantages
- Explain the features, merits and demerits of Rorsrachch Inkblot test as personality test
- Explain the features, merits and demerits of Ruder test as interest inventory
- Describe the features, advantages and limitations of observation as a technique of evaluation
- Describe the features, types, advantages and disadvantages of questionnaire as a tool
- Describe the features, types, advantages and disadvantages of interview as a tool
- State the features and utility of cumulative record card in evaluation

2.3 Tests: Essay type and Objective type, Short answer type and Oral type

2.3.1 Techniques and Tools of Evaluation

The technique of evaluation is a method or procedure in carrying out a scientific or mechanical operation or a degree of expertness shown in the method or procedure or carrying out an operation. The technique of evaluation in the field of education is the degree of skillfulness dexterity shown in the process of collecting information or evidences with regard to a characteristic quality of a person or performance process.

The term tool literally means implement or appliance for an operation. In educational evaluation it is a means of collecting evidences for students' performance.

The techniques and tools of evaluation is collated in Fig. 2.1.

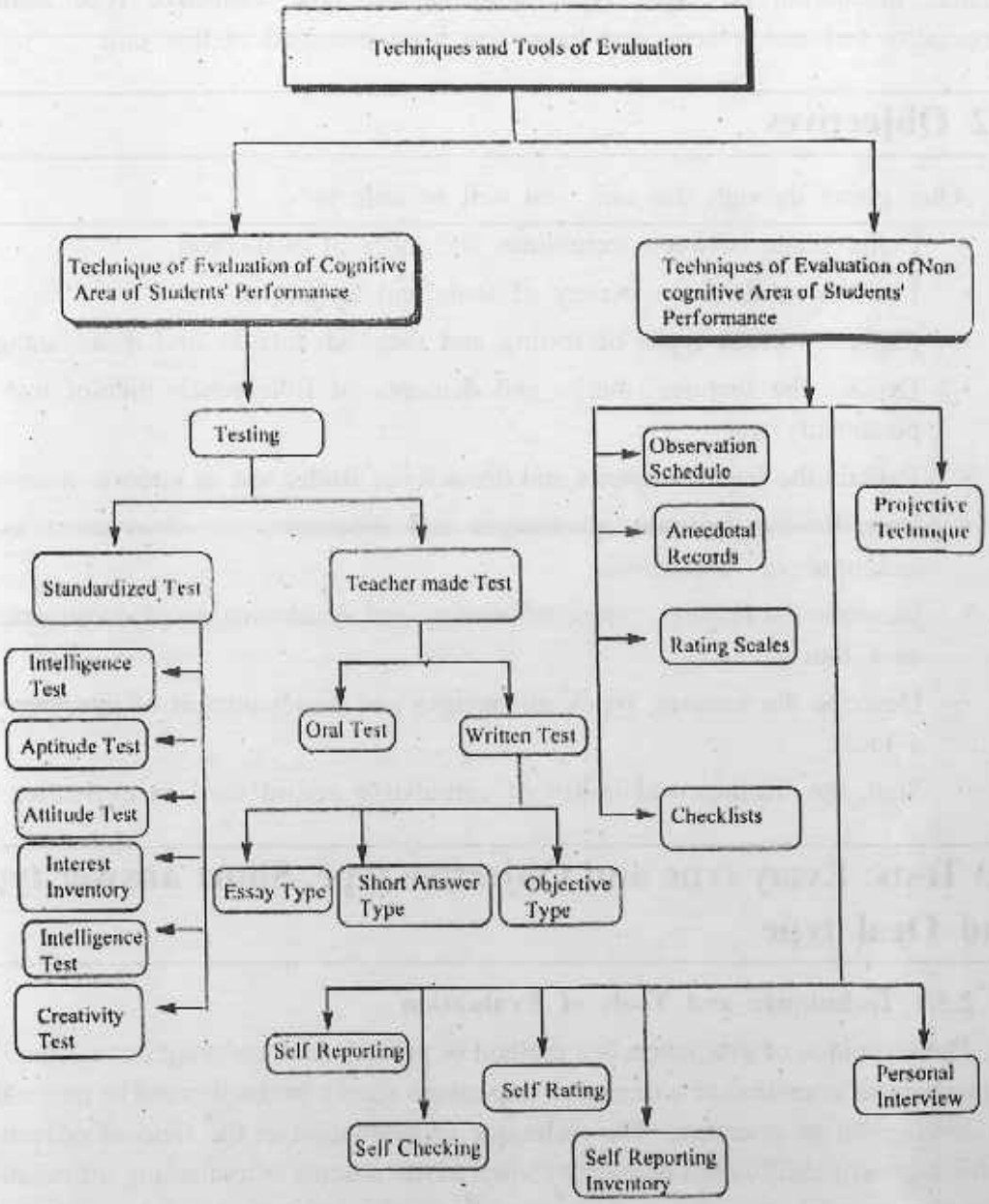


Fig. 2.1 Techniques and Tools of Evaluation

2.3.2 Features of a Good Evaluation Tool

The features of a good evaluation tool is depicted in Figure 2.2.

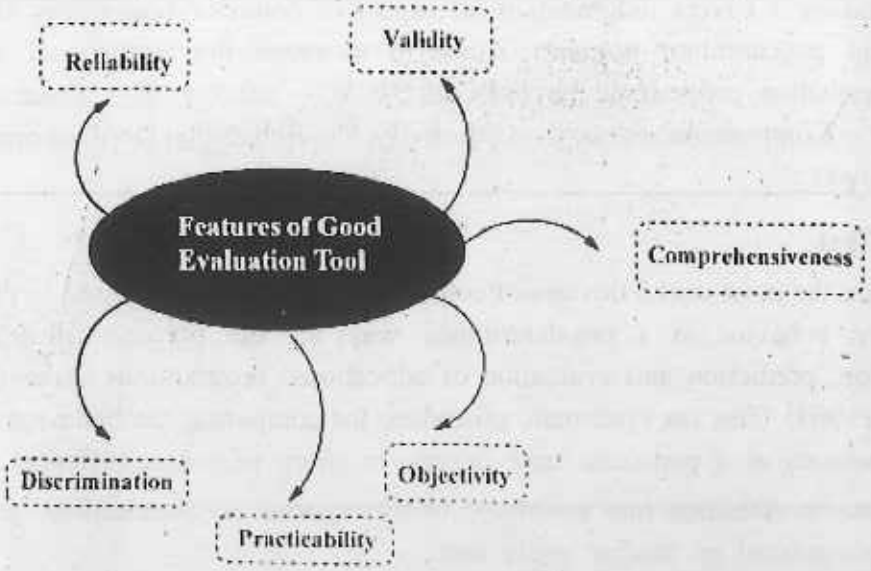


Fig. 2.2 Features of a good evaluation tool

a) **Reliability:** Reliability refers to consistency of an evaluation. If the evaluation tool reveals consistency while measuring at different intervals then the tool is said to be reliable.

b) **Validity:** Validity of an evaluation means truthfulness. If a tool measures what it intends to measure, it is said to be valid.

N.B: Validity implies reliability but the reverse is not true. Thus, a reliable tool may not be valid but the valid tool is reliable.

c) **Objectivity:** An evaluation is objective when the scorer's personal judgment does not affect the scoring.

d) **Discrimination:** One of the most important features of a good evaluation tool is discrimination, i.e., it should be able to separate the good from bad or it should be able to divide the big group into small sub-groups of similar characteristics.

e) **Practicability:** An evaluation tool should be practicable in the sense that it can be administered easily with minimum outlay of time with minimum

cost and it should be easily scorable.

f) **Comprehensiveness:** A good evaluation programme should be comprehensive one as it covers judgment of all aspect of behavior (cognitive, affective and psychomotor domain). Acronym to recall the features of a good evaluation programme is: (VPCORD): V = Validity, P = Practicability, C = Comprehensiveness, 0 = Objectivity, R = Reliability, D = Discrimination power.

2.3.3 Test

Tests are the most useful devices of collecting data. They are devised to evaluate or measure behavior in a pre-determined way for the purpose of selection, classification, prediction and evaluation of educational programmes. According to Cornbach (1964) 'Test isa systematic procedure for comparing the behavior of two or more persons at a particular time or one or more personsat different times.

Test can be classified into mainly of two categories: a) Standardized test and b) Non-standardized or teacher made test.

a) **Standardized Test:** These tests are constructed by specialists. These tests are constructed after intensive verification of validity, reliability, objectivity, discriminating power etc. The standardized tests include carefully prepared directions for its administration and scoring and interpretations by means of norms that have been arrived at through application of the test to a large number of subjects. Tests measuring psychological aspects e.g., intelligence, aptitude, attitude, interest, personality traits, achievement, creativity etc. are examples of standardized test.

b) **Non standardized** Test: These are also called teacher made tests as they are prepared by the teachers to meet the immediate demands of the situation. These tests lack all the good aspects of standardized tests. They have a limited scope of use. The test may be either verbal (oral type test) or non verbal (performance type). Non verbal tests are two types: 1) Written test and 2) Practical test. Written tests can be classified as (a) Essay type test, (b) Short answer type test, (c) Very short answer type test and (d) Objective type test.

2.3.3.1. Essay Type Test

2.3.3.1.1 Meaning of Essay Type Test

The essay tests are still commonly used tools of evaluation, despite the increasingly wider applicability of the short answer and objective type questions.There are certain outcomes of learning (e.g., organising, summarising, integrating ideas and expressing in one's own way) which cannot be satisfactorily measured through objective type tests. The importance of essay tests lies in the measurement of such instructional outcomes.

An essay test may give full freedom to the students to write any number of pages. The required response may vary in length. An essay type question requires the pupil to plan his own answer and to explain it in his own words. The pupil exercises considerable freedom to select, organize and present his ideas. Essay type tests provide a better indication of pupil's real achievement in learning. The answers provide a clue to nature and quality of the pupil's thought process.

That is, we can assess how the pupil presents his ideas (whether his manner of presentation is coherent, logical and systematic) and how he concludes. In other words, the answer of the pupil reveals the structure, dynamics and functioning of pupil's mental life. The essay questions are generally thought to be the traditional type of questions which demand lengthy answers. They are not amenable to objective scoring as they give scope for halo-effect, inter-examiner variability and intra-examiner variability in scoring.

2.3.3.1.2 Types of Essay Test:

There can be many types of essay tests:

Some of these are given below with examples from different subjects:

1. Selective Recall.

e.g. What was the religious policy of Akbar?

2. Evaluative Recall.

e.g. Why did the First War of Independence in 1857 fail?

3. Comparison of two things—on a single designated basis,

e.g. Compare the contributions made by Dalton and Bohr to Atomic theory.

4. Comparison of two things—in general.

e.g. Compare Early Vedic Age with the Later Vedic Age.

5. Decision—for or against.

e.g. Which type of examination do you think is more reliable? Oral or Written. Why?

6. Causes or effects.

e.g. Discuss the effects of environmental pollution on our lives.

7. Explanation of the use or exact meaning of some phrase in a passage or a sentence.

e.g., Joint Stock Company is an artificial person. Explain 'artificial person' bringing out the concepts of Joint Stock Company.

8. Summary of some unit of the text or of some article.

9. Analysis

e.g. What was the role played by Mahatma Gandhi in India's freedom struggle?

10. Statement of relationship.

e.g. Why is knowledge of Botany helpful in studying agriculture?

11. Illustration or examples (your own) of principles in science, language, etc. e.g. Illustrate the correct use of subject-verb position in an interrogative sentence.

12. Classification.

e.g. Classify the following into Physical change and Chemical change with explanation. Water changes to vapour; Sulphuric Acid and Sodium Hydroxide react to produce Sodium Sulphate and Water; Rusting of Iron; Melting of Ice.

13. Application of rules or principles in given situations.

e.g. If you sat halfway between the middle and one end of a sea-saw, would a person sitting on the other end have to be heavier or lighter than you in order to make the sea-saw balance in the middle. Why?

14. Discussion.

e.g. Partnership is a relationship between persons who have agreed to share the profits of a business carried on by all or any of them acting for all. Discuss the essentials of partnership on the basis of this partnership.

15. Criticism—as to the adequacy, correctness, or relevance—of a printed statement or a classmate's answer to a question on the lesson.

e.g. What is the wrong with the following statement?

The Prime Minister is the sovereign Head of State in India.

16. Outline.

e.g. Outline the steps required in computing the compound interest if the principal amount, rate of interest and time period are given as P, R and *T* respectively.

17. Reorganization of facts.

e.g. The student is asked to interview some persons and find out their opinion on the role of UN in world peace. In the light of data thus collected he/she can reorganise what is given in the text book.

18. Formulation of questions-problems and questions raised.

e.g. After reading a lesson the pupils are asked to raise related problems-questions.

19. New methods of procedure

e.g. Can you solve this mathematical problem by using another method?

2.3.3.1.3 Advantages of the Essay Tests:

1. It is relatively easier to prepare and administer a six-question extended-response essay test than to prepare and administer a comparable 60-item multiple-choice test items.
2. It is the only means that can assess an examinee's ability to organise and present his ideas in a logical and coherent fashion.
3. It can be successfully employed for practically all the school subjects.
4. Some of the objectives such as ability to organise idea effectively, ability to criticise or justify a statement, ability to interpret, etc., can be best measured by this type of test.
5. Logical thinking and critical reasoning, systematic presentation, etc. can be best developed by this type of test.
6. It helps to induce good study habits such as making outlines and summaries, organising the arguments for and against, etc.
7. The students can show their initiative, the originality of their thought and the fertility of their imagination as they are permitted freedom of response.
8. The responses of the students need not be completely right or wrong. All degrees of comprehensiveness and accuracy are possible.
9. It largely eliminates guessing.

10. They are valuable in testing the functional knowledge and power of expression of the pupil.

2.3.3.1.4 Limitations of Essay Tests:

1. One of the serious limitations of the essay tests is that these tests do not give scope for larger sampling of the content. You cannot sample the course content so well with six lengthy essay questions as you can with 60 multiple-choice test items.
2. Such tests encourage selective reading and emphasise cramming.
3. Moreover, scoring may be affected by spelling, good handwriting, coloured ink; neatness, grammar, length of the answer, etc.
4. The long-answer type questions are less valid and less reliable, and as such they have little predictive value.
5. It requires an excessive time on the part of students to write; while assessing, reading essays is very time-consuming and laborious.
6. It can be assessed only by a teacher or competent professionals.
7. Improper and ambiguous wording handicaps both the students and valuers.
8. Mood of the examiner affects the scoring of answer scripts.
9. There is halo effect-biased judgement by previous impressions.
10. The scores may be affected by his personal bias or partiality for a particular point of view, his way of understanding the question, his weightage to different aspect of the answer, favouritism and nepotism; etc.

Thus, the potential disadvantages of essay type questions are:

(i) Poor predictive validity,

(ii) Limited content sampling,

(iii) Scores unreliability, and

(iv) Scoring constraints.

2.3.3.1.5 Suggestions for Improving Essay Tests:

The teacher can sometimes, through essay tests, gain improved insight into a student's abilities, difficulties and ways of thinking and thus have a basis for guiding his/her learning.

(A) While Framing Questions:

1. Give adequate time and thought to the preparation of essay questions, so that they can be re-examined, revised and edited before they are used. This would increase the validity of the test.
2. The item should be so written that it will elicit the type of behaviour the teacher wants to measure. If one is interested in measuring understanding, he should not ask a question that will elicit an opinion; e.g., "What do you think of Buddhism in comparison to Jainism?"
3. Use words which themselves give directions e.g. define, illustrate, outline, select, classify, summarise, etc., instead of discuss, comment, explain, etc.
4. Give specific directions to students to elicit the desired response.
5. Indicate clearly the value of the question and the time suggested for answering it.
6. Do not provide optional questions in an essay test because—
   (i) It is difficult to construct questions of equal difficulty;
   (ii) Students do not have the ability to select those questions which they will answer best;
   (iii) A good student may be penalised because he is challenged by the more difficult and complex questions.
7. Prepare and use a relatively large number of questions requiring short answers rather than just a few questions involving long answers.
8. Do not start essay questions with such words as list, who, what, whether. If we begin the questions with such words, they are likely to be short-answer question and not essay questions, as we have defined the term.
9. Adapt the length of the response and complexity of the question and answer to the maturity level of the students.
10. The wording of the questions should be clear and unambiguous.
11. It should be a power test rather than a speed test. Allow a liberal time limit so that the essay test does not become a test of speed in writing.
12. Supply the necessary training to the students in writing essay tests.
13. Questions should be graded from simple to complex so that all the testees can answer atleast a few questions.

14. Essay questions should provide value points and marking schemes.

(B) While Scoring Questions:

1. Prepare a marking scheme, suggesting the best possible answer and the weightage given to the various points of this model answer. Decide in advance which factors will be considered in evaluating an essay response.
2. While assessing the essay response, one must:
   a. Use appropriate methods to minimise bias;
   b. Pay attention only to the significant and relevant aspects of the answer;
   c. Be careful not to let personal idiosyncrasies affect assessment;
   d. Apply a uniform standard to all the papers.
3. The examinee's identity should be concealed from the scorer. By this we can avoid the "halo effect" or "biasness" which may affect the scoring.
4. Check your marking sch ial responses.
5. Once the assessment has begun, the standard should not be changed, nor should it vary from paper to paper or reader to reader. Be consistent in your assessment.
6. Grade only one question at a time for all papers. This will help you in minimising the halo effect in becoming thoroughly familiar with just one set of scoring criteria and in concentrating completely on them.
7. The mechanics of expression (legibility, spelling, punctuation, grammar) should be judged separately from what the student writes, i.e. the subject matter content.
8. If possible, have two independent readings of the test and use the average as the final score.

2.3.3.2 Short Answer Type Test

2.3.3.2.1 Meaning of Short Answer Type Test:

The modern trend is to include more short answer questions in the question papers in order to improve their reliability, validity and sampling capacity. Short answer questions generally require exact answers and, although taking many forms, they share the following distinctive features:

(i) They usually take less than five minutes to read and answer, many take less than a minute.

(ii) Short answer questions permit larger sampling of content.

(iii) They tend towards greater objectivity in scoring.

(iv) More reliable and valid than essay questions.

(v) The answer is supplied by the pupil, not pre-selected as in objective questions.

(vi) Precise and specific as to the scope and length of answers.

2.3.3.2.2 Classification of Short Answer Type Test:

They can be grouped into two broad categories:

(a) Extended answer.

(b) Insert and completion.

(a) Extended Answer Type:

The extended answer version includes questions which require pupils to write a brief description, draw a map, make a list, perform a calculation, translate a sentence. Write down a definition or formula and so on.

They are probably the commonest form of questions used in schools and are frequently used in examining Boards. They are deceptively easy to set and usually difficult to mark with any degree of speed and consistency.

Examples 1:

Give the titles of two novels of Prem Chand and for each the names of three of the principal character.

(a) Title:

Character: (i)..... (ii)..... (iii).....

(b) Title:

Character: (i)..... (ii)..... (iii).....

2. Describe briefly two factors which contributed to the outbreak of the First World War. Do not write more than thirty words about each factor.

3. Name the two types of chemical compounds which combine to form salts, and then name one inorganic and one organic salt.

4. Make diagrams to show the structure of

(i) An atom.

(ii) A human cell.

(iii) Methyl alcohol.

5. Write two separate sentences, one containing the word 'eminent' and the other containing 'imminent'.

(b) Insert or Completion Type:

The commonest form of completion questions is one where the pupil is required to add one or two words to complete an incomplete statement correctly. Where the missing words are in the body of the statement to be completed, it is usually called an insert type.

A completion type is where the words are required at the end of the statement. The use of insert or completion questions is not, however, limited to written statements and can be used to prepare extremely good questions based on incomplete maps, drawings, diagrams, formulae, calculations, and the like.

Examples 2:

1. Complete the missing words in this paragraph.

That night there was so little hotel a...tion that they had to take an expensive..... of rooms. After paying the bill they were almost p..... less.

(A useful technique for testing vocabulary and spelling).

2. Complete the following formulae:

Ammonia : N.....

Sulphuric : H.....

Sodium carbonate : ..... CO

2.3.3.2.3 Merits of Short Answer Type Test:

1. Objectivity of scoring can be better ensured in short answer type question in comparison to long answer question.
2. Besides, the question-setter can ask a number of such questions as compared to long-answer type question within the same time limit. Thus, there can be a greater coverage of content (course).
3. They are more reliable than the long-answer type-questions.
4. There is less chance of guessing by the students.
5. Preparation and administration are easy.
6. It is a compromise between the essay and the objective form of test items.

2.3.3.2.4 Demerits of Short Answer Type Test:

There are no such significant demerits of the short answer type test items. Handwriting, language, expression and the way of organization of answer may affect the scores. However, the use of short-answer type tests are to be preferred to long-answer type tests.

2.3.3.3 Objective Type Test

2.3.3.3.1 Meaning of Objective Type Test:

Simply, an objective type test is one which is free from any subjective bias either from the tester or the marker. It refers to any written test that requires the examinee to select the correct answer from among one or more of several alternatives or supply a word or two and that demand an objective judgment when it is scored.

2.3.3.3.2 Objective-Centered Test/Objective based Test:

When questions are framed with reference to the objectives of instruction, the test becomes objective-based. This type of test may contain essay type and objective type test items.An essay test may be objective-centered or objective-based, though it may be difficult to score it objectively. An objective type test, on the other hand, can always be scored objectively, though it may not be objective-centered if it is not planned with reference to the objectives of instruction.

2.3.3.3.3 Objective-type tests have two characteristics viz.:

1. They are pin-pointed, definite and so clear that a single, definite answer is expected.
2. They ensure perfect objectivity in scoring. The scoring will not vary from examiner to examiner.

2.3.3.3.4 Merits of Objective Type Test:

1. Objective type test gives scope for wider sampling of the content.
2. It can be scored objectively and easily. The scoring will not vary from time to time or from examiner to examiner,
3. This test reduces (a) the role of luck and (b) cramming of expected questions. As a result, there is greater reliability and better content validity.
4. This type of question has greater motivational value.
5. It possesses economy of time, for it takes less time to answer than an essay test. Comparatively, many test items can be presented to students. It also

saves a let of time of the scorer.

6. It eliminates extraneous (irrelevant) factors such as speed of writing, fluency of expression, literary style, good handwriting, neatness, etc.
7. It measures the higher mental processes of understanding, application, analysis, prediction and interpretation.
8. It permits stencil, machine or clerical scoring. Thus scoring is very easy.
9. Linguistic ability is not required. :

2.3.3.3.5 Limitations of Objective Type Test:

1. Objectives like ability to organise matter, ability to present matter logically and in a coherent fashion, etc., cannot be evaluated.
2. Guessing is possible. No doubt the chances of success may be reduced by the inclusion of a large number of items.
3. If a respondent marks all responses as correct, the result may be misleading.
4. Construction of the objective test items is difficult while answering them is quite easy.
5. They demand more of analysis than synthesis.
6. Linguistic ability of the tester is not at all tested,
7. Fruiting cost considerably greater than that of an essay test.

2.3.3.3.6 Guidelines for Constructing Better Objective Type Test Items:

To be a good item writer, one should have:

(a) A thorough understanding of the subject matter;

(b) A thorough understanding of the pupils tested;

(c) Perseverance; and

(d) A little creativity to prepare fertile kind of items.

It is of paramount importance for him to be cognizant of the pitfalls involved in writing objective type test items.

We shall now offer some general guidelines for the writing of objective type test items:

1. Each item must be clearly expressed i.e. there must be precision in writing the test items.
2. Test for important facts and knowledge and not for trivial details; e.g.

(a) Give the name of the ship that Columbus was on when he discovered America.

(b) Give the date (and/or time) when Edison invented the light bulb.

These items test the ability to recall or supple trivial details and therefore are unsound.

3. Avoid ambiguous statements. Each item should be subjected to one and only one interpretation.

Poor:

Rabindranath Tagore wrote Gitanjali in..... The item is ambiguous because the examinee does not know whether the teacher wants to know the year, the date, the language or the place.

Better:

In which language did Rabindranath Tagore write Gitanjali?

4. Quantitative rather than qualitative words should be used. Words such as few, many, low, high, large, etc. are vague; indefinite, and, therefore, should be avoided.

Poor:

TF Many people are literate in Odisha.

Better:

TF About 85% of the people are literate in Odisha.

5. Use good grammar and sentence structure to improve clarity.

Poor:

TF In a triangle, whose one of the angle's measure is 90°, the hypotenuse is equal to the square root of the sum of the squares of the other two sides.

Better:

TF In a right-angled triangle, the square on the hypotenuse is equal to the sum of the squares on the other two sides.

6. Avoid lifting statements verbatim from the text-book. The use of text book language in a test encourages a pupil to memorise rather than to understand the subject matter.

7. There should be only one correct answer.

Poor:

Fill in the blank by inserting an operational symbol.

2.....2 = 4

Here, some students may write +, others may write X.

8. Avoid negative questions whenever possible. An indiscriminate use of the negative should be avoided. It takes more time to answer.

TF The longitude of Bombay is not 73°E.

Better:

TF The longitude of Bombay is 73 °E.

9. Directions to questions should be specific. Ambiguous wording and double negatives should be avoided in questions.

2.3.3.4 Oral Type Test

2.3.3.4.1 Meaning of Oral Type Test

Oral test is a face-to-face question-answer activity between the examiner and the examinee. It is generally used by teachers in classroom situations to measure the actual level of knowledge gained by the students.

2.3.3.4.2 Merits of Oral Type Tests

The advantages of oral tests are:

i) Help to probe into pupils insight.

ii) Help to identify and analyze pupils presence of mind.

iii) Pupils' cognitive, affective and psychomotor abilities can be evaluated.

iv) Pupil's spontaneity and mannerism can be evaluated.

2.3.3.4.3 Limitations of oral Type Test

The limitations of oral tests are as follows

i) Oral tests tend to be subjective.

ii) They are usually unplanned.

iii) They are time consuming.

iv) Oral examinations may opens doors of prejudice, partiality and discrimination.

2.4 Personality Test-Rorschach Ink Blot Test

2.4.1 Meaning of Personality Test

Personality of an individual cannot be measured, whereas it can only be analyzed, appraised or assessed. This is because that, personality is an abstract, multidimensional, holistic in assumption, dynamic, complex, individual and situation specific. As a whole, it is the sum total of many traits or qualities possessed to different degrees in different individuals.

Accurate assessment of personality depends on the following: i) Nature of the thing to be measured, ii) Instrument/device to be employed, and iii) the measure. There are three types of methods of assessing personality. They are: i) Subjective, ii) Objective and iii) Projective. In this section we will discuss about projective technique, e.g., The Rorschach Ink-blot Test of assessing personality.

2.4.2 Projective Technique of Assessing Personality

Projective techniques are used when covert or overt or unconscious behavioural dispositions are to be assessed. They are based on the phenomena of projection to test the total personality and not in fragments. A few commonly used and significant protective techniques are: i) The Rorschach Ink-blot Test, ii) Thematic Apperception Test (TAT), iii) Children's Apperception Test (CAT), iv) Picture Completion Test (PCT) etc.

2.4.3 The Rorschach Ink-blot Test

This technique was developed by Herman Rorschach (1884-1922) a Swiss Psychiatrist The material used in this test consists 10 cards on which there are inkblots. Five of them are black and white and the other five are multi coloured. These ink-blots are completely unstructured-the shapes of the blots do not have any specific meaning.

The test is administered in the following sequences:

1) The cards are presented one at a time in a specific order. When the subject gets seated, the examiner gives him the first card with necessary instructions and asks him to say what he sees in it, what it looks like to him, etc.
2) The subject is allowed as much time as he wants for a given card and is permitted to give as many responses he wishes. She/he is also allowed to turn the card around and look at it from any angle he wants.

3) Besides keeping a record of the responses of the subject concerning these inkblots and separate pieces of paper, the examiner notes the time taken for each response. The position in which the cards are being held, emotional expression and other factors of incidental behavior of the subject during the test etc.

4) After all the cards have been presented, the second phase of inquiry which is intended to look for clarification or addition to the original responses follows. The responses are served and analyzed in accordance with the standardized procedure and interpreted in the light of occurrence of certain kind of responses.

2.5 Interest Test-Kuder Test

2.5.1 Meaning of the term Interest

Interest is a term which is commonly used by educationists and laymen. The activity of focusing perception in a single direction is called attention and attention is an important aspect of achieving anything. Interest aids attention. Interest is an enduring attitude which engages the individual's attention to make it selective towards the object of interest.

2.5.2 Key Aspects of Interest

The key aspects of identifying student's interest are: i) Keen observation of their activities, ii) Analyzing the questions they ask, iii) Listening to their conversation, iv) Recording about the books they read, v) Observing their spontaneous drawings, vi) Asking their likes and dislikes, vii) Reading their self-reports about their interests.

2.5.3 Interest Inventory and Kuder's Test

Interest Inventories are available to know the interest of the individuals. Kuder's interest inventory is one such, which records the preferences of an individual. The inventory has forms covering 10 vocational interests, 38 occupational interest areas where 5 are personal interest areas. The preferences are given weight age and the inventory can be quantified by scoring them. The qualities thus yielded can be computed to provide a profile.

2.6 CRC (Cumulative Record Card)

2.6.1 Meaning of Cumulative Records:

Cumulative records are a sort of permanent document of an individual pupil's educational history. This record serves as the basis of reports to other schools, colleges or prospective employers.Basically a Cumulative Record Card is a document in which it is recorded cumulatively useful and reliable information about a particular pupil or student at one place. Hence presenting a complete and growing picture of the individual concerned for the purpose of helping him during his long stay at school. And at the time of leaving it helps in the solution of his manifold problems of educational, vocational and personal-social nature and thus assisting him in his best development.

According to Jones, a Cumulative Record is, "A permanent record of a student which is kept up-to-date by the school; it is his educational history with information about his school achievement, attendance, health, test scores and similar pertinent data," If the Cumulative Record is kept together in a folder it is called Cumulative Record Folder (CRF). If the Cumulative Record is kept in an envelop it is called a Cumulative Record Envelop (CRE). If the cumulative Record is kept in a card it is called a Cumulative Record Card (CRC).

2.6.2 Need for School Record:

The modern type of Cumulative Record was first made available in 1928 by the American Council on education. The need for such a record was felt in view of inadequate information that was contained in the various forms as available. The Secondary Education Commission has made the following observations regarding the need for School records "neither the external examination singly or together can give a correct and complete picture of a pupils all round progress at any particular age of his education, yet it is imparted for us to assess this in order to determine his future course of study or his future vocation."

For this purpose, a proper system of school records should be maintained for every pupil indicating the work done by him in the school from day to day, month to month, term-to-term and year to year. Such a school record will present a clear and continuous statement of the attainment of the child in different intellectual pursuits through-out the successive stages of his education. It will also contain a

progressive evolution of development in other directions of no less importance such as the growth of his interest, aptitudes and personal traits, his social adjustments, the practical and social activities in which he takes part.

2.6.3 Characteristics of Cumulative Record:

The Cumulative Record is characterized in the following grounds:

(i) The Cumulative Record is a permanent record about the pupil or student.

(ii) It is maintained up-to-date. Whenever any new information is obtained about the pupil it is entered in the card.

(iii) It presents a complete picture about the educational progress of the pupil, his past achievements and present standing.

(iv) It is comprehensive in the sense that it contains all information about the pupil's attendance, test scores, health etc.

(v) It contains only those information's which are authentic, reliable, pertinent, objective and useful.

(vi) It is continuous in the sense that it contains information about the pupil from the time he enters for pre-school education or kindergarten system till he leaves the school.

(vii) Whenever any information is desired by any-body concerned with the welfare of the child he should be given the information but not the card itself.

(viii) Confidential information about the pupil is not entered in the CRC but kept in a separate file.

2.6.4 Data contained in the cumulative record card (CRC) should be:

1. Accurate
2. Complete
3. Comprehensive
4. Objective.
5. Usable
6. Valid

(i) Keeping of record is a continuous process and should cover the hole history from pre-school or kindergarten to the college and this should follow the

child from school. The Card will furnish valuable information's about the growth of a child and the new school can place him and deal with him to a greater advantage.

(ii) All the teachers and the guidance workers should have access to these records. Matters too confidential may be kept at a separate place. The child concerned may have an opportunity to study his own Cumulative Record in consultation with the counseller.

(iii) The essential data should be kept in a simple, concise and readable form so that it may be convenient to find out the main points of life of the child at a glance.

(iv) Records should be based on an objective data. They should be as reliable as possible.

(v) The record system should provide for a minimum of repetition of items.

(vi) It should contain reliable, accurate and objective information.

(vii) A manual should be prepared and directions for the guidance of persons, feeling out of using the records given in it.

(viii) The record should be maintained by the counsellor and should not be circulated throughout the faculty for making entries on it by other members of the staff. These entries should made by them on other forms and the entry in this card should be made very carefully by counsellor.

2.6.5 Types of Information Maintained in the-CRC

The types of information which are collected and entered or included in the CRC are as follows;

1. Identification Data

Name of the pupil, sex, father's name, admission No., date of birth, class, section, any other information that helps in easy location of the card.

2. Environmental and Background Data

Home-neighbourhood influences, socio-economic status of the family, cultural status of the family, number of brothers and sisters, their educational background, occupations of the members of the family.

3. Physical Data

Weight, height, illness, physical disabilities, etc.

4. Psychological Data

Intelligence, aptitudes, interests, personality qualities, emotional and social adjustment and attitudes.

5. Educational Data

Previous school record, educational attainments, school marks, school attendance.

6. Co-curricular Data

Notable experiences and accomplishment in various fields-intellectual, artistic, social, recreational, etc.

7. Vocational Information Vocational ambitions of the student.

8. Supplementary Information It is obtained by the use of standardized tests.

9. Principal's overall remarks.

2.6.6 Sources of Collection of Information:

Information about every pupil or child for the maintenance in the CRC should be collected from the following sources

1. Parents or guardian's data form:

Family background and the personal history of the child may be gathered from the parents who are asked to fill in the form.

2. Personal data form

In order to obtain information regarding the pupils interest and participation in extra-curricular activities and his vocational preferences the personal data is of great use. The pupil may be asked to give details of himself. This will supplement the information obtained from the parents data form.

3. School records

These include:

(i) Records of achievement tests.

(ii) Records of other tests.

(iii) Admission and withdrawal record.

4. Other sources:

These include:

(i) Personal visits by the teachers

(ii) Observations made by the teachers.

2.6.7 Maintenance of the Record:

The maintenance of the Cumulative Record Card should begin when the student enters school and should follow the student from class to class within a school and from school to school as he continues his progress.

The class teacher will maintain the Cumulative Record. In view of the fact that he spends much time with the students he will be in a greater position to judge them from different aspects. He will maintain a diary or note-book in which he will note down from time to time his observations about his students. At the end of the year he will make the necessary entries in the Cumulative Record Card (CRC). It is very desirable that he consults his colleagues who also know the pupils. These entries should be made after careful consideration.

2.7 Observation as a technique

2.7.1 Meaning of Observation:

Observation is one of the most ancient and widely used instruments of assessing personality observation have been defined as "Measurement without instruments". In education, observation is the most commonly employed at all measurements technique.

Observation can be done in partially controlled situation.

Observation can be done in free situations too.

2.7.2 Advantages of Observation:

i) As the actual behavior of the child is recorded it is more reliable and objective.

ii) It can be used in every situation.

iii) It is adaptable to the both the individuals and groups.

iv) This method can be used with children of all ages.

v) It can be used with some training and experiences.

2.7.3 Limitations of Observation Technique

i) There is a great scope for personal prejudice

ii) Only overt behavior of an individual is observed. This overt behavior does not provide reliable information regarding the internal process.

iii) Observation is subjected to two kinds of errors e.g., sampling error and observation error.

iv) Problem-solving skill in cognitive domain is difficult to evaluate in this system.

2.8 Self Reporting

2.8.1 Meaning of Self Reporting

Self reporting is an instrument of evaluation which depends upon the individual to be evaluated for opinions regarding his own behavior and traits. Self-reporting includes the description of an individual's likes, dislikes, conflicts and problems. Self reports with their emphasis on self-description play an important role in evaluation. The self-report provides important clues for understanding the individual and his problems.

The self reports are those, which reveal a participant's view of experiences in which he has been involved. The documents are the firsthand account of the experiences of the writer; it is a spontaneous first person description by an individual of his own actions. The self reports have certain advantages. These are: i) The self reports throw light on the motives, ideas and self conceptions of the writer; ii) They are self justification of an individual's experiences. These reports suffer from the limitations of being not an objective document. The self reports may highlight only the positive aspects of a person's personality by hiding the negative side. Self reports are of two types, e.g., i) Questionnaire and ii) Interview.

2.8.2. Interview

A data collection tool for collecting data from the participants directly through a verbal interaction by making them respond to the purposefully framed questions aimed at serving the objectives of the study.

2.8.2.1 Definition of Interview

Tuckman (1972): By providing access to what is "inside a person's head", an interview makes it possible to measure what a person knows, what a person likes or dislikes, and what a person thinks.

Dyer (1995:56): An interview is constructed rather than naturally occurring situation, and this renders it different from an ordinary everyday conversation. .

Best and Kahn (2006:335): The interview is in a sense an oral questionnaire. Instead of writing the response, the subject or interviewee gives the needed information orally and face-to-face (or via the telephone).

2.8.2.2 Types of Interview

1) Structured Interview: It involves the use of predetermined questions which are standardized. The responses got by such interviews will also be structured one.
2) Unstructured Interview: As the name itself indicates the questions asked in such interviews are unstructured; the interviewer is free to ask questions depending on the situations.
3) Non-Directive Interview: This is a type of counseling interview. The purpose of this is to help oneself. Such type of interview aims at getting information from the client about his problems and needs. This type of interview is used in psychotherapy. It expects the interviewer to have certain skills which can help the clients to give their opinions. .
4) Focused Interview: Such type of interview is conducted under some concrete circumstances. This interview is done on the basis of an interview guide in which the field of enquiring and hypothesis are clearly studied. The specialty of the focused interview is that the personal reactions, emotional and intellectual orientation of the persons to be interviewed towards specific issues can be studied,
5) Repeated Interview: Repeated interview is an aspect of directive interview. This type of interview resembles the focused interview. The interview for the same individual or any social unit is repeated several times to study the development process of behavioural changes and changes in attitude.
6) Life History Interview: This type of interview is the method of eliciting life histories by question and answer. A psychologist gathers information from his client by a series of interviews about the life history. Here the interviewer adopts a free association technique in which the respondent tells about himself in a manner chosen by him or he may adopt an active technique in which the interviewer elicits information with the help of an interview guide.
7) Personal Interview: In personal interview, single individual is interviewed.

Personal interview helps to establish close personal contacts between the interviewers and gather detailed knowledge about intimate and personal aspects of the individual.

8) Mass Interview: Mass interview is generally used to collect factual information. Here information is collected by interviewing the mass to know the opinions or many contemporary things of interest.

2.8.2.3 Advantages of Interview

i) The interviews yield a high percentage of returns compared to the questionnaire.

ii) The data collected is reliable as it is done personally contacting the person,

iii) Supplementary information can also be gathered by this method.

iv) The interviewer can observe the facial expressions and gestures etc of the interviewee. This observation helps the interviewer to evaluate the meaning of the verbal replies given by informants as the hesitation, inhibited reactions can be observed.

v) The use of interview method ensures greater number of usable returns compared to other methods.

vi) The individual can be oriented towards the specific topic as needed by the interviewer.

vii) The interview method allows for many facilities which aid on the spot adjustments and thus ensure rich material.

viii) The language of the interview can be adapted to the ability or the educational level of the person interviewed. Therefore, it is comparatively east to avoid misinterpretations or misleading questions.

ix) The interview is a more appropriate technique for revealing information about emotionally sensitive subjects or for probing the sentiments underlying an expressed opinion.

2.8.2.4 Limitations of the Interview

i) It is a time consuming and expensive method.

ii) The success of interview depends on the efficiency of interviewer. The interviewer needs special skills and training for conducting interview.

iii) The adequacy of the data suffers in case the interviewer is biased,

iv) The respondent sometimes may give exaggerated information or account of the presence of the interviewer.

v) Interview needs proper selection and organization of questions,

vi) The interviewer should develop proper rapport with the respondent. Sometimes it may be difficult for the interviewer to do so.

vii) The time selected for the interview should be conducive to the respondents so that they can give information with ease and frankness,

viii) The interview may not be successful in bringing out the hidden feeling of the respondents.

2.8.3 Questionnaire

2.8.3.1 Meaning of Questionnaire

A questionnaire is a device consisting of a series of questions dealing with some psychological, social, educational topics sent or given to an individual or a group of individuals with the objective of obtaining data with regard to some problems under investigation.

2.8.3.2 Definition of Questionnaire

Good and Hatt (1952): Questionnaire may be defined as 'a device for securing answers to a series of questions by using a form which the respondent ; fills in himself.

Barr et al (1953): Questionnaire is defined as 'a systematic compilation of questions that are administered to a sample of population from which information is desired'.

2.8.3.3 Types of Questionnaire

Questionnaire is the most popular means of collecting data in educational research.

Questionnaires are classified into two types: i) Open ended questionnaires ii) Close ended questionnaires.

i) Open ended questionnaires: Open ended questionnaires contain such type of questions which expect the respondent's open response.

ii) Close ended questionnaires: This is called restricted type questionnaire. They include a set of questions to which respondents can reply in a limited

number of ways. The respondent is invariably permitted to reply only with 'yes' or 'no' or no opinion or she/he is requested to select an answer from *a* short list of possible responses and put a tick ( ) mark in the space provided. The investigator can use any one type or a combination of both for collection of data.

2.8.3.4 Advantages of Questionnaires

The advantages of questionnaire are:

i) Questionnaire is helpful in collecting factual information.

ii) With its help, first hand information and opinion can be secured from the persons concerned through their own writing.

iii) It can work as an economic tool for collecting the information from a large number of respondents.

iv) In the administration of a questionnaire on a personal bias the researcher can have a face-to-face interaction with the respondents available at a particular location.

v) In comparison to other tools of data collection like interview and testing devices, questionnaires are more familiar and comprehensible to most of the people.

vi) Questionnaire as a tool of data collection provides greater opportunity to the respondents for providing the informat n their own ways by enjoying full freedom of doing so.

vii) A bid advantage of questionnaire lies in its quality of providing security and secrecy cover.

viii) Questionnaires, especially sent through mail or handed over personally for being returned afterwards, carry a special advantage to the respondents as the respondents can complete it according to their convenience and availability of time,

ix) Written questionnaires enjoy greater possibility of reducing the researcher's bias account of the uniformity lying in the question presentation.

2.8.3.5 Limitations of Questionnaire

i) The questionnaire cannot be used for small children and illiterates.

ii) Sometimes the mailed questionnaire may result in poor response,

iii) The respondents may not also be the representative sample of the group required.

iv) There is a possibility of misinterpretation of items in the questionnaire by the respondents, if goes unchecked in case of mailed questionnaires.

v) There could be inconsistency in responses to items, which measure the same objectives.

2.9 Let us sum up

In this Unit, you studied about the meaning and concept of techniques and tools of evaluation and features of a good tool. You studied the types of tests, their advantages and limitations. You have achieved idea about the key aspects of personality test and interest test. Cumulative record card, a useful data gathering tool for student's profile has been discussed in detail. You also learnt about self reporting technique and associated tools (Interview and questionnaire) for collecting information. Observation technique, a non cognitive tool for student's performance has also been discussed.

2.10 Check your progress (10 questions: 5 SA type, 5 objective type)

A) Short Answer type questions:

1) Distinguish between techniques and tools of evaluation.
2) State the advantages of observation as a technique in evaluation.
3) Write the advantages of CRC.
4) Describe in brief the Rorschach Ink-blot test of assessing personality.
5) Discuss the advantages and limitations of questionnaire and interview.

B) Objective Answer type questions:

1) Teacher made tests are also called ..... (fill in the blanks).
2) Most widely used data collection device is ———— (fill in the blank).
3) Self-reports include the opinions about oneself (Say True otr False).

4) Which of the following is psychological test? a) Interest b) Aptitude c) Intelligence d) Attitude

5) Match the following:

A	B
Reliability	Personal prejudices
Objectivity	Wide coverage
Validity	Consistency
Comprehensiveness	Truthfulness

2.11 References

1. Aggarwal, J.C. *Essentials of Examination System, Evaluation, Tests and Measurement.* Vikash Publication Pvt. Ltd., New Delhi, 1997.
2. Gronlund, W. E. *Measurement and Evaluation in Teaching.* McMillan Publishing Co, New York, 1976.
3. Bhatia, K. K. *Measurement and Evaluation in Education.* Prakash Brothers, Ludhiana, 1973.
4. Dandekar, W.M. *Evaluation in Schools,* Srividya Prakashana, Poova, 1970.
5. Rawat. D.S. *Measurement, Evaluation and Statistics in Education.* New Raj Book Depot, New Delhi, 1970.
6. NSOU Study Material (B.Ed. ODL Programme, *Teacher Functions,* School of Education, Kolkata, 2013.
7. IGNOU Study Material (MESE-057), *Measurement and Evaluation in Education,* School of Education, New Delhi.

Unit:3 ❑ Characteristics of a good tool

Course Content

3.1 Introduction

3.2 Objectives

3.3 Objectivity

3.3.1 Meaning

3.3.2 Nature

3.4 Reliability

3.4.1 Concept

3.4.2 Nature

3.4.3 Causes of low Reliability

3.4.4 Determination of Reliability

3.5 Validity

3.5.1 Concept

3.5.2 Types of validity

3.6 Norms

3.6.1 Meaning

3.6.2 Importance

3.7 Usability/Practicability

3.8 Let us sum up

3.9 Let us check our progress

3.10 References : Books

3.1 Introduction

In all measurements some sort of instruments are needed. A tool can be defined as an implement to facilitate or extend the work of the hand and eye. Generally,

we think of it as something that enables us to perform manual operation more easily and effectively. It adds strength and precision to the work of our hands. In the evaluation of achievement of any subject, tools and techniques are needed to facilitate measuring and recording the characteristics of students. Therefore, it can be said that tools are indispensable for evaluation. In educational measurement, if we tend to find out who is better out of two students, some evaluation tool is definitely required. This is especially due to intangibility of characteristics we are intending to measure.

An effective evaluation, therefore, needs a good evaluation tool. This requirement leads to a number of questions. For example, what constitutes a good evaluation device? What steps should be taken at the time of construction of the test so that it works efficiently? What aspects should be given special attention to? The answer to this questions lead us to constitute the criteria for a good device.

In this unit, let us look at the specific characteristics of a good tool. These will be discussed under the heads: Objectivity, Reliability, Validity, Norms and Practicability.

3.2. Objectives:

After going through the unit we shall be able to :

1. Describe the characteristics of a good tool,
2. Discuss meaning and nature of objectivity,
3. Define the concept of Reliability; discuss the nature, causes of low reliability and determination of reliability,
4. Define the concept of validity, discuss the types of validity,
5. Explain the meaning and importance of norms,
6. Discuss the criteria of usability or practicability of the tool.

3.3. Objectivity:

A test must have the trait of objectivity in terms of both item and scoring. It means it must be free from the subjective elements so that there can be complete

interpersonal agreement among the experts regarding the item and scoring of the test.

3.3.1. Meaning:

Objectivity of items refers to the phrasing of items in such a manner that they are interpreted in exactly the same way by all those who take the test. For that items must have uniformity of order of presentation. It should be placed either in ascending or descending order. On the other hand, objectivity of scoring means that the scoring method of the test should be standard one to maintain the complete uniformity when the test is scored by different experts at different times. The term Objectivity refers to the extent the judgment or opinion of the scorer is eliminated from the scoring process. However, a test is said to be an objective if same scores of results are obtained by administering the test to a particular group or students on different occasions and scored either by the same scorer or different scorer. A test that is objective measures without reference to outside influences. For example, an objective test of personality will return the same answers regardless of whether the person completing the test uses a pen or pencil. Irrelevant, unrelated factors do not influence the test results if a test is objective.

3.3.2. Nature:

In objective type tests the resulting scores are in no way influenced by the mental condition, judgment, personal bias or feelings of the scorer. Items are prepared in such a way that there will be only one correct answer of the scorer. Whoever may administer or score a test on a group of students the test score will always tend to be the same. This particular characteristic is said to be the objectivity of the test. For example, name the present capital city of India. The obvious and one and only correct answer is New Delhi. The score will always be the same under any circumstances.

Therefore, it is essential to preserve and maintain the objectivity of the test. Sometimes it becomes difficult to retain this characteristic. Following are the causes of low objectivity of a test

- Lack of comprehensiveness and too much generalisation reduce the objectivity
- Ambiguity in question framing leads to confusion and misinterpretation which may reduce objectivity of the item.

- Sometimes a student who gives the best answer at the beginning of the test is likely to get more marks in the subsequent questions and vice versa.
- Sometimes a particular student gets marks not only on the basis of his or her writing ability but in comparison to ability of the previous students.
- Lack of uniformity in scoring by the examiners in absence of a scoring key can be detrimental for objectivity.
- Lack of consistency in judgment even among competent examiners may lead to subjectivity.
- Subjectivity may also occur due to application of language, quality of handwriting of the examinee.
- Sometimes the examiners are influenced by the volume and length of the answer rather than depth of the content.

To minimize these subjective elements it is required to reform the conventional essay type questions and to introduce new type of objective tests. The term essay implies a written response that may consist of many sentences to several pages. The students are allowed to enjoy freedom with respect to the content, construction, wording, length and organization. Essay type questions, therefore, definitely have some unique characteristics to judge several qualities of students that otherwise are difficult to evaluate. These are ability to select relevant facts from a body of acquired knowledge, establish relationship between various aspects of knowledge, analysis, synthesis, assimilation, expression, interpretation, organization, critical thinking and so on. In spite of so many advantages essay type questions are discarded due to its lack of objectivity in both item and scoring process. However, some precautions can be taken to reduce the subjectivity of a test:

a. Construction of question: The scope and range of wording should be clear. It is needed to define and restrict the area to be covered by the question. The questions must be related with the instructional objectives of the units in order to achieve maximum content validity. The value points in a question should be intimated to students. Optional questions should not be provided because it reduces the comparability and also difficult to score papers on a common basis.

b. Administration of the test: The test should be administered under proper conditions.

c. Scoring system: An answer key should be provided by the paper setter with which all answers can be compared for evaluation. For each question the factors to be considered in evaluating the response should be given. The students' identity should not be disclosed in order to avoid subjectivity.

d. Interpretation of result: Result should be interpreted objectively.

3.4 Reliability

Reliability refers to the consistency of a measure. A test is considered reliable only if we get the same result repeatedly. If a test is applied to an individual or to a group of students on two or more occasions and the results obtained in all occasions do not differ at all or differ In a very little degree it can be said that the test is a reliable one.

3.4.1. Concept

Reliability in statistics. Education and psychometrics is the overall consistency of a measure. A measure is said to have a high reliability if it produces similar results under consistent conditions. It is the characteristic of a set of test scores that relates to the amount of random error from the measurement process that might be embedded in the scores. Scores that are highly reliable are accurate, reproducible, and consistent from one testing occasion to another. That is, if the testing process were repeated with a group of test takers, essentially the same results would be obtained. Various kinds of reliability coefficients, with values ranging between 0.00 (much error) and 1.00 (no error), are usually used to indicate the amount of error in the scores. For example, measurements of people's height and weight are often extremely reliable. According to Anastasi (1968) reliability refers to the "consistency of scores obtained by the same individuals when re-examined with test on different occasions or with different sets of equivalent items, or under other variable examining conditions". Consistency in results obtained in a single administration is the index of internal consistency of the test. Whereas, consistency in results obtained upon testing and retesting is an index of temporal consistency. Reliability thus includes both internal consistency as well as temporal consistency. A test can be called sound only when it is reliable because reliability indicates the extent to which the scores obtained in the test are free from such internal defects of

standardization that are likely to produce errors of measurement. In other words, the test is free from variable errors. Therefore, a serviceable degree of reliability is another essential criterion of a good test. There are several general classes of reliability estimates:

- Inter-rater reliability assesses the degree of agreement between two or more raters in their appraisals.
- Test-retest reliability assesses the degree to which test scores are consistent from one test administration to the next. Measurements are gathered from a single rater who uses the same methods or instruments and the same testing conditions. This includes intra-rater reliability.
- Inter-method reliability assesses the degree to which test scores are consistent when there is a variation in the methods or instruments used. This allows inter-rater reliability to be ruled out. When dealing with forms, it may be termed as parallel or alternate or equivalent form reliability.
- Internal consistency reliability assesses the consistency of results across items within a test.
- The reliability of a test can also be defined from another angle. Whenever we measure something either in social science or physical sciences the measurement involves some degree of error. The error is caused due to either imperfection in the instrument or environmental and personality factors of examinees being tested. If the measurement is perfectly accurate, free from all kinds of error then the reliability will be perfect. The reliability coefficient will be +1.00. But this is an ideal situation which is rarely achieved either in social science or in physical science. Reliability is never perfect as each measurement contains some errors. Thus, it can be said that each individual score or measurement is made up of two types of scores. These are true score and error score. It can be expressed in terms of an equation:
- $X_T = X_\infty + X_e$
- Where, X_T is the actual obtained score
- X_∞ is the true score (∞ is the sign of infinity which represents the true score)

- X_e is the error score
- The true score is the score,

True score is the score which is free from errors occurring due to chance factor and other kinds of error. It is indicated by the mean of a large number of scores obtained by the same number of person on the same test. Usually, a person's true score remains same but its obtained score may vary with every trial as error score contributes to obtained score in each trial. Error score may be of two kinds. These are random or chance and systematic or constant. Random error scores affect the score sometimes by inflaming or by depressing the score. Therefore, it works randomly in both the positive and negative directions. Chance errors or random errors are also known as the errors of measurement. In the long run, the positive and negative errors tend to cancel each other and therefore the mean of all these errors of measurement will become zero. Systematic errors work constantly in one direction. Therefore, they would either inflate or depress the score. So the mean of such errors of measurement will never become zero. For example, if we are weighing 10 girls who are all 15 years old. The weighing machine may sometimes give high readings or sometimes give low readings due to some mechanical error. If each girl is weighed 10 times we get 100 readings. There is a possibility that high readings and low readings would cancel each other and the mean of these errors becomes zero, leaving the true readings of weight alone. On the other hand, it may happen that the weighing machine is either high reading each time or low reading each time, a case of systematic error. In this case, errors will go on accumulating either on the positive or on the negative side and the mean of such errors would not be zero. When such errors creep into measurement, it is difficult to estimate the true score. When a test is administered many factors are likely to contribute to the error score. Such as, error in scoring, error in test administration, fluctuations in the examinee's motivational and emotional status, guessing, misunderstanding of instruction, misleading items etc. the reliability is directly related to the size of the error score. The smaller the error score the more reliable is the test or the measuring instrument.

Since, any obtained score is divided into the true score and the error score, the total variance of the test is also divided into components-the true variance and

the error variance. Variance is defineds standard deviation squared or SD^2.

However, students should understand any test is neither perfectly reliable nor perfectly unreliable.. Thus, reliability is not an absolute principle, rather it is always a matter of degree.

3.4.2. Nature

Reliability in its simplest sense refers to the precision or accuracy of the measurement or score. A well constructed scientific instrument should yield accurate results both at present as well as over time. In other words, Reliability refers to a type of consistency. This consistency of scores or measurement is reflected in the reproducibility of the scores. A test is said to be consistent over a considerable period of time when all the examinees retain their same relative ranks of two separate testing with the same test. A test is also consistent; if the examinees obtain high scores on one set of items also score high on an equivalent set of items and similarly those who obtain high scores on one set of items also score low on an equivalent set of items.

Reliability is never the property of the test itself. Rather, it is the property of a test when it is administered to the examinees. In other words, it is the property of test scores. Thus, Reliability refers to the results obtained with an evaluation instrument and not to the instrument itself.

The correlation coefficient indicating temporal stability and internal consistency are known as coefficient of stability and coefficient of internal consistency or alpha coefficient respectively. Any statistical measure of reliability must indicate both the coefficient of stability as well as the alpha coefficient. For obtaining the coefficient of stability, the two sets of measurements or scores, found upon testing and retesting are correlated with each other. Similarly, for obtaining the alpha coefficient the two sets of measurements or scores by two equivalent sets of items of the same test after its single administration, are correlated with each other. That is why; Reliability refers to self-correlation of the test

Reliability of a device is the degree to which the device and its scores reflect true or non-error variance. Or it can be said that, it reflects the degree to which the scores are free from chance or random errors. To make the devices more reliable we should avoid the occurrence of chance error in the scores. This nature of

reliability also subsumes objectivity because the latter is due to one type of chance errors at the time of passing judgment. That is why objectivity of measurement is studied with reliability.

Reliability and validity are the two dimensions of test efficiency. Reliability is a matter of stability of test scores whereas; validity is the correlation of the test with certain outside independent criteria. Therefore, Reliability is a necessary but not a sufficient condition for validity, because, tests possessing poor reliability will yield low validity. Again, a test may be reliable but not valid. It may yield consistent score but score need not be representing what exactly we want to measure. Reliability merely provides the consistency that makes validity possible.

3.4.3. Causes of low Reliability:

The Reliability of test scores is influenced by a large number of factors. These factors can be categorized into three heads: Extrinsic, Intrinsic and method used for determining reliability.

1. **Extrinsic factors:** Extrinsic factors are those which lie outside the test itself but tend to make the test reliable or unreliable. These are as follows:
   a. **Group variability:** When the group of examinees being tested is homogeneous in ability, the reliability of the test scores is likely to be lowered. The effect of variability on reliability can be examined by seeing what happens when variability is zero. In that case, individuals receive the same score; standard deviation and Zscore become zero.
   b. **Guessing and chance error:** Guessing in a test is an important source of unreliability. In two alternative response options there is a 50 percent chance of answering the items correctly on the basis of the guess. In multiple choice items the chances of getting the answering correct purely by guessing are reduced. Guessing has two important effects upon the total test score. First, it tends to raise the total score making reliability coefficient high. Second, guessing contributes to the measurement error since the examinees differ in exercising their luck over guessing the correct answer.
   c. **Environmental conditions:** It is preferable to maintain a uniform testing environment. Infrastructure facilities like light, sound and other

comforts should be arranged equally and uniformly to all the examinees. Otherwise, it will tend to lower the reliability of the test scores.

d. **Momentary fluctuations:** Momentary fluctuations in the examinees influence the test score. Hence, they tend to affect reliability. A broken pencil, momentary distraction by the sudden sound of some microphone, anxiety regarding the non completion of test, mistake in giving answer, feeling fatigue or other emotional and physical problem are some of the factors that explain momentary fluctuation in the examinee, leading to lower reliability of test.

2. **Intrinsic factors:** It refers to those factors which lie within the test itself and influence the reliability of the test. These are as follows:

   a. **Length of the test:** The length of the test or size of the sample is directly proportional with the reliability coefficient. Generally, shorter tests lead to lower reliability due to increasing sampling error. Lengthening the test or averaging the total test scores obtained from several repetitions of the same test tends to increase the reliability.

   b. **Homogeneity of items:** Homogeneity of item is an important factor in reliability. The concept of homogeneity of items includes two things.- item reliability and homogeneity of function or trait measured from one item to another. When the test is heterogeneous one the reliability is zero or very low.

   c. **Range of the total score:** If the obtained total scores on the test are very close to each other or if there is lesser variability among them the reliability of the test is lowered. The standard deviation of the total score is low, the reliability is also low.

   d. **Difficulty value of items:** In general items having indexes of difficulty at 0.5v or close to it, yield higher reliability than items of extreme indexes of difficulty. When items are too easy or too difficult the test yields very poor reliability, as these items do not contribute to the reliability.

   e. **Discrimination value:** when items do not discriminate well between superior and inferior, or when items have poor discrimination values

the item total correlation is affected, this leads to low reliability of the test.

f. **Scorer reliability:** It, also called reader reliability, Is an important factor which affects the reliability of the test. It means how closely two or more scores agree in scoring or rating the same set of responses. If they do not agree, the reliability is likely to be lowered.

g. **Ambiguity of language:** Vague language coupled with unstructured nature of tasks presented to the students with no or improperly planned directions for scoring become the root cause of personal errors and subjectivity. The degree of seriousness of the problem obviously varies from subject to subject. For example, it is definitely less serious in Mathematics than in Social Science.

3. Methods of Determining Reliability: Reliability also depends on the method used for determining reliability coefficient. The methods are discussed in the following subunit.

3.4.4. Determination of Reliability:

The ability of a test to give consistence result is called reliability. Unfortunately; it is very difficult to calculate reliability. The goal of estimating reliability is to determine how much of the variability in test scores is due to errors in measurement and how much is due to variability in true scores.

Four practical strategies have been developed that provide workable methods of estimating test reliability.

1. Test-retest reliability method: This method directly assesses the degree to which test scores are consistent from one test administration to the next. It involves the followings:

- Administering a test to a group of individuals
- Re-administering the same test to the same group at some later time
- Correlating the first set of scores with the second

The correlation between scores on the first test and the scores on the retest is used to estimate the reliability of the test using the Pearson product-moment correlation coefficient. This method has its own limitations. Very short time interval

between two administrations of the test may lead to repetition of responses by students due to transfer and memory effect. Again too long interval may lead to changes in the students' behaviour due to growth and maturity effect. Therefore, time interval between two administrations should neither be too short nor too long.

2. Parallel-forms method: The key to this method is the development of alternate test forms that are equivalent in terms of content, response processes and statistical characteristics. For example, alternate forms exist for several tests of general intelligence, and these tests are generally seen equivalent.

With the parallel test model it is possible to develop two forms of a test that are equivalent in the sense that a person's true score on form A would be identical to their true score on form B. If both forms of the test were administered to a number of people, differences between scores on form A and form B may be due to errors in measurement only. It involves the followings:

- Administering one form of the test to a group of individuals.
- At some later time, administering an alternate form of the same test to the same group of people.
- Correlating scores on form A with scores on form B

The correlation between scores on the two alternate forms is used to estimate the reliability of the test. This method provides a partial solution to many of the problems inherent in the test-retest reliability method. For example, since the two forms of the test are different, carryover effect is less of a problem. Reactivity effects are also partially controlled; although taking the first test may change responses to the second test. However, it is reasonable to assume that the effect will not be as strong with alternate forms of the test as with two administrations of the same test.

However, this technique has its disadvantages:

- It isyery difficult to prepare two sets whose items are equivalent. ^
- It may very difficult to create several alternate forms of a test
- It may also be difficult if not impossible to guarantee that two alternate forms of a test are parallel measures

3. Split-half method:

This method treats the two halves of a measure as alternate forms. It provides

a simple solution to the problem that the parallel-forms method faces: the difficulty in developing alternate forms. It involves the followings:

- Administering a test to a group of individuals
- Splitting the test in half
- Correlating scores on one half of the test with scores on the other half of the test

The correlation between these two split halves is used in estimating the reliability of the test. This halves reliability estimate is then stepped up to the full test length using the Spearman-Brown prediction formula.

There are several ways of splitting a test to estimate reliability. For example, a 40-item vocabulary test could be split into two sub tests, the first one made up of items 1 through 20 and the second made up of items 21 through 40. However, the responses from the first half may be systematically different from responses in the second half due to an increase in item difficulty and fatigue.

Reliability of the whole test = 2x reliability on half test/1+ reliability on half test

In splitting a test, the two halves would need to be as similar as possible, both in terms of their content and in terms of the probable state of the respondent. The simplest method is to adopt an odd-even split, in which the odd-numbered items form one half of the test and the even-numbered items from the other. This arrangement guarantees that each half will contain an equal number of items from the beginning, middle, and end of the original test. This method is frequently used by test makers. According to some psychologist split half method is the best method of calculating reliability coefficient.

4. Internal consistency:

This method assesses the consistency of results across items within a test. The most common internal consistency measure is Cronbach's alpha, which is usually interpreted as the mean of all possible split-half coefficients. Cronbach's alpha is a generalization of an earlier form of estimating internal consistency, **Kuder-Richardson formula 20.** They devised a method where single administration of the test was required. They developed some formulas for determining reliability coefficient of a test which is as follows:

$r = N/(N-1)\times\sigma T^2 - \bullet pq/\sigma^2$

$$r = \frac{N}{N-1}[1 - \frac{\Sigma_1^N pq}{\sigma^2}]$$

r = Reliability coefficient of the whole test

σ = Standard deviation of the test score

N = Number of test items included in the test

p = proportion of the group of subjects answering a particular test item correctly

q = proportion of the group of subjects answering a particular test item incorrectly = 1–p

Calculation: p and q should be calculated for each item to get pq for each item. Then all summation of all pq should be derived, *a* can be calculated from the distribution. Now substituting all these values in the above formula Reliability coefficient can be found. Further, a less accurate Reliability coefficient can be calculated by using the **Kuder-Richardson formula 21 as stated below:**

$$r = \frac{N}{N-1}[1 - \frac{\{M(N-M)\}}{N\sigma^2}]$$

Here,

r = Reliability coefficient of the whole test

N = Number of test items included in the test

M = Arithmetic Mean of the test scores

σ = Standard deviation of the test score

These measures of reliability differ in their sensitivity to different sources of error and so need not be equal. Also, reliability is a property of the *scores of a measure* rather than the measure itself and are thus said to be *sample dependent* Reliability estimates from one sample might differ from those of a second sample (beyond what might be expected due to sampling variations) if the second sample is drawn from a different population because the true variability is different in this second population. (This is true of measures of all types-yardsticks might measure houses

well yet have poor reliability when used to measure the lengths of insects.)

Reliability may be improved by clarity of expression (for written assessments), lengthening the measure, and other informal means. However, formal psychometric analysis, called item analysis, is considered the most effective way to increase reliability. This analysis consists of computation of item difficulties and item discrimination indices, the latter index involving computation of correlations between the items and sum of the item scores of the entire test. If items that are too difficult, too easy, and/or have near-zero or negative discrimination are replaced with better items, the reliability of the measure will increase.

3.5. Validity

Validity is another characteristic of a scientific instrument. The term "validity" means truth. Thus, validity refers to the degree to which a test measures what it claims to measure. Validity is not the self correlation of test like reliability; rather it is correlation with some outside independent criteria.

3.5.1. Concept

Validity has been defined in different ways. Anastasi (1968) has defined "the validity of a test concerns what the test measures and how well it does so." Lindquist (1951) has defined validity of a test "as the accuracy with which it measures that which is intended to measure or as the degree to which it approaches infallibility in measuring what it purports to measure." These two definitions reveal the fact that for determining the validity of a test the test must be compared with some ideal independent measures of criteria. Thus, correlation coefficient computed between the test and the ideal measures of criteria is known as the validity coefficient.

As we have already noticed that Validity reflects the degree to which an evaluation device approaches infallibility in measuring in what it intends to measure. It means to determine how valid a tool is one should compare the reality of what it does measure with some ideal conception of what it ought to measure. The measurement should be consistent. Thus, validity has two aspects: Reliability and Relevance. Reliability is discussed earlier. It appears that reliability is a necessary, though not a sufficient, condition for validity. An evaluation tool can not be valid

unless it is reliable, but it may be reliable still not valid. The latter may happen if the tool lacks relevance. Relevance concerns the closeness of the agreement between what the device measures and the function it is used to measure. Generally we need a criterion to define what the evaluation instrument should measure. The validity of the tool can be estimated by the correlation between the raw scores from it and the true criterion scores. The relevance of the test is an estimate of the correlation between the true scores from the evaluation device and the true criterion scores. It is interesting to note that no device is valid for all purposes in all situations and for all groups of students. In fact, validity is always related to the purpose. A device which is valid for one purpose may not be equally valid for any other purpose. Validity is not a problem in physical measurement. The relevance is not questionable because the features and traits are well defined. So, they are concerned with reliability of measurement. But in behavioral science like Education and Psychology, relevance becomes tremendously important as the functions to be measured can seldom be well defined and measurement has to be mainly indirect.

In broader sense, validity is concerned with generalizability. When a test is a valid one its conclusion can be generalized in relation to the general population. From the above discussion it can be said that Validity has three important properties. These are:

- Validity is a relative term. A test is valid for a particular purpose. For example, a test of mathematical ability will be worthless for measuring the knowledge of History.
- Validity is not a fixed property of the test. Validation is a changing and unending process. The old content of the test becomes less meaningful with the discovery of new concepts and formulation of hew meanings. Therefore, they need to be modified radically in the light of the new meaning.
- Validity, like reliability is a matter of degree and not an all-or-none property. A test meant for measuring a particular trait or ability cannot be said either perfectly valid or not valid at all.

3.5.2. Types of validity:

Since there are many purposes of testing, it naturally follows different types

of validity representing each purpose of the test. Generally, there are mainly three main purposes of testing. These are:

- Representation of a certain specified area of content: The tester may wish to determine how an examinee performs at present in a sample of situation that the test claims to represent.
- Establishment of a functional relationship with a variable at present or in future: The tester may wish to predict an examinee's future based on a certain variable. For example, the tester may measure mechanical aptitude of an examinee and predict his future performance in a job of mechanic.
- Measurement of a hypothetical trait or quality: A tester may determine the extent to which an examinee possesses some traits as measured by the test performance. For example, a tester may wish to know the how much an examinee scores on some abstract measures like intelligence.

Corresponding to each of these purposes, there are three main types of validity. These are as follows:

A. **Content or curricular validity:** Content validity is most appropriately considered in connection with the achievement testing. If the achievement test faithfully represents the objectives and reflects the emphasis on both content and objectives during the instruction, then the achievement test possesses content validity. If we have a valid test in English language, it means that the task in the test represent the pertinent body of content in the prescribed course. When we criticize a test in Mathematics not fairly representing the actual content of the course, we are actually opening about the test's content validity. This can be easily estimated by pooling the rational judgments of Mathematics teacher, educators and subject experts. They can analyse the test and estimate how much the test agrees with the instructional plan.

   The measure is represented subjectively after a careful process of inspection comparing the content of the test with the instructional objectives of the course of instruction. The teacher has to match the test items with the content and to check whether all the specific instructional objectives are represented in the test. Therefore, in order to determine content validity, it is convenient to construct a table of specification as shown below.

Table of specification for Achievement Test

Subject: Geography

Class: VI

Unit: Drainage of West Bengal

Sub-units	Instructional Objectives				Total Number
	Knowledge based based	understanding	Application based	Skill based	
1. Introduction of drainage of West Bengal	4	7	3	0	14
2. Southward flowing rivers of West Bengal	6	7	5	5	23
3. East/South-Eastward flowing rivers of West Bengal	7	8	12	5	32
4. Landward/Bay ward flowing rivers of West Bengal	8	8	10	5	31
Total	25	30	30	15	100

The table reflects the sample of learning tasks to be measured. The closer the test items correspond to the specified sample the possibility of getting content validity is greater.

B. Construct validity: Construct validity is the third important type of validity. The term construct validity was first introduced in 1954 in technical recommendations of American Psychological Association. It concerns the extent to which a device tells us something about the trait or construct that we intend to measure through it Construct validation is more complex compared to the other two. Construct validity is computed only when the

scope for investigating criterion related validity and content validity is bleak. Construct validity is concerned with the extent to which a device tells us something about the trait or construct which we intend to measure through it. Since educational achievement is itself an important construct, construct validity is much more involved in achievement testing as content validity. Construct validity is also largely determined on rational basis.

Psychologist N. E. Gronlund suggested that there is no adequate single method of establishing construct validity. It is a matter of accumulating evidences from many sources and its stability depends upon number of evidences acquired. Generally when an experimenter wants to construct psychological test it begins with the working concept of the test which is to be measured here. Then some test items are prepared which will be able to reveal the traits and appears to be relevant. After that the test items are submitted to a group of teachers and subject experts for criticism and judgment. Some of the items may be dropped after selection if not approved by the experts as to be relevant enough. Then all the approved items are to be selected to construct the test. The next thing is to select some valid standardized tests which also measure the same traits. Now both the newly constructed test and the standardized test are applied to a group of students. The test scores of the group of students in the two tests are arranged into two frequency distributions. With the help of a statistical method coefficient of correlation of the two sets of scores is calculated. This gives the value of the new test.

C. **Criterion related validity:** Criterion related validity is a common and popular type of test validity. It obtained by comparing or correlating the test scores with scores obtained on a criterion available at present or to be available in the future. The criterion is defined as an external and independent measure of essentially the same variable that the test claims to. There are two sub types of criterion related validity: concurrent and predictive validity.

D. **Concurrent validity:** concurrent validity occurs when the criterion measures are obtained at the same time as the test scores. This indicates the extent to which the test scores accurately estimate an individual's current state with regards to the criterion. For example, on a test that measures the utility

of objective type test, the test will have concurrent validity if it measured the same thing with currently used Essay type test. To determine concurrent validity of a test a tester will prepare some test items which he thinks relevant. The next step is to select valid standardized test which also measure the same traits. Now both the newly constructed test and selected tests are applied to students. The test scores of the group of students in the two tests are arranged into two frequency distributions. Coefficient of correlation of the two sets of score is computed with the help of the statistical method. This gives the concurrent validity of the new test. This coefficient of correlation is considered , $ as the validity coefficient of the new test. The standardized test with whose scores the score of the new tests are compared is known as criterion.

E. **Predictive validity:** Predictive validity may be defined as the extent to which test performance is accurate in predicting some future performance. In other words, it refers to the ability of a test to predict someone's performance on something. If we design a tool to detect to which student should be encouraged to pursue higher Physics on the basis of their scores on the test, we should concentrate on predictive validity of the test. This can also be determined empirically like the concurrent validity. The only difference is that in concurrent validity we need criterion of current success and in predictive validity we need a criterion of future success. Therefore, examples of tests with predictive validity are aptitude tests helpful in career counseling and guidance. Firstly, a test will be prepared whose predictive validity will be determined. Then it will be applied to a group of students. Next we need to select a valid standardized test which may measure the same traits. After a considerable time this standardized test is applied to the same group of students. The test scores of the group of students in the two tests are arranged into two frequency distributions. Then, the coefficient of correlation of the two sets of scores is computed applying the statistical method. This gives the predictive validity of the new test.

3.6. Norms

A test must also be guided by certain norms. Norms refer to the average

performance of a representative sample on a given test. Norms increases the usability of a test.

3.6.1. Meaning:

Two important definitions are discussed here. According to **Frank S. Freeman,** norm is the average or standard score in a particular test made by a specified population. **Thorndike** and **Hagen** defined norm as the average performance on a particular test made by a standardization sample.

Norms help in interpretation of scores. The term norm conveys several meanings. Some important meanings are discussed in the following paragraph

- It is a statistical procedure to minimise the interpretive error in a test score. It is a device of transforming raw score to standard scores.
- It is meant to determine the position of its score
- Norms are average score or values determined by actual measurement in a group of persons who are representative of a specified population.
- Norms reflects development under condition that may be or often less than optimal
- Norm is the standard level against which the value of certain activity can be compared. The performance or score of an individual is compared against determined level.

There are four common types of norms-age norm, grade norm, and percentile and standard score norms. A test constructor prepares any one of these norms depending upon the purpose and use for his test.

Age norms are based on average scores earned by pupils at different ages. They are interpreted in terms of age equivalent. Test performance is expressed in age level. Age norms are used in mental ability tests (mental age), achievement test in arithmetic (arithmetic age). Age norm represents test performance in unequal units. The school year is divided into 12 months. Therefore, age equivalence for age 5 ranges from 5.0 to 4.11. Age units do not have uniform meaning due to variation in growth pattern in different age levels. Age norms can be more meaningful in elementary level, where growth of children's ability is more regular and continuous, than at the secondary level of education. Age norms do not always represent the child's mental ability as the child's mental age does not always confer its chronological age.

Grade norms are widely used in standardised achievement tests as the tests are employed within academic setting. These educational tests are based on average scores earned by pupils in each grade and interpreted in terms of grade equivalence. Grade equivalence is expressed in two numbers. The first is year and second is month. The school year is divided into 10 months assuming little or no changes occur during vacation. Therefore, grade equivalence for grade 4 ranges from 4.0 to 4.9. During interpretation, we should always remember that grade equivalent indicates the average performance of the pupil in each grade level. For every grade level, 50 percent of the pupils in the standardization groups are above and 50 per cents are below. Therefore, we should not interpret this particular grade norm as something all our students should attain.

The percentile rank of a score is the percentage of scores in its frequency distribution that are equal to or lower than it. For example, a test score that is greater than or equal to 75% of the scores of people taking the test can be said to be at the 75th percentile, where 75 is the percentile rank. Percentile norms are more convenient to use than age norm or grade norm as they are easy to understand. A percentile rank indicates the percentage of individuals who fall below a particular score. The only problem with this norm is percentile units are unequal and therefore, not easily comparable.

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. Standard scores indicate the number of standard deviation units a raw score fall above or below the mean. They can be converted to percentile. They have equal units that are comparable. However, persons having little training in statistics find it difficult to use standard scores.

In a distribution, deviations of the scores from its mean expressed in sigma (symbol of sigma to be inserted) are called standard scores. Observed values above the mean have positive standard scores, while values below the mean have negative standard scores. The standard score is a dimensionless quantity obtained by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. This conversion process is called standardizing or normalizing. Standard scores are also called z-values, z-scores, normal scores, and standardized variables. They are most frequently used to compare

an observation to a standard normal deviate, though they can be defined without assumptions of normality.

Computing a z-score requires knowing the mean and standard deviation of the complete population to which a data point belongs. The standard score of a raw score X is

$$Z = \frac{X - \mu}{\sigma}$$

Where:

μ is the mean of the population.

σ is the standard deviation of the population.

The absolute value of z represents the distance between the raw score and the population mean in units of the standard deviation, z is negative when the raw score is below the mean, positive when above.

3.6.2. Importance:

Norms refers to information regarding the group performance of a particular reference on a particular measure for which a person can be compared to. Norms mean standardized score. Scores on psychological test are most commonly interpreted by reference to norm that represents the test performance on standardization sample. Norms always represent the best performance. Norms increases the usability of a test.

Basically norms are important because there are two purposes of norms:

1. Norms indicate the individual's relative standing in the normative sample and thus permit evaluation of his/her performance in refer to other persons.
2. Norms provide compared measures that permitted a direct comparison of the individual performance on difference test.

Norms help in interpretation of scores. Sir Francis Gallon at the first time developed the logic for norm based testing in the 18th century. Statistical concept of interpreting is discussed here.

1. **Frequency distribution:** A major object of statistical method is to organize and summarize quantitative data in order to facilitate their understanding. A list of 1000 test scores can be an overwhelming sight. In that form it conveys little meaning. A first step in bringing order into such a chaos

of raw data is to tabulate the scores into a frequency distribution. A distribution is prepared by grouping the scores into convenient class intervals and tallying each score in the appropriate interval. When all scores have been entered the tallies are counted to find the frequency, or number of cases, in each class interval. The sum of these frequencies will equal N, the total number of cases in the group.

2. **Graphical representation:** The information provided by a frequency distribution can also be presented graphically in the form of a distribution curve. On the baselines, or horizontal axis, are the scores grouped into class intervals; on the vertical axis are the frequencies or number of cases falling within each class interval. The graph has been plotted in two ways. In the histogram, the height of the column erected over each class interval corresponds to the number of persons scoring in that interval. In the frequency polygon, the number of persons in each interval is indicated by a point in the center of the class interval and across from the appropriate frequency. The successive points are then joined by straight lines.
3. **Central Tendency:** A group of scores can also be described in terms of some measure of central tendency. The most familiar of these measures is the average, more technically known as the mean (M), and it is found by adding all sores and dividing the sum by the number of cases (N). Another measure is the mode, or most frequent score. In a frequency distribution, the mode is the midpoint of the class interval with the highest frequency. A third measure of central tendency is the median or middlemost score when all scores have been arranged in order of size. The median is the point that bisects the distribution, half the cases falling above it and half below.
4. **Variability:** Further description of a set of test scores is given by measures of variability, or the extent of individual differences around the central tendency. The most obvious and familiar way for reporting variability is in terms of range between the highest and lowest score. T range, however, is extremely crude and unstable, for it is determined by only two scores. A single unusually high or low score would thus markedly affect its size. A more precise method of measuring variability is based on the difference between each individual's score and the mean of the group.

A much more serviceable measure of variability is the standard deviation (symbolized by either SD) in which the negative signs are legitimately eliminated by squaring each deviation. The sum of this column divided by the number of cases is known as the variance, or mean square deviation. The variance has proved extremely useful in sorting out the contributions of different factors lo individual differences in test performance. The SD also provides the basis for expressing an individual's scores on different tests in terms of norms.

Developmental Norms

One way in which meaning can be attached to test scores is to indicate how far along the normal developmental path an individual has progressed. Developmental systems utilize more highly qualitative descriptions of behavior in specific functions, such as sensory-motor activities or concept formation.

Mental Age: The term "mental age' was widely popularized through the various translations and adaptations of the Binet-Simon scales, although Binet himself had employed the more neutral term "mental Level". In age scales such as the Binet and its revisions (prior to 1986), items were grouped into year levels. For example, those items passed by the majority of 7-years olds in the standardization sample were placed in the 7-year level, and so forth. A child's score on the test would then correspond to the highest year level that he or she could successfully complete in actual practice; the individual's some tests below their mental age and passed some above it. For this reason, it was customary to compute the basat age, that is, the highest age at and below which all tests were passed. Partial credits, in months, were then added to this basaf age for all tests passed at higher year levels. Mental age norms have also been employed with tests that are not divided into year levels. In such case, the child's raw score is first determined. The mean raw scores obtained by the children in each year group within the standardization sample constitute the age norms for such a test. The mean raw score of the eS-year old children, for example, would represent the 8-year old raw score then her or his mental age on the test is 8 years. All raw scores on such a test can be transformed in a similar manner by reference to the age norms.

Grade Equivalents: Scores on educational achievement tests are often interpreted in terms of grade equivalents. Grade norms are found by computing the mean raw score obtained by children in each grade. Thus, if the average number of problems solved correctly on an arithmetic test by the fourth graders in the standardization sample is 23, then a raw score of 23 corresponds to grade equivalents of 4.

Intermediate grade equivalents, representing fractions of a grade, are usually found by interpolation, although they can also be obtained directly by testing children at different times within the school years. For example, 4.0 refer to average performance at the beginning of the fourth grade. Grade norms are also subject to misinterpretation unless the test user keeps firmly in mind the manner in which they were derived.

Ordinal Scales: Ordinal scales are designed to identify the stage reached by the child in the development of specific behavior functions. Although scores may be reported in terms of approximate age levels, such scores are secondary to qualitative description of the child's characteristics behavior. The ordinarily of such scales refers to the uniform progression of development through successive stages, in so far as these scales typically provide information about what the child is actually able to do(e.g. climbs stairs without assistance; recognizes identity in quantity of liquid when poured into differently shaped containers), they share important features with the domain-referenced tests.

3.7. Usability/Practicability:

Practicability is another important characteristic of a good tool. It stands on a different area from the technical considerations discussed so far, but it has its own significance. Practicability imposes conditions on reliability and validity. In other words, reliability and validity depends on practicability. As we go on increasing the length of the test the reliability of the test goes on increasing. Although we may not be able to increase the length of the test beyond a certain level due to practical limitations. Thus the evaluation tools and techniques are likely to be ineffective if it lacks^ practicability. Some of the aspects of practicability which are to be taken into consideration at the time of construction of the test are economy, administrability, scorability, interpretability, utility and acceptability. These are discussed in details in the following paragraphs.

- **Economy:** Economy related to time, cost and effort should be kept in mind during construction of the test. The device should be undoubtedly be prepared by the teachers. A difficult tool will immediately invite resistance from the teachers for its use. It should not take away too much of teaching time so that the instruction itself suffers. As far as printing and other charges are concerned the device should be well within the easy reach of the budget of the institution.

- **Administrability:** The tools and techniques of evaluation should be easy to administer. They should not require expertly trained personnel to administer the test. It should be teacher friendly. So that the classroom teachers are able to handle the test efficiently. Moreover, the device should be such that it can be administered within the framework of institution's timetable. It should have clear and unambiguous directions both for examiners and pupils. It should be in conformity with the format. ^
- **Scorability:** Objectivity and ease are the two aspects of scorability. Objectivity of scoring is an important part of reliability. Objectivity of scoring is an important part of reliability and therefore, should be taken into consideration while ensuring reliability as already discussed earlier. Steps will have to be taken to make easy scoring in the evaluation tool. This is mainly concerned with degree of expertness and time required for scoring. The degree of expertness is determined by the form of questions. Form of question is controlled by the instructional objectives, based on which the tool is constructed. It also depends on the purposes of evaluation. Therefore, the constructor has to determine the difficulty level of the tasks keeping in mind the purpose, ability of the pupils and instructional objectives of the test . Some steps can be carefully taken to reduce the time for scoring.
- **Interpretability:** Ease of interpretation of the score is another crucial factor that must be considered at the time of construction of the test. A good device possesses interpretability only when its scores can be given meaning in objective terms by the use of derived scores, norms, ratings or other methods. This is an essential feature of standardization of the test.
- **Utility:** An evaluation tool will be having utility if it gives rich dividend to the users for example students, teachers, institutions etc. If the scores can be employed effectively for the guidance of pupils, improvement of instruction, improvement of educational programme, management, administration and betterment of the institution.
- **Acceptability:** Acceptability of the evaluation device among students, teachers, administrators, policy makers and public at large is very important consideration under practicability of a test. Evaluation is visualized as a cooperative process. The test may have an excellent technical and academic base, but if theteachers/ students and other concerned stakeholders express resistance to accept a certain evaluation device, it certainly loses its

practicability. For example, internal assessment procedures, although theoretically technically and academically very sound are not always acceptable to students, teachers and public in general. Acceptability is mainly a function of the environment. The use of certain types of evaluation devices or certain forms of question in them should be governed by the prevalent socio economic condition and reactions. For example, if general atmosphere of malpractice prevails, multiple choice question tests should be avoided as they are highly amenable to code signaling.

3.8. Let us sum up:

The construction of a good tool is very important. Failure to construct a good evaluation tool may result in damage to the educational process, in all measurements some sort of instruments are needed. A tool may be defined as an implement which facilitates or extends the work of the hand and eye. In educational evaluation and measurement a good test is required that is free from variable error, personal error, constant error and interpretive error. Therefore, a good tool must possess objectivity, reliability, validity, norms and practicability.

A test must be free from personal errors by making it objective. It must yield same scores whether it administered by different persons or by the same person at different times on the same individual. If a test does so it is called an objective test. Objectivity must be retained during construction of questions and scoring of items.

A test may suffer from another form of variable error or chance error. Special care should be taken in order to avoid variable errors and to make the test reliable. Reliability therefore, refers to the consistency of a measure. A test is considered reliable if we get the same result repeatedly. If a test is applied to an individual or to a group of students on two or more occasions and the results do not differ at all or in a very little degree, we say that the test is reliable one and is free from variable error. Therefore, some extrinsic and intrinsic factors are crucial for low reliability of a test. There are four methods of determining reliability. These are, test-retest method, equivalent forms method, split half method and Kuder-Richardson method.

A test is said to be valid when it measures what it exactly intends to measure. The validity of a test depends upon the fidelity with which it measures the desired ability. There may have some constant errors in our instrument that may ineffectual

our test scores. For the results to be accurately applied and interpreted it is vital for a test to be valid or free from constant error. There are three types of validity. Content validity, construct validity and criterion related validity. Criterion related validity may further be subdivided into concurrent validity and predictive validity.

Errors of interpretation may also influence score of measurement. In order to enable an examiner to interpret properly the raw score of an individual with the others all tests should be standardized. Therefore, norms are to be calculated for comparing of scores within the group. So norms are important criteria for standardization of the test.

Finally, a good test must have utility and practicability to fulfill the purpose of the test. Practicability depends on economy of time, cost and effort, administrability, scorability, interpretability, utility and acceptability.

3.9. Let us check our progress

Short answer type questions:

1. Name the criteria of a good test
2. Define Reliability.
3. What is content validity?
4. State briefly the relationship between reliability and validity,
5. State the importance of norm.
6. What is mean?

Objective type questions:

1. Deviations of the scores from its mean expressed in sigma (symbol of sigma to be inserted) are called
   I. zed scores
   II. standard scores
   III. central tendency
   IV. T- Scores.
2. Which one of the following indicates item stability?
   I. Correlation by parallel test method
   II. Correlation by Split half method
   III. Correlation by Test-retest method

IV. Correlation by Standard Error method

3. Which one of the following indicates item sampling?
   I. Correlation by Alternative test method
   II. Correlation by Split half method
   III. Correlation by Test-retest method
   IV. Correlation by Standard Error method
4. Name any one extrinsic factor of low reliability.

3.10. References: Books:

- Mangal, S.K. (2010): *Statistics in Psychology and Education,* New Delhi, PHI Learning Private limited, ISBN-978-81-203-2088-8, page no 71
- Oriondo, Leonora Loyola (1984): *Evaluating educational outcomes (Test, Measurement and Evaluation),* Manila, Phillipines, Eleanor M. Dallo-Antonio Rex book store, ISBN 971-23-0330-6, pages 164 to 167,177
- Sarkar, Amal kanti (2010): *Measurement and Evaluation in Education,* Kolkata, India, Rita Book House, page no 48 to 66
- Singh, A.K. (2002): Tests, *Measurements and Research Methods in Behavioural Science,* Patna, Bihar, India, Bharati Bhawan Publishers and Distributors, page no 3 to 87
- Singha, H.S. (1999); *Modern Educational Testing,* New Delhi, India, Sterling Publication Private limited, ISBN812-07-217-13, page no 79 to 99
- Thammishetty, manichander (2016): *Educational Evaluation,* solapur, Maharashtra, india, Laxmi Book Publication, ISBN132-93-1229-5, page no. 53,54

Web references:

- https://en.wikipedia.org/wiki/ReliabilityJstatistics) (visited on 31.10.17)
- *A Primer on Psychology a new approach to an ancient question.* https://funpsychology. wordpress.com/psychological-testing/norms/ (visited on 02.11.2017)
- https://en.wikipedia.org/wiki/Percentile rank (visited on 31.10.17)
- https://en.wikipedla.org/wiki/Standard score (visited on 31.10.17)

Unit : 4 ❑ Evaluation Process

Course Content

4.1 Introduction

4.2 Objectives

4.3 Evaluation Process

- **4.3.1 Concept**
  - **4.3.1.1 Steps of evaluation**
  - **4.3.1.2 Purpose of Evaluation**
- **4.3.2 Types of Evaluation**
  - **4.3.2.1 Formative Evaluation**
  - **4.3.2.2 Summative Evaluation**
  - **4.3.2.3 Difference between Formative and Summative Evaluation**

4.4 Concept of Norm-referenced test and criterion referenced test

- **4.4.1 Norm Referenced Test (NRT)**
  - **4.4.1.1 Meaning and definition of NRT**
  - **4.4.1.2 Chief characteristics of NRT**
  - **4.4.1.3 Process for developing NRT**
  - **4.4.1.4 Reliability of NRT**
  - **4.4.1.5 Validity of NRT**
  - **4.4.1.6 Interpretation of NRT**
  - **4.4.1.7 Uses of NRT**
  - **4.4.1.8 Advantages and limitations of NRT**
- **4.4.2 Criterion Referenced Test (CRT)**
  - **4.4.2.1 Definition of meaning of CRT**
  - **4.4.2.2 Characteristics of CRT**
  - **4.4.2.3 Phases in development of CRT**
  - **4.4.2.4 Reliability of CRT**
  - **4.4.2.5 Validity of CRT**
  - **4.4.2.6 Interpretation of CRT**
  - **4.4.2.7 Uses of CRT**
  - **4.4.2.8 Limitations of CRT**
- **4.4.3 Differences between CRT and NRT**

4.5.1 Concept of Grading
4.5.1.1 Meaning
4.5.1.2 Significance
4.5.1.3 Objectives
4.5.1.4 Mechanism
4.5.2 Concept of Credit System
4.5.2.1 Meaning
4.5.2.2 Advantages
4.6 Let us sum up
4.7 Let us check our progress
4.8 References

4.1. Introduction:

Evaluation is a systematic determination of a subject's merit, worth and significance, using criteria that are governed by a set of standards. It can assist an organization, programme, project or any other intervention or initiative to assess any aim, realizable concept or proposal, or any alternative, to help in decision making or to ascertain the degree of achievement or value in regard to the aim and objectives and results of any such action that has been completed. The primary purpose of evaluation, in addition to gaining insight into prior or existing initiatives, is to enable reflection and assist in the identification of future change. Evaluation is often used to characterize and apprise subjects of interest in a wide range of human enterprises. It is long term and done at the end of a period of time. The contemporary education recognizes evaluation as an integral part of instructional process. Modern educational process has given equal importance to planning, directing and evaluating instruction. Even the most skilful teacher has to evaluate the actual results of his instructional procedure in terms of pupil achievement to carry his task to its ultimate conclusion. Therefore, evaluati6n process, its types, methods will be discussed in this unit.

4.2. Objectives:

After going through the unit, we shall be able to :

1. Describe the concept of evaluation process,
2. Classify and discuss formative and summative evaluation,

3. Define the concept of norm reference test,
4. Define the concept of criterion reference test,
5. Discuss the concept of grading system.
6. Define the concept of credit system in educational evaluation,

4.3. Evaluation Process:

Evaluation is a continuous process. It is concerned with more than the academic achievement of students. It intends in development of the individual in terms of desirable behavioural changes in relation to feeling, thinking and action. Effective learning must bring about some behavioural changes in students. These desired changes are set during planning of instructional objectives. Educators were interested previously in measuring the achievement of students based on students score. Today, educational measurement is considered as an important aspect of evaluation. Educational evaluation is the evaluation process of characterizing and appraising some aspects of an educational process. There are two common purposes in educational evaluation which are, at times, in conflict with one another. Educational institutions usually require evaluation data to demonstrate effectiveness to funding authority and other stakeholders, and to provide a measure of performance for marketing purposes. Educational evaluation is also a professional activity that individual educators need to undertake if they intend to continuously review and enhance the learning they are endeavoring to facilitate.

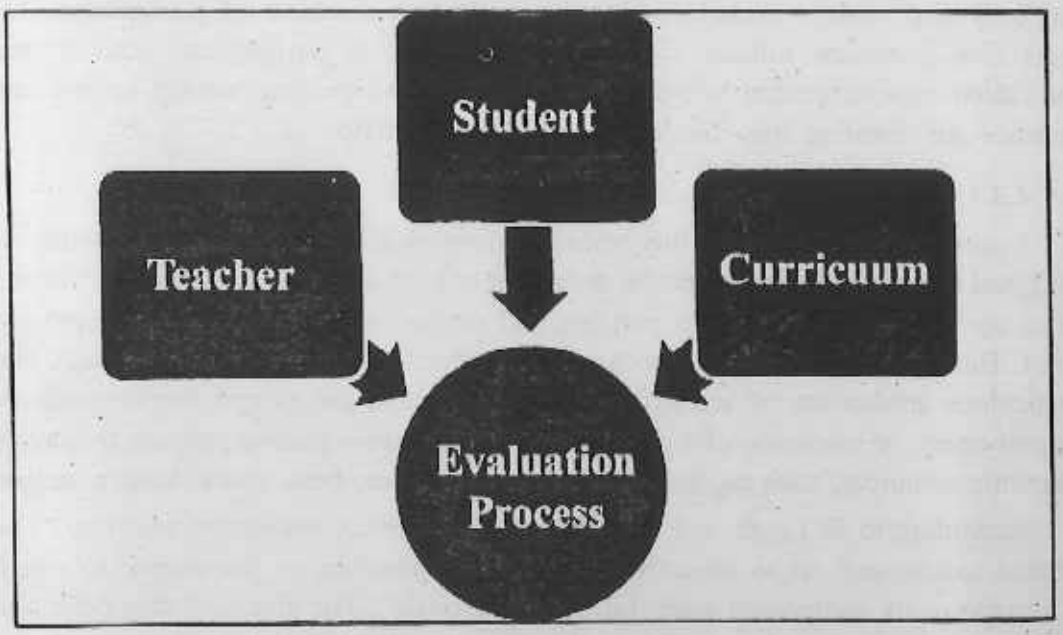
4.3.1. Concept:

Evaluation is inherently a theoretically informed approach (whether explicitly or not), and consequently any particular definition of evaluation would have been tailored to its context- the theory, needs, purpose, and methodology of the evaluation process itself. Having said this, evaluation has been defined as: A systematic, rigorous, and meticulous application of scientific methods to assess the design, implementation, improvement, or outcomes of a program. It is a resource-intensive process, frequently requiring resources, such as, evaluates expertise, labour, time, and a sizable budget.

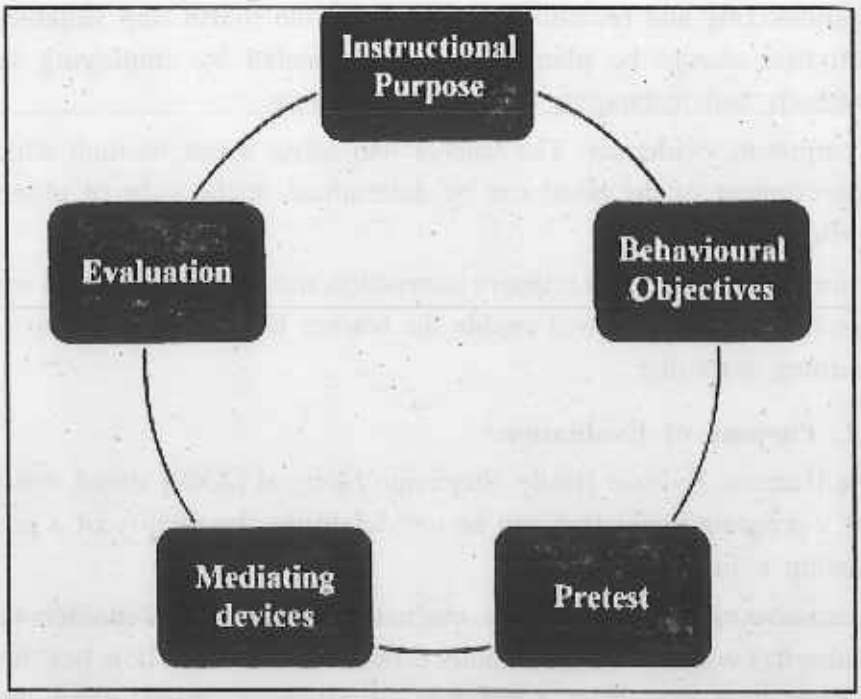
According to St Leger and Wordsworth-Bell (1992), evaluation refers to "The critical assessment, in as objective a manner as possible, of the degree to which a service or its component parts fulfils stated goals". The focus of this definition is on attaining objective knowledge, and scientifically or quantitatively measuring predetermined and external concepts.

According to Stufflebeam (2001) "A study designed to assist some audience to assess an object's merit and worth". In this definition the focus is on facts as well as value laden judgments of the programs outcomes and worth.

Today we do not only test but also evaluate for appraisal of the effects of the learning experiences. Evaluation designates a process of appraisal which involves the acceptance of specific values and the use of variety of instruments of observation, including measurement, as the basis of value judgment. From the functional point of view it involves identification and formulation of a range of major objectives of a curriculum, their definition in terms of students' behaviour and construction of valid, reliable and practical instruments. These instruments or tools are used for observing the specific, phases of students' behaviour such as knowledge, skill, attitude, appreciation, personal and social adaptability, interest and work habits. While the child is acquiring knowledge and skill, learning in attitude, appreciation and interest are also taking place. This approach indicates a shift from a narrow conception of subject matter outcomes to a broader, conception of growth and development of individuals. Evaluation refers to the assessment of students' progress toward stated objectives, efficiency of our teaching as well as effectiveness of the curriculum. Therefore, evaluation has three dimensions, which are represented diagrammatically:



The nature of evaluation as a conscious process is illustrated in following diagram:



4.3.1.1. Steps of evaluation:

Evaluation is the process of determining the extent to which objectives have been achieved. It includes all the procedure used by the teacher, students, administrators and other school personnel to appraise outcomes of instruction. Therefore, evaluation involves the following steps:

1. Formulating general objectives: The first step involved in evaluation is to find out what to evaluate. General objectives should be stated in terms of knowledge, understanding, skills, application, interest, appreciation.
2. Formulating specific objectives: Specific objectives of every topic provide direction to teaching learning process. It determines two things-the various types of learning situations to be provided by the class teacher to the learner and the method to be employed to evaluate both objectives and learning experiences.
3. Securing evidences on the achievement of objectives in selected situations: Framing contents (curriculum, syllabi, and courses) and content analysis are included in the third step. Curriculum framers are philosophers and educationists. Content analysis is done by the teachers.

4. Summarizing and recording evidences: In the fourth step suitable learning activities should be planned and implemented by employing variety of methods and techniques of teaching learning.
5. Interpreting evidences: The teacher will select a test through which actual achievement of the pupil can be determined in the light of objectives set forth.
6. Using interpretation to improve instruction and students progress report: The results of evaluation will enable the teacher to introspect and to recognize learning activities.

4.3.1.2. Purpose of Evaluation:

Marthe Hurteau, Sylvain Houle, Stephanie Mongiat (2009) stated that the main purpose of a program evaluation can be to "determine the quality of a programme by formulating a judgment".

An alternative view is that projects, evaluators, and other stakeholders (including funding authority) will all have potentially different ideas about how best to evaluate a project since each may have a different definition of 'merit'. The core of the problem is thus about defining what is of value. From this perspective, evaluation "is a contested term", as "evaluators" use the term evaluation to describe an assessment, or investigation of a programme, while others simply understand evaluation as being synonymous with applied research.

There are two functions considering to the evaluation purpose. Formative evaluations provide the information on the improving a product or a process. Summative evaluations provide information of short-term effectiveness or long-term impact to decide the adoption of a product or process. Not all evaluations serve the same purpose, some evaluations serve a monitoring function rather than focusing solely on measurable programme outcomes or evaluation findings and a full list of types of evaluations would be difficult to compile. However, the strict adherence to a set of methodological assumptions may make the field of evaluation more acceptable to a mainstream audience but this adherence will work towards preventing evaluators from developing new strategies for dealing with the myriad problems that programmes face.

4.3.2. Types of Evaluation:

Evaluation plays an important role in the field of education. We evaluate

student's skills and abilities in order to select students for special programme and to diagnose student's weakness and strength. We also take the help of evaluation to determine the effectiveness of instruction, usefulness of special programme or a new curricular, and mastery of certain learning objectives. So evaluation can be **formative** that is taking place during the development of a concept or proposal, project or organization, with the intention of improving the value or effectiveness of the proposal, project, or organisation. It can also be **summative,** drawing lessons from a completed action or project or an organisation at a later point in time or circumstance. Honer, M. Scriven (1967) coined the terms of summative and formative evaluation. Evaluation also helps the teachers to report student progress to parents. The report is also used for guidance by counsellors in career counselling. Evaluation helps the researcher to investigate the effectiveness of teaching methods used by classroom teachers. Evaluation can also be diagnostic in nature. Depending on the purpose of evaluation a classroom teacher usually uses four types of evaluation.

- **Placement evaluation:** Placement evaluation occurs before instruction. The objective of placement evaluation is to match students with different programmes. Its function is to place students in programmes that are suitable to their interest and abilities. With the help of standardised test, teachers' observation and previously administered summative evaluations, placement evaluation takes place. For example, a student to be placed in a class for mentally retarded students, an intelligent test will be administered to find out the score. If the student scores below a predefined minimum level of intelligence quotient then that student will be placed in a special class.
- **Formative evaluation:** Formative evaluation of a programme is designed to monitor the instructional process to determine whether the learning is actually taking place as planned earlier. This type of evaluation provides continuous information that can be used to modify the programme to improve its effectiveness and efficiency. In a classroom situation, purpose of formative evaluation is to provide feedback to the teacher and students about progress in unit test, measures of interest and attitude. Interview or conferences with students and parents during the programme can provide important indications for improvement and betterment of the programme by redirecting it. These are mostly teacher made tests. At the end of the

teaching of a chapter the teacher takes a test or quiz. Then the teacher reviews the test result to determine which material requires further discussion.

- **Diagnostic evaluation:** As the name suggests diagnostic evaluation is used to determine an individual's strength and weakness that are left unresolved by the standard corrective prescription of formative evaluation in order to improve performance of the teachers and the students. In this type of evaluation both standardised and teacher made tests are used. This type of evaluation often used to point out the difficulty in learning in spite of using alternative methods of instruction. Gronlund (1985) stated that formative evaluation provides first aid treatment for simple learning problems and diagnostic evaluation searches for underlying causes of those problems that do not respond to the first aid treatment. Therefore, diagnostic evaluation is much more comprehensive and detailed in nature. It depends upon vision test, hearing test and other tests that determine the learning approach of the students. It usually occurs at the beginning of the school year or before a unit of instruction. Its main purpose is to identify students who lack pre-requisite knowledge, understanding, skills so that remedial help can be arranged, to identify gifted learners to ensure that they are being sufficiently challenged and to identify student interest.
- **Summative evaluation:** Summative evaluation usually occurs after instruction. It is designed to determine how well the instructional objectives are met. Formal classroom tests, such as unit test or final examination are the most frequently used tools used in this evaluation. The term summative means summing up of all the available information regarding a programme at its terminal point. Such information can be valuable way of assessing the effectiveness of the whole programme. It also provides correction if the programme continues. The techniques used here are teacher made, evaluation product such as achievement test, research, report, themes, drawing etc. It also depends on rating on different types of performances such as oral report, laboratory experiments etc.

4.3.2.1. Formative Evaluation: Formative evaluations evaluate a program during development in order to make early improvements, help to refine or improve programme. Different scholars have defined formative evaluation. Such as,

According to A. J. Nitko (1983) "formative evaluation is concerned with judgment made during the design and or development of a programme which are directed towards modifying, forming or otherwise improving the programme before it is completed."

Gronlund (1985) defined formative evaluation as "formative evaluation is used to monitor learning progress during instruction and to provide continuous feedback to both pupil and teacher concerning learning successes and failures. Feedback to pupils reinforces successful learning and identifies the learning errors that need correction. Feedback to the teacher provides information for modifying instruction and prescribing group and individual remedial work".

R. L Ebel and D. A. Frisbie (1986) opined that "formative evaluation is conducted to monitor the instructional process, to determine whether learning is taking place as planned."

Gilbert Sax (1989) stated that "formative evaluation takes place during instruction by letting the teacher or evaluator know if students are meeting instructional objectives, if the programme is on time and if there are ways that the programme might be improved. Formative evaluation helps current students to learn more effectively"

W. Wiersma and S. G. Jurs (1990) have written that "formative evaluation occurs over a period of time and monitor students progress."

After going through definitions by different scholars, some characteristic features of formative evaluation are as follows:

1. Formative evaluation is cause seeking.
2. It relatively focuses on molecular analysis.
3. It is interested in the broader experiences of the programme users.
4. Its design is exploratory and flexible in nature.
5. It often ignores the local effects of a particular programme.
6. It seeks to identify influential variables.
7. It requires detailed analysis of instructional material for mapping the hierarchical structure of the learning tasks and actual teaching of the course for a certain period of time.

Following are the implications of formative evaluation for the classroom teacher:

Formative evaluation is done during an instructional programme.

1. The instructional programme should aim at attainment of certain objectives, also during the implementation of the programme.
2. Formative evaluation is used to monitor learning and modifying the programme if needed before its completion.
3. Formative evaluation is used for current students.

4.3.2.2. Summative Evaluation: Summative evaluation provides information on program effectiveness conducted after the completion of the program design. Following definitions are provided by different scholars.

According to A. J. Nitko (1983) "summative evaluation describes judgements about the merits of an already completed programme."

Gronlund (1985) stated that "summative evaluation typically comes at the end of a course or unit of instruction. It is designed to determine the extent to which the instructional objectives have been achieved and is used primarily for assigning course grades or certifying pupil mastery of the intended learning outcomes."

R. L Ebel and D, A. Frisbie (1986) defined that "summative evaluation is conducted at the end of an instructional segment to determine if learning is sufficiently complete to warrant moving the learner to the next segment of instruction."

Gilbert Sax (1989) stated that "a summative evaluation can provide evidence that the programme is satisfactory and should be continued for next year students or that student learning and learning attitudes are so negative that a new programme is needed."

W. Wiersma and S. G. Gurs (1990) have opined that "summative evaluation is done at the conclusion of instruction and measures the extent to which students have attained desired outcomes"

After going through definitions by different scholars, the following elements of the summative evaluation are identified:

1. There must be some instructional programmes prior to summative evaluation.
2. The instructional programme should be set for attainment of some objectives.
3. Summative evaluation is done at the end or completion of a particular programme whose duration may vary from semester to whole year.
4. Summative evaluation should check whether there has been learning or not. If the learning is found to take place the quantity and quality of that learning

in relation to pre-determined objectives should be kept in mind.

5. Summative evaluation provides feedback to the classroom teacher for the success or failure of the programme or of the student.

The followings are the chief characteristics of summative evaluation:

1. Summative evaluation uses well defined evaluation designs.
2. It also focuses on analysis.
3. It provides descriptive analysis.
4. It takes into account the local factors.
5. It is non reactive and unobtrusive as far as possible.
6. It is concerned with broad range of issues.
7. Reliable and valid instruments are used here.

3.3.2.3. Difference between Formative and Summative Evaluations: Initially the terms were applied for evaluation of curricular work only. M. Scriven in his book Evaluation Thesaurus (1980) differentiated between these terms. Formative evaluations are conducted during development or improvement of a programme or product. It is conducted for internal or in-house staff and normally done by an internal or combination of internal and external evaluators. Whereas, summative evaluation is conducted after completion of a programme or a course of study. It is done for the benefit of some external audience or decision maker by an internal or external evaluator or a combination of both. Followings are the main differences between these two types of evaluations:

1. The evaluations differ in purpose, nature and timing.
2. Summative evaluation is the terminal assessment of performance at the end of the instruction, whereas formative evaluation is the assessment made during the instructional phase to inform the teacher about progress in learning and taking necessary steps.
3. The use of record and profile of achievement are frequently used in formative evaluation, lesser used in summative evaluation.
4. The formative evaluation is termed as process and the summative evaluation is termed as product evaluation. Summative evaluation determines the extent to which the examinee has mastered the knowledge and skill associated with the course, but formative evaluation deals with the process by which the examinee achieved these outcomes.

5. In formative evaluation both students and teachers are being evaluated. Whereas, in summative evaluation assessment is done to test the learning outcomes of the students against a set of pre-determined criteria without revealing the route to the teacher which the students followed in reaching the point.
6. Formative evaluation uses a variety of instruments which are either locally developed or standardized. It relies on observation and informal data collection devices. In contrast, summative evaluation tends to use well defined evaluation devices. These are non-reactive, comparative and concerned with broad range of issues. These are publicly accepted, reliable and valid instruments.
7. In formative evaluation scoring is based on criterion referenced approach, but in summative evaluation scoring is normally based on norm referenced approach. Criterion reference test can also be applied in summative evaluation.
8. The method used in formative evaluation of reporting score is individual pattern of pass-fail scores in each task in hierarchy, whereas in summative evaluation attainment is reported in terms of total score.
9. In formative evaluation immediate and continuous feedback is provided to the students. Thus it forces and reinforces learning mastery by providing data that can direct remedial teaching. Summative evaluations are in real sense 'final' tests of students' achievement typically covering large blocks of instructional materials.

Table-1: Broad differences between Formative and Summative evaluations

Criteria	Formative Evaluation	Summative Evaluation
Purpose	To monitor progress of students by getting feedback	To check final status of students
Content focus	Detailed, narrow scope	General, broader scope
Methods used	Daily assignments, teachers' observation	Tests, projects
Tools used	Teacher made test	Standardized test
Frequency	Daily	Weekly, quarterly, monthly, half-yearly, annual term examination

4.4. Concept of Norm-referenced test and Criterion referenced test:

Generally we get a score after administering a test. On the basis of scores higher achievers are praised and lower achievers are identified. Marks obtained by a student in an achievement test are called raw scores. Raw scores do not indicate the merit of a student From raw scores we cannot determine that children in school A are better in English than children in school B. For this comparison we need some broader, uniform, objective and stable standard of reference or unit of measurement in educational and psychological testing. The establishment of such reference point is important for both standardized test and teacher made achievement test. It is important for all sorts of test, like intelligent test, mathematical aptitude, personality test etc. If the scores need value that is meaningful, significant, objective and stable, we require comparing an individual pupil's performance with that of a group, either his own or a standardized test, a group similar to him upon whom the test was developed. A score is not high or low, it is higher or lower than other scores. The better type of units to express test result than raw score of a crude percentage is called norms. In other words, the medium or average performance on standardized test of pupils of a teenage group or a placement or of a school group is termed as Norm. Such average is obtained by administering the test to a large group of students who represent different geographical areas in a country comprising both boys and girls, rural and urban. Among these students there may be children from private and public schools. Such a group will provide a frame of reference for interpreting individual scores. Robert Glaser (1963) originally coined the terms norm-referenced test and criterion-referenced test.

4.4.1. Norm Referenced Test (NRT):

A norm-referenced test is not measured against defined criteria. This type of test is relative to the student group undertaking the assessment. It is a very effective method of comparing students. The IQ test is the best known example of NRT. Many entrance tests are examples of NRT, permitting a fixed proportion of students to pass.

4.4.1.1. Meaning and definition of NRT: Raw scores are expressed in terms of different units, such as number of trials taken within a specified period to reach

a criterion. When we compare an examinee's test score with the core of a specific score of examinees on that test the method is called Norm referenced test. When raw scores are compared to the norms, it becomes scientific. Norms may be defined as the average performance on a particular test made by a standardization sample. A standardization sample is the true representative of the population. It takes the test to provide data for comparison and subsequent interpretation of the test scores. To have an adequate representation the sample must include cross sectional representation of different parts of the population with a view to compare the raw scores with performance of the standardization sample. When these are converted then they are called derived scores.

Gronlund (1985) defined NRT as the tests are "designed to rank students in order of achievement, from high to low, so that decisions are based on relative achievements (e.g., selection, grouping, grading) can be made with greater confidence". Bormuth (1970) stated that NRT is designed "to measure the growth in a student's attainment and to compare his level of attainment with levels reached by other students and norm groups".

This test is primarily used for comparing achievement of an examinee to that of a large representative group of examinees at the same grade level. The representative group is known as the 'norm group'. Norm group may comprise of examinees at the local, district, state or national level. Since the development of NRT is expensive and time consuming, they are produced by commercial test publishers. The following derived scores are generally reported for these published tests:

- Percentile Rank: A number telling the percentage of examinees in a defined group scoring lower than the particular raw score.
- Linear standard score: A number indicating the location of a particular raw score in relation to the mean and standard deviation of a defined group.
- Normalized Standard Score: A number showing the location of the particular raw score in relation to a normal distribution defined in terms of a particular group.
- Grade Equivalent Score: A number telling the grade placement for which the particular raw score is the average for a defined group.

4.4.1.2. Chief characteristics of NRT:

1. The basic purpose of NRT is to measure students' achievement in curriculum based skills.
2. It is prepared for a particular grade level.
3. It is administered after instruction.
4. It is used to form homogenous or heterogeneous class groups.
5. It classifies achievement as above average, average and below average for a given grade.
6. Generally it is reported in the Percentile rank. Linear standard score, Normalized standard score, Grade equivalent score.

4.4.1.3. Process for developing NRT: According to L M. Carey (1988), the NRT development process can be divided into following stages:

1. **Design stage:** This is done through
   I. Curriculum analysis
   II. Selecting objectives to be measured
   III. Analyzing objectives for determining pre- requisite skill
   IV. Developing table of specification for test
   V. Determining specification for item.
2. **Development stage:** This stage consists of the followings:
   I. Writing items following the specifications
   II. Developing needed art work and illustrations
   III. Writing response, directions and examples
   IV. Writing administrative directions
   V. Reviewing item illustration and directions
   VI. Developing test lay out
   VII. Developing simple test.
3. **Conducting Field Test:** At this stage, test is tried out through the following steps:
   I. Selecting representative group
   II. Administering test

III. Scoring

IV. Analyzing information

V. Analyzing data and selecting items

VI. Developing final test form.

4. **Developing Test Norm:** Norms of the test are developed through

I. Describing characteristics of population

II. Selecting representative norm group

III. Administering test to norm group

IV. Scorings

V. Converting raw scores to standard scores

VI. Creating norm tables.

5. **Writing Test Manual:** The test manual is written through

I. Describing the design process and skill measured

II. Describing the field test process

III. Describing the development process

IV. Describing criteria to select item

V. Describing norm group selection procedure

VI. Describing norm group characteristics

VII. Describing test characteristics-reliability and standard error of measurement

VIII. Describing standard administrative procedure

IX. Describing scoring procedure and derivation of standard scores

X. Describing score interpretation procedure.

4.4.1.4. Reliability of NRT: The following factors can increase the reliability:

I. As we know that test length affects reliability, other things being equal, the reliability of the test can be increased by increasing its length.

II. Homogeneous items in content increases reliability.

III. Moderate difficulty value of items increase reliability over the extremely easy or difficult value.

IV. Increased range of performance of the examinees being tested tends to increase reliability.

4.4.1.5. Validity of NRT: The validity of NRT can be increased by;

I. Constructing items of proper difficulty level.

II. Increasing the test length.

III. Increasing the heterogeneity of the group which is being tested.

IV. Administering the test under proper conditions.

4.4.1.6. Interpretation of NRT:

a) If the scores obtained by students deviate largely from expected score by a regular classroom teacher, then the teacher should explore by instructional implication.

b) If the range of scores obtained by the student in different subjects varies largely, the teacher should try to locate the reason for it motivation, special interest and abilities or difficulties.

c) If the test results do not match the expectation of the teacher for individual students the validity of the test is said to be threatened.

d) Small differences in sub-test scores may be due to chance error also for which the standard error of measurement should be used.

e) Information collected from various tests should be used to explain performance on other tests.

4.4.1.7. Uses of NRT: N. Vasantha Ram Kumar and K. N. Lalithamal (1990) stated that NRT is useful in the following cases:

1. In aptitude testing for making differential prediction.
2. To get a reliable rank ordering of the pupils with respect to the achievement that is measured.
3. To identify the pupils who have mastered the essentials of the course more than others.
4. To select the best of the applicants for a particular programme.
5. To find out the effectiveness of a programme in comparison to other programmes.

Many college entrance examinations and nationally used school tests use norm-referenced tests. The SAT, Graduate Record Examination (GRE), and Wechsler Intelligence Scale for Children (WISC) compare individual student performance to the performance of a normative sample. Test takers cannot "fail" a norm-referenced

test, as each test taker receives a score that compares the individual to others that have taken the test, usually given by a percentile. This is useful when there is a wide range of acceptable scores, and the goal is to find out who performs better.

IQ tests are norm-referenced tests, because their goal is to see which test taker is more intelligent than the other test takers. The median IQ is set to 100, and all test takers are ranked up or down in comparison to that level.

Theatre auditions and job interviews are norm-referenced tests, because their goal is to identify the best candidate compared to the other candidates, not to determine how many of the candidates meet a fixed list of standards.

4.4.1.8. Advantages and limitations of NRT: The primary advantage of norm-referenced tests is that they can provide information on an individual's performance on the test compares to others in the reference group.

A serious limitation of norm-referenced tests is that the reference group may not represent the current population of interest. The Oregon Research Institute's International Personality Item Poolwebsite has noted that, one should be very wary of using canned "norms" because it isn't obvious that one could ever find a population of which one's present sample is a representative subset. Most "norms" are misleading, and therefore they should not be used. Far more defensible are local norms, which one develops oneself. For example, if one wants to give feedback to members of a class of students, one should relate the score of each individual to the means and standard deviations derived from the class itself. To maximize informativeness, one can provide the students with the frequency distribution for each scale, based on these local norms, and the individuals can then find (and circle) their own scores on these relevant distributions.

Test items that are answered correctly by most of the pupils are not included in these tests because of their inadequate contribution to response variance. They will be the items that deal with important concepts of course content.

There is lack of congruence between what the test measures and what is stressed in a local curriculum. Norm-referencing does not ensure that a test is valid (i.e. that it measures the construct it is intended to measure).

Norm referencing promotes unhealthy competition and is injurious to self concepts of low scoring students.

Another disadvantage of norm-referenced tests is that they cannot measure

progress of the population as a whole, only where individuals fall within the whole. Rather, one must measure against a fixed goal, for instance, to measure the success of an educational reform program that seeks to raise the achievement of all students.

With a norm-referenced test, grade level was traditionary set at the level set by the middle 50 percent of scores. By contrast, the National Children's Reading Foundation believes that it is essential to assure that virtually all children redd at or above grade level by third grade, a goal which cannot be achieved with a norm-referenced definition of grade level. Norms do not automatically imply a standard.

A norm-referenced test does not seek to enforce any expectation of what test takers should know or be able to do. It measures the test takers' current level by comparing the test takers to their peers. A rank-based system produces only data that tell which students perform at an average level, which students do better, and which students do worse. It does not identify which test takers are able to correctly perform the tasks at a level that would be acceptable for employment or further education.

4.4.2. Criterion Referenced Test (CRT):

When we follow a second method of interpreting a test score is to establish an external standard or criterion and compare the examinee's test score with it. This process is known as Criterion Referencing. In this test there is a fixed performance criterion. When an examinee passes some predetermined number of items or answers them correctly, it is said that he is capable of the total performance demanded by the test. Glaser (1963) first used the term CRT to highlight the need for test that can describe the position of a learner on a performance continuum, rather than the learner's rank within the group. A criterion-referenced test is a style of test which uses test scores to generate a statement about the behaviour that can be expected of a person with that score. Most tests and quizzes that are written by school teachers can be considered criterion-referenced tests. In this case, the objective is simply to see whether the student has learned the material. Criterion-referenced assessment can bo contrasted with norm-referenced assessment. Criterion-referenced testing was a major focus of psychometric research in the 1970s. The word criterion refers to a domain of behaviour. In CRT one is interested in referencing of an examinee's test performance to a well defined domain of behaviour measuring an objective or skill.

4.4.2.1. Definition and meaning of CRT:

We can define CRT as one in which the test performance is linked or related to some behavioural measures. An obvious question may come to our mind is that from where do the criteria come with which a test is referenced? According in Cox and Vargas (1966), a major criterion for referencing a test is the training that increases the skill or proficiency. The test scores of results can be interpreted as an indication of increased skill or proficiency.

A common misunderstanding regarding the term is the meaning *of criterion.* Many, if not most, criterion-referenced tests involve a cut score, where the examinee passes if their score exceeds the cut off smre and fails if it does not often called a mastery test). The *criterion* is not the cut off score; the criterion is the domain of subject matter that the test is designed to assess. For example, the criterion may be "Students should be able to correctly add two single-digit numbers," and the tut off score may be that students should correctly answer a minimum of 80 percent of the questions to

The criterion-referenced interpretation of a test score identifies the relationship to the subject matter. In the case of a mastery test, this does mean identifying whether the examinee has "mastered" a specified level of the subject matter by comparing their score to the cut off score. However, not all criterion-referenced tests have a cut score, and the score can simply refer to a person's standing on the subject domain. The American College Testing (ACT) is an example of this; there is no cut off score, it simply is an assessment of the student's knowledge of high-school level subject matter. Because of this common misunderstanding, criterion referenced tests have also been called standards-based assessments by some education agencies, as students are assessed with regards to standards that define what they "should" know, as defined by the state.

Gronlund, N. E. (1985) stated that CRT is a "test designed to provide a measure of performance that is interpretable in terms of a clearly defined and delimited domain of learning tasks.

Ivon (1970) defines a CRT a;: "one consisting of items keyed to a set of behavior objectives.

Glaser and Nitko in the book 'Educational Measurement' (1971) stated that "a CRT is one that is deliberately constructed to yield measurements that are directly interpretable in terms of special performance standards".

From above definitions it is concluded that CRT is meant to measure the achievement of an examinee on a certain domain to find out his level of achievement in that domain. It has nothing to do with the achievement level of other examinee.

4.4.2.2. Characteristics of CRT:

1. Its main objective is to measure students' achievement of curriculum based skills.
2. It is prepared for a particular course or grade level.
3. It has balanced representation of both goals and objectives,
4. It is used to evaluate the curriculum plan instruction progress and group students' interaction.
5. It can be administered before and after instruction.
6. It is normally reported in terms of
   - I. Minimum scores for partial and total mastery of main skill areas
   - II. Number of correct items
   - III. Percent of correct items
   - IV. Derived score based on correct items and other factors.

4.4.2.3. Phases in development of CRT:

The following phases can be followed for development of CRT:

1) Identification of subject area
2) Selection of unit of topic
3) Delineation of domain and its description
4) Specification of domnin objective
5) External review of steps 3 and 4.
6) Internal review of step 5
7) Framing test items form A and B
8) Internal review of step 7
9) External review of step 7
10) Field try out of test form A and B
11) Internal review of step 10
12) Final test form A and B

13) Using test in the classroom

14) Finding reliability and validity.

4.4.2.4. Reliability of CRT:

Reliability of CRT can be increased by taking the following measures :

I. Test length should be sufficient through to find out test score reliability.

II. The sample of examinees to finding out reliability should be adequate and representative.

III. The reliability information should be provided in the test for each intended use of the test score.

IV. The reliability information provided in the test should be appropriate for the use of the scores of the test.

4.4.2.5. Validity of CRT:

I. The validity evidence should be adequate for the intended use of the test score.

II. The test manual should provide an appropriate discussion on the factor affecting the validity of the scores.

4.4.2.6. Interpretation of CRT:

I. Interpretation should be based on what the items actually measure.

II. For each type of interpretation there should be sufficient items. If these are less than 10 items, they should be combined with other items through lengthening content items.

III. The test should contain both easy and difficult items. Otherwise it would be difficult to describe what low achievers could do.

4.4.2.7. Uses of CRT:

1. To discover the inadequacies in learner's learning and assist the weaker section of learners to reach the level of other students through a regular programme of remedial teaching or instruction.
2. To identify the master learners and average and slow learners in a class.
3. To find out the level of attainment of various objectives of instruction.
4. To find out the level at which a particular concept has been learnt.

5. For better placement of concept at different grade level.
6. To make instructional decisions for teaching a learner in individually prescribed instruction programme.

4.4.2.8. Limitations of CRT:

According to Chase (1974), CRT has following limitations:

1. CRT only shows whether a learner has reached proficiency in a task area but it does not indicate the extent of quality of learner personel of ability.
2. Task incluyded in the CRT may be highly influenced by a given teacher's interest or biases, leading to general validity problem.
3. CRT includes partial areas for listing behavioural objectives and this may be a constructing element for teachers.
4. CRT is important for only a small fractrion of important educational achievements. On the contrary, for promotion and assessment of various skills which are important functions of school need NRT.

4.4.3. Differences between CRT and NRT:

Serial no	NRT	CRT
1	NRT covers a large domain of learning tasks with just a few items measuring each specific task.	CRT facuses on a delimited domain of learning tasks with a relatively large number of items measuring each specific task.
2	It gives emphasis on discrimination among individuals.	It what examines can do and what contain.
3	Contains items of average difficulty.	Contains both easy and difficult items.
4	Interpretation of test needs a defined	Interpretation needs defined as well as delimited achievement domain.
5	In NRT student is tested after each unit of the new material presented.	In CRT the student is tested after each unit for mastery of objectives.

Serial no	NRT	CRT
6	Students are assigned marks of grade to indicate its performance.	Students are showed to proceed to new material if mastery is obtained.
7	A student is allowed to go to the next unit and presented new materials along with the whole class.	A student is allowed to proceed to the next unit if mastery is obtained after remedial instruction.

In India work on CRT is yet in the stage of infancy. There seems to be more emphasis on NRT in the evaluation of students. In spite of the dissimilarities among NRT and CRT, the tests also possess some common elements. Both the tests need reliability, validity, measured achievement domain, similarity of items; same rules are followed for writing items excepting the items of difficulty. Moreover, sample of test items should be relevant and representative in both the tests. Therefore, we cannot develop a clear distinction between NRT and CRT. These two types of test may be considered complementary. The only difference between these two test is CRT is always based upon a predetermined cutoff score whereas a NRT is always based upon the performance of a normative group or standardization sample.

4.5.1. Concept of Grading:

Grading in education is the process of applying standardized measurements of varying levels of achievement in a course. Grades can be assigned as letters generally A through F), as a range (for example 1 to 6), as a percentage of a total number of questions answered correctly, or as a number out of a possible total (for example out of 20 or 100).

4.5.1.1. Meaning: Grading is a means of reporting the result of evaluation. It is a better means than traditional marking system which is very common in use to report result in evaluation. We award individual students for their performance by giving marks. But marks don not indicate anything about the quality of the performance, rather it shows only quantity. Grades provide the quality of performance. Grading is doing either using symbols (A, B, C, D, E) or using verbal description (very good, good, average, poor, very poor). There are different scales which can be employed in grading system.

Five point scale-A, B, C, D and E

Seven point scale-O, A, B, C, D, E and F

The five point scale provides wide variation within the grade points. Whereas, the seven point scale keeps the evaluation within reasonable limits of accuracy and 'spacing' within a grade point.

4.5.1.2. Significance: Grading system is meant to being replaced the traditional marking system. Grading system is superior due to following reasons:

1. The traditional marking system employs 0-100 scale. Based on the performance of a Student, the examiner has to place the student on a suitable point in the 0-100. The logic behind placing a student securing 59 in second division or class and just 60 in first division or class is not justified. The grading system reduces 101 (0-100) point scale to a variable short range of five points or seven point scale.
2. The traditional marking system suffers from lack of objectivity, reliability and validity. There lies 50 percent chance of error in marking Grading helps in removing errors in marking.
3. The lack of reliability clearly indicates that the cut off marks used for award of division or class and for determining pass of fail is completely arbitrary. Yet so much importance is given to the percentage of marks, class or division for the entire academic career of the student. Grading system also solves this problem.
4. There is no uniformity in computing cut off points for declaring students pass or fail or placing them in divisions and classes. Different universities use different cut off points as minimum qualification for entrance examination. This creates a lot of complications during migration from one university to others. On the contrary, grading system provides uniformity in terms of the scale adopted by them.
5. Marking system creates problem for inter subject comparison. For example a score of 60 is pretty good in literature, but the same score may not be considered very well in Mathematics. Grading system helps us to compare the performance of a student in different subjects.

4.5.1.3. Objectives: The grading system has the following objectives:

1. To indicate the quality of performance of students in examination.

2. To ensure greater uniformity in the assessment of students' performance.
3. To facilitate better comparability of students' performance in various subjects and among various universities.
4. To facilitate the inter university migration of students.
5. To enable students to make correct choice of subjects in accordance with their abilities.
6. To remove the defects of marking system.

4.5.1.4. Mechanism: In some countries, all grades from all current classes are averaged to create a **grade point average (GPA)** for the marking period. The GPA is calculated by taking the number of grade points a student earned in a given period of time of middle school through high school. GPA's are also calculated for undergraduate and graduate students in most universities. The GPA can be used by potential employers or educational institutions to assess and compare applicants. **A cumulative grade point average (CGPA)** is a calculation of the average of all of a student's total earned points divided by the possible number of points. This grading system calculates for all of his or her complete education career. Grade point averages can be unweighted {where all classes with the same number of credits have equal influence on the GPA} or weighted {where some classes are given more influence than others}.

4.5.2. Concept of Credit System:

A credit system is a systematic way of describing an educational programme by attaching credits to its components. The definition of credits in higher education systems may be based on different parameters, such as student workload, learning outcomes and contact hours.

4.5.2.1. Meaning: The credit system is basically a system for simplifying the process of 'academic book keeping'. A student earns his diploma by accumulating a specified number of academic credits instead of attending for a specified number of years. Thus, a part time student may spread out his studies over several more years than normal, keeping beside his other activities. While a brilliant student may take a heavier load than normal and shorten his education by a semester or so. A university usually prescribes that a certain number of minimum credits must be obtained in the students' main department. The distribution of remaining credits may be unrestricted or may be restricted in accordance with any scheme which

the university may desire. Credit system has the effect of adding considerably to the flexibility allowed by the semester organization of the academic schedule. The semester credit hour is defined as a unit for expressing quantitatively the content of a course at the level of higher education.

The most widely accepted meaning of a credit is that a student earns one credit by attending one period of 50 minutes of lecture class instructions and doing 2 to 3 hours of home study, each week though out one semester {15 to 18 weeks}. Thus the terms credit and semester hours are often used interchangeably. The credit system allows greater flexibility in fitting course time to the particular subject taught. Most semester courses carry 2 to 4 credits each. Credits do not always require a particular number of class periods. Since one hour of theory class is expected to be supplemented by 2 to 3 hours of home study, it is considered equivalent to 3 hours of laboratory practical work. Similarly, 2 to 3 credits per week may be allowed for a thesis that involves no class attendance at all. A full time student at a normal pace will complete 30 to 32 semester credit hours of course work in an academic year of 9 months. To obtain a Bachelors degree in a three year college, students must normally accumulate 90 to 94 semester hours. Another important feature of this system is greater reliance on home study.

The most important feature is perhaps the flexibility it permits to students intensively as well as extensively to cover a wide range of subjects, coping up with needs and abilities of the individual students. The selection of course is done with utmost care and in consultation with the adviser concerned. It is also ensured that nearly 2/3rds of course credits is from the subject in whom the students intend to major at the end of first 2 years of general education programme. The students may also allow offering some courses for which no credit may be given. Generally, a student is required to offer not less than 9 credit courses and not more than 16 credit courses in one semester.

4.5.2.2. Advantages:

1. The flexibility it permits to students to cover a wide range of subjects is perhaps the most important advantage of credit system in education.
2. The course can be adjusted keeping in mind the need and abilities of the individual student.
3. The selection of course is done with utmost care and in consultation with the advisor who ensure that preparatory courses are taken in the first or second semester.

4. It is beneficial for the part time students who can spread out his studies over several more years than normal according to his convenience.
5. Especially brilliant students also may shorten his education by a semester based on his need and ability.
6. Credit do not always require any particular class period.
7. Finally, an important advantage associated with credit system is ample opportunity for home study.

4.6. Let us sum up:

Evaluation is a continuous process that is designed to provide information that will help us make judgment about a given situation. It pertains to the estimation of quality. In the instructional context this term is used widely to indicate the quality of learning. Evaluation process has the purpose of enabling each pupil to actualize one's capacities to the optimum possible context and to diagnose his strength and weakness with a view to further guidance. Sometimes the purpose of evaluation is to predict the educational practices which a particular student teacher can participate in or organize.

Based on the. purpose, evaluation can be of four types: Placement, Formative, Diagnostic and Summative. To enable people and agencies make judgment about the work undertaken, to identify their knowledge, attitude and skill, and to understand the changes that have occurred in these and to increase their ability to assess their learning and performance formative evaluation is used. This type of evaluation is an ongoing classroom process that keeps students and educators informed of students' progress towards progress learning objectives. The main purpose of formative education is to improve instruction and student learning. It provides teachers with information which can be used to modify instructional objectives. Students are provided direction for future learning and are encouraged to take responsibility for their own progress.

On the other hand, summative evaluation is used to enable people and agencies to demonstrate that they have fulfilled the objectives of the programme or to demonstrate that they have achieved the standard required. Summative evaluation occurs most often at the end of the unit of study. Its primary purpose is to determine what has been learned over a period of time, to summa rise student progress, and

to report on progress relative to curriculum objectives to students, parents and educators.

Students should be compared in terms of norm. In its simplest form, norm of a certain test implies average performance on that test. Generally there are two reference points that are applied in interpreting the test scores. A norm-referenced test (NRT) is a type of test, which yields an estimate of the position of the tested individual in a predefined population, with respect to the trait being measured. The estimate is derived from the analysis of test scores and possibly other relevant data from a sample drawn from the population. That is, this type of test identifies whether the test taker performed better or worse than other test takers, not whether the test taker knows either more or less material than is necessary for a given purpose. The term normative assessment refers to the process of comparing one test-taker to his or her peers. Norm-referenced assessment can be contrasted with criterion-referenced assessment. In a criterion-referenced assessment, the score shows whether or not test takers performed well or poorly on a given task, not how that compares to other test takers. The same test can be used in both ways.

As alternatives to normative testing, tests can be criterion-referenced assessments. In criterion-referenced assessment, the individuals' performance is compared only to their previous performances. For example, a person on a weight-loss diet is judged by how his current weight compares to his own previous weight, rather than how his weight compares to an ideal or how it compares to another person.

A test is criterion-referenced test (CRT) when the performance is judged according to the expected or desired behavior. Tests that judge the test taker based on a set standard (e.g., everyone should be able to run one kilometer in less than five minutes) are criterion- referenced tests. The goal of a criterion-referenced test is to find out whether the individual can run as fast as the test giver wants, not to find out whether the individual is faster or slower than the other runners. Standards-based education reform focuses on criterion- referenced testing. Most everyday tests and quizzes taken in school, as well as most state achievement tests and high school graduation examinations, are criterion-referenced. In this model, it is possible for all test takers to pass or for all test takers to fail. In CRT emphasis is given on improvement of students' achievement, whereas, in NRT measurement of achievement is more important. In India terminal examination system based on NRT still holds good.

The traditional examination system is being replaced by modern evaluation processes. Grading is one of the important methods of replacing the traditional marking system. Grades describe the level of achievement. It addresses the limitations on marking system. Marks denote a qualification of a performance, while grades denote qualitative banding of clustering of performance. The grading can be done by using five point or seven point scale.

Semester credit system is another important method of modern evaluation process. The semester credit hour is defined as a unit for expressing quantitatively the content of a course at the level of higher education.

4.7. Let us check our progress

Short answer type questions:

1. State the purposes of evaluation.
2. Mention the steps for evaluation process.
3. Differentiate between formative and summative evaluation.
4. What is NRT?
5. Write down the significance of grading system.
6. How many credit hours are generally offered to a student in one semester?

Objective type questions:

A. Choose the correct alternative:

1. Term-end examination is an example of
   a. placement evaluation
   b. summative evaluation
   c. diagnostic evaluation
   d. formative evaluation
2. IQ test is based on CRT (T IF)

B. Fill in the blanks:

1. All grades from all current classes are averaged to create a..... for the marking period.
2. A number telling the percentage of examinees in a defined group scoring lower than the particular raw score is.....

4.8. References:

- Aggarwal, J.C. (2001). Essentials of Examination System: Evaluation, Tests and Measurements, Vikash Publishing, Delhi, ISBN 9788125900511, pp. 73-76, 190-195, 280-291.
- Bormuth, J.R. (1970). On the Theory of Achievement Test Items. University of Chicago Press, Chicago.
- Carey, L.M. (1988). Process for Developing Norm-Referenced Achievement Test. A.P.H. Publishing, New Delhi.
- Dash, B.N., Dash, Nibedita and Rath, M.P. (2003). A Textbook of Educational Psychology, Delhi, Dominant Publishers, pp. 483 to 492.
- Dash, M. (2004). Education in India: Problems and Perspectives, Atlantic Publishers & Distributors Pvt Ltd, Delhi, ISBN-10: 8171568815, pp. 204-208.
- Ebel R.L. and Frisbie D.A. (1986). Essentials of Educational Measurement, 4th Ed. Englewood Cliffs, NJ: Prentice-Hall.
- Glaser (1963). Assessment and Evaluation of Developmental Learning, Praeger, London. Glaser and Nitko(1971). in R. L. Thorndike (Ed.), Educational Measurement. (2nd. ed.) Washington, D.C.: American Council on Education, pp.443-03.
- Gronlund (1985) Measurement and Evaluation in Teaching. MacMillan, New York.
- Honer, M. Scriven (1967). Perspectives of Curriculum Evaluation. American Educational Research Association Monograph Series on Curriculum Evaluation, Vol. 1.
- Kumar, N. Vasantha Ram and Lalithamal, K. N. (1990). Evaluation and Measurement, D.P.H., New Delhi.
- Marthe Hurteau, Sylvain Houle, Stephanie Mongiat (2009). Educational Evaluation; Judgment (Psychology); Vol. 15 Issue 3, pp. 307-319.
- Nitko A. J. (1983) Educational Tests and Measurement, Harcourt Brace Jovanovich, New York.
- Sarkar, Amal Kanti (2010). Measurement and Evaluation in Education, Kolkata, India, Rita Book House, pp. 22-26, 217-219.

- Sax, Gilbert. (1989). Principles of Educational and Psychological Measurement and Evaluation (3rd ed.). Belmont, CA: Wadsworth, p.678.
- Siddiqui, Majibul Hasan (2009). Educational Evaluation, A.B.H. Publishing Corporation, New Delhi.
- Singh, A.K. (2002). Tests, Measurements and Research Methods in Behavioural Science, Patna, Bihar, India, Bharati Bhawan Publishers and Distributors, pp. 3 - 87.
- Singh a, H.S. (1999). Modern Educational Testing, New Delhi, India, Sterling Publication Private limited, ISBN812-07-217-13, pp. 79 - 99.
- St Leger, A.S., Shneiden, H. and Walsworth-Bell, J.P. (1992). Evaluating Health Services' Effectiveness, Buckingham, Open University Press.
- Stufflebeam (2001). The Metaevaluation Imperative. American Journal of Evaluation. 22; 183.
- Wiersma, W. and Jurs, S.G. (1990). Educational Measurement and Testing. 2nd Editon. Boston: Allyn and Bacon.

Web pages:

- https:!len.wikipedia.org/wiki/Educational evaluation (visited on 17.11.2017)
- https://en.wikipedia.org/wiki/Grading (education) (visited on 17.11.2017)
- https:!len.wikipedia.org/wiki/Criterion-referenced test(visited on 17.11.2017)
- https://en.wikipedia.org/wiki/Norm-referenced test(visited on 17.11.2017)
- https:!len.wikipedia.org/wiki/Evaluation#Methods and technigues(visited on 17.11.2017)
- https://www.upjs.sk/en/university/education/credit-system/(visitedon 17.11.2017)

Unit : 5 ❑ Educational Statistics

STRUCTURE:

5.1 Introduction

5.2 Objectives

5.3 Concept, Scope and Need of Educational Statistics

5.3.1 Concept

5.3.2 Scope

5.3.3 Need

5.4 Organization and Tabulation of Data-Frequency Distribution Table

5.4.1 Organization of data

5.4.2 Tabulation of data

5.5 Graphical Representation of Data-Pie diagram, Bar Graph, Histogram, Frequency Polygon and Ogive (Using 75% rule)

5.5.1 Pie diagram

5.5.2 Bar Graph

5.5.3 Histogram

5.5.4 Frequency Polygon and

5.5.5 Ogive (Using 75% rule)

5.6 Let Us Sum Up

5.7 Check our progress

5.8 References.

5.1 Introduction

Data in raw form may have little meaning to the researcher until they have been classified, organized and analyzed with the help of various frequency distribution table or graphical presentation. In this unit we shall be able to understand the meaning, scope, and need of educational statistic; the tabulation of data in a meaningful way and various types of graphical representation to make the data easily comprehensible.

5.2 Objectives

After going through this unit, we should be able to:

- understand the concept, scope, and need of educational statistic;
- understand the meaning and nature of data;
- organize and tabulate the data-frequency distribution table; and
- explain the methods of graphical representation of data with the help of pie diagram, bar graph, histogram, frequency polygon and ogive (Using 75% rule).

5.3 Concept, Scope and Need of Educational Statistics

For better understanding the educational statistics, it becomes necessary to study about the various aspects of educational statistics as given below:

5.3.1 Concept

5.3.2 Scope

5.3.3 Need

5.3.1 Concept

Statistics is a familiar word now-a-days. The word **'statistics'** has been derived from the Latin word **'status'**. It has two different meanings. In the plural sense the word implies a set of numerical figures called **'data'** obtained by counting, or, measurement. In the singular sense 'Statistic' refers to the subject of scientific activity which deals with the theories and method of collection, classification, presentation, analysis, comparison, and meaningful interpretation of **'raw data'**.

In order to understand the meaning of statistics few definitions of different authors are stated below:

'Statistics are numerical statement of facts in any department of enquiry placed interrelation to each other'.- Bouly

Croxton and Cowdon defined it as 'the science which deals with the collection, analysis and interpretation of numerical data'.

Lovitt: 'Statistics is that which deals with the collection, classification and tabulation of numerical facts as the basis for explanation, description and comparison

of phenomena.'

Webster: 'Statistics are the classified facts representing the conditions of the people in a state especially those facts which can be stated in numbers or in tables of numbers or in any tabular or classified arrangements.

Educational statistics is the measurement and evaluation of teaching learning process. In this process we obtain scores and then interpret these scores in order to take decisions. Educational statistics enables us to study these scores objectively. It makes the teaching learning process more efficient. : ,

5.3.2 Scope

Educational statistics is not a mere device for collecting numerical data, as a means of developing sound techniques for their handling, analyzing and drawing valid inferences from them. Educational Statistics is applied in every sphere of educational process (i.e. planning, implementing, and evaluating) for upHftment of education.

5.3.3 Need

The knowledge of educational statistics is needed for the following purposes:

a) Statistics presents the data in a precise and meaningful way and express it in a understandable and interpretable manner.

b) Teachers can get the most exact type of description of the pupil's performance or trait from educational statistics.

c) It is needed to the teacher for predicting a thing happens under conditions we know and have measured.

d) It gives complex data in a simplified manner and helps us to study the trends and relationships of different phenomena and compare them.

e) Educational statistics needs to draw conclusions. It helps us about how much faith should be placed in any conclusion and about how far we may extend our generalization.

f) It is needed for analyzing some of the causal factors underlying complex and otherwise bewildering events.

g) It needs for classifying numerical data, measuring uncertainty, testing the hypothesis, formulating policies and taking valid inferences.

h) It needs for predicting the future plan of action of the phenomena.

5.4 Organization and Tabulation of Data-Frequency Distribution Table

For statistical analysis and graphical representation of raw scores, the following two steps have to be followed:

5.4.1 Organization of data ;

5.4.2 Tabulation of data

5.4.1 Organization of data

For organization of data, it becomes necessary to study about various forms of data as shown below:

a) Primary and Secondary data

b) Qualitative and Quantitative data

c) Discrete and Continuous data

d) Temporal and Spatial data

a) Primary and Secondary data: When data is collected by a particular organization or person from the primary source for specific purpose of use, called as **'Primary data'.** The various methods of collection of primary data are: (/) Mailed questionnaires; (ii) Results of experiments; (iii) Data from local agents and correspondents; (iv) Direct personal investigation (interview/observation); (v) Indirect oral investigation; (vi) Questionnaires to be filled in by enumerators, etc. Data collected in this manner are called **'raw data'.**

Data collected and published by one organization and subsequently it is used by other organizations are called **'Secondary data'.** The various sources of collection for secondary data are: periodicals and newspapers; official publications of local, state, central and foreign governments, publications of research papers published by different university departments; etc. Secondary data should be used with care. These are generally voluminous and have to be arranged properly before use.

The raw data does not give a clear picture as they are collected in real situations and arranged haphazardly. For this reason we have to organize or classify the data systematically according to their similarities for reducing our mental strain. Classification make the data condense by dropping out unnecessary details. It

provides the data for comparison between different sets of data clearly marking the different points of agreement and disagreement. It gives us for studying the relationship between several traits and make further statistical treatment like tabulation, etc. e.g. During population census, people in our country are classified according to their sex (males/ females), Age (0-5 years, 6- 10 years, 11-15 years, etc.), marital status (married/unmarried), place of residence (rural/urban), profession (agriculture, transport, doctor, production, commerce, others), residence in states (Mumbai, Delhi, West Bengal, Bihar, etc.), etc. After organization of raw data into frequency distributions or into another form, called group data.

b) Qualitative and Quantitative data: Qualitative data is related to attributes or non-measurable characteristics; like nationality, social status, sex, occupation, etc. For example, the population of the whole state can be divided into four categories as widowed, divorced, married, and unmarried. Such data is called as **categorical or qualitative data.** When only one trait, e.g., social status, is used for classification, it is called **simple classification.** When more than one trait, e.g., sex, deafness, and religion, are used for classification, it is called **manifold classification.**

The **quantitative data** gathered with the help of various psychological and educational tests are expressed in interval scale of measurement. This data are unorganized and hence, a researcher is required to organize/classify them in a systematic manner for analysis and interpretation. It is done according to numerical size like weights in kg or heights in cm.

Let us consider a set of data given in Table 5.1

Table 5.1: Presentation of qualitative and quantitative data

Qualitative data		Quantitative data	
Management	No. of Schools	Enrolment	No. of Schools
Government	5	Up to - 50	24
Private	10	51-100	15
Private Aided	12	101-200	29
Local Body	3	201-300	12
Total	30	Total	80

From the above table it is shown that management of school has been classified

into 4 categories such as Government, Private, Private Aided, and Local Body. Again number of schools has been shown according to their enrolment of students. The first one is based on category and the second one is based on enrolment i.e. numbers.

c) Discrete and Continuous data: Numerical or quantitative data may be continuous or discrete depending on the nature of the elements or being observed.

Discrete data denotes the exact values and not any fractional values. Number of births in a certain year, number of workmen in a factory, members of a family, students in a class, number of telephone calls in a month, etc., are examples of discrete data.

Continuous data represents any numerical value (fractional/integral) within a certain range. Height, rainfall, weight, time, temperature, etc., are examples of continuous data. Age of students in a school is a continuous variable as it can be measured to the nearest fraction of time, i.e. days, months, years, etc.

d) Temporal and Spatial data: Temporal data is related to time e.g., historical events arranged over a period of time, sex ratio of a state for several years, purchase and sales of a company for five years, etc.

Spatial data denotes space or places, e.g., production of wheat in quintals in various districts, population growth of a state according to districts, etc.

5.4.2 Tabulation of data

When statistical data are presented in a tabular form then it is called data tabulation. The data should be in a frequency distribution form. Frequency distribution is a set of classes together with the number of occurrences [frequencies] of scores in each class in a given set of data or it is grouping of data into mutually exclusive classes showing no. of observations in each class. Frequency is statistical data that may consist of a list of numbers which may be repeated twice and even more than twice related to any kind of work. The repetition of number is a data set, termed as frequency of that particular number or the variable in which that number is assigned. The frequencies of variables in a data are to be listed in a table. This table is known as frequency distribution table and the list is referred to as frequency distribution.

Croxton and Cowden defined frequency distribution as "a statistical table which shows the sets of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side".

There are three types of frequency distribution. These are

i. Simple or Ungrouped frequency distribution

ii. Grouped frequency distribution

iii. Cumulative frequency distribution

i] Simple or Ungrouped frequency distribution: The data is presented in a simple or ungrouped form when the scores or values are presented with their number of occurrences in a tabular form. Number of occurrences is to be counted by applying the tally mark method. The example is given below:

Example 5.1: Let us consider the marks obtained by 100 students of a class in Economics.

Table 5.2: Marks of 100 Students of a Class in Economics

57	61	63	65	62	68	69	64	65	67
58	56	60	66	62	57	72	67	65	71
64	66	71	73	67	65	64	64	60	69
64	62	69	67	65	63	63	59	61	64
65	72	66	71	78	70	67	66	60	62
65	58	63	68	64	61	62	65	66	62
62	65	65	60	64	61	64	69	62	64
59	63	68	67	65	62	65	68	61	63
62	72	62	66	66	65	63	67	66	63
63	66	65	63	62	62	66	64	62	62

If the raw-data of Table 5.2 are arranged in either ascending, or, descending order of magnitude, we get a better way of presentation, usually called an "array" (Table 5.3).

Table 5.3: Array of Marks Shown in Table 5.2

56	60	62	62	63	64	65	66	67	69
57	61	62	63	64	64	65	66	67	69
57	61	62	63	64	65	65	66	67	70
58	61	62	63	64	65	65	66	68	71
58	61	62	63	64	65	65	66	68	71
59	61	62	63	64	65	65	66	68	71
59	62	62	63	64	65	66	67	68	72
60	62	62	63	64	65	66	67	68	72
60	62	62	63	64	65	66	67	69	72
60	62	62	63	64	65	66	67	69	73

Now let us present the above data in the form of a simple (or, ungrouped) frequency distribution using the tally marks. A tally mark is an upward slanted stroke (/) which is put against a value each time it occurs in the raw data. The fifth occurrence of the value is represented by a cross tally mark (\) as shown across

the first four tally marks. Finally, the tally marks are counted and the total of the tally marks against each value is its frequency.

Let us now represent the data in Table 5.3 as simple (or, ungrouped) frequency distribution.

Table 5.4: Simple Frequency Distribution of Marks of 100 Students

Marks	Tally marks	Frequency	Marks	Tally marks	Frequency
6	/	1	65	𝍸 𝍸 ////	14
57	//	2	66	𝍸 𝍸	10
58	//	2	67	𝍸 //	7
59	//	2	68	𝍸	5
60	////	4	69	////	4
61	𝍸	5	70	/	1
62	𝍸 𝍸 𝍸	15	71	///	3
63	𝍸 𝍸	10	72	///	3
64	𝍸 𝍸 /	11	73	/	1
Total	-	52	Total	-	48

Total Frequency 52 + 48 = 100

ii] Grouped Frequency Distribution:

Further, Table 5.2 can be condensed by making them into smaller groups, or, classes called **"Class-Intervals"**. The numbers fall in a class-interval is called its **"Class Frequency"**.

When the raw data is arranged into a number of classes by dividing the whole range of observations and indicating the corresponding class-frequencies against the class-intervals is called **"grouped frequency distribution"**.

Let us now represent the data in Table 5.2 as grouped frequency distribution. We find that the lowest value is 56 and the highest value is 73. Thus for approximately 10 classes the difference of values between two consecutive classes will be $\frac{73-56}{10}=\frac{17}{10}=1.7=2$ and the nine class-intervals will be 56-57, 58-59,..., etc. (Table 5.5).

Table 5.4: Simple Frequency Distribution of Marks of 100 Students

Class-intervals (Marks)	Tally marks	Frequency (No. of students)
56-57	///	3
58-59	////	4
60-61	𝍸 ////	9
62-63	𝍸 𝍸 𝍸 𝍸 𝍸 /	25
64-65	𝍸 𝍸 𝍸 𝍸 ////	25
66-67	𝍸 𝍸 𝍸 //	17
68-69	𝍸 ////	9
70-71	////	4
72-73	////	4
	Total	100

Thus the steps for preparing the grouped frequency distribution are :

1. to determine the class intervals.
2. to record the data using tally marks.
3. to find frequency of each class by counting the tally marks.

Several Important Terms :

(a) **Class-mark, or, Mid-value :** the mid-value or, class-mark of the class-interval lies exactly at the middle of the class-interval and is given by :

$$\text{Class-mark, or, Mid-vlue} = \frac{(\text{lower class limit} + \text{upper class limit})}{2}$$

$$\text{or, Lower class limit} + \frac{1}{2}(\text{upper class limit} - \text{lower class limit}$$

$$\text{or, } \frac{(\text{lower class boundary} + \text{upper class boundary}}{2}$$

(b) Class-limits: The minimum and maximum values of a class-interval are called lower class-limit and upper class limit respectively. In Table 1.5 the upper class-limits are 57, 59, 61, 63, 65, 67, 69, 71, 73 and the lower class-limits of nine classes are 56, 58, 60, 62, 64, 66, 68, 70, 72.

(c) Class boundaries: Class boundaries are associated with grouped frequency distribution, are the true-limits of a class interval. It, where there is a gap between the lower class-limit and the upper class-limit of the next class. This can be determined by applying the formula:

$$\text{Upper class boundary} = \text{upper class-limit} + \frac{1}{2}\,d$$

$$\text{Lower class boundary} = \text{lower class limit} - \frac{1}{2}d$$

Where d = common difference between the upper class-limit of a class-interval and the lower class limit of the next higher class interval.

The class-boundaries of the class-intervals of Table 1.5 will be 55.5 –57.5; 57.5 – 59.5; 59.5 – 61.5; etc., since d = 58 – 57 = 60 – 59 = ... = 1. The class-boundaries convert a grouped frequency distribution (inclusive type) into a continuous frequency distribution.

(d) Type of class-interval: different types of class-intervals with their class limits ate given below:

(A) Exclusive type

Upper limit excluded			Lower limit excluded
i	ii	iii	
20 – 25	150 – 200	20–	Above 20 but not more than 25
25 – 30	200 – 250	30–	Above 25 but not more than 30
30 – 35	250 – 300	40–	Above 30 but not more than 35
35 – 40	300 – 450	50 – 60	Above 35 but not more than 40

(B) Inclusive type	**(C) Open-end type**		**(D) Unequal class-interval**
	i	**ii**	
40 – 49	0 – 30	Below 30	0 – 10
50 – 59	30 – 60	30 – 40	10 – 30
60 – 69	60 – 90	40 – 50	30 – 80
70 – 79	90 – 120	Above 50	80 – 200

Class-intervals, **Exclusive type (A)** like 20 – 25, 25-30; 150 – 200, 200 – 250; 20–, 30–; are upper limit exclusive type, i.e., an item exactly equal to 25, 200 and 30 are put in the intervals 25-30, 200-250 and 30-, respectively and not in intervals 0 – 25, 150 – 200 and 20–, respectively. Similarly, 25 is included and 20 excluded (lower limit) in "above 20 but not more than 25" class-interval. In the exclusive type the class-limits are continuous, i.e., the upper-limit of one class-interval is the lower limit of the next class-interval and class limits of a class-interval coincide with the class boundaries of that class-interval. It is suitable for continuous variable data and facilitates mathematical computations.

Again class-intervals, **Inclusive type (B)** like 40-49, 50-59, 60-69, etc., are inclusive type. Here both the upper and lower class-limits are included in the class-intervals, e.g., 40 and 49 both are included in the class-interval 40-49. This is suitable for discrete variable data. There is no ambiguity to which an item belongs but the idea of continuity is lost. To make it continuous, it can be written as 39.5-49.5, 49.5-59.5, 59.5-69.5 etc.

In **'open-end' class-interval (C)** either the upper limit of the last class- interval, or, lower limit of the first class-interval, or, both are missing. It is difficult to determine the mid-values of the last and first class-intervals without an assumption. If the other closed class-intervals have equal width, then we can assume that the open-end class-intervals also have the same common width of the closed class-intervals.

Grouped frequency distributions are kept open ended when there is limited number of items scattered over a long interval. **Unequal class-intervals (D)** are preferred only when there is a great fluctuation in the data.

(e) Width or Length (or size) of a Class-interval: It is the difference between the upper class boundary and the lower class boundary i.e.

Width of a class-interval = Upper class boundary – Lower class-boundary

Common width of a class-interval is equal- to the difference between two successive upper Class-limits or, two successive lower class-limits (when the class-intervals have equal widths). The primary rules of classification are given below:

(i) The classes should be exhaustive, i.e., each value of the raw data should be included in them.

(ii) All the classes should preferably be equal length or width. Only in some special cases, it uses classes of unequal width.

(iii) Width of class-interval is determined by first fixing the no. of class-intervals and then dividing the total range by that number.

(iv) The class-limits (integral or fractional) should be made in such a way that no value of the item in the raw data coincides with the value of the limit.

(v) There should not be any ambiguity in the definition of classes. It will eliminate all doubts while including a particular item in a class.

(vi) The classes should be mutually exclusive and non-overlapping, i.e.,each item of the raw data should fit only in one class.

(vii) The classification must be suitable for the object of inquiry.

(viii) The classification should be flexible and items included in each class must be homogeneous.

(ix) The number of classes should normally be between 10 and 20, i.e., neither too small nor too large.

(f) Frequency density:

$$\text{Frequency density of a class-interval} = \frac{\text{class frequency}}{\text{width of the class}}$$

e.g., the frequency density of class-interval 62-63 in Table 5.5 is $\frac{25}{2} = 12.5$. If the class-intervals of a frequency distribution are of unequal width, the frequency densities of the class-intervals can be used to compare the concentration of frequencies in the class-intervals and construct the histogram.

(g) Percentage frequeneye :

$$\text{Percentage frequency of a class-interval} = \frac{\text{class frequency}}{\text{total frequency}} \times 100 \text{ e.g.,}$$

$$\text{the percentage frequency of above class-interval} = \frac{25}{100} \times 100 = 25$$

(h) Relative frequency:

$$\text{Relative frequency} = \frac{\text{class frequency}}{\text{total frequency}}$$

e.g., the relative frequency of class-interval 62–63 in Table 5.5 is $\frac{25}{100} = 0.25$

iii] Cumulative frequency distribution: When the frequency of each class-interval is cumulative then a frequency distribution becomes cumulative. Cumulative frequency of a class-interval can be made by adding the frequency of that class-interval to the sum of the frequencies of the preceding class-intervals.

Often we want to know the number of cases which fall above, or, below a certain value. Hence, cumulative frequencies are of two types, i.e., (1) more than (or, from above) cumulative frequencies and (2) less than (or, from below) cumulative frequency. In the more than type the cumulative frequency of each class-interval is obtained by adding the frequencies of the given class and the succeeding classes. In the less than type the cumulative frequency of each class-interval is obtained by adding the frequencies of the given class and all the preceding classes, when the classes are arranged in the ascending order of the value of the variable. For grouped frequency distribution, the cumulative frequencies are shown against the class-boundary points.

Example 1.2. Construct the cumulative frequency distribution (both "more than" and "less than" types) from the following Table 5.6:

Table 5.6: Height [in inch] of students

Height (in inch):	30-39	40-49	50-59	60-69	70-79	Total
No. of students:	8	12	27	17	14	78

Solution: Here, d = Gap between two consecutive classes = 1, $\frac{1}{2}d = 0.5$

∴ Lower class-boundary points are 29.5, 39.5, 49.5, etc. and the last upper class-boundary point is 79.5. Hence, the class boundary points are 29.5,39.5,..., 79.5.

Table 5.7: Cumulative Frequency Distribution of Height of 68 Students

Class-boundary points (Height in inch)	Cumulative Frequency	
	"Less than"	"More than"
29.5	0	70 + 8 = 78
39.5	0 + 8 = 8	58 + 12 = 70
49.5	8 + 12 = 20	31 + 27 = 58
59.5	20 + 27 = 47	14 + 17 = 31
69.5	47 + 17 = 64	0 + 14 = 14
79.5	64 + 14 = 78	0

Note: In less than type, no. of students of height less than 29.5 inch is nil; hence 0 is written against 29.5; no. of students of heights less than 39.5 is 8, hence 8 is written against 39.5 inch and so on. Similarly, in more than type, the number of students of heights 79.5 inch and above is nil, hence 0 is written against 79.5 inch; number of students of weights 69.5 inch or more is 14, hence 14 is written against 69.5 inch and so on.

Median, Quartiles, Deciles and percentiles can be obtained from cumulative frequency distributions.

5.5 Graphical Representation of Data-Pie Diagram, Bar Graph, Histogram, Frequency Polygon and Ogive (Using 75% Rule)

To demonstrate to our readers the characteristics of this frequency distribution more clearly, we could draw up a pictorial representation of the data. One of the advantages of doing this is that the mode will be immediately apparent, as will other features, such as the rate at which numbers fall off to either side and any specially interesting clusters of data. A graphical presentation can also be justified by its immediate appeal to the eye.

The diagrams commonly used to depict statistical data, given in the form of frequency distribution, are:

5.5.1 Pie diagram

5.5.2 Bar Graph

5.5.3 Histogram

5.5.4 Frequency Polygon and

5.5.5 Ogive (Using 75% rule).

5.5.1 Pie diagram:

Pie diagram is simple diagram for displaying categorical or grouped data. Pie diagram is circle whose area is divided proportionately among the different components by straight lines drawn from the centre to the circumference of the circle. They are best used when there are only a handful of categories to display and we are interested in the comparison of the various categories or between a

part and the whole, such a diagram is very helpful in effectively displaying the data.

A pie chart consists of a circle divided into segments, one segment for each category. The size of each segment is determined by the frequency of the category and measured by the angle of the segment. As the total number of degrees in a circle is 360, the angle given to a segment is 360 times the fraction ofthe data in the category, i.e.

$$\text{Angle} = \frac{\text{Number in category}}{\text{Total number in sample (n)}} \times 360$$

Example 5.3: Make a pie diagram considering the data in Table 5.S

Table 5.8: Sale of different news papers

Paper	Frequency
The Times	20
The Sun	10
The Statesman	30
A The Telegraph	40
Total	100

Solution: The pie diagram is constructed by first drawing a circle and then dividing it up into segments whose angles are calculated using the following formula which is given in Table5.9.

$$\text{Angle} = \frac{\text{Number in category}}{\text{Total number in sample (n)}} \times 360$$

Table 5.9: Sale of different news papers

Paper	Frequency	Angle [in degree]
The Times	20	72
The Sun	10	36
The Statesman	30	108
The Telegraph	40	144
Total	100	360

Pie Diagram

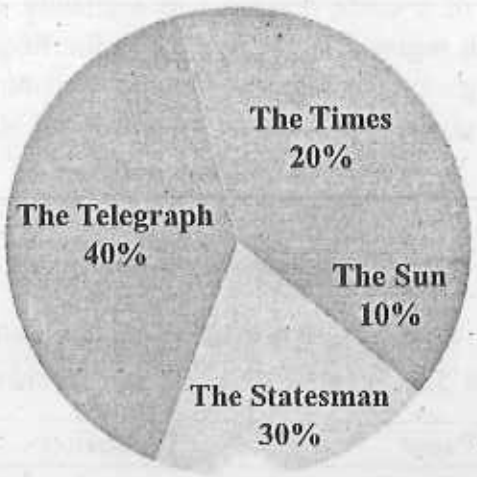


Figure 5.1: Showing the sale of different news papers

5.5.2 Bar Graph

Bar graph consists of a group of equispaced rectangular bars, one for each category (or class) of given statistical data. The bars, starting from a common base line, must be of equal width and their lengths represent the values of statistical data. They are shaded of co loured suitably. Bar graph may be used for discrete variable where a histogram cannot be constructed as the classes are not comparable in terms of magnitude.

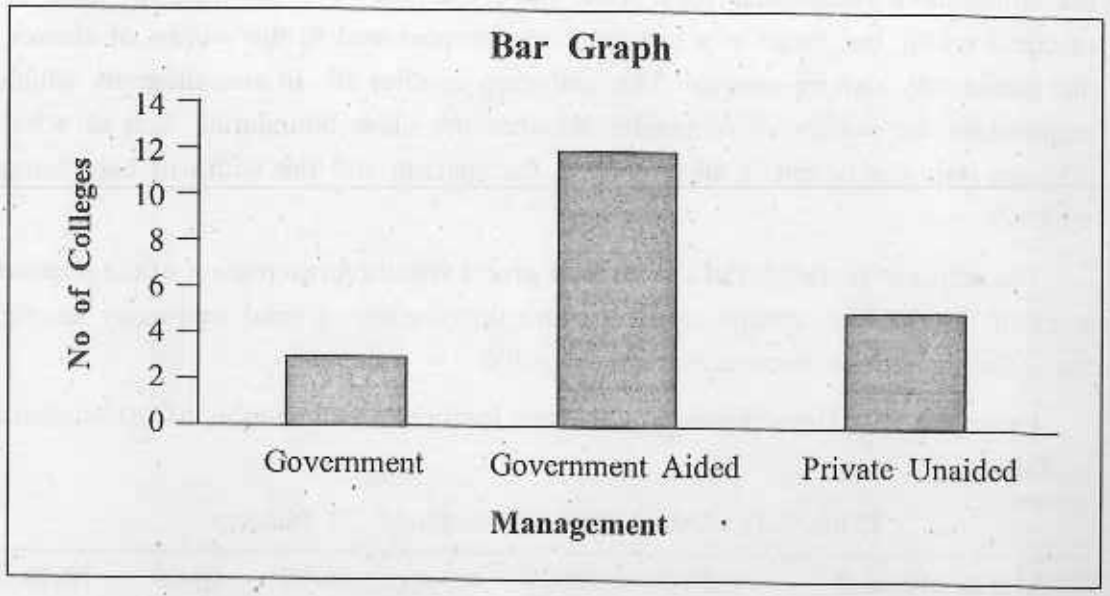
Example 5.4: Draw a bar graph for a particular district where the number of managementwise colleges is 20. The distribution of colleges is shown in Table 5.10

Table 5.10: Management wise colleges

Management	No. of Colleges
Government	3
Government Aided	12
Private Unaided	5

Solution:

Figure 5.2: Showing the bar graph of different management-wise colleges



Bars (rectangles) of uniform width are drawn with equal spaces in between them, on the horizontal axis-called x-axis represents the management of colleges. The heights of the rectangles are shown along the vertical axis-called y-axis and are proportional to their respective frequencies (number of Colleges). The width of the rectangle has no special meaning except to make it pictorially more attractive.

5.5.3 Histogram:

Histogram is the most common form of diagrammatic representation of grouped frequency distribution. It consists of a set of adjoining rectangles drawn in a horizontal base line, with areas proportional to class frequencies. The width of rectangles, one for each class, extends over the class boundaries (not class limits) shown on the horizontal scale. When all classes have equal width, the heights of rectangles will be proportional to the class frequencies and it is then customary to take the heights numerically equal to the class frequencies. If, however, the classes are of unequal width the rectangles will also be of unequal width, and therefore the heights must be proportional to the frequency densities.

Although vertical bar chart and histogram may appear somewhat alike, the main

point of distinction between them is that the consecutive rectangles in a histogram have no space in between, but the bar diagram must have equal spaces left between the consecutive rectangular bars. Also, the rectangles of a bar diagram must be of equal width, but those in a histogram are proportional to the widths of classes, and hence may also be unequal. The histogram is, after all, an area diagram, which emphasizes the widths of rectangles between the class boundaries. But in a bar diagram only the height is all-important, the spacing and the width of bars being arbitrary.

The series of rectangles in a histogram give a visual representation of the relative sizes of the, various groups and the entire distribution of total frequency among the different classes becomes at once visible.

Example 5.5: Draw histogram for the frequency distribution of 20 students in Table 5.11.

Table 5.11: Distribution of marks of 20 students

Marks obtained	20-29	30-39	40-49	50-59	60-69	70-79
Number of students	1	3	1	6	4	5

Solution: We go through the following steps for drawing a histogram:

Step 1: On a graph paper, draw two perpendicular lines and call them as horizontal (x-axis) and vertical (y-axis) axes,

Step 2: Along the' horizontal axis, we take classes (marks) 19.5-29.5, 29.5-39.5, ... (Here each is of equal width 10)

Step 3: Choose a suitable scale on the vertical axis to represent the frequencies (number of students) of classes.

Step 4: Draw the rectangles as shown in Fig. 5.3.

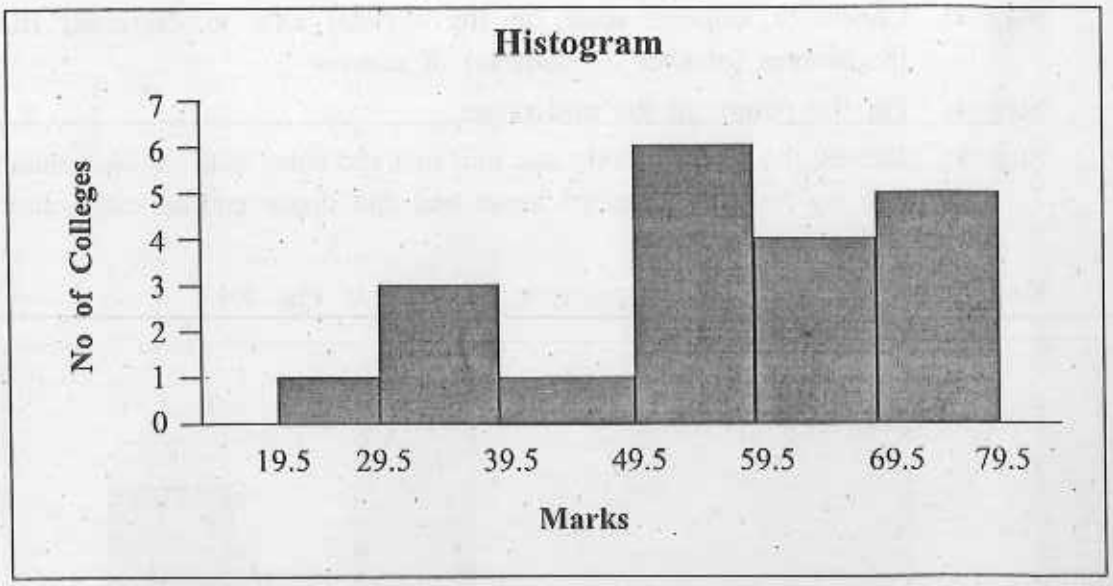


Figure 5.3: Showing the histogram of the marks of students

5.5.4 Frequency Polygon

Frequency polygon is the graphical representation alternative to histogram and may be looked upon as derived from histogram by joining the mid-points of the tops of consecutive rectangles. It is generally used in cases when all the classes have a common width. In actual construction, therefore, .the frequency polygon is obtained by joining the successive points whose abscissa represent the mid-values and ordinates represent the corresponding class frequencies. The two end-points are joined to the base line at the mid-values of the empty classes at each end of the frequency distribution. Thus the frequency polygon has the same area as the histogram, provided the width of all classes is the same. The frequency polygon is particularly useful in representing simple frequency distributions of a discrete variable.

Example 5.6: Draw frequency polygon considering the Table 5.11and figure5.3.

Solution: We go through the following steps for drawing a frequency polygon:

Step 1: On a graph paper, draw two perpendicular lines and call them as horizontal (x-axis) and vertical (y-axis) axes,

Step 2: Along the horizontal axis, we take the mid-points i.e. 24.5, 34.5, 44.5 ... and so on,

Step 3: Choose a suitable scale on the vertical axis to represent the frequencies (number of students) of classes.

Step 4: Put the points on the mid-values,

Step 5: Join all the points orderly and join two end sides with the horizontal line on the mid-values of lower end and upper end of extra class limits.

Step 6: Draw the frequency polygon as shown in Fig. 5.4.

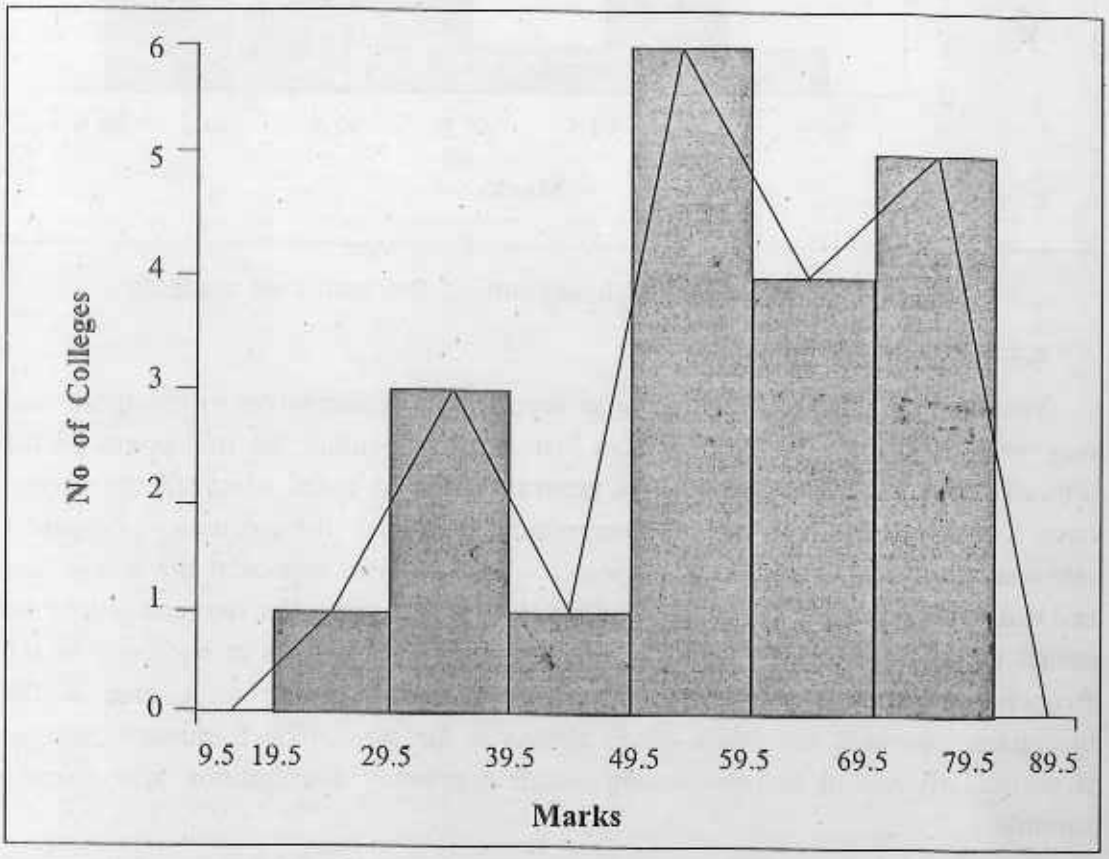


Figure 5.4: Showing frequency polygon of marks of the students

5.5.5 Ogive [using 75% rule]

Ogive is the graphical representation of cumulative frequency distribution, and hence is also called Cumulative Frequency Polygon. When cumulative frequencies are plotted against the corresponding class boundaries and the successive points

are joined by straight lines, the line diagram obtained is known as Ogive or Cumulative Frequency Polygon. It starts from the lowest class boundary on the horizontal axis and gradually rising upward ends at the highest class boundary corresponding to the cumulative frequency N, i.e. the total frequency. It looks like an elongated letter S. Unequal widths of classes in the frequency distribution do not cause any difficulty in the construction of an ogive. For drawing ogive using 75% [$\frac{3}{4}$ th] rule means if vertical axis is 3 then horizontal axis will be 4.

The difference between the construction of the frequency polygon and Ogive is that for frequency polygon, one takes the mid points of the class interval on horizontal axis, while for Ogive one takes the upper boundary of the class interval on horizontal axis. Again on the vertical axis, in case of Ogive one takes cumulative frequency/cumulative frequency percentage instead of frequency only.

Example 5.7: Draw Ogive considering the Table 5.11.

Solution: We go through the following steps for drawing an Ogive [cumulative frequency curve or percentage cumulative frequency curve].

Step 1: On a graph paper, draw two perpendicular lines and call them as horizontal (x-axis) and vertical (y-axis) axes,

Step 2: Along the horizontal axis, we take the upper boundary of the class interval i.e. 19.5,29.5,39.5, ... and so on.

Step 3: Choose a suitable scale on the vertical axis to represent the cumulative frequency or percentage cumulative frequency (number of students) of classes.

Step 4: Put the points on the upper boundary.

Step 5: Join all the points orderly and join the lower end with the horizontal line on the lower boundary of the lower class limit.

Step 6: Draw the Ogive [cumulative frequency as shown in Fig. 5.5 and percentage cumulative frequency as shown in fig. 5.6] with the help of the following Table: 5.12.

Table 5.12: Cumulative frequency distribution or percentage cumulative frequency distribution of marks

Marks obtained	20-29	30-39	40-49	50-59	60-69	70-79
Number of students	1	3	1	6	4	5
Cumulative Frequency	1	4	5	11	15	20
% Cumulative Frequency	5	20	25	55	75	100

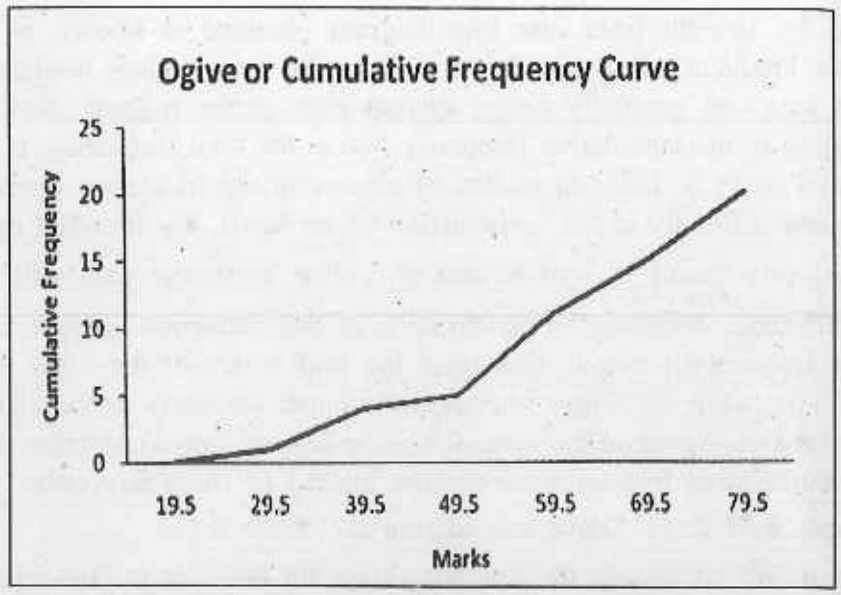


Figure-5.5: Showing the Ogive or cumulative frequency curve

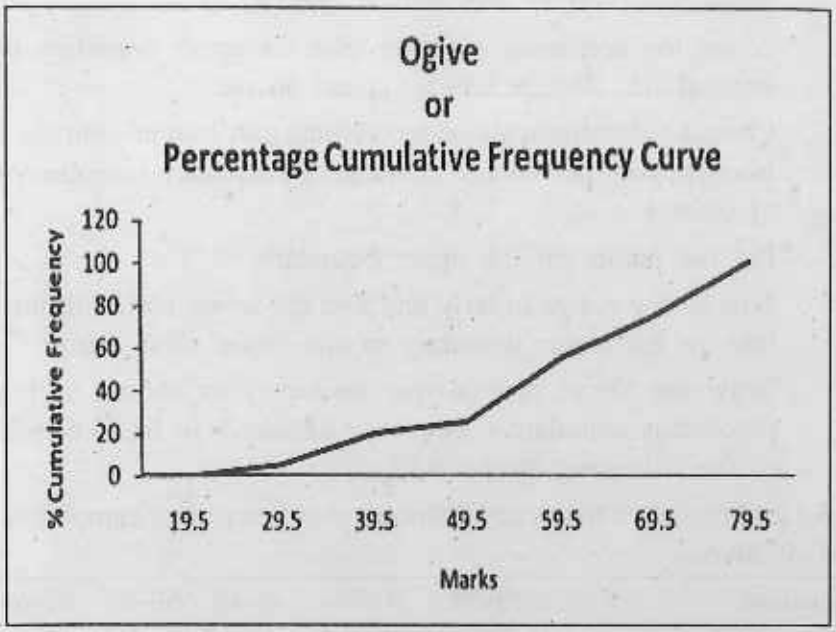


Figure 506: Showing the Ogive or cumulative frequency percentage curve

5.6 Let us sum up :

- Statistic derived from Latin word 'status' means the activity which deals with the theory and method of collecting, analyzing, and interpreting the data. When it occurs in education system for planning, implementing, and evaluating for upliftment of education then it is called educational statistics.
- Data may be primary or secondary; it may be qualitative or quantitative; and it may be discrete or continuous, depending on the nature of elements.
- Primary data mean data collected by himself or herself, secondary data represent data not collected by himself or herself, qualitative data are those data where elements are classified into categories, quantitative data are the form of number or numerical data, discrete data related to exact values and not any fractional values, continuous data signifies any numerical value (integral/fractional) within a certain range. Temporal data are related to time. Spatial data denote space or places, In this Unit, we discussed how are the data organized and tabulated in a frequency distribution table. For the preparation of frequency distribution three steps should be followed, first one is the selection of non-overlapping classes; the second one is enumeration of data values in each class; and the third one is construction of table.
- The methods of presenting the data graphically using bar chart which was used mostly to represent the qualitative data; histogram, frequency polygon, pie diagram, and Ogive which were explained and used mostly to represent the quantitative data. The graphical presentation of the data often facilitates understanding of a set of data. It is easier to read and interpret data from graphs.

5.7 Check Our Progress (10 Questions : 5 SA Type 5 Objective Types)

A. Short Answer type questions:

1. Define educational statistics.
2. Write two needs of educational statistics.
3. Defme primary data.

4. The following data are the scores of 40 students obtained in Bengali.

5	6	7	8	9	10	11	12	13	14
15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34
35	36	37	38	39	40	41	42	43	44

(a) Arrange the above data in frequency distribution using i) an interval of 5 units, and ii) an interval of 10 units. Let the first interval begins with score 40.

5. Construct i) histogram or ii) frequency polygon or iii) Ogive using the above data.

B. Objective type questions

1. What will be the lower and upper class boundaries of 25 - 29 where next class limit. is 30-34?
2. What will be the mid-value of the score 25 - 29?
3. Height of persons is type of data (Primary / Secondary) (Fill in the blank).
4. Scores on a test of intelligence are type of data (Continuous / Discrete) (Fill in the blank) .
5. List two difficulties faced by anyone while preparing the histogram.

5.8 References :

Best, John W. and Kahn, James V. (1992). Research in education. Delhi: Prentice Hall of India.

Garrett, H.E. (1962). Statistics In psychology and education. Bombay: Allied Publishers Pvt. Ltd.

Guilford, J.P. (1965). Fundamental statistics in psychology and education. New York: McGraw-Hill Book Co.

Koul, Lokesh (1997). Methodology of educational research (3rd revised edn.). New Delhi: Vikas Publishing House Pvt. Ltd.

Lovitt, W. V. and Holtzclaw H. F. (1929). Statistics. New York: Prentice-Hall.

Mangal, S. K. (2002). Statistics in psychology and education. Delhi: Prentice Hall of India.

Unit: 6 ❑ Descriptive Statistics

STRUCTURE:

6.1 Introduction

6.2 Objectives

6.3 Meaning of central tendency- mean, median, and mode-their properties, calculation, and application.

6.3.1 Mean

6.3.2 Median

6.3.3 Mode

6.4 Measure of variability- Range, Average Deviation [AD] , Standard Deviation [SD] and Quartile Deviation [QD]· - their properties, calculation and application

6.4.1 Range

6.4.2 Average Deviation

6.4.3 Standard Deviation

6.4.4 Quartile Deviation

6.5 Percentile and Percentile Rank- defmition, calculation, application, graphical determination

6.5.1 Percentile

6.5.2 Percentile Rank

6.6 Let Us Sum Up

6.7 Check our progress

6.8 References.

6.1 Introduction

In the previous unit we have studied about the educational statistics. We have also studied about the graphical representation of data. In this unit we shall learn how to classify the data for measuring different descriptive statistics which are used to describe the characteristics of -a sample or population in totality. Generally, statistical measures are classified into two types - descriptive and inferential. Descriptive statistics are computed to describe the characteristics of a sample or

population in totality and thus limit generalization to particular sample. The measures of descriptive statistics most commonly used ill educational research are: i) measures of central tendency and ii) measures of variability.

In this unit, we shall focus upon the descriptive statistical measures.

6.2 Objectives

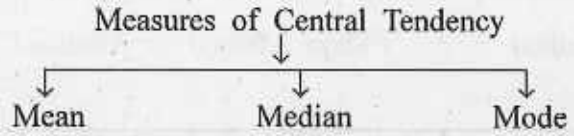
After going through this unit, we should be able to:

- understand the meaning of measures of central tendency;
- learn about the properties of mean, median, and mode;
- calculate mean, median, and mode;
- apply mean, median, and mode in the appropriate situations;
- understand the meaning of measures of variability;
- learn about the properties of range, average deviation, standard deviation, and quartile deviation;
- calculate range, average deviation, standard deviation, and quartile deviation;
- apply range, average deviation, standard deviation, and quartile deviation in the appropriate situations;
- Define percentile and percentile rank;
- Calculate percentile and percentile rank;
- Determine percentile and percentile rank graphically and
- Apply percentile and percentile rank in appropriate situations.

6.3 Meaning of Central Tendency— mean, median and Modetheir Properties, Calculation and Application.

Data are collected from different in order to know the performance of the individual or a group of individuals for erstanding of the individuals or a group. The most effective and efficient way of describing a set of data is by means of the single point which describes the entire series of obse ations with their varying sizes. The central location of a sample is a point on the measurement scale around which most of the scores are found, so that some observations are larger and some others are smaller than it. A measure of central tendency denotes average and

provides concise description of the performance of the whole group for comparison of groups in terms of their performance and it is used to interpret the nature of the data of that particular group. The three most commonly used measures of central tendency are i) Mean, ii) Median, and iii) Mode.



This section will discuss the measures of central tendency such as Mean, Median, and Mode and their properties, calculation, and application.

6.3.1 Mean

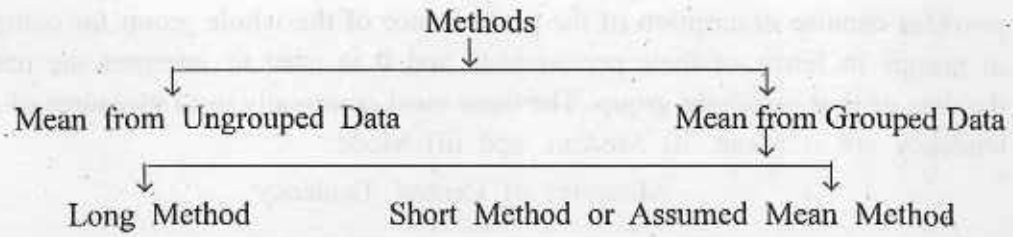
In nonnallanguage we use the term 'average' for what is technically known as the 'arithmetic mean'. The mean is the most widely used indicator of central location. Mean is calculated when the data are presented on equal interval scale. It is the most familiar, most frequently used, and well understood average.

Properties: The properties of an arithmetic mean are given below:

1. Arithmetic mean is the simplest and easiest to understand.
2. It is rigidly defmed.
3. It is based on all observations.
4. It is easy to calculate.
5. Arithmetic mean is also the most stable of all the measure of central tendency, so far as sampling fluctuations are concerned.
6. The total of a set of observations is equal to the arithmetic mean.
7. The sum of the deviations of a set of observations from their arithmetic mean is always zero.
8. Mean represents the centre of gravity of the scores.
9. All observations are taken into consideration for calculating mean.
10. The result of mean has its high reliability and applicability to the inferential statistics.

Calculation:

It can be calculated by following two methods: i) from ungrouped data and ii) from grouped data. Again mean from grouped data can be calculated by following two methods: a) long method and b) short method or assumed mean method.



i] Mean from Ungrouped Data or Simple Ungrouped Mean: The mean of a data (X) is obtained by getting sum of all observations and dividing the sum ($\sum X$) by the number of observations (N). The formula for computing the mean from ungrouped data is given by:

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

Where, X = Individual score,

$\sum X$ = Sum of all scores, and

N = Number of total frequencies.

Example 6.1: Calculate mean from the scores 10, 12, 11, 25, 20, 26 and 18 obtained by 7 students in test of mathematics.

Solution:

The mean will be:

$$\text{Mean} = M = \frac{10+12+11+25+20+26+18}{7} = 17.43$$

ii] Mean from Grouped Data or Grouped Mean

When data are given in grouped frequency distribution then we may compute by long method or by the short method using assumed mean. These are low:

A. Mean by Long Method:

In long method, mean is calculated by the following formula:

$$\text{Mean} = \frac{\Sigma f X i}{N}$$

Where, Xi = Midpoint of each Group,

f = Frequency,

N = Total number of frequencies.

Example 6.2: Compute mean from the following frequency distribution Table 6.1.

Table 6.1: Frequency distribution of 63 students

Class Interval	13-17	18-22	23-27	28-32	33-37	38-42	43-47	48-52	53-57
Frequency	2	22	10	14	3	4	6	1	1

Solution:

Table 6.2: Mean from grouped frequency distribution [long method]

Class Interval	Frequency (f)	Mid value (X)	fX
13-17	2	15	30
18-22	22	20	440
23-27	10	25	250
28-32	14	30	420
33-34	3	35	105
38-42	4	40	160
43-47	6	45	270
48-52	1	50	50
53-57	1	55	55
Total	63		ΣfX= 1780

$$\text{Mean} = \frac{\Sigma fXi}{N} = \frac{1780}{63} = 28.25$$

Hence, Mean is 28.25

B. Mean by Short Method or Assumed Mean Method or Coding Method

In short method or assumed mean method, mean can be calculated by the following formula:

$$\text{Mean} = \text{A.M.} + C\ i = \text{A.M.} + \frac{\Sigma fx'}{N} i$$

Where, A.M. = Assumed Mean,

$C = \frac{\Sigma fx'}{N}$,

$x' = (X\text{-}M)/i$ = Deviation of mid value from assumed mean and dividing by the size of the class interval,

i = Size of the class interval,

f = Frequency,

N = Total number of frequencies.

Example 6.3: Compute mean from the following frequency distribution Table: 6.3.

Table 6.3: Frequency distribution of 63 students

Class Interval	13-17	18-22	23-27	28-32	33-37	38-42	43-47	48-52	53-57
Frequency	2	22	10	14	3	4	6	1	1

Solution:

Table 6.4: Mean from grouped frequency Distribution [short method]

Class Interval	Frequency (f)	Mid value (X)	$x' = \frac{X - M}{i}$	fx'
13-17	2	15	–3	–6
18-22	22	20	–2	–44
23-27	10	25	–1	–10
28-32	14	30	0	0
33-37	3	35	1	3
38-42	4	40	2	8
43-47	6	45	3	18
48-52	1	50	4	4
53-57	1	55	5	5
Total	N=63			$\Sigma fx' = -22$

A.M. = Assumed Mean $=\frac{28+32}{2}=30$

Mean = A.M. + Ci = A.M. + $\frac{\Sigma fx'}{N}$i

$= 30+\frac{-22}{63}\times 5 = 30\text{-}1.75 = 28.25$

Hence, Mean is 28.25

Application:

Mean is computed when:

i) the centre of gravity of a sample is desired;

ii) the most stable measure of central tendency is desired;

iii) the scores are distributed symmetrically about the centre of the distribution;

iv) an accurate description of the sample and indirectly, that of the population is required; and

v) the additional statistics are to be computed later.

6.3.2 Median

Using the median gets us around the difficulty for the mean outlined just above. The median is the central value or middle-most value of set of observations when they are arranged in order of magnitude. The median is the point that divides a set of ordered observations in such a way that half of the observations are placed below the median value and half are placed above that value.

Properties: The properties of median are as follows:

1. Median is rigidly defined.
2. It is based on all observations.
3. It is easy to calculate.
4. It is not so well-understood as arithmetic mean or mode.
5. It shows greater fluctuation from sample to sample, and hence is less reliable than arithmetic mean.

Calculation: It can be calculated from ungrouped data and grouped data. Median is, in certain sense, the real measure of central tendency as it gives the value of the most central observation. It is unaffected by extreme values, and can be easily calculated from frequency distributions with open-end classes.

i] Median from Ungrouped Data: When there is an **odd number** of scores and scores are arranged (ascending or descending order) according to their ranks then median is the middle score. The median can be obtained by finding out the $\frac{(x+1)}{2}$th term. For example, considering the scores: 3, 4, 9, 7, 11, 10, and 14. Here we have seen seven scores. On arranging them in ascending (or descending) or we may have the sequence of scores as follows: 3,4,7,9,10,11, and 14. Here, the $\frac{(x+1)}{2}$th term will be $\frac{7+1}{2}$th term = 4^{th} term, which happens to be 9. We find that there are 3 cases above and below 9. Thus 9 divides the whole distribution into two equal halves. Therefore, the median of the data would be 9.

When there is an **even number** of scores such as 10, 15, 16, 14, 18, and 21. The first step towards finding the median would be to arrange the given scores serially (ascending or descending order). We shall get 10, 14, 15, 16, 18, and 21. The median is a point halfway between the two middle values 15.5 ($\frac{6+1}{2}$th term = 3.5^{th} term or the average of 3^{rd} term and 4^{th} term, i.e $\frac{15+16}{2}=15.5$).

After calculating the median, we can check that half the scores do fall above and below the scores we have identified as median. This will help us to avoid making errors.

Median from Grouped Data: Median from a grouped frequency distribution is that value which corresponds to cumulative frequency $\frac{N}{2}$. Median from a grouped frequency distribution can be calculated by the following formula:

$$\text{Median} = L_L + \frac{\frac{N}{2} - f_b}{f_m} \times i$$

Where L_L = Lower limit of Median class interval

N = Total number of cases

f_b = Cumulative frequency below the median class

fm = Frequency in median class

i = Size of the class interval

Example 6.4: Calculate median from the following table 6.5.

Table 6.5: Frequency distribution of 40 students

Class Limit	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	Total
Frequency	2	3	5	8	10	6	4	2	40

Solution:

Table 6.6: Calculation of Median

Class Limit	Frequency [f]	Cumulative Frequency [cf]	Class Boundaries
5-9	2	2	4.5-9.5
10-14	3	5	9.5-14.5
15-19	5	10	14.5-19.5
20-24	8	18	19.5-24.5
25-29	10	28	24.5-29.5
30-34	6	34	29.5-34.5
35-39	4	38	34.5-39.5
40-44	2	40	39.5-44.5

Median $= L_L + \frac{\frac{N}{2} - f_b}{f_m} \times i$ [Where $L_L = 24.5$, $f_b = 18$, $f_m = 10$, $i = 5$, $N=40$]

$$24.5 + \frac{20-18}{10} \times 5 = 24.5 + 1 = 25.5$$

Therefore, the median is 25.5

Special cases for calculation of median:

Special case 1:

If it is found that $\frac{N}{2}$ is equal to the cumulative frequency then the median will be the upper limit of class (where cumulative frequency is less than type) and the lower limit of the class (where cumulative frequency is more than type).

Example 6.5: Find the median of the following frequency distribution Table 6.7.

Table 6.7: Frequency distribution of 40 students

Scores	20-29	30-39	40-49	50-59	60-69
Freguency	3	7	10	15	5

Solution:

From the above distribution the median will be the upper limit of the class interval 40-49 i.e 49.5 as $\frac{N}{2}=20\left(\frac{40}{2}\right)$ is equal to the cumulative frequency 20 in less than type.

Another way, we can say that the median will be the lower limit of the class interval 50-59 i.e 49.5 as $\frac{N}{2}=20\left(\frac{40}{2}\right)$ is equal to the cumulative frequency 20 in more than type.

Special case 2:

If it is found that there are no cases with the class interval containing the median, or $\frac{N}{2}$ lie in two class interval then the median will be the mid value of that class where the frequency is zero.

Example 6.6: Find the median of the following frequency distribution

Table 6.8: Frequency distribution of 15 students

Scores	20-29	30-39	40-49	50-59	60-69
Frequency	3	5	0	7	1

Solution:

From the above distribution, it is found that $\frac{N}{2}=8\left(\frac{16}{2}\right)$ lie against two class intervals showing the cumulative frequencies (less than type) i.e 30-39 and 40-49. The median will be the mid value of 40-49 i.e 44.5.

Application:

Median is computed when:

i) we want to locate the exact midpoint of the distribution i.e.exact 50% point of the distribution;

ii) all the observations are not known, median can be applied, provided the general location of all observations and values near the middle are available. Median can also be applied without difficulty from grouped frequency distributions with classes of unequal width or with open-end classes i.e incomplete distribution;

iii) qualitative data in psychological and social studies, where numerical measurements may not be available, but it is possible to rank the objects in some order;

iv) there are extreme scores which would markedly affect the mean; and

v) we are interested in whether cases fall within the upper or lower halves of the distribution and not particularly in how far they are from central point.

6.3.3 Mode

The most frequently occurring value or with maximum frequency of a set of observations is called mode arid therefore even easier to find than the mean or median. It is the most typical value and at times represents the true characteristic of the frequency distribution as a measure of central tendency. The mode is very helpful in understanding nominal data where mean and median provide almost no help. Here we can count the number of cases in each category and obtain the frequency and note down the category which is most popular.

Properties

The properties of mode are the following:

1. It is rigidly defined.
2. It is based on all observations.
3. It is very difficult to calculate accurately.
4. It is less reliable and less stable as regards sampling fluctuations.
5. It is well-understood and easily comprehensible.

Calculation:

For continuous variables mode gives a quick measure which is les dependable

and less precise as compared to other measures of central tendency. If we draw a histogram or frequency polygon, we will find a highest pick that represents the mode of this distribution. Sometimes there are two distinct highest picks that represent the maximum frequency then we can say that it is bimodal distribution andthe values or scores with highest frequency can be said the modes of this distribution.

i) Mode from ungrouped data: In the case of a simple ungrouped distribution, mode can be found by inspection only. for example, if the scores of five students are 12, 13 11, 13, and 10 the most frequent score is 13 as it has been obtained by two students. Therefore, mode is 13.

When all observations occur with equal frequency, mode does not exist. Forexample, the scores of four students are 12, 12, 12, and 12. Then we can say that there is no mode in this distribution.

Again there is more than one mode, it two or more values occur with the maximum frequency. For example, the scores of six students are 6, 5,4, 6, 7, and 5 there are two modes i.e 5 and 6 and this is bi-modal distribution.

ii] Mode from grouped data: When the data are arranged in grouped frequency distribution in terms of class intervals, the mode will be mid value of that class interval where the frequency is high. Or it can be calculated by the formula:

Mode = 3 Median - 2 Mean

Example 6.7: Calculate mode from the table no 6.6.

Form the distribution of the table no 6.6, it is found that there are maximum (10) frequencies in the class interval 25 - 29. So the midpoint, i.e. 27 is the mode. Or mode can be obtained by applying the formula Mode=3median-2mean.

Example 6.8: Calculate mode from the following distribution.

Table 6.9: Frequency distribution of 50 students

Class interval	Frequency
10-19	4
20-29	15
30-39	10
40-49	15
50-59	6

Solution:

The above distribution is bimodal distribution, has two modes as there are two highest frequencies (15) which lie in the class interval 20-29 and the class interval 40-49. The mode will be the two, i.e. the two midpoint of the two class interval. The first one will be the midpoint of the class 20-29 i.e. 24.5 and the next one will be the midpoint of the class 40-49 i.e. 44.5

Application:

Mode may be applicable in the following type of educational situations when:

i) we want a quick and approximate measures of central tendency;

ii) unlike arithmetic mean, it can be calculated from frequency distributions, with open-end classes. In fact, if the modal class and its two adjoining classes together with their class frequencies are available, mode can be determined, provided it is known that all classes are of equal width;

iii) a rough estimate of central tendency will serve the purpose;

iv) the purpose is to know the most typical case; and

v) the distribution is skewed or the data is incomplete and most of the frequency are towards the extremes.

6.4 Measure of variability—Range, Average Deviation [AD], standard deviation [SD] and quartile deviation [QD]—their properties, calculation and application

Measure of variability indicates in quantitative terms the extent to which the scores in a distribution scatter from the centre of the distribution. For describing a distribution adequately we need not only the measures of central tendency but also the measures of variability because the former only locates the centre of the distribution and tells nothing about how the scores or measurements are arranged in relation to the central tendency. The variability of data reveals much more than what the measure of central tendency alone can do. For comparing the relative degree of variability with regard to some trait among the individuals in two or more groups, the measure of variability of the trait in each group is needed.

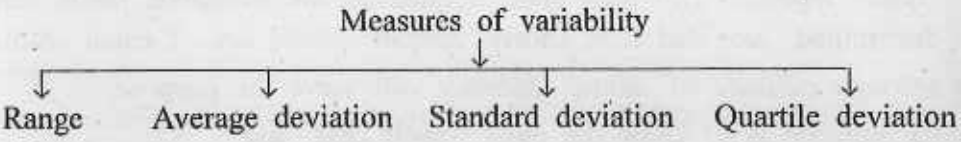
The measures of variability give information about whether or not the group is homogeneous. It is an important characteristic indicating the extent to which observations vary among themselves. The variability of a given set of observation

will be zero, only when all of them are equal. The wider the discrepancy from one observation to another, the larger will be the variability. Where central tendency indicates the size of a typical value representing a whole set of observations, variability denotes the degree of variation. Two series of observations may have same arithmetic mean but unequal variation. For example:

Series 1: 15, 19,22,24, 30 Total 110

Series 2: 5, 8, 22, 25, 50 Total 110

These series have equal number of observations and have the same mean (22) and median (22). However, it may be noticed that observations in series 1 are close to one another, while those in series 2 is more than that of series 1. Therefore, necessity of variability is must where the averages alone are not sufficient to reveal all the characteristics of data. Important measures of variability are as under:



6.4.1 Range

The simplest way to measure the variation among a set of observations is to use that is called the range. The range is the difference between the highest and the lowest scores of a distribution. It is easy to calculate the range of a data but it is difficult to measure the nature of distribution of data around the central location as the value of range is purely depends only on two values (the lowest and the highest) that are independent of central value and distribution of data. It is not reliable measure of variability. Range can fluctuate to a large extent from one sample to another, representing the same population. It is usually used as a quick measureof variability, which is to be supported by a more representative measure.

Properties: The properties of range are given below:

1. It is the difference between the largest and smallest observations.
2. It is simple to understand and easy to calculate.
3. It is not generally used because of its many disadvantages.
4. It does not depend on allobservations.
5. Range cannot be calculated from frequency distributions with open-end classes.
6. It is highly affected by extreme values.

Calculation:

Range can be calculated by applying the formula given below:

Range = $U_L - L_L$

Where, U_L = the upper limit of the distribution

L_L = the lower limit of the distribution

Example 6.9: Calculate the range of the scores, 12, 15,69,25, and 45 of 5 students.

Solution: Range = $U_L - L_L = 69 - 12 = 57$

Application: Range can be applied in the following cases:

1. When the data are too scattered and need to more precise measure of variability.
2. When we want to know the extreme scores and the spread of the distributions.

6.4.2 Average Deviation

The distance of a score from a central point is called deviation or it can be said that it is the difference between any particular value and the mean. It is a measure of how far that value deviates from the mean. The average of the deviation of all values from the arithmetic mean is known as mean deviation or average deviation. In formal terms:

Xi = Xi-M [where Xi = the i^{th} value of the set]

The sum of these deviations is zero and therefore the mean of the deviations would also be zero. This isn't what we wanted. If we use the plus and minus signs to represent direction away from the mean then all the pluses and minuses will cancel each other out when we add the deviations. The answer is to take the mean of all the deviation sizes, and to ignore any minus signs. This is known as taking the absolute value and is represented mathematically by two vertical bars (I) either side of a number. So, for the absolute value of deviation score we would write |X - M| or |x|.

Properties:

The properties of average deviations are the following:

1. It depends on all observations.
2. It is the arithmetic mean of absolute deviations (i.e differences) from mean or any other specified value.

Calculation:

Average deviation can be obtained from two types of data:

i. Average deviation or mean deviation from ungrouped data.

ii. Average deviation or mean deviation from grouped data.

Calculation of two types of data is discussed below:

i. Average Deviation or Mean Deviation from Ungrouped Data

Example 6.10: Find the average deviation for the scores given below: 10, 13, 12, 15,23, and 17

Solution:

Table 6.10: Calculation of average deviation

Scores	x= |X-M|
10	5
13	2
12	3
15	0
23	8
17	2
Total = 90	Σ|X-M| = 20

The mean of the above scores is ($\frac{90}{6}$) = 15 . For calculating the mean deviation we have to subtract the mean from each score and the result should be written in the absolute values which are shown in the second column. Average deviation will be calculated by the following formula:

$$\text{Mean Deviation} = \frac{\Sigma|(X-M)|}{N} = \frac{\Sigma|x|}{N} = \frac{20}{6} = 3.33$$

Where, X = Scores,

x = X–M,

|X - M| = Absolute values of deviation,

$\sum|x - M|$ = Sum of the absolute values of deviation,

N = Total number of scores.

ii Average Deviation or Mean Deviation from Grouped Data

Example 6.11: Calculate mean deviation from the following grouped data.

Table 6.11: Frequency distribution of 63 students

Class Interval	13-17	18-22	23-27	28-32	33-37	38-42	43-47	48-52	53-57
Frequency	2	22	10	14	3	4	6	1	1

Solution:

Table 6.12: Calculation of Average Deviation or Mean Deviation

Class Interval	Frequency (f)	Mid value (X)	$\|X-M\|$	$f\|X-M\|$
13-17	2	15	13.25	26.50
18-22	22	20	8.25	181.5
23-27	10	25	3.25	32.5
28-32	14	30	1.75	24.5
33-37	3	35	6.75	20.25
38-42	4	40	11.75	47
43-47	6	45	16.75	100.5
48-52	1	50	21.75	21.75
53-57	1	55	26.75	26.75
Total	N=63			$\sum f\|X-M\|$ =481.25

Average Deviation or Mean Deviation from Grouped Data $= \frac{\sum f|X-M|}{N}$

$= \frac{481.25}{63} = 7.64$

Application:

Average deviation may be used for the following purpose:

1. When we want to know all deviations from the mean according to their size.
2. When standard deviation are influenced unduly by the extreme deviation.
3. When we want to use the simplest measure of dispersion that takes into account all the values in a given distribution.
4. Those people who have a little knowledge about statistics can apply this method as it is easily comprehensible.

6.4.3 Standard Deviation

Of all the measures of variability, standard deviation symbolized by Greek small letter σ (read as sigma), is the most widely used statistic. The standard deviation is important method to determine and access the scatter-ness of variability in the data around its central value (mean). The standard deviation is closely associated with each other and measures the same trait. It is the square root of the arithmetic mean of the squared deviations of scores taken from their mean. Due to this reason, it has also been called the root-mean-square- deviation. The standard deviation is the square root of the variance and accordingly, variance is the square of the standard deviation.

Properties:

The properties of standard deviation are the following;

1. Standard deviation is rigidly defined.
2. It is based on all observations.
3. It is calculated fairly easily.
4. It is least affected by fluctuations of sampling.
5. The significance of standard deviation is not easy to understand.
6. It is a round-about measure obtained by squaring deviations from mean, finding the average of these squares, and then taking the square-root (called root-mean-square-deviation).

Calculation:

Standard deviation can be calculated from group data as well as from ungrouped data.

i] Standard Deviation of Ungrouped Data: If the data are ungrouped, it may be calculated with either of the following formulas:

a] Standard deviation from mean

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\Sigma(X-M)^2}{N}} = \sqrt{\frac{\Sigma x^2}{N}}$$

[Where X = Scores, M= Mean, x = X-M, and N= Total number of scores]

b] Standard deviation from raw score

$$\text{Standard deviation } (\sigma_x) = \frac{1}{N}\sqrt{N\Sigma X^2 - (\Sigma x)^2}$$

[Where X = Scores and N= Total number of scores]

a] Standard Deviation from Mean: From the following example, it will be cleared how to calculate the standard deviation from mean.

Example 6.12: Calculate the standard deviation from the following distribution:

20, 25, 30, 27, and 23

Solution:

Table 6.13: Calculation of Standard Deviation from Mean

Scores	X= (X–M)	$(X-M)^2$
20	–5	25
25	0	0
30	5	25
27	2	4
23	–2	4
Total = 125		$\Sigma(X-M)^2 = 58$

From the above distribution, Mean = $\frac{\Sigma X}{N} = \frac{125}{5} = 25$

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\Sigma(X-M)^2}{N}} = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{58}{5}} = 3.41$$

b] Standard Deviation from Raw Scores: From the following example, it will

be cleared how to calculate the standard deviation from the raw scores.

Example 6.13: Calculate the standard deviation of the above scores.

Solution:

Table 6.14: Calculation of Standard Deviation from Raw Scores

Scores [X]	X^2
20	400
25	625
30	900
27	729
23	529
Total = 125	$\Sigma X^2 = 3183$

Standard deviation $(\sigma_x) = \frac{1}{N}\sqrt{N\Sigma X^2 - (\Sigma x)^2}$

$= \frac{1}{5}\sqrt{5 \times 3183 - (125)^2}$

$= \frac{1}{5}\sqrt{15915 - 15625}$

$= \frac{1}{5}\sqrt{290}$

$= \frac{17.03}{5} = 3.41$

ii] Standard Deviation from Grouped Data: Standard deviation of grouped data may be calculated by the following two formula:

a] Long Method:

Standard deviation $(\sigma_x) = \sqrt{\frac{\Sigma fx^2}{N}}$,

[Where, x = (X – M), N = Total frequency]

b] Short Method or Assumed Mean Method or step deviation method

Standard deviation $(\sigma_x) = i \times \sqrt{\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2}$

Where, i = Length of class interval,

$x' = \frac{X - AM}{i}$ i.e. deviation of mid-point from the assumed mean [X - AM] and divided by the length of class interval (i),

N = Total frequency

a] Long Method:

For application of long method, the following example may be cited.

Example 6.14: Calculate standard deviation from the following distribution.

Table 6.15: Frequency distribution of 70 students

Class Interval	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84
Frequency	5	7	11	15	12	10	6	4

Solution:

Table 6.16: Calculation of standard deviation from grouped data [Long Method]

Class Interval	Frequency (f)	Mid value (X)	fX	x =(X-M)	x^2	fx^2
5-14	5	9.5	47.5	-32.29	1042.64	5213.2
15-24	7	19.5	36.5	-22.29	496.84	3477.88
25-34	11	29.5	324.5	-12.29	151.04	1661.44
35-44	15	39.5	592.5	-2.29	5.24	78.6
45-54	12	49.5	594	7.71	59.44	713.28
55-64	10	59.5	595	17.71	313.64	3136.4
65-74	6	69.5	417	27.71	764.84	4589.04
75-84	4	79.5	318	37.71	1422.04	5688.16
Total	N=70		$\sum fX$= 2925			$\sum fx^2$= 24558

Solution: The following steps should be followed for calculation of standard deviation:

First step: to find the mean of this distribution.

$$\text{Mean} = \frac{\Sigma fX}{N} = \frac{2925}{70} = 41.79$$

Second step: to find the deviation of the score [midpoint] from the mean.

Third step: to square the deviation and

Fourth step: to find the product of the square deviation and the corresponding frequency.

$$\text{Hence, standard deviation } (\sigma_x) = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{24558}{70}} = \sqrt{350.83} = 18.73$$

Here, we must see that the there are so many values which are. in decimal numbers. Therefore, it is very difficult to calculate. For this reason, we can follow the shortcut method. Here, instead of taking deviations from actual mean, we take deviations from assumed mean. As a result, the formula for finding the standard deviation is given below.

b| Standard Deviation by Assumed Mean Method or Step Deviation Method:

Example 6.15: Calculate standard deviation from the above table 6.15 by assumed mean method or step deviation method

Solution:

Table 6.17: Calculation of standard deviation from grouped data [Short Method]

Class Interval	Frequency (f)	Mid value (X)	$x' = \frac{X - AM}{i}$	fx'	fx'2
5-14	5	9.5	-3	-15	45
15-24	7	19.5	-2	-14	28
25-34	11	29.5	-1	-11	11
35-44	15	39.5	0	0	0
45-54	12	49.5	1	12	12
55-64	10	59.5	2	20	40
65-74	6	69.5	3	18	54
75-84	4	79.5	4	16	64
	N=70			$\Sigma fx' = 26$	$\Sigma fx'^2 = 254$

Here, assumed mean is the midpoint of the class interval 35-44 i.e. 39.5.

Hence, standard deviation $(\sigma_x) = i \times \sqrt{\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2}$

$$= 10 \times \sqrt{\frac{254}{70} - \left(\frac{26}{70}\right)^2}$$

$$= 10 \times \sqrt{3.63 - .14}$$

$$= 10 \times 1.87 = 18.7$$

Here also the standard deviation is 18.7

Application:

Standard deviation may be used for the following purposes:

1. Standard deviation is applicable for greatest stability of statistic.
2. When we want to know the greater effect on variability by the extreme deviations of the data proportionally.
3. For wider use of other statistics.

6.4.4 Quartile Deviation

Before taking up the quartile deviation, we must know the meaning of quartiles. Quartiles may be defined as those values of the variate which divide. the total frequency into four equal parts. Quartiles are denoted by QI, Q2, and Q3. The quartile deviation, symbolized by Q, is defined as half the distance between QI or the 25th percentile and Q3 or the 75th percentile. Q is one half of the distance between these two quartile points [(Q3 - QI) is known as inter-quartile], it is also known as the semi-interquartile range. By definition, QI refers to that point in the distribution, which has 25% of the scores below it and Q3 refers to that point in the distribution, which has 75% of scores below it. Q2 is the 50th percentile or median of the distribution. The value of the variate which divides the lower half below the median, into two equal parts on the basis of frequency, is called Lower Quartile. It is denoted by QI. Similarly, when the upper half which is above the median, is divided on the basis of frequency into two equal parts, the value of the variate is called Upper quartile, denoted Q3. In terms of a formula, Q may be written as under:

$$\text{Quartile Deviation (Q)} = \frac{Q_3 - Q_1}{2}$$

Where Q_3 = Third quartile or 75th percentile

Q_I = First quartile or 25th percentile

Properties:

The properties of the quartile deviation are the following:

1. It is rigidly defined.
2. It is based on all observations.
3. It can be calculated with great ease and has a very simple meaning.
4. It is half the difference between the two quartiles QI and Q3, which include the central 50% of observations.
5. It can be calculated from frequency distributions with open-end classes.
6. It is difficult to decide sampling fluctuations through quartile deviation.

Calculation:

Already we are aware of calculation of the median by using the formula given below:

$$\text{Median} = L_L + \frac{\frac{N}{2} - f_b}{f_m} \times i$$

Where, L_L = Lower limit of the class interval in which median lies

N = Total number of cases

f_b = Cumulative frequency below this class

f_m = Frequency in this class

i = Size of the class interval.

For computation of quartile deviation, the following formula will be used:

$$\text{Quartile Deviation (Q)} = \frac{Q_3 - Q_1}{2}$$

Where Q_3 = third quartile or 75th percentile and

Q_1 = first quartile or 25th percentile.

For computing the quartile deviation (Q) we have to calculate first Q_1 and Q_3

by applying the formula of median. The only difference is that for computing the median we use or consider $\frac{N}{2}$ cases, while for the Q_1 and Q_3 we have to take $\frac{N}{4}$ and $\frac{3N}{4}$ cases respectively. The following example will clear the matter.

Example 6.16: Calculate the quartile deviation from the following frequ distribution.

Table 6.18: Frequency distribution of 64 students

Class Interval	13-17	18-22	23-27	28-32	33-37	38-42	43-47	48-52	53-5
Frequency	2	22	10	13	4	4	6	2	1

Solution:

Table 6.19: Calculation of Quartile Deviation

Class Interval	Frequency (f)	Cumulative frequency (cf) (less than)
13-17	2	2
18-22	22	24
23-27	10	34
28-32	13	47
33-37	4	51
38-42	4	55
43-47	6	61
48-52	2	63
53-57	1	64
Total	N=64	

Here N = 64, for Q_3 we have to take $\frac{3N}{4}$ = 48 cases, and for QI we have to take $\frac{N}{4}$ = 16 cases. It is found that cf 48 will be included in class interval 33-

37. Whose real limits are 32.5-37.5. Again cf 16 will be included in class interval 18-22 whose real limits are 17.5-22.5.

The value of Q_3 will be

$$Q_3 = L_L + \frac{\frac{3N}{4} - f_b}{f_m} \times i$$

$$= 32.5 + \frac{48-47}{4} \times 5$$

$$= 32.5 + 1.25$$

$$= 33.75$$

The value of Q_1 will be

$$Q_1 = L_L + \frac{\frac{N}{4} - f_b}{f_m} \times i$$

$$= 17.5 + \frac{16-2}{22} \times 5$$

$$= 17.5 + 3.18$$

$$= 20.68$$

Therefore the quartile deviation (Q) will be

$$\text{Quartile Deviation (Q)} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{33.75-20.68}{2} = 6.54$$

Application:

Quartile deviation may be used for the following purposes:

1. When the median is used for the best measurement of central tendency.
2. When standard deviation are influenced disproportionately by the scattered or extreme scores.
3. When our primary interest is to know the middle 50% of the scores.

6.5 Percentile and Percentile Rank — Definition, Calculation, Application, and Graphical Determination

6.5.1 Percentile

We know that median divides the total frequency into two equal parts, quartiles divides the total frequency into 4 equal parts, but in case of percentile, the total frequency is divided into 100 equal parts. Therefore, it can be said that percentile is the value of the variate which divide the total frequency into 100 equal parts. It is denoted as PI, P2, P3, ... P 100 . Where P I denotes that 1 percent of the cases lies below this area, in this way for we can say that P50 means 50 percent of the cases lies below this area. P25, P50 and P75 represent the first quartile i.e QI, 2nd quartile i.e Q2 or Median, and third quartile i.e.Q, respectively. Similarly, first, second, third, ninth deciles are represented by PIO, P20, P30, P90 respectively. Hence, we can say that percentile point is the point or value or score of the variates that denotes a certain percentage of cases lies under this area.

Calculation:

For calculating the percentile, first we have to know the points for which the specified percent of cases to be obtained. By using the following formula we can find the percentile.

$$P_P = L_L + \frac{\frac{Np}{100} - f_b}{f_m} \times i$$

Where P = Percentile

p = the relevant percentile point required

L_L = Lower limit of the class

N = Total number of cases

f_b = Cumulative frequency below the class

fm = Original frequency in this class

i = Size of the class interval

Suppose if we want to find P_{10} and P_{25} then the formula will be

For P_{10}

$$P_{10} = L_L + \frac{\frac{10N}{100} - f_b}{f_m} \times i$$

For P_{25}

$$P_{25} = L_L + \frac{\frac{25N}{100} - f_b}{f_m} \times i$$

and so on.

Example 6.17: Find out P_{80} from the following frequency distribution.

Table 6.20: Frequency distribution of 24 students

Class Interval	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54
Frequency	1	2	2	5	6	4	2	1	1

Solution:

Table 6.21: Calculation of P_{80}

Class Interval	Frequency (f)	Cumulative frequency (cf)
10-14	1	1
15-19	2	3
20-24	2	5
25-29	5	10
30-34	6	16
35-39	4	20
40-44	2	22
45-49	1	23
50-54	1	24
Total	N=24	

We can obtain the value of P_{80} by applying the following formula:

$$P_{80} = L_L + \frac{\frac{80N}{100} - f_b}{f_m} \times i$$ [Where, L = 34.5, N=24, f_b = 16, fm= 4, and i=5]

$$= 34.5 + \frac{19.2 - 16}{4} \times 5$$

$= 34.5 + 4 = 38$

Hence, we can say that 38 is the value of Pso that represents 80 percent cases lie below the score 38.

Graphical Determination:

By using the above example, determination of Pso from graph which 1S presented below:

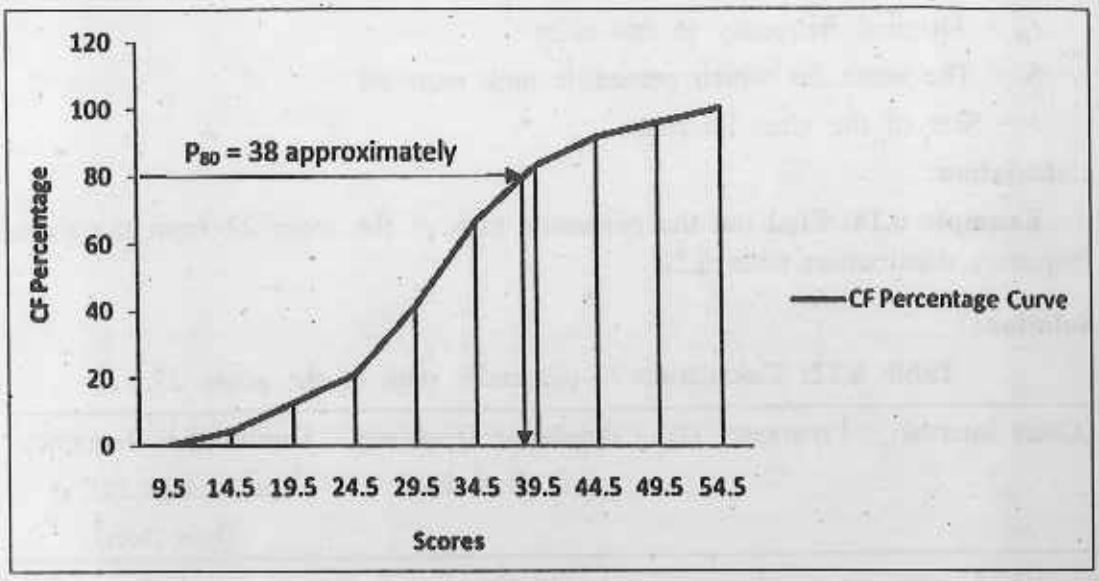


Figure 2.1: Showing the score of P_{80}

To find the percentile point of 80 [P_{80}] we locate it on the y-axis and cast horizontally towards until the curve in intersected. Then we read vertically towards the x-axis and the point reached is the percentile point. From the above graph it is observed that the percentile point of 80 [P_{80}] is 38 approximately that means 80 percent of the individual lies under the score 38.

6.5.2 Percentile Rank

When any score or value of the variates is expressed in percentage area then it is called percentile rank. Suppose, the score of 40 is to be converted into its percentile rank. After calculation it is found that percentile rank is P_{85} of the score 40 that means score 40 is expressed in percentile rank P_{85}.

The percentage cumulative frequency may be used in determining the percentile rank of a score.

$$P_R = f_b + \left(\frac{S - L_L}{i} \times f_m \right) \times \frac{100}{N}$$

Where P_R = Percentile Rank

L_L = Lower limit of the class

N = Total number of cases

f_b = Percentage cumulative frequency below the class

f_m = Original frequency in this class

S = The score for which percentile rank required

i = Size of the class interval.

Calculation:

Example 6.18: Find out the percentile rank of the score 27 from the above frequency distribution table 6.20.

Solution:

Table 6.22: Calculation of percentile rank of the score 27.

Class Interval	Frequency (f)	Cumulative frequency (cf) (less than)	Cumulative frequency percentage (cf%) (less than)
10-14	1	1	4.17
15-19	2	3	12.5
20-24	2	5	20.83
25-29	5	10	41.67
30-34	6	16	66.67
35-39	4	20	83.33
40-44	2	22	91.67
45-49	1	23	95.83
50-54	1	24	100
Total	N=24		

We can obtain the percentile rank of score 27 by applying the following formula:

$$P_R = F_b + \left(\frac{S-L}{i} \times f_m\right) \times \frac{100}{N}$$

[Where F_b= 20.83, S = 27, i = 5, L = 24.5, fm = 5, N = 24]

$$= 20.83 + \left(\frac{27-24.5}{5} \times 5\right) \times \frac{100}{24}$$

$$= 20.83 + 2.5 \times \frac{100}{24}$$

$$= 20.83 + 10.42$$

$$= 31.25$$

Graphical Determination:

By using the above example, determination of the score 27 from graph which is presented below:

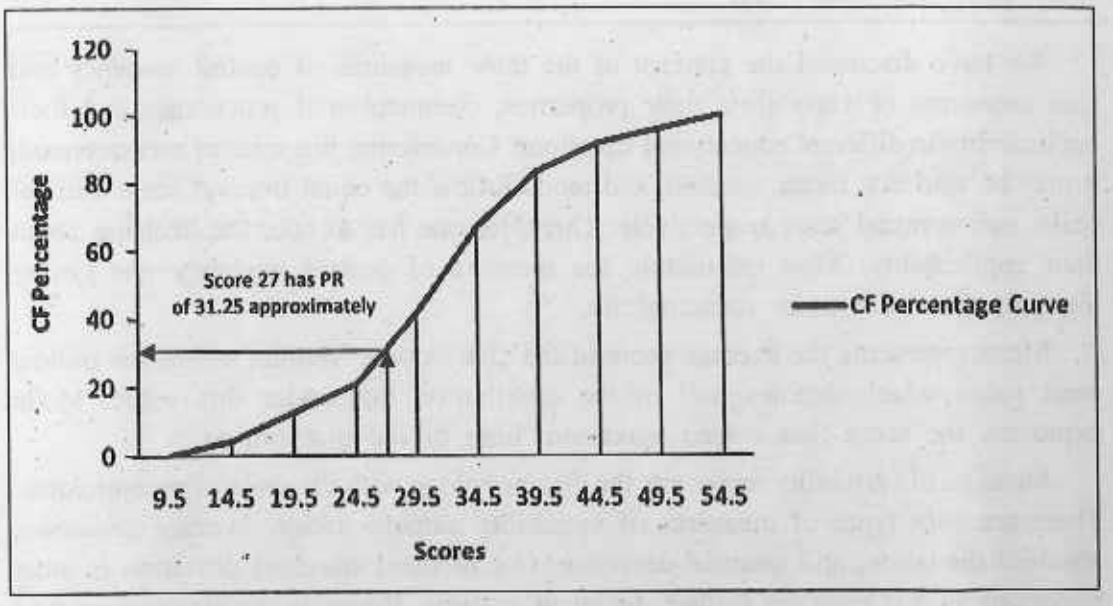


Figure 6.2: Showing the percentile rank of the score 27

At first, to find the percentile rank of the score 27 we locate it on the x-axis [represents the score] and cast vertically upward until the curve in intersected. Then

we read horizontally towards the y-axis [represents the cumulative frequency percentage] and the point reached is the percentile rank. From the above graph it is observed that the percentile rank of the score of 27 is 31.25 approximately that means 31 percent of the individual lies under the score 27.

Application:

The applications of the percentile and percentile rank are the following:

1. Percentile use frequently for interpreting the test score e.g. PIO is equal to score 40 that means 10 percent of the students are below the score 40.
2. When we want to know the certain percentage of cases under the score then percentile rank be used.
3. Both are used in a standardized test as a percentile norm for interpreting the test scores.
4. Both are used for comparison of two groups or individual.

6.6. Let us sum up

We have discussed the concept of the three measures of central tendency and four measures of variability, their properties, computational procedure, and their applicability in different educational situations. Considering the scale of measurement, it may be said that mean, median, and mode follow the equal interval scale, ordinal scale, and nominal scale respectively. Therefore one has to take the decision about their applicability. After calculating the measure of central tendency one has to interpret the result more meaningfully.

Mean represents the average score of the distribution. Median means the middle most value which denotes half of the distribution lies under this value. Mode expresses the score that comes maximum time of the distribution

Measure of variability represents the distance along with the scale of measurement. There are four types of measures of variability namely- range, average deviation, standard deviation, and quartile deviation. Out of them standard deviation is most important as it is used for further statistical analysis. Range is the simplest method among all the measures of variability.

Percentile and percentile rank are used for describing score and the position respectively in a set of observation which denotes that a certain percentage lies under the score.

6.7 Cheque Our Progress (10 questions : 5 SA Type 5 Objective Type)

Short answer [SA) type questions:

1. Define mean. .
2. List two conditions for best use of mean, median, and mode.
3. Compute mean or median of following frequency distribution of 75 students:

Scores	20-24	25-29	30-34	35-39	40-44
f	9	16	25	15	10

4. Define standard deviation.
5. Defme percentile and percentile rank.

Objective type questions:

1. Nominal scale of measurement is appropriate to which measure of central tendency?
2. Median is associated to which scale of measurement?
3. Find the mean, median, and mode of the following scores:
   10, 12, 11, 13,5, 15, 10
4. Compute range for the following data:
   25,56,42,20,35,41 392831,23
5 .....is called the root-mean-square-deviation (Fill in the blank).

6.8 References

Best, John W. and Kahn, James V. (1992). Research in education. Delhi: Prentice Hall of India.

Garrett, H.E. (1962). Statistics in psychology and education. Bombay: Allied Publishers Pvt. Ltd.

Guilford, J.P. (1965). Fundamental statistics in psychology and education. Ne York: McGraw-Hill Book Co.

Koul, Lokesh (1997). Methodology of educational research (3rd revised edn.). New Delhi : Vikas Publishing House Pvt. Ltd.

Mangal, S. K. (2002). Statistics in psychology and education. Delhi: Prentice-Hall of India.

Unit-7 ❑ Normal Distribution and Derived Score

STRUCTURE:

7.1 Introduction

7.1.1 Definition of Normal Probability Curve (CPC)

7.1.2 Major Characteristics of Normal Probability Curve (NPC)

7.2 Skewness

7.2.1 Types of Skewness

7.2.2 Measures of skewness

7.3 Kurtosis

7.3.1 Types of Kurtosis

7.3.2 Measurement of Kurtosis

7.3.3 Limitations of NPC

7.4 Derived Score

7.4.1 Types of Derived Score

7.4.2 Standard scores (Z-score and T-score)

7.5 Z-Scores

7.5.1 How to interpret z-scores

7.5.2 Charecteristics of a Z-score distribution

7.5.3 Uses of Z-score

7.5.4 Advantage and disadvantage of Z-score

7.6 T-Scores

7.6.1 Uses of T-score

7.6.2 Advantage and disadvantage of T-score

7.6.3 Difference between z-score and t-score

7.7 Bibliography

7.1 Introduction:

The main focus of statistical treatment is to make raw data meaningful. There are various statistical techniques and devices which are employed for the interpretation and presentation of data. The interpretation is usually done on the basis of statistical analysis and presentation made by using graphical devices. The graphical devices are employed for the visual presentation of data, which is easy to understand. It serves mainly two purposes of statistical treatment.

- To understand the nature of the frequency distribution of data in terms of skewness, kurtosis and modality.
- To transform the raw data into standard scores in the form of percentiles and z-score or (σ-score which are easily interpretable

Similarly normal distribution is another graphical device for transforming raw scores into standard score or z-score or-o-score whose mean is zero and (J is one. In this chapter normal distribution curve or normal probability curve (NPC) has been discussed in detail.

NPC is the frequency polygon of any normal distribution. It is an ideal symmetrical frequency curve and is supposed to be based on the data of a population. It is bell shaped curve and a graph representing a distribution of scores.

7.1.1 Definition of Normal Probability Curve (NPC):

According to 1.P Guilford

"Normal probability curve is a well-defined, wellstructured, mathematical curve, have a distribution of three scores mean and S.D. one the curve itself is a mathematical conception. It does not occur in nature it is not biological or psychological curve."

Laplace and Gauss (1777-1855), derived the normal probability curve independently, so the curve is also known as gaussian curve in the honor of Gauss.

First recorded in 1890-1895 "a bell-shaped curve showing a particular distribution of probability over the values of a random variable. Also called Gaussian curve, probability curve. Origin of normal curve.

The NPC can be defined with the help of its major features-

"The distribution which is unimodel, skewness is zero and kurtosis is 263, mean is zero "and σ is one is known as normal probability curve."

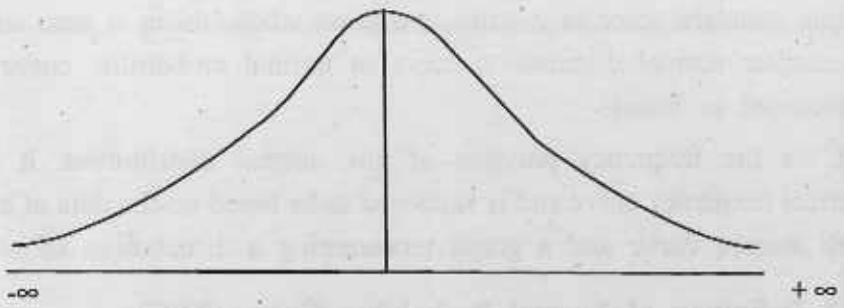
7.1.2 Major Characteristics of Normal Probability Curve (NPC):

The major characteristics ofNPC are limited which can be easily remembered which are as follows-

1. The curve is bilaterally symmetrical: The curve is symmetrical to its ordinate of the central point of the curve. It means the size, shape and slope of the curve on side of the curve is identical to the other side of the curve. If the curve is bisected then its right hand side completely matches to the left hand side.

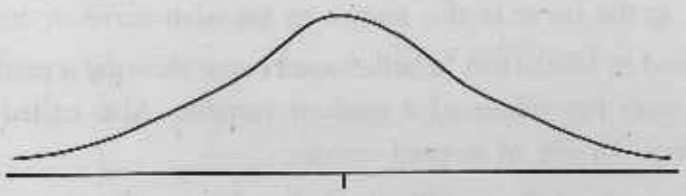
2. The curve is asymptotic: The normal probability curve approaches the horizontal axis and extends from

$-\infty$ to $+\infty$. Means the extreme ends of the curve tends to touch the base line but never touch it. It is depicted in figure no. 1 given below.



3. It is an uni-modal curve:

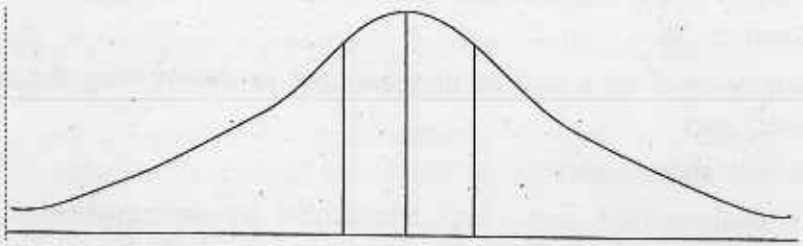
The main feature of frequency distribution is modality or mode. A distribution has one mean, one median but may have more than one mode. The uimodel curve which has its mode .3989 or 39.89 percent frequency lie at mean is called normal curve. Uni-Modal Distribution



Uni-Modal Distribution

4. The points of inflection occur at ± 1 standard deviation unit:

The points of influx in a NPC occur at ±1 unit above and below the mean. Thus at this point the curve changes from convex to concave in relation to the horizontal axis.

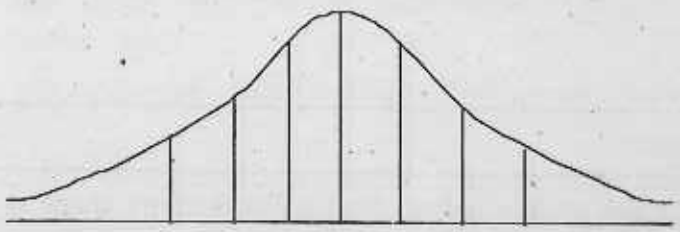


5. The measures of central tendency are equal: The mean,median and mode fall at the middle point and they are numerically equal (Mean=Median=Mode).

6. The height of the curve symmetrically declines: The height of the curve is .3989. The 39.89 percent frequencies concentrate at its centre. So the height of the curve decline to both the directions symmetrically from the central point. Means the M+ 0' and M-O' are equal if the distance from the mean is equal.

7. The total area of NPC is divided into± 3 standard deviations:

The total-area of NPC is divided into six standard deviation units. From the centre it is divided into three '+ve' standard deviation units and three '-ve' standard deviation units. This ± 3 of NPC include different number of cases separately. Between ± 1σ lie the middle 2/3rd cases or 68.26%, between ± 2σ lie 95.44% cases and between ± 3σ lie 99.73% cases and beyond ± 3σ only 0.37% cases fall.



8. In Normal Probability Curve the standard deviation is the 50% larger than the Q: In NPC the Q is generally called the probable error or PE. The relationship- between PE and σ can be stated as following.

9. The mean of NPC is μ and the standard deviation is σ : As the mean of the NPC represents the population mean. So it is represented by μ (meu). The standard deviation of the curve is represent by the Greek letter, σ.

Other Characteristics of Normal Probability Curve are—

- The mean of the normal distribution is zero SD is equal to one. The mean is equal to 30'
- Q can be used as a unit of measurement in determining the area within a given part.
- It is bell shaped curve.
- It is mathematical curve and represented by an equation. The X-axis indicates σ-scores and V-ordinate represents frequency of the normal distribution.
- The curve has its maximum height or ordinate at the starting point.The mean of the distribution.
- The total area under the curve is taken arbitrarily to be 10,000, for a greater ease in the computation.

Measuring Divergence from normality:

We have discussed above the Normal Probability Curve is a symmetrical curve, means the Mean, Median and Mode coincide together. So there is a perfect balance between the right side and the left side of the curve. But in a natural situation this may not be the case always. In the distribution the Mean, Median and Mode may not coincide together. This type of distribution cannot be called as a normal distribution. This is a distribution which diverges from normality. These diverges are of two types. .

- Skewness
- Kurtosis

7.2 Skewness:

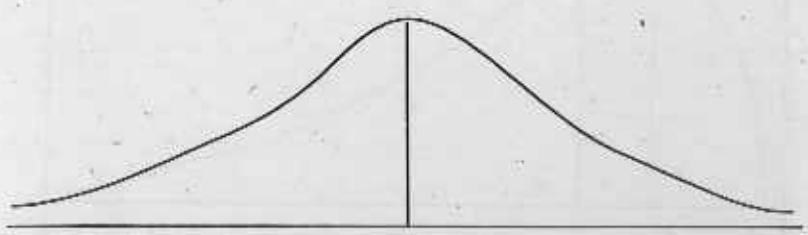
The first thing you usually notice about a distribution's shape is whether it has one mode (peak) or more than one: If its unimodal (has just one peak), like most data sets, the next thing you notice is whether it's skewed to one side. If the bulk ofthe data is at the left and the right tail is longer, we say that the distribution is skewed right; if the peak is toward the right and the left tail is longer, we say that the distribution is skewed left.

Skewness refers to lack of symmetry. A normal curve is a perfect symmetrical curve. In many distributions which deviate from the normal, the value of mean, median and mode are different and there is no symmetry between the two halves of the curve. Such distributions are said to be skewed.

The skewness statistic is sometimes also called the skewedness statistic. Normal distributions produce a skewness statistic of about zero. (I say "about" because small variations can occur by chance alone). So a skewness statistic of -0.0 1819 would be an acceptable skewness value for a normally distributed set oftest scores because it is very close to zero and is probably just a chance fluctuation from zero. As the skewness statistic departs further from zero, a positive value indicates the possibility of a positively skewed distribution (that is, with scores bunched up on the low end of the score scale) or a negative value indicates the possibility of a negatively skewed distribution (that is, with scores bunched up on the high end of the scale). Values of 2 standard errors of skewness (ses) or more (regardless of sign) are probably skewed to a significant degree.

Skewness is asymmetry in a statistical distribution, in which the curve appears distorted or skewed either to the left or to the right. Skewness can be quantified to define the extent to which a distribution differs from a normal distribution. In a normal distribution, the graph appears as a classical, symmetrical "bell-shaped curve." The mean or average, and the mode or maximum point on the curve, are equal.

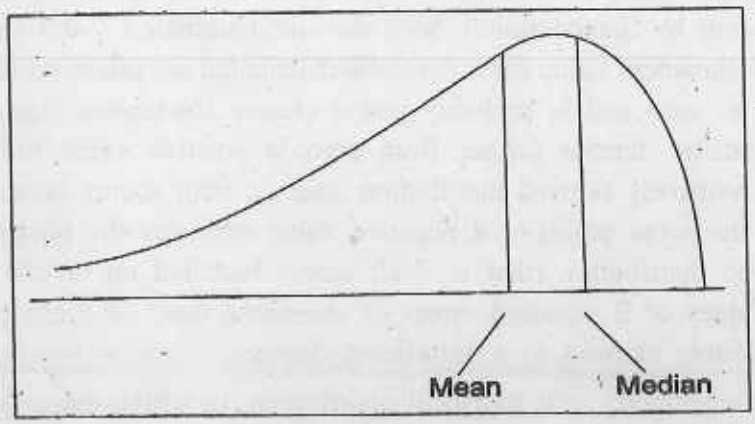
Skewness denotes the tendency of distribution to depart from symmetry. According to Simpson, "Skewness or asymmetry is the attribute of a frequency distribution thatextends further on one side of the class with the highest frequency that on the other"!'



Mean = Median = mode

7.2.1 Types of Skewness:

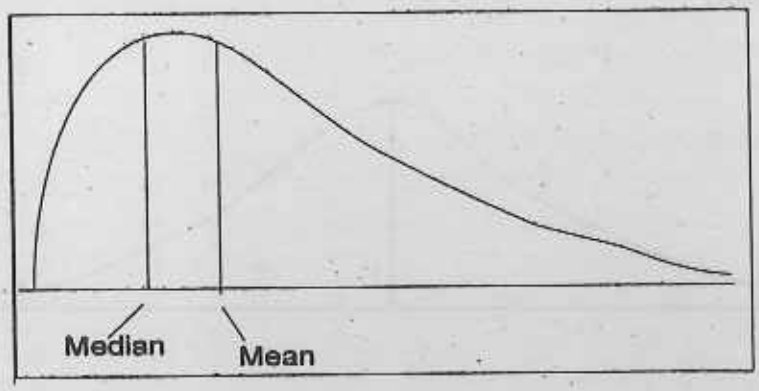
When in a distribution the scores are massed at the high end of the scale to the right end and are spread out more gradually towards the left side, at that time the distribution is said to be negatively skewed. Also in this case, Mean ≤ Median < Mode.



Negatively Skewed (median > mean)

In the negatively skewed distribution the Median is greater than the Mean. So when the skewness is negative the Mean lies to the left of the median.

Similarly when in a distribution the scores are massed at the low end of the scale i.e. to the left end and are spread out more gradually to the right side at that time, the distribution is said to be positively skewed.



Positively Skewed (Median < Mean)

In the positive skewed distribution the Median is less than the mean. So when the skewness is positive the Mean lies to the right of the median. Positively skewed data is also called right skewed, right-tailed, skewed to the rig

Skewness	Distribution Shape	Calculated Value
Positive Skewness	Tail to the Right, values extend further to the right but concentrated in left	Mean > Median > Mode
Zero Skewness	Bell shaped or symmetrical	Mean = Median = Mode
Negative Skewness	Tail to the left, values can extend further to the left but concentrated in the right	Mean < Median < Mode

7.2.2 Measures of skewness:

The direction and extent of skewness can be measured in variousbways. We shall discuss three measures of skewness in this section.

1. Karl Pearson's coefficient of skewness
2. Bowley's coefficient of skewness
3. Kelly's coefficient of skewness

1. Karl Pearson's coefficient of skewness:

• Karl Pearson's Measure of Skewness:

The skewness can be observed and is also calculated by using the following formula.Hence Karl Pearson's coefficient of skewness is defined in t m s of median as

$$SK = \frac{3(Mean - Median)}{\sigma}$$

Where: SK = Coefficient of Skewness.

σ = Standard deviation distribution.

• Measure of Skewness in Terms of Percentiles:

Another formula is also used for computing coefficient of skewness of a

distribution. In this method we can compute skewness from percentile.

$$SK = \frac{P_{90} + P_{10}}{2} = P_{50}$$

Where SK = Skewness

P_{90} = 90^{th} Percentile

P_{10} = 10^{th} Percentile

P_{50} = 50^{th} Percentile or median.

Advantage and disadvantage of Karl Pearson's coefficient of skewness:

- **Advantage:**

  SK is independent of scale. Because (mean-mode) and standard deviation have same scale and it will be cancelled out when taking the ratio.

- **Disadvantage:**

  SK depends on the extreme values.

2. Bowley's coefficient of skewness:

- This measure is based on quartiles. For a symmetrical distribution, it is seen that Q_1 and Q3 are equidistant ftom median. Thus (Q3 - Md) - (M, - Q_1) can be taken as an absolute measure of skewness. A relative measure of skewness, known as Bowley's coefficient (SQ),is given by

$$S_Q = \frac{(Q_3 - M_D) - (M_D - Q_1)}{(Q_3 - M_D) + M_D - Q_1)}$$

$$= \frac{Q_3 - 2M_D + Q_1}{Q_3 - Q_1}$$

Advantage and disadvantage offiowley's coefficient of skewness:

Advantage:

SQdoes not depend on extreme values.

Disadvantage:

SQdoes not utilize the data fully.

3. Kelly's coefficient of skewness:

- Bowley's measure of skewness is based on the middle 50% of the observations because it leaves 25% of the observaticins on each extreme of the distribution.

As an improvement over Bowley's measure, Kelly has suggested a measure based on P_{10} and, P_{90} so that only 10% of the observations on each extreme are ignored. Kelly's coefficient of skewness, denoted by Sp, is given by

$$S_P = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{50}) + (P_{50} - P_{10})}$$

$$= \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

7.3 Kurtosis:

In Greek language kurtosis means 'bulgines'. kurtosis indicates the nature of the vertex of the curve. Several statisticians defined kurtosis. Some of these definitions are:

"In statistics, kurtosis reefers to' the degree of flatness of peaked ness in the regionabout the mode of frequency curve. The degree of kurtosis of a distribution is measuredrelative to the peaked ness of normal curve."

When there are very few individuals whose scores are near to the average score for their Group, the curve representing such a distribution becomes 'flattened' in the middle. On the other hand, when there are too many cases in the central area, the distribution curve becomes too 'peaked' in comparison with the normal curve. Both these characteristics of being flat or peaked are used to describe the term kurtosis.

"kurtosis characterizes the relative peakedness or flatness of a distribution compared to the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution" (Microsoft, 1996).

"A measure of kurtosis indicates the degree to which a curve of the frequency distribution is peaked or flat-topped."

Balanda and MacGillivray say the same thing in another way: increasing kurtosis is associated with the "movement of probability mass from the shoulders of a distribution into its center and tails."

Kurtosis means the 'Peakedness' or flatness of a frequency distribution compared to the normal distribution. The Collins Dictionary of Statistics defines kurtosis as "the sharpness of a peak on a curve of a probability density function." The Normal

Probability Curve is moderately peaked or flatter than the NPC.

7.3.1 Types of Kurtosis: There are three types of Kurtosis. Surch as—

a. Leptokurtic (KU > O.263)

b. Mesokurtic (KU = O.263)

c. Platykurtc (KU < O.263)

a. Leptokurtic:

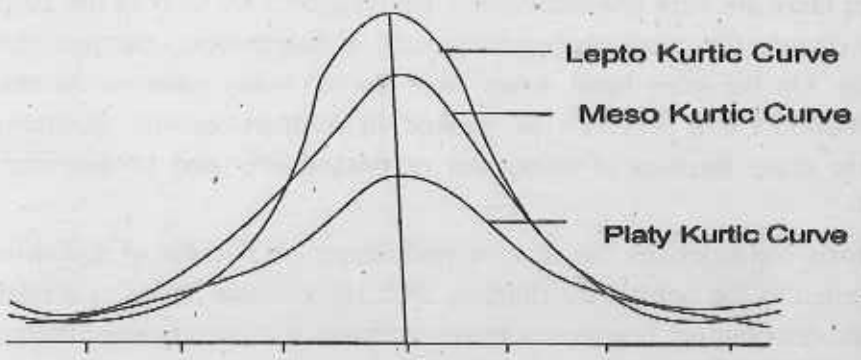
When the frequency distribution is more peaked at the centre then the normal curve is called as Leptokurtic. The value of kurtosis of a Leptokurtic Curve is greater than 0.263.

b. Mesokurtic:

When the frequency distribution is normally distributed the curve is Mesokurtic. The kurtosis of a normal curve is 0.263.

c. Platykurtc:

When a frequency distribution is flatter than the normal curve it is called as platykurtic. The value of kurtosis of a platykurtic curve is less than 0.263.



Term	Distribution Shape	Kurtosis
Leptokurtic	Peaked	Greater than 0.263
Mesokurtic	Normal	0.263
Platykurtic	Flat	Less than 0.263

The quantitative value for kurtosis is calculated by finding the fourth power of the deviation of each data value from the mean value.

- A normal (mesokurtic) distribution will have a zero value for kurtosis.
- The flatter playkurtic distribution will have a negative value for kurtosis.
- The more pointed, leptokurtic distribution will have a positive value for kurtosis.

To compute kurtosis, we use the following formula:

$$Ku = \frac{Q}{P_{90} - P_{10}}$$

Where Ku = Kurtosis
Q = Quartile deviation
P_{90} = 90^{th} Percentile
P_{10} = 10^{th} Percentile

7.3.2 Mesurement of Kurtosis:

Kurtosis is another measure of the shape of a distribution. Whereas skewnessmeasures the lack of symmetry of the frequency curve of a distribution, kurtosis is a measure of the relative peakedness of its frequency curve. Various frequency curves can be divided into three categories depending upon the shape of their peak. The three shapes are termed as Leptokurtic, Mesokurtic and Platykurhc as shown

Mesure of kurtosis is denoted by β_2 and in a normal distribution β_2 = 3.If β_2 is greater than 3, the curve is more peaked and is named as leptokurtic. If β_2 is less than 3, the curve is flatter at the top than the normal, and is named as platykurtic. Thus kurtosis is measured by β_2 where x = (X–) R.A. Fisher had introduced another notation Greek letter gamma symbolically.$Y_2 = \beta_2 - 3 = 3$. In this case of a normal distribution, Y_2 is zero. Y_2 is more than zero (positive), then the curve is platykurtic and if Y_2 is less than 0 (negative) then the curve is leptokurtic.

It may be noted that $\mu 4$ = is an absolute measure of kurtosis but β_2 = is a relative measure of kurtosis. Larger the value of Y_2 in a frequency distribution, the greater is its departure from normality. β_1 and β_2 are measures of symmetry and normality respectively. If β_2 = 0, the distribution is symmetrical and if β_2 = 3, the distribution curve is mesokurtic.

Example: From the following data given below, calculate the value of kurtosis and find out the nature of distribution:

X	0-10	10 - 20	20-30	30-40	40-50
F	5	10	15	10	5

Solution:

Calculation of Mean = Calculation of β_2 = where, $\mu 4 = \mu 2 = \beta_2$ = since the value of $\beta_2 = 3$, the distribution curve is mesokurtic.

Why is the normal distribution useful?

- Many things actually are normally distributed, or very close to it. For example, height and intelligence are approximately normally distributed; measurement errors also often have a normal distribution
- The normal distribution is easy to work with mathematically. In many practical cases, the methods developed using normal theory work quite well even when the distribution is not normal.
- There is a very strong connection between the size of a sample N and the extent to which a sampling distribution approaches the normal form. Many sampling distributions based on large N can be approximated by the normal distribution even though the population distribution itself is definitely not normal.

Applications of the normal Probability Curve:

It has a wide application in behavioural science, bio-sciences in field of measurement, research statistics and administration.

I. **In Administration:** The normal probability curve (NPC) is used in classifying the group and assigning grades to the individuals. The raw scores are transformed into standard scores for comparing two or more groups. A group is divided in to sub-groups according to abilities of the individuals.

II. **In Statistical Treatment or Analysis:** The use of statistical technique depends on the nature of distribution of data. Parametric techniques are used in normal distribution and non - parametric techniques are used for skewed distribution or free distribution of data.

III. **In Developing Normal or Standard Score:** The NPC indicates the σ-scores or standard scores by transforming raw scores. Thus the norms of a test are developed in the form' of σ-scores. It is known as Z-scores

norms or σ–score norms. The percentiles are also obtained by the use of NPC.

IV. **In Research Work:** In behavioural science a representative sample is required because the findings of the study are based on sample observations. A representative sample may yield data which are normally distribution. The application of statistical treatment is based on the distribution of data. The accuracy of the inference depends on the use of proper statistical analysis.

V. **Measurement and Test- Construction:** The standardization oftest implies that it yields normal distribution otherwise test may be very easy or vary difficulty. The obtained score are normalized for developing norms of the test.

VI. **Comparison of Two Distributions:** The NPC is used to compare two distributions terms of overlapping on the assumption that these distributions are normal.

VII. **Items Difficulty Values:** The Z- scores are also used to determine the relative difficulty values of test items, problems and other test items.

VIII. **Levels of Significance:** The levels of significance of statistical results are determined in terms of NPC limits. The .05 level of significance indicates the 95 percent confidence interval, similarly .01 level of significance means 99 percent confidence interval.

The principles of Normal Probability Curve are applied in the behavioural science in different areas. Some of them are as following:

- **NPC is useful to normalize frequency distribution:**

  In order to normalize a frequency distribution we use Normal probability Curve. For the process of standardizing a psychology test this process is very much necessary .

- **NCP is used to generalize about population from the sample:**

  We compute standard error of mean, standard error of standard deviation and other statistics to generalize about the population from which the sample is drawn. For this computation we use the table area under NPC .

- **NCP is used to determine the percentage of cases in a normal distribution within given limits**

  The Normal Probability Curve helps us to determine:

- What percent of cases fall between two scores of a distribution?
- What percent of scores lie above a particular score of a distribution?
- What percent of scores lie below a particular score of a distribution?

Example 1.

Given a distribution of scores with a mean of 24 and o of 8. Assuming normality, what percentage of the-cases will fall between 16 and 32?

Solution:

Here first of all we have to convert both the scores 16 and 32 into a standard score (Z score; discussed in next section).

$$Z = \frac{X}{\sigma}$$

Where Z = Standard score

X =Deviation of the score (X-M).

σ = standard deviation.

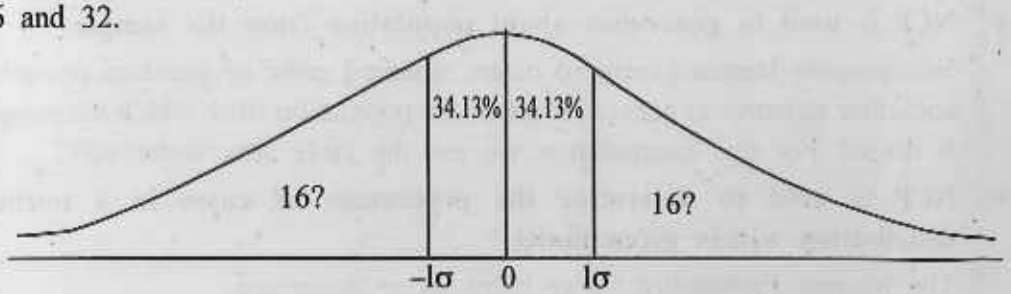
Z score for 16 =

$$\frac{X}{\sigma} = \frac{16-24}{8} = \frac{-8}{8} = -1\sigma$$

Z score for 32 =

$$\frac{X}{\sigma} = \frac{32-24}{8} = \frac{8}{8} = +1\sigma$$

In the NPC table 1σ is located in the first column. There are 34.13% cases between mean and 1σ as shown in figure above the mean. Similarly the case 34.13% are between - 1σ and mean. The total cases between -1σ and + 1σ = 34.13% + 34.13% = 68.26%. The cases above 1σ are (5000 – 3413) = 1587 and also below -1σ. The percent of cases above + 1σ is 16%. So that the 68.26% case will fall between 16 and 32.



Example 2.

Given a distribution of scores with a mean of 40 and o of 8. Assuming normality what percentage of cases will lie above and below the score 36?

Solution:

First of all we have to convert the raw score 36 into standard score.

$$Z = \frac{X}{} = \frac{36-40}{8} = \frac{-4}{8} = -0.5$$

In the table area under the NPC it is found that 19.15% cases fall between mean and -0.5σ.

Therefore the total percentage of cases above the score 36 is 50 + 19.15 = 69.15% and below the scores are above the score 36 and 30.85% score are below the score 36 .

- **NPC is used to determine the value of a score whose percentile rank is given:**

  By using NPC table we can determine the raw score of the individual if the percentile rank is given.

Example 03. In a distribution of scores of a class Pinky's percentile rank in statistics is 65. The mean of the distribution is 55 with a standard deviation of 10. Find out the raw score of Pinky in statistics.

Solution: As Pinky's percentile rank is 50 in a normal distribution, her position is 35% above the mean. By entering in the NPC table, we found that 35% from the mean is + 1.04σ.

By putting the value in 'Z' score

$$\frac{X-M}{\sigma} = Z$$

$$\frac{X-55}{10} = 1.04$$

$X - 55 = 10 \times 1.04$

$X - 55 = 10.40$

$x = 10.40 + 55 = 64.4$ or 65

So Prinky's raw score in statistics is 65.

- **NPC help to determine the relative difficulty of test items problems:**
  It is known that what percentage of students successfully solved a problem we can determine the difficulty level of the item or problem by using table area under NPC .
- **NPC is used to find the limits in a normal distribution which include a given percentage of cases:**
  When a distribution is normally distributed and what we know about the distribution is mean and the standard deviation at that time by using the table area under NPC we can determine the limits which include a given percentage of cases.

Example 04.

Given a distribution of scores with a mean of 20 and a 0' of 5. If we assume normality what limits will include the middle 75% of cases.

Solution:

In a normal distribution the middle 75% cases include 37.5% cases above the mean and 37.5% cases below the mean. From the table NPC we can say that 37.5% cases cover 1.15σ units.Therefore the middle 75% cases lie between mean and ± 1.15σ units.

As $\sigma = 5$

So $5 \times 1.15\sigma = \pm 5.75$ units

Adding the value to mean we can get:

$20 + 5.75 = 25.75$

$20 - 5.75 = 14.25$

So in this distribution middle 75% cases will include the limits 14.25 to 25.75

- **To test the significance of observations in experiments we use NPC:**
  In an experiment we test the relationship among variables are due to chance fluctuations or errors of sampling procedure or it is real relationship. This is done with the help of able area under NPC.
- **It is used to compare two distributions in terms of overlapping:**
  If scores of two groups on a particular variable are normally distributed, what we know about the group is the mean and standard deviation of both the groups. And we want to know how much the first group over-laps the

second group or vice-versa at that time we can determine this by using the table area under NPC.

Example 05:

A student obtains 80 marks in Maths and 50 in English. If the mean and SO for the scores in Maths are 70 and 20 and for the score in English are 30 and 10 find out in which subject, Maths or English, he did better?

Solution:

Here, from the given data, direct comparison of his status in Maths and in English cannot be made because the marks achieved do not belong to the same scale of measurement.

For putting them into a common scale, let us convert these two raw scores into z scores.

Here,

Raw scores in maths (X_1) = 80, M_1 = 70 and σ_1 = 20, and

Raw score in Englisht X_2) = 50, M_2 = 30 and σ_2 = 10

Therefore,

Z score in Maths,

$$= \frac{X_1 - M_1}{\sigma_1} = \frac{80-70}{20} = 0.5$$

Z score in English,

$$= \frac{X_2 - M_2}{\sigma_2} = \frac{50-30}{10} = 2.0$$

We can thus conclude that he did better in English than in Maths .

- **NPC helps us in dividing a group into sub-groups according to certain ability and assigning the grades:**

  When we want to divide a large group in to certain sub-groups according to some, specified ability at that time we use the standard deviation units of a NPC as units of scale.

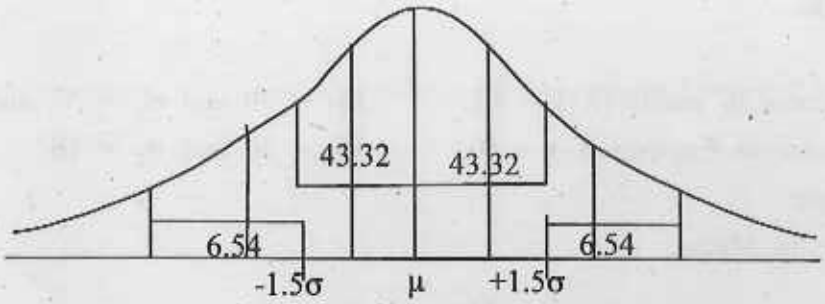
Example 06:

An achievement test was administered to the 600 students of 8th grade. The teacher wants to assign these students in to grades namely A, B, C, and D according to their performance in the test. Assuming the normality of the distribution of scores calculate the number of students can be placed in each group.

Solution:

The area under a NPC is divided into ±3cr or 6cr units. Here we have to divide the students into 4 sections.

So each section has $\frac{6\sigma}{4} = 1.5\sigma$ Units.



So if we shall distribution the section in order of merit.

Tile Section 'A' will be within 1.5σ to 3σ.

Section 'B' will be within mean to -1.5σ.

Section 'C' will be within mean to -1.5σ.

Section 'D' will be within -1.5σ to - 3σ.

Entering into the Table 'A' , we found that:

3σ covers the 49.86% cases.

1.5σ covers the 43.32% cases.

Grade 'A' will cover = 49.86 - 43.32 = 6.54% cases.

Table area within μ to + 1.5σ covers 43.32% cases.

So grade B will cover 43.32% cases.

Table area with in μ and -1.5σ covers 43.32% cases.

So Grade C will cover 43.32%.

Table area within μ and -3σ covers 49.86% and -1.5σ covers 43.32% cases. So between -1.5σ and -3σ, 49.86% - 43.32% =6.54% cases lie.

7.3.3 Limitations of NPC:

The NPC has the following limits as a statistical technique in analysing and interpreting scores.

- It is a well-defined, well-structured mathematical curve; therefore it cannot be used with so accuracy as mathematics is an exact and precise. There is approximation in behavioural sciences.
- The σ-values do not reveal the position or rank of an individual in his group as percentil does.
- The σ-values are in negative while human measurement is not done in negative score. The negative value does not make sense.
- The obtained distribution of data needs to be normalized before transforming score in σ-score.
- The NPC cannot be used without the NPC table. The table assumes the area of the curve 10,000.
- No variable or behavioural science, social sciences and bio-science is normally distributed. Simply it is assumed that the distribution is normal. Thus it involves error.

7.4 Derived Score:

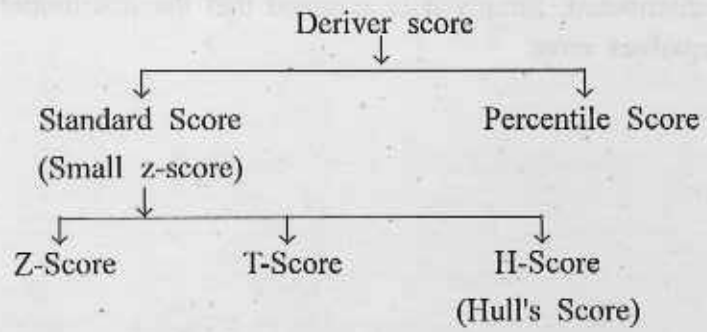
When an individual's performance is expressed numerically it is called 'row scores'. According to Gronlund and Linn (1995, p338) "A row score is simply the number of points received on a test when the test has been scored according to the direction". From a row score we can't say anything about 'how good a score is'? 'What was the difficulty level of the test?' Therefore in order to make a row score meaningful we must convert it into a 'derived score'. A derive score is the "score resulting when a row score is converted to some system of comparable measure, preferably a system having a standard reference point and equal units."

The derived scores help us to know the position of an individual in his group and we can compare the performance with others. "A derived score is a numerical description of an individual's performance in terms of norms."

According to Overton (2012)"Derived scores are obtained by using a raw score and expectancy table".

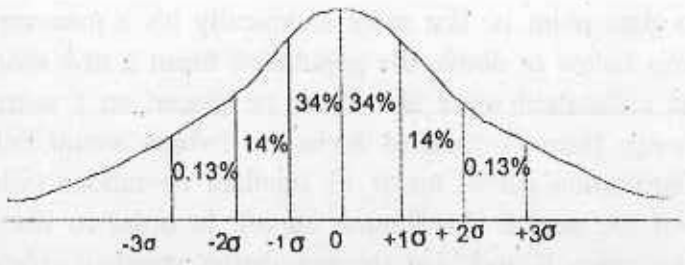
Derived scores can be several types of scores which are percentile ranks, standard scores, grade equivalents, age equivalents, or language quotients etc.

7.4.1 Types of Derived Score:



7.4.2 Standard scores (Z-score and T-score):

Standard scores also indicate the relative position of pupil in a group by showing how far the row score is above or below average. The standard scores express the performance of pupils in standard deviation units. The meaning of standard deviation and standard scores are based on Normal Probability Curve (NPC).



Percentage of cases coming under ±3σ units.

NPC is a systematical bell-shaped curve that has many useful mathematical properties. One of such property is that when it is divided in to standard deviation (σ) units each portion under the curve contains a fixed percentage of cases. This property helps for the interpretation of test scores.

In NPC between mean and ±1σ, 34% cases fall, between ±1σ to 2σ, 14% cases fall, between ± 2σ to 3σ, 2% of the case fall and only 0.13% cases fall beyond ± 3σ.

7.5 Z-Scores:

Z-Scores are a transformation of individual raw scores into a standard form, where the transformation is based on knowledge about the standardization sample's mean and standard deviation. The formula for computing Z-scores is the individual raw score (X) minus the mean of the scores obtained by the standardization sample (M), divided by the standard deviation of scores obtained by the standardization sample (σ). Z-scores have a mean of 0 and a standard deviation of I. A score that is one standard deviation below the mean has a Z-score of -1. A score that is at the mean would have a Z-score of O. The formula for transforming a raw score into a Z-score is a follows:

$$Z = \frac{X - M}{\sigma}$$

Because of the fact that the pulse (+) and minus (-) signs can easily get lost when looking at this type of standard score, Z-scores are frequently converted into other types of standard scores. Specifically they are often transformed into Deviation IQ scores, T -scores, and scaled scores.Z- score is the number of standard deviations

from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is. A z-score is also known as a standard score and it can be placed on a normal distribution curve. Z-scores range from -3 standard deviations (which would fall to the far left of the normal distribution curve) up to +3 standard deviations (which would fall to the far right of the normal distribution curve). In order to use a z-score, you need to know the mean μ and also the population standard deviation (σ).

Z-score (or SO-score) = (observed value - median value of the reference population) / standard deviation value of reference population.

For example, in a test of mathematics Susmita has secured 60 marks and in a test of English she has secured 65 marks. The mean of the mathematics test is 50 and σ = 6. The mean of English test is 62 and σ = 5. In which subject Susmita has a better performance?

Z score of Mathematics is

$$Z = \frac{60-50}{6} = 1.67\sigma$$

Z score of English is

$$Z = \frac{65-62}{5} = 0.6\sigma$$

7.5.1 How to interpret z-scores:

- A z-score less than 0 represent an element less than the mean.
- A z-score greater than 0 represents an element greater than the mean.
- A z-score equal to 0 represents an element equal to the mean.
- A z-score equal to 1 represents an element that is 1 standard deviation greater than the mean; a z-score equal to 2, 2 standard deviations greater than the mean; etc.
- A z-score equal to -I represents an element that is 1 standard deviation less than the mean; a z-score equal to -2, 2 standard deviations less than the mean; etc.
- If the number of elements in the set is large, about 68% of the elements have a z-score between -1 and 1; about 95% have a z-score between -2 and 2; and about 99% have a z- score between -3 and 3.

7.5.2 Characteristics of a Z-score distribution:

1. When Z-score distribution is based on the sample mean and sample standard deviation, then the mean and standard deviation of the Z-score distribution will equal zero and one respectively.
2. When Z-score distribution is based on the population mean and population standard deviation, then the mean and the standard deviation of the Z-score distribution will only approximate to zero and one if the sample is random.
3. The shape of a distribution of z scores is reflective; the shape is the same as the shape of the distribution of the original.
4. If the original distribution is normal, then the Z-score distribution will be normal.
5. If the original distribution is skewed, then the Z-score distribution will also be skewed.
6. The Z-scores are not based upon the known population mean and standard deviation, but on an external reference population. In this situation the Z-scores are used to identify those individuals in the sample falling below a specified Z-score.
7. The variance of a distribution of z scores is one.

7.5.3 Uses of Z-scores:

1. **To identify the position of observation(s) in a population distribution**

   This is the comrnones use of Z-scores. Converting a measurement to a Z-score indicates how far from the mean the observation lies in units of standard deviations. The basic data are age, sex, weight and height. The three preferred indices are:

   a. Height-for-age:

   Low height for age is an indicator of stunting, which is usually associated with poor economic conditions or repeated exposure to food shortage caused by drought or war. In other words chronic malnutrition. -

   b. Weight-for-height:

   This is an indicator of wasting and is associated with failure to gain weight or a loss of weight. In other words, it is called acute malnutrition.

   c. Weight-for-age:

   This reflects the effects of either wasting or stunting, or both.

The Z-score of a child's 'weight-for-height' (for example) is computed using the child's weight together with the population mean weight and standard deviation for that height derived from a set of reference values.

The WHO has strongly advocated the use of Z-scores in nutrition studies, although other methods are still used in some countries.

2. **To eliminate a factor in the analysis by expressing data relative to the mean**

   In field studies, it is quite common to find that the response variable you are studying is affected by the one (or more) explanatory variables in addition to the variable you are most interested in. One way to remove the effect of an explanatory variable is to standardize the data to the mean value of each level of that variable.

3. **To standardize data for subsequent display and/or analysis.**

   When wish to compare the distributions of two variables that are measured in quite different units, then converting measurements to Z-scores enables them to be displayed on the same axes. There are also a number of analyses you will meet (for example major axis regression, described in Unit and cluster analysis) where you convert measurements to Z-scores prior to analysis.

7.5.4 Advantages. and disadvantages of Z-score:

Advantages of Z scores:

- They permit us to convert raw scores into a common scale which has equal units and which can be readily interpreted.
- They give us an idea of how well a teacher-made test is. A good teacher-made test designed to discriminate among students will generally have a range between 4 and 5 SDs, i.e., 2.0 to 2.5 SDs on either side of the mean.
- The advantage of the z score transformation is that it takes into account both the mean value and the variability in a set of raw scores.
- Z-scores are also sex-independent, thus permitting the evaluation of children's growth status by combining sex and age groups.
- These characteristics of Z-scores allow further computation of summary statistics such as means, standard deviations, and standard error to classify a population's growth status.

- Z-scores have the same statistical relation to the distribution of the reference around the mean at all ages, which makes results comparable across age) groups and indicators.

Disadvantages of Z scores:

- They involve the use of decimals and negative numbers.
- The main disadvantage of standard scores is that they always assume a normal distribution. But if this assumption is not met, the scores cannot be interpreted as a standard proportion of the distribution from which they were calculated. For example, if the distribution is skewed, the area with the standard deviation of 1 to the left of the mean is not equal to the area within the same distance to the right of the mean.

7.6 T-Scores:

A t score is one form of a standardized test statistic (the other you'll come across in elementary statistics is the z-score). The t score formula enables you to take an individual score and transform it into a standardized form>one which helps you to compare scores.

T-scores refer to "any set of normally distributed standard scores that has a mean of 50 and a standard deviation of 10".

Z-scores can be transformed into T -scores scores by multiplying the given Z-score by 10 (the standard deviation of the distribution ofT-scores), and adding 50 (the mean of the distribution of T-scores) to this product. The formula used to compute 'T' is as following:

T-score = 50 + 10 (Z)

Form our earlier example we have Z score of 1.67 in Mathematics and 0.60 in English. By converting these two into T - score:

T-score of Mathematics = 50 + (10 × 1.67) = 66.7

T-score of English = 50 + (10 × 0.6) = 56

From the above data we can say that the performance of Susmita in Mathematics is certainly better than the performance in English. One important merit of reporting test results in T - score is that only positive integers are produced. Therefore the interpretation in T -score is very much simple.

Characteristics of T score:

- The mean of the distribution is equal to O.
- The variance is equal to v / (v - 2), where v is the degrees of freedom (see last section) and v > 2.
- The variance is always greater than 1, although it is close to 1 when there are many degrees of freedom.

7.6.1 Uses of T scores:

The general rule of thumb for when to use a t score is when your sample:

- Has a sample size below 30,
- Has an unknown population standard deviation.

7.6.2 Advantages and disadvantages of T score:

Advantages of T score:

The T score has two major advantages:

- It is always a whole number, and
- It is never a negative number.
- The range ofT scores is from 1 to 1 00 with a mean of 50.

Disadvantages of T score:

You need to know. the original raw/obtained test's scores mean and standard deviation to get back to your true raw score. On most tests (including the SASSI-3) the fixed mean is 50 and fixed standard deviation is 10 both constants.

7.6.3 Difference between z-score and t-score:

Difference between z-score and t-score are—

- The Z score is scaled down by the population standard deviation. The T score is scaled down by the sample standard deviation.
- The z-distribution is the "standardized" normal distribution. The normal distribution is the familiar bell-shaped curve, symmetric, with almost all of the data falling within three standard deviations around the mean and the t-distribution is also symmetric and mound shaped but its exact shape depends on the sample size.
- With smaller samples the t distributions has "fatter tails" and is more shallow and flat than Z score.

7.7 Bibliography:

Elhance, D. N. and V. lhance (1988) Fundamentals of Statistics, Kitab Mahal, Allahabad.

Nagar, A. L. and R. K. Dass (1983) Basic Statistics, Oxford University Press, Delhi.

Mangal, S.K (2010), Statistics in Psychology and Education, PHI Learning Private Limited, New Delhi.

Nanda, G.C & Khatoi, P. K (2005) Fundamentals of Educational Research and Statistics, Kalyani Publishers, New Delhi.

Garrett, H.E. (1981) Statistics in Psychology and Education, Vakils, Feffer and Simons Ltd., Bombay.

Sharma, R.A (2008), Elementary Statistics in Educational and Psychology, Meerut: R. Lall book depot.

Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: Harper Collins.

Shenoy, G.V., Srivastava V.K., and Sharma, S.C., 1989, Business Statistics, Wiley Eastern: New Delhi.

Winer, BJ (l971). Statistical Principles in Experiment at Designs, London: McGraw Hill Kogakusha Ltd.

Swain, S. K and Khatoi, P.K (2000): Educational Measurement, Statistics and Guidance,Kalyani Publishers, Ludhiana.

Sancheti, D.C., and Kapoor, V.K., (1989), Statistics Theory Methods and Applications, Sultan Chand &Sons: New Delhi.

Fisher, R.A. (1970): Statistical Methods for Research Works, 14th ed. Hanfner Publishing Co. Inc., New York.

Koul, L:(l990): Methodology of Educational Research 2nd Ed. Vikas Publishing House Pvt. Ltd., New Delhi.

HTML: http://wwwjalt.org/testlbro_l.htm/PDF:

http://www.jalt.org/testlpDF/Brownl.pdf

Unit : 8 ❑ Relationship and Inferential Statistics

STRUCTURE:

8.1 Introduction

8.2 Objectives

8.3 Concept and Definition of Correlation

8.3.1 Concept of Correlation

8.3.2 Definitions of Correlation

8.3.3 Characteristics of Correlation

8.3.4 Need/Uses of Correlation

8.4 Types of Correlation

8.4.1 Positive Correlation (+)

8.4.2 Negative Correlation (–)

8.4.3 Zero Correlation or no Correlation

8.4.4 Linear Correlation

8.5 Coefficient of Correlation

8.5.1 Degree of Correlation

8.5.2 Method of Studying Correlation

8.5.2.1 Spearman's Rank difference method

8.5.2.2 Karl Pearson's coefficient of correlation

8.5.2.3 Computation of r when the data are grouped

8.6 Interpretation of Pearson's Correlation Coefficient

8.7 Let us sum up

8.8 Unit End Exercises

8.9 Bibliography

8.1 Introduction:

We have so far discussed the statistical description of a single variable. Now we shall study the problem of describing the degree of simultaneous variation

between two variables. The data in which we obtain measures of two variables for each individual is called a bivariate data. For example, we get bivariate data if we have scores oftests of Mathematics and English for a group of school children. The essential tenets of the bivariate data are that one measure can be paired with another measure for each member of the group.

When we study bivariate data we may like to know the degree of relationship between variables of such data. This degree of relationship between variables is known as 'correlation' which is represented by the co-efficient of correlation. This co-efficient may be identified by either the 'r', the Greek symbol rho (p), other symbols depending upon the data distributions and the way the coefficient has been calculated.

8.2 Objectives:

After careful study the module, we will be able to:

1) Realize the meaning and definition of correlation.
2) Understand the characteristics of correlation.
3) Be acquainted with different the types of correlation.
4) Understand the need or uses of correlation
5) Understand the various measures of relationship.
6) Understand the various methods of computing coefficient of correlation

8.3 Concept and Definition of Correlation:

8.3.1 Concept of Correlation:

We have studies some characteristics of one variable only, which are mean of the distribution of height, standard deviation of weight, skewness of the distribution of income. But, man situation arise in which may have to study two variables simultaneously, say x and y. For example the variables may be :

- Ages of husband and wife.
- The height and weight.
- Intelligence and achievement.

There are main problems involved in such studies, the data may reveal some association between x and y, and we may be interested to measure numerically the strength of this association between the variables. Such the measure will determine how well a linear or other equation explains the relationship between the variables. This is problem of correlation.

Correlation. is another way of assessing the relationship between variables. To be more precise, it measures the extent of correspondence between the ordering of two random variables.

Correlation refers to the relationship between two variables. Two variables are said to be correlated when change in one variable appears to be accompanied by a change in other variable. It is frequently necessary to determine the degree of relationship that exists between sets of scores representing two or more traits or abilities, or between sets of scores obtained for other reason. For this purpose the statistical techniques of correlation is used. Correlation is a statistical tool that helps to measure and analyze the degree of relationship between two variables. Correlation analysis deals with the association between two or more variables. According to Conner, *"If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in the other(s) than they are said to be correlated".*

8.3.2 Definitions of Correlation:

Croxton and Cowden:

"When the relationship is of a qualitative nature, the approximate statistical tool for discovering and measuring the relationship and expressing in the brief formula is known as correlation."

Collins Dictionary of Statistics:

"If two variables are such that, when one changes, the other does so in a related manner, they are said to be correlated."

Dictionary of Education C.V. Good:

"Correlation is the tendency for corresponding observations in two or more series to vary together from the averages of their respective series that is to have similar relative position."

A.M. Tutte said "Correlation is an analysis of the co variation between two or more variables."

W.A. Neiswanger said "Correlation analysis contributes to the understanding of economic behaviour, aid is locating the critically important variables on which others depend, may reveal to the economic the connections by which disturbances spread and suggest to him the paths through which stabilizing forces may become effective."

According to Tippett "The effect of correlation is to reduce the range of uncertainty of our prediction."

According to Simpson and Kofka "Correlation analysis deals with the association between two or more variables."

According to English & English Dictionary, "Correlation is a relationship or dependence. It is the fact that two things. or variables are so related that change in one is accompanied by a correspondingor parallel change in the other."

According to Ferguson "Correlation is concerned with describing the degree of relation between two variables."

According to Guilford "A coefficient of correlation is a single number that tells us to what extent two things are related, to what extent variation in one goes with variations in the other."

According to Lathrop "Correlation is ajoint relationship between two variables."

Whenever two variables of the same group are so related that the increase or decrease are correspond to the increase or decrease to another or conversely, increase or decrease corresponds to the increase or decrease to another, they are said to correlated

8.3.3 Characteristics of Correlation: Main characteristics of correlation are

a) Correlation is a measures of relation to indicate the strength of association between variables;
b) Coefficient of correlation is a ratio ranging over a scale from -1.00 to + 1.00 through 0 (zero relation);
c) Measures of relationship deals with more than one variable of the same sample;

d) It needs the simultaneous variation of the variables;

e) Positive correlation indicates the large amount of the one variable tend to accompany large amount of the other;

f) Negative correlation indicates that small amount of the one variable tend to accompany large amount of the other;

g) Zero correlation indicates no consistent relationship.

8.3.4 Need/Uses of Correlation:

Correlation gives meaning to a construct. Co relational analysis is very much essential for educational and psychological research. Following are the some major fields where it is idel used.

- It is used to determine the reliability and validity oftest results.
- Testing whether certain data is consistent with hypothesis.
- Predicting one variable on the basis of other related variable(s).
- Building psychological and educational models and theories.
- Grouping variables/measures for parsimonious interpretation of data.
- Carrying multivariate statistical tests (Hoteling's T^2; MANOVA, MANCOVA, Discriminant analysis, Factor Analysis).
- To identify extraneous variable(s) and to isolate their effect in an experiment.
- To calculate further statistics based on the co-efficient of correlation.

8.4. Types of Correlation:

To have a clear understanding of the concept of correlation we must discuss different types of correlation. In a bivariate distribution the relationship may be categorized in to different types.

1. Positive Correlation (+).
2. Negative Correlation (-).
3. Zero Correlation or no correlation.
4. Linear Correlation.
5. Non-linear or curvilinear Correlation.

8.4.1 Positive Correlation (+):

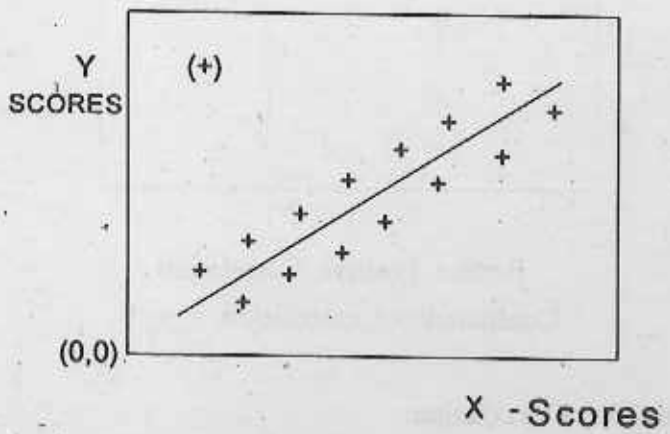
When increase or decrease in one variable brings corresponding increase or decrease in the other variable or when two variables move in the same direction then the correlation between these two variables is said to be Positive Correlation.

Positive Correlation Example:

Training and performance are two variables, which are positively correlated. When training increases, performance also increases; when training decreases, performance decreases. Other example in education field are—

a. High school students who had high grades also had high scores on the Scholastic Assessment Test.
b. As attendance at school drops, so does achievement.
c. When enrollment at college decreases, the number of teachers decreases.
d. As a student's study time increases, so does his test average.
e. The more years of education you complete, the higher your earning potential will be.
f. College students with higher Scholastic Assessment Test scores. also have higher grades in college.

Positive Correlation Graph:



Types of Positive Correlation:

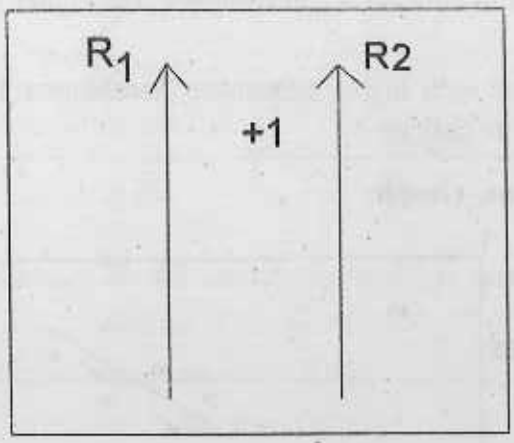
We see two types positive correlation. Which are -

I. Perfect Positive Correlation.

II. Weak Positive Correlation

I. Perfect Positive Correlation: When every unit increase or decrease in one variable is followed by proportional increase or decrease in the other variable the relationship is perfect positive correlation. A positive relationship ranges from 0 to + 1 when it is + 1 the correlation is perfect positive correlation. So a perfect positive correlation has variables that have the same changes, but the point is more close together and forms a line.

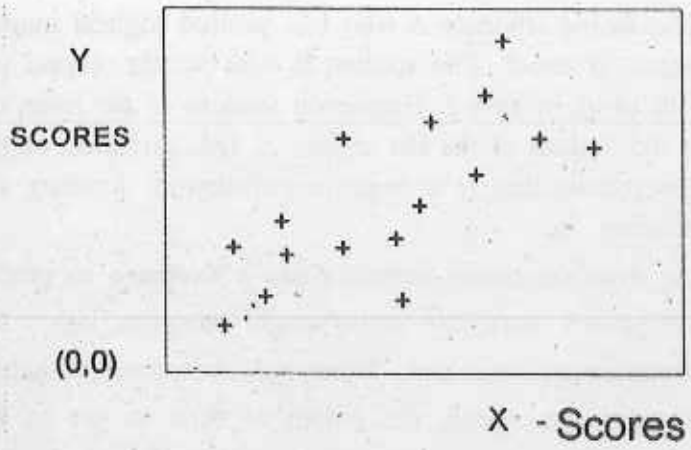
Perfect Positive Correlation Example: Suppose 100 students have exactly the same standing in two tests, the students who scores first in the test scores first, in other the student who ranks second in the first also rank second in the second test. This one to one correspondence holds throughout the entire list. So the relation is perfect positive correlation, since the relative position of each subject is exactly the same in one test as the other. So the coefficient of correlation is + 1.



Perfect Positive Correlation
Coefficient of correlation = + 1

II. Weak Positive Correlation:

A weak positive correlation has variables that have the same changes but the points on the graph are dispersed.



A weak correlation means that as one variable increases or decreases, there is a lower likelihood of there being a relationship with the second variable. In visualization with a weak correlation, the angle of the plotted point cloud is flatter. If the cloud is very flat or vertical, there is a weak correlation. Earthquake magnitude and the depth at which it was measured is therefore weakly correlated, as you can see the scatter plot is nearly flat.

8.4.2 Negative Correlation (-) : In this type of correlation, the two variables move in the opposite direction. So when a high degree of one trail or variable is associated with a low degree of another is called negative correlation. Where increase in one variable result in decrease in other variable and vice-versa the relationship is said to be negative correlation. The negative correlation may range from 0 to -1.

Example of Negative Correlation: Suppose in a test 5 students A, B, C, 0, and E have secured, 85, 77, 74, 68 and 67 marks. In the second test they have secured 41,44,51, 55 and 59 respectively.

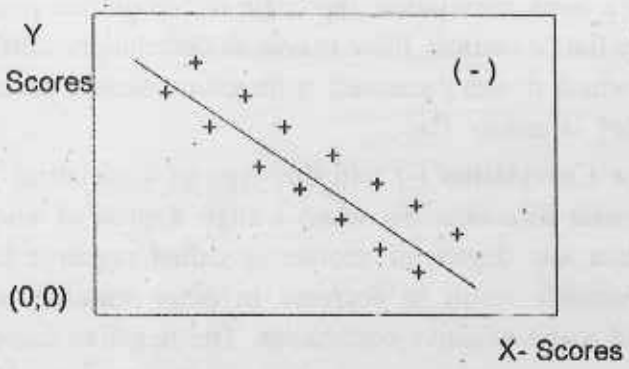
Table No. : 1

Students	Scores in Test-l	Score in Test-2
A	85	41
B	77	44
C	74	51
D	68	55
E	67	59

In the above example students A who has secured highest marks in test-1 has secured lowest mark in test-2. The student B who stands second in Test-1 ranks next to the bottom (4th) In Test-2. Here each student as far from the top the list in Test-1 as from the bottom of the list in Test-2. This direction of the relationship is inverse. So this relationship is a negative correlation. Another some examples of negative correlation are-

- A student who has many absences has a decrease in grades.
- As one exercises more, his body weight becomes less.
- As the temperature increases, fewer hot chocolate products are sold.
- If a train increases speed, the length of time to get to the final point decreases.

Graph of Negative Correlation:



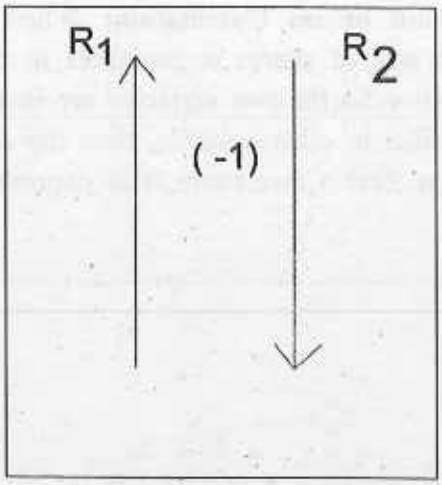
Types of Negative Correlation:

Generally the negative correlation can be classified in two ways-

a. Perfect negative Correlation.

b. Weak negative Correlation.

a. Perfect negative Correlation: When every unit of increase in one variable brings proportionate decrease in the other variable the relationship is called perfect negative correlation. A perfect negative correlation means the relationship that exists between two variables is negative 100% of the time. The coefficient of correlation is indicated by-1.

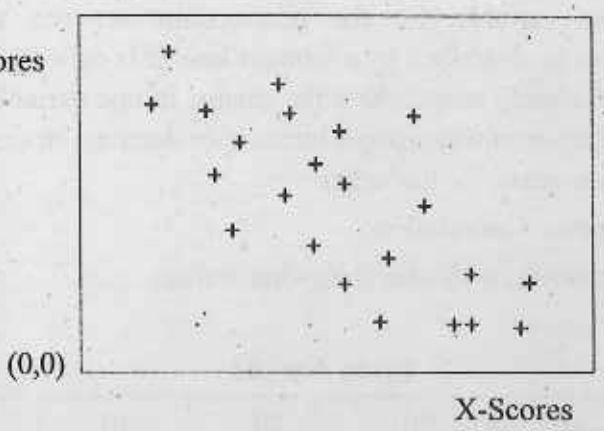


Perfect Negative Correlation

Coefficient of correlation = –1

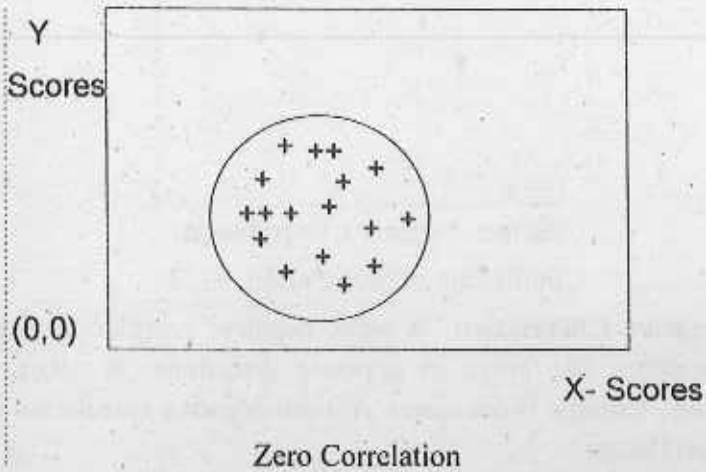
b. Weak negative Correlation: A weak negative correlation is a relationship between two variables that move in opposite directions. In other words, when variable A increases, variable B decreases. A weak negative correlation is also known as an inverse correlation.

Two variables can have varying weak of negative correlation. The variable A could be weak negatively correlated with B, and may have a correlation coefficient of -0.2 means that for every unit change in variable B, variable A experiences a decrease, but only slightly, by 0.2.



Weak negative Correlation

8.4.3 Zero Correlation or no Correlation: When there is no systematic relationship between two sets of scores or variables in that case it is known as zero-agreement or no relation.So the two variables are independent and the change in one variable has no effect in other variable, then the correlation between these two variables is known as Zero Correlation. It is expressed by the coefficient of 00.



Zero Correlation

Coefficient of correlation = 0

8.4.4 Linear Correlation: When the change in one variable result in the constant change in the other variable. So the relationship between two variables is proportional and it can be described by a straight line, it is called Linear Correlation. Suppose a correlation clearly reveals how the change in one variable is accompanied by a change in the other or to what extent increase or decrease in one is accompanied by the increase or decrease in the other.

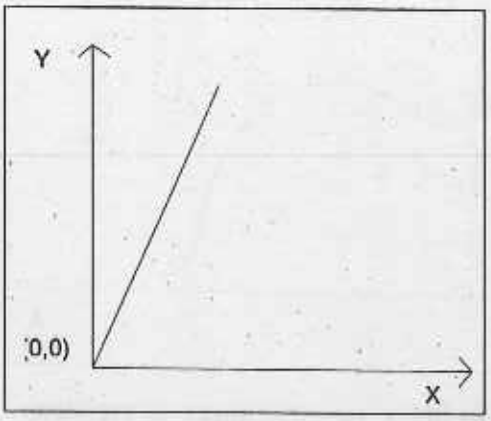
Example of Linear Correlation:

Consider the variables with the following values.

Table No: 02

X	10	20	30	40	50
Y	20	40	60	80	100

Here, there is a linear relationship between the variables. There is a ratio 1:2 at all points. Also, if we plot a graph showing them they will be in a straight line.



Linear Correlation

2 Non-linear or Curvilinear Correlation:

When the relationship between two variables is not proportional throughout the series and it can be described by a curve line is called as Non-line or Curvilinear Correlation. When the amount of change in one variable is not in a constant ratio to the change in the other variable, we say that the correlation is non-linear.

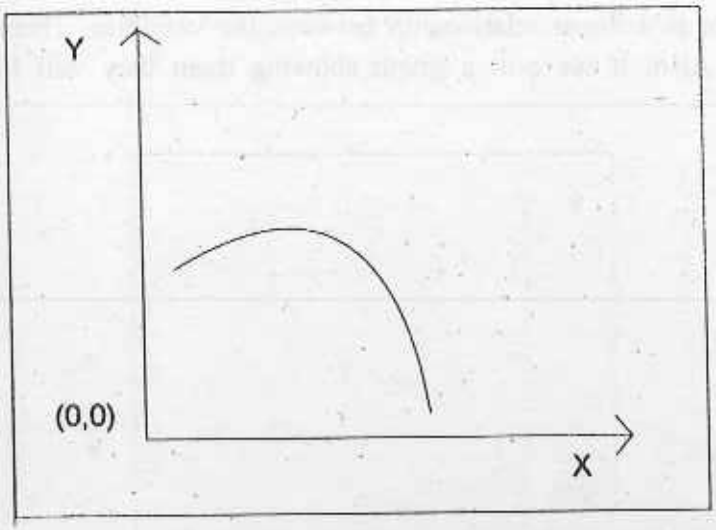
Example of Non-linear or Curvilinear Correlation:

Consider the variables with the following values

Table No: 03

X	10	20	30	40	50
Y	10	30	70	100	130

Here there is a non-linear relationship between the variables. The ratio between them is not fixed for all points. Alsoif we plot them on the graph, the points will not be in a straight line. It will be a curve.



Non-linear or curvilinear Correlation

8.5 Coefficient of Correlation:

The correlation coefficient is a statistical measure that calculates the strength of the relationship between the relative movements of the two variables. It is a kind of ratio which expresses the extent to which changes in one variable are accompanied with changes in the other variable. According to C. V. Good "Coefficient of correlation is a pure number, varying usually from + 1 through 0 to -1, that denotes the degree of relationship existing between two (or more) series of observations. So we have discussed—

- A correlation coefficient of 1 means that for every positive increase in one variable, there is a positive increase of a fixed proportion in the other. For example, shoe sizes go up in (almost) perfect correlation with foot length
- A correlation coefficient of -1 means that for every positive increase in one variable; there is a negative decrease of a fixed proportion in the other. For example, the amount of gas in a tank decreases in (almost) perfect correlation with speed .
- Zero means that for every increase, there isn't a positive or negative increase. The two just aren't related.

8.5.1 Degree of Correlation: The degree of correlation is expressed by the

value of coefficient of correlation which ranges between + 1 and -1. The value of coefficient of correlation directly indicates the degree of correlation as detailed in the following table no 04:

Table No: 04

Value of Correlation Coefficient:

Sl. No.	Value of correlation of Coefficient	Degree of Correlation
01	+1 Value	Perfect positive correlation
02	-1 Value	Perfect negative correlation
03	Value between I to 0.75	High degree of correlation
04	Value between 0.75 to 0.5	Moderate degree of correlation
05	Value between 0.5 to 0.0	Low degree of correlation
06	0 (Zero)	No correlation

8.5.2 Method of Studying Correlation: There are seven methods of computing coefficient of correlation from a bivariate distribution. But we discuss first three types method.

1. Spearman's rank difference method.
2. Karl Pearson's coefficient of correlation(Covariance method)
3. Scatter diagram or datagram method.
4. Coefficient of correlation by concurrent deviation.
5. Graphic method.
6. Two-way frequency table (Bivariate correlation method)
7. Kendall's Tau (non-parametric, can be non-linear)

8.5.2.1 Spearman's Rank difference method: Sometime we come across statistical series in which the variables under consideration are not capable of quantitative measurement but can be arranged in serial order. This happens when we are dealing with qualitative characteristics (attributes) such as honesty, beauty, character, morality, etc, which cannot be measured quantitatively but can be arranged serially.In such situations Charles Edward Spearman, a British psychologist,

developed a formula in 1904 which consists in obtaining the correlation coefficient between the ranks of individuals in the two attributes under study.

In computing the correlation between two sets of ranks, special methods have been devised. When we have only a few scores (n is too small) having two sets, at that time it is advisable to ranks these scores and to compute the coefficient of correlation by Spearman's Rank difference Method (p)

Symbolically

$$\rho = 1 - \frac{6 \times \Sigma D^2}{n(n^2 - 1)}$$

Where

ρ = Rho, Spearman's rank correlation coefficient.

D = R_1 – R_2 (Difference between rank-1 and Rank-2).

ΣD^2 = Sum of the squares of difference in rank

N = Number of pairs.

Assumptions of ρ

- The data are badly skewed or on' is too small.
- When quantitative measurement is not possible.
- Data are free or independent of some characteristics of the population distribution.
- Data are in ordinal scale.

Example:

A group of 7 students has been administered mathematics and science tests, their obtained scores are given in the following table. Find out the coefficient of correlation between two tests of scores by rank difference method.

Students	A	B	C	D	E	F	G
Mathematics(test-l)	32	28	35	26	22	20	30
Science(test-2)	27	25	26	22	15	18	24

Solution:

Step 01: Rank the I st test of scores, starting from Rank 1 to the highest score, and write the ranks under R_1 (Col-4)

Step 02: Rank the 2nd test of scores starting Rank lto the highest score and write the ranks under R_2 column (Col-5)

Table No: 05

1	2	3	4	5	6	7
Students	**Scores of Mathematics**	**Scores of Science**	R_1	R_2	$D(R_1-R_2)$	D^2
A	32	27	2	1	1	1
B	28	25	4	3.5	.5	.25
C	.35	26	1	2	1	1
D	26	22	5	5	0	0
E	22	15	6	7	1	1
F	20	18	7	6	1	1
G	30	24	3	3.5	.5	.25
N=7						$\sum D2 = 4.50$

Step 03: Find out D by deducting R_2 from RI (R_1- R_2) (Col-6).

Step 04: Find out D^2 by squaring the D (Col-7) then compute $\sum D^2$ adding the values in Col-7.

Step 5: Put the values in the formula and get the result.

$$\rho = 1 - \frac{6 \times \Sigma D^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6 \times 4.50}{7(7^2 - 1)}$$

$$= 1 - \frac{6 \times 4.50}{7 \times 48}$$

$$= 1 - \frac{4.50}{56}$$

$$= \frac{56 - 4.50}{56}$$

$$= \frac{51.50}{56} = 0.92$$

So the coefficient of correlation between the scores of mathematics and science is 0.92.

Merits of Rank Difference Method:

- It provides a quick and convenient way of estimating correlation when sample is very small.
- It is easy to calculate.
- It can be applied to any type of data. Qualitative or Quantitative. Hence correlation with qualitative data such as honesty, beauty can be found.
- This is most suitable in case there are two attributes.
- When the data are in ordinal scale at that time we use this method.

Demerits of Rank Difference Method:

- It makes no allowance for gaps between adjacent scores.
- It is only an approximately calculate measure as actual values are not used for calculations.
- Combined r of different series cannot be obtained as in case of mean and S.D.
- Cannot be used for finding out correlation in a grouped frequency distribution.
- It cannot be treated further algebraically.
- It is difficult to compute ρ from data when N is large say more than 30.

8.5.2.2 Karl Pearson's coefficient of correlation (Covariance method): The

Pearson correlation coefficient is a very helpful statistical formula that measures the strength between variables and relationships. This is popularly known as Product Moment Coefficient of correlation. It called Product Moment because "The sum of the deviations from the mean (raised to some power) and divided by N is called a moment. When the corresponding deviations in 'x' and 'y' are multiplied together, summed and divided by N the term product moment is used".Symbolically the product moment coefficient of correlation is as 'r'.

The coefficient of correlation in product moment is:

$$r = \frac{\Sigma xy}{N\sigma_x s\sigma_y}$$

Where:

r = Product Moment.

x = Deviation of scores from mean of X variable from its mean.

Y = Deviation of scores from mean of Y variable from its mean.

Σ = Sum total of products.

O_x = S.D of score of X variable.

O_y = S.D of score of y variable.

N = Number of pairs taken.

Assumptions of the Pearson correlation coefficient:

The assumptions of the Pearson product moment correlation can be easily overlooked. The assumptions are as follows:

- Normal distribution.
- Linearity in correlation.
- Continuous series.
- Level of measurement.
- Related pairs.
- Absence of outliers.
- Homoscedasticity

Peat & Barton 2005 (p.156) list the following assumptions:

- Both variables must be normally distributed.

- The sample must have been selected randomly from the general population.
- The observations (i.e. each pair) are independent of one another.
- The relation between the two variables is linear.
- The variance is constant over the length of the data.

Computation of Product Moment Correlation: Pearson's the product moment correlation can be computed in two different ways .

- ❖ When the data are ungrouped .
- ❖ When the data are grouped.

When the data are ungrouped: The ungrouped data is generally two ways computation of coefficient of correlation. They are

1. When deviations are taken from means.
2. Calculation from Raw scores or original score.

1. Estimating product moment correlation is when deviations are taken from means: The formula used to compute r from ungroup data when deviation are taken from the means of the two distributions X and Y reads like this.

$$rxy = \frac{\Sigma xy}{N\sigma_x.\sigma_y}$$

Where:

r_{xy} = Product Moment Coefficient of correlation between X variable and Y variables.

x = Deviation of scores from mean of X variable.

Y = Deviation of scores from mean of Y variable.

Σ = Sum total of products.

O_x = S.D of score of X variable.

O_y = S.D of score of y variable.

N = Number of pairs taken.

We can also derive a more simplified formula by substituting $\sqrt{\frac{\Sigma x^2}{N}}$ for ax

and $\sqrt{\frac{\Sigma y^2}{N}}$ for σ_y.

After simplification the formula read like this:

$$rxy = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

Example 01:

Compute the coefficient of correlation of the scores of 15 students in a test of Science and Mathematics in product moment method.

Student	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Marks in Science	48	62	75	56	68	42	35	44	54	67	61	54	48	54	35
Marks in Mathematic	55	50	85	48	62	35	45	50	45	61	59	48	50	60	35

Solution:

Step 01: Find out mean of scores in Science(X) and mean of scores in Mathematics (Y). Here M_x = 53.5 and M_y = 52.5

Step 02: Find out the deviation(x) of each score on Science test (Table: 01, Col-4) and deviation(y) of each score on Mathematics test (Table: 06, Col-5)

Table No: 06

1	2	3	4	5	6	7	8
Student	Marks in Science(X)	Marks in Mathematics(Y)	X	y	X^2	Y^2	xy
A	48	55	-5.5	2.5	30.25	6.25	-13.75
B	62	50	8.5	-2.5	72.25	6.25	-21.25
C	75	85	21.5	32.5	462.25	1056.25	698.75
D	56	48	2.5	-4.5	6.25	20.25	-11.25
E	68	62	14.5	9.5	210.25	90.25	137.75
F	42	35	-11.5	-17.5	132.25	306.25	201.25
G	35	45	-18.5	-7.5	342.25	56.25	138.75
H	44	50	-9.5	-2.5	90.25	6.25	23.75

1	2	3	4	5	6	7	8
Student	Marks in Science(X)	Marks in Mathematics(Y)	X	y	X^2	Y^2	xy
I	54	45	0.5	-7.5	0.25	56.25	-3.75
J	67	61	13.5	8.5	182.25	72.25	114.75
K	61	59	7.5	6.5	56.25	42.25	48.75
L	54	48	0.5	-4.5	0.25	20.25	-2.25
M	48	50	-5.5	-2.35	30.25	6.25	13.75
N	54	60	0.5	7.5	0.25	56.25	3.75
O	35	35	-18.5	-17.5	342.25	306.25	323.75
					$\sum x^2$= 1957.75	$\sum y^2$= 2107.75	$\sum XY$= 1652.75

Step 03: Square all of the x and all of the y and find out x^2 and y^2. Add the x^2 in col-6 and y^2 in col-7 and find out $\sum x^2$ and $\sum y^2$.

Step 04: Multiply the deviations of X variable (col-4) with deviation of Y variable (col-5) with due regard to algebraic sing to get xy (col-8). Then add the values in eel-S and get $\sum xy$.

Step 05: Put the values in the result.

$$r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$= \frac{1652.75}{\sqrt{1957.75 \times 2107.75}}$$

$$= \frac{1652.75}{2031.37} = 0.81$$

So the coefficient of correlation between the scores in Science and scores in Mathematics of the 15 students is 0.81.

2. Calculation of product moment coefficient of correlation from raw scores or original score.

We can also compute the r from raw scores or directly from original scores. In this case we apply following formula.

$$r_{xy} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{\left\{N\Sigma X^2 - (\Sigma x)^2\right\}\left\{N\Sigma Y^2 - (\Sigma Y)^2\right\}}}$$

Where

r_{xy} = Product moment coefficient of correlation between X and Y variables.

ΣXY = Sum of the product of X and Y.

ΣX = Sum of the scores of X variable.

ΣY = Sum of the scores of Y variable.

ΣX^2 = Sum of the square of X.

ΣY^2 = Sum ofth-e square ofY.

Example 02: Compute the coefficient of correlation of the following two sets of scores obtained from a test of Bengali and English often students in product moment method.

Students	A	B	C	D	E	F	G	H	I	J
Scores in Bengali	18	12	15	8	10	7	6	14	9	13
Scores in English	18	11	20	6	10	8	7	16	7	15

Solution:

Step 01: Square all the X and Y, and find out X^2 and Y^2.

Step 02: Find the product of X and Y by multiplying each X with corresponding Y.

Step 03: Add the x (Col-2), Y (Col-3), X^2 (Col-4), Y^2 (Col-5) and XY (Col-6) to get ΣX, ΣY, ΣX^2, ΣY^2, and ΣXY respectively.

Table No: 07

1	2	3	4	5	6
Students	Scores in Bengali(X)	Scores in English(Y)	X^2	Y^2	XY
A	18	18	324	324	324
B	12	11	144	121	132
C	15	20	225	400	300
D	8	6	64	36	48
E	10	10	100	100	100
F	7	8	49	64	56
G	6	7	36	49	42
H	14	16	196	256	224
I	9	7	81	49	63
J	13	15	169	225	195
	$\sum X=112$	$\sum Y = 118$	$\sum X^2 = 1388$	$\sum Y^2 = 1624$	$\sum XY= 1484$

$$r_{xy} = \frac{N\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{\left\{N\Sigma X^{2}-(\Sigma x)^{2}\right\}\left\{N\Sigma Y^{2-(\Sigma Y)^{2}}\right\}}}$$

$$= \frac{10\times1484-112\times118}{\sqrt{\{10\times1388-(112)^2\}\{10\times1624-(118)^2\}}}$$

$$= \frac{14840-13216}{\sqrt{\{13880-12544\}\{16240-13924\}}}$$

$$= \frac{1624}{\sqrt{1336\times2316}} = \frac{1624}{\sqrt{3094176}}$$

$$= \frac{1624}{1759.03} = 0.92$$

So the coefficient of correlation between the two sets of scores is 0.92.

Merits of Pearson's Method of Studying Correlation: The following are the chief points of merit that go in favor of the Karl Pearson's method of correlation:

- The reliability of test is calculated in terms of Pearson(r) .
- This method also, determines the exact extent, or degree to which they are correlated.
- Under this method, we can also ascertain the direction of the correlation i.e. whether the correlation between the two variables is positive, or negative.
- This method enables us in estimating the value of a dependent variable with reference to a particular value of an independent variable through regression equations.
- This method has a lot of algebraic properties for which the calculation of co-efficient of correlation, and a host of other related factors viz. co-efficient of determination, are made easy.
- The cut of score is determined empirically with the help of scatter plot.
- In guidance service, especially for prediction purpose Pearson's correlation used.

Demerits of Pearson's Method of Studying Correlation: Despite the above points of merits, this method also suffers from the following demerits:

- It is comparatively difficult to calculate as its computation involves intricate algebraic methods of calculations.
- It is very much affected by the values of the extreme items.
- It is based on a large number of assumptions viz. linear relationship, cause and effect relationship etc. which may not always hold good.
- It is very much likely to be misinterpreted particularly in case of homogeneous data.
- In comparison to the other methods, it takes much time to arrive at the results.
- It is subject to probable error which itspropounded himself admits, and therefore, it is always advisable to compute it probable error while interpreting its results.

8.5.2.3 Computation of r when the data are grouped.

Bivariate Frequency Distribution Using Scatter Diagram: When number is large, computing r in the above method is laborious and time consuming. We can

overcome this difficulty by arranging the data in the form of a diagram or chart known as 'scatter diagram'. It is also known as two way frequency distribution or bivariate frequency distribution.

How to Prepare a Scatter Diagram:

For example 50 students of grade of a High School achieved the following scores upon a group intelligence test (X) and algebra test (Y). Let us construct a scatter diagram for the scores given below.

X	Y	X	Y	X	Y	X	Y	X	Y
48	173	40	95	47	153	32	111	35	160
38	134	57	146	39	150	50	164	48	149
26	179	23	175	32	134	29	119	40	149
37	159	51	126	37	184	41	160	43	143
34	167	35	120	26	154	32	149	38	159
51	136	41	154	40	90	49	149	37	157
46	152	28	146	31	96	58	143	40	120
35	160	35	159	50	159	26	164	36	140
41	135	39	138	29	175	33	127	43	141
49	105	32	154	41	164	44	144	48	143

Let us take the class intervals of Intelligence test along the left hand margin, from top to bottom of the diagram(Table No: 08) and class intervals of Algebra test along the top of the diagram from left to right.

SCATTER DIAGRAM OR CORRELATION TABLE ALGEBRA TEST SCORES (Y VARIABLE)

	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180-189	f_x
20-24									/1		1
25-29			/1			/1	/1	/1	//2		6
30-34	/1		/1	/1	/1	/1	/1				7
35-39				/1	//2	/1	//V/5	//2			12

	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180-189	f_x
40-44	//2			/1	/1	///3	/1	//2		/1	10
45-49		/1				////4	//2		/1		8
50-54				/1	/1		/1	/1			4
55-59						//2					2
f_y	3	1	2	4	5	12	11	7	4	1	50

Method for Preparing Scatter Diagram: Suppose we want to plot the scores of the 1st student in the diagram. The 1st student has a intelligence Scores of 48 and algebra score of 173. Here we have to put a tally vin the cell correspondingt to the class intervals, 45-49 in intelligence and 170-179 in algebra test. Like wise we have to put tallies for all 50 students in accordance with the two scores, intelligence test and algebra test. Then the tallies of each cell will be counted and translated into the number. There after the number of each row will be added and frequency for each class interval of intelligence test (X variable) f_x will be found out. .For example in the (Table No: 08). The f_x for the first row is 1, 2nd row 6, 3rd row 7 and the hke wise 8th row 2. In the same manner the cell numbers of each column will be added and frequency for each class interval of algebra test (Y variable) f_y will be determained. For example the f_y for lst column is 3, 2nd column 1, 3rd column 2 and like wise l0th column is 1. After al of the tallies have been listed, the frequency in each cell is added and entered on the diagram, the diagram is then a correlation table.

The scatter diagram is prepared for a large sample, because the techniques discussed earlier for un grouped data are used for small sample. The techniques are not applicable for calculating coefficient or correlation of large samples. The modified formula of assumed mean deviation is used for computing the value of 'r'.

$$r_{xy} = \frac{\frac{\Sigma x' y'}{N} - C_x . C_y}{\sigma_x \sigma_y}$$

Where

r_{xy} = Product moment correlation between variable X and variable Y.

$\sum x^l y^l$ = Sum of the products of the devaitions of the X and y variable from the assumed mean.

x^l = Deviation of the scores (Mid-points) of X from the assumed mean.

y^l = Deviation of the scores (Mid-points) ofY from the assumed mean.

C_x = Correlation for X variable.

Symbolically: $C_x = \frac{\Sigma f x^1}{N}$

Cy = Correlation for Yvariable.

Symbolically: $C_y = \frac{\Sigma f y^1}{N}$

σ_x = S.D of the X variable.

σ_y = S.D of the y variable.

By substituting the values of C_x, C_y, σ_x and ay we can also derive a simplified formula as following:

$$r_{xy} = \frac{\frac{\Sigma x^1 y^1}{N} - C_x . C_y}{\sigma_x \sigma_y}$$

$$= \frac{N\Sigma fx'y' - \Sigma fx'.\Sigma fy'}{\sqrt{\left\{N\Sigma fx'^2 - (\Sigma fx')^2\right\}\left\{N\Sigma fy'^2 - (\Sigma fy')^2\right\}}}$$

Step 01: Add the frequencies of each column of algebra scores and get f_y. Then add the frequencies of each row of intelligence test and get f_x.

Step 02: Assume a mean for the intelligence test scores (as we have discussed in computing mean in assumed mean method) and draw a double line of that column to make it distinct. Like-wise assume a mean for the algebra test scores and draw a double line of that row to make it distinct. In this present problem for intelligence test the mid-point of c.i. 40-44 i.e 42 and for algebra test the mid-point of c.i. 140-149 i.e. 144.5 are taken as assumed means. Now we can take x and y' from this point as indicated in the Table no: 09

Step 03: Multiply the x with f_x and find out f_x and in the same way multiply

the y' with fyand find out f_y.

Step 04: Multiply the fx' column with x column and get fx'^2 and fy row with y' and get f_y^2.

Step 05: The next task is find out *fx'y'*. Multiply the *x'* of the column with the *y* of the row of a particular cell giving due weightage to the algebraic sings. Write the product to the top corner of the cell within bracken. Then multiply the cell frequency with the product and get the value of *fx'y'* of that cell and write it to the lower left corner of the cell. For example the frequency of cell 20-24 and 170-179 is 1. Here *x'* is -4 andy is +3 the product of *x'* and *y'* is -12. By multiplying the product -12 with cell frequency 1 we get *fxy'* = –12 for that cell. Like-wise we can calculate the *fx'y'* for all the cells. Adding the values of cells row wise we can get the values of *fx'y'* column. Adding these values get $\sum fx'y'$. To check the correctness adds the values of *fx'y'* column wise get *fx'y'* row and adding these values we can also get $\sum fx'y'$ (See table no:09)

Table No 09:

SCATTER DIAGRAM ALGEBRA TEST SCORES (Y VARIABLE)

Step 06: Add the values of *fx'*, fx^2, *fy'* and fy'^2 and get $\sum fx'$, $\sum fx'^2$, $\sum fy'$ and $\sum fy'^2$ respectively.

	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180 189	f_x	X_1	fx^1	fx^{12}
-24									12^1 (12)		1	–4	–4	16
-29			9^1 (9)			0^1 (0)	-3^1 (–3)	-6^1 (–6)	-18^2 (–9)		6	–3	–18	54
–34	10^1 (10)		6^1 (6)	4^1 (4)	2^1 (2)	0^1 (0)	-2^1 (–2)	-4^1 (–4)			7	–2	–14	28
–39				2^1(2)	2^2(1)	0^1(0)	-5^5 (1)	-4^2 (–2)		-4^1 (–4)	12	–1	–12	12
–44	0^2(0)			0^1(0)	0^1(0)	0^3(0)	0^1(0)	0^2(0)			10	0	0	0
–49		-4^1				0^4(0)	2^2(1)		3^1(3)		8	+1	8	8

	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180 189	f_x	X_1	fx^1	fx^{12}
–54				-4^1 (–4)	-2^1 (–2)		2^1 (2)	4^1 (4)			4	+2	8	16
–59						$0^2(0)$					2	+3	6	18
	3	1	2	4	5	12	11	7	4	1	50		$\sum fx^1$ =–26	$\sum fx^{12}$ =
	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	$\sum fx^1$ =3			
	-15	-4	-6	-8	-5	0	11	14	12	4	$\sum fx^1$ =221			
	10	4	15	2	2	00	-6	-10	-27	-4	$\sum fx^1 y^1$= -22			

Step 07: Put the values in the formula and get the result.

$$r = \frac{N\Sigma fx'y' - \Sigma fx'.\Sigma fy'}{\sqrt{\left\{N\Sigma fx'^2 - (\Sigma fx')^2\right\}\left\{N\Sigma fy'^2 - (\Sigma fy')^2\right\}}}$$

$$= \frac{(50\times -22) - (-26\times 3)}{\sqrt{\{(50\times 152) - (-26)^2\}\{(50\times 221) - (3)^2\}}}$$

$$= \frac{-1100 - (-78)}{\sqrt{(7600 - 676)(11050 - 9)}}$$

$$= \frac{-1022}{\sqrt{6924\times 11041}}$$

$$= \frac{-1022}{8743.45}$$

$$= -116 \text{ or, } -12$$

Therefore the r for both the variables is -12

Advantages of Scatter Diagram:

1. Simple & Non Mathematical method.
2. Not influenced by the size of extreme item.
3. First step in investing the relationship between two variables

Disadvantage of scatter diagram:

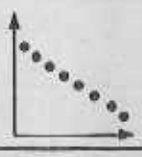
1. Cannot adopt the an exact degree of correlation

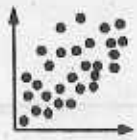
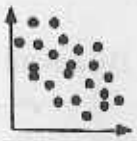
Uses of Coefficient of Correlation (r):

1. To find out the degree of relationship or inter dependence between two variables r is used.
2. To take decisions in educational and vocational guidance r is used.
3. To determine the reliability of a test result r is used.
4. To determine the validity of test scores r is used.
5. Can be used to make predictions about the variables under study.
6. Can be used in many places, including natural settings, libraries, etc.
7. Gain quantitative data which can be easily analyzed.

8.6 Interpretation of Pearson's Correlation Coefficient

Pearson's correlation coefficient gives a measure of the relationship between two variables on a scale from -1 to 1. Word descriptors based on r-values seem doubtful at the best of times and the majority of texts on this subject do not include them. Many texts and Internet sites vary on the advice they give. Here is one possible interpretation.

r	Description	*r*	Description
1	perfect positive correlation 	−1	perfect negative correlation 
0.75 to 1	strong positive correlation 	−1 to −0.75	strong negative correlation 

0.50 to 0.75	moderate positive correlation 	−0.75 to −0.50	moderate negative correlation 
0.25 to 0.50	weak positive correlation 	−0.50 to −0.25	weak negative correlation 

Branches of Statistics: The field of statistics is divided into two major divisions: descriptive statistics and inferential statistics.

Descriptive statistics: Descriptive statistics are brief descriptive coefficients that summarize a given data set, which can be either a representation of the entire or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability (spread). Measures of central tendency include the mean, median, and mode, while measures of variability include the standard deviation, variance, the minimum and maximum variables, and the kurtosis and skewness.

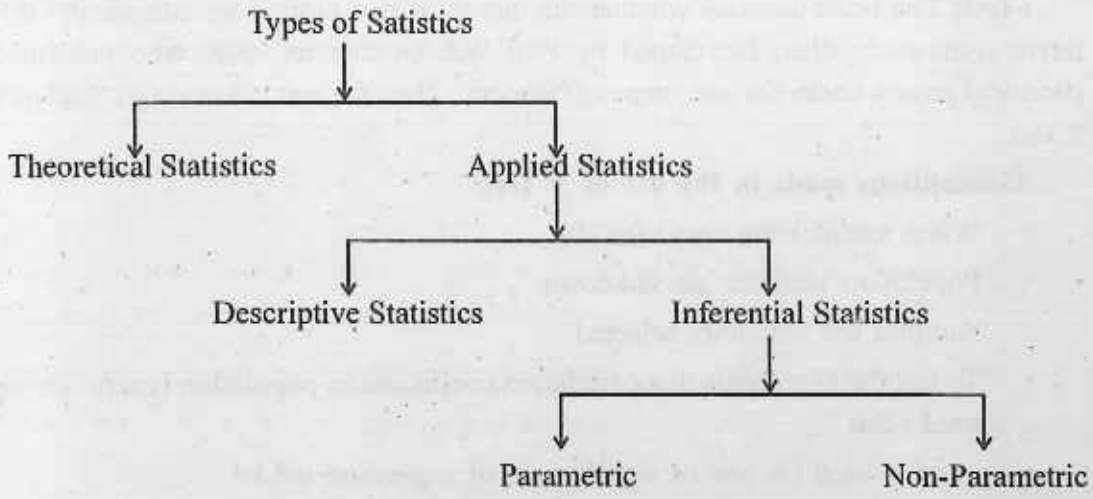
Inferential statistics: Inferential statistics is a random sample of data taken from a population to describe and make inferences about the population. Inferential statistics are valuable when examination of each member of an entire population is not convenient or possible. For example, to measure the diameter of each nail that is manufactured in a mill is impractical. You can measure the diameters of a representative random sample of nails. You can use the information from the sample to make generalizations about the diameters of all of the nails.

Difference between descriptive statistics and inferential statistic

SI. No	Descriptive Statistics	Inferential Statistics
1	Concerned with the describing the target population	Make inferences from the sample and generalize them to the population
2	Organize, analyze and present the data in a meaningful manner	Compares, test and predicts future outcomes
3	Final results are shown in form of charts, tables and grasp	Final result is the probability scores

Sl. No	Descriptive Statistics	Inferential Statistics
4	Describes the data which is already known	Tries to make conclusions about the population that is beyond the data available.
5	Tools-Measures of central tendency (mean/median/mode), Spread of data (range, standard deviation etc.)	Tools-hypothesis test, Analysis of variance etc.

Different Branches of Statistics are given following diagram



Inferential Statistics: There are two main types of inferential statistics: parametric statistics and non-parametric statistics

Parametric test: A parametric statistical test is one that makes assumptions about the parameters (defining properties) of the population distribution(s) from which one's data are drawn.

Assumptions for parametric test:

- The observations are independent. The selection of one case is not dependent on the selection of any other case.
- The variables involved must have been measured in interval or ratio scale.
- The samples have equal, or nearly equal. Variances. This condition is particularly important to determine when samples are small.

- The samples have been drawn should be normally distributed.
- Used for quantitative data
- Used for continuous varialves

Some Parametric tests:

1. t-test (n<30)
2. F-test
3. Z-test (n>30)
4. Pearson's coefficient of correlation
5. ANOVA (analysis of variance)

t-test: The t-test assesses whether the means of two groups are statistically different form each other. Developed by Prof W.S Gossett in 1908, who published statistical papers under the pen name of 'Student'. Thus the test is known as Student's 't' test.

Assumptions made in the use of 't' test:

- When samples are small (n<30)
- Population variance are unknown
- Samples are randomly selected
- To test the hypothesis that correlation coefficient in population is zero the we used t-test
- t-test is used for test of significance of regression model
- Data utilized is quantitative
- Variable follow normal distribution

$$t = \frac{M1-M2}{SED}$$

F-test: The name was coined by George W. Snedecor, in honour of Sir Ronald A. Fisher. Fisher initially developed the statistic as the variance ratio in the 1920. Any statistical test that uses F-distribution can be called an F-test. It is used when the sample size is small i.e. n < 30. For example, suppose one is interested to test if there is any significant difference between the mean height of male and female students in a particular college. In such a situation, a t-test for difference of means can be used.

Assumption for the F-test:

- Tow sample have same variance
- F-test is small sample test
- F-test is design to test if two population variance are equal.
- The variance ratio $S1^2/S2^2$
- Degrees of freedom for larger population variance is VI and smaller population variance is V2
- Comparing to variance
- The sample must be independent
- Value does't can be negative (O-infinitive)

Non-parametric test: Non-parametric or distribution free test is a statistical procedure whereby the data does not match a normal distribution. The data used in non-parametric test is frequently of ordinal data type, thus implying it does not depend on arithmetic properties. Consequently, all tests involving the ranking of data are non-parametric and also no statement about the distribution of data is made.

Characteristics and Features of Non Parametric Test

1. They do not suppose any particular distribution and the consequential assumptions.
2. They are rather quick and easy to use i.e., they do not require laborious computations since in many cases the observations are replaced by their rank order and in many others we simply use signs.
3. They are often not as efficient or 'sharp' as tests of significance or the parametric tests. An interval estimate with 95% confidence may be twice as large with the use of nonparametric tests as with regular standard methods. The reason being that these tests do not use all the available information but rather use groupings or rankings and the price we pay is a loss in efficiency. In fact, when we use non-parametric tests, we make a tradeoff: we loose sharpness in estimating intervals, but we gain the ability to use less information and to calculate faster.
4. When our measurements are not as accurate as is necessary for standard tests of significance, then non-parametric methods come to our rescue which can be used fairly satisfactorily.

5. Parametric tests cannot apply to ordinal or nominal scale data but non-parametric tests do not suffer from any such limitation.
6. The parametric tests of difference like 't or 'F' make assumption about the homogeneity of the variances whereas this is not necessary for non-parametric tests of difference.

Advantage of non-parametric test:

1. Easier to learn and aapply.
2. Based on a model that specific very general condition.
3. No specific form of the distribution form which the sample was drawn is needed.
4. Hence non-parametric tests are also known as distribution free test.

Disadvantage of non-parametric test:

1. Losing precision.
2. Low power.
3. False sense of security.
4. Where a parametric test would be appropriate, non-parametric test has less power.
5. A large sample size can be required to draw conclusions with the some degree of confidence.

Difference between parametric test and non-parametric test

Sl. No	Parametric	Non-Parametric
1	Assumed normal distributions	No assumed shape/distribution
2	Handles interval data or ratio data	Handles Ordinal data, Nominal, ranked data
3	Results can be significantly affected by outliers	Results cannot be seriously affected by outliers
4	Perform well when the spread of each group is different, might not provide valid results if groups have a same spread	Perform well when the spread of each group is same, might not provide valid results if groups have a different spread
5	Have more statistical power	It is not so powerful like parametric test

8.7 Let us sum up:

In this unit, we presented the direction and magnitude of a relationship between the two variables is measured with the help of co-efficient of correlation. The product moment correlation co-efficient is computed when the variables (X and Y) are expressed in an interval or ratio scales of measurement and the distritbution of the variables have a linear relationship. In case of ordinal(ranked) data we compute rank difference correlation. The procedure for computing these co-efficients has been discussed in this unit.

8.8 Unit End Exercises (10 questions: 5SA type, 5 Objective type)

I. Answer the following questions (Short answer type):

a) What is correlation?

b) Define correlation.

c) What do you mean by correlation of coefficient?

d) What are the main characteristics of correlation of coefficient?

e) List the situations in which product moment correlation is used.

f) Discuss the types of correlation used in educational research.

g) What is Rank correlation?

h) Compute product moment correlation for the following data:

X: 50, 54, 56, 59,60,62,61,65,67, 71, 71, 74

Y:22,25, 34,28,26, 30, 32,30,28, 34,34,36,

i) Compute rank difference correlation for the data give under (h). Compare the result with the one obtained using product moment correlation.

2. Answer the following questions (Objective type):

a) Coefficient of correlation ranging from.....

b) Correlation analysis deals with the association between.....variables

c) What indicate by a positive correlaion?

d) Who invented product moment method?

e) Who had developed the concept of Rank Correlation?

f) When r = 1 then it is called.....

8.9 Bibliography

1. S. K. Mangal- Statistics in Education and Psychology
2. A. K. Singh - Test, Measurement and Research Methods in Behavioral Sciences
3. H.E. Garret- Statistics in Education and Psychology
4. R. A. Sharma- Mental Measurement and Evaluation
5. Y. P. Aggarwal- Statistics Methods Concepts. Application and Computation

মানুষের জ্ঞান ও ভাবকে বইয়ের মধ্যে সঞ্চিত করিবার যে একটা প্রচুর সুবিধা আছে, সে কথা কেহই অস্বীকার করিতে পারে না। কিন্তু সেই সুবিধার দ্বারা মনের স্বাভাবিক শক্তিকে একেবারে আচ্ছন্ন করিয়া ফেলিলে বুদ্ধিকে বাবু করিয়া তোলা হয়।

—রবীন্দ্রনাথ ঠাকুর

ভারতের একটা mission আছে, একটা গৌরবময় ভবিষ্যৎ আছে, সেই ভবিষ্যৎ ভারতের উত্তরাধিকারী আমরাই। নূতন ভারতের মুক্তির ইতিহাস আমরাই রচনা করছি এবং করব। এই বিশ্বাস আছে বলেই আমরা সব দুঃখ কষ্ট সহ্য করতে পারি, অন্ধকারময় বর্তমানকে অগ্রাহ্য করতে পারি, বাস্তবের নিষ্ঠুর সত্যগুলি আদর্শের কঠিন আঘাতে ধূলিসাৎ করতে পারি।

—সুভাষচন্দ্র বসু

Any system of education which ignores Indian conditions, requirements, history and sociology is too unscientific to commend itself to any rational support.

—Subhas Chandra Bose

Price : Rs. 225.00
(NSOU-র ছাত্র-ছাত্রীদের কাছে বিক্রয়ের জন্য নয়)

Published by Netaji Subhas Open University, DD-26, Sector-I, Salt Lake, Kolkata - 700 064 & Printed at Prabaha, 45, Raja Rammohan Roy Sarani, Kolkata - 700 009.