Bachelor of Education
SEMESTER II
Course Title: Assessment of Learning
Credits: 4
Course Code: BED106 **MM: 100**

Unit I Concept of Evaluation

· Concept of Measurement, Assessment and Evaluation

· Need and Scope of Evaluation

· Distinction between the following: Measurement, Examination, Assessment and Evaluation

· Evaluation Approaches: Formative -Summative

· Continuous Comprehensive Evaluation: Need, Relevance, Implementation Procedure, Problems

Unit - II Tools and Techniques of Evaluation

· Characteristics of good measuring instruments and factors affecting them.

· Reliability and Validity of Tools

· Tools of evaluation:- Quantitative ñ Written, Oral and Practical(Types of Questions: Short, Long, MCQs covering all three domains of Learning-Cognitive, Affective and Psychomotor)

· Qualitative Observation, Introspection, Projection and Sociometry

· Use of these tools for internal assessment & maintaining cumulative records of learners in School

· Planning and Preparation of test (including blue print)

Unit- III Statistical Methods and Interpretation of scores

· Need & Importance of Statistics in Evaluation

· Graphical Representation Histogram, Frequency Polygon, Pi Charts,

· Measures of Central Tendency:- Mean, Median, Mode. (Meaning, Characteristics, use only)

· Measures of Variability :(Meaning, Characteristics, Use only) Range, Quartile deviation, Standard deviation

· Normal Probability Curve:-Properties and Uses. (Skewness and Kurtosis (Meaning & Reasons)

· Coefficient of Correlation-Spearman's Rank Rule Method

· Percentile & Percentile rank (Meaning & Uses)

Unit IV New Trends in Evaluation (Need and Use)

· Question bank

· Grading system

· Online Examination

· Open Book Examination

· Credit System

· Exam on Demand (meaning & uses only)

Concept of Measurement, Assessment and Evaluation

Test: A method to determine a student's ability to complete certain tasks or demonstrate mastery of a skill or knowledge of content. Some types would be multiple choice tests, or a weekly spelling test. While it is commonly used interchangeably with assessment, or even evaluation, it can be distinguished by the fact that a test is one form of an assessment.

Assessment: The process of gathering information to monitor progress and make educational decisions if necessary. As noted in my definition of test, an assessment may include a test, but also includes methods such as observations, interviews, behavior monitoring, etc.

Evaluation: Procedures used to determine whether the subject (i.e. student) meets a preset criteria, such as qualifying for special education services. This uses assessment (remember that an assessment may be a test) to make a determination of qualification in accordance with a predetermined criteria.

Measurement, beyond its general definition, refers to the set of procedures and the principles for how to use the procedures in educational tests and assessments. Some of the basic principles of measurement in educational evaluations would be raw scores, percentile ranks, derived scores, standard scores, etc.

Educational assessment involves gathering and evaluating data evolving from planned learning activities or programs. This form of assessment is often referred to as evaluation (see section below on Assessment versus Evaluation). Learner assessment represents a particular type of educational assessment normally conducted by teachers and designed to serve several related purpose (Brissenden and Slater, n.d.). These purposed include:

• motivating and directing learning

• providing feedback to student on their performance

• providing feedback on instruction and/or the curriculum

• ensuring standards of progression are met Learner assessment is best conceived as a form of two-way communication in which feedback on the educational process or product is provided to its key stakeholders (McAlpine, 2002).

Specifically, learner assessment involves communication to teachers (feedback on teaching); students (feedback on learning); curriculum designers (feedback on curriculum) and administrators (feedback on use of resources).

Types and Approaches to Assessment Numerous terms are used to describe different types and approaches to learner assessment. Although somewhat arbitrary, it is useful to these various terms as representing dichotomous poles (McAlpine, 2002).

Formative <--------------------------------> Summative

Informal <--------------------------------> Formal

Continuous <----------------------------------> Final

Process <--------------------------------> Product

Divergent <--------------------------------> Convergent

Dimension	**Assessment**	**Evaluation**
Timing	Formative	Summative
Focus of Measurement	Process-Oriented	Product-Oriented
Relationship Between Administrator and Recipient	Reflective	Prescriptive
Findings and Uses	Diagnostic	Judgmental
Modifiability of Criteria, Measures	Flexible	Fixed
Standards of Measurement	Absolute (Individual)	Comparative
Relation Between Objects of A/E	Cooperative	Competitive

Qualities of Good Measuring Instruments

I. **Validity:** The extent to which the instrument really measures what it is intended to measure. The validity of the test concerns what the test measures and how well it does so. A valid measurement tool does a good job of measuring the concept that it purports to measure. It is important to remember that the validity of an instrument only applies to a specific purpose with a specific group of people. For example, a scale is not considered simply "valid" or "invalid"—but it might be considered valid for measuring social responsibility outcomes with college freshmen. **Factors Affecting Validity**

- Unclear direction.
- Reading vocabulary.
- Difficult sentence construction.

- Poorly constructed test items.
- Use of in appropriate items.
- Medium of expression.
- Difficulty level of items.
- Influence of extraneous factors.
- In appropriate time limit.
- Inadequate coverage

I. Content Validity: Content Validity means the extent to which the content of the test is truly representative of the content of the course. Content Validity Establishes that the instrument includes items that comprise the relevant content domain. (For example, a test of English grammar might include questions on subject-verb agreement, but should not include items that test algebra skills.)

II. Concurrent validity: Concurrent validity (simultaneous, parallel) is the degree of which the test agrees with or correlates with a criterion which is a set up as an acceptable measure. The criterion is always available at the time of testing. (For example, researchers give a group of students a new test,designed to measure mathematical aptitude. They then compare this with the test scores already held by the school, a recognized and reliable judge of mathematical ability. Cross referencing the scores for each student allows the researchers to check if there is a correlation, evaluate the accuracy of their test, and decide whether it measures what it is supposed to)

Weakness: Concurrent validity is regarded as a fairly weak type of validity, and is rarely accepted on its own. The problem is that the benchmark test may have some inaccuracies and, if the new test shows a correlation, it merely shows that the new test contains the same problems

II Reliability: if a test is used by any person and anytime for repeated evaluation, conclusion are same then test is said to be reliable. Factors affecting

- Length of test : it varies from the number of questions and syllabus to be covered by them
- Difficulties level of the questions
- Ambiguity in the question
- Hallo effect
- Sample
- Optional question in the test

III Objectivity: when a evaluation is done with a certain tool by many evaluators and conclusion are same then test is objective. Factors affecting

- Type of question
- Ambiguity

- Evaluation method
- Design of test

IV Adequacy: a test with which evaluation of all objectives and competencies are done. Factors affecting

- Design of test
- Experience of paper setter
- Types of questions
- Nature of content
- Nature and type of objectives

V Usability/ Practicability: when a test is east to construct, easy to administer and easy to interpret. Factors affecting

- Instruction to test administrator
- Checking malpractices
- Scoring key and guidelines
- Length of test
- Cast of test
- Ease of interpretation and application

Reliability is the degree to which an assessment tool produces stable and consistent results.

Types of Reliability

1. **Test-retest reliability** is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for stability over time. Example: A test designed to assess student learning in psychology could be given to a group of students twice, with the second administration perhaps coming a week after the first. The obtained correlation coefficient would indicate the stability of the scores
2. **Parallel forms reliability** is a measure of reliability obtained by administering different versions of an assessment tool (both versions must contain items that probe the same construct, skill, knowledge base, etc.) to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions. Example: If you wanted to evaluate the reliability of a critical thinking assessment, you might create a large set of items that all pertain to critical

thinking and then randomly split the questions up into two sets, which would represent the parallel forms

3. **Inter-rater reliability** is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Inter-rater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge of the construct or skill being assessed. Example: Inter-rater reliability might be employed when different judges are evaluating the degree to which art portfolios meet certain standards. Inter-rater reliability is especially useful when judgments can be considered relatively subjective. Thus, the use of this type of reliability would probably be more likely when evaluating artwork as opposed to math problems.
4. **Internal consistency reliability** is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results.
    - Average inter-item correlation is a subtype of internal consistency reliability. It is obtained by taking all of the items on a test that probe the same construct (e.g., reading comprehension), determining the correlation coefficient for each pair of items, and finally taking the average of all of these correlation coefficients. This final step yields the average inter-item correlation.
    - Split-half reliability is another subtype of internal consistency reliability. The process of obtaining split-half reliability is begun by "splitting in half" all items of a test that are intended to probe the same area of knowledge (e.g., World War II) in order to form two "sets" of items. The entire test is administered to a group of individuals, the total score for each "set" is computed, and finally the split-half reliability is obtained by determining the correlation between the two total "set" scores.

Validity refers to how well a test measures what it is purported to measure.

While reliability is necessary, it alone is not sufficient. For a test to be reliable, it also needs to be valid. For example, if your scale is off by 5 lbs, it reads your weight every day with an excess of 5lbs. The scale is reliable because it consistently reports the same weight every day, but it is not valid because it adds 5lbs to your true weight. It is not a valid measure of your weight.

<u>Types of Validity</u>

1. **Face Validity** ascertains that the measure appears to be assessing the intended construct under study. The stakeholders can easily assess face validity. Although this is not a very "scientific" type of validity, it may be an essential component in enlisting motivation of

stakeholders. If the stakeholders do not believe the measure is an accurate assessment of the ability, they may become disengaged with the task.

Example: If a measure of art appreciation is created all of the items should be related to the different components and types of art. If the questions are regarding historical time periods, with no reference to any artistic movement, stakeholders may not be motivated to give their best effort or invest in this measure because they do not believe it is a true assessment of art appreciation.

2. Construct Validity is used to ensure that the measure is actually measure what it is intended to measure (i.e. the construct), and not other variables. Using a panel of "experts" familiar with the construct is a way in which this type of validity can be assessed. The experts can examine the items and decide what that specific item is intended to measure. Students can be involved in this process to obtain their feedback. Example: A women's studies program may design a cumulative assessment of learning throughout the major. The questions are written with complicated wording and phrasing. This can cause the test inadvertently becoming a test of reading comprehension, rather than a test of women's studies. It is important that the measure is actually assessing the intended construct, rather than an extraneous factor.

3. Criterion-Related Validity is used to predict future or current performance - it correlates test results with another criterion of interest. Example: If a physics program designed a measure to assess cumulative student learning throughout the major. The new measure could be correlated with a standardized measure of ability in this discipline, such as an ETS field test or the GRE subject test. The higher the correlation between the established measure and new measure, the more faith stakeholders can have in the new assessment tool.

4. Formative Validity when applied to outcomes assessment it is used to assess how well a measure is able to provide information to help improve the program under study. Example: When designing a rubric for history one could assess student's knowledge across the discipline. If the measure can provide information that students are lacking knowledge in a certain area, for instance the Civil Rights Movement, then that assessment tool is providing meaningful information that can be used to improve the course or program requirements.

5. Sampling Validity (similar to content validity) ensures that the measure covers the broad range of areas within the concept under study. Not everything can be covered, so items need to be sampled from all of the domains. This may need to be completed using a panel of "experts" to ensure that the content area is adequately sampled. Additionally, a panel can help limit "expert" bias (i.e. a test reflecting what an individual personally feels are the most important or relevant areas). Example: When designing an assessment of learning in the theatre department, it would not be sufficient to only cover issues related to acting. Other areas of theatre such as lighting, sound, functions of stage managers should all be included. The assessment should reflect the content area in its entirety.

<u>The Purposes of Evaluation</u> According to Oguniyi (1984), educational evaluation is carried out from time to time for the following purposes:

(i) to determine the relative effectiveness of the programme in terms of students' behavioural output; Measurement and Evaluation in Education
(ii) to make reliable decisions about educational planning
(iii) to ascertain the worth of time, energy and resources invested in a programme
(iv) to identify students' growth or lack of growth in acquiring desirable knowledge, skills, attitudes and societal values
(v) to help teachers determine the effectiveness of their teaching techniques and learning materials
(vi) to help motivate students to want to learn more as they discover their progress or lack of progress in given tasks
(vii) to encourage students to develop a sense of discipline and systematic study habits; (viii)
(viii) to provide educational administrators with adequate information about teachers' effectiveness and school need)
(ix) to acquaint parents or guardians with their children's performances
(x) to identify problems that might hinder or prevent the achievement of set goals
(xi) to predict the general trend in the development of the teaching-learning process
(xii) to ensure an economical and efficient management of scarce resources
(xiii) to provide an objective basis for determining the promotion of students from one class to another as well as the award of certificates
(xiv) to provide a just basis for determining at what level of education the possessor of a certificate should enter a career.

<u>TYPES OF EVALUATION</u> There are two main levels of evaluation viz: programme level and student level. Each of the two levels can involve either of the two main types of evaluation – formative and summative at various stages.

Programme evaluation has to do with the determination of whether a programme has been successfully implemented or not. **Student evaluation** determines how well a student is performing in a programme of study.

<u>Formative Evaluation</u>

The purpose of formative evaluation is to find out whether after a learning experience, students are able to do what they were previously unable to do. Its ultimate goal is usually to help students perform well at the end of a programme.

Formative evaluation enables the teacher to:

1. Draw more reliable inference about his students than an external assessor, although he may not be as objective as the latter;

2. Identify the levels of cognitive process of his students;

3. Choose the most suitable teaching techniques and materials;

4. Determine the feasibility of a programme within the classroom setting;

5. Determine areas needing modifications or improvement in the teaching-learning process; and
6. determine to a great extent the outcome of summative evaluation

SUMMATIVE EVALUATION Summative evaluation often attempts to determine the extent the broad objectives of a programme have been achieved.

It is concerned with purposes, progress and outcomes of the teaching-learning process. Summative evaluation is judgmental in nature and often carries threat with it in that the student may have no knowledge of the evaluator and failure has a far reaching effect on the students. However, it is more objective than formative evaluation.

Some of the underlying assumptions of summative evaluation are that:

1. The programme's objectives are achievable;

2. The teaching-learning process has been conducted efficiently;

3. The teacher-student-material interactions have been conducive to learning

4. The teaching techniques, learning materials and audio-visual aids are adequate and have been judiciously dispensed

5. There is uniformity in classroom conditions for all learners.

PURPOSE OF ASSESSMENT

Assessment involves deciding how well students have learnt a given content or how far the objective we earlier set out has been achieved quantitatively. The data so obtained can serve various educational functions in the school viz:

(a) Classroom function this includes

(i) Determination of level of achievement

(ii) Effectiveness of the teacher, teaching method, learning situation and instructional materials

(iii) Motivating the child by showing him his progress i.e. success breeds success.

(iv) It can be used to predict students performance in novel situations.

(b) Guidance functions Assessment procedure can give the teacher diagnostic data about individual pupils in his class. These will show the pupils' strength, weaknesses and interests. It can also help to decide on which method to use or what remedial activities that are necessary. Parents and pupils can also be rightly guided in terms of career choice.

(c) Administrative functions

(i) Assessing can serve as communication of information when data collected are used in reports to parents

(ii) It could form the basis upon which streaming, grading, selection and placement are based.

(iii) Making appropriate decisions and recommendations on curricula packages and curricula activities.

For any form of assessment to be able to serve the above functions, it cannot be a one shot kind of assessment. It has to be an on-going exercise throughout the teaching and learning processes. This is why continuations assessment is advocated in the classroom.

THE CONCEPT OF CONTINUOUS ASSESSMENT

By continuous assessment, we mean assessing or weighing performance of students periodically to be able to determine progress made in teaching-learning activities. Continuous assessment tests are used to evaluate the progress of students periodically. Continuous assessment tests can be done daily, weekly, monthly, depending on the goals of teaching and learning. Continuous assessment is defined in the Federal Ministry of Education handbook as: "A mechanism whereby the final grading of a student in the cognitive, affective and psychomotor domains of behaviour takes account in a systematic way, of all his performances during a given period of schooling. Such an assessment involves the use of a great variety of models of evaluation for the purpose of finding and improving the learning and performance of the students."

Continuous assessment thus is a veritable tool in assessment in that it is comprehensive, systematic, cumulative and guidance oriented. Many schools in the country have since embarked on the implementation of continuous assessment. It is not surprising therefore to find teachers testing their pupils weekly, at the end of each unit or module etc. In recent times however, these tests have assumed disciplinary status to check noise making, absenteeism etc. At this juncture, Continuous Assessment in practice ceases to be a tool for aiding learning. One can only call it what it is – "Continuous Testing". I urge you to be aware of the practice of continuous testing in our school system which is injurious to learning as against Continuous Assessment that is being advocated.

The following are the advantages of a continuous assessment:

→ It provides useful information about the academic progress of the learner;

→ It makes the learner to keep on working in a progressive manner

→ It informs the teacher about the teaching-learning effectiveness achieved;

→ It gives a true picture of the student academic performance since it is a continuous process rather than one duration type of test which may be affected by many variables such as sickness, fatigue, stress, etc; and

→ It makes learning an active rather than a passive process.

USING CONTINUOUS ASSESSMENT TO IMROVE TEACHING AND LEARNING

(a) **Motivation:** The effectiveness of efforts to help people learn depends on the learner's activities and the achievement that results. Feedback regarding one's effectiveness is positively associated with perceived locus of causality, proficiency and intrinsic motivation (Deci, 1980). When assessment is carried out systematically and in a purposive manner and the feedback of such is given immediately, it can go a long way in correcting any anomaly in the teaching-learning continuum. In the past, students often do hasty and last minute preparation towards final examinations. This neither helps them to have a thorough grasp of the learning experiences nor does it allow the teacher to apply remedial measures to the areas of deficiency or improve on his teaching methods. However, using Continuous Assessment appropriately, students study more frequently and retain what they study for longer period of time. This generally improves their learning which goes a long way in motivating them to study further.

(b) **Individual Differences**: The classroom is an admixture of the slow learners, average, gifted, extroverts, introverts, early bloomers etc. Each of these categories of students should be given a particular attention by the teacher. Using Continuous Assessment, the teacher will be able to identify these differences and apply at the appropriate time, the necessary measure to improve not only his teaching but the learning of the students and hence their performances.

(c) **Record-Keeping:** Continuous Assessment affords the teacher the opportunity to compile and accumulate student's record/performances over a given period of time. Such records are often essential not only in guidance and counseling but also in diagnosing any problem that may arise in future.

(d) **Examination Malpractice:** This is an endemic problem at all levels of our educational system. In practice, continuous assessment had been able to minimize this to a tolerable level and the fear of using one single examination to judge performance of a wide range of course(s) is removed.

CHARACTERISTICS OF CONTINUOUS ASSESSMENT TESTS

i. In most cases, continuous assessment tests are periodical, systematic, and well planned. They should not be tests organized in a haphazard manner.
ii. Continuous Assessment tests can be in any form. They may be oral, written, practical, announced, or unannounced, multiple choice objective, essay, or subjective and so on.
iii. Continuous assessment tests are often based on what has been learnt within a particular period. Thus, they should be a series of tests.
iv. In Nigerian educational system, continuous assessment tests are part of the scores used to compute the overall performance of students. In most cases, they are 40% of the final score. The final examination often carries 60%.
v. Invariably, continuous assessment tests are designed and produced by the classroom teacher. Some continuous assessment tests are centrally organized for a collection of schools or for a particular state.
vi. All continuous assessment tests should meet the criteria stated in Units three and five for a good test: validity, reliability, variety of tests items and procedure, etc.

PROBLEMS OF CONTINUOUS ASSESSMENT

Plausible and important as the above discussion is, Continuous Assessment is not without its own problems in the classroom. However, real as the problems may be, they are not insurmountable. Some of these problems include:

1. Inadequacy of qualified teachers in the respective fields to cope with the large number of students in our classroom. Some time ago, a Minister of Education lamented the population of students in classrooms in some parts of the country.
2. The pressure to cover a large part of the curricula, probably owing to the demand of external examinations, often makes teachers concentrate more on teaching than

Continuous Assessment. There is no doubt that such teachings are not likely to be very effective without any form of formative evaluation.

3. The differences in the quality of tests and scoring procedures used by different teachers may render the results of Continuous Assessment incomparable.

TAXANOMY OF EDUCATIONAL OBJECTIVES

Benjamin Bloom et al classified all educational objectives into three, namely: cognitive, affective and psychomotor domains.

Cognitive domain involves remembering previously learnt matter.

Affective domain relates to interests, appreciation, attitudes and values.

Psychomotor domain deals with motor and manipulative skills.

The focus of Assessment is on these three domains of educational objectives. However, researches have shown that the emphasis has been on the cognitive than the others. This may be because of the difficulty associated with writing of objectives in the other areas. This paper will also focus more on the cognitive domain because most of the Teacher-made Tests will focus on this. However, the methods of assessing all the three domains will be discussed. For emphasis, the main areas of the cognitive domain are reproduced below:

Bloom's Cognitive Domain

1.0.0 Knowledge of specifics

1.1.1 Knowledge of terminology

1.1.2 Knowledge of specific facts.

1.2.0 Knowledge of Ways and Means of Dealing with specifics

1.2.2 Knowledge of trends and sequences.

1.2.3 Knowledge of classification and categories

1.2.4 Knowledge of criteria

1.2.5 Knowledge of methodology

1.3.0 Knowledge of universal and abstractions

1.3.1 Knowledge of Principles and Generalizations.

1.3.2 Knowledge of Theories and Structures.

2.0.0 Comprehension

2.1.0 Translation

2.2.0 Interpretation

2.3.0 Explanation

3.0 Application

4.0.0 Analysis

4.1.0 Analysis of Elements

4.2.0 Analysis of Relationships

4.3.0 Analysis of Organizational principles

5.0.0 Synthesis

5.1.0 Production of a unique communication.

5.2.0 Production of a plan or proposed set of operations.

6.0.0 Evaluation

6.1.0 Judgment in terms of internal evidence

6.2.0 Judgment in terms of External Criteria.

Stages in Assessment Practice

(i) Understand and state the instructional outcomes you wish to assess.
(ii) Formulate the specific behavior you wish to assess
(iii) Formulate and create situations which will permit such behavior to be demonstrated by the pupils.
(iv) Use appropriate device or instrument to assess the behavior.
(v) Take appropriate actions on the outcome of assessment carried out.

Stages in the Assessment of Cognitive Behaviors

A. Preparation:

(i) Break curriculum into contents (tasks) to be dealt with weekly.

(ii) Break contents into content elements
(iii) Specify the performance objectives

B. Practice:

i. Give quality instruction
ii. Engage pupils in activities designed to achieve objectives or give them tasks to perform.
iii. Measure their performance and assess them in relation to set objectives.

C. Use of Outcome:

i. Take note of how effective the teaching has been; feedback to teacher and pupils.
ii. Record the result
iii. Cancel if necessary
iv. Result could lead to guidance and counseling and/or re-teaching.

STAGES IN ASSESSMENT OF PSYCHOMOTOR OUTCOMES These learning outcomes cannot be accessed through achievement tests or class work. The learning outcomes stretch from handling of writing materials to activities in drama, practicals, laboratory activities, technical subjects, games and athletics. Some of the learning outcomes are subject based or non-subject based, e.g. subject based outcomes. - Drawing and painting from art - Fluency in speech from language - Saying prayers from religious studies - Measuring quantities and distance from mathematics.

- Laboratory activities in sciences

- Manipulative skills in subjects involving the use of apparatus and equipment.

- Planting of crops/Experiments However, elements of cognitive behavior are also present in these activities because for one to do something well, one must know how since the activities are based on knowledge.

STAGES IN ASSESSING AFFECTIVE LEARNING OUTCOMES IN SCHOOLS

These learning outcomes include feelings, beliefs, attitudes, interests, social relationships etc. which, at times are referred to as personality traits. Some of these that can be assessed indirectly include:

i. Honesty – truthfulness, trustworthiness, dependability, faithfulness etc. ii.

ii. Respect – tolerance, respect for parents, elders, teachings, constituted authority, peoples' feelings. iii.
iii. Obedience – to people and law iv.
iv. Self-control – temperamental stability, non-aggression, use of decent language etc.
v. Social relationship – kindness, leadership and social skills.

The most appropriate instrument for assessment here is observation. Others like self reporting inventories; questionnaires, interviews; rating scales, projective technique and sociometric technique may as well be used as the occasion demands. In assessing students' personality traits, it is necessary to assume that every student possesses good personality characteristics until the contrary is proved. Note that the purpose of assessing students' personality traits in the school is to give feedback to the students to help them adjust in the right direction rather than the assignment of grades.

Other personality traits which can be accessed directly are:

i. Attendance behavior – regularity and punctuality to school and other activities.
ii. Participatory behavior in non-academic activities.
iii. Appearance: personal cleanliness, in clothes, and for materials handled.
iv. Conduct: based on reported observed behaviors or incidents involving a display of exemplary character or culpable behavior
v. Cooperative behavior in groups

OLD AND MODERN PRACTICES OF ASSESSMENT

A. Old Assessment Practices. Comprised mainly of tests and examinations administered periodically, either fortnightly or monthly. Terminal assessment were administered at the end of term, year or course; hence its being called a 'one-shot' assessment. This system of assessment had the following shortcomings:-

i. It put so much into a single examination.
ii. It was unable to cover all that was taught within the period the examination covered
iii. Schools depended so much on the result to determine the fate of pupils.
iv. It caused a lot of emotional strains on the students.
v. It was limited to students cognitive gains only.
vi. Its administration periodically, tested knowledge only
vii. Learning and teaching were regarded as separate processes in which, only learning could be assessed.

viii. It did not reveal students' weakness early enough to enable teacher to help students overcome them.
ix. Achievements were as a result of comparing marks obtained.
x. It created unhealthy competition which led to all forms of malpractices.

B. Modern Practice

i. This is a method of improving teaching and learning processes.
ii. It forms the basis for guidance and counseling in the school.
iii. Teaching and learning are mutually related.
iv. Teaching is assessed when learning is and vice-versa.
v. Assessment is an integral and indispensable part of the teaching-learning process.
vi. The attainment of objectives of teaching and learning can be perceived and confirmed through continuous assessment.
vii. It evaluates students in areas of learning other than the cognitive.

CONSTRUCTION OF TESTS IN THE CLASSROOM

Teacher-made tests are indispensable in evaluation as they are handy in assessing the degree of mastery of the specific units taught by the teacher. The principles behind the construction of the different categories of Tests mentioned above are essentially the same. These shall now be discussed. Planning for the Test Many teacher-made tests often suffer from inadequate and improper planning. Many teachers often jump into the classroom to announce to the class that they are having a test or construct the test haphazardly.

It is at the planning stage that such questions as the ones listed below are resolved:

(i) What is the intended function of this test? Is it to test the effectiveness of your method, level of competence of the pupils, or diagnose area of weakness before other topics are taught? (ii)
(ii) What are the specific objectives of the content area you are trying to achieve? (iii)
(iii) What content area has been taught? How much emphasis has been given to each topic? (iv)
(iv) What type of test will be most suitable (in terms of effectiveness, cost and practicality) to achieve the intended objectives of the contents?

Defining Objectives: As a competent teacher, you should be able to develop instructional objectives that are behavioral, precise, realistic and at an appropriate level of generality that will serve as a useful guide to teaching and evaluation. This job has been made easier as these are already stated in the various curriculum packages designed by the Federal Ministry of Education, which are available in schools. However, when you write your behavioral objectives, use such action verbs like define, compare, contrast, draw, explain, describe, classify, summarize, apply, solve, express, state, list and give. You should avoid vague and global statements involving the use of verbs such as appreciate, understand, feel, grasp, think etc. It is important that we state objectives in behavioral terms so as to determine the terminal behaviour of a student after having completed a learning task.

Martin Haberman (1964) says the teacher receives the following benefits by using behavioral objectives:

1. Teacher and students get clear purposes.
2. Broad content is broken down to manageable and meaningful pieces.
3. Organizing content into sequences and hierarchies is facilitated.
4. Evaluation is simplified and becomes self-evident.
5. Selecting of materials is clarified (The result of knowing precisely what youngsters are to do leads to control in the selection of materials, equipment and the management of resources generally).

Specifying the Content: to be covered You should determine the area of the content you want to test. It is through the content that you will know whether the objectives have been achieved or not.

Preparation of the Test Blueprint: Test blueprint is a table showing the number of items that will be asked under each topic of the content and the process objective. This is why it is often called Specification Table. Thus, there are two dimensions to the test blueprint, the content and the process objectives.

As mentioned earlier, the content consists of the series of topics from which the competence of the pupils is to be tested. These are usually listed on the left hand side of the table. The process objectives or mental processes are usually listed on the top-row of the table

. The process objectives are derived from the behavioral objectives stated for the course initially. They are the various mental processes involved in achieving each objective. Usually, there are about six of these as listed under the cognitive domain via: Knowledge, Comprehension, Analysis, Synthesis, Application and Evaluation.

i. **Knowledge or Remembering;** This involves the ability of the pupils to recall specific facts, terms, vocabulary, principles, concepts and generalizations from memory. This may involve the teacher asking pupils to give the date of a particular event, capital of a state or recite multiplication tables. Examples:

Behavioral objectives: To determine whether students are able to define technical terms by giving their properties, relations or attributes. Question: Volt is a unit of (a) weight (b) force (c) distance (d) work (e) volume You can also use picture tests to test knowledge of classification and matching tests to test knowledge of relationships.

ii. **Comprehension and Understanding:** This is testing the ability of the pupils to translate, infer, compare, explain, interpret or extrapolate what is taught. The pupils should be able to identify similarities and differences among objects or concepts; predict or draw conclusions from given information; describe or define a given set of data i.e. what is democracy? Explain the role of chloroplast in photosynthesis.

iii. **Application**: Here you want to test the ability of the students to use principles; rule and generalizations in solving problems in novel situations, e.g. how would you recover table salt from water?

iv. **Analysis**: This is to analyze or break an idea into its parts and show that the student understands their relationships.

v. **Synthesis**: The student is expected to synthesize or put elements together to form a new matter and produce a unique communication, plan or set of abstract relations.

vi. **Evaluation:** The student is expected to make judgments based upon evidence.

Weighting of the Content and Process Objectives

The proportion of test items on each topic depends on the emphasis placed on it during teaching and the amount of time spent. Also, the proportion of items on each process objectives depends on how important you view the particular process skill to the level of students to be tested. However, it is important that you make the test a balanced one in terms of the content and the process objectives you have been trying to achieve through your series of lessons. Percentages are usually assigned to the topics of the content and the process objectives such that each dimension will add up to 100%. (See the table below). After this, you should decide on the type of test you want to use and this will depend on the process objective to be measured, the content and your own skill in constructing the different types of tests.

Determination of the Total Number of Items

At this stage, you consider the time available for the test, types of test items to be used (essay or objective) and other factors like the age, ability level of the students and the type of process objectives to be measured. When this decision is made, you then proceed to determine the total number of items for each topic and process objectives as follows:

(i) To obtain the number of items per topic, you multiply the percentage of each by the total number of items to be constructed and divide by 100. This you will record in the column in front of each topic in the extreme right corner of the blueprint. In the table below, 25% was assigned to soil. The total number of items is 50 hence 12 items for the topic (25% of 50 items = 12 items).

(ii) To obtain the number of items per process objective, we also multiply the percentage of each by the total number of items for test and divide by 100. These will be recorded in the bottom row of the blueprint under each process objective. In the table below:

(a) The percentage assigned to comprehension is 30% of the total number of items which is 50. Hence, there will be 15 items for this objective (30% of 50 items).

(b) To decide the number of items in each cell of the blue print, you simply multiply the total number of items in a topic by the percentage assigned to the process objective in each row and divide by 100. This procedure is repeated for all the cells in the blue print. For example, to obtain the number of items on water under knowledge, you multiply 30% by 10 and divide by 100 i.e. 3.

In summary, planning for a test involves the following basic steps:

(1) Outlining content and process objectives.
(2) Choosing what will be covered under each combination of content and process objectives.
(3) Assigning percentage of the total test by content area and by process objectives and getting an estimate of the total number of items.
(4) Choosing the type of item format to be used and an estimate of the number of such items per cell of the test blue print.

Basic Principles for Constructing Short-Answer Tests

Some of the principles for constructing multiple choice tests are relevant to constructing short-answer tests

(1) The instructions must be clear and unambiguous. Candidates should know what to do.
(2) Enough space must be provided for filing in gaps or writing short answers.

(3) As much as possible the questions must be set to elicit only short answers. Do not construct long answer-question in a short answer test.
(4) The test format must be consistent. Do not require fill in gaps and matching in the same question.
(5) The questions should be related to what is taught, what is to be taught or what to be examined. Candidates must know beforehand the requirements and demands of the test

Question Bank Defined

A Question Bank is a database of questions that can be shared between and among Courses, Learning Object Repositories, and Master Courses. Question Banks are searchable, so that questions meetings specific criteria can be drawn from them to create Assessments.

A question bank is a planned library of test items designed to fulfill certain predetermined purposes. Question bank should be prepared with at most care so as to cover the entire prescribed text. Question bank should be exhaustive and cover entire content with different types.

PURPOSES:

- To improve the teaching learning process
- Through instructional efforts the pupils growth will be obtained
- To improve evaluation process
- A pool of test items can be used for formative and summative evaluation of the pupils performance

It is a pool of readymade quality question is made available to teachers and examiners. So that they may select appropriate questions to assess pre determined objectives.

SPECIFICATION OF PAPER SETTINGS: This is a two dimensional tables by concerned university given quantum of questions where objective tested are also noted. Paper setting should have a model answer where marks for the points specified .Estimated time should be sufficient for answering all questions.

QUESTION PAPER: This is usually supplied by universities to question paper setter which specifies the requirement. This is also called item card to its contains of items.

PROCEDURE OF QUESTION PAPER SETTING:

a) Plan, design for question paper
b) Make list of specification
c) View model question paper sent by the university
d) Specify types of questions, marking and possible model answer
e) Read again for any corrections
f) Give final review of question paper

g) Write, dispatch and maintain confidentiality

Planning and design: This needs syllabus review, duration of question paper, marks allotted scheme of examination and objectivity of recall, interpretation as well as problem solving, weighing for all types of questions and for content topic given. Guideline and instruction are noted.

Specification: Objectives sorted out to elicit knowledge, understanding and application.

Model question paper: After view of model question paper sent by university, tally with your design and specifications, at this stage we can rectify our design and specification suiting to affiliated universities.

Types of questions and marking: By reforming sample U.G and P.G question paper you can give this information.

Corrections: This helps in refining, editing to question paper. Clarify and spelling mistakes are got corrected.

Review: This helps in checking options, sections, difficulty level of the paper, grouping, numbering and instructions are checked.

Confidentiality: All precautions for maintaining confidentiality as required. Suggested checklist to review the question paper

- within curriculum
- Syllabus covered
- Weight age to topics
- Not too easy
- Not difficult
- Performance possible by category of learners
- What is the format
- What about clarify
- Any repetition
- Answerable in given time duration
- Standardized Above lines of approach, systematically done, definitely improve question paper setting in the examination

What are open book exams?

Open book exams allow you to take notes, texts or resource materials into an exam situation. They test your ability to find and apply information and knowledge, so are often used in subjects requiring direct reference to written materials, like law statutes, statistics or acts of parliament.

Open book exams usually come in two forms:

- Traditional sit-down / limited-time exams, with varying degrees of access to resources and references.
- Take home exams–open book exams you do at home. Question(s) are handed out, answers are attempted without help from others, and the exam is returned within a specified period of time (often the next day).

What kinds of material can be used?

The materials you can take into an Open Book exam can vary. Some restrict the type of materials (e.g. formula sheets and tables or a limited number of texts), others may be totally unrestricted (any quantity of any material).

Materials might be:

- your notes
- readings, reference materials or textbooks
- equipment like calculators, drafting tools etc.

Materials used in Take Home exams are usually unrestricted. Check your course guide or with your lecturer to find out what you can use. The main restriction for Take Home exams is that they must be your work–you must attempt them by yourself without any help from others.

Why some exams are 'open book'?

Because they test for more than just rote-learning. At university, simply memorizing and repeating information is not enough to get you a good mark. Higher education is supposed to equip you with intellectual abilities and skills. Open Book exams test your ability to quickly find relevant information and then to understand, analyze, apply knowledge and think critically.

What kinds of questions will an open book exam have?

Open Book Exams don't test your memory. They test your ability to find and use information for problem solving, and to deliver well-structured and well-presented arguments and solutions.

Open Book exam questions usually require you to apply knowledge, and they may be essay-style questions or involve problem solving or delivering solutions. The style of question depends on the faculty or school setting the exam. For example in Law, the questions may set up a hypothetical fact situation that you will need to discuss.

Misconceptions about open book exams

1) Open Book exams are a breeze

Open Book exams are not an easy option. Answering the questions well requires more than just copying information straight from texts. For example, having access to a textbook can stop you from giving a wrong answer if you can't remember a fact or formula, but just getting that fact correct won't get you good marks. In Open Book exams, it's how you locate, apply and use the information that is important.

2) You don't have to study

Probably the biggest misconception about Open Book exams is that there is no need to study anything. However, you should study just as you would for any other exam. Having books and notes to refer to might mean you don't have to memories as much information, but you still need to be able to apply it effectively.

This means you must fully understand and be familiar with the content and materials of your course so you can find and use the appropriate information. In Open Book exams, you need to quickly find the relevant information in the resources you have. If you don't study you won't be able to–you won't know where it is.

3) You can just copy straight from the book!

You can't copy chunks of text directly from textbooks or notes. This is plagiarism. In Open Book exams, the resource materials are made available to you, so you are expected to do more than just reproduce them. You must be able to find, interpret and apply the information in your sources to the exam questions. You usually need to reference as well, just as you would for any other assignment.

4) The more materials the better!

Don't get carried away and overload with materials and resources in the exam. Only take what you need. Stacks of books won't necessarily guarantee your performance, and you won't have time for extensive reading. Too many materials can end up distracting you and crowding up your work space. Carefully select your materials and organize them for quick reference.

Advantages of Open-book Examination

- Less demanding on memory (regurgitation of memorized materials) because it is no longer necessary for students to cram a lot of facts, figures and numbers for open-book examination
- Provides a chance for students to acquire the knowledge during the preparation process of gathering suitable learning materials rather than simply recalling or rewriting it
- Enhances information retrieval skills of students through finding the efficient ways to get the necessary information and data from books and various resources
- Enhances the comprehension and synthesizing skills of students because they need to reduce the content of books and other study materials into simple and handy notes for examination

Disadvantages of Open-book Examination

- Difficult to ensure that all students are equally equipped regarding the books they bring into the exam with them, because the stocks of library books may be limited and also some books may be expensive to students
- More desk space is needed for students during the examination because students often need lots of desk space for their textbooks, notes and other reference materials
- Sometimes students may spend too much time on finding out which parts of the books to look for answers instead of applying the knowledge, practical skills and reasoning ability

- A lot of students are unfamiliar with open-book examinations. They must be provided with clear procedures and rules.

How to design a good Open-book Examination Assessment?

1. Set questions that require students to do things with the information available to them, rather than to merely locate the correct information and then summarize or rewrite it
2. Make the actual questions straightforward and clear to understand. Students usually read the questions quickly because they often want to save their time searching answers from textbooks and notes
3. Arrange a bigger venue to hold the examinations because students may need larger desks for examinations
4. Make sure there is enough time for students taking the examination. The length of open-book examination is usually longer than the traditional examination because students need extra time for searching information and data from their notes and textbooks.
5. Set up the appropriate marking criteria for open-book examinations as the aspects to be assessed in open-book examinations may be different from those in traditional examinations. For example, the assessment criteria may have to weigh more on the application of knowledge, comprehension skills and critical thinking skills, rather than recalling knowledge from textbooks and notes.

Marking Rubrics

MARKING RUBRICS	**Excellent**	**Proficient**	**Average**	**Poor**
Comprehension:	Demonstrated complete knowledge of concepts or principles of the course; showed a thorough and excellent understanding in interpretation of the content from textbooks, notes and other	Reflected most of the knowledge or main points of concepts or principles; showed a good understanding in interpretation of the content from textbooks, notes, and other learning	Showed partial knowledge of some points of the concepts or principles; showed a basic understanding in interpretation from textbooks, notes, and other learning materials	Showed minimal knowledge of concepts or principles; showed a poor understanding in interpretation from textbooks, notes, and other learning materials

	learning materials	materials		
Synthesis:	Demonstrated excellent ability to look at an issue from different dimensions, and generated innovative ideas apart from searching from textbooks	Showed good ability to investigate an issue from various dimensions; attempted to generate ideas apart from searching from textbooks	Showed fair ability to look at an issue from different dimensions, but mostly base on the resources from textbooks	Showed very limited ability to investigate an issue from different dimensions
Application:	Demonstrated competent ability to elaborate and reflect on what they have learned and applied it in the context of the questions	Attempted to elaborate, but mostly summed up what they have learned and applied it in the context of the questions	Showed a general description of what they found from textbooks; attempted to apply what they have learned in the context of questions	Showed a poor understanding of what they have learned and failed to apply it in the context of questions

Online Examination

Online examinations, sometimes referred as e-examinations, are the examinations conducted through the internet or in an intranet (if within the Organization) for a remote candidate (s). Most of the examinations issue results as the candidate finish the examination, when there is an answer processing module also included with the system. Candidate is given a limited time to answer the questions and after the time expiry the answer paper is disabled automatically and answers sent to the examiner. The examiner will evaluate answers, either through automated process or manually and the results will be sent to the candidate through email or made available in the web site.

Today many organizations are conducting online examinations worldwide successfully and issue results online.

There are advantages and disadvantages in online examinations. The main advantage is that it can be conducted for remote candidates and evaluation of answers can be fully automated for MCQ questions and other essay type questions can be evaluated manually or through automated system, depending on the nature of the questions and the requirements. Also online examinations can be conducted at any time and does not incur higher cost as traditional exam scenario as there is no paper work involved(e.g.: printing exam papers, prepare paper admissions etc) , there are no invigilators, also no need of arrangement of exam centers. When comparing with traditional exam scenario the cost for an online examination will be almost zero.
The disadvantage of the e-examination is the inability of invigilating. There are methodologies used in these examinations, when registering candidates and presentation of questions, so that to test candidates knowledge and skills. However with a limited time, candidate cannot be completely tested with his skill-test.

The system

This solution will provide assistance in creating, conducting and evaluating examinations. Large organizations will be able to have a centralized database of questions, from which the tests will be prepared. Internet will be used as a media for disseminating and conducting tests, thus maintaining a uniform pattern for all the examinees throughout the organization.
The task of maintaining the record of scores and the tests for which a particular candidate has appeared will be done by the portal. The progress reports can be printed at any point of time by just providing the necessary details of a candidate. The solution can also be used by faculty members to create question papers. They can provide a question bank consisting of questions of varied difficulty levels. Numerous sets of distinct question papers can then be printed, consisting of all type of questions in equal proportion. The same can be put to use in corporate organizations and universities or colleges.

For schools, colleges, universities managing the examination have never been easy. Now school teachers and college/university professors can manage question papers very easily.

This system aims to be a powerful tool for eLearning and online education. You can create quiz, question bank, certification examination questions in any language. Useful for school, college, university, teachers and professors for managing question papers and examinations. Recruiting agencies, companies can use it for candidate's skills evaluation by conducting online test. It is very useful for parents in the academic development of kids to improve their educational skills.

The system is consisting of a web server with a database facility. This server is configured with proper security measures. Clients (candidates) can connect through the internet with a web browser (e.g.: Internet Explorer, Mozilla Firefox etc) to the server and take the exam. Examiners too can connect to the server through the internet or through the intranet for setting up papers and to do other related tasks.

• **<u>NEED: Theory about examinations and why the need for online exams</u>:**

The traditional approach to measuring a person's level of knowledge in a topic has been the examination. These days there is often more emphasis on "internal" assessments, which may consist of assignments and projects given out by the teacher and then marked or assessed by the same teacher. Examinations have the advantages of:

• Confidence that a large number of students are all being assessed equally.
• Reduced opportunity for cheating.
• Less marking work, where an entire year's assessment can be made based on the output of students over a 2 or 3 hour period.

This solution will provide assistance in creating, conducting and evaluating examinations. Large organizations will be able to have a centralized database of questions, from which the tests will be prepared. Internet will be used as a media for disseminating and conducting tests, thus maintaining a uniform pattern for all the examinees throughout the organization.

The task of maintaining the record of scores and the tests for which a particular candidate has appeared will be done by the portal. The progress reports can be printed at any point of time by just providing the necessary details of a candidate. The solution can also be used by faculty members to create question papers. They can provide a question bank consisting of questions of varied difficulty levels. Numerous sets of distinct question papers can then be printed, consisting of all type of questions in equal proportion. The same can be put to use in corporate organizations and universities or colleges.

The Internet creates opportunities for making examinations both more reliable and cheaper than they are already. To see why, we need to look at what the components of an examination process are. Ignoring the efforts of the examinees, there are three major components of effort required to administer an examination:

1. *Create the examination*, and keep the contents secure until the examination starts.

2. *Supervise the examination* – supervise the entry of students into the examination environment, identify them if necessary, ensure that they do not have prohibited materials or communication devices, make sure they do not communicate with each other, and collect the examination papers at the end of the examination period.

3. *Conduct the examinations.*

Using computer technology, it should be possible to streamline these three phases, and also disconnect them from each other.

Also, online examination system offers the following features that throw light on its need today:

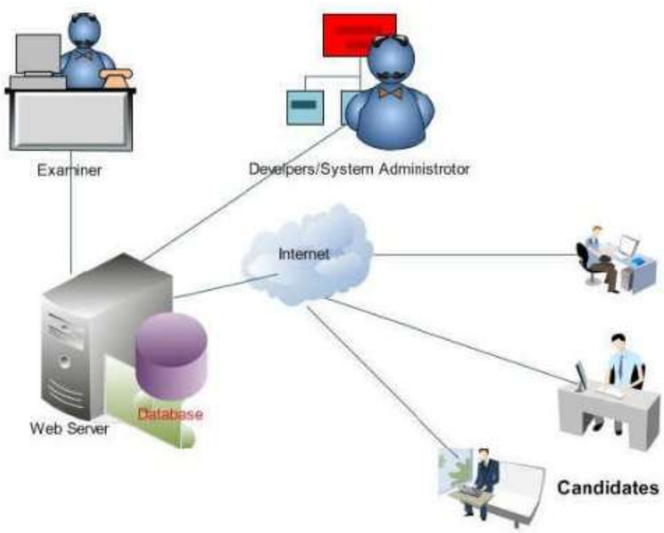
A) Reliability - An examination result is only useful if it can be trusted. If a potential employer judges a potential employee on the basis of an examination result, the employer needs to trust the creation process, the supervision process and the marking process. Where these are separated from each other, the examination result must explicitly show the parties responsible for each component of the process, i.e. who created the examination, who supervised it and who marked it (or, if it was marked automatically, what algorithm was used to mark it). Anyone judging the reliability of a result must take into account all identified responsible parties.

B) User details - The software maintain the details for each student as well as provides a facility for editing the student details if required. There is Password based access for Individual Answer Sheet, General Mark Lists and Trends Graphs.

<u>Functioning of the Online Examination System</u>:

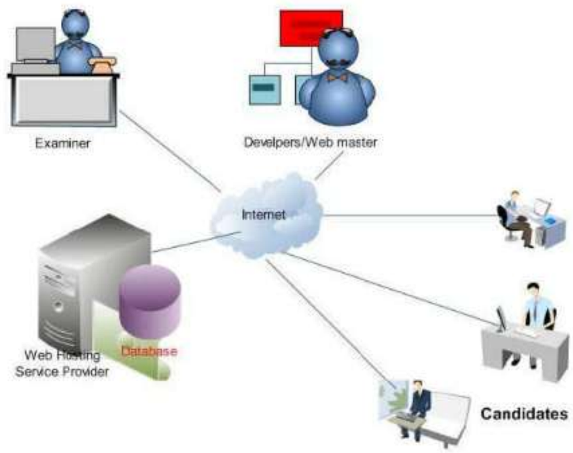
The system is consisting of a web server with a database facility. This server is configured with proper security measures. Clients (candidates) can connect through the internet with a web browser (eg: Internet Explorer, Mozilla Firefox etc) to the server and take the exam. Examiners too can connect to the server through the internet or through the intranet for setting up papers and to do other related tasks. There are two options available to have a web server facility

- **Option 1 -Having an in-house web server**



In this system the web server and database management system is installed for management of databases, these databases could be Candidates database, Questions database and if the automated evaluation is carried out a separate database for evaluation purposes. The web server can be on windows platform which uses Microsoft Windows Server with Internet Information System (IIS) as the web server software.

- **Option 2 - Obtain web hosting facility with server side programming and database facility from a web hosting service provider**



Under this setup the developers upload locally developed server scripts to the web server through internet and set up the databases remotely with the obtained server facility and there is no system administration involved.

2. The Overall Description

2.1 Product Perspective

The Online Examination System is a web based application. The system **can be modified and customized to suit the need of any Educational Institutions, Primary and Secondary Schools, Colleges, Professional and Vocational Institutes, Universities or Training Academies.** This software is particularly suited to conduct competitive exams like recruitment exams and Common Entrance Tests (CET) of various states as it will save valuable time spent on assessing the answer books and the results can be obtained immediately. This system software offers the following:

- Fully automated web based examination software
- Register students online
- Automated test creation / randomized questions
- Centralized administrator controls
- All the features are customizable

How the software (system) works: The software maintains the details for each student as well as provides a facility for editing the student details if required. There is **Password based access for Individual Answer Sheet, General Mark Lists and Trends Graphs**. Authentication is provided Via Sessions and the password is stored in encrypted form (MD5 hash) in the database.

A brief description of how the system works:

- Applicants (Candidates) apply for the exam online or through other medium
- After evaluating the applications admissions are sent to the accepted candidates
- Candidates register and take the exam and submit the answers.
- After evaluating answers examiner issues result.

The product (system) aims at reducing costs associated with conducting exams over a period of time and achieving total automation of examination system- related tasks like registration, publication of results, which leads to a very high degree of system efficiency.

2.2 Product Features

Some basic facilities provided by the system are listed below. They will be discussed further in greater detail.

Few of them are:

1. In built Question database for exam questions
2. Access anywhere, anytime Application
3. Exam Format intuitive and easy to navigate.

4. Administrators load the questions into the database
5. Examinations are generated automatically as per student selections
6. Exams and questions can be edited, deleted, and re-used anytime
7. Faculty can also upload required study material and references for test takers.
8. Communities or groups can be formed for specific organizations.
9. The results can be viewed online and printed.
10. A variety of analysis options based on different parameters or combinations of different parameters for faculty/examiners conducting the test.

2.3 User Characteristics

The users of this system are precisely students and teachers. Students need to know how to access internet and make their profiles with the system. Teachers need to possess a good skill of certain languages that have to be used while uploading of questions etc. overall users of this system do not need to have a very high technical know-how. The teachers also need to identify the need to upload study material and practice tests. The assessment which can be done in various forms requires understanding of basic mathematical functions and graphs. This system is user-friendly.

2.4 Operating Environment

The software will be a web-based application, meaning it runs from a browser. The product should be able to be run from a remote client machine with an Internet connection. This decision is made depending on where the data file is stored. If it is stored locally, then the program will run on the local machine. The external interface with be through the browser, through ASP.Net version 3.5 HTML and VC#, Ajax along with some JavaScript. The hardware running the software will be Pentium 4 2.4 GHz, running Windows Vista or later using SQL Server 2008 Express Edition as the Database system and the development environment will be Visual Studio 2008. Along with this the inbuilt Microsoft IIS server will be used in the process.

2.5 Design and Implementation Constraints

- The system will not contain any audio component to enable the visually impaired to use the product.
- The medium of instruction will be English only.
- The internet connection used by the organization should have a speed of 1 gbps or higher as the navigation between questions would be done in acceptable time limits.2.6

User Documentation
An online product guide and sitemap will be embedded in the product website.

2.7 Assumptions and Dependencies

All the software and hardware products mentioned are assumed to be available with the developers. To fulfill server space constraints any freely available format converters could be used.

3. System Features

3.1 Basic Features

• Type of Questions:
Questions can be multiple choice, true/false, short answer (fill in the blank), multiple answer (checkboxes), and essay.

• Exams can be timed or untimed.
i. Exams can be turned on or off and have a date range for each exam.
ii. The system can automatically create exams by randomly choosing questions from your exam question database.
iii. The system will randomly order the questions for each student.

• The total number of questions can vary from paper to paper.

Answers processed and marks awarded instantly. Results can be viewed immediately upon Submission. Generation of Individual Answer Sheets, General Mark Lists Summary (Total marks Obtained) as well as Detailed (Topic Wise marks Obtained).

• The topics in the detailed General Mark List can be specified by you along with the minimum pass marks for the subject.
• Can include pictures, graphics files in the questions of the online examination.
• The computerized examination system has a facility to add students details, View Student List as well as Edit student details. The Student details that you wish to store in the database can be specified by you.
• The Customized Online examination system has a password based authentication system for students as well as System Administrator. The passwords are encrypted and stored in SQL Server database.
• The System Administrator as well the Student can change their own passwords. In addition the administrator can edit all details of the student.

• Practice Tests - You'll have the option of simply providing a large bank of practice test questions that users can practice with, or to provide a tutorial area as well, making your product full test prep "course" as opposed to a bank of practice questions.
• Results: Students and instructors get instant results. Results can also be printed as soon as the test finishes.

3.2 Advanced Features

• The examiner who uploads the questions gets an option of specifying whether the questions can be used for practice tests or not. • The candidate should be able to register online for the examination online after accepting the application (Optional) • the system should be capable of issuing online-admission cards (index numbers/ or passwords) for exam entry. (Optional)

4. Requirements

4.1 User Interface

the external user interface forms are designed using standard tools available in Microsoft Visual Studio 2008 Professional Edition. Some interfaces may also contain Macromedia Flash components. Error messages will be displayed in message boxes.

4.2 Hardware Requirements

Hardware requirements for the system to work are:
• CPU: Any CPU which can be in the form of external (CPU for PC) or internal (for laptops)
• Monitor: any monitor which is capable of displaying the signals sent by the computer
• Keyboard: A standard QWERTY keyboard for data entering.
• Mouse: Any standard mouse
• Printer: Any printer it could be Laser printer, Inkjet Printer or even a Dot- matrix printer will work fine for the software

4.3 Software Requirements

Software requirements for the system to work are:
• Microsoft Windows (95, 98, ME, NT, XP or Vista): The software will work on any of the Microsoft Windows OS.
• ASP.Net

4.4 Communications Requirements

Communication Requirements for the system to work are:
• Web Browsers: Internet Explorer (7, 8), Mozilla Firefox, Safari, Opera.

- LAN Connection

5. Non-Functional Requirements:

5.1. Performance Requirements

- Should be capable of giving access to concurrent users without degrading the system performance and accept answers.
- Sessions of each candidate should be synchronized with server and duration calculations should be done according to the server time.

5.2. Safety Requirements

- The system should be designed in as a secured system applying safety measures
- Special exception handling mechanism should be in place to avoid system errors.
- In case of scenarios where data integrity can be compromised, measures should be taken to ensure that all changes are made before system is shut down.

5.3. Security Requirements

- Only registered students are able to access a test.
- The test can be made available on specific dates.
- The test can be made available at specific times.
- The number of times students access tests can be set.
- The login time per student is available.
- Logoff time per student is available.
- The system should be synchronized with the server time and should be capable of disable answer sheet automatically after time out.

5.4. Software Quality Attributes Availability:

The system should be available 24/7.It should always provide real time information.

Reliability: The system should be robust enough to have a high degree of fault tolerance. The system should not crash and should identify the invalid input and produce a suitable error message. It should be able to recover from hardware failures, power failures and other natural catastrophes and rollback the databases to their most recent valid state. Usability: The system should provide an easy-to-use graphical interface similar so that the users do not have to learn a new style of interaction. The web interface should be intuitive and easily navigable Users should be able to understand the menu and options provided by the system. Any notification or error messages generated, should be clear, succinct, polite and free of jargon. Integrity: Only system administrator has the right to change system parameters, such as pricing policy etc. The system should be secure and must use encryption to protect the databases. Users need to be authenticated before having access to any personal data.

EXISTING SYSTEMS - There are many online examination systems available in the market today. In this part of the discussion we will briefly describe the existing systems and undermine the features of the proposed system. These features are the improvements or advantages of our proposed system over the existing systems. Online test plus, Sify-itest, Exam-Pro Software, Exam 9 products, iSummation Technologies and many more are the widely used online examination software today.

Credit System

Why are academic credit systems useful?

Here are some benefits of academic credit systems:

- Keep track of student process and determine when he or she has met the requirements for awarding a specific academic degree.
- Offer a good estimate of the workload of a programme and how it is distributed between the different higher education activities.
- Some academic credit systems permit students to transfer to different programmes and keep part or all of their previously earned credit points.
- If a student has not graduated a study programme, but wants to find a job, earned academic credits may be presented to employers as proof of previous studies.
- Some universities use academic study credits to set programme costs.

Exam on Demand

Introduction

National Institute of Open Schooling has been working on the concept of On-Demand Examination (ODE) at Secondary Level since 2003 in the area of its feasibility and operationalisation. The novel concept of ODE is a great step in the direction of flexibility to the open and distance learning. This will make the total system of examination independent of the time frame and will help the student to take up the examinations as per their wish and preparation. The basic concept of On Demand Examination is that NIOS Student can walk into the examination center as and when he/she feel ready for the examination. NIOS re-introduced the ODES at Secondary level in 2005. With the success of ODE at Secondary level, NIOS started the On-Demand Examination in Sr.Secondary also from October 2007. At present, ODES is being conducted at NIOS HQ at NOIDA and at its 13 Regional Centres/Sub Centres of NIOS in the following subjects at Secondary and Sr.Secondary level.

- **Secondary :** Hindi (201), English (202), Sanskrit (209), Mathematics (211), Science & Technology (212), Social Science (213), Economics 214), Business studies (215), Home Science (216), Data Entry Operations (229), sychology (222), Indian Culture & Heritage (223) & Painting (225).
- **Sr. Secondary :** Hindi (301), English (302), Sanskrit (309), Mathematics (311), Physics (312), Chemistry (313), Biology (314), History (315), Geography (316), Political Sc. (317), Economics (318), Business Studies (319), Accountancy (320), Home Science (321), Psychology (328) ,Sociology (331), Painting (332),Environmental Science (333) and Data entry Operations (336).
- **Medium of Question Paper:** The Question Paper will be in English and Hindi Medium Only, However the Regional Medium Candidates want to appear through the On Demand Examination will have a choice of Answering the Question Paper in the respective Regional Medium and no Regional Medium Question Paper will be provided in On Demand Examination.

The Features

A set of Question Paper is generated randomly by the computer out of already developed Question Bank based on the blue print and paper design of the subject as and when demanded. All such generated Question Papers are different with same difficulty level. For this a huge database of items based on the blue print and question paper design has been developed A number of items having comparable difficulty level are developed for each activated/marked cells of the blue print. These questions covered in the item bank test learning objectives under knowledge, understanding, application and skill competencies of a Student. For identification, all these items are given a code indicating the subject, the content area to which the item belong, the objective being tested, the type of question, the marks allocated to the item and the serial number of the item.

Advantages

The advantages of ODES can be stated as follows:

- Allows the Student to get assessed when he/ she is ready. Readiness depends on the Student and not on the institution.
- Attempts to remove the stress of appearing in examination(s), whether for all subjects or in one subject at a fixed time and schedule.
- Attempts to remove the threat of failure in examination.
- Removes frustration, loss of self esteem, peer group ridicule, depression etc.
- Knowledge of results is almost immediate and success, even in one subject, is a strong motivating factor.
- Degree and level of performance is decided by the Student who can reappear in the examination as many times as one wants, till satisfied.

- # Malpractices will be reduced, as it is a system where the tools for evaluation are unique to

Venue for theory paper	**At all the Regional Centres and at the HQ at NOIDA**

an individual Student. Every question paper for each Student is different having comparable difficulty level.

- Respects the individuality and sovereignty of each Student.

Every Month through NIOS web site www.nios.ac.in for the ODE conducted during last month.

- After due processing the ODES result will be declared and published through NIOS web site during last week of every month for the examinations conducted during the preceded month.
- The mark sheet and other documents will be printed only for overall pass candidates by M&M unit and will be sent to the concerned Regional Centre for dispatch to the individual candidate.
- The other candidates wanting mark sheet can send their request through an application attaching the Internet print out and bank draft of Rs.100/- in favour of Secretary, NIOS, payable at NOIDA to Section Officer (M&M unit), National Institute of Open Schooling A-24-25, Institutional Area, Sector-62, NOIDA, U.P. 201309
- For the purpose of Mark Sheets and certification, all the candidates appearing during May till October will be considered under April Examination of that year. And all the candidates appearing during November till April will be consider under preceeding October Examination.
- "Re-Checking" or "Re-Evaluation" as per NIOS examination norms is allowed in case of ODES also. The candidate may apply to Director (Evaluation) for this on the prescribed proforma available on the NIOS web site along with requisite fees mentioned in the proforma.

Examination Days for theory paper	Tuesday, Wednesday & Thursday (For Regional Centres) and Tuesday,Wednesday,Thursday & Friday for HQ NOIDA 2.00 pm to 5.00 pm
Venue for Practical Examination	NIOS AI (study centre) or at another reputed school already identified for the practical Examination
Examination Days for Practical	Every Friday/Saturday on weekly basis or as Announced on the day of the Theory Paper.
Result Declaration	Every Month through NIOS web site www.nios.ac.in for the ODE conducted during last month

Measures of Central Tendency

Introduction

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the median and the mode.

The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. In the following sections, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.

Mean (Arithmetic)

The mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data (see our Types of Variable guide for data types). The mean is equal to the sum of all the values in the data set divided by the number of values in the data set. So, if we have n values in a data set and they have values x_1, x_2, ..., x_n, the sample mean, usually denoted by $\bar{x}$ (pronounced x bar), is:

$$\bar{x} = \frac{(x_1 + x_2 + \cdots + x_n)}{n}$$

This formula is usually written in a slightly different manner using the Greek capitol letter, $\sum$, pronounced "sigma", which means "sum of...":

$$\bar{x} = \frac{\sum x}{n}$$

You may have noticed that the above formula refers to the sample mean. So, why have we called it a sample mean? This is because, in statistics, samples and populations have very different meanings and these differences are very important, even if, in the case of the mean, they are calculated in the same way. To acknowledge that we are calculating the population mean and not the sample mean, we use the Greek lower case letter "mu", denoted as μ:

$$\mu = \frac{\sum x}{n}$$

The mean is essentially a model of your data set. It is the value that is most common. You will notice, however, that the mean is not often one of the actual values that you have observed in your data set. However, one of its important properties is that it minimises error in the prediction of any one value in your data set. That is, it is the value that produces the lowest amount of error from all other values in the data set.

An important property of the mean is that it includes every value in your data set as part of the calculation. In addition, the mean is the only measure of central tendency where the sum of the deviations of each value from the mean is always zero.

When not to use the mean

The mean has one main disadvantage: it is particularly susceptible to the influence of outliers. These are values that are unusual compared to the rest of the data set by being especially small or large in numerical value. For example, consider the wages of staff at a factory below:

Staff	1	2	3	4	5	6	7	8	9	10
Salary	15k	18k	16k	14k	15k	15k	12k	17k	90k	95k

The mean salary for these ten staff is $30.7k. However, inspecting the raw data suggests that this mean value might not be the best way to accurately reflect the typical salary of a worker, as most workers have salaries in the $12k to 18k range. The mean is being skewed by the two large salaries. Therefore, in this situation, we would like to have a better measure of central tendency.

As we will find out later, taking the median would be a better measure of central tendency in this situation.

Another time when we usually prefer the median over the mean (or mode) is when our data is skewed (i.e., the frequency distribution for our data is skewed). If we consider the normal distribution - as this is the most frequently assessed in statistics - when the data is perfectly normal, the mean, median and mode are identical. Moreover, they all represent the most typical value in the data set. However, as the data becomes skewed the mean loses its ability to provide the best central location for the data because the skewed data is dragging it away from the typical value. However, the median best retains this position and is not as strongly influenced by the skewed values. This is explained in more detail in the skewed distribution section later in this guide.

Median

The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by outliers and skewed data. In order to calculate the median, suppose we have the data below:

65	55	89	56	35	14	56	55	87	45	92

We first need to rearrange that data into order of magnitude (smallest first):

14	35	45	55	55	**56**	56	65	87	89	92

Our median mark is the middle mark - in this case, 56 (highlighted in bold). It is the middle mark because there are 5 scores before it and 5 scores after it. This works fine when you have an odd number of scores, but what happens when you have an even number of scores? What if you had only 10 scores? Well, you simply have to take the middle two scores and average the result. So, if we look at the example below:

65	55	89	56	35	14	56	55	87	45

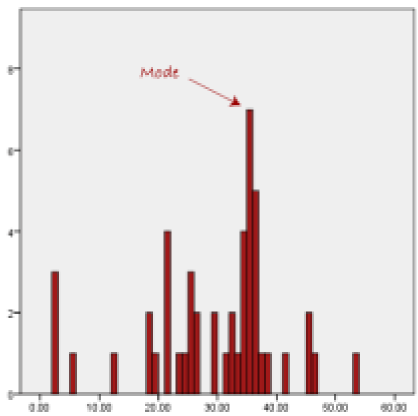
We again rearrange that data into order of magnitude (smallest first):

14	35	45	55	**55**	**56**	56	65	87	89

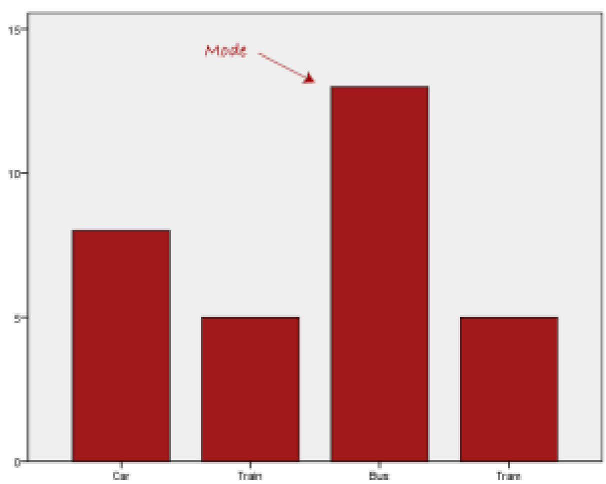
Only now we have to take the 5th and 6th score in our data set and average them to get a median of 55.5.

Mode

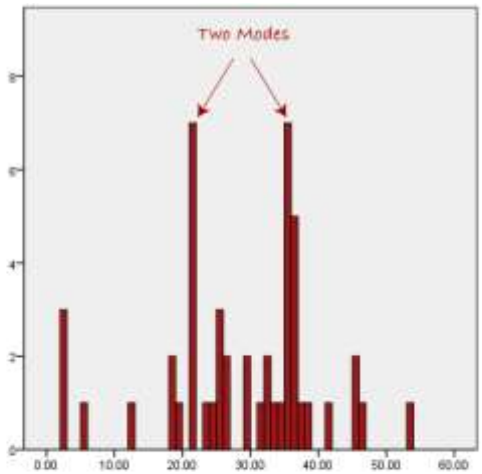
The mode is the most frequent score in our data set. On a histogram it represents the highest bar in a bar chart or histogram. You can, therefore, sometimes consider the mode as being the most popular option. An example of a mode is presented below:



Normally, the mode is used for categorical data where we wish to know which is the most common category, as illustrated below:

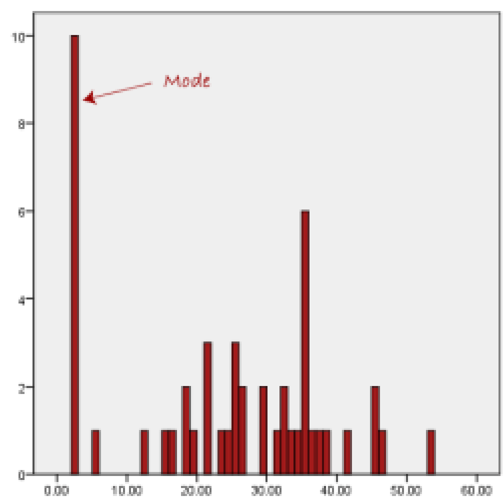


We can see above that the most common form of transport, in this particular data set, is the bus. However, one of the problems with the mode is that it is not unique, so it leaves us with problems when we have two or more values that share the highest frequency, such as below:



We are now stuck as to which mode best describes the central tendency of the data. This is particularly problematic when we have continuous data because we are more likely not to have any one value that is more frequent than the other. For example, consider measuring 30 peoples' weight (to the nearest 0.1 kg). How likely is it that we will find two or more people with **exactly** the same weight (e.g., 67.4 kg)? The answer, is probably very unlikely - many people might be close, but with such a small sample (30 people) and a large range of possible weights, you are unlikely to find two people with exactly the same weight; that is, to the nearest 0.1 kg. This is why the mode is very rarely used with continuous data.

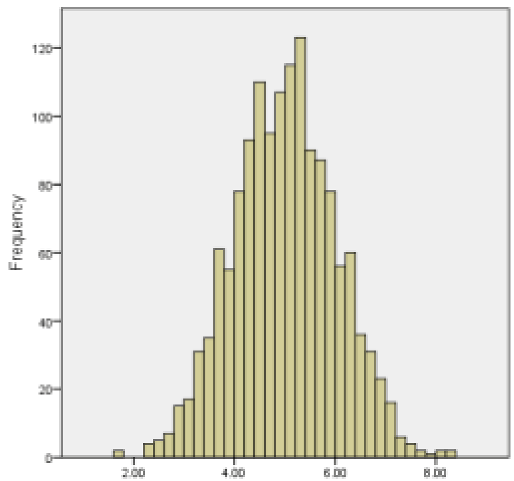
Another problem with the mode is that it will not provide us with a very good measure of central tendency when the most common mark is far away from the rest of the data in the data set, as depicted in the diagram below:



In the above diagram the mode has a value of 2. We can clearly see, however, that the mode is not representative of the data, which is mostly concentrated around the 20 to 30 value range. To use the mode to describe the central tendency of this data set would be misleading.

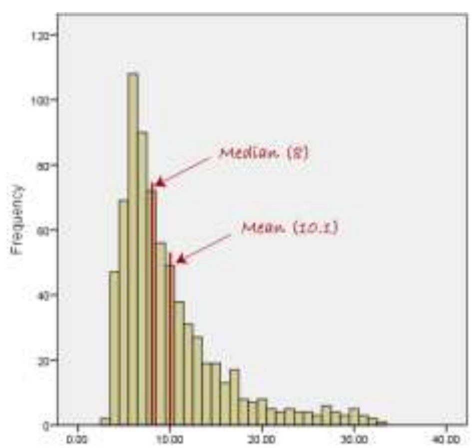
Skewed Distributions and the Mean and Median

We often test whether our data is normally distributed because this is a common assumption underlying many statistical tests. An example of a normally distributed set of data is presented below:



When you have a normally distributed sample you can legitimately use both the mean or the median as your measure of central tendency. In fact, in any symmetrical distribution the mean, median and mode are equal. However, in this situation, the mean is widely preferred as the best measure of central tendency because it is the measure that includes all the values in the data set for its calculation, and any change in any of the scores will affect the value of the mean. This is not the case with the median or mode.

However, when our data is skewed, for example, as with the right-skewed data set below:



we find that the mean is being dragged in the direct of the skew. In these situations, the median is generally considered to be the best representative of the central location of the data. The more skewed the distribution, the greater the difference between the median and mean, and the greater emphasis should be placed on using the median as opposed to the mean. A classic example of the above right-skewed distribution is income (salary), where higher-earners provide a false representation of the typical income if expressed as a mean and not a median.

If dealing with a normal distribution, and tests of normality show that the data is non-normal, it is customary to use the median instead of the mean. However, this is more a rule of thumb than a strict guideline. Sometimes, researchers wish to report the mean of a skewed distribution if the median and mean are not appreciably different (a subjective assessment), and if it allows easier comparisons to previous research to be made.

Summary of when to use the mean, median and mode

Please use the following summary table to know what the best measure of central tendency is with respect to the different types of variable.

Type of Variable	**Best measure of central tendency**
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

REPRESENTATION OF DATA

Besides the tabular form, the data may also be presented in some graphic or diagrammatic form.

"The transformation of data through visual methods like graphs, diagrams, maps and charts is called representation of data."

The need of representing data graphically:

Graphics, such as maps, graphs and diagrams, are used to represent large volume of data.

They are necessary:

· If the information is presented in tabular form or in a descriptive record, it becomes difficult to draw results.

· Graphical form makes it possible to easily draw visual impressions of data.

· The graphic method of the representation of data enhances our understanding.

· It makes the comparisons easy.
· Besides, such methods create an imprint on mind for a longer time.
· It is a time consuming task to draw inferences about whatever is being presented in non–graphical form.
· It presents characteristics in a simplified way.
· These makes it easy to understand the patterns of population growth, distribution and the density, sex ratio, age–sex composition, occupational structure, etc.

General Rules for Drawing Graphs, Diagrams and Maps

1. Selection of a Suitable Graphical Method

Each characteristic of the data can only be suitably represented by an appropriate graphical method. For example,
· To show the data related to the temperature or growth of population between different periods in time line graph are used.
· Similarly, bar diagrams are used for showing rainfall or the production of commodities.
· The population distribution, both human and livestock, or the distribution of the crop producing areas are shown by dot maps.
· The population density can be shown by choropleth maps.
Thus, it is necessary and important to select suitable graphical method to represent data.

2. Selection of Suitable Scale

Each diagram or map is drawn to a scale which is used to measure the data. The scale must cover the entire data that is to be represented. The scale should neither be too large nor too small.

3. Design

The diagram or map should have following design:
· Title: The title of the diagram/map must be clear and include -

- The name of the area,
- Reference year of the data used and
- The caption of the diagram.

These are written with different font sizes and thickness. The title, subtitle and the corresponding year is shown in the centre at the top of the map/diagram.
· Legend or Index: The index must clearly explain the colours, shades, symbols and signs used in the map and diagram. A legend is shown either at the lower left or lower right side of the map sheet.
· Direction the maps should show the direction North and properly placed on the top.

<u>Types of Diagrams</u>

The diagrams and the maps is of following types:
(i) O ne-dimensional diagrams such as line graph, poly graph, bar diagram, histogram, age, sex, pyramid, etc
(ii) Two-dimensional diagram such as pie diagram and rectangular diagram;
(iii) Three-dimensional diagrams such as cube and spherical diagrams.
The most commonly drawn diagrams and maps are:

• Line graphs
• Bar diagrams
• Pie diagram
• Wind rose and star diagram

• Flow Charts

1. Line Graph

The line graphs are usually drawn to represent the time series data related to the temperature, rainfall, population growth, birth rates and the death rates.

Construction of a Line Graph

1st step: Round the data to be shown upto the 1 digit of even numbers.

2nd step: Draw X and Y-axis. Mark the time series variables (years/months) on the X axis and the data quantity/value to be plotted on Y axis.

3rd step: Choose an appropriate scale to show data and label it on Y-axis. If the data involves a negative figure then the selected scale should also show it.

4th step: Plot the data to depict year/month-wise values according to the selected scale on Yaxis, mark the location of the plotted values by a dot and join these dots by a free hand drawn line.

Polygraph is a line graph in which two or more than two variables are shown on a same diagram by different lines. It helps in comparing the data. Examples which can be shown as polygraph are:

· The growth rate of different crops like rice, wheat, pulses in one diagram.

· The birth rates, death rates and life expectancy in one diagram.

· Sex ratio in different states or countries in one diagram.

Construction of a Polygraph

All steps of construction of polygraph are similar to that of line graph. But different lines are drawn to indicate different variables.

3. Bar Diagram

It is also called a columnar diagram. The bar diagrams are drawn through columns of equal width. Following rules were observed while constructing a bar diagram:

(a) The width of all the bars or columns is similar.

(b) All the bars should are placed on equal intervals/distance.

(c) Bars are shaded with colours or patterns to make them distinct and attractive.

Three types of bar diagrams are used to represent different data sets:

· The simple bar diagram

· Compound bar diagram

· Polybar diagram.

Simple Bar Diagram

A simple bar diagram is constructed for an immediate comparison. It is advisable to arrange the given data set in an ascending or descending order and plot the data variables accordingly. However, time series data are represented according to the sequencing of the time period.

Construction Steps: Draw X and Y-axes on a graph paper. Take an interval and mark it on Y-axis to plot data. Divide X-axis into equal parts to draw bars. The actual values will be plotted according to the selected scale.

4. Line and Bar Graph

The line and bar graphs as drawn separately may also be combined to depict the data related to some of the closely associated characteristics such as the climatic data of mean monthly temperatures and rainfall.

•

Construction:

(a) Draw X and Y-axes of a suitable length and divide X-axis into parts to show months in a year.

(b) Select a suitable scale with equal intervals on the Y-axis and label it at its right side.

(c) Similarly, select a suitable scale with equal intervals on the Y-axis and label at its left side.

(d) Plot data using line graph and columnar diagram.

5. Multiple Bar Diagram

Multiple bar diagrams are constructed to represent two or more than two variables for the purpose of comparison. For example, a multiple bar diagram may be constructed to show proportion of males and females in the total, rural and urban population or the share of canal, tube well and well irrigation in the total irrigated area in different states.

Construction

(a) Mark time series data on X-axis and variable data on Y-axis as per the selected scale.

(b) Plot the data in closed columns.

6. Compound Bar Diagram

When different components are grouped in one set of variable or different variables of one component are put together, their representation is made by a compound bar diagram. In this method, different variables are shown in a single bar with different rectangles.

Construction

(a) Arrange the data in ascending or descending order.

(b) A single bar will depict the set of variables by dividing the total length of the bar as per percentage.

7. Pie Diagram

Pie diagram is another graphical method of the representation of data. It is drawn to depict the total value of the given attribute using a circle. Dividing the circle into corresponding degrees of angle then represent the sub– sets of the data. Hence, it is also called as Divided

Circle Diagram. The angle of each variable is calculated using the following formulae.

If data is given in percentage form, the angles are calculated using the given formulae.

Calculation of Angles

(a) Arrange the data on percentages in an ascending order.

(b) Calculate the degrees of angles for showing the given values

(b) It could be done by multiplying percentage with a constant of 3.6 as derived by dividing the total number of degrees in a circle by 100, i. e. 360/100.

(c) Plot the data by dividing the circle into the required number of divisions to show the share different regions/countries

Construction

(a) Select a suitable radius for the circle to be drawn. A radius of 3, 4 or 5 cm may be chosen for the given data set.

(b) Draw a line from the centre of the circle to the arc as a radius.

(c) Measure the angles from the arc of the circle for each category of vehicles in an ascending order clock-wise, starting with smaller angle.

(d) Complete the diagram by adding the title, sub – title, and the legend. The legend mark be chosen for each variable/category and highlighted by distinct shades/colours.

Precautions

(a) The circle should neither be too big to fit in the space nor too small to be illegible.

(b) Starting with bigger angle will lead to accumulation of error leading to the plot of the smaller angle difficult.

8. Flow Maps/Chart

Flow chart is a combination of graph and map. It is drawn to show the flow of commodities or people between the places of origin and destination. It is also called as Dynamic Map.

Transport map, which shows number of passengers, vehicles, etc., is the best example of a flow chart. These charts are drawn using lines of proportional width. Many government agencies prepare flow maps to show density of the means of transportation on different routes.

The flow maps/ charts are generally drawn to represent two the types of data as given below:

1. The number and frequency of the vehicles as per the direction of their movement
2. The number of the passengers and/or the quantity of goods transported.

Requirements for the Preparation of a Flow Map:

(a) A route map depicting the desired transport routes along with the connecting stations.

(b) The data pertaining No. of trains of selected routes of to the flow of goods, Delhi and adjoining areas services, number of vehicles, etc., along with the point of origin and destination of the movements.

(c) The selection of a scale through which the data related to the quantity of passengers and goods or the number of vehicles is to be represented.

Construction

(a) Take an outline map of Delhi and adjoining areas in which railway line and the nodal stations are depicted.

(b) Select a scale to represent the number of trains. Here, the maximum number is 50 and the minimum is 6. If we select a scale of 1cm = 50 trains, the maximum and minimum numbers will be represented by a strip of 10 mm and 1.2 mm thick lines respectively on the map.

(c) Plot the thickness of each strip of route between the given rail route.

(d) Draw a terraced scale as legend and choose distinct sign or symbol to show the nodal points (stations) within the strip.

Thematic Maps

Varieties of maps are drawn to understand the patterns of the regional distributions or the characteristics of variations over space these maps are known as the distribution maps or thematic maps.

Requirements for Making a Thematic Map

(a) State/District level data about the selected theme.

(b) Outline map of the study area along with administrative boundaries.

(c) Physical map of the region. For example, physiographic map for population distribution and relief and drainage map for constructing transportation map.

Rules for Making Thematic Maps

(i) The drawing of the thematic maps must be carefully planned. The final map should properly reflect the following components:

a. Name of the area
b. Title of the subject-matter
c. Source of the data and year
d. Indication of symbols, signs, colours, shades, etc.
e. Scale
(ii) The selection of a suitable method to be used for thematic mapping.
Classification of Thematic Maps based on Method of Construction
The thematic maps are generally, classified into quantitative and non-quantitative maps.
The quantitative maps are drawn to show the variations within the data. For example, maps depicting areas receiving more than 200 cm, 100 to 200 cm, 50 to 100 cm and less than 50
cm of rainfall are referred as quantitative maps. These maps are also called as statistical maps.
The non-quantitative maps, on the other hand, depict the non–measurable characteristics in the distribution of given information such as a map showing high and low rainfall-receiving areas. These maps are also called as qualitative maps.
The construction of quantitative maps: There are three types of quantitative maps -
(a) Dot maps
(b) Choropleth maps
(c) Isopleth maps

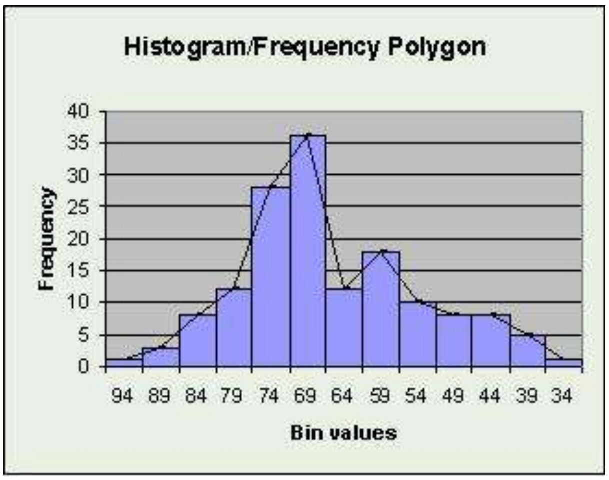
9. D ot Maps

The dot maps are drawn to show the distribution of phenomena such as population, cattle, types of crops, etc. The dots of same size as per the chosen scale are marked over the given administrative units to highlight the patterns of distributions.
Requirement
(a) An administrative map of the given area showing state/district/block boundaries.
(b) Statistical data on selected theme for the chosen administrative units, i.e., total population, cattle etc.
(c) Selection of a scale to determine the value of a dot.
(d) Physiographic map of the region especially relief and drainage maps.
Precaution
(a) The lines demarcating the boundaries of various administrative units should not be very thick and bold.
(b) All dots should be of same size.
Construction
a) Select the size and value of a dot.
(b) Determine the number of dots in each state using the given scale. For example, number of dots in Maharashtra will be 9,67,52,247/100,000 = 967.52. It may be rounded to 968, as the fraction is more than 0.5.
(c) Place the dots in each state as per the determined number in all states.

Frequency Polygon

A *frequency polygon* is another way to show the information in a frequency table. It looks a little bit like a line graph. To make a frequency polygon, you just need to plot a few points and then join the points by straight lines. So what points do you need to plot? Well, first you have to find the *midpoints* of each class. The midpoint of a class is the point *in the middle* of the class. So for instance, if I have a class "10 – 19", then the midpoint is 14.5. A class of "0 – 5" has a midpoint of 2.5.

Frequency Polygon



This is a histogram with an overlaid frequency polygon.

Midpoints of the interval of corresponding rectangle in a histogram are joined together by straight lines. It gives a polygon i.e. a figure with many angles. it is used when two or more sets of data are to be illustrated on the same diagram such as death rates in smokers and non smokers, birth and death rates of a population etc One way to form a frequency polygon is to connect the midpoints at the top of the bars of a histogram with line segments (or a smooth curve). Of course the midpoints themselves could easily be plotted without the histogram and be joined by line segments. Sometimes it is beneficial to show the histogram and frequency polygon together.

Unlike histograms, frequency polygons can be superimposed so as to compare several frequency distributions.

Frequency polygons were created in the 9th century as a way of not only storing data, but making it easily accessible for people who are illiterate.

Histogram represents data in the form of a diagram and it becomes easy for the reader to analyse and interpret to come to various conclusions and decisions. In histogram the figures are given the size proportionally so that analysis can be done by non qualified and non technical person too.

Histogram consist a series of erect drawn bars in form of the rectangles on the x axis with base equal to the corresponding class intervals and with equal base sections drawn on it. In drawing the histogram of a given continuous frequency distribution we first mark off along the x-axis all the class intervals on a suitable class. On each class interval intersect rectangles with heights proportional to the frequency of the corresponding class interval so that the area of the rectangle is proportional to the frequency of the class. If, however, the classes are of unequal width then the height of the rectangle will be proportional to the ratio of the frequencies to the width of the classes, the diagram of continuous rectangles so obtained is called histogram.

To draw the histogram for an ungrouped frequency distribution of a variable we shall have to assume that the frequency corresponding to the variate value x is spread over the interval x - $\frac{h}{2}$ to x + $\frac{h}{2}$, where h is the jump from one value to the next.

If the grouped frequency distribution is not continuous first it is to be converted into continuous distribution and then the histogram is drawn. Although the height of each rectangle is proportional to the frequency of the corresponding class, the height of a fraction of the rectangle is not proportional to the frequency of the class so that histogram cannot be directly used to reach frequency over a fraction of a class interval.

Histogram Definition

A two dimensional frequency density diagram is called a histogram. Histograms are diagrams which represent the class interval and the frequency in the form of a rectangle. There will be as many adjoining rectangles as there are class intervals and is a graphical representation showing a visual impression of data, first introduced by karl pearson. It can estimate the probability distribution of a continuous variable. Frequency distributions can be represented graphically using histograms. The midpoint occurs when about half the area is to the left and half is to the right. Histogram is a type of graph which is also called as frequency density diagram

Histogram Formula

If 'd' is the gap between the upper limit of any class and the lower limit of the succeeding class the class boundaries for any class are then given by:
Upper class boundary = Upper class limit + $\frac{d}{2}$
Lower class boundary = Lower class limit + $\frac{d}{2}$

Types of Histograms Individual data points are grouped together in classes, so that you can get an idea of how frequently data in each class occur in the data set. A histogram with more number of class intervals is more effective in depicting the structure of the frequency distribution.

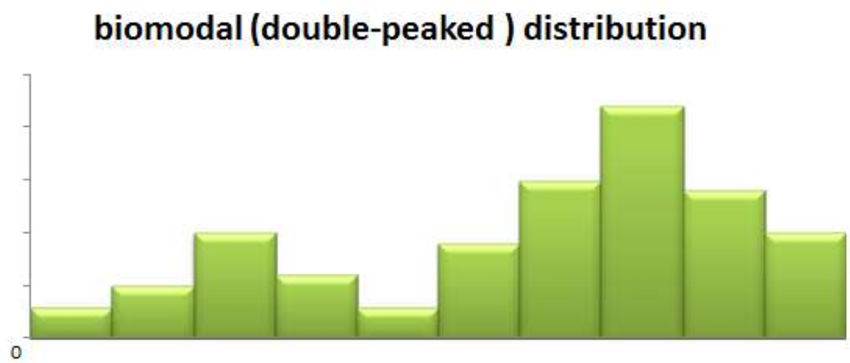
Different types of histograms are explained below.

Frequency histogram

Histogram is a graphical display of data using bars of different heights. The relative frequency table is a compact numerical way to present how the data is distributed. If the frequencies are plotted as columns, the resulting plot is called a histogram.

Double peaked distribution

Looks like the back of a two humped camel. Outcomes of two processes with different distributions are combined in one set of data. In some cases bimodal histogram indicates that the sample can be divided into two sub samples that differ from each other in some way.It will have two peaks.



J shaped distribution

Extreme case of negative skew. The reverse J or 'ski-jump' shape is an extreme case of positive skew. Cases are scoring tight up against a fixed end limit of a scale. J is like a heap which has one half missing, because the scale comes to an end so there is nowhere for the expected other half of the distribution to fall.

Skewed Histogram

A non symmetric histogram is called skewed if it is not symmetric. If the upper tail is longer than the lower tail then it is positively skewed. If the upper tail is shorter than it is negatively skewed. A skewed distribution is one that is not symmetrical, but rather has a long tail in one direction.

Uniform Histogram

A uniform distribution often means that the number of classes is too small, each class has about the same number of elements. It may describe a distribution which has several peaks. Uniform histogram have all the bars of same height.

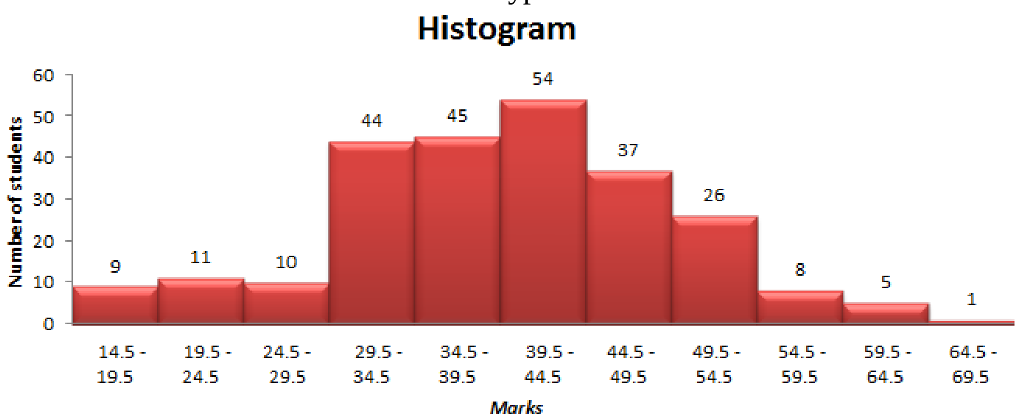
<u>Histogram Example Problems</u>

Example 1: Distribution of marks of 250 students is obtained as follows. **Draw a histogram for the given data.**

Marks	*Number of students*
14.5 - 19.5	9

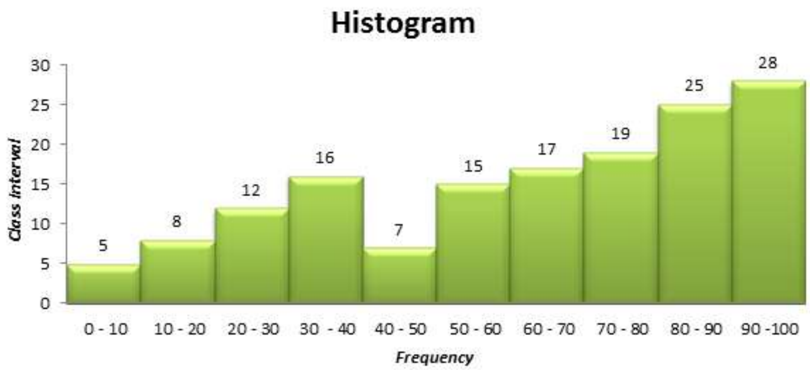
19.5 - 24.5	11
24.5 - 29.5	10
29.5 - 34.5	44
34.5 - 39.5	45
39.5 - 44.5	54
44.5 - 49.5	37
49.5 - 54.5	26
54.5 - 59.5	8
59.5 -64.5	5
64.5 - 69.5	1

Solution: Since the grouped frequency distribution is not continuous, we first convert it into a continuous distribution with exclusive type classes as given below. The upper and lower class limits of the new exclusive type classes are known as class boundaries.



Example 2: Draw a histogram for the following data

Class Interval	*Frequency*
0 - 10	5
10 - 20	8
20 - 30	12
30 - 40	16
40 - 50	7
50 - 60	15
60 - 70	17
70 - 80	19
80 - 90	25
90 -100	28



Histogram Vs Bar Graph

The table shown below compares histogram and bar graph.

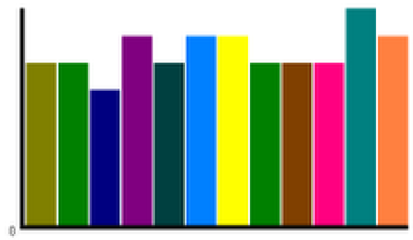
Histogram	**Bar Graph**
1. It consists of rectangles touching each other	It consists of rectangles normally separated from each other with equal space
2. The frequency is represented by the area of each rectangle	The frequency is represented by height. The width has no significance.
3. It is two dimensional	It is one dimensional

Types of Histograms

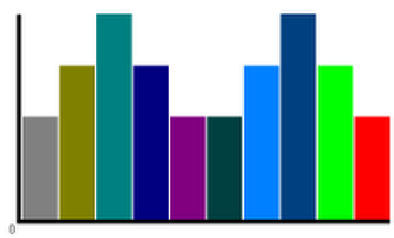
Histograms help us to determine which causes dominate. Histogram gives a clear picture of the location and variation in a data set. However, histograms can be manipulated to show different pictures, If too many bars are used then the data can be misleading. In a histogram, frequency is measured by the area of the column. In a vertical bar graph, frequency is measured by the height of the bar.

Given below are different types of histograms

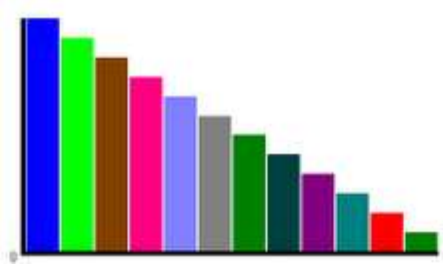
- **Uniform:** Uniform distribution gives very little information about the data set. For example when tossed a coin there may not be frequent heads so when a tail occurs there will be change in the pattern. Here we can see that the number of classes will be too less.



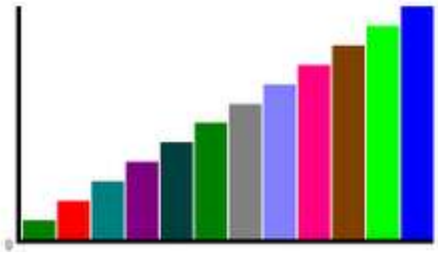
- **Bimodal:** A bimodal shape will have two peaks which gives us an insight that the data is from two different systems. The two sources are then individually analyzed.



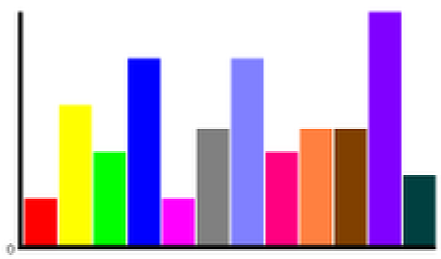
- **Symmetric:** A histogram is said to be symmetric if it's right half is exactly similar to it's left half.
- **Skewed Right:** The graph given below shows the distribution is skewed to the right and we call it as positively skewed.Here the values will be greater than zero.



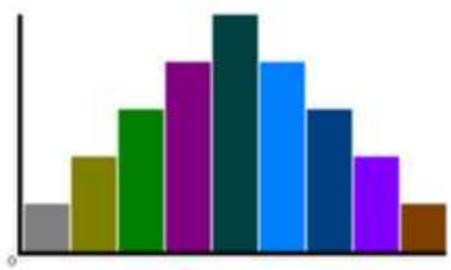
- **Skewed Left:** The graph given below shows the distribution is skewed to the left and we call it as negatively skewed.All the collected data has values less than 100.



- **Random:** A random distribution will not follow any pattern for the data set and it will have several peaks where the sources of variation are combined together and we analyze it separately. In this type of distribution we can come across different classes.If there are no multiple sources of variations then we group the pattern, and find if there is any kind of useful information in that data.



- **Bell shaped:** Bell shaped curve mostly looks like a Normal distribution and statistical calculations must be used to verify whether the given data follows the normal distribution or not.



Histogram statistics:

For histograms, the following statistics are calculated:

Mean	***The average of all the values.***
Median	***Median is the middle number.If it is an even number then the median is the mean of***

	the two middle numbers.
Mode	***The value that occurs most often***
Minimum	***The smallest value.***
Maximum	***The biggest value.***
Std Deviation	***An expression of how widely spread the values are around the mean.***
Class Width	***Difference between the two boundaries of a class.***
Number of Classes	***The number of bars (including zero height bars) in the histograms.***
Skewness	***Skewness is the lack of symmetry, a data set is symmetric if it looks the same to the left and right of the center point***
Kurtosis	***Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution.***

A **pie chart** (or a **circle chart**) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice (and consequently its central angle and area), is proportional to the quantity it represents. While it is named for its resemblance to a pie which has been sliced, there are variations on the way it can be presented. The earliest known pie chart is generally credited to William Playfair's *Statistical Breviary* of 1801.[1][2]

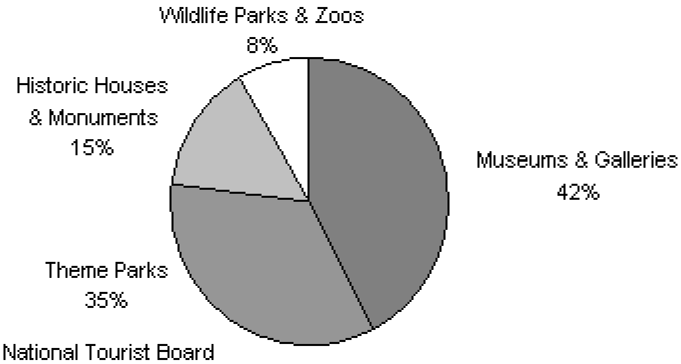
Pie charts are very widely used in the business world and the mass media.[3] However, they have been criticized,[4] and many experts recommend avoiding them,[5][6][7][8] pointing out that research has shown it is difficult to compare different sections of a given pie chart, or to compare data across different pie charts. Pie charts can be replaced in most cases by other plots such as the bar chart, box plot or dot plots.

Pie charts

What is a pie chart?

A pie chart is a circular graph that shows the relative contribution that different categories contribute to an overall total. A wedge of the circle represents each category's contribution, such that the graph resembles a pie that has been cut into different sized slices. Every 1% contribution that a category contributes to the total corresponds to a slice with an angle of 3.6 degrees.

Distribution of visitors by types of tourist attractions, Britain 1991



Source: National Tourist Board

What data can be presented using a pie chart?

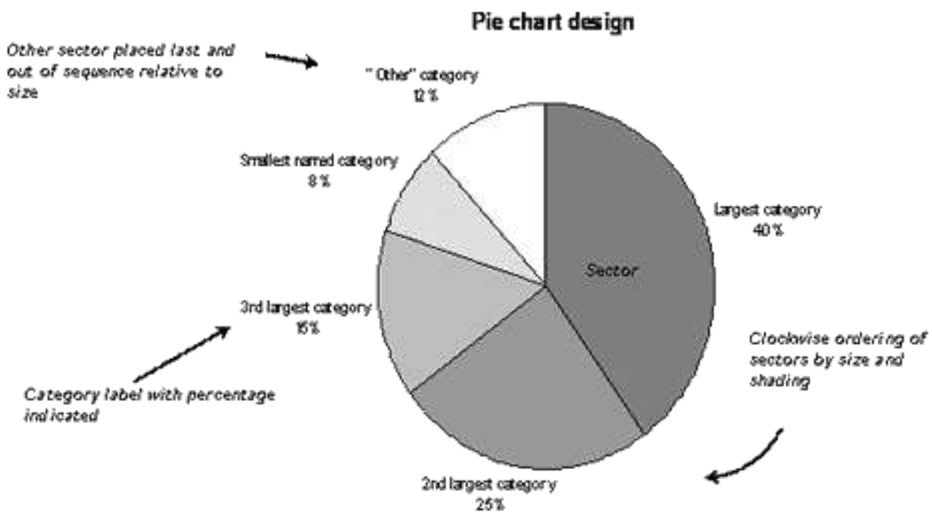
- Pie charts are a visual way of displaying data that might otherwise be given in a small table.
- Pie charts are useful for displaying data that are classified into nominal or ordinal categories. Nominal data are categorised according to descriptive or qualitative information such as county of birth or type of pet owned. Ordinal data are similar but the different categories can also be ranked, for example in a survey people may be asked to say whether they classed something as very poor, poor, fair, good, very good.
- Pie charts are generally used to show percentage or proportional data and usually the percentage represented by each category is provided next to the corresponding slice of pie.
- Pie charts are good for displaying data for around 6 categories or fewer. When there are more categories it is difficult for the eye to distinguish between the relative sizes of the different sectors and so the chart becomes difficult to interpret.

Design issues

Pie charts provide a good visual representation of the data when the categories show some variation in size. When there are several similar-sized categories, a pie chart can end up looking cluttered and it may be difficult to interpret the data. In such cases consider whether a table would present the information more effectively.

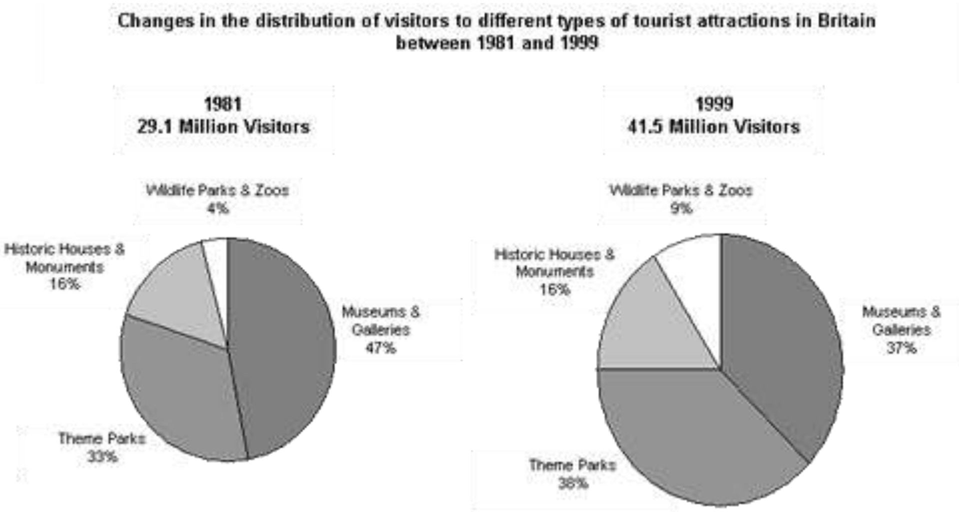
It is usual for the different sectors of the pie chart to be arranged clockwise in order of magnitude. If there is a slice that does not contain a unique category of data but summarises several, for example "other types" or "other answers", then even if it is not the smallest category it is usual to display it last in order that it does not detract from the named categories of interest.

It is helpful to colour or shade the different slices so that they grade from dark to light tones as you move from the first to the last slice.



Comparing pie charts

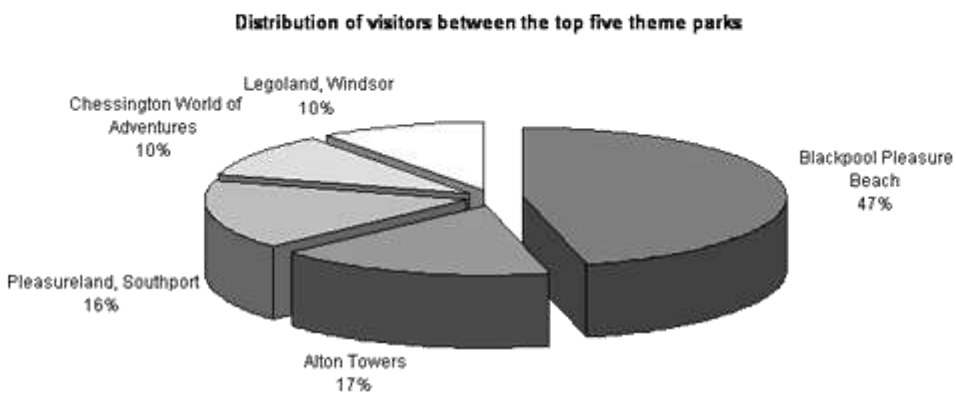
Two or more pie charts can be used to compare two sets of data where the categories are the same or similar but there is a change in another variable, such as time or age. In such cases it is helpful to maintain the same ordering and colouring of slices in the second pie chart as the first in order to facilitate comparison. This may mean that in the second pie chart the slices are no longer presented in order of magnitude. If the two charts represent different sized totals show this by making the areas of the pie chart proportional to the totals they each represent.



3D effects and exploding pie charts

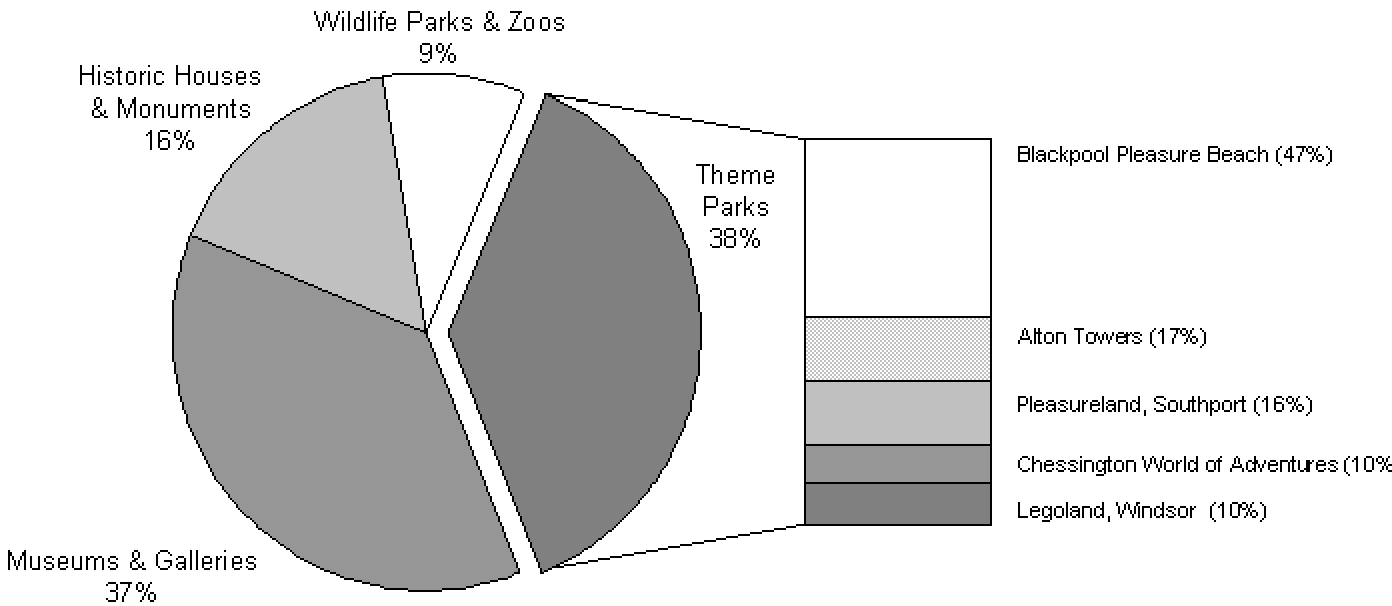
Avoid design elements that detract from the message you are trying to convey such as 3D effects and exploding slices. These can produce optical effects that make it hard to compare different categories.

For example, in the 3D- exploding pie chart below, the slice representing Alton Towers appears quite a bit larger than that for Pleasureland although there is actually only a 1% difference between the two categories. Also, both slices for Chessington World and Legoland represent 10% but again they appear different sizes due to the 3D perspective. In this example separating (exploding) the different wedges of the pie chart also adds to the difficulty of interpreting the data and estimating the relative size of different sectors.



There are however occasions when the use of an exploding pie chart can enhance the presentation of the data. For example if you wish to highlight information in one category/wedge in particular, or when additional information is provided about a particular category as is shown in the example below.

Distribution of visitors to different types of tourist attractions in Britain, 1999



Where next?

This guide has outlined the various ways in which pie charts can be used to present data and has also provided design and presentation advice. Information about other graph and chart types and any specific design issues related to them can be found in the companion study guides: Bar charts and Histograms. The study guide Presenting numerical data provides guidance on when to use graphs to present information and compares the uses of different graph and chart types.

<u>Measures of Variability</u>

Measures of variability indicate the degree to which the scores in a distribution are spread out. Larger numbers indicate greater variability of scores. Sometimes the word *dispersion* is substituted for variability, and you will find that term used in some statistics texts. We will divide our discussion of measures of variability into four categories: range measures, the average deviation, the variance, and the standard deviation.

Range Measures

The **range** is the distance from the lowest score to the highest score. We noted that the range is very unstable, because it depends on only two scores. If one of those scores moves further from the distribution, the range will increase even though the typical variability among the scores has changed very little.

This instability of the range has lead to the development of two other range measures, neither of which rely on only the lowest and highest scores. The **interquartile range** is the distance from the 25th percentile and the 75 percentile. The 25th percentile is also called the **first quartile**, which means that it divides the first quarter of the distribution from the rest of the distribution. The 75th percentile is also called the **third quartile**, because it divides the lowest three quarters of the distribution from the rest of the distribution. Typically, the quartiles are indicated by uppercase *Q*s, with the subscript indicating which quartile we are talking about (Q_1 is the first quartile and Q_3 is the third quartile). So the interquartile range can be computed by subtracting Q_1 from Q_3 [i.e., $Q_3 - Q_1$]. There is a variation on the interquartile range, called the **semi-interquartile range** or **quartile deviation**. This value is equal to half of the interquartile range.

Using this notation, the median is the second quartile (the 50th percentile). That means that we can use a variation of the formula for the median to compute both the first and third quartiles. Looking at the equation below for the median, we would make the following changes to compute these quartiles.

- To compute Q_1, n_{median} becomes n_{Q1}, which is equal to .25*N. We then identify the interval that contains the n_{Q1} score. All of the other values are obtained in the same way as for the median.
- To compute Q_3, n_{median} becomes n_{Q3}, which is equal to .75*N. We then identify the interval that contains the n_{Q3} score. All of the other values are obtained in the same way as for the median.
- To compute the interquartile range, subtract Q_1 from Q_3.
- To compute the quartile deviation, divide the interquartile range by 2.

$$Median = LRL + i\left[\frac{(n_{median} - n_{LRL})}{f_i}\right]$$

It is common to report the range, and many computer programs routinely provide the minimum score, maximum score, and the range as part of their descriptive statistics package. Nevertheless, these are not widely used measures of variability. The same computer programs that give a range, will also provide both a standard deviation and variance. We will be discussing these measures of variability shortly, after we have introduced the concept of the average deviation.

The Average Deviation

The average deviation is not a measure of variability that anyone uses, but it provides an understandable introduction to the variance. The variance is not an intuitive statistic, but it is very useful in other statistical procedures. In contrast, the average deviation is intuitive, although pretty much worthless for other statistical procedures. So we will use the average deviation to introduce the concept of the variance.

The **average deviation** is, as the name implies, the average deviation (distance) from the mean. To compute it, you start by computing the mean, then you subtract the mean from each score, ignoring the sign of the difference, and sum those differences. You then divide by the number of scores (*N*). The formula is shown below. The vertical lines on either side of the numerator indicate that you should take the absolute value, which converts all the differences to positive quantities. Therefore, you are computing deviations (distances) from the mean.

$$Average-Deviation = \frac{\sum |X - \overline{X}|}{N}$$

Chapter 5 in the textbook walked you through the computation of the average deviation. The reason we take the absolute value of these distances from the mean is that the sum of the differences from the mean, some positive and some negative, will always equal zero. We can prove that fact with a little algebra, but you can take our word for it.

As we mentioned earlier, the average deviation is easy to understand, but it has little value for inferential statistics. In contrast, the next two measures (variance and standard deviation) are useful in other statistical procedures. So we now turn our attention to them.

The Variance

The variance takes a different approach to making all of the distances from the mean positive so that they will not sum to zero. Instead of taking the absolute value of the difference from the mean, the variance squares all of those differences. The notation that is used for the variance is a lowercase s^2. The formula for the variance is shown below. If you compare it with the formula for average deviation, you will see two differences instead of one between these formulas. The first is that the differences are squared instead of taking the absolute value. The numerator of this formula is called the **sum of squares**, which is short for sum of squared differences from the mean. See if you can spot the second difference.

$$s^2 = \frac{\sum (X - \overline{X})^2}{N-1}$$

Did you recognize that the variance formula does not divide by *N*, but instead divides by *N-1*? The denominator (*N-1*) in this equation is called the **degrees of freedom**. It is a concept that you will hear about again and again in statistics. If you would like to know more about degrees of freedom, you can click on this link. This link provides a conceptual explanation of this concept.

The reason that the variance formula divides the sum of squared differences from the mean by *N-1* is that dividing by *N* would produce a biased estimate of the population variance, and that bias is removed by dividing by *N-1*. You can learn more about the concept of biased versus unbiased estimates of population parameters by clicking on this link.

The Standard Deviation

The variance has some excellent statistical properties, but it is hard for most students to conceptualize. To start with, the unit of measurement for the mean is the same as the unit of measurement for the score. For example, if we compute the mean age of our sample and find that it is 28.7 years, that mean is on the same scale as the individual ages of our participants. But the variance is in squared units. For example, we might find that the variance is 100 years^2. Can you even imagine what the unit of years squared represents? Most people can't. But there is a measure of variability that is in the same units as the mean. It is called the **standard deviation**, and it is the square root of the variances (see the formula below). So if the variance was 100 years^2, the standard deviation would be 10 years. Since we used the symbol s^2 to indicate variance, you might not be surprised that we use the lowercase letter *s* to indicate the standard deviation. You will see in our discussion of relative scores how valuable the standard deviation can be.

$$s = \sqrt{s^2}$$

The Range

The range is the most obvious measure of dispersion and is the difference between the lowest and highest values in a dataset. In figure 1, the size of the largest semester 1 tutorial group is 6 students and the size of the smallest group is 4 students, resulting in a range of 2 (6-4). In semester 2, the largest tutorial group size is 7 students and the smallest tutorial group contains 3 students, therefore the range is 4 (7-3).

- The range is simple to compute and is useful when you wish to evaluate the whole of a dataset.
- The range is useful for showing the spread within a dataset and for comparing the spread between similar datasets.

An example of the use of the range to compare spread within datasets is provided in table 1. The scores of individual students in the examination and coursework component of a module are shown.

Table 1: Comparison of coursework and examination marks for 14 students

Student	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Coursework mark	27	44	39	23	41	48	37	34	40	43	30	43	29	27
Examination mark	12	47	26	25	38	45	35	35	41	39	32	25	18	30

To find the range in marks the highest and lowest values need to be found from the table. The highest coursework mark was 48 and the lowest was 27 giving a range of 21. In the examination, the highest mark was 45 and the lowest 12 producing a range of 33. This indicates that there was wider variation in the students' performance in the examination than in the coursework for this module.

Since the range is based solely on the two most extreme values within the dataset, if one of these is either exceptionally high or low (sometimes referred to as outlier) it will result in a range that is not typical of the variability within the dataset. For example, imagine in the above example that one student failed to hand in any coursework and was awarded a mark of zero, however they sat the exam and scored 40. The range for the coursework marks would now become 48 (48-0), rather than 21, however the new range is not typical of the dataset as a whole and is distorted by the outlier in the coursework marks. In order to reduce the problems caused by outliers in a dataset, the inter-quartile range is often calculated instead of the range.

The Inter-quartile Range

The inter-quartile range is a measure that indicates the extent to which the central 50% of values within the dataset are dispersed. It is based upon, and related to, the median.

In the same way that the median divides a dataset into two halves, it can be further divided into quarters by identifying the upper and lower quartiles. The lower quartile is found one quarter of the way along a dataset when the values have been arranged in order of magnitude; the upper quartile is found three quarters along the dataset. Therefore, the upper quartile lies half way between the median and the highest value in the dataset whilst the lower quartile lies halfway between the median and the lowest value in the dataset. The inter-quartile range is found by subtracting the lower quartile from the upper quartile.

For example, the examination marks for 20 students following a particular module are arranged in order of magnitude.

Lower quartile — Median — Upper quartile

Student	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Mark	43	48	50	50	52	53	56	58	59	60	62	65	66	68	70	71	74	76	78	80

The median lies at the mid-point between the two central values (10th and 11th)

= half-way between 60 and 62 = 61

The lower quartile lies at the mid-point between the 5th and 6th values

= half-way between 52 and 53 = 52.5

The upper quartile lies at the mid-point between the 15th and 16th values

= half-way between 70 and 71 = 70.5

The inter-quartile range for this dataset is therefore **70.5 - 52.5 = 18** whereas the range is: **80 - 43 = 37.**

The inter-quartile range provides a clearer picture of the overall dataset by removing/ignoring the outlying values.

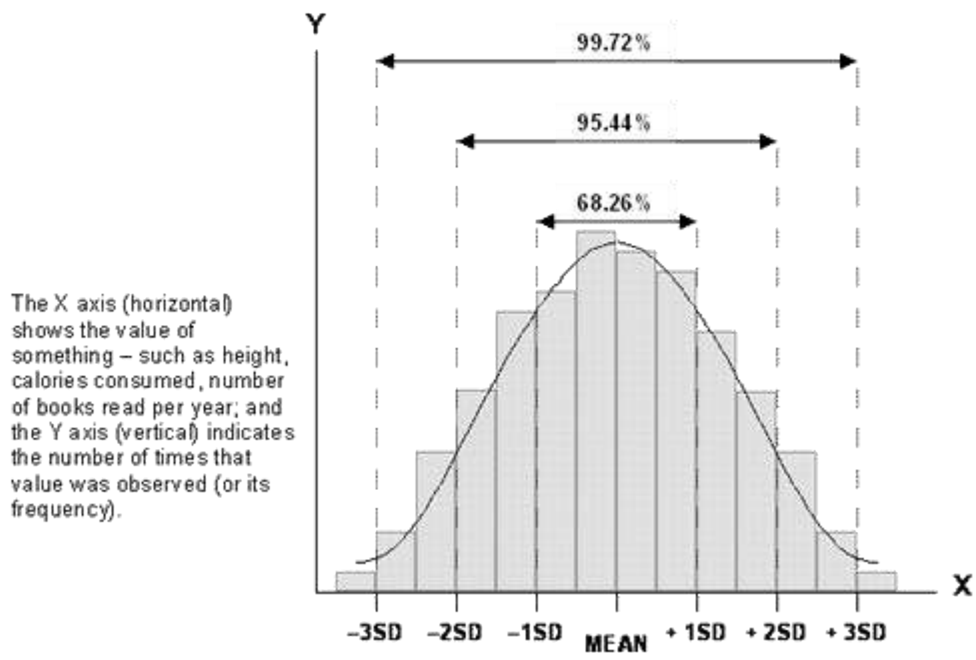
Like the range however, the inter-quartile range is a measure of dispersion that is based upon only two values from the dataset. Statistically, the standard deviation is a more powerful measure of dispersion because it takes into account every value in the dataset. The standard deviation is explored in the next section of this guide.

The Standard Deviation

The standard deviation is a measure that summarises the amount by which every value within a dataset varies from the mean. Effectively it indicates how tightly the values in the dataset are bunched around the mean value. It is the most robust and widely used measure of dispersion since, unlike the range and inter-quartile range, it takes into account every variable in the dataset. When the values in a dataset are pretty tightly bunched together the standard deviation is small. When the values are spread apart the standard deviation will be relatively large. The standard deviation is usually presented in conjunction with the mean and is measured in the same units.

In many datasets the values deviate from the mean value due to chance and such datasets are said to display a normal distribution. In a dataset with a normal distribution most of the values are clustered around the mean while relatively few values tend to be extremely high or extremely low. Many natural phenomena display a normal distribution.

For datasets that have a normal distribution the standard deviation can be used to determine the proportion of values that lie within a particular range of the mean value. For such distributions it is always the case that 68% of values are less than one standard deviation (1SD) away from the mean value, that 95% of values are less than two standard deviations (2SD) away from the mean and that 99% of values are less than three standard deviations (3SD) away from the mean. Figure 3 shows this concept in diagrammatical form.



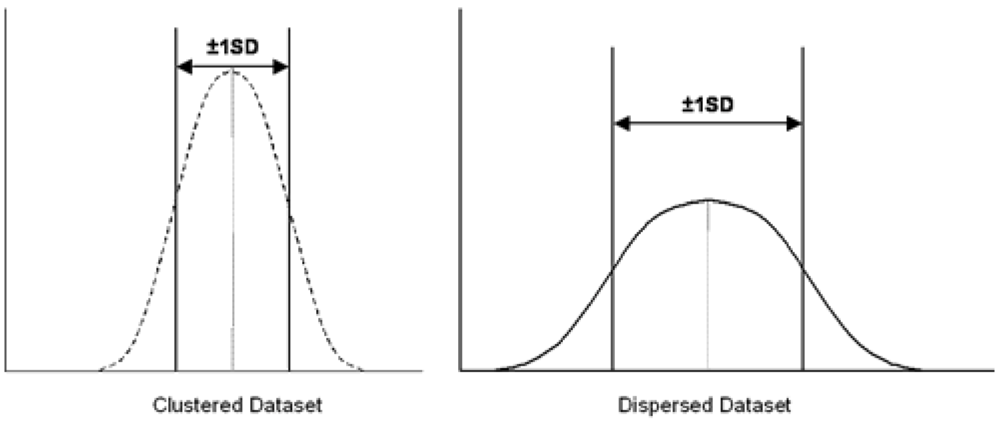
If the mean of a dataset is 25 and its standard deviation is 1.6, then

- 68% of the values in the dataset will lie between **MEAN-1SD** (25-1.6=23.4) and **MEAN+1SD** (25+1.6=26.6)

- 99% of the values will lie between **MEAN-3SD** (25-4.8=20.2) and **MEAN+3SD** (25+4.8=29.8).

If the dataset had the same mean of 25 but a larger standard deviation (for example, 2.3) it would indicate that the values were more dispersed. The frequency distribution for a dispersed dataset would still show a normal distribution but when plotted on a graph the shape of the curve will be flatter as in figure 4.

Example 4: Typical frequency distributions for clustered and dispersed datasets



Population and sample standard deviations

There are two different calculations for the Standard Deviation. Which formula you use depends upon whether the values in your dataset represent an entire population or whether they form a sample of a larger population. For example, if all student users of the library were asked how many books they had borrowed in the past month then the entire population has been studied since all the students have been asked. In such cases the population standard deviation should be used. Sometimes it is not possible to find information about an entire population and it might be more realistic to ask a sample of 150 students about their library borrowing and use these results to estimate library borrowing habits for the entire population of students. In such cases the sample standard deviation should be used.

Formulae for the standard deviation

Whilst it is not necessary to learn the formula for calculating the standard deviation, there may be times when you wish to include it in a report or dissertation.

The standard deviation of an entire population is known as **σ** (sigma) and is calculated using:

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

Where **x** represents each value in the population, **μ** is the mean value of the population, **Σ** is the summation (or total), and **N** is the number of values in the population.

The standard deviation of a sample is known as **S** and is calculated using:

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

Where **x** represents each value in the population, **x** is the mean value of the sample, **Σ** is the summation (or total), and **n-1** is the number of values in the sample minus 1.

Normal Probability Curve

The normal distribution is the most important and most widely used distribution in statistics. It is sometimes called the "bell curve," although the tonal qualities of such a bell would be less than pleasing. It is also called the "Gaussian curve" after the mathematician Karl Friedrich Gauss. As you will see in the section on the history of the normal distribution, although Gauss played an important role in its history, Abraham de Moivre first discovered the normal distribution.

Strictly speaking, it is not correct to talk about "**the** normal distribution" since there are many normal distributions. Normal distributions can differ in their means and in their standard deviations. Figure 1 shows three normal distributions. The green (left-most) distribution has a mean of -3 and a standard deviation of 0.5, the distribution in red (the middle distribution) has a mean of 0 and a standard deviation of 1, and the distribution in black (right-most) has a mean of 2 and a standard deviation of 3. These as well as all other normal distributions are symmetric with relatively more values at the center of the distribution and relatively few in the tails.

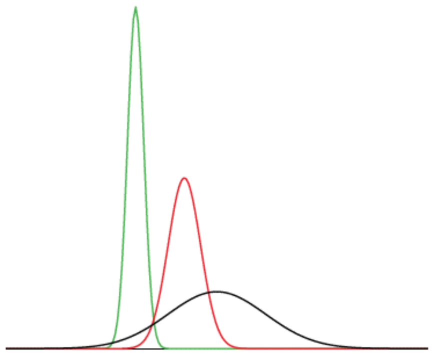


Figure 1. Normal distributions differing in mean and standard deviation.

The density of the normal distribution (the height for a given value on the x axis) is shown below. The parameters μ and σ are the mean and standard deviation, respectively, and define the normal distribution. The symbol *e* is the base of the natural logarithm and π is the constant pi.

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

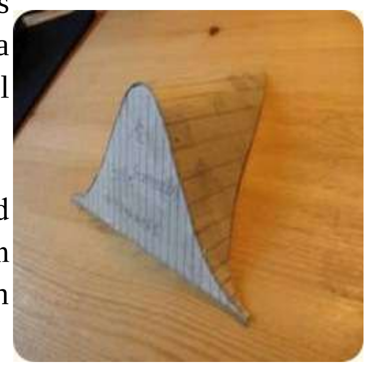
Since this is a non-mathematical treatment of statistics, do not worry if this expression confuses you. We will **not** be referring back to it in later sections.

Seven features of normal distributions are listed below. These features are illustrated in more detail in the remaining sections of this chapter.

1. Normal distributions are symmetric around their mean.
2. The mean, median, and mode of a normal distribution are equal.
3. The area under the normal curve is equal to 1.0.
4. Normal distributions are denser in the center and less dense in the tails.
5. Normal distributions are defined by two parameters, the mean (μ) and the standard deviation (σ).
6. 68% of the area of a normal distribution is within one standard deviation of the mean.
7. Approximately 95% of the area of a normal distribution is within two standard deviations of the mean.

<u>Most important Characteristics of the Normal distribution :</u>

Actually we can say that Normal distribution is the most widely known and used of all distributions. Because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference for many probability problems So Normal distribution characteristics is :



• Symmetric & bell shaped • Continuous for all values of X between -∞ and ∞ so that each conceivable interval of real numbers has a probability other than zero. • -∞ ≤ X ≤ ∞ • Two parameters, μ and σ. Note that the normal distribution is actually a family of distributions, since μ and σ determine the shape of the distribution. • The rule for a normal density function is e21 =)

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

where μ is the location parameters and σ is the scale parameters. The case where μ= 0 and σ= 1 is called the **standard normal distribution**. The equation for the standard normal distribution is

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

• The notation N(μ, σ2) means normally distributed with mean μ and variance σ2. If we say X ~ N(μ, σ2) we mean that X is distributed N(μ, σ2). • About 2/3 of all cases fall within one standard deviation of the mean, that is P(μ - σ ≤ X ≤ μ + σ) = .6826. • About 95% of cases lie within 2 standard deviations of the mean, that is P(μ - 2σ ≤ X ≤ μ + 2σ) = .9544

PROPERTIES OF NORMAL DISTRIBUTION

A normal distribution with parameters μ and σ has the following properties.
1. The curve is Bell –shaped
a. It is symmetrical (Non-skew).
That is β1 = 0
b. The mean, media and mode are equal

2. The curve is asymptotic to the X-axis. That is, the curve touches the X-axis only at $-\infty$ and $+\infty$.

3. The curve has points of inflexion at $\mu - \sigma$ and $\mu + \sigma$.

4. For the distribution ….

a. Standard deviation = σ

b. Quartile deviation = 2/3 σ (approximately)

c. Mean deviation = 4/5 σ (approximately)

5. For the distribution –

a. The odd order moments are equal to zero.

b. The even order moments are given by –

Thus, $\mu_2 = \sigma^2$ and $\mu_4 = 3\sigma^4$

6. The distribution is mesokurtic. That is, $\beta_2 = 3$.

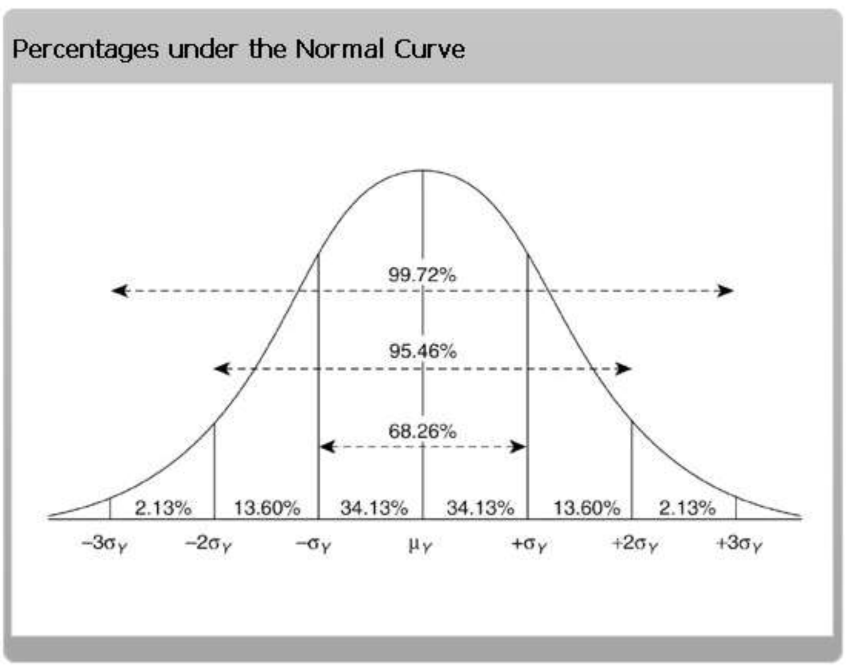
7. Total area under the curve is unity.

$P[a < X \leq b]$ = Area bounded by the curve and the ordinates at a and b

a. $P[\mu - \sigma < X \leq \mu + \sigma] = 0.6826 = 68.26\%$

b. $P[\mu - 2\sigma < X \leq \mu + 2\sigma] = 0.9544 = 95.44\%$

c. $P[\mu - 3\sigma < X \leq \mu + 3\sigma] = 0.9974 = 99.74\%$



<u>**Spearman's rank correlation coefficient**</u>

Spearman's rank correlation coefficient or **Spearman's rho**, named after Charles Spearman and often denoted by the Greek letter ρ(rho) or as r_s, is a nonparametric measure of statistical dependence between two variables. It assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

Spearman's coefficient, like any correlation calculation, is appropriate for both continuous and discrete variables, including ordinal variables.[1][2] Spearman's $\rho$and Kendall's τcan be formulated as special cases of a more general correlation coefficient.

Definition and calculation

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables.[3] For a sample of size *n*, the *n* raw scores $X_i, Y_i$are converted to ranks x_i, y_i, and ρ is computed from:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

where $d_i = x_i - y_i$, is the difference between ranks. See the example below. Identical values (rank ties or value duplicates) are assigned a rank equal to the average of their positions in the ascending order of the values. In the table below, notice how the rank of values that are the same is the mean of what their ranks would otherwise be:

Variable X_i	**Position in the ascending order**	**Rank x_i**
0.8	1	1
1.2	2	$\frac{2+3}{2} = 2.5$
1.2	3	$\frac{2+3}{2} = 2.5$
2.3	4	4
18	5	5

In applications where duplicate values are known to be absent, a simpler procedure can be used to calculate ρ.

This method should not be used in cases where the data set is truncated; that is, when the Spearman correlation coefficient is desired for the top X records (whether by pre-change rank or post-change rank, or both), the user should use the Pearson correlation coefficient formula given above.

The standard error of the coefficient (σ) was determined by Pearson in 1907 and Gosset in 1920. It is

$$\sigma = \frac{0.6325}{\sqrt{n-1}}$$

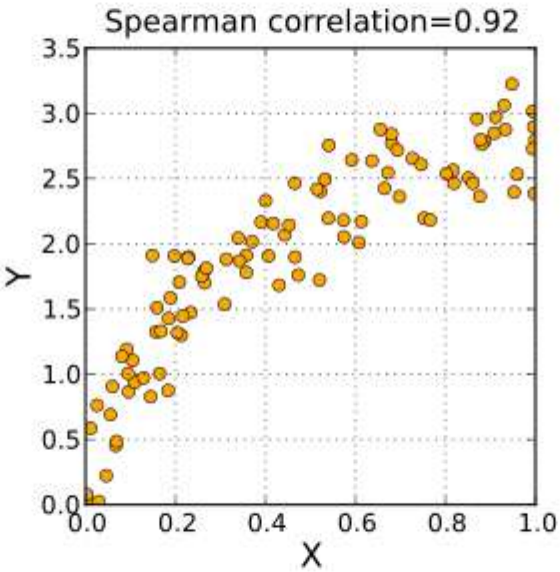
Related quantities

There are several other numerical measures that quantify the extent of statistical dependence between pairs of observations. The most common of these is the Pearson product-moment correlation coefficient, which is a similar correlation method to Spearman's rank, that measures the "linear" relationships between the raw numbers rather than between their ranks.

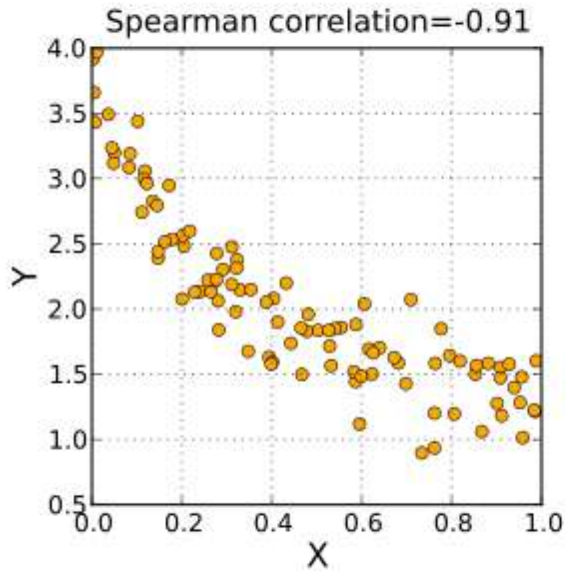
An alternative name for the Spearman rank correlation is the "grade correlation";[5] in this, the "rank" of an observation is replaced by the "grade". In continuous distributions, the grade of an observation is, by convention, always one half less than the rank, and hence the grade and rank correlations are the same in this case. More generally, the "grade" of an observation is proportional to an estimate of the fraction of a population less than a given value, with the half-observation adjustment at observed values. Thus this corresponds to one possible treatment of tied ranks. While unusual, the term "grade correlation" is still in use.[6]

Interpretation

Positive and negative Spearman rank correlations



A positive Spearman correlation coefficient corresponds to an increasing monotonic trend between *X* and *Y*.



A negative Spearman correlation coefficient corresponds to a decreasing monotonic trend between *X* and *Y*.

The sign of the Spearman correlation indicates the direction of association between *X* (the independent variable) and *Y* (the dependent variable). If *Y* tends to increase when *X* increases, the Spearman correlation coefficient is positive. If *Y* tends to decrease when *X* increases, the Spearman correlation coefficient is negative. A Spearman correlation of zero indicates that there is no tendency for *Y* to either increase or decrease when *X* increases. The Spearman correlation increases in magnitude as *X* and *Y* become closer to being perfect monotone functions of each other. When *X* and *Y* are perfectly monotonically related, the Spearman correlation coefficient becomes 1. A perfect monotone increasing relationship implies that for any two pairs of data values X_i, Y_i and X_j, Y_j, that $X_i - X_j$ and $Y_i - Y_j$ always have the same sign. A perfect monotone decreasing relationship implies that these differences always have opposite signs.

The Spearman correlation coefficient is often described as being "nonparametric". This can have two meanings. First, the fact that a perfect Spearman correlation results when *X* and *Y* are related by any monotonic function can be contrasted with the Pearson correlation, which only gives a perfect value when *X* and *Y* are related by a linear function. The other sense in which the Spearman correlation is nonparametric in that its exact sampling distribution can be obtained without requiring knowledge (*i.e.*, knowing the parameters) of the joint probability distribution of *X* and *Y*.

Example

In this example, the raw data in the table below is used to calculate the correlation between the IQ of a person with the number of hours spent in front of TV per week.

IQ, X_i	Hours of TV per week, Y_i
106	7
86	0
100	27
101	50
99	28
103	29
97	20
113	12
112	6
110	17

Firstly, evaluate d_i^2. To do so use the following steps, reflected in the table below.

1. Sort the data by the first column (X_i). Create a new column $x_i$and assign it the ranked values 1,2,3,...*n*.
2. Next, sort the data by the second column (Y_i). Create a fourth column $y_i$and similarly assign it the ranked values 1,2,3,...*n*.
3. Create a fifth column $d_i$to hold the differences between the two rank columns (x_i and y_i).
4. Create one final column d_i^2 to hold the value of column $d_i$squared.

IQ, X_i	Hours of TV per week, Y_i	rank x_i	rank y_i	d_i	d_i^2
86	0	1	1	0	0
97	20	2	6	−4	16
99	28	3	8	−5	25

100	27	4	7	−3	9
101	50	5	10	−5	25
103	29	6	9	−3	9
106	7	7	3	4	16
110	17	8	5	3	9
112	6	9	2	7	49
113	12	10	4	6	36

With d_i^2 found, add them to find $\sum d_i^2 = 194$. The value of *n* is 10. These values can now be substituted back into the equation : $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$ to give

$$\rho = 1 - \frac{6 \times 194}{10(10^2 - 1)}$$

which evaluates to ρ = -29/165 = −0.175757575... with a P-value = 0.627188 (using the t distribution)

This low value shows that the correlation between IQ and hours spent watching TV is very low, although the negative value suggests that the longer the time spent watching television the lower the IQ. In the case of ties in the original values, this formula should not be used; instead, the Pearson correlation coefficient should be calculated on the ranks (where ties are given ranks, as described above).

Determining significance

One approach to test whether an observed value of ρ is significantly different from zero (r will always maintain $-1 \le r \le 1$) is to calculate the probability that it would be greater than or equal to the observed r, given the null hypothesis, by using a permutation test. An advantage of this approach is that it automatically takes into account the number of tied data values there are in the sample, and the way they are treated in computing the rank correlation.

Another approach parallels the use of the Fisher transformation in the case of the Pearson product-moment correlation coefficient. That is, confidence intervals and hypothesis tests relating to the population value ρ can be carried out using the Fisher transformation:

$$F(r) = \frac{1}{2}\ln\frac{1+r}{1-r} = \operatorname{artanh}(r).$$

If $F(r)$ is the Fisher transformation of r, the sample Spearman rank correlation coefficient, and n is the sample size, then

$$z = \sqrt{\frac{n-3}{1.06}}F(r)$$

is a z-score for r which approximately follows a standard normal distribution under the null hypothesis of statistical independence ($\rho = 0$).

One can also test for significance using

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

which is distributed approximately as Student's t distribution with $n-2$ degrees of freedom under the null hypothesis. A justification for this result relies on a permutation argument.[10]

pv rank] is a very recent R package that computes rank correlations and their p-values with various options for tied ranks. It is possible to compute exact Spearman coefficient test p-values for $n \le 26$.

A generalization of the Spearman coefficient is useful in the situation where there are three or more conditions, a number of subjects are all observed in each of them, and it is predicted that the observations will have a particular order. For example, a number of subjects might each be given three trials at the same task, and it is predicted that performance will improve from trial to trial. A test of the significance of the trend between conditions in this situation was developed by E. B. Page and is usually referred to as Page's trend test for ordered alternatives.

Percentile

A **percentile** (or a centile) is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. For example, the 20th percentile is the value (or score) below which 20 percent of the observations may be found.

The term percentile and the related term percentile rank are often used in the reporting of scores from norm-referenced tests. For example, if a score is in the 86th percentile, it is higher than 86% of the other scores. The 25th percentile is also known as the first quartile (Q_1), the 50th percentile as the median or second quartile (Q_2), and the 75th percentile as the third quartile (Q_3). In general, percentiles and quartiles are specific types of quantiles.

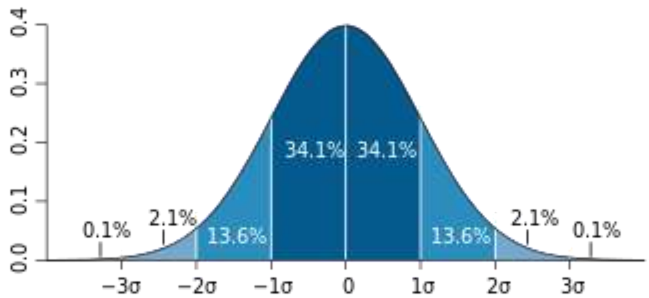
Applications

When ISPs bill "burstable" internet bandwidth, the 95th or 98th percentile usually cuts off the top 5% or 2% of bandwidth peaks in each month, and then bills at the nearest rate. In this way infrequent peaks are ignored, and the customer is charged in a fairer way. The reason this statistic is so useful in measuring data throughput is that it gives a very accurate picture of the cost of the bandwidth. The 95th percentile says that 95% of the time, the usage is below this amount. Just the same, the remaining 5% of the time, the usage is above that amount.

Physicians will often use infant and children's weight and height to assess their growth in comparison to national averages and percentiles which are found in growth charts.

The 85th percentile speed of traffic on a road is often used as a guideline in setting speed limits and assessing whether such a limit is too high or low.

The normal distribution and percentiles



Representation of the three-sigma rule. The dark blue zone represents observations within one standard deviation (σ) to either side of the mean (μ), which accounts for about 68.2% of the population. Two standard deviations from the mean (dark and medium blue) account for about 95.4%, and three standard deviations (dark, medium, and light blue) for about 99.7%.

The methods given in the Definitions section are approximations for use in small-sample statistics. In general terms, for very large populations following a normal distribution percentiles may often be represented by reference to a normal curve plot. The normal distribution is plotted along an axis scaled to standard deviations, or sigma units. Mathematically, the normal distribution extends to negative infinity on the left and positive infinity on the right. Note, however, that only a very small proportion of individuals in a population will fall outside the −3 to +3 range. For example, with human heights very few people are above the +3 sigma height level.

Percentiles represent the area under the normal curve, increasing from left to right. Each standard deviation represents a fixed percentile. Thus, rounding to two decimal places, $-3\ \sigma$ is the 0.13th percentile, $-2\ \sigma$ the 2.28th percentile, $-1\ \sigma$ the 15.87th percentile, 0 the 50th percentile (both the mean and median of the distribution), +1 σ the 84.13th percentile, +2 σ the 97.72nd percentile, and +3 σ the 99.87th percentile. This is known as the 68–95–99.7 rule or the three-sigma rule. Note that in theory the 0th percentile falls at negative infinity and the 100th percentile at positive infinity, although in many practical applications, such as test results, natural lower and/or upper limits are enforced.

Definitions

There is no standard definition of percentile, however all definitions yield similar results when the number of observations is very large In the limit, as the sample size approaches infinity and the data points become so densely spaced they appear continuous, the $100p^{th}$ percentile ($0<p<1$) approximates the inverse of the cumulative distribution function (CDF) thus formed, evaluated at *p*, as *p* approximates the CDF. Some methods for calculating the percentiles are given below.

The Nearest Rank method

One definition of percentile, often given in texts, is that the *P*-th percentile $(0 < P \leq 100)$ of a list of *N* ordered values (sorted from least to greatest) is the smallest value in the list such that P percent of the data is less than or equal to that value. This is obtained by first calculating the ordinal rank and then taking the value from the ordered list that corresponds to that rank. The ordinal rank *n* is calculated using this formula

$$n = \left\lceil \frac{P}{100} \times N \right\rceil$$

Note the following:

- Using the Nearest Rank method on lists with fewer than 100 distinct values can result in the same value being used for more than one percentile.
- A percentile calculated using the Nearest Rank method will always be a member of the original ordered list.
- The 100th percentile is defined to be the largest value in the ordered list.

Worked examples of the Nearest Rank method

Example 1:

Consider the ordered list {15, 20, 35, 40, 50}, which contains five data values. What are the 30th, 40th, 50th and 100th percentiles of this list using the Nearest Rank method?

Percentile P	Number in list	Ordinal rank n	Number from the ordered list	Percentile value	Notes

	N		**list that has that rank**		
30th	5	$n = \left\lceil \frac{30}{100} \times 5 \right\rceil = \lceil 1.5 \rceil = 2.$	the second number in the ordered list, which is 20	20	20 is an element of the list
40th	5	$n = \left\lceil \frac{40}{100} \times 5 \right\rceil = \lceil 2.0 \rceil = 2.$	the second number in the ordered list, which is 20	20	In this example it is the same as the 30th percentile.
50th	5	$n = \left\lceil \frac{50}{100} \times 5 \right\rceil = \lceil 2.5 \rceil = 3.$	the third number in the ordered list, which is 35	35	35 is an element of the ordered list.
100th	5	Last	50, which is the last number in the ordered list	50	The 100th percentile is defined to be the largest value in the list, which is 50.

So the 30th, 40th, 50th and 100th percentiles of the ordered list {15, 20, 35, 40, 50} using the Nearest Rank method are {20, 20, 35, 50}

Example 2:

Consider an ordered population of 10 data values {3, 6, 7, 8, 8, 10, 13, 15, 16, 20}. What are the 25th, 50th, 75th and 100th percentiles of this list using the Nearest Rank method?

Percentile	**Number**	**Ordinal rank**	**Number**	**Percentile**	**Notes**

P	**in N**	**list n**	**from the value ordered list that has that rank**	
25th	10	$n = \left\lceil \frac{25}{100} \times 10 \right\rceil = \lceil 2.5 \rceil = 3.$	the third number in the ordered list, which is 7	7	7 is an element of the list
50th	10	$n = \left\lceil \frac{50}{100} \times 10 \right\rceil = \lceil 5.0 \rceil = 5.$	the fifth number in the ordered list, which is 8	8	8 is an element of the list.
75th	10	$n = \left\lceil \frac{75}{100} \times 10 \right\rceil = \lceil 7.5 \rceil = 8.$	the eighth number in the ordered list, which is 15	15	15 is an element of the list.
100th	10	Last	20, which is the last number in the ordered list	20	The 100th percentile is defined to be the largest value in the list, which is 20.

So the 25th, 50th, 75th and 100th percentiles of the ordered list {3, 6, 7, 8, 8, 10, 13, 15, 16, 20} using the Nearest Rank method are {7, 8, 15, 20}

Example 3:

Consider an ordered population of 11 data values {3, 6, 7, 8, 8, 9, 10, 13, 15, 16, 20}. What are the 25th, 50th, 75th and 100th percentiles of this list using the Nearest Rank method?

Percentile P	Number in list N	Ordinal rank n	Number from the ordered list that has that rank	Percentile value	Notes
25th	11	$n = \left\lceil \frac{25}{100} \times 11 \right\rceil = \lceil 2.75 \rceil = 3.$	the third number in the ordered list, which is 7	7	7 is an element of the list
50th	11	$n = \left\lceil \frac{50}{100} \times 11 \right\rceil = \lceil 5.50 \rceil = 6.$	the sixth number in the ordered list, which is 9	9	9 is an element of the list.
75th	11	$n = \left\lceil \frac{75}{100} \times 11 \right\rceil = \lceil 8.25 \rceil = 9.$	the ninth number in the ordered list, which is 15	15	15 is an element of the list.
100th	11	Last	20, which is the last number in the ordered list	20	The 100th percentile is defined to be the largest value in the list, which is 20.

So the 25th, 50th, 75th and 100th percentiles of the ordered list {3, 6, 7, 8, 8, 9, 10, 13, 15, 16, 20} using the Nearest Rank method are {7, 9, 15, 20}

The Linear Interpolation Between Closest Ranks method

An alternative to rounding used in many applications is to use linear interpolation between adjacent ranks.

Commonalities between the Variants of this Method

All of the following variants have the following in common. Given the order statistics

$$\{v_i, i = 1, 2, ..., N : v_{i+1} \geq v_i, \forall i = 1, 2, ...N - 1\},$$

we seek a linear interpolation function that passes through the points (v_i, i). This is simply accomplished by

$$v(x) = v_{\lfloor x \rfloor} + (x\%1)(v_{\lfloor x \rfloor+1} - v_{\lfloor x \rfloor}), \forall x \in [1, N] : v(i) = v_i, \text{ for } i = 1, 2, ..., N,$$

where $\lfloor x \rfloor$uses the floor function to represent the integral part of positive x, whereas $x\%1$uses the mod function to represent its fractional part (the remainder after division by 1). (Note that, though at the endpoint$x = N$, $v_{\lfloor x \rfloor+1}$is undefined, it does not need to be because it is multiplied by$x\%1 = 0$.) As we can see, $x$is the continuous version of the subscript$i$, linearly interpolating $v$between adjacent nodes.

There are two ways in which the variant approaches differ. The first is in the linear relationship between the *rank*x, the *percent rank* $P = 100p$, and a constant that is a function of the sample size N:

$$x = f(p, N) = (N + c_1)p + c_2.$$

There is the additional requirement that the midpoint of the range $(1, N)$, corresponding to the median, occur at $p = 0.5$:

$$f(0.5, N) = \frac{N + c_1}{2} + c_2 = \frac{N + 1}{2} \therefore 2c_2 + c_1 = 1,$$

and our revised function now has just one degree of freedom, looking like this:

$$x = f(p, N) = (N + 1 - 2C)p + C.$$

The second way in which the variants differ is in the definition of the function near the margins of the $[0, 1]$range of$p$: $f(p, N)$should produce, or be forced to produce, a result in the range $[1, N]$, which may mean the absence of a one-to-one correspondence in the wider region.

First Variant, $C = 1/2$

$$x = f(p) = \begin{cases} Np + \frac{1}{2}, \forall p \in [p_1, p_N], \\ 1, \forall p \in [0, p_1], \\ N, \forall p \in [p_N, 1]. \end{cases} ,$$

where

$$p_i = \frac{1}{N}\left(i - \frac{1}{2}\right), i \in [1, N] \cap \mathbb{N}$$

$$\therefore p_1 = \frac{1}{2N}, p_N = \frac{2N-1}{2N}.$$

Furthermore, let

$$P_i = 100p_i.$$

The inverse relationship is restricted to a narrower region:

$$p = \frac{1}{N}\left(x - \frac{1}{2}\right), x \in (1, N) \cap \mathbb{R}.$$

Worked Example of the First Variant

Consider the ordered list {15, 20, 35, 40, 50}, which contains five data values. What are the 5th, 30th, 40th and 95th percentiles of this list using the Linear Interpolation Between Closest Ranks method? First, we calculate the percent rank for each list value.

List value v_i	**Position of that value in the ordered list** i	**Number of values** N	**Calculation of percent rank**	**Percent rank,** P_i	**Notes**
15	1	5	$\frac{100}{5}\left(1 - \frac{1}{2}\right) = 10.$	10	

20	2	5	$\frac{100}{5}\left(2-\frac{1}{2}\right)=30.$	30
35	3	5	$\frac{100}{5}\left(3-\frac{1}{2}\right)=50.$	50
40	4	5	$\frac{100}{5}\left(4-\frac{1}{2}\right)=70.$	70
50	5	5	$\frac{100}{5}\left(5-\frac{1}{2}\right)=90.$	90

Then we take those percent ranks and calculate the percentile values as follows:

Percent Rank P	**Number of values** N	**Is** $P < P$ **?**	**Is** $P > P$ **?**	**Is there a percent rank equal to** P **?**	**What do we use for percentile value?**	**Percentile value** $v(f(p)$	**Notes**
5	5	Yes	No	No	We see that P=5, which is less than the first percent rank p1=10, so use the first list value v1, which is 15	15	15 is a member of the ordered list
30	5	No	No	Yes	We see that P=30 is the same as the second percent rank p2=30, so use the second list value v2, which is 20	20	20 is a member of the ordered list

40	5	No	No	No	We see that P=40 is between percent rank p2=30 and p3=50, so we take k=2, k+1=3, P=40, pk=p2=30, vk=v2=20, vk+1=v3=35, N=5. Given those values we can then calculate v as follows: $v = 20 + 5 \times \frac{40-30}{100}(35-20)$	27.5	27.5 is not a member of the ordered list
95	5	No	Yes	No	We see that P=95, which is greater than the last percent rank pN=90, so use the last list value, which is 50	50	50 is a member of the ordered list

So the 5th, 30th, 40th and 95th percentiles of the ordered list {15, 20, 35, 40, 50} using the Linear Interpolation between Closest Ranks method are {15, 20, 27.5, 50}

Second Variant, $C = 1$

$$x = f(p, N) = p(N-1) + 1,\ p \in [0, 1]$$

$$\therefore p = \frac{x-1}{N-1},\ x \in [1, N].$$

Note that the $x \leftrightarrow p$ relationship is one-to-one for $p \in [0, 1]$, the only one of the three variants with this property; hence the "INC" suffix, for *inclusive*, on the Excel function.

Worked Examples of the Second Variant

Example 1:

Consider the ordered list {15, 20, 35, 40, 50}, which contains five data values. What is the 40th percentile of this list using this variant method?

First we calculate the rank of the 40th percentile:

$$x = \frac{40}{100}(5-1)+1 = 2.6$$

So, x=2.6, which gives us $\lfloor x \rfloor = 2$and $x\%1 = 0.6$. So, the value of the 40th percentile is

$$v(2.6) = v_2 + 0.6(v_3 - v_2) = 20 + 0.6(35-20) = 29.$$

Example 2:

Consider the ordered list {1,2,3,4} which contains four data values. What is the 75th percentile of this list using the Microsoft Excel method?

First we calculate the rank of the 75th percentile as follows:

$$x = \frac{75}{100}(4-1)+1 = 3.25$$

So, x=3.25, which gives us an integral part of 3 and a fractional part of 0.25. So, the value of the 75th percentile is

$$v(3.25) = v_3 + 0.25(v_4 - v_3) = 3 + 0.25(4-3) = 3.25.$$

Third Variant, $C = 0$

Adopted by Microsoft Excel since 2010 by means of PERCENTIL.EXC function. However, as the "EXC" suffix indicates, the Excel version *excludes* both endpoints of the range of *p*, i.e., $p \in (0,1)$, whereas the "INC" version, the second variant, does not; in fact, any number smaller than 1/(N+1) is also excluded and would cause an error.)

$$x = f(p,N) = \begin{cases} p(N+1), \, p \in \left(0, \frac{N}{N+1}\right] \\ N, \, p \in \left[\frac{N}{N+1}, 1\right] \end{cases} \quad .$$

The inverse is restricted to a narrower region:

$$p = \frac{x}{N+1}, \, x \in (0, N).$$

Worked Example of the Third Variant

Consider the ordered list {15, 20, 35, 40, 50}, which contains five data values. What is the 40th percentile of this list using the NIST method?

First we calculate the rank of the 40th percentile as follows:

$$x = \frac{40}{100}(5+1) = 2.4$$

So *x*=2.4, which gives us $\lfloor x \rfloor = 2$and $x\%1 = 0.4$. So the value of the 40th percentile is calculated as:

$$v(2.4) = v_2 + 0.4(v_3 - v_2) = 20 + 0.4(35 - 20) = 26$$

So the value of the 40th percentile of the ordered list {15, 20, 35, 40, 50} using this variant method is 26.

The Weighted Percentile method

Definition of the Weighted Percentile method

In addition to the percentile function, there is also a *weighted percentile*, where the percentage in the total weight is counted instead of the total number. There is no standard function for a weighted percentile. One method extends the above approach in a natural way.

Suppose we have positive weights $w_1, w_2, w_3, \dots, w_N$associated, respectively, with our *N* sorted sample values. Let

$$S_n = \sum_{k=1}^{n} w_k,$$

the n-th partial sum of the weights. Then the formulas above are generalized by taking

$$p_n = \frac{100}{S_N}\left(S_n - \frac{w_n}{2}\right)$$

and

$$v = v_k + \frac{P - p_k}{p_{k+1} - p_k}(v_{k+1} - v_k).$$

The 50% weighted percentile is known as the weighted median