Business Statistics

Teachers' Guide

Grade 12

(Implemented from 2017)

Depatment of Commerce
Faculty of Science and Technology
National Institute of Education
Maharagama

web : www.nie.lk
Email : info@ nie.lk

Business Statistics

Teachers' Guide

Grade 12

Depatment of Commerce
Faculty of Science and Technology
National Institute of Education
Maharagama

Business Statistics

Grade 12 - Teachers' Guide

First Print 2017

ISBN :

Depatment of Commerce
Faculty of Science and Technology
National Institute of Education

Printed By :

Content

Message from the Director General …

With the primary objective of realizing the National Educational Goals recommended by the National Education Commission, the then prevalent content based curriculum was modernized, and the first phase of the new competency based curriculum was introduced to the eight year curriculum cycle of the primary and secondary education in Sri Lanka in the year 2007

The second phase of the curriculum cycle thus initiated was introduced to the education system in the year 2015 as a result of a curriculum rationalization process based on research findings and various proposals made by stake holders.

Within this rationalization process the concepts of vertical and horizontal integration have been employed in order to build up competencies of students, from foundation level to higher levels, and to avoid repetition of subject content in various subjects respectively and furthermore, to develop a curriculum that is implementable and student friendly.

The new Teachers' Guides have been introduced with the aim of providing the teachers with necessary guidance for planning lessons, engaging students effectively in the learning teaching process, and to make Teachers' Guides will help teachers to be more effective within the classroom. Further, the present Teachers' Guides have given the necessary freedom for the teachers to select quality inputs and activities in order to improve student competencies. Since the Teachers' Guides do not place greater emphasis on the subject content prescribed for the relevant grades, it is very much necessary to use these guides along with the text books compiled by the Educational Publications Department if, Guides are to be made more effective.

The primary objective of this rationalized new curriculum, the new Teachers' Guides, and the new prescribed texts is to transform the student population into a human resource replete with the skills and competencies required for the world of work, through embarking upon a pattern of education which is more student centered and activity based.

I wish to make use of this opportunity to thank and express my appreciation to the members of the Council and the Academic Affairs Board of the NIE the resource persons who contributed to the compiling of these Teachers' Guides and other parties for their dedication in this matter.

Dr. (Mrs.) Jayanthi Gunasekara
Director General
National Institute of Education

Message from the Deputy Director General

Education from the past has been constantly changing and forging forward. In recent years, these changes have become quite rapid. The Past two decades have witnessed a high surge in teaching methodologies as well as in the use of technological tools and in the field of knowledge creation.

Accordingly, the National Institute of Education is in the process of taking appropriate and timely steps with regard to the education reforms of 2015.

It is with immense pleasure that this Teachers' Guide where the new curriculum has been planned based on a thorough study of the changes that have taken place in the global context adopted in terms of local needs based on a student-centered learning-teaching approach, is presented to you teachers who serve as the pilots of the schools system.

An instructional manual of this nature is provided to you with the confidence that, you will be able to make a greater contribution using this.

There is no doubt whatsoever that this Teachers' Guide will provide substantial support in the classroom teaching-learning process at the same time. Furthermore the teacher will have a better control of the classroom with a constructive approach in selecting modern resource materials and following the guide lines given in this book.

I trust that through the careful study of this Teachers Guide provided to you, you will act with commitment in the generation of a greatly creative set of students capable of helping Sri Lanka move socially as well as economically forward.

This Teachers' Guide is the outcome of the expertise and unflagging commitment of a team of subject teachers and academics in the field Education.

While expressing my sincere appreciation for this task performed for the development of the education system, my heartfelt thanks go to all of you who contributed your knowledge and skills in making this document such a landmark in the field.

M.F.S.P. Jayawardhana

Deputy Director General

Faculty of Science and Technology

Guidence and Approval	Academic Affairs Board National Institute of Education
Subjec Cordinator	Ms. M. A. Indra Pathmini Perera Senior Lectuer, National Institute of Education

Curriculum Committee

Mr. K. R. Pathmasiri	Director, Department of Commerce, National Institute of Education
Ms. M. A. Indra Pathmini Perera	Senior Lectuer, Department of Commerce, National Institute of Education
Dr. H. M. L. K. Herath	Senior Lecturer Wayamba University of Sri Lanka.
Mr. K. A. Darmasena	Senior Lecturer University of Sri Jayawardenepura.
Mr. S. A. C. Stanley Silva	Senior Lecturer University of Sri Jayawardenepura.
Ms. M. Kamani Perera	Education Director Ministry of Education.
Mr. W. M. P. G. Edirisingha	SLTS 1 Vishaka Vidyalaya, Colombo 07
Ms. M. L. S. L. Perera	SLTS 11 Ananda Vidyalaya, Colombo 07
Ms. M. Neranjan	SLTS 1 Ramanadan Hindu College, Colombo 04
Ms. Gayani Arunika Perera	SLTS 11 Panadura Balika Vidyalaya, Panadura.

Mr. Prabaharan	Lecturer Department of Commerce National Institute of Education
Mr. Ananda Maddumage	Lecturer Department of Commerce National Institute of Education
Mr. D.L.C.R.Ajith Kumara	Lecturer Department of Commerce National Institute of Education
Mr.S.R.Rathnajeewa	Assistant Lecturer Department of Commerce National Institute of Education
Mr.S.K.Rathnasiri Silva	Senior Lecturer University of sri Jayawardhanapura
Mr. Hemantha Diunugala	Senior Lecturer University of sri Jayawardhanapura

Writting Pannel :

Mr. W. M. P. G. Edirisingha	SLTS 1 Vishaka Vidyalaya, Colombo 07
Ms. M. L. S. L. Perera	SLTS 11 Ananda Vidyalaya, Colombo 07
Mr. W. M. B. Jayasingha	SLTS 1 (Retired) Nalanda College, Colombo 10
Ms. K. V. Abrew	SLTS 1 (Retired) St. Pauls Balika Vidyalaya, Bambalapitiya
Ms. M. E. M. Fernando	SLTS 1 (Retired) St. Jhoshep Vass Vidyalaya, Wennappuwa.
Ms. Gayani Arunika Perera	SLTS 11 Panadura Balika Vidyalaya, Panadura.
Mr. C. L. M. Navas	ISA, Zonal Education Office, Ibbagamuwa.
Mr. M. H. M. Buhari	ISA (Retired) Zonal Education Office, Kegalle.

Editers Pannel :

Ms. M. A. Indra Pathmini Perera	Senior Lecturer Department of Commerce, National Institute of Education
Dr. H. M. L. K. Herath	Senior Lecturer Wayamba University of Sri Lanka.
Mr. K. A. Darmasena	Senior Lecturer University of Sri Jayawardenepura.
Mr. S. A. C. Stanley Silva	Senior Lecturer University of Sri Jayawardenepura.
Mr. Hemantha Diunugala	Senior Lecturer University of sri Jayawardhanapura

Instructions of Referring to the Teachers Instructional Manual (Teachers Guide)

The Business Statistics syllabus for General Certificate of Education (Advanced Level) is implemented from 2017 onwards after being revised under the national curriculum revising policy which is implemented once in every eight years. The Business Statistics which was introduced as an Advanced Level subject in 1997 was undergone to the first revision on competency based in 2009 has been orderly listed out under 11 competencies for both grade 12 and grade 13 in this manual. A practical teaching-leaning process that can be implemented in the classroom for 42 competency levels from the first competency in grade 12 syllabus has been proposed in this manual.

This teacher instruction manual has been complied all the competency levels prescribed for grade 12 in Business Statistics syllabus of General Certificate of Education (Advanced Level).

The relevant competency, competency levels, the number of periods allocated for the competency level, the learning outcomes expected to have been achieved at the end of learning the subject matters under the competency level, are contained first and then proposed instructions for lesson planning, followed by a guideline to explain the subject matters are also contained in this manual in great details. Proposed activities for assessment and evaluation are also associated with many competency levels at the end.

The classroom teacher-learning process is expected to be planned in a manner of Business Statistics will be developed parallel to the growth of attitudes, skills and practices of the students. An adequate guidance is expected to be gained for that purpose through this Teacher Instruction Manual.

Every teacher should lead the students for practical learning through planning the lessons for classroom teacher – learning process with reference to the proposed benchmarks under instructions for lesson planning and the detailed facts contained in the guidelines to explain the subject matters.

Since Business Statistics is a practically important subject, the lesson plans are expected to be prepared by the teachers expanding the boundaries of the scope of their comprehension, reviewing the updated subject matters simultaneous to the prospective changes that may possibly take place in the business field.

Project Leader

Learning Outcomes and Model Activities

Competency 1.0 : Studies the nature and scope of Business Statistics

Competency Level 1.1 : Enquires Business Statistics and its Limitations

No. of Periods : 04

Learning outcomes :

- 'Defines' 'Statistics'
- Explains the role of Business Statistics.
- Differentiates the Descriptive Statistics and Inferential Statistics.
- Points out Importance of Statistics.
- Explains the limitations of Statistics.
- Describes the mis-uses of Statistics.

Instructions for Lesson Planning :

- Involve in a brain-storming exercise inquiring in to the knowledge of students about 'Statistics'
- Write down on the board the terms/statements/clauses which are supportive to construct a definition for Statistics using the responses forwarded by the students.
- Inquire into the subject matters learnt related to Statistics in G. C. E. Ordinary Level class in learning Mathematics. Note down those facts also on the board.
- Construct an appropriate definition for 'Statistics' with the students.
- Discuss the well known definitions with the students.
- Present the following table containing a summery of the results of few Arts subjects of G. C. E. (A/L) examination in a selected school.

Subject	No. of candidates appeared	No of candidates passed	Percentage (%)
Statistics	90	60	67
Sinhala	25	20	80
History	40	30	75

Raise the following questions from the students

- From which subject have the highest number of students got through?
- From which subject have the smallest number of students got through?

- Hold a discussion highlighting the following facts.
  - In order to launch a study in a particular field, the data related to that field should be collected .
  - Accurate decisions can not be made considering data absolutely (without a proper compression).
  - Data should be organized for a meaningful comparison.
  - Data should be analyzed to achieve at correct decisions regarding the relevant field.
- In accordance with the table mentioned above.
  - The details related to the number of students appeared for each subject and the number of students passed are collected data.
  - The number of students passed in Statistics, Sinhala and History being 60, 20 and 30 respectively are considered as absolute data.
  - Tabulating the number of candidates appeared at the examination and number of students passed is considered as organizing of data.
  - Percentage of students passed in Statistics, Sinhala and History being 66.7%, 80% and 75% respectively is considered as analysis of data.
  - Accordingly deciding that the subject recorded with best results as Sinhala is considered as coming to conclusions.
- Accordingly, explain the steps of 'Statistics'
- Foreword the following statements one by one to the class and highlight the role of Statistics and its limitations.
  - "The profit of the business has increased in 2015 compared to 2014."
  - "The profit of the business has increased by Rs. 50 000 in 2015 compared to 2014. "
  - The number of units produced in five days in a farm of manufacturing school bags respectively are as follows.

    58 42 70 66 44
  - The average number of school bags produced in a day in a firm of manufacturing school bags is 56.
  - The employees of ' Hotel Lavila' are honest.
  - Neela's weight is 60 kg, height is 166cm and her monthly salary is Rs. 40,000/-
  - " There is a 90% confidence about receiving an ordered batch of materials in proper time."
- Explain the Role and Limitations of Statistics, in accordance with the above mentioned statements.

- Discuss the importance of Statistics with the students.
- Make use of the following details to explain the Descriptive Statistics and Inferential Statistics.
- Given below is the amount of money spent by 10 students drawn in random to buy food items from school canteen in terms of Rs. 25, 30, 40, 50, 60, 40, 20, 22, 28, 40
  - Arrange these data in ascending order.
  - Find mean, median and mode of these data.
  - How much do you assume that the average amount of money that a student in this school spending in a day to buy food items from the canteen.?
  - Explain that to take down the amount of money spent by each student in the selected group of 10 students is data collection, to arrange them in ascending order is data organizing and to compute mean, mode and median is analysis of data.
  - Point out that studying a sample in great details is 'Descriptive Statistics.'
  - Point out that assuming the average amount of money spent by a student in this scool in a day is 'Inferential Statistics.?
  - Explain the miss-uses of Statistics.

A guideline to explain the subject matters :

" Statistics is a method of decision making in the face of uncertainty on the basis of numerical data and calculated risk." Prof. Ya-Lum-Chou.

"Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data." Croxtion and Cowden

"Statistic is the science of measurement of social organizsm regarded as a whole in all its manifestations."

Hence " Statistics is a study the techniques of collection of data related to various subjets organization of those data, presentation of those data and coming to conclusions through proper analysis of those data."

- The steps of a statistical study are as follows :
  - Collection of relevant data
  - Organization and presentation of data
  - Analysis of data
  - Coming to conclusions
- The role of Statistics

- Presentation of data in a significant manner
- Presentation of complicated data simply for easy understanding
- Being a technique of comparison
- Ability to analyse individual experience scientificaly in a broad range
- Providing with proper guidance for furnishing the business policies
- Ability to quantify the size/strength of a certain phenomena
- Being supportive to highlight cause and effect relations

Limitations of Statistics

- Application of only the quantitative data
- Statistics does not deal with individual observations
- Statistical conclusions being true only on general situations
- Statistical information being miss-used due to negligence or ignorance
- Not being able to ascertain everything using Statistics.
- Statistical conclusions being involved in uncertainly
- Outcomes of a statistical study are not viable for ever

Importance of Statistics

- Leading towards optimal decisions before uncertainly.
- Ability of achieving at optimal decisions for entirety through outcomes of a sample survey
- Ability of forecasting the potential movement of a variable
- Ability of recognizing the interrelations among variables
- Ability of being aware of the relative importance of diverse variables
- Ability of scrutinizing complecated systems in simply
- Colletion of data, orgainzation of data, presentation of data and analses of data is known as Descriptive Statistics
- Coming to a conclustion about a whole using the results of sample study is Interental Statistics

Some of the misuses of Statistics are given below.

- Falsely interpretation of the results of an analysis

  Ex : when two students were selected to the university out of only three candidates sat for the (A/L) exam in a particular school, stating that more than 66% of the students get the chance of university entrance in that school

- Using insufficient data for making comparisons

  Ex : When the profit of X Company is Rs. 100 000 and that of Y company is Rs. 150 000 stating the X company earns less profit than Y company without knowing the business volume of each company

- Making bias interpretations for statistical data

  Ex : Predicting that a particular candidate is winning in the general election based on the results of a sample survey launched, drawing a sample from a more favourable area for that candidate

- Making recommendations without using a sufficient and unbiased sample

  Ex : When four doctors out of five have recommended a particular drug, popularizing that the drug hase been recommended by 80% of the doctors

- Selecting sample infavour of a particular party (Drawing bias sample)

  Ex : Analysting the data collected from a personal desire of the invigilator

Competency 1.0 : Studies the nature and scope of Business Statistics

Competency Level 1.2 : Manipulates the Contribution of Statistics in Business Field.

No. of Periods **:** 04

Learning outcomes :

- Highlights the situations where the subject Statistics is used in the Business filed.
- Lists various techniques used in Statistics.
- Explains how those techniques are applicable in business field.
- Evaluates the contribution of Statistics in other subjects.

Instructions for Lession Planning :

- Present the following statements to the class related to ABC firm.
  1. The best 10 candidates have been recruited to the service based on the results of the interview held in 2017.
  2. The profit of the year 2015 has increased by 10% compared to the year 2014.
  3. "Feasibility of opening a new breach in Ampara is required to be assessed."
  4. A batch of goods sent by a supplier for an order has been rejected since it did not meet the expected quality level.
- Present these statements to the class one by one and highlight the business field revealed from each.
- Statistical techniques which can be assumed to have been used in each situation related to each statement above can be discussed as follows :
  1. Each candidate has been assigned marks based on various criteria in the interview and 10 have been recruited considering the average marks having analyzed the data using techniques such as mean, range or the weighted mean etc.
  2. The techniques as time series analysis and index numbers have been used in concluding the profit of 2015 has increased by 10% compared to 2014, considering the net profit of 2014 and 2015 seperately.
  3. The techniques such as sample surveys, hypothesis testing, statistical estimations, regression & correlation analysis and the probability law have been used in the feasibility study about opening a new branch in Ampara.
  4. Statistical quality control techniques have been used in making a decision to reject the batch of goods sent by the supplier.
- Hence explain the techniques used in Statistics.
- Point out that the statistics are used in some other fields rather than the business filed.

A guideline to explain the subject matters :

- Statistical techniques are used to make decisions in business field in following situations.
  - When an industry is established at a suitable location
  - When the goods and services are manufactured
  - In production planning
  - In market surveys
  - In product control
  - In human resources management
  - In financial management

- Following statistical techniques are used in decision making.
  - Probability law
  - Sample surveys
  - Regression and correlation analysis
  - Time serious analysis
  - Statistical quality control
  - Statistical inference
  - Index numbers
  - Hypothesis testing

- Statistical technuqes are used in following specific situations in the business field.
- In Production Management
  - A feasibility study is launched before an industry being located in an area for evaluating the eligibility.
- Application of Statistical Quality Control (QC)
  - Technices to examine the quality standard of goods and services manufactured.
- In marketing management
  - Ability of applying sample survey techniques to evaluate the consumer taste
  - Time series analysis can be applied to study the variations take place in sales for a considerable period of time.
- In Human Resources Management
  - Data analysis techniques are used for analyzing a bio-data sheet in enrolment of employees to an institution.

- Hypothesis Testing can be used to check whether the employee effectiveness depends on the salary.

- In Financial Management
    - Ability of applying Index Numbers to estimate the cash flows in a proposed project.
    - Ability to apply Regression & Correlation analysis to estimate the prospective profits in a business.

- Statistical techniques are applied not only in business field but also in some other fields.
    - In medical science in order to diagnose the diseases according to the growing pattern of the related symptoms and to test and prescribe various treatments.
    - In order to conduct various researches in the field of engineering
    - In order to conduct researches in the field of Agriculture.
    - In order to study the relationship between the market demand and price and to select the optimum production opportunity among various opportunities in Economic field.
    - In order to select the optimum production opportunity from the alternation production opportunities in the field of production management
    - In order to analyze examination marks and to assess the intelligent quotient in education field.
    - In order to make weather forecasting in the filed of Meteorology.
    - In order to analyze various social phenomena and to identify individual behaviour in Sociology and Psychology.
    - Various research institutes apply statistical techniques to analyze data and come to rational conclusions through hypothesis tastings and estimations.

Competency 2.0 : Organises and presnets business data.

Competency Level 2.1 : Studies various data sources.

No. of Periods : 06

Learning outcomes :

- Introduces statistical data.
- Describes the need of data for statistical studies.
- Introduces what a population is.
- Introduces what a sample is.
- Introduces quantitative data and gives examples.
- Introduces qualitative data and gives examples.
- Introduces internal data and gives examples.
- Introduces external data and gives examples.
- Gives examples for internal and external data sources seperately.
- Introduces primary and secondary data sources and differentiates between them.
- Points out the reliability of primary data sources.
- Points out the merits and demerits of primary data.
- Lists out the secondary data sources.
- Points out the merits and demerits of using secondary data sources.
- Categorizes data on accordance with measuring scales.

Instructions for Lesson Planning :

- Inquire the weight and height of few students in the class and write on the board.
- Explain that those are the data related to weight and height of the students in the class.
- Hence explain the 'characteristic' and 'variable'.
- Inquire how the data related to the way of students coming to school can be derived.
- Assure that all the students in school can be inquired or a few selected students can be inquired about their way of coming to school.
- Hence explain the 'population' and 'sample'.
- Explain with adequate examples that the sample should very well represent the population.
- Inquire the grade received by each student in the class for Mathematics at G. C. E. (Ordinary Level) examination and write on the board.
- Inquire the number of honest students, number of obedient students, number of good students and number of clever students in the class.

- Explain that statistical data can be categorized as quantitative data and qualitative data.
- Suppose that the data are required to study whether the exam results of the students can be improved by participating supportive tuition classes.
- Point out that these data can be collected from,
  - The pupils in the school.
  - Pupils in other schools
  - Parents
  - School teachers
  - Teachers who conduct supportive tuition classes.
- Hence explain the internal data and external data.
- Suppose that the data related to the results of G. C. E. (O/L) examination are needed.
- Point out that these data can be collected either by inquiring from the pupils in the class or with reference to the results sheet available in the office.
- Hence explain 'primary data' and 'secondary data'.
- Suppose that data are needed to study about the educational background in a rural area.
- Point out that these data can be collected from parents and teachers in the area as well as observing the physical resource available there.
- Explain that those are primary data sources.
- Suppose that the data related to the composition of population in few years are needed to be collected.
- Inquire the sources from which these data can be collected in this purpose and write down on the board.
- Such sources are ;
  - Annual report of the Central Bank of Sri Lanka.
  - Statements and reports of the department of Census and Statistics and in any other public or private institute.
  - Magazines
  - News papers
  - Inter-net etc. and explain that these ae the secondary data sources.
- Explain the advantages and disadvantages of primary data.
- Explain the advantages and disadvantages of secondary data.

Activity 1 :

Categorize the following variables considering the values assigned on them under the four heading.

- Nominal scale
- Ordinal scale
- Interval scale
- Ratio scale

Variables

Age, Gender, Academic level, Consumer taste for a product, Social status, Complexion, Price of a product, Mass of a product, Temperature in Colombo city. Spectators' interest for a T.V. programme. Leaves obtain by workers in a factory.

Answer :

Nominal Scale	Ordinaly Scale	Interval Scale	Ratio Scale
- Gender - Complexion -	- Academic Level - Social status - Specters' interest for a T.V. programme - Consumer taste for a product	- Temperature in Colombo city	- Age - Price of a commodity - Mass of a product - Leaves obtained by workers

Activity 2 :

- Let the students to study the following case carefully.

  The principal of Jayasumana Vidyalaya called upon Mrs. Anumani, the teacher-in-charge of Business Statistics and assigned her to forward a report to him about the late comers to school daily.

  The teacher called upon the prefects who have been involved in duty at the gate in the morning and revealed the following details referring to their field note books.

No. of daily late comers in each week in last month.

Day	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
Monday	38	40	36	42
Tuesday	25	24	23	26
Wednesday	27	26	25	21
Thursday	28	22	25	21
Friday	22	26	24	22

- Further she got revealed the following facts as well, from the late comers.
    - Some of those students usually travell in common school service buses.
    - Late comers are more frequent on Mondays since heavy traffic congestions are reported usually on Mondays.
    - Some of the students get late to school, because they have to support their parents' daily works.
    - Five students who travel from more than 15 km away in public passenger transport buses, frequently get late to school.
- Next Anumani teacher called upon the parents of the students who come from more than 15km away and revealed the following facts through the discussion.
    - They can't afford the private bus fare between Rs. 3 000 – Rs. 5 000 per month, bus less than Rs. 100 is to be paid for monthly season ticket in SLCTB.
- After that she called upon the parents of two students who are frequently late to school because they have to support their parents and revealed the following reasons.
    - Taking the cattle away from the pen and tie them in grass lands.
    - Going to the paddy field and supplying water.
    - Supporting the younger siblings in the absence of parental care.

After providing with a sufficient time to study this case carefully put the following questions on the board.

1. What are the data collected from school premises?
2. What are the data collected out side the school premises?
3. What can be considered as statistical data among them?
4. What are the qualitative data among them.
5. What are the primary data revealed through this case study?
6. What are the secondary data revealed through this study?

A guideline to explain the subject matters :

- The set of all the elements / members included in the field of a statistical study is termed as '**population**'.

    Ex : in a study launched to analyze the problems related to the employees in a particular factory, all the workers employed in the factory are included to the 'population'.

- In order to come to conclusions regarding the 'population' a representative potion drawn in random (few elements representing the population) is termed as '**a sample**'.

    Ex : A set of ten workers drawn in random from factory

- Data collected from a representative sample are supportive to achieve at accurate and reliable conclusions regarding the population considered in a particular study.
- Unless those data are accurate and reliable the conclusions achieved based on those data may be false and invalid.
- Quantifiable data are considered as 'quantitative data'.

  Ex : exam marks of students.
- Unquantifiable data are considered as qualitative data.

  Ex : honesty and capability of students.

  Those qualitative data are not considered as statistical data.
- Data collected inside the business premises for the purpose of the study such data are called internal data.

  Ex : data collected from the labourers in the firm.

  Data collected from the managers of the firm to study the problems related to workers in a particular company.
- Data collected from any outside person or outside institution for a particular study such data are called external data.

  Ex : in a study launched to investigate in to the problems related to the employees of a company, the data collected from any legal act or ordinance or from any external institution like the bank, insurance companies etc.
- In addition to the data available inside the premises, the data collected from outside institutes or individuals are called 'external data' those external and internal data are categorized as primary data and secondary data.
- Data collected by the invigilators for the first time directly from the field focusing the objectives of the study, are known as **primary data.**

  Ex : 1. data collected from villagers in a study launched regarding the poverty alleviation in a rural area

  2. data collected by department of Census Statistics
- When the data collected by an individual or a firm for a particular study are used for another study, then those data are known as Secondary data.

  Ex : data extracted form the annual report of Sri Lanka Central Bank for a study launched regarding the povety ellevation in rural areas
- The individuals interviewed and any other physical factors observed in collecting primary data for a particular study are considered as primary data sources.

  Ex : in a study launched regarding the education facilities available in a rural area

  The individuals such as the teachers, students, parents and physical recourses such as school buildings, classroom conditions, sanitary facilities etc.

- The books, magazines, newspapers articles and the inter-net from which secondary data are extracted are known as secondary data sources.

  Ex : 1. annual report of the Central Bank of Sri Lanka

  2. Voters register declared by the elections secretariat annually
  3. Past financial statements in a business firm
  4. Pay sheets attendance registers etc. in a firm

- Merits and demerits of using primary data are mentions below

  Merit

  1. Holding a grater reliability
  2. Considering of a higher accuracy
  3. Relevancy for the objectives of the study
  4. Being up dated

  Demerits

  1. Incurring higher cost
  2. Spending much time
  3. Inability to collect primary data in an unfriendly atmosphere to launch a survey

- Merits and demerits of using secondary data are as follows :

  Merits :

  1. Ability to collect data incurring a lower cost.
  2. Ability to finalize the study during a short period.

  Demerits:

  1. Possibility of data being irrelevant to the objectives of the current study
  2. Possibility of data being out dated
  3. Possibility of some assumptions on which those data were collected at primary stage being disagreed with the objectives of the current study

Nominal Data

- Once a particular characteristic or an attribute related to any other categorical variable is available in nominal values, only with the purpose of categorizing such attributes or characteristics, numbers or any other symbols are used and the data derived in that manner are called nominal data.
- No any mathematical operation can be launched on such numbers.
- The data derived in this manner are known as 'Nominal data' and only a separation of data on the relevant attribute or characteristic is expected.

Ex : (i) Gender of an individual can be inquired in various ways as follows.

Put the relevant code in the cage

- Sex ☐

If male 'M' female 'F' ☐

If male '1' female '2' ☐

If male 'B' female 'G' ☐

Put a '√' male ☐ female ☐

Ex : Residential district can be required as follows.

Put the relevant letter in the cage

Residential District ☐

Colombo – C Gampaha – G Kalutara – K ☐

Colombo – 1 Gampaha – 2 Kalutara – 3 ☐

Put a '√' in the relevant cage

Colombo ☐ Gampaha ☐ Kalutra ☐

Ordinal data

- The codes assigned in categorical variables in accordance with a particular attribute not only with the purpose of categorically identifying but also for a meaningful comparison order or ranks and such data are known as ordinal data.
- There isn't a specific scale among these orders (ranks)
- By categorizing the data ordinary in this manner identifying the data categorically as well as a rough idea about the quantitative(extent / magnitude etc.) strength of data also can be given.

Ex : Various ways on which the programme for Mathematics can be inquired are as follows :

- Put the code related to your choice in the cage
- Like very much - 4 • Like - 3 • Hate - 2 • Extremely hate- 1

Put a '√' in the relevant cage

Like very much	☐
Like	☐
Hate	☐
Extremely hate	☐

Put the relevant code in the cage ☐

- Like very much - A • Like - B • Hate - C • Extremely hate - D

Interval Data

The data that can be undergone to mathematical operations, indicating equal deviations, associated with relatively an order, including a zero, but not a true zero are known as Interval Scale data.

Ex : degrees of Celsius (C^0) and degree of Fahrenheit (F^0) are used for measuring the temperature

- Both these measuring scales can take a zero value, but zero in each scale does not give the same meaning.

$0\ C^0 = 32\ F^0$ Where as

$0\ F^0 = -17.7746\ C^0$

- The difference between any two successive values in each scale is the same.

Ex : $34\ C^0 - 33\ C^0 = 2\ C^0 - 1\ C^0$

or

$$98 F^0 - 97 F^0 = 2 F^0 - 1 F^0$$

but

$$2 F^0 - 1 F^0 \neq 2 C^0 - 1 C^0$$

Ratio Scale Data

The data can be undergone to all the mathematical operations, having an equal deviation between any two successive values, ratio of any two values is meaningful (rational) and consists of a true zero, are known as, Ratio Scale data. Hence the characteristics of Ratio Scale Data can be stated as

- Identification of categorical variable
- Indenting the quantity / extent / magnitude
- Availability of specific intervals
- Containing a true zero

Examples for Ratio Scale data

- Marks (scores)
- Height
- Mass
- Age / Life expectation
- Income/expenditure

Competency 2.0 : Organizes and Presents Business Data

Competency level 2.2 : Construct the instruments required to collect data.

No. of periods : 10

Learning outcomes :

- Lists the methods of collecting data.
- Explains the personal interview method.
- Points out merits and demerits of personal interview method.
- Explains self enumeration method.
- Points out merits and demerits of self enumeration method.
- Introduces telephone interview method.
- Points out merits and demerits of telephone interview method.
- Explains how to collect data using direct observation method.
- Points out merits and demerits of direct observation method.
- Explains how to collect data using electronic media (e-data).
- Explains how to collect data using focused groups.
- Comparatively analyses the various instruments used in data collection.
- States the facts to be considered in preparing the 'questionnaire' and the 'schedule'.
- Introduces the 'pre-test'.
- Introduces the 'editing of data' and its requirement.

Instructions for Lesson Planning :

- Inquire the students about the methods that can be used to collect data regarding the occupation and income of the people in your area.
- Note down the methods suggested by the students on the board.
- List the methods of collecting primary data from those methods.
- Inform the students to name the appropriate method to collect data at each situation mentioned below and inquire the reason for naming that method.
  - In a study launched regarding the reading interest of the Advanced Level students in all the schools in the district
  - In a study of health and sanitary issues of the people in a rural area
  - The price of the shares exchanged in Colombo Stock Exchange at this moment

- In a study about the types of vehicles moving in a particular express way
  - In a study about the efficiency of a newly introduced software
  - At the end of a training programme organized by an institute to collect data from the participants to assess the success of that programme
  - In an inquiry from the people scattered globally about the efficiency of a newly introduced product
- Using the responses of the students explain the relative merits and demerits of each method.
- Presents the following questionnaire to the class and guide the students to complete it.

1. Grade
2. Date of birth
3. Which amount of money do you spend monthly for each item given below.
   - to buy meals from the canteen
   - travelling expenses
   - clothes, foot ware etc.
   - stationary items
   - other necessities
4. Number of days you are absent from school in a week is.
   - 0 • 1-2 • 2-3 • More than 3
5. Mention the number of subjects for which you received each grade at G. C. E. (O/L) exam.
   - A • B • C • S • W
6. Your preference for the school base assessments.
   - Like • Dislike
7. How much did you spend for books and stationary in last months ? Rs. ....
8. You suppose that the supportive tuition is essential don't' you?
   - Yes • No • Can't say
   - Some times needed • No need at all
9. Come out with your points of view regarding the Competency the Development Projects.

   ....

   ....

10. Express the amount of money you spend to buy meals from the canteen for a month, as a percentage of total monthly income of your family.

11. Number of members in your family

- 1
- 2
- 3
- 4
- 5
- more than 5

- Explain the following using this questionnaire.
  - Qualities expected in a good questionnaire
  - Types of questions in a questionnaire with the merits and demerits of each type.
  - Steps of preparing a good questionnaire.
  - Pre – test

- Explain the shortcomings of the above questionnaire on following facts.
  - Weak-points in the structure of the questionnaire.

    1st and 2nd questions are related to the personal information, but again 11th question inquires about the privacy. The questions focused the same area should be arranged together.
  - Difficultly answerable questions are contained.

    3rd question inquires the expenditure for clothing.
  - Containing biased questions (a leading questions)
    - 8th question is biased, since it contains a clue to the answer.
  - Containing complicated calculation. 10th question is complicated
  - Containing the questions focusing the past memories.

    Question no. 07 inquires about the past memory.

A guideline to explain the subject matters :

- There are various methods of collecting Primary data
  - Direct observation method
  - Self enumeration method (questionnaire method)
  - Telephone interview method
  - Personal interview method
  - Electronic data collection method
  - Focused group interview method

Personal interview method.

- The method of collecting data meeting the respondents personally and discussing with them, is known as the 'personal interview method'
- Here the data revealed by the respondents are taken down in the schedule by the interviewer/invigilater/investigator/examiner.
- Collecting data visiting few selected families and discussing with the Chief Occupant in a survey launched to study about the living condition of a rural community, can be taken as an instance.

- Few advantages of using personal interview method are as follows :
    - Ability to expect relatively a grater rate of responses
    - Ability to get data with a higher reliability
    - Ability to ensure the confidence in respondent about the study
    - Ability to collect data from illiterated people as well
    - Ability to ensure the accuracy of data revealed by the respondent.

- Few disadvantages in personal interview method are as follows :
    - Being an expensive method
    - Difficulty to collect data on extremely sensitive matters
    - Possibility of being biased on the investigator/interviewer/invigilator/examiner
- The efficiency of personal interview method depends on the training and experience of the investigator/interviewer/invigilator/examiner

Direct observation method

- The method of collecting data observing the relevant field by the invigilators being involved personally in that field is known as the 'direct observation method.' Any suitable instruments or appliances can be used with this regard.

  i.e. electronic cameras, (CCTV) speed meters etc.

- Few advantages of direct observation method are as follows.

    - The accuracy being at a higher level
    - The response rate being at a higher level
    - The reliability being at a higher standard
    - No any other evidence being needed to verify the validity of data

- Few disadvantages of direct observation method are as follows :

    - Usage being limited
    - Spending much time and money
    - Possibility of being biased
    - Accuracy of data being depend on the quality of technical instruments

Self enumeration method :

- The method of collecting data, giving a questionnaire to the respondent to be completed by himself, is known as the 'self-enumeration method'
- Here the data are reported by the respondent him/her self.
- A copy of the questionnaire can be sent to the respondent in mail or in another way.
- Few relative advantages of using self enumeration method are as follows.
    - Ability to collect data from a vast geographical area
    - Ability to collect data from a large group of people under lower cost
    - Ability of getting reponses for sensitive questions
    - Ability of finalizing the survey in a short period

- Few relative disadvantages in self enumeration method are as follows :

    - Response rate being at a lower level.
    - Inability to assure the accuracy of data collected.
    - Reliability of data being in a poor condition.
    - Inability to use this method successfully for illiterated people.
    - Inability of supporting the respondents when necessary to have a clear understanding about any question.
- Self enumerations method is more efficient if all the respondents are in a similar academic back-ground.

Telephone Interview Method

- The method of collecting data by questioning over the phone is known as the telephone interview method.
- The data received are noted down in a schedule by the invigilator.

- Few relative advantages of telephone interviews method are as follows :
  - Ability of collecting data immediately
  - Ability of using this method in national level studies as well as in international level studies
- Being a cost minimizing method compared to the personal interview method
- Few disadvantages in telephone interview method compared to the other methods are as follows:
  - Knowledge of the contact numbers of the respondents being compulsory
  - Possibility of sample not being representative, when some of the respondents are not access to the telephone
  - Inability to ensure the accuracy of data provided by the respondent
  - Possibility of data being distorted, due to the communication barriers in network

Focused Group Discussion Method

- The method of collecting data discussing with a small group of individuals, who are well experienced and knowledgable in connection with the field of study is known as focused group discussion method.
- Here in this method the investigator instructs the focused group regarding the data needed to be collected.

Few advantages of focused group discussion method are as follows.

- Ability of investigating in great details
- Being a more appropriate method to collect qualitative data such as attitudes, beliefs and experiences etc
- Being a cost minimizing method compared to the other methods of collecting data
- Rate of responses being in a higher level, since the ideas are exchanged among the group members
- Reliability of data being in a greater level
- Ability of using as a method of collecting additional data in details, after collecting the quantitative data
- This method is supportive to increase the sample size since discussions are held among the group members at once

Few disadvantages of focused group discussion method are as follows :

- Receiving various opinions for a single objective and therefore it is difficult to analyse the data
- Possibility of coming to conclusions being delayed

- The focused group discussion methods is frequently used in the marketing field to get feedback about newly introduced products.

- **The Method of Electronic Data Collecting**
  - Collecting of data using modern electronic technological devices is known as Electronic Data Colleting Method (E-Research)
    - There are few techniques applicable in collecting data under this method.
    - Computer Assisted Personal Interview (CAPI)
    - Computer Assisted Self Interview (SCASI)
    - e-mail survey
    - web site survey
  - Few advantages of collecting data using electronic techniques are as follows.
    - Convenience in usage
    - Quick access of data
    - Being a cost minimizing method
    - Convenience in organization of data
    - Ability of using globally
  - Few disadvantages of collecting data using electronic techniques are as follows :
    - Possibility of the rate of response being unsatisfactory, before lack of computer literacy of the respondent
    - Since the data revealed by the respondents are computerized by the enumerator there is a possibility of being less reliable
    - Inability of getting a representative sample in the absence of modern technological facilities

- A schedule should be used in personal interview method. A schedule may contain, detailed in which the data can be systematically reported, while they are being collected from respondents.
- A questionnaire is used in self-enumeration method. The questions in a questionnaire should demonstrate the following qualities.
  - Being relevant to the objectives
  - Not being complicated, biased and double meaning
  - Not arousing the past memories
  - Clarity and simplicity

Following qualitative characteristics are expected in a good questionnaire.

- Clarification of the objective of the study.
- The questionnaire should not be too long.
- The questions should be arranged in an orderly manner.
- Proper instructions should be stated for responding.
- Leading (biased) questions not being included.
- Double / multi meaning terms / questions not being included.
- Questions arousing the post memories not being included.
- Qustions with complicated calculations not being included.

- A questionnaire should consist of following types of questions.
  - Dichotomous questions
  - Multiple choice questions (MCQs)
  - Open (free answerable) questions
  - Direct questions
- Respondents find it difficult to answer on their own way to dichotomous, multiple choice and direct questions, but in open questions they get the chance to come out with their own responses.
  - There is also a chance of deriving new opinions though open questions.
  - It is easy to analyse the response received for dichotomous,, multiple choice and direct questions, but the responses received for open questions are not that easy to analyse.

- A schedule is a chart (instrument) which is used to note down the data collected on personal interview method or telephone interview method. A guideline is provided to the invigilator through the schedule.
- Both the questionnaire and schedule should be undergone to a pre-test and the identified weak-points should be adjusted.
- Few copies of the questionnaire are given to few members in the same category (not to the chosen sample members) and get them answered to check whether the accurate responses are received or whether there are any shortcomings in it, and then the questionnaire should be re-adjusted before giving it to the properly drawn sample.
- Data collected in any of the above methods, should be edited before they are analysed.
- Necessary adjustments are made for accuracy, completeness, relevency and consistency under editing of data.
- In editing,
  - If some of the responses are illegible they are re-written legibly,
  - Necessary adjustments are made to secure the consistency.

    Ex : correcting the distance mentioned in various units in 'km's
    - Adjusting the real codes for relevant responses.
    - If sarcastical or humorous responses are found or responded by any other persons or not responded, those parts or the entire questionnaire should be rejected.

Competency 2.0 : Organizes and Presents Business Data.

Competency 2.3 : Organizes Business Data.

No. of Periods : 12

Learning Outcomes :

- Differentiates between raw data and organized data.
- Prepares the array of data
- Points out the merits and demerits of the array of data.
- Constructs the Stem & Leaf diagram.
- Explains merits and demerits of organizing data in a stem & leaf diagram
- Names the characteristics expected in a perfect Table.
- Presets data using a perfect Table.
- Constructs grouped and ungrouped frequency distributions using given data.
- Constructs relative frequency distribution and grouped frequency distribution for a given frequency distribution.
- Constructs a cumulative relative frequency distribution using a relative frequency distribution.

Instructions for Lession Planning :

- Provide with the sudents the following details related to the number of over-time hours completed by 30 employees in a manufacturing company in last month.

12	20	08	15	21	21	09	09	10	15
15	16	23	15	12	08	05	05	24	20
16	10	15	22	16	20	25	15	09	16

- Rais the following questions after presenting the above data set.
    - What is the minimum number of O.T hours completed by an employee during this period of 30 days?
    - What is the maximum number of O.T hours completed by an employee during this period of 30 days?
    - Which number of O.T. hours have been covered by the most number of employees during this period?

- Point out the importance of organizing these data in a particular way rather than keeping them as they have been collected, in order to answer these questions easily.
- Lead the students to arrange these data in ascending order or descending order,
- Ensure that the students find it easy to answer the same questions when they are repeated after preparing the array of data.
- Enquire the students about various methods that can be applied in organizing data and note them down on the board.
- Define the' array of data'.
- Introduce the Stem & leaf diagram as an optional method to the array of data and guide the students to prepare a Stema leaf diagram.
- The Stem & leaf diagram related to the above data set is given below.

Stem	Leaf
0	5, 5, 8, 8, 9, 9, 9
1	0, 0, 2, 2, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6
2	0, 0, 0, 1, 1, 2, 3, 4, 5

Key : 0/5 = 05

- Explain that not only the data in two digit numbers, but also the data in any other numerical form can be presented in a Stem & leaf diagram.
- Give the following numbers to students and ask them to prepare a Stem & leaf diagram.
105, 145, 176, 105, 112, 123, 154, 163, 155, 147
- The relevant stem & leaf diagram is as follows.

Stem	Leaf
10	5, 5
11	2
12	3
14	5, 7
15	4, 5
16	3
17	6

Key : 10/5 $\rightarrow$ 105

- Point out that the Stem & leaf diagram can be used when the numbers are given in decimals as well :

  20.3   21.2   20.5   21.7

Stem	Leaf
20	3, 5
21	2, 7

Key : 20/3 $\rightarrow$ 20.3

- Discuss the merits and demerits of organizing data in a Stem & leaf diagram and note them down on the board.
- Present the following paragraph to the class.
- The attendence of the students in grade 12C class in a week has been quoted from the Daily Attendance Register as follows.

  25 boys and 15 girls were present on Monday; the total number of students is 40. Total number of students attended on Tuesday was 45 out of which 28 were girls and 17 were boys. On Wednesday, there were 20 boys and 30 girls where the total number was 50. There were 16 boys and 30 girls present on Thursday. The total number o students present on Friday was 35 out of which 20 were girls.
- Point out the importance of organizing these data and guide the students to organize them in a suitable table.

Day	Attendence		
	Girls	Boys	Total
Monday	25	15	40
Tuesday	28	17	45
Wednesday	30	20	50
Thursday	30	16	46
Friday	20	15	35

Source : Class Attendence Register

- Discuss the qualitative characteristics that can be expected in a perfect table.
- Inquire the students how a data set consists of large number of observations and when the same observation repeat for several times, can be organized.
- Give the following set of observations to the students.

  10, 15, 10, 12 13, 14, 10, 12, 15, 12

  13, 14, 15, 13, 12, 13, 10, 11, 12 13

- Construct the ungrouped frequency distribution with the students related to the above data set.

observation	tally marks	frequency
10	////	4
11	/	1
12	~~////~~	5
13	~~////~~ /	5
14	//	2
15	///	3
Total frequency		20

- Involve the students in following activity to construct a grouped frequency distribution.
- Assuming that these data are the marks obtained by 30 students in the term test for Statistics and note them down on the board.

02, 49, 23, 19, 75, 99, 65, 39, 45, 62
25, 55, 70, 50, 35, 60, 72, 40, 42, 45
63, 50, 59, 48, 64, 65, 78, 79, 80, 79

- Highlight through a discussion with the students the fact that to prepare an ungrouped frequency distribution is not suitable to organize these data.
- Give instructions to the students to prepare a grouped frequency distribution in organizing these data.
- Construct a grouped frequency distribution following the steps mentioned below.

Range = 99 - 02 = $\underline{\underline{97}}$

- Find the width of a class interval considering 5 class intervals are to be prepared.

Class width = 97 ÷ 5 = $\underline{\underline{19.4}}$

- Prepare the class intervals taking the class width as 20 and enter the data to relevant class intervals using tally marks.

Marks obtained	tally marks	No. of students (f)
01 - 20	//	02
21 - 40	𝍸	05
41 - 60	𝍸 𝍸	10
61 - 80	𝍸 𝍸 //	12
81 - 100	/	01
Total frequency		30

- Discuss the term "relative" with the students and explain that the relative frequency of a class interval is the ratio of the frequency of that class interval to the total frequency of the distribution.
- Present the following frequency distribution to the class and lead the students to prepare a relative frequency distribution.

Class interval	Frequency	Relative frequency
10 - 20	02	$\frac{2}{40}$ = 0.050
20 - 30	05	$\frac{5}{40}$ = 0.125
30 - 40	09	$\frac{9}{40}$ = 0.225
40 - 50	15	$\frac{15}{40}$ = 0.375
50 - 60	06	$\frac{6}{40}$ = 0.150
60 - 70	03	$\frac{3}{40}$ = 0.075
	40	1.000

- Point out the fact that when the relative frequency is expressed as a percentage, that is called percentage relative frequency.
- Discuss the meaning of the term ' cumulative' and then introduce 'cumulative frequency'.
- Explain that the 'cumulative frequency is in two types as "less than" and "or more".
- Lead the students to build up 'less than" cumulative frequency distribution and "or more" cumulative frequency distribution using the following grouped frequency distribution.

Class interval	Frequency
21 - 25	02
26 - 30	05
31 - 35	12
36 - 40	04
41 - 45	02

- Point out that , when the "les than" cumulative frequency distribution is prepared, an additional class internal with the frequency zero is to be assumed before the "lowest" class interval in the given distribution.
- "Less than" cumulative frequency distribution for the above frequency distribution is as follows :

Class interval	Frequency	Less than upper boundary	Cumulative frequency
15.5 - <u>20.5</u>	00	Less than 20.5	00
20.5 - <u>25.5</u>	02	Less than 25.5	02
25.5 - <u>30.5</u>	05	Less than30.5	07
30.5 - <u>35.5</u>	12	Less than 35.5	19
35.5 - <u>40.5</u>	04	Less than 40.5	23
40.5 - <u>45.5</u>	02	Less than 45.5	25

- Point out that when the 'or more' cumulative frequency distribution is prepared, an additional class interval with frequency 'zero' is to be assumed after the highest class interval in the given distribution.
- 'or more' cumulative frequency distribution for the above frequency distribution is as follows :

Class interval	Frequency	or more Lower boundary	Cumulative frequency
<u>20.5</u> - 20.5	02	20.5 or more	25
<u>25.5</u> - 305	05	25.5 or more	23
<u>30.5</u> - 35.5	12	30.5 or more	18
<u>35.5</u> - 40.5	04	35.5 or more	06
<u>40.5</u> - 45.5	02	40.5 or more	02
<u>45.5</u> - 50.5	00	45.5 or more	00

- Explain that the sum of the relative frequencies related to each class boundary is the cumulative relative frequency and guide the students to prepare the cumulative relative frequency distribution for the above distribution.

Class interval	Frequency	Relative Frequency	Cumulative ralative frequency
15.5 - 20.5	00	0.00	0.00
20.5 - 25.5	02	0.08	0.08
25.5 - 30.5	05	0.20	0.28
30.5 - 35.5	12	0.48	0.76
35.5 - 40.5	04	0.16	0.92
40.5 - 45.5	02	0.08	1.00

A guideline to explain the subject matters :

- unorganized data collected through a survey are known as 'Raw Data'
- Following techniques are applied in organizing of data.
  - Array of data
  - Stem & leaf diagram
  - Ungrouped frequency distribution
  - Grouped frequency distribution
  - Relative frequency distribution
  - Cumulative frequency distribution
  - Cumulative relative frequency distribution
- Processing of unorganized data in ascending order or descending order is known as an 'Array' of data.
- 'Array of data' or 'Stem & leaf diagram can be used as the initial step of organizing 'raw data'.
- Several advantages of preparing a Stem & leaf diagram are as follows.
  - Ability to make out the minimum value, maximum value and range of data easily
  - Ability to derive the measures like mode and median easily
  - Facilitating calculate quartiles
  - Ability to understand the shape of the distribution roughly
- Several disadvantages of a Stem & leaf diagram are as follows :
  - Data not being summarized
  - Difficulty in constructing in the presence of a great deal of data

- Characteristics expected in a complete Table are as follows :
  - Main heading
  - Stating the unit of data
  - Stating headings and subheadings for rows and columns
  - Table number
  - Stating the total and grand total when necessary.
  - Stating the source of data
  - Stating a foot note
  - A table containing a set of observations associated with the respective frequency of each observation is known as an ungrouped frequency distribution.
- When there is a great deal of discrete data spread in a small range, the ungrouped frequency distribution is more suitable.
- Once the observations are separated into groups (class intervals) and a table containing those class intervals associated with their corresponding frequencies, is known as a grouped frequency distribution.
- When the data are spread in a vast range and to be organized them in a grouped frequency distribution, the width of a class interval should be determined first of all.
- For that purpose the range of observations should be divided by the number of class intervals determined.
- According to the width of a class interval determined, a set of class intervals should be prepared so as to insert each and every observation in the given data set. Normally the number of class intervals should be minimize 5 and maximum 20.
- The grouped frequency distribution is more appropriate, when more data are dispersed in a vast range.
- The two concepts class limits' and 'class boundaries are applied in connection with a grouped frequency distribution. Further when some of the observations overlap on boundaries 'class limits' and class boundaries are the same.
- The identify of original data cannot be revealed after a grouped frequency distribution is prepared. Further, since it is assumed that all the observations inserted in a class interval overlap on the class mark (class midpoint) the 'grouping error' is occurred.
- Once the frequency of a class interval is expressed as a ratio to the total frequency of the distribution, the relative frequency of that particular class interval can be derived.
- The distribution containing the class intervals and the corresponding relative frequencies is known as a Relative Frequency Distribution.
- Relative frequency can be motioned as a fraction or a decimal or a percentage.
- When the relative frequency of each class interval is expressed as a percentage that distribution is known as a percentage relative frequency distribution.
- When a cumulative frequency distribution is prepared using a grouped frequency distribution the class boundaries should be first determined.

- The cumulative frequency distribution is in two types as 'lest than' cumulative frequency distribution and 'or more' cumulative frequency distribution.
- When 'less than' cumulative frequency distribution is prepared an additional class interval with zero frequency is assumed below the first class interval in the same size in the given distribution.
- The sum of the frequencies below the upper class boundary of a class interval of a distribution is known as the 'less than' cumulative frequency of that class interval
- When 'or more' cumulative frequency distribution is prepared an additional class interval with the frequency zero is assumed above the highest class interval of the given frequency distribution with the same size.
- The sum of the frequencies of a distribution more than or equal to the lower class boundary of a particular class interval is called the 'or more' cumulative frequency of that class interval.
- A distribution with 'or more' cumulative frequency of every class interval is known as 'or more' cumulative frequency distribution.
- Once the cumulative frequency of a class interval is divided by the total frequency of that distribution that is the cumulative relative frequency of that class interval.
- A table consists of cumulative relative frequency of each and every class interval is known as a cumulative relative frequency distribution.

Assessment and Evaluation :

- After all the above discussions involve the students in the following activity.

1. Guide the students to prepare the 'array of data' ' stem & leaf diagram and ungrouped frequency distribution giving a set of data spread in a small range.
2. Provide with a set of data spread in a vast range and inquire an appropriate frequency distribution.
    - Let them build up that frequency distribution.
    - Using that grouped frequency distribution lead them to prepare
        - The relative frequency distribution
        - 'less than' cumulative frequency distribution
        - 'or more' cumulative frequency distribution

Competency 2.0 : Organizes and presents the Business Data

Competency Level 2.4 : Presents Busieness data using Charts

No. of Periods : 12

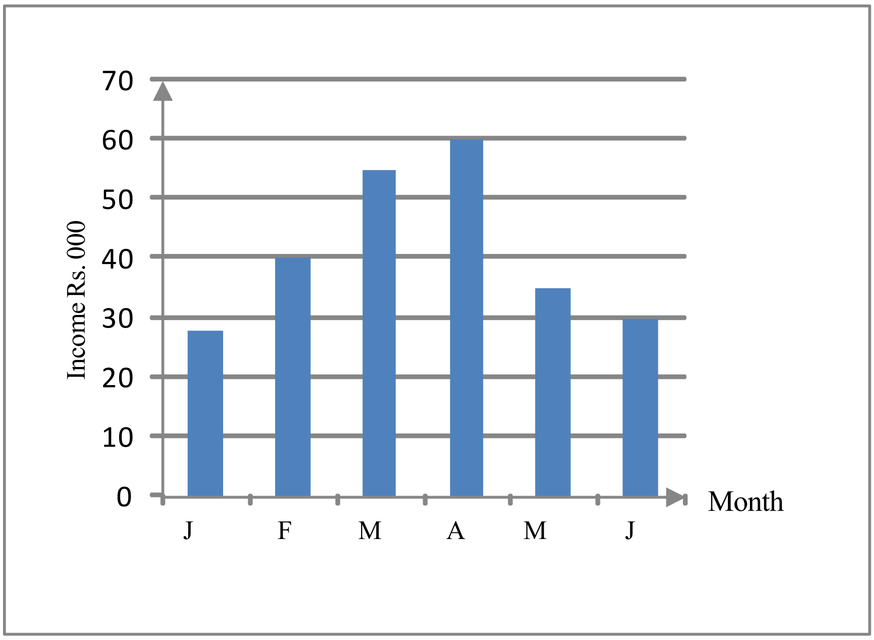
Learning outcomes :

- Describes the facts to be considered in constructing a chart.
- Points out the need of a chart as a technique of presentating data.
- Introduces the 'Simple Bar Chart'.
- Points out the requirements of a Sample Bar Chart.
- Constructs a Sample Bar Chart using that data given.
- Introduces the 'Component Bar Chart'.
- Points out the requirement of Component bar Chart.
- Constructs a Component bar Chart using the data given.
- Introduces the 'Percentage Component Bar Chart'.
- Points out the requirement of a Percentage Component Bar Chart.
- Constructs a Percentage Component Bar Chart using the data given.
- Introduces the 'Multiple Bar Chart.
- Point out the need of a Multiple Bar Chart.
- Constructs a Multiple Bar Chart using the data given.
- Introduces the 'Pictogram'.
- Points out the need of a pictogram.
- Constructs a pictogram using the data given.
- Introduces the pie-chart'
- Points out the need of a 'pie chart'.
- Constructs a pie chart using the data given.
- Introduces a 'Profile Chart'.
- Points out the need of a profile chart.
- Constructs a profile chart using the data given.
- Points out the problems arrisen in representing data using charts.
- Explains the merits and demerits of the techniques of representing data using charts relative to each other.

Instructions for Lesson Planning :

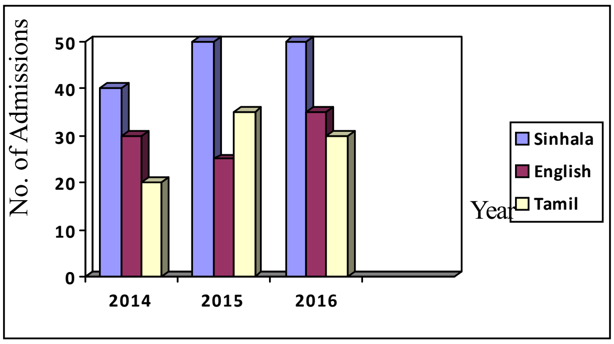
- Presents (Display) the following diagrams to the class and inquire the opinion of the students.

Monthly Sales Income of a business Jan-June, 2017 (Rs. 000')

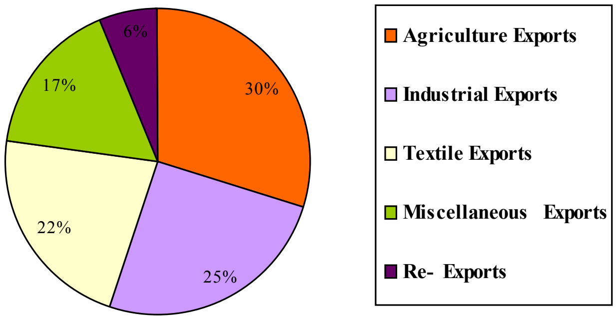


No of students admitted to

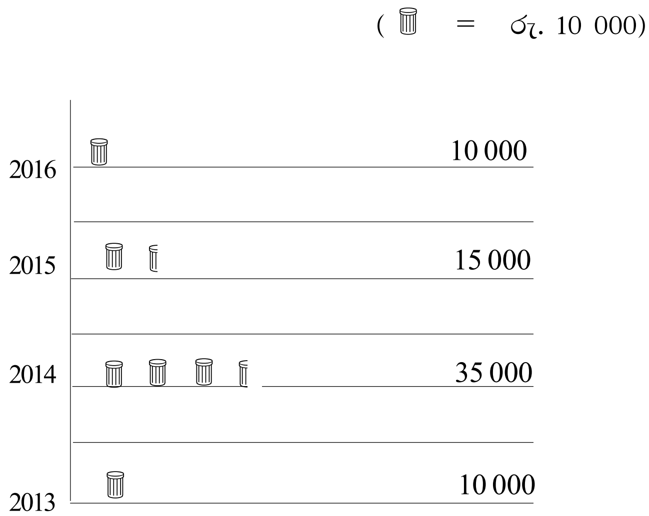
A/L Class in Vijaya Maha Vidyalaya



Composition of Exports in ‘A’ country in 2016



Annual savings deposited in a Rural Development Bank collected through home saving tills



- Hold a discussion highlighting the following facts.
  - Explain that the statistical data can be represented using graphs and charts for convenience of understanding simply.
  - Point out the fact that a rough idea about the data can be derived just at a glance on a graph or a chart even without reading any details about the data.
  - Point out the fact that in accordance with the type of data and the purpose of presentation of data, the most appropriate technique should be selected from simple bar charts, component bar-charts, multiple bar charts, pie charts and pictograms etc.

- Provide with each data set given below and lead the students to construct the diagram associated.

(i) Simple vertical bar-chart.

No. of units produced in product
X- in ABC company (pvt) Ltd 2012 – 2016 (units' 000)

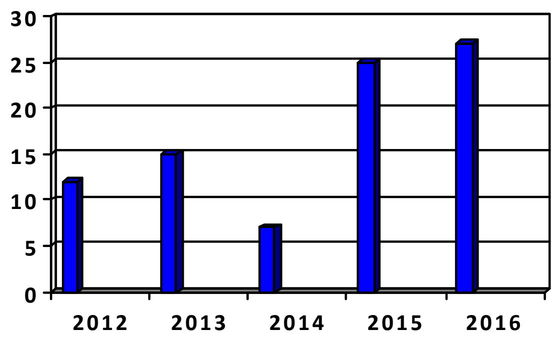
Year	No. of units produced (units' 000)
2012	12
2013	15
2014	07
2015	25
2016	27

Annual profit in ABC company (pvt) Ltd.

Year	Profit/loss (Rs. Million)
2012	24
2013	4.6
2014	(1.2)
2015	6.8
2016	5.4

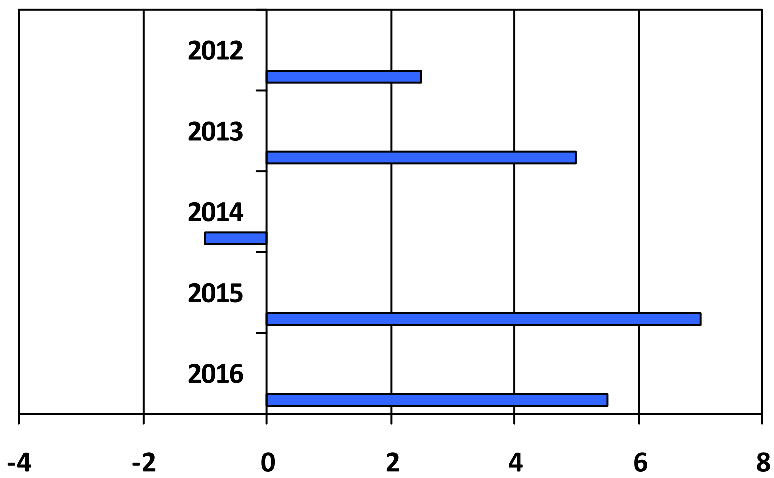
- Lead the students to construct the Simple Bar Chart for this data set.
- Discuss the merits and demerits of simple bar charts.
- The vertical bar charts constructed for the above data set is as follows.

ABC company (Pvt.) Ltd
Annual production of item - X



- The simple horizontal bar-charts representing the annual profit/loss of ABC company (Pvt.) Ltd. is given below.

Annual profit/loss of ABC company (Pvt.) Ltd. (2012-2016)



Source : Income statement (2011-2015)

2. Component Bar-chart

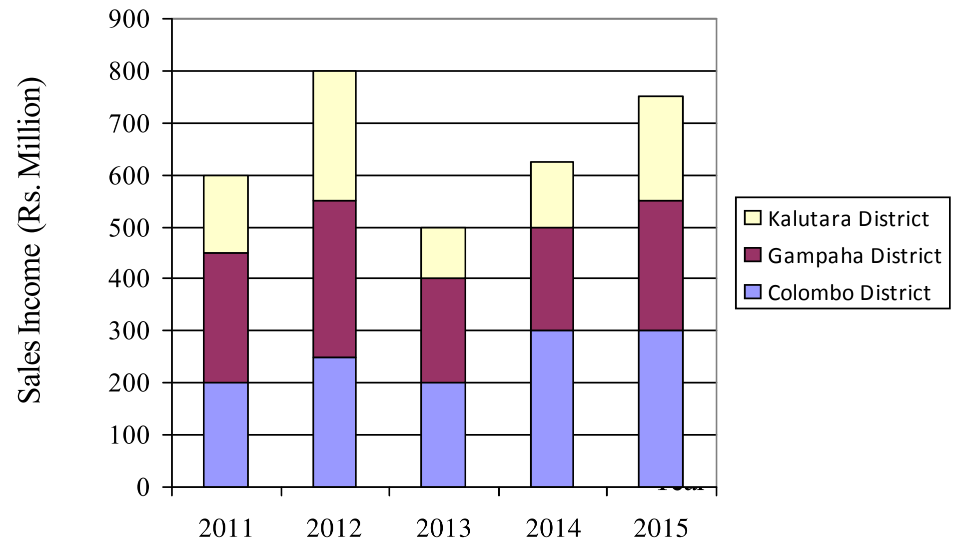
The annual sales income of XYZ company (Pvt.) Ltd carrying out business affairs in Western Province is given below.

Year	Sales Income (Rs. million)			
	Colombo	Gampaha	Kalutara	Total
2011	200	250	150	600
2012	250	300	250	800
2013	200	200	100	500
2014	300	200	125	625
2015	300	250	200	750

- Provide with these data to the students and lead them to construct a component bar-chart and a percentage component bar-chart.
- Discuss the merits and demerits of component bar-charts with the students.
- The component bar-chart related to the above mention data set is as follows :

xyz company (Pvt.) Ltd

Annual Sales Income – 2011-2015



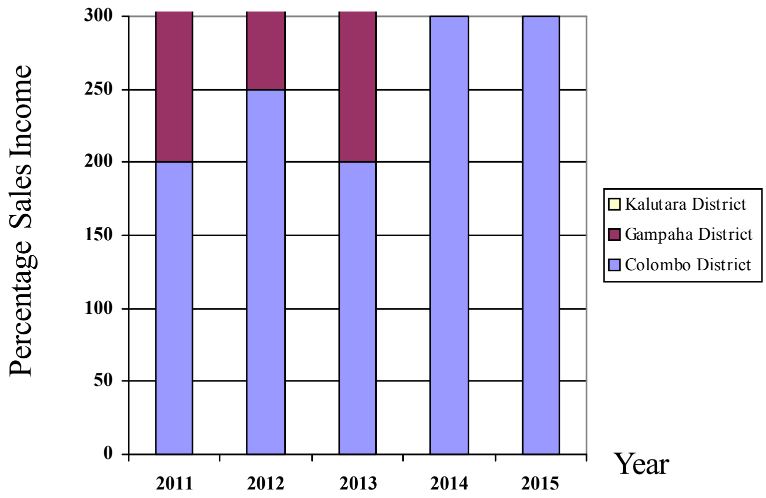
Source : Income statement (2011-2015)

The annual sales income of XYZ company (Pvt.) Ltd carrying out business affairs in Western Province is given below.

Income (Rs. million)

Year	Colombo			Gampaha	Kalutara		Total
2011	200	$\frac{200}{600}\times 100 = 33.33$	250	$\frac{250}{600}\times 100 = 41.67$	150	$\frac{150}{600}\times 100 = 25$	600
2012	250	$\frac{250}{800}\times 100 = 31.25$	300	$\frac{300}{800}\times 100 = 37.5$	250	$\frac{250}{800}\times 100 = 31.25$	800
2013	200	$\frac{200}{500}\times 100 = 40$	200	$\frac{200}{500}\times 100 = 40$	100	$\frac{100}{500}\times 100 = 20$	500
2014	300	$\frac{300}{625}\times 100 = 48$	200	$\frac{200}{625}\times 100 = 32$	125	$\frac{125}{625}\times 100 = 20$	625
2015	300	$\frac{300}{750}\times 100 = 40$	250	$\frac{250}{750}\times 100 = 33.33$	200	$\frac{200}{750}\times 100 = 26.67$	750

Percentage Component Bar - chartchart



3. Multiple Bar chars

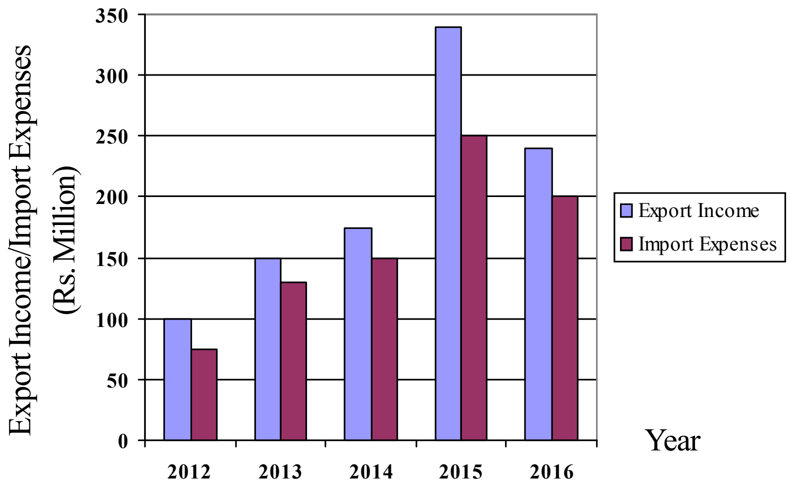
The export income and the import expresses of A country are stated in the following table.

Year	Import expenses (Rs. million)	Export income (Rs. million)
2010	100	75
2011	150	130
2012	175	150
2013	340	250
2014	240	200

- Lead the students to construct a multiple bar-chart using these data.
- Discuss the merits and demerits of multiple bar-charts.

The multiple bar-chart constructed using the above data set is mentional below

Annual Export Income and Import Expenditure : 2012-2016



4. pie-chart

The number of permanent employees in LMN company (Pvt.) Ltd, in each department is mentioned in the following table.

Department	No. of Employees
Factory	240
Stores	60
Office	40
Show room	20
Maintenance	40

- Let the student to complete the following table in-order to construct the pie-chart.

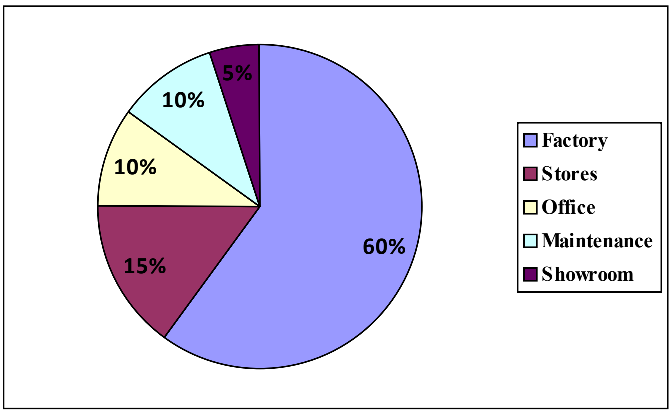
Department	No. of Employees	Percentage of Employees	Magnitude of the sector (Angle)
Total			

- Lead the students to construct those sectors on a circle drawn on a suitable diameter.
- Guide the students to use suitable colours or designs for clear identification of each sector representing the number of employees involved in each department and to finish the work stating the Heading, Key, Chart No. and data sources etc.
- Discuss the merits and demerits of the pie-chart.
- The completed table is as follows.

Department	No. of Employees	Percentage of Employees	Magnitude of the sector (Angle)
Factory	240	60	216^0
Stores	60	15	54^0
Office	40	10	36^0
Show Room	20	5	18^0
Maintenance	40	10	36^0
Total	400	100	360^0

The pie chart constructed for the above data is given below.

Comopistion of Employees in LMN Company Ltd.



Source : Employee Register

5. Profile Chart

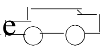
KLM company (Pvt.) Ltd is a textile business firm which consist of an island wide net work. The average departmental income of a branch of the company and the departmental income of Maharagama branch are as follows.

Item	Average departmental income of a branch	Departmental sales income in Maharagama branch (Rs.million)
Baby suits	188	200
Shirts - 15.5"	175	180
Shirts - 16.0"	210	190
T - Shirts - L	240	300
T - Shirts - M	215	245
Trousures	230	225

6. pictogram

- Provide with the following table to the students, containing the data related to the manufacture of motor vehicles in Motor Asia Company Ltd. during the last 5 (five) years.

Year	2012	2013	2014	2015	2016
No. of motor vehicle manufactured	1 000	1 500	2 250	2 500	2 000

- Guide the students to construct a pictogram on the scale  = 500 vehicle.

A guideline to explain the subject matters :

- Various techniques of representing data can be used to present the data systematically for facilitating the process of data analysis and approaching at accurate conclusions.
- A rough idea about the changing pattern of data can be gained quickly though graphical/ pictorial representation of data.
- Various graphs and charts have been introduced in appropriate with each situation considering the factors such as the type of data, quantity of data and the purpose of representation of those data.

Those are.

1. Simple bar – charts
   (a) Simple vertical bar – charts
   (b) Simple horizontal bar – charts
2. Component bar – charts
3. Multiple bar – charts
4. Pie charts
5. Profile charts
6. Pictograms

- A chart through which the value of a sigle variable is represented by the 'height' or ' length' of a bar (rectangle), is known as a simple bar – chat. Simple bar charts are in two types as simple vertical bar – charts and simple horizontal bar – charts.
- Following facts should be taken into consideration in constructing a simple bar chart.
  - Keeping the width of every bar equally
  - Keeping an equal gap between every two bars which is not equal to the width of a bar
  - The same gap between two bars should be kept between the first bar and the vertical/horizontal axis as well.
  - When the scale used is small the value should be mentioned in writing on the top/ end of each bar.
  - Stating the headline, source of data, chart number etc.
- Profit/loss, budget difficit /surplus, balance of payments etc. taking plus / minues value can be represented using a deviation bar chart.
- Once a variable consists of few components the simple bar-charts are not sufficient to be used in representing data.
- Once a variable consists of more than one component in order to show the change in each component as well as the change in the total value of the variable the 'component bar charts' are used.
- Percentage component bar chart is constructed representing the percentage of each component.

- Having constructed the bars with equal height to the total value of the variable, each bar should be divided for each component according to the scale used.
- Each subdivision in every bar for separate components should be designed appropriately and a key guidance should be displayed outside the chart.
- Once it is needed to emphasize the change in total value of the phenomenon the component bar-chart is constructed, where as the change in each component is to be emphasized the multiple bar-chart is constructed, combining the componental bars together.
- Each and every bar for each component is constructed starting from the origine, (base-line) so that the comparison is quick and convenient.
- The chart, which is constructed on a circular plane in order to represent the relative importance of each component in the **entirety** related to a certain phenomenon of a particular variable, is known as a 'pie-chart'.
- There are few steps to be followed in constructing a pie chart as mentioned below.
- Expressing the value of each component as a percentage of the total value.
  - Computing the magnitude of each sector of the circle by multiplying each percentage value by 3.6 since $\left(\frac{1}{100} \times 360^0 = 3.6^0\right)$
- Constructing a circle with an appropriate diameter.
- Measuring each angle using a protractor.
- Constructing those angles (sectors) accurately.
- Stating the percentage of each sector in side or outside the chart.
- Constructing designs on the area of each sector representing each component.
- Completing the pie-chart with an appropriate headline, key guidance and chart No. etc…\

Profile chart

- A dioramic representation constructed to compare a specific condition of a particular variable against the general condition of that variable in the same co-ordinate plane is known as a 'Profile Chart'.

Ex : to represent the pattern of average monthly income of an employee working in a firm in comparison to the pattern of monthly expenses of Mr. Perera, who is working in the same institute.

Pictograms

- A diagramic representation constructed on the same co-ordinate plane using a suitable scaled figure (picture) which is very closer to the relevant variable, in such a manner that even the iliterated persons also can understand is known as a 'Pictogram'. Pictograms are also known as picture graphs.

Ex : 1. In a company where motor vehicles are manufactured, a pictogram can be constructed using a scaled picture of a vehicle for representing 100 vehicles.

2. In order to represent the manufacturing of bread in a bakery a pictogram can be constructed using a picture of 1 loaf of bread for 100 loaves of bread.

- Some practical instances where each of these charts and graphs are applicable in representing data are mentioned below.

- Simple vertical bar-charts.

(i) To represent the number of defectives found in each day from a manufacturing process during a period of 5 (five) days

(ii) To represent the annual total production of a factory for five years

Simple Horizontal bar-chart

- To represent the number of vehicles registered during the year 2016 in nine provinces in Sri Lanka
- To represent the number of employees working in each department of a company

N:B: In such situations where data are in categorical type once the terms are written horizontally, it facilitates the data users

Component bar-charts

1. To represent the number of tea bags in 3 categories for last 5 years by a tea bag seller.
2. To represent the annual income earned in five years from export industry of tea, rubber, and coconut.

Percentage component bar charts.

1. To represent the A/L results of few subjects in a school in 2016 with the percentage of each grade of A, B, C, S, F
2. To represent the composition of export income of a country for few years depicting the contribution of each export field in the entire annual export income.

Multiple bar – charts

1. To represent the relative changes in sales income of he first six months in 2016 in a leather ware manufacturing company where foot ware, travelling bags and hand bags are manufactured.

Pie-charts

1. To represent the A/L results of Business statistics in the year 2016 with the relative contribution of each grade as A, B, C, S, F.

- To represent the relative contribution of each sector in the total export income of a country in 2016.

Profile chart

- To represent the average daily sales revenue in 7 days of the week of Lottery Board and the average daily sales income of Lottery agencies in Colombo district in the same week.

Pictogram

- To represent the number of deaths reported in Western province from Dengu in the first six months of the year.

Assessment and Evaluation :

Provide with each set of data to the students and lead them to represent each using an appropriate technique.

1. Number of units provided annually in a machine with a life time of 5 years is mentioned below.

Year	number of units manufactured
2012	1 800
2013	2 400
2014	900
2015	1 200
2016	900

2. The values of production is 'Suwa Sala' Gament factory in each quarter in the year, 2015 are given in the following table.

Production Department	The Value of Production			
	Quarter I	Quarter II	Quarter III	Quarter IV
Baby suits	0.8	1.8	1.2	1.4
Ladies suits	1.4	2.0	1.6	1.5
Male suits	1.2	2.2	1.8	1.6
Total	3.4	6.0	4.6	4.5

3. The pattern of monthly expenditure of a normal family and a sophisticated family living in a particular colony is given below.

Types of expenses	Percentage of monthly expenditure	
	Of a normal family	Of a sophisticated family
1. Food & beverage	40	25
2. Health & protection	10	15
3. Social & cultural affairs	08	20
4. Transportation	12	16
5. Child education	23	14
6. Entertainment	07	10
Total	100	100

(4) Quantity of loaves of bread sold daily during a week in a bakery are as follows :

Day	Quantity of leaves of bread
Monday	150
Tuesday	160
Wednesday	250
Thursday	200
Friday	240
Saturday	150
Sunday	180

Competency 2.0 : Organizes and Reprersents the Business Data.

Competency Level 2.5 : Represents Business data graphically

No. of Periods : 12

Learning outcomes :

- Processes the data approproately in representing graphically.
- Constructs linear graphs naming the axes accurately.
- Describes the variations of data using linear graphs.
- Constructs the histogram and frequency polygon in relevant to the frequency distributions.
- Compares the histogram and frequency polygon.
- Constructs a curve (ogive) with relevant to cumulative frequency distributions.
- States the median of the distribution through observing the ogive.

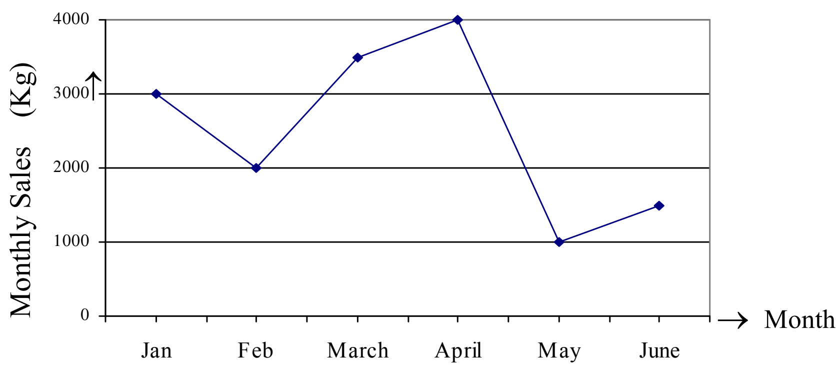
Instructions for lesson planning :

- Present the following assumed data to the class related to monthly sales of fish in the first 6 months of the year 2016 in Rajagiriya Sales Outlet of Fisheries Cooperation.

Month	Jan	Feb	March	April	May	June
Fish Sale (Kg)	3 000	2 000	3 500	4 000	1 000	1 500

- Make a discussion highlighting the following facts.
    - The given information have been processed through daily sales data of the outlet.
    - In order to construct a linear graph for given data a co-ordinate plane with horizontal axis – **X** for months and the vertical axis – **Y** for sales should be prepared according to an appropriate scale.
    - The graph should be plotted out marking **(x, y)** co-ordiantes on the co-ordinate plane.
    - The graph should be named using an appropriate heading.

Fisheries Corporation Sales Outlet Rajagiriya
Monthly Sales of Fish from January – June – 2016



- Let the students to construct the graph on a graph paper following the above guideline and inquire the following questions.
  - Name the month during which the highest quantity of sales is reported.
  - Name the month during which the lowest quantity of sales is reported.
  - What are the other facts that can be mentioned regarding the sales of fish ?

Answers :

  - The month with the highest sales is April
  - The month with the lowest sales is May
  - Other facts : Monthly sales of fish have been highly fluctuated.
- Call upon a student before the class and let him construct a histogram on the board for a given grouped frequency distribution.
- Lead him also to derive the frequency polygon on that histogram.

Activity 1 :

The data collected from 200 students questioning about the time taken to finish the school home – work at home in last evening, are mentioned in the following table.

Time (minutes)	25-30	30-35	35-40	40-45	45-50
Frequency	35	39	68	42	16

- Construct the histogram and the frequency polygon for the above data.
- Comment on the shape of the distribution.

Activity 02

Given below is a frequency distribution inculding data related to the height of 295 students.

Height (cm)	135-145	145-150	150-155	155-160	160-175
Frequency	40	40	75	65	75

- Construct the histogram and the frequency polygon for the above data.
- Comment on the shape of the distribution.

Answer : Activity - 02

Height (cm)	Frequency	Class width	Adjustment Frequency
135 - 145	40	10	$\frac{40}{10} \times 5 = 20$
145 - 150	40	05	$\frac{40}{5} \times 5 = 40$
150 - 155	75	05	$\frac{75}{5} \times 5 = 75$
155 - 160	65	05	$\frac{65}{5} \times 5 = 65$
160 - 175	75	15	$\frac{75}{15} \times 5 = 25$

Activity – 03

Construct the relative frequency histogram and relative frequency polygon using the data given in following frequency distribution.

Observed value-x	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	12	15	27	16	06	04

Answer : Activity - 03

Class Interval	Frequency (f)	
20 - 30	12	$\frac{12}{80} = 0.1500$
30 - 40	15	$\frac{15}{80} = 0.1875$
40 - 50	27	$\frac{27}{80} = 0.3375$
50 - 60	16	$\frac{16}{80} = 0.2000$
60 - 70	6	$\frac{6}{80} = 0.0750$
70 - 80	4	$\frac{4}{80} = 0.0500$
Total	80	= 1.000

- Pay attention of the students to the following situation to explain the cumulative frequency curve (ogive).
- Construct a graph curve on the board discussing with students assuming the total runs scored by Sri Lankan team at the end of each over in a cricket match with 20 overs.
- Inquire the following questions from the students with reference to that graph constructed on the board.
  - What is the total runs scored at the end of the 5^{th} over?
  - What is the total runs scored at the end of the 10^{th} over?
  - How many runs scored through the last 5 overs?
  - Is the run rate greater through the first 5 overs or the last 5 overs?
- Carry on a discussion highlighting the following facts.
  - Graphical representations similar to this type can be used for frequency distributions.
  - Such a graphical representation is known as a Cumulative Frequency distribution?

- Such graphical representations can be used for "Less than" cumulative frequency distributions as well as 'or more' 'cumulative Frequency Distributions.
- This graph is also termed as 'Ogive'

A Guideline to explain the subject matters :

- Representation of the movement of a variable in a linear formation in accordance with a given sacle is known as a graph
- Data in the long run can be represented very effectively using a linear graph
- The variation, trend etc. of a variable can be identified using a graph
- Linear graph can be used to identify the type of relationship between two variables
- Graphs are considered as bi-dimensional charts.
- Frequency distribution also can be represented graphically using histograms, frequency polygons and ogive curves etc.
- The set of joined rectangles constructed using the class width and frequency of a frequency distribution is known as a histogram.
- When the histogram is constructed for a frequency distribution with unequal size class intervals the adjusted frequency or the frequency density can be used.
- A frequency polygon can be constructed with or without the histogram considering the class mark and the frequency of each class interval.
- The area of the histogram and the area converted by the frequency polygon should be equal.
- Histogram and frequency polygon can be used to identify the shape of a frequency distribution.
- When it is needed to compare few frequency distributions the relative frequency polygon for each distribution can be plotted out on the same co-ordinate plane. (Relative frequency means that the ratio of the frequency of a class interval to the total frequency of the distribution)
- 'Less than' ogive and 'or more' ogive can be constructed using cumulative frequency distributions.
- 'Less than' ogive inclines from left to right.
- 'Or more' ogive declines from right to the left.
- 'Less than' ogive and or more ogive both intersect at median of the distribution, and median can be identified by drawing a perpendicular line from that intersected point to the horrizontal axis.
- An idea about the rate of cumulating the value of a distribution can be derived using ogive curve.

Assessment and Evaluation :

1. Fill the blanks in the 'Less than" Cumulative Frequency Column and 'or more' Frequency Column in the following frequency distribution.

Class Interval	Frequency	Cumulative frequency Less than"		Cumulative frequency or more'	
0-0	--	Less than 45.5	0	45.5 or more	60
45-55	08	Less than 55.5	8	55.5 or more	52
55-65	12	Less than 65.50	20	65.5 or more	....
65-75	20	Less than 75.5	....	75.5 or more	....
75-85	18	Less than 85.5	....	85.5 or more	....
85-95	02	Less than 95.5	....	95.5 or more	....
	60				

- Construct the "less than" Ogive using the "less than" cumulative frequency distribution.
- Construct the 'or more' Ogive using 'or more' cumulative frequency distribution.
- Plot both the Ogive curves on the same co-ordinate plane.

2. Following frequency distribution prepared using the data related to the daily sales income of a business firm has been given to you.

Sales income (Rs.000)	10-19	20-29	30-39	40-49	50-59	60-69
No. of days	12	20	36	28	18	06

(i) Plot the histogram on a graph paper for this data set.

(ii) Derive the number of dates less than Rs: 36 000 of daily sales.

(iii) Derive frequency polygon on the histogram you have constructed

(iv) Comment on the shape of the distribution using the histogram and frequency polygon.

(v) Prepare the 'less than' cumilative frequency distribution and 'or more' cumulative frequency distribution.

(vi) Construct both the ogives on the same co-ordinate plane.

(vii) Using those ogive curves find the minimum daily sales of the highest earning 30 days and the maximum daily sales of the lowest earning 30 days

Competency 2.0 : Organizes and Reprersents the Business Data.

Competency Level 2.6 :

No. of Periods : 08

Learning outcomes :

- Introduces the variables that can be represented using the 'Lorenze Curve".
- Constructs a "Lorenze Curve", processing data appropriately.
- Explains the abnormality (deviation) of distributing the relevent variable using a "Lorenze Curve".
- Describes how to compute the "Ginni" co-efficient in decision making.
- Gives instances where Lorenze Curve is applicable.
- Introduces the variables that can be depicted using an Z-chart.
- Interprets the 'Z-chart.'
- Constructs the Z-chart processing the data appropriately.
- Describes the uses of Z-chart in decision making.

Instructions for Lesson Planning

- Produce the following notices to the class.
    - There is a great deviation (anomality) in the distribution of income among various citizens in under-developed countries.
    - There is a great deviations in dispersion of the citizens of a country among various provinces.
    - There are anomalities of salaries among various employees and officers according to their posts.
    - There are disparities in distributing the market quota of an industry among the various firms involved in it.

- Hold a discussion highlighting the following facts.
    - National income of a country is suitable to be distributed fairly/uniformaly among the citizens of that country.
    - Uniformal distribution of citizens of a country among the provinces does not create problems in utilizing the resources.
    - Presence of severe anomalities in salary scales of the employees in institutions, controlling these institutions becomes problematic.
    - If the market quota of an industry distributes uniformly among all the firms involved in that industry the market competition can be kept under control.

- Therefore it would be better to study to which extent the variables that should be uniformaly distribted are apart from uniformity.
- Involve the students involve in the following activity to explain the chart of Lorenze Curve.
- Suppose that an income of Rs. 10 000/- is distributed among 10 members in four families as follows.

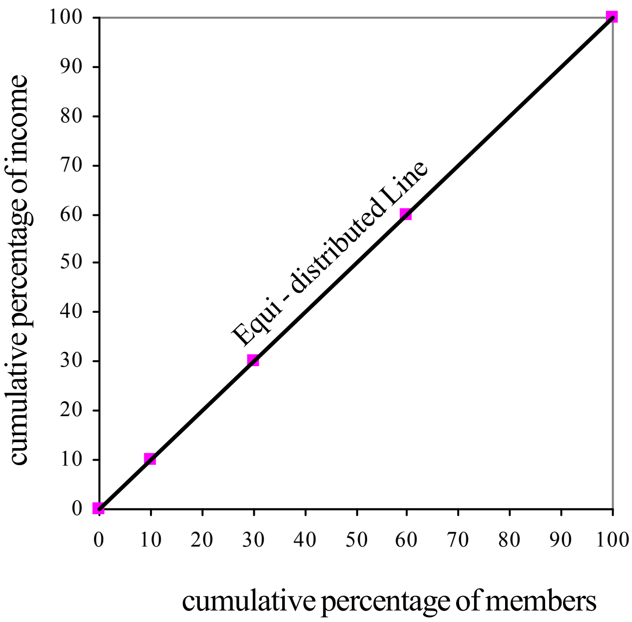
Family	Member	Amount
W	A	1 000
X	B	1 000
	C	1 000
Y	D	1 000
	E	1 000
	F	1 000
Z	G	1 000
	H	1 000
	I	1 000
	J	1 000
		10 000

- Inquire the followings from the students.
    - Express the number of members in 'W' family as a percentage of total number of members in these four families. Which percentage of the total amount distributed, has that family received?
    - Express the number of members in 'X' family as a percentage of total number of members in these four families. Which percentage of the total income distributed, has that family received ?

- Accordingly complete the following table with the students using the above mentioned data.

Family	No. of members	As a percentage of the total members %	Cumulative percentage of members %	Amount received Rs.	As percentage to the total amount %	Cumulative percentage of income %
X	1	10	10	1000	10	10
W	2	20	30	2000	20	30
Y	3	30	60	3000	30	60
Z	4	40	100	4000	40	100

- Plot the ordered pairs of cumulative percentage of members and the cumulative percentage of income in a square shaped co-ordinate plane and join those points as follows.



- Explain that the line drawn by the students is known as 'equi-distributed Line' and that is the main diagonal of the square shaped co-ordinate plane.
- Explain that it is hardly seen such distributions practically with no any deviation and most probably the distributions with abnormalities can be seen.
- Inform the students that Rs. 10 000 income has been actually distributed among those family members as follows.

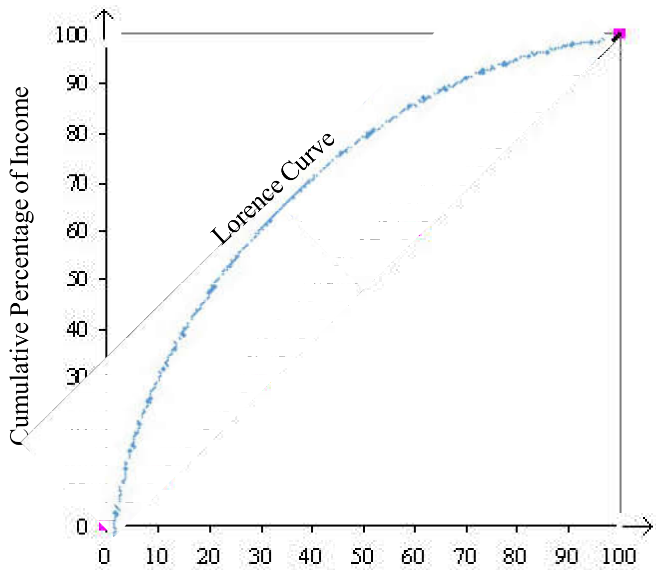
Family	Members	Amount (Rs.)
W	A	2 000
X	B	1 800
	C	1 700
Y	D	500
	E	600
	F	900
Z	G	400
	H	800
	I	600
	J	700
		10 000

- Inquire the following facts from the students.
  - Express the number of members in W family as a percentage of the total number of members and which percentage of total income has been distributed to that family?
  - Express the number of members in 'X' family as a percentage to the total number of members and which percentage of the total income has been 'distributed to the 'X' – family?
- Accordingly complete the following table discussing with the students.

Family	No. of members	Percentage of members to total members %	Cumulative percentage %	Amount received Rs.	percentage of amount received %	Cumulative percentage of income %
W	1	10	10	2 000	20	20
X	2	20	30	3 500	35	55
Y	3	30	60	2 000	20	75
Z	4	40	100	2 500	25	100

- Plot the ordered pairs of cumulative percentage of members and the cumulative percentage of amount received on the same co-ordinate plane used above and draw a smooth line joining the original and end points (0,0 and 100, 100) as well.

- Explain the fact that the line constructed accordingly is known as the Lorenze Curve.



- Divide the students into groups appropriately and involve them in the following activity.
  - Pay your attention to the set of data mentioned below.
  - Details related to the distribution of farming land in a country among the citizens living in various provinces.

Province	Population (Millions)	Extent of farming land distributed (Hectares)
A	8.0	400
B	5.0	640
C	2.0	800
D	1.5	320
E	4.5	200
F	9.0	520
G	4.0	640
H	16.0	4 480
	50.0	8 000

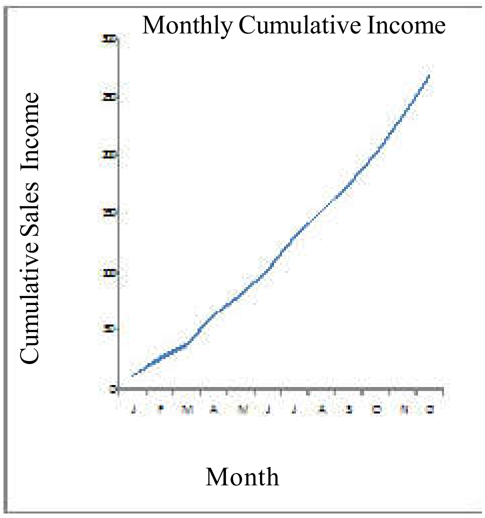
- Identify the two variables related to the data in this table.
- Express each value in the first variable as a percentage of the total value.
- Construct the cumulative value column adding each value in percentage value column step by step.
- Express each value in the second variable as a percentage of the total value of that variable.
- Construct the cumulative value column adding each value in percentage value column step by step.
- Prepare a square shaped co-ordinate plane on 0-100 scale on a graph paper available with you
- Name each axis separately representing one variable on horizontal axis and the other variably on vertical axis.
- Join the co-oridanates (0,0-100,100) and name that line as "Equi-distributed Line".
- Plot the points of cumulative percentage value of the variable represented on **Y** axis against the corresponding cumulative percentage value of the variable represented on **X** axis.
- Construct a smooth curve joining those points together with (0,0) and (100, 100) co-ordinates as well.
- Name that curve as the "Lorenze Curve.
- Write the values on vertical axis related to the points marked on the Lorenze curve against 20%, 50% and 80% on the horizontal axis separately.
- Comment on the disparities in distribution of the considered variable at each point.

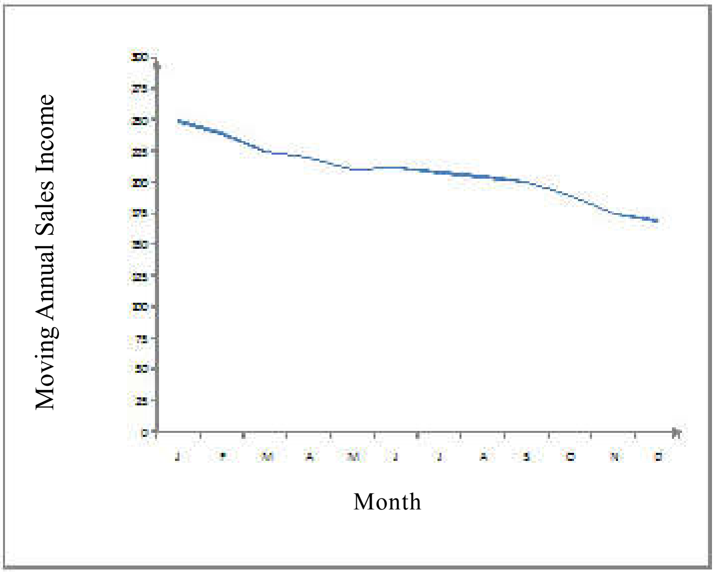
The table containing cumulative values and cumulative percentage values is given below.

Model Solution

Province	Population			Extent of Agro-lands distributed (Hectare)		
	No. of individual (million)	Percentage (%)	Cumulative percentage	Extent of land (hectare)	Percentage (%)	Cumulative percentage
A	8.0	16.0	16	400	5.0	5.0
B	5.0	10.0	26	640	8.0	13.0
C	2.0	4.0	30	800	10.0	23.0
D	1.5	3.0	33	320	4.0	27.0
E	4.5	9.0	42	200	2.5	29.5
F	9.0	18.0	60	520	6.5	36.0
G	4.0	8.0	68	640	8.0	44.0
H	16.0	32.0	100	4480	56.0	100.0
	50.0	100	-	8000	100.0	-

- Plot the Lorenze Curve on a graph paper using the details in this table.
- Produce the following three diagrams to the class and pay attention of the students to inquire the variations in them.





Carry on a discussion heighlighting the following facts.

- The fluctuations in monthly sales income of a business firm can be vividly demonstrated using the first diagram.
- Cumulative monthly income of a business firm can be demonstrated in the second diagram and that curve always inclines from left to right as shown.
- The moving annual income of a business firm is demonstrated in the third diagram and the progress of the business in the long run is evaluated using monthly values in two successive years.
- A diagram in the shape of an Z letter in English alphabet can be derived by plotting those three curves together on the same co-ordinate plane.
- Involve the students in following activity.
    - The number of vehicles sold monthly in Chandana traders company Ltd, during the last two years is mentioned in the following table.

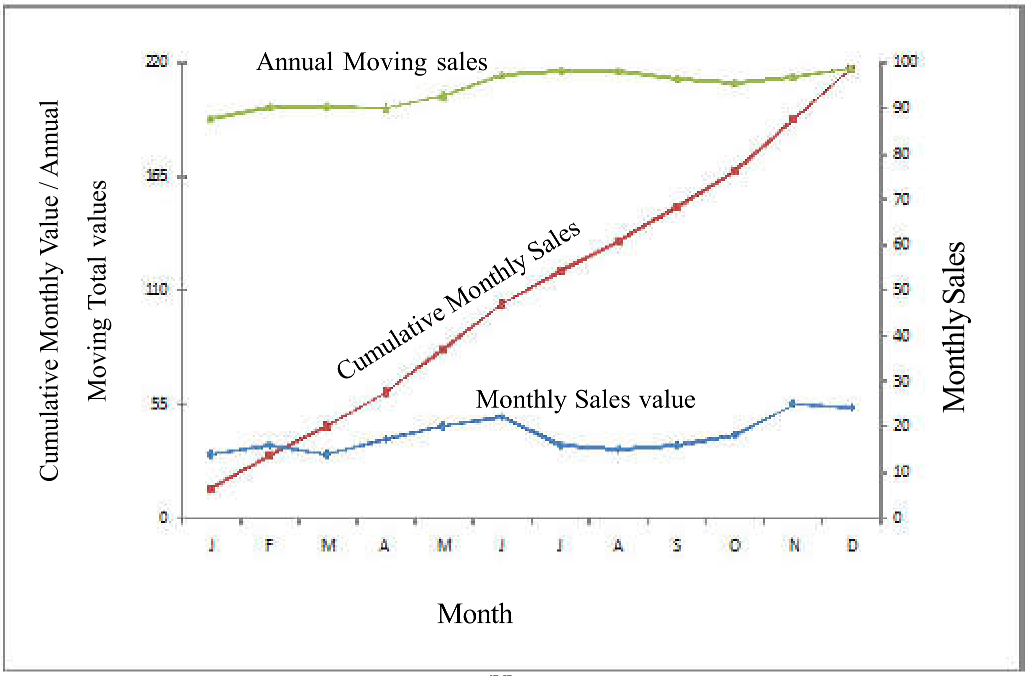
Month	Number of vehicles sold	
	2015	2016
January	08	14
February	10	16
March	14	14
April	18	17
May	14	20
June	12	22
July	14	16
August	15	15
September	20	16
October	20	18
November	22	25
December	20	24

- Compute the cumulative values in the year 2016 of the set of data received by you.
- Compute the total value of every year ended at the end of every month in the year 2016 (moving annual total value)

  Ex : The Moving Annual Total at the end of 31st January 2016 means that the total value for the year ended 31st January 2016.
- That means the sum of the values of the 12 months from February 1st 2015 to January 31st 2016.
- Follow the steps given below to construct the Z chart on a graph paper.
  - Divide the horizontal axis for the 12 months from January to December in equal ranges.
  - Number the left hand side vertical axis for cumulative monthly values and annual moving total values on a suitable scale.
  - Prepare a suitable scale line on the right hand side as well for normal monthly values in 2016. (it would be better to use a scale to represent 1 unit on right hand side vertical axis equal to the height of 5 units in left hand side vertical axis).
  - Plot the normal monthly values in 2016 on the co-ordinate plane following the right vertical scale and derive a curve joining those points.
  - Plot the corresponding cumulative monthly values on the co-ordinate plane following the left vertical axis and derive a curve joining those points.

- Plot the corresponding annual moving total value for each month on the co-ordinate plane following the left vertical axis.
- Mention what is communicated through each curve constructed by you according to the above guide lines.
- Produce the student findings to the entire class collectively and creatively.

Month	No. of vehicles sold		Cumulative monthly value	Annual Moving Total values
	2015	2016		
January	8	14	14	$187-8+14=193$
February	10	16	30	$193-10+16=199$
March	14	14	44	$199-14+14=199$
Apri	18	17	61	= 198
May	14	20	81	= 204
June	12	22	103	= 214
July	14	16	119	= 216
August	15	15	134	= 216
Septermbr	20	16	150	= 212
October	20	18	168	= 210
November	22	25	193	= 213
December	20	24	217	= 217
	187	217		



- A remarkable variation can be seen in monthly sales. From March to June a considerable inclinnings where as a declining in. July and August and again a rapid inclining towards the end of the year can be very well observed.
- A rate of inclining in sales can be seen according to the Cumulatve Monthly Value curve.
- According to the annual moving total curve a gradual inclining trend of annual sales is obvious.

A Guideline to explain the subject matters.

- The Lorenze Curve can be constructed by plotting out the cumulative percentage values on a square shaped co-ordinate plan.
- The chart, which was introduced by Lorenze to demonstrate the extent of a variable like national income, production quota of an industry and profit etc. being deviated from uniformity is known as the Lorenze curve.
- The extent of deviating the variable from its equi-distributed functioning can be vividly represented through plotting out the Lorenze curve in a square shape co-ordinate plane with (0-100) scale on both the axes.
- By plotting out several Lorenze curves on the same co-ordiante plane the deviation in each distribution can be compared.
- The main diagonal of this square joining the pairs of co-ordinates (0,0-100, 100) is termed as the 'Equi-distributed Line' and the perpendicular distance from any point on that line to horizontal axis and vertical axis is the same. (The angle to the horizon of the Equi-distributed line being 45^0 provides with the theoretical basis for this characteristic)
- The area between the Lorenze curve and the Equi-distributed line is expressed as a ratio to the area of the triangle containing the Lorenze curve and that ratio is known as the Gine co-efficient.

$$\text{Gini Co-efficent} = \frac{\text{Area between the Equi-distributed line and the Lorenze Curve}}{\text{The entire area of the triangle containing the Lorenze curve}}$$

- Some of the instances for variables that can be depicted through the Lorenze curve are as follows.
    - To represent the deviations in distributing the Gross Domestic Product of a country among the citizens of that country.
    - To represent the deviations in distributing the total cultivated land in a particular colony among the farmer families living in it.

- To represent how the business quota of an industry has been distributed among the individuals and firms involved in it.
- To represent the deviations in distributing the volume of ordinary shares issued by a company among the share holders.

- Once monthly data related to any single variable in two successive years are available, in order to analyse the movement of that variable in three different aspects, the linear chart constructed on the same co-ordinate plane is known as the Z-chart since the three curves are appeared in formation itself closer to the shape of the letter “Z” in English alphabet.
- The Z chart consists of the following three curves.

  1. Normal monthly value curve
  2. Cumulative value curve
  3. Moving annual summations curve (Moving annual Total value curve)

- Monthly fluctuations of the relevant variable are depicted by the normal monthly value curve.
- The cumulative (total) value of the variable as at any time unit given can be easily represented through the cumulative value curve. This curve always inclines from left to right.
- The moving annual total value (summation) curve is considered as the most significant curve in this chart. That curve creatively depicts the long term trend of the variable as if the business has been continued for 12 years using the details of only two years. The long term progress of the variable can be evaluated very significantly using this curve. There by an opinion about the prospective movement of the variable in the long run can be analyzed very clearly.

Assessment and Evaluation :

- Details about the distribution of the total product of the garment industry of a country among the firms involved in that industry in various regions.

Region	No of firms involved in grament filed	Annual income in the garment industry
A	54	80
B	72	128
C	90	240
D	36	80
E	27	64
F	27	64
G	29	80
H	25	64
	360	800

1. A mother and infant nutritional programme was launched in a province where the majority is poor, during the last two years. The monthly infant – mortality rate during the period of three months after child delivery has been observed as follows.

Month	Number of infant deaths occured monthly (for 1000 births)	
	2015	2016
January	28	20
February	26	23
March	24	22
April	22	23
May	22	20
June	18	20
July	16	14
August	15	12
September	12	10
October	12	08
November	10	08
December	08	06

(2) The number of child births occured monthly during the last two years in a particular hospital has been experienced as follows :

Month	Number of births	
	2015	2016
January	54	55
February	56	55
March	54	53
April	58	57
May	60	58
June	64	66
July	62	60
August	60	58
September	56	58
October	58	60
November	61	59
December	63	61

Competency 3.0 : Analyses the Business Data using the Techniques of Descriptive Statistics.

Competency Level 3.1 : Uses the measures of Central Tendency for analysis of data.

No. of Periods : 12

Learning outcomes :

- Interprets the 'Central Tendency'.
- Explains the qualities of a good measures of Central Tendency.
- Interprets the 'Mean' as a measures of Central Tendency.
- Computes the Mean for ungrouned and grouped data using the relevant formulae.
- Interprets the "Median' as a measure of Central Tendency.
- Computes the Median for ungrouped and grouped data using the relevant formulae.
- Interprets the 'Mode'.
- Computes the Mode for ungrouped and grouped data.
- Lists out the significant features in Mean as a good measures of Central Tendency.
- Lists out the significant features in Median.
- Lists out the significant features in Mode.
- Comparatively explains the advantages and disadvantages of the measures of Central Tendency.
- Provides instances for which each measure of Central Tendency is applicable.
- Explains the empirical relationships among Mean Median and Mode.

Guidelines for Lesson Planning.

- Hold a discussion with students regarding the following matters.
- Number of occupants at each house in a street having 21 houses is as follows.
- 2, 3, 4, 3, 3, 3, 3, 3, 4, 4, 8, 7, 8, 8, 5, 4, 7, 6, 6, 7, 7
- Prepare the array of data for the above data set.
- 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 6, 6, 7, 7, 7, 7, 8, 8, 8
- Inquire the number of occupants in the most number of families
- 3 members.
- Inquire the average number of occupants in a house.

  $\frac{105}{21} = 5$

- Referring to the above array of data what is the observation in the middle?

- The observation in the middle of the array is 4.
- Accordingly point out that a single value can be stated to represent a set of data and such values are known as the measures of Central Tendency.
- Explain that
  - The observation lies in a data set for most of the time (with the highest frequency) is 'Mode'
  - The value derived by dividing the sum of all the observations from the number of observations in the data set is the 'Mean'.
  - The observation lies in very middle of an array of data is 'Median'.
- Group the students appropriately and involve them in the following activity.

Activity – I

Data set – I

- The length of 12 iron nails drawn in random from a batch of iron nails manufactured in a machine is given below (cm).
  4.0, 3.8, 4.1, 4.0, 3.9, 3.8
  4.0, 4.0, 3.8, 3.9, 4.0, 3.8

Data set – 2

- Leaves obtained by 10 workers in a firm during the period of one year are as follows.
  15, 13, 15, 16, 14, 16, 15, 14, 17, 14
- Study the set of data you received very carefully.
- What is the Mode of that data set?
- Construct an ungrouped frequency distribution using that set of data.
- State the Mode using that ungrouped frequency distribution,
- Point out that there are data sets with no Mode or with two modes (bi-model) or with more than two modes (multi model)
- Involve them in the following activity to compute the Mode of a grouped frequency distribution.

Activity – 02

A frequency distribution containing the daily wages of 100 labourers working in a factory on day's pay basis is given below.

Wage Rs.	No. of labourers
501 - 550	04
551 - 600	15
601 - 650	35
651 - 700	29
701 - 750	10
751 - 800	07
	100

- Points out the fact that the Mode of such a frequency distribution can not be identified directly and easily, since the identify of original data is not revealed.
- Highlight the fact that it would be better to assume that the observation considered to be Mode will be contained in the class interval with the highest frequency.
- Point out how the frequencies of the Model class preceeding class and succeeding class can change based on the size of those class intervals.
- Point out that the following algebraic formula has been constructed to compute the Mode using the information related to Model class and the other two class intervals on either side.

$$\text{Mo} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) C$$

Mo – Mode

L_1 - Lower Class boundary of the Mode containing class.

Δ_1 - Difference between the frequencies of Mode containing class and the preceeding class. (35 - 15)

Δ_2 - Difference between the frequencies of Mode containing class and the succeeding class. (35 - 29)

C - Width of the Mode containing class. (650.5 - 600.5)

- Instructed the students to compute the Mode of the above frequency distribution using this formula.
- The Mode containing class interval is 601-650.

$$Mo = L_1 + \left(\frac{A_1}{A_1 + A_2}\right)C$$

$$= 600.5 + \left(\frac{20}{20+6}\right)50$$

$$= 600.5 + 38.46$$

$$= \underline{\underline{638.96}}$$

- Explain how to derive the Mode of a grouped frequency distribution using a histogram.

- A value in the range 600.5 – 650.5 is derived as Mode. Then the Mode is about 639.
- Involve the students in following activity in order to make them aware of the 'Median'.
- Provide with the data set which was used to find Mode in Activity I itself to the students having grouped them appropriately.
- Provide with the following instructions.
  - Prepare an array of data.
  - Find the Median.
- Prepare theungrouped Frequency distribution
- Derive the Median using the cumulative frequency of it.

- How to find Median constructing an ungrouped frequency distribution is as follows.

Data set – 1

Length of a nail (cm)	No. of Nails (f)	Cumulative Frequency (cf)
3.8	4	4
3.9	2	6
4.0	5	11
4.1	1	12
	12	

$$Md = \frac{(n+1)^{th}}{2} \textit{observation}$$

$$= \frac{13}{2} = 6.5^{th} \quad \textit{observation}$$

$$= 3.9 + 0.5 \quad (4.0 - 3.9)$$

$$= 3.9 + 0.5 \times 0.1$$

$$= 3.9 + 0.05$$

$$= \underline{\underline{3.95}}$$

Data set - 2

Number of Leaves	No. of. Workers (f)	Cumulative Frequency (f)
13	1	1
14	3	4
15	3	7
16	2	9
17	1	10
	10	

$$Md = \frac{(n+1)^{th}}{2} \textit{term}$$

$$= \frac{11}{2} = 5.5^{th} \textit{term}$$

$$Md = \underline{\underline{15}}$$

Provide with the following ungrouped frequecny distribution to the students.

Assessment marks x	No. of students Frequency x
3	02
4	03
5	07
6	14
7	08
8	03
9	02
10	01

- Guide them to prepare the less than cumulative frequency column.
- Point out that it is quite sufficient to consider the $\frac{n}{2}^{th}$ term as the Median of this distribution since the total number of observations is exceeding 30.
- Hence, lead the students to derive the Median of this distribution.

Answer :

Assessment marks x	No. of students Frequency f	Cumulative Frequency cf
3	02	02
4	03	05
5	07	12
6	14	26
7	08	34
8	03	37
9	02	39
10	01	40

- Total number of observations in this distribution = 40

- The Median Value $= \frac{n}{2}^{th}$ value

$= \frac{40^{th}}{2}$ value

$= 20^{th}$ value

- When we go through the cumulative frequency column, we can find that the first value exceeding half of the total number of observations is 26. Then all the 14 observations after the 12^{th} observation are sixes. Hence the Median of this distribution is 6.

- Point out the fact that to find the Median of a grouped frequency distribution is not that easy, it is too different to this method.

- Explain that the value at which the frequency distribution is seperated into two equal parts is the Median and further that the Median value can not be directly identified since the identity of original data is not revealed.

- In order to find the Median of a grouped frequency distribution provide with the distribution in Activity – 2 above to the students and compute the Median discussing with them.

Wage (Rs.)	No. of Workers fi	frequency cf
501 - 550	04	04
551 - 600	15	19
601 - 650	35	54
651 - 700	29	83
701 - 750	10	93
751 - 800	07	100
	100	

- Derive the cumulative frequency column
- Deriving the Median containg class

$$Md = \frac{n}{2} = \frac{100}{2} = 50^{th}$$

$= \underline{\underline{601 - 650}}$ worker containing class

- Find the Median using this formula.

$$Md = L_1 + \frac{\left(\frac{n}{2} - Fc\right)\ C}{fm}$$

$$= 600.5 + \left(\frac{\frac{40}{2} - 19}{35}\right) 50$$

$$= 600.5 + \frac{31}{35} \times 50$$

$$= 600.5 + 44.29$$

$$= \underline{\underline{644.79}}$$

L_1 = Lower class boundary of Median class

n = Total number of observations

F_c = Cumulative Frequency of the preceeding class

F_m = Frequency of Median class

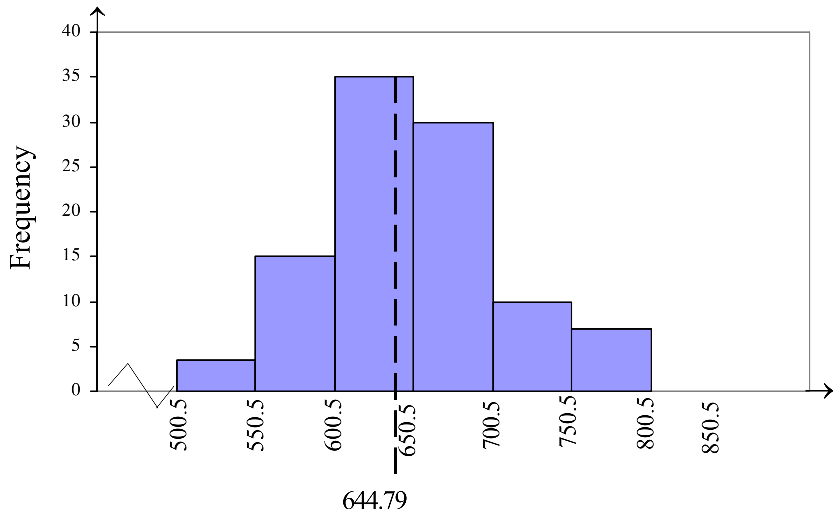
C = Size of the median class

- Explain how to derive the Median using the 'Ogive' as well.
- Construct the ' Less than' cumulative frequency distribution and derive the Median as follows.

Wage (Rs.)	Cumulative Frequency
Less than 500.5	0
Less than 550.5	4
Less than 600.5	19
Less than 650.5	54
Less than 700.5	83
Less than 750.5	93
Less than 800.5	100

- Construct the cumulative frequency distribution (Ogive) representing class boundary on the horizontal axis and the cumulative frequency on vertical axis.
- $\frac{n}{2} = \frac{100}{2}$ The salary of the employee at the 50^{th} place.
- The median is a value between 600.5 and 650.5.
- In the same way the Median can be derived using the 'or more' ogive as well which is constructed for the 'or more'cumulative freqecny distribution, referring to the 50^{th} place.
- Further both the ogive curves can be plotted on the same co-ordiante plane and the median can be derived by drawing a perpendical to the horizontal axis from the point of intersection.

- Explain the technique of deriving the Median using the histogram as well.
- Construct a histogram for the data set 2 and derive the Median as follows.



$$Md = 600.5 + \frac{31}{35} \times 50$$
$$= 600.5 + 44.29$$
$$= \underline{\underline{644.79}}$$

- Involve the students in the following activity to make them aware of the "Mean".
- Provide with the two sets of data given in the activity I above, having grouped the students in an appropriate way.
- Lead them to get the average value, dividing the sum of those data by total number of data.
- Introduce that this arithmetic average is defined as the 'Mean' in Statistics.

Answer :

- 1^{st} data set

4, 3.8, 4.1, 4, 3.9, 3.8, 4, 4, 3.8, 3.9, 4, 3.8

$$\bar{x} = \frac{\sum_{i=1}^{n_1} x_i}{n_1}$$

$$x_1 = \frac{4+3.8+4.1+4+3.9+3.8+4+4+3.8+3.9+4+3.8}{12}$$

$$= \frac{47.1}{12} = \underline{\underline{3.925}}$$

When considered 2^{nd} data set

$x_1 =$ 15, 13, 15, 16, 14, 16, 15, 14, 17, 14

$$\bar{X} = \frac{\sum X_i}{n_2} \qquad = \frac{149}{10} = \underline{\underline{14.9}}$$

- Construct the ungrouped frequency distributions for the above two data sets.
- Compute the mean of each distribution.
- Ungrouped frequency distribution for the first data set.

Length of a nail (X)	No. of nails (f)	fx
3.8	4	15.2
3.9	2	7.8
4.0	5	20.0
4.1	1	4.1
	12	47.1

$$\bar{X} = \frac{\sum fx}{\sum f}$$

$$= \frac{47.1}{12}$$

$$\bar{X} = \underline{\underline{3.925}}$$

- Ungrouped frequency distribution for the second data set.

No of days on leave (X)	No. of workers (f)	*fx*
13	1	13
14	3	42
15	3	45
16	2	32
17	1	17
	10	149

$$\overline{X} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{149}{10}$$

$$\overline{X} = \underline{\underline{14.9}}$$

- Explain that the mean of an ungrouped frequency distribution also can be computed as follows :
  - Lead the students to choose any observation as the assumed mean.
  - Guide them to derive the deviation of each observation from that assumed mean.

    $(d_i = x_i - A)$
  - Guide them to complete the $f_i \times d_i$ column.
  - Lead them to derive the mean of the distribution adjusting that mean deviation value to the assumed mean.
  - Explain how to compute the composite mean of the two data sets.
  - Composite mean is symbolized as $\overline{\overline{X}}$

$$\overline{\overline{X}} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

$$= \frac{12 \times 3.925 + 10 + 14.9}{12 + 10}$$

$$= \frac{47.1 + 149}{22}$$

$$\overline{\overline{X}} = \underline{\underline{8.91}}$$

- Provide with the grouped frequency distribution considered in Activity – 2 above to the students.
- Compute the Mean of that distribution discussing with the students.

Wage (Rs.)	No. of workers (f_i)	Mid point (x_i)	$f_i x_i$
501 - 550	04	525.5	2 102.0
551 - 600	15	575.5	8 632.5
601 - 650	35	625.5	21 892.5
651 - 700	29	675.5	19 589.5
701 - 750	10	725.5	7 255.0
751 - 800	05	775.5	3 877.5
801 - 850	02	825.5	1 651.0
	100		65 000

$$\overline{X} = \frac{\sum fx}{\sum f}$$

$$= \frac{65000}{100}$$

$$= \underline{\underline{650}}$$

- Direct the students to select any one of the class marks as the Assumed Mean (A) and to find the deviations from that Assumed Mean to each class mark as $d_i = x_i - A$
- Direct the students to recompute the Mean of this distribution using the formula $\overline{X} = A + \frac{\sum f_i d_i}{\sum f_i}$ having completed $f_i \times d_i$ column.

Solution :

Wages (Rs.)	No. of worksrs (f_i)	Class Mark x_i	$d_i = x_i \times A$	$f_i \times d_i$
501 - 550	04	525.5	- 150	- 600
551 - 600	15	575.5	- 100	- 1 500
601 - 650	35	625.5	- 50	- 1750
651 - 700	29	675.5	0	0
701 - 750	10	725.5	50	500
751 - 800	05	775.5	100	500
801 - 850	02	825.5	150	300
				- 3 850 + 1 300
				$\sum f_i d_i$ = - 2 550

Mean = $A + \frac{\Sigma f_i d_i}{\Sigma f_i}$

$= 675.5 + \left(\frac{-2550}{100}\right)$

$= 675.5 - 25.5$

$= \underline{\underline{650}}$

- Get the students exposed to calculate the Mean of a grouped frequency distribution with the same size class intervals very conveniently using the code system through the formula $\overline{X} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) C$ taking the class width as the common factor as $u_i = {d_i}/{c}$

Solution

Wages (Rs.)	No. of worksrs (f_i)	Class Mark x_i	$u_i = {d_i}/{c}$	$f_i \times u_i$
501 - 550	04	525.5	- 3	- 12
551 - 600	15	575.5	- 2	- 30
601 - 650	35	625.5	- 1	- 35
651 - 700	29	675.5	0	0
701 - 750	10	725.5	1	10
751 - 800	05	775.5	2	10
801 - 850	02	825.5	3	06
				- 77 + 26
				$\Sigma f_i u_i = \underline{\underline{-51}}$

Mean $\overline{X} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) C$

$= 675.5 + \left(\frac{-51}{100}\right) 50$

$= 675.5 - 25.5$

$= \underline{\underline{650}}$

- Discuss with the students the Significant properties of Mode, Median and Mean separately as a measure of central Tendency.
- Median and Mean separately measure of Central Tendency.
- Highlight the properties of a good measure of Central Tendency through the discussion.
- Discuss the relative advantages and disadvantages of the measures of Central Tendency.
- Discuss the most appropriate instance where each measure of measures of Central Tendency is applicable.
- Discuss the empirical relations among he measures of Central Tendency.

A Guideline to explain the subject matters :

- The Central Tendency which is a significant property representing the nature of the distribution of a particular variable is the trend of scattering the data about a certain point.
- The measures of Central Tendency are as follows :

  (i) Mean (ii) Median (iii) Mode

Deriving the 'Mode'.

- The observation appeared with the highest frequency in a data set or frequency distribution is known as 'Mode.'
- There may be data sets with a single mode bi-model or multi-model or no model in connection to a set of data or a distribution.
- The following formula can be used to calculate the Mode of a Grouped Frequency distribution.

$$M_o = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) C$$

Li = Lower Class Boundary of the Model Class.

Δ_1 = The difference between the frequency of the model class and frequency of the preceeding class.

Δ_2 = The difference between the frequency of the model class and the frequency of the succeeding class.

C = Width of the model class.

- Significant features of Mode as a measure of Central Tendency are as follows.
  - Mode is not influenced by extreme and end values.
  - Mode can be calculated even in a distribution with open class intervals.
  - Being a measure that can be derived graphically.
  - Applicability in expressing the average of qualitative data.
  - Being a categorical measure.

Deriving the Median

- The place value at which an array of data is separated into two equal parts is known as the Median.
- The Median of a data set consists of less than 30 observations is the value at $\left(\frac{n+1}{2}\right)^{th}$ place.
- Below mentioned steps can be followed to compute the Median of an ungrouped Frequency distribution.
  - deriving the cumulative frequency
  - the value at $\left(\frac{n}{2}\right)^{th}$ place as the Median
- Following formula can be used to compute the Median of a grouped frequency distribution.

$$Md = L_1 + \left(\frac{\frac{n}{2} - Fc}{fm} \right) C$$

- The class interval containing the $\left(\frac{n}{2}\right)^{th}$ value is the Median class.

L_1	=	Lower Boundary of the Median class
n	=	Number of observations
fm	=	Frequency of the Median class
Fc	=	Cumulative Frequency of the preceding class to the Median class
C	=	Width of the Median class

- Median can be derived using the Ogive curve
- Median can be derived using the histogram as well
- Several characteristics of Median as a measure of Central Tendency are as follows.
  - Being an identical (unique) measures which is always available
  - Not being a affected by end and extreme values
  - Applicability in interpreting the Median in ordinal data
  - Ability of computing when all the observations of a data set are unknown; in distributions with open class intervals
  - Being the most appropriate measures of Central Tendency for extremely skewed distributions

Computing the Mean

- The single value derived by dividing the sum of all the observations of a data set from the number of data in it, is known as the Mean of that data set.
- When each observation of a population consists of N number of data is given by $x_1, x_2, x_3..........x_N$ the mean of that population can be derived as

$$\mu = \frac{\sum_{i=1}^{n} x_i}{N}$$

- When each observation of *n* numbr of data is given as $x_1, x_{2..........} x$ the mean of the sample can be derived as

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Once the observations of an ungrouped frequency distribution are given as $x_1, x_2, x_3,......x_n$ associated with the corresponding frequencies as $f_1, f_{2..........} f_n$ the mean of such a distribution can be computed using the following formulae.

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \qquad \overline{X} = A + \frac{\sum_{i=1}^{n} f_i d_i}{n \sum_{i=1}^{n} f_i}$$

A = Assumed Mean

$d_i = x_i - A$ (deviations)

- Considering the class mid value of each class interval of a grouped frequency distribution as 'X" the mean can be computed using the following formulae.

$$\overline{X} = \frac{\sum_{i=1}^{K} f_i x_i}{\sum_{i=1}^{K} f_i} \qquad \overline{X} = A + \frac{\sum_{i=1}^{K} f_i d_i}{\sum_{i=1}^{K} f_i}$$

A = Assumed Mean

$d_i = x_1 - A$ (deviations)

- When the width of all the class intervals of a grouped frequency distribution are equal, the Median can be computed using the following formulae.

$$\overline{X} = A + \left(\frac{\sum_{i=1}^{K} f_i u_i}{\sum_{i=1}^{K} f_i} \right) C$$

A = Assumed Mean

$$u_i = \frac{d_i}{c} = \frac{x - A}{c}$$

C = Class width

- Significant features of Mean as a good measures of Central Tendency are as follows.
    - Representing all the observations in the data set
    - Ability of using as an algebraic function
    - Once the Means of few data sets are known separately a single Mean can be derived by combining them as follows

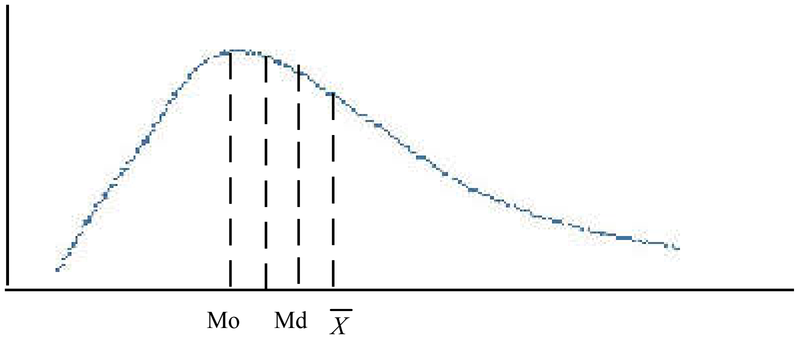
    $$\overline{X} = \frac{n_1 \overline{x} + n_2 \overline{x}_2 + ...n_k \overline{x}_k}{n_1 + n_2 + ... + n_k}$$

    - Being an identical measure
    - Being relatively a reliable measures of Central Tendency
- Following properties should be contained in a good measure of Central Tendency.
    - Representing all the data
    - Ability to be used as an algebra function
    - Being an identical measure
- Relative advantages and disadvantages of the measures of Central Tendency are as follows.
    - The Mean is a good representative measure since it considers all the observations of a data set when compared to the Mode and Median.
    - The Mean and Median are identical measures compared to Mode.

- Once qualitative data are available, it is meaningless to compute the Mean and therefore Mode and Median are suitable to measure the Central Tendency at such situations.
- When there are open classes in frequency distributions the Mean can not be computed, but the Mode and Median are useful at such situations.
- Mode and Median are not influenced by extreme and end values, but the Mean is highly influenced by those values.
- Mean can be used as an algebraic function but Mode and Median cannot be used in that manner.
- Mode and Median can be derived graphically, but the Mean cannot be derived in that manner.

- Instances where the Mean is applicable as a good measure of Central Tendency.
    - When the data are available in the form of a qualitative variable (Ratio scale).

      Ex : mass or length of objects, examination marks, sales income ….. etc.
    - In the absence of extreme and end values in the data set.
- Instances where the Median is applicable as a good measure of Central Tendency.
    - When the data related to qualitative variables are available.
        - Ex : Consumer desires, attitudes etc. …
    - When a measure free from extreme and end values is requires as an average,
    - When a measure is required to show the average of a frequency distribution with open class intervals.
- Instances where the Mode is applicable as a good measure of Central Tendency.
    - In case of a decision should be made the abundance of data.
      Ex : determining the shoe size.
    - When data related to qualitative variables are available (normal / ordinal data)
      Ex : Consumer taste, attitudes etc. …
    - When data are highly scatterred around a particular value.
- Empirical relationship existing among Mode, Median and Mean.
    - When data are symetrically spread.
      $\bar{x}(\mu) = Md = Mo$



- When the data are spread moderately in an asymmetrical shape.

  $\bar{x} - Mo = 3(\bar{x} - Md)$

Asseessment & Evaluation :

Activity - 1

- Collect data about the number of siblings having for the students in an Advanced Level class in the school.
- Collect data about the number of siblings having for the students in all the Advanced Level classes in the school.

- Prepare,

  - The array of data

  - Ungrouped frequency distribution of those data.

- Compute the mode, median and mean of those data

Activity – 2

- Collect data related to the height of the students in A/L classes in the school.
- Prepare a grouped frequency distribution using those data.
- Compute the mode, median and mean of that distribution.

Competency 3.0 : Analyses business data using Descriptive Statistical Techniques
Competency level 3.2 : Analyses data using specific measures of Central Tendency.
No. of Periods : 12

Learning outcomes :

- Points out the need of specific measures of Central Tendency for analysis of data.
- Interprets the Geometric Mean.
- Explains the instances where Geometric Mean is applicable.
- Computes the Geometric Mean.
- Points out the problems arisen in computing the Geometric Mean.
- Interprets the Harmonic Mean.
- Explains the instances where the Harmonic Mean is applicable.
- Computes the Harmonic Mean.
- Intepretes the Weighted Mean
- Explains what weighing is.
- Explains the instances the Weighted Mean is applicable.
- Computes the Weighed Mean.
- Arrays the Geometric Mean, Harmonic Mean and Arithmetic mean for a same set of data in accordance with the extent.
- Makes business decisions using Specific Measures of Central Tendency.

Instructions for lesson planing :

- Present the following data related to the sales volume of a business firm during the last three months.

Month	Qty. of Sales (units)
January	2 000
February	4 000
March	32 000

- Highlight the fact that the sales volume in February is twice that of in January and further the sales volume in March is eight times that of in February.
- Explain that in 'twice' and the 'eight times' are rates of increment.
- Let the students to find the average of those rates as they have already learnt.
- Point out that the sales volume of each month calculated using that average is not realistic as follows :

Average rate of increment	$= \frac{2+8}{2} = \underline{\underline{5}}$
Sales volume in January	= 2 000 units
Sales volume in February	= 2 000 x 5
	= 10 000 units
$\therefore$ Sales volume in March 10 000 x 5	= 50 000 units

- point out that it is more realistic to use the **n**th root of the products of 'n' number of increment rates as the Mean in such an instance.

The Square root of the product of the increments rates is $\sqrt{2 \times 8} = \underline{\underline{4}}$

Accordingly sales in January	= 2 000 units
Sales in February	= 2 000 x 4
	= 8 000 units
Sales in March	= 8 000 x 4
	= 32 000 units

- Explain that the average computed in this manner is known as Geometric Mean.
- Ensure that the Geometric Mean should be used to compute the Mean of data available in increment rates.
- Accordingly explain how to compute the Geometric Mean.
- Involve the students in the following activity.

Data set – 1

Annual increment rates of a particular industry in last four years are as follows.

0.2%, 2%, 5%, 8%

Data set – 2

Annual increment rates in Gross Domestic Product of a particular country in last three years are as follows :

1.5% 3.25% 5.12%

Data set – 3

Annual increment rates in the sales income of a business firm in last four years are as follows

5% 10% 20% -10%

- Compute the Geometric Mean of each data set.
- Guide the students to convert the percentage growth in to increments ratios.

  Ex : An increment of 2% means that the value 100 increased up to 102 and therefore the increment ratio is $\frac{102}{100}=1.02$
- Point out that usage of logarithmic tables would facilitate the calculation of Geomantic Mean.
- Point out that 10% minus increase rate means that 0.90 $\left(\frac{90}{100}=0.90\right)$ in the 3rd data set.

Solutions : Data set – 1

Since the increment rates are 0.2%, 2%, 5% and 8%.

$$(0.2\% = \frac{100.2}{100} = 1.002,\ \ 2\% = \frac{102}{100} = 1.02)$$

Hence, $x_1=1.002 \qquad x_2=1.02 \qquad x_3=1.05 \qquad x_4=1.08$

$$\therefore G=\sqrt[4]{x_1\times x_2\times x_3\times x_4}$$

$$G=\sqrt[4]{1.002\times1.02\times1.05\times1.08}$$

$$\log G=\frac{1}{4}\left(\lg 1.002+\lg 1.02+\lg 1.05+\lg 1.08\right)$$

$$\log G=\frac{1}{4}\left(0.0009+0.0086+0.0212+0.0334\right)$$

$$\log G=\frac{1}{4}\times 0.0641$$

$$=0.0160 \quad anti\log$$

$$=1.0375$$

$$\therefore G=\underline{\underline{3.75\%}}$$

$\therefore$ Annual average growth rate is 3.75%

Solutions : data set – 2

Since the increment rates are 1.5%, 3.25%, 5.12%

- Process the data as follows :

$$1.5\% = \frac{101.5}{100} = 1.015$$

$$3.25\% = \frac{103.25}{100} = 1.0325$$

$$5.12\% = \frac{105.12}{100} = 1.0512$$

x_1 = 1.015 x_2 = 1.0325 x_3 = 1.0512

$$G = \sqrt[3]{x_1 \times x_2 \times x_3}$$

$$G = \sqrt[3]{1.015 \times 1.0325 \times 1.0512}$$

$$\log G = \frac{1}{3}(\lg 1.015 + \lg 1.0325 + \lg 1.0512)$$

$$\log G = \frac{1}{3}(0.0065 + 0.0139 + 0.0217)$$

$$\log G = \frac{1}{3} \times 0.0421$$

$$= 0.0140 \quad anti \log$$

$$G = 1.033$$

$$G = \underline{\underline{3.3\%}}$$

$\therefore$ Annual average growth rate is 3.3%

Solutions data set – 3

Since the increment rates are 5%, 10%, 20% and -10%

Process the data as follows.

$$5\% = \frac{105}{100} = 1.05$$

$$10\% = \frac{110}{100} = 1.10$$

$$20\% = \frac{120}{100} = 1.20$$

$$-10\% = \frac{90}{100} = 0.90$$

$x_1 = 1.05$ $\quad x_2 = 1.10$ $\quad x_3 = 1.20$ $\quad x_3 = 0.90$

$$G = \sqrt[3]{x_1 \times x_2 \times x_3 \times x_4}$$

$$G = \sqrt[4]{1.015 \times 1.10 \times 1.20 \times 0.90}$$

$$\log G = \frac{1}{4}\left(\lg 1.015 + \lg 1.10 + \lg 1.20 + \lg 0.90\right)$$

$$\log G = \frac{1}{4}\left(0.0212 + 0.0414 + 0.0792 + \overline{1}.9542\right)$$

$$\log G = \frac{1}{4} \times 0.0960$$

$$G = 0.0240 \quad anti\log$$

$$G = 1.057$$

$$G = \underline{\underline{5.7\%}}$$

$\therefore$ Annual average growth rate is 5.7%

- Present the following problems to the students.

  If a similar amount is spent to buy pencil per Rs. 60 a dozen $\left(60^{Rsd-1}\right)$ and per Rs. 40 a dozen $\left(40^{Rsd-1}\right)$. Compute the average price of a dozen of pencils. Guide the students to derive the average price of a dozen of pencils on simple arithmetic mean

  that is $\frac{60 + 40}{2} = 50$
- Inquire in to the opinions of the students about the suitability of the answer derived.

- Point out that the total amount to be spent to buy both types of pencils is Rs. 240 if Rs. 120 are spent to buy each type of pencils.
- Point out that five dozens of pencils can be bought, when two dozens of pencils are bought for Rs. 120 from the first type of pencils and 3 dozens of pencils are bought for Rs. 120 from the second type of pencils.
- Accordingly, point out that the average price of a dozens of a pencils is . Rs, 48

$$\frac{Rs.240}{5}$$

- Point out that the total amount to be spent according to the average derived early is Rs. 250 (Rs. 50 x 5) and that method is not suitable for calculating the Mean for these data.
- Point out how to calculate the Harmonic Mean using the relevant formula.

$$H = \frac{2}{\frac{1}{60} + \frac{1}{40}}$$

$$H = \underline{\underline{48}}$$

- Hence, assure that the total amount spent to buy 5 dozens of pencils is Rs. 240 as per Harmonic Means.

$$\left(Rs.48 \times 5 = Rs.240\right)$$

- Ensure that the Harmonic Mean is more suitable to compute the mean of data available in reciprocals

- Present the following data set to the students and involve them in the Activity.

Data set – A

The speed of 4 vehicles moved accross a 100m long bridge are as follows $\left(Kmph^{-1}\right)$

30 60 20 50

Find the average speed of a vehicle on this journey in kilometers per hour. $\left(Kmph^{-1}\right)$

Data set – B

A particular business that involve in selling only 3 products as A B and C annually increases the unit price of each product by Rs. 10, Rs, 50 and Rs 20. What is the average annual increasing of the price of this firm?

Solution – Data set – A

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

$$= \frac{4}{\frac{1}{30}+\frac{1}{60}+\frac{1}{20}+\frac{1}{50}}$$

$$= 4 \times \frac{300}{36}$$

$$= \underline{\underline{33.33 kmph^{-1}}}$$

Solution – Data set – B

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

$$= \frac{3}{\frac{1}{10}+\frac{1}{50}+\frac{1}{20}}$$

$$= 17.647$$

- Present the following problem to the students.
  40% of the requirements for a particular job is theoretical knowledge and 60% consists of practical skills. Marks received by two candidates A and B who applied for this job are as follows.

Test	Marks Received	
	A	B
Theoretical Test	38	25
Practical Test	35	45
Total Marks	73	70

- Who is in the front according to the total marks of the two tests?
- Do you suppose that he is the best person for the relevant job?
- Lead a discussion highlighting the following facts using the responses of the students.
  - A greater importance should be given to the practical skills more than to the theoretical knowledge for this job.
    (For theoretical knowledge 40% for practical skills 60%)
  - The simple arithmetic mean which is calculated on the total marks is not suitable at such situations.
  - An average giving a higher weight for the data with relatively a higher importance should be computed.
  - The concept of 'Weighted Mean' is very important in this context.

Solutions :

- Average Marks of the candidate – A

$$= 38 \times \frac{40}{100} + 35 \times \frac{60}{100}$$

$$= 15.2 + 21 \quad = 36.2$$

- Average Marks of the candidate – B

$$= 25 \times \frac{40}{100} + 45 \times \frac{60}{100}$$

$$= 10 + 27 = 37$$

- Explains that computing the Mean in this manner is known as the 'Weighted Mean".
- Point out that the following formula can be used to compute the weighted Mean.

$$\bar{x}_w = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

- Weighted Mean Marks for the candidate – A

$x_1 = 38$ $\quad x_2 = 35$ $\quad w_1 = 40$ $\quad w_2 = 60$

$$= \frac{(38 \times 40) + (35 \times 60)}{40 + 60}$$

$$= \underline{\underline{36.2}}$$

- Weighted Mean Marks for the candidate – B

$x_1 = 25$ $\quad x_2 = 45$ $\quad w_1 = 40$ $\quad w_2 = 60$

$$= \frac{(25 \times 40) + (45 \times 60)}{40 + 60}$$

$$= \underline{\underline{37}}$$

- The candidate - B should be enrolled

- Involve the students in the following activity to understand the relationship among Arithmetic Mean. Geometric Mean, and Harmonic Mean.

Activity – 01

- Compute the Simple Arithmetic Mean, Geometric Mean and Harmonic Mean of the following data set.
  4 6 9
- Arrange those Means in ascending order.

Activity – 02

- Compute the Simple Arithmetic Mean, Geometric Mean and Harmonic Mean of the following data set.
  5 5 5
- Point out the relationship among Arithmetic Mean, Geometric Mean and Harmonic Mean according the result received.

Solution for Activity – 1

Simple Arithmetic Mean. $\bar{X} = \frac{4+6+9}{3} = \underline{\underline{6.33}}$

Geometric Mean $G = \sqrt[3]{4 \times 6 \times 9}$

$= 6$

$$H = \frac{3}{\frac{1}{4} + \frac{1}{6} + \frac{1}{9}}$$

$= \underline{\underline{5.68}}$

Accordingly $\bar{X} > G > H$

Solution of Activitiy - 2

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

$H = \frac{3}{\frac{1}{5} + \frac{1}{5} + \frac{1}{5}} = 5$

$G = \sqrt[3]{5 \times 5 \times 5} = 5$

Accordingly $\bar{X} = G = H$

A guidelines to explain the subject matters :

- Significant techniques should be followed to compute the Mean of different types of data.
- Geometric Mean can be used to compute the Mean of rates. The following formula is used to compute the Geometric Mean of **n** number of observations in a data set.

$$G = \sqrt[n]{x_1 \times x_2 \times .... x_n}$$

$$G = \sqrt[n]{\pi_{i=1}^{n} X_i}$$

n = Number of observations $\qquad$ $\pi =$ Product of n observations.

The following formula is used to compute the Geometric Mean using Logarithmic tables.

$$\log G = \sqrt[n]{\lg x_1 + \lg x_2 + ... \lg x_n}$$

$$\therefore \log G = \frac{1}{n}\left(\sum_{i=1}^{n} \log x_i\right)$$

- The Harmonic Mean can be used to compute the Mean of ratios presented as reciprocals. The following formula is used.

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

n = number of observations.

- Weighted Mean is used to compute the Mean of data in the different Relative Importance. The following formula is used.

$$\overline{X}_w = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

Wi – Weights assigned for each observation.

- Weighing means that assigning a significant value for each observation according to the relative importance of each.

- When calculated for the same data set.

Arithmetic Mean $\geq$ Geometric Mean $\geq$ Harmonic Mean
$\overline{X} \geq G \geq H$

Assessment and Evaluation

- A particular business firm involved in selling only the three products A, B and C, increase the price of each item by 10%, 50% and 20% respectively. When consided three of these products what is the percentage of annual price increasing of the products in this firm?

- When considered the consumption of meals in a family it seems to be consists of 40% grains 25% vegetables, 20% of fruits and 15% fish.

Grains	Rs. 100
Vegetables	Rs. 60
Fruits	Rs. 180
Fish	Rs. 300

What is the average price of one Kilogram of a meal in the above family.

Competency 3.0 : Analyses business data using Descriptive Statistical Techniques

Competency level 3.3 : Uses the measures of Relative Location to analyse the location of data.

No of Periods : 12

Learning outcomes :

- Introduces "Relative Location."
- Describes the measures of computing the Relative location.
- States the uses of the measures of Relative location.
- Computes the Quartiles, Deciles and Percentiles for ungrouped data sets and grouped frequency distributions.
- Makes decisions through, Quartiles Deciles and Percentiles.

Instructions for lesson planning :

- Inform the students that a particular student named **A** has scored 40 marks for a certain subject of the term test and inquire their opinions regarding that mark.
- After the responses of the students, state that number of students sat for that test was 23 and present the following mark list in ascending order to the class.

  4, 5, 8, 10, 13, 16, 17, 19, 20, 22, 25, 25, 25, 28, 28, 30, 33, 34, 36, 37, 39, 40, 42
- Inquire once again the opinion of the students regarding the 40 marks received by **A**.
- Point out that the student who received 40 marks is included in the group of highest quarter, if this series of marks separated in to 4 equal parts.
- Point out that the student who received 40 marks is contained the highest $\frac{1}{10}th$ of students if this series of marks is separated in to 10 equal parts.
- Point out that it is unable to have a good understanding about the data, referring to a single datrum in that data set absolutely.
- Explain that the knowledge about the Relative Location of data is required to have a correct awareness about the data set.
- Point out that the measures of Relative Location can be used to evaluate the significant points in a data set.
- Explain that the Quartiles, Deciles and Percentiles are used as the measures of Relative Location.
- Involve the students in the following activity to explain the Quartiles.
  - Produce the following list of scores gained by 11 cricketers in a game at a certain inning.

  38 16 24 40 58 90 30 14 41 39 61

- Call up on one of the students before the class and let him/her to arrange these scores in ascending order.

  14 16 24 30 38 39 40 41 58 61 90

- Conduct a discussion highlighting the following matters.
- The mid value of the array of data (6^{th} value) is the Median, (Md=39)
- The mid value in the part of data set separated to the left hand side of the Median is 24 and that value is called the first Quartile (Q_1)
- The mid value on the part of the data set separated to the right hand side of the Median is 58 and that value is known as the third Quartile (Q_3)
- The Median value is also named as the second Quartiles (Q_2)

- Involve the students in the following activity to compute the quartiles of an array of data.

Activity – 01

- Produce the following observations before the class, regarding the number of coconuts sold by a retailer during a period of 10 days and guide the students to compute the Quartiles.

Date	1	2	3	4	5	6	7	8	9	10
No. of coconut sold	28	34	30	19	26	42	38	50	44	43

Solution for Activity – 01

Array of data 19 26 28 30 34 38 42 43 44 50

Fist Quartile

$$Q_1 = \frac{1}{4}(n+1)^{th} \text{ observation}$$

$$Q_1 = \frac{1}{4}(10+1)^{th} \text{ observation}$$

$$= \frac{1}{4} \times 11^{th} \text{observation}$$

$$= 2.75^{th} \text{ observation}$$

$$\therefore Q_1 = 26 + 0.75(28-26)$$

$$= 26 + 0.75 \times 2$$

$$= \underline{\underline{27.5}}$$

$$Q2 = \frac{2}{4}(n+1)^{th} \text{ observation}$$

$$= \frac{2}{4}(10+1)^{th} \text{ observation}$$

$$= \frac{1}{2} \times 11^{th} \text{ observation}$$

$$= 5.5^{th} \text{ observation}$$

$$Q_2 = 34 + 0.5(38-34)$$

$$= 34 + 0.5 \times 4$$

$$= 34 + 2$$

$$= \underline{\underline{36}}$$

$Q_3 = \frac{3}{4}(n+1)^{th}$ observation

$= \frac{3}{4}(10+1)^{th}$ observation

$= \frac{3}{4} \times 11^{th}$ observation

$= 8.25^{th}$ observation

$Q_3 = 43 + 0.25(44 - 43)$

$= 43 + (0.25 \times 1)$

$= \underline{\underline{43.25}}$

- Involve the students in the following activity to compute the quartiles of an ungrouped frequency distribution.

Activity – 02

- Produce the following ungrouped frequency distribution to the class containing the marks of a school Based Assessment received by 25 students who study Business Statistics and lead them to compute Q_1, Q_2 and Q_3.

Marks	3	4	5	6	7	8	9	10
No. of students	2	3	3	7	4	3	2	1

Solution - Activity - 2

Marks x	No. of students f	Cumulative frequency
3	2	2
4	3	5
5	3	8
6	7	15
7	4	19
8	3	22
9	2	24
10	1	25

First Quartile

$Q_1 = \frac{1}{4}(n + 1)^{th}$ observation

$= \frac{1}{4}(25+1)^{th}$ observation

$= \frac{1}{4} \times 26^{th}$ observation

$= 6.5^{th}$ observation

When the cumulative frequency column is carefully checked, we can understand that every observation up to the 8^{th} observation after the 5^{th} is 5.

$Q_1 = 5 + 0.5(5-5)$ observation

$= 5 + 0.5 + 0$

$= 5$

(N:B: since both the 6^{th} and 7^{th} observations are the same (5)

$Q_2 = \frac{2}{4}(n+1)^{th}$ observations

$= \frac{1}{2}(25+1)^{th}$ observations

$= \frac{1}{2} \times 26^{th}$ observations

13^{th} observations

$\underline{\underline{Q_2 = 6}}$

$$Q_3 = \frac{3}{4}(25+1)^{\text{th}} \text{ observations}$$

$$= \frac{3}{4} \times 26^{\text{th}} \text{ observations}$$

$$= 19.5^{\text{th}} \text{ observations}$$

$$\therefore Q_3 = 7 + 0.5(8-7)$$

$$= 7 + 0.5 \times 1$$

$$\underline{\underline{Q_3 = 7.5}}$$

- Involve the students in the following Activity to compute the quartile of a grouped frequency distribution.

Activity – 3

- Produce the following set of data to the class containing the number of customers came to a bank in a period of 50 days, and lead the students to compute the quartiles of that distribution.

No of customers	No. of days
26-50	4
51-75	5
76-100	7
101-125	11
126-150	9
151-175	8
176-200	6

Activity – 03 : Solution

Class interval	Frequency	Cumulative frequency
26 - 50	4	4
51 - 75	5	9
76 - 100	7	16
101 - 125	11	27
126 - 150	9	36
151 - 175	8	44
176 - 200	6	50

Finding the Q_1 containing class $\frac{1}{4} \times n$ th observation containing class.

$\frac{1}{4} \times 50$ th observation containing class.

12.5 th observation containing class.

$= \underline{\underline{76 - 100}}$

$$Q_1 = L_1 + \left(\frac{\frac{n}{4} - Fc}{fQ_1} \right) C$$

$$Q_1 = 75.5 + \left(\frac{\frac{50}{4} - 9}{7} \right) 25$$

$$= 75.5 + \left(\frac{12.5 - 9}{7} \right) 25$$

$$= 75.5 + \frac{3.5}{7} \times 25$$

$$= 75.5 + 12.5$$

$$= \underline{\underline{88.0}}$$

$$Q_2 = L_1 + \left(\frac{\frac{2n}{4} - Fc}{fQ_2} \right) C$$

Finding the Q_2 containing class

$\frac{2}{4} \times 50$ observation containing class

25th observation containing class

101-125 – class interval

$$Q_2 = 100.5 + \left(\frac{\frac{2}{4} \times 50 - 16}{11} \right) 25$$

$$= 100.5 + \frac{9 \times 25}{11}$$

$$= 100.5 + 20.45$$

$$= \underline{\underline{120.95}}$$

$$Q_3 = L_1 + \left(\frac{\frac{3}{4} n - Fc}{fQ_3} \right) C$$

$$= \underline{\underline{155.19}}$$

Finding Q_3 containing class

$\frac{3}{4} \times 50$ observation contain class

37.5th observation contain

151-175 class interval

$$= 150.5 + \left(\frac{37.5 - 36}{8} \right) 25$$

$$= 155.19$$

Deciles

- The array of marks received by 29 students for a particular subject at the term test given below.

4	5	8	10	13	16	17	20	22	25
25	25	28	28	30	34	37	39	40	42
43	45	45	46	48	50	51	52	55	

- Write the above marks on the board.
- Call upon a student and let him/her to mark the places where the data set is divided (separated) in to 10 equal parts.
- Inquire the student how many such places are there.
- Inquire the students each of those place values.
- Explain that those are the Deciles.
- Involve the students in the following activity to compute the Deciles of an array of data.

Activity – 04

The observations related to the number of coconuts sold by a retailer are as follows.

28 25 20 34 30 19 26 30 42 38 40
20 43 50 44

* Compute the following deciles of the above data set.

The third Decile (D_3)

The fifth Decile (D_5)

The seventh Decile (D_7)

Activity – 4 : Solution

The array of data is as follows.

19 20 20 25 26 28 30 30 34 38 40
42 43 44 50

$$D_3 = \frac{3}{10} \times (n+1)^{\text{th}} \text{ obervation}$$

$$= \frac{3}{10} \times 16^{\text{th}} \text{ obervation}$$

$$= 4.8^{\text{th}} \text{ obervation}$$

$$= 25 + (26 - 25) \times 0.8$$

$$= 25 + 0.8 \times 1$$

$$= \underline{\underline{25.8}}$$

$D_5 = \frac{5}{10} \times (n+1)$ th observation

$= \frac{5}{10} \times 16$ th observation

$= 8$ th observation

$D_5 = \underline{\underline{30}}$

$D_7 = \frac{7}{10} \times (n+1)$ th observation

$= \frac{7}{10} \times 16$ th observation

$= 11.2$ th observation

$= 40 + 0.2\ (42 - 40)$

$D_7 = \underline{\underline{40.4}}$

- Involve the students in the following activity to compute the deciles in an ungrouped frequency distribution.

Activity – 5

- Produce the following ungrouped frequency distribution to the class, containing the assessment marks of 40 students.

Marks received	3	4	5	6	7	8	9	10
No. of students	4	4	7	12	6	3	2	2

Let them compute D_1, D_2, D_5, D_9

Activity - 5 : Solution

Marks	No. of students	Cumulative frequency
3	4	4
4	4	8
5	7	15
6	12	27
7	6	33
8	3	36
9	2	38
10	2	40

$D_1 = \frac{1}{10} \times (n+1)$ th observation

$= \frac{1}{10} \times 40^{th}$ observation

$= 4$ th bservation

$= \underline{\underline{3}}$

$D_2 = \frac{2}{10} \times 40^{th}$ th observation

$= 8^{th}$ observation

$= \underline{\underline{4}}$

$D_5 = \frac{5}{10} \times 40^{th}$ observation

$= 20^{th}$ observation

$D_5 = \underline{\underline{6}}$

$D_9 = \frac{9}{10} \times 40^{th}$ observation

$= 36^{th}$ observation

$D_9 = \underline{\underline{8}}$

- Involve the students in the following activity to compute the deciles of a grouped frequency distributions.
- Pay the attention of students to the grouped frequency distribution considered in the above activity 3 and lead them to compute D_1, D_5, and D_8

Activity – 6 : Solution

D_1 – Containing class

$= \frac{1}{10} \times 50^{th}$ observation containing class

$= 5^{th}$ observation containing class

= 51-75 class interval.

$$D_1 = L_1 + \left(\frac{\frac{n}{10} - Fc}{fD_1} \right) C$$

$$= 50.5 + \left(\frac{\frac{50}{10} - 4}{5} \right) 25$$

$$= 50.5 + 5.0$$

$$= \underline{\underline{55.5}}$$

D_5 - Containing class

$= \frac{5}{10} \times 50^{th}$ observation containing class

$= 25^{th}$ observation containing class

= 101-125 class interval

$$D_5 = L_1 + \left(\frac{\frac{5n}{10} - Fc}{FD_5} \right) C$$

$$= 100.5 + \left(\frac{25 - 16}{11} \right) 25$$

$$= 100.5 + \frac{9}{11} \times 25$$

$$= \underline{\underline{120.95}}$$

$$D_1 = L_1 + \left(\frac{\frac{8n}{10} - fc}{FD_8} \right) C$$

$$150.5 + \left(\frac{\frac{8 \times 50}{10} - 36}{8} \right) 25$$

$$= 150.5 + \frac{4}{8} \times 25$$

$$= 150.5 + 12.5$$

$$= \underline{\underline{163}}$$

Finding the D_8 - Containing class

$= \frac{8}{10} \times 50^{th}$ observation containing class

$= 40^{th}$ observation containing class interval

= 151-175 class interval

Percentiles

- Raise the following questions to the students to introduce 'percentiles.'
  - How many place values are there at which a distribution is seperated into 100 equal parts?
  - How can those place values be named as?
- Involve the students in following activity to compute the percentile of an ungrouped frequency distribution.

Activity – 7

- Assessment marks of 50 students are given in the following ungrouped frequency distribution.

Marks	3	4	5	6	7	8	9	10
No. of students	4	5	8	12	10	6	3	2

Compute the following percentiles.

(i) Tenth percentile (P_{10})

(ii) Fiftyth percentile (P_{50})

(iii) Ninetyth percentile (P_{90})

Activity – 07 : Solution

Marks	No. of students	Cumulative frequency
3	4	4
4	5	9
5	8	17
6	12	29
7	10	39
8	6	45
9	3	48
10	2	50

$P_{10} = \frac{10n}{100}$ th observation

$= \frac{10 \times 50}{100}$ = 5^{th} observation

$= \underline{\underline{4}}$

$P_{50} = \frac{50n}{100}$ th observation

$= \frac{50 \times 50}{100}$ = 25^{th} observation

$= \underline{\underline{6}}$

$9_{90} = \frac{90n}{100}$ th observation

$= \frac{90 \times 50}{100}$ = 45^{th} observation

$= \underline{\underline{8}}$

- Involve the students in following activity to compute the percentiles of a grouped frequency distribution.
- Provide with the grouped frequency distribution in activity 3 and guide them to compute the following percentiles.
- 25th percentile (P_{25})
- 50th percentile (P_{50})
- 75th percentile (P_{75})

- Compare the first quartile second quaritle and the third quartile computed in activity-3 with the percentile values computed above.

Solution - Activity 08

No of customers (class intervals)	No. of days (frequency)	Cumulative frequency
26-50	4	4
51-75	5	9
76-100	7	16
101-125	11	27
126-150	9	36
151-175	8	44
176-200	6	50

$$P_{25} = L_1 + \left(\frac{\frac{25n}{100} - Fc}{fp_{25}} \right) C$$

P_{25} - Containing class

$$= \frac{25 \times 50}{100} = 12.5$$

$$= 75.5 + \left(\frac{\frac{25 \times 50}{100} - 9}{7} \right) 25$$

= 12.5th observation containing class interval

= That is 76 - 100

$$= 75.5 + \frac{3.5}{7} \times 25$$

$$= \underline{\underline{88}}$$

P_{50} containing class $= \dfrac{50 \times 50}{100}^{th}$ observation containing class interval

$= 25^{th}$ observation containing class interval

101 - 125

$$P_{50} = L1 + \left(\frac{\frac{50n}{100} - Fc}{fP_{50}} \right) C$$

$$P_{50} = 100.5 + \left(\frac{\frac{50 \times 50}{100} - 16}{11} \right) 25$$

$$P_{50} = \underline{\underline{120.95}}$$

P_{75} containing class $= \dfrac{75 \times 50}{100}^{th}$ observation containing class

$= 37.5^{th}$ observation containing class

$= 101$ - 125 class interval

$$P_{75} = L_1 + \left(\frac{\frac{75n}{100} - Fc}{fP_{75}} \right) C$$

$$= 150.5 + \left(\frac{37.5 - 36}{8} \right) 25$$

$$= \underline{\underline{155.19}}$$

Hence $Q_1 = P_{25}$, $Q_2 = P_{50}$ $Q_3 = P_{75}$

A guideline to explain the subject matters :

- In addition to the measures of central tendency, the measures of relative location are also required to identify (the non-central locations) relative to all the observations.
- The measures which are used to assess the significant points in a data set are known as measures of relative location.
- Accordingly the measures used to determine the relative location of a single value in respect of a group of values are known as the measures of relative location.
- Three types of such measures of relative location are as follows.
  - Quartiles
  - Deciles
  - Percentiles

Quartiles

- The three place values at which a particular array of data or a frequency distribution is separated into four equal parts are called quartiles.

First Quartile $Q_1 = \frac{1}{4}$ (n+1)th observation

Second Quartile $Q_2 = \frac{1}{2}$ (n+1)th observation

Third Quartile $Q_3 = \frac{3}{4}$ (n+1)th observation

(N.B. **n** is the total number of observations in the data set)

- If there are 30 or more observations in a data set, the quartiles of an array of data or an ungrouped frequency distribution can be derived as follows :

$$Q_1 = \frac{1}{4} \times n^{th} \text{ observation}$$

$$Q_2 = \frac{1}{2} \times n^{th} \text{ observation}$$

$$Q_3 = \frac{3}{4} \times n^{th} \text{ observation}$$

- The following formulae are used to compute the quartiles of a grouped frequency distribution.

$$Q_1 = L_1 + \left(\frac{\frac{n}{4} - Fc}{fQ_1} \right) C \qquad Q_2 = L_1 + \left(\frac{\frac{2n}{4} - Fc}{fQ_2} \right) C$$

$$Q_3 = L_1 + \left(\frac{\frac{3n}{4} - Fc}{fQ_3} \right) C$$

L_1 = Lower class boundary of the class interval containing the relevant quartile

n = Total number observations in the distribution

fc = The cumulative frequency less than the lower class boundary of the class interval containing the relevant quartile

f_Q = The frequency of the class interval containing the relevant quartile

C = Size (width) of the class interval containing the relevant quartile

Deciles

The nine place values where a frequency distribution is separated into ten equal parts are known as deciles. The deciles less than 30 observations of data or in an ungrouped frequency distribution can be computed as follows.

The first decile $D_1 = \frac{1}{10}(n+1)^{th}$ Observation

The second decile $D_2 = \frac{2}{10}(n+1)^{th}$ Observation

The ninth decile $D_9 = \frac{9}{10}(n+1)^{th}$ Observation

- Deciles in an array of data or ungrouped frequency distribution with 30 or more observations can be found as follows.

The first decile $D_1 = \frac{1}{10} \times n^{th}$ Observation

The second decile $D_2 = \frac{2}{10} \times n^{th}$ Observation

The ninth decile $D_9 = \frac{9}{10} \times n^{th}$ Observation

- Following formulae can be used to derive the deciles of a grouped frequency distribution (The principle of deriving the Median of such a distribution is applied here as well)..

- The first decile $$D_1 = L_1 + \left(\frac{\frac{n}{10} - Fc}{fD_1} \right) C$$

- The second decile $$D_2 = L_1 + \left(\frac{\frac{2n}{10} - Fc}{fD_2} \right) C$$

- The ninth decile $$D_9 = L_1 + \left(\frac{\frac{9n}{10} - Fc}{fD_9} \right) C$$

N.B.

L_1 = Lower class boundary of the class interval containing the relevant decile

n = Total number of observations

F_c – Cumulative frequency less than the lower boundary of the class interval containing the relevant decile

f_D = Frequency of the class interval containing the relevant decile

C – Width of the class interval containing the relevant decile

Percentiles

- 99 place values at which a frequency distribution is separated to 100 equal parts are called percentiles.
- The percentiles of an ungrouped frequency distribution can be computed as follows.
- The first percentile $P_1 = \frac{1}{100} \times n^{th}$ observation

  Accordingly any percentile can be found.

- The percentiles of grouped frequency distribution can be found as follows.

$$P_1 = L_1 + \left(\frac{\frac{In}{100} - Fc}{fp_1} \right) C$$

- Accordingly this formula can be used to find any percentile with necessary adjustments.
  Eg : 64th percentile

$$P_{64} = L_1 + \left(\frac{\frac{64n}{100} - Fc}{fp_{64}} \right) C$$

L_1 = Lower class boundary of the class interval containing the relevant Percentile

n = Total number of observations in the relevant distribution

F_c – Cumulative frequency less than the lower boundary of the class interval containing the relevant Percentile

F_p - Frequency of the class interval containing the relevant Percentile

C – Width of the class interval containing the relevant Percentile

- Following relations are found in measures of Relative Location.

(I) $Q_2 = D_5 = P_{50}$

(2) $Q_1 = P_{25}$

(3) $Q_3 = P_{75}$

(4) $D_1 = P_{10}$

(5) $D_2 = P_{20}$ etc . . .

Competency 3.0 : Analyses Business Data using the Techniques of Desciptive Statistics.

Competency Level 3.4 : Uses the Measures of Dispersion for Analysis of Data.

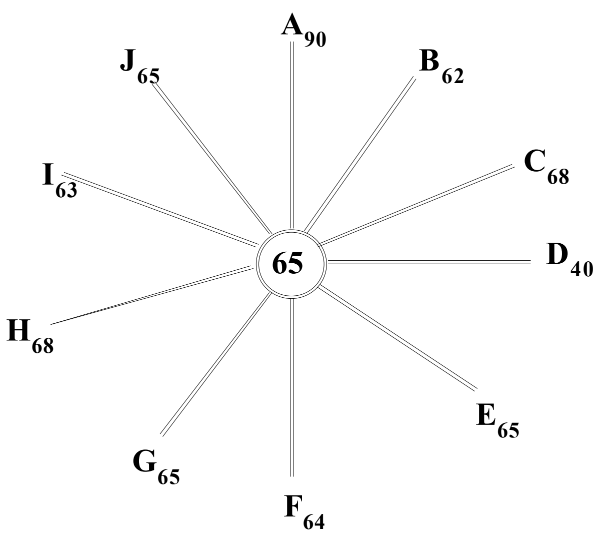
No. of Periods : 12

Learning outcomes :

- Interprets the' Dispersion"
- Points out the uses of computing the Dispersion.
- Lists out the measures used to evaluate the Dispersion.
- Computes Range, Quartile Deviation, Mean Deviation, Variance and standard Deviations in ungrouped frequency distributions and grouped frequency distributions.
- Introduces the 'Relative Dispersion'.
- Explains the need of evaluating the Relative Dispersion.
- Evaluates the Relative Dispersion using the co-efficient of variation.
- Standardizes the data using-z- value.
- Makes Business decisions using the measures of dispersion.

Instructions for Lesson Planning :

- Produce the following diagram related to the marks of 10 students for statistics.



- Compute the mean marks of the 10 students.
- Ascertain that the value mentioned in the middle of the diagram is the mean.
- State the difference between the marks of each student and the Mean.

90 - 65 = 25	64 - 65 = -1
62 - 65 = -3	65 - 65 = 0
68 - 65 = 3	68 - 65 = 3
40 - 65 = -25	63 - 65 = -2
65 - 65 = 0	65 - 65 = 0

- Compute the Mean of those differences.
- Ascertain that Mean of those differences is Zero.

$$\sum_{i-1}^{n} (x_1 - \bar{x}) = 0$$

- Ascertain that there are three students who received marks equal to the Mean and the marks 90 and 40 are highly deviated from the Mean.
- Mention the uses of knowing about the dispersion of data.
- Descirbe the measures that can be used to measure the deviation or dispersion of data.

Give out the following two sets of data to the students and let them find the Range of each.

Data set – I

Given below are marks received for Mathematics by 10 students in a class.

65, 70, 62, 90, 92, 50, 48, 32, 60, 71

Data set – 2

Monthly salaries of 100 employees working in a Business firm are given in the following distribution (Rs. 000).

Salary (Rs. 000)	No. of employees f_i
5 - 9	11
10 - 14	20
15 - 19	35
20 - 24	20
25 - 29	08
30 - 34	06
	100

- Derive the Range of each data set mentioned above.
- List out uses and limitations of Range.

Solution for Activity – I

- **Data set – I**

  Range (R) = Maximum value – Minimum value

  = 92 – 32 = 60

- **Data set – 2**

  The Range – (R) can be derived using two ways.

  1. Using class boundaries

     UCB of the highest class interval – LCB of the lowest class interval

     34.5 – 4.5 = 30.0

  2. Using the class mid points.

  Mid points of the highest class – Mid point of the lowest class

  32 – 7 = 25

- Pay the attention of students about the need of an another measure to evaluate the dispersion of a data set pointing out the weak point of Range, since it refers only for two values.
- Therefore point out that the Quartile Deviation is a better measure of dispersion relative to Range.

Activity – 2

- Derive the first Quartile (Q_1) and the third Quartile (Q_3) for the above mentioned two data sets.
- Derive the Quartile Deviation by dividing the answer received after subtracting the first quartile from third quartile by 2.
- Explain the uses and limitations in Quartile Deviation relative to the Range.

Solution for Activity – 2

Data set – 1

32, 48, 50, 60, 62, 65, 70, 71, 90, 92

$Q_1 = \frac{1}{4}(n+1)$ th observation

$= \frac{1}{4} \times 11 = \frac{11}{4}$ th observation

$= 2.75$ th observation

$Q_1 = 48 + (50 - 48)0.75$

$= 48 + 1.5$

$= \underline{\underline{49.5}}$

$Q_3 = \frac{3}{4}(n+1)$ th observation

$= \frac{3}{4} \times 11$ th observation

$= 8.25$ th observation

$Q_3 = 71 + (90 - 71)0.25$

$= 71 + 4.75$

$= \underline{\underline{75.75}}$

Quartile Deviation

$$QD = \frac{Q_3 - Q_1}{2}$$

$$= \frac{75.75 - 49.5}{2}$$

$$= \underline{\underline{13.125}}$$

Data set – 2 Quartile Deviation can be found as follows.

Salaries (Rs.000')	No. of employees f_1	Cumulative Frequency Fc
5 - 9	11	11
10 - 14	20	31
15 - 19	35	66
20 - 24	20	86
25 - 29	08	94
30 - 34	06	100

Q_1 Containing class interval

$$= \frac{1}{4} \times n^{th}$$

$$= \frac{1}{4} \times 100^{th}$$

$$= 25^{th}$$

$$= L_1 + \left[\frac{\frac{n}{4} - fc}{fQ_1} \right] C$$

$$= 9.5 + \left(\frac{25 - 11}{20} \right) 5$$

$$= 9.5 + \frac{14 \times 5}{20}$$

$$= 9.5 + 3.5$$

$$= \underline{\underline{13}}$$

Q_3 Containing Class Interval

$= \frac{3}{4} \times 11^{th}$ observation Containing Class

$= \frac{3}{4} \times 100^{th}$ observation Containing Class

$= 75^{th}$ observation Containing Class is 20-24

$$Q_3 = L_1 + \left[\frac{\frac{3n}{4} - fc}{fQ_3} \right] C$$

$$= 19.5 + \left(\frac{75 - 66}{20} \right) 5$$

$$= 19.5 + \frac{9 \times 5}{20}$$

$$= 19.50 + 2.25$$

$$= \underline{\underline{21.75}}$$

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$= \frac{21.75 - 13.00}{2}$$

$$= \underline{\underline{4.375}}$$

Activity – 3

- Derive the Means for each data set above separately.
- Ascertain that the sum of the real Mean diviations derived by subtracting Mean from each observation is always equal to zero.
- Derive the deviation of each value from the Men absolutely.
- Derive a quantitative value for Mean Deviations by dividing the sum of those absolute deviations by the total number of observations.
- Explain the advantages and disadvantages of Mean Deviation Relative to the Range and Quartile Deviation.

Solution for Activity – 3

Data set 1 :

X	$[X_i - \overline{X}]$
65	1
70	6
62	2
90	26
92	28
50	14
48	16
32	32
60	4
71	7
640	136

$$\overline{X} = \frac{\sum X_i}{n}$$

$$= \frac{640}{10}$$

$$\overline{X} = \underline{\underline{64.0}}$$

$$\text{MD} = \frac{\sum([x - \bar{x}])}{n}$$

$$= \frac{136}{10}$$

$$= \underline{\underline{13.6}}$$

Salaries Rs. 000	No. of Employees	Mid value (x_i)	$[x_i - \bar{x}]$	$f_i[x_i - \bar{x}]$
5 - 9	11	07	11	121
10-14	20	12	06	120
15-19	35	17	01	35
20-24	20	22	04	80
25-29	08	27	09	72
30-34	06	32	14	84
	100			512

$$M.D = \frac{\Sigma(1 - \bar{x}1)f}{\Sigma f}$$

$$\bar{X} = \frac{\sum X_i}{n}$$

$$= \frac{512}{10}$$

$$= \underline{\underline{5.12}}$$

$$\bar{X} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{1760}{100}$$

$$= 17.6$$

$$= \sim 18$$

Activity - 4

- Highlight the need of a more suitable measure than Mean Deviation to evaluate the dispersion of a data set.
- Point out that a more suitable measure to evaluate the dispersion can be derived by dividing the sum of the squares of those Mean deviations form the total number of observations contained in the data set.
- Explain that, that measure is known as Variance and the variance can be computed using following formulae.

Variance of an Array of data

$$\left(s^2\right) = \frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n}$$

Variance of a frequency distribution

$$S^2 = \frac{\sum_{i-1}^{n} f_i(x_i - \bar{x})^2}{\Sigma f_i}$$

- Use is $\left(S^2\right)$ highly appropriate to derive the Variance of a sample.

Point out further that below mentioned alternative formulae also can be used to compute variance.

Array of data

$$S^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$$

A frequency distribution

$$X_i = \frac{\sum fx_i^2}{\sum f_i} - \left(\frac{\sum fx_i}{\sum f_i}\right)^2$$

- Accordingly guide the students to compute the variance of the above two data sets.

Data set – 1

X_i	$\left(X_i - \overline{X}\right)^2$
65	1
70	36
62	4
90	676
92	784
50	196
48	256
32	1 024
60	16
71	49
640	3 046

$$\overline{X} = \frac{\sum X_i}{n}$$

$$= \frac{640}{10}$$

$$= \underline{\underline{64}}$$

$$\left(S^2\right) = \frac{\Sigma(x - \overline{x})^2}{n}$$

$$= \frac{3042}{10}$$

$$= \underline{\underline{304.2}}$$

$$S = \sqrt{\frac{\Sigma\left(x - \overline{x}\right)^2}{n}}$$

$$= \sqrt{304.2}$$

$$= \underline{\underline{17.4}}$$

Salaries Rs. 000	No. of Employees	X_i	f_iX_i	$f_iX_i^2$
5 - 9	11	7	77	539
10 - 14	20	12	240	2 880
15 - 19	35	17	595	10 115
20 - 24	20	22	440	9 680
25 - 29	08	27	216	5 832
30 - 34	06	32	192	6 144
	100	-	1 760	35 190

$$S^2 = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx_i}{\Sigma f}\right)^2}$$

$$= \frac{35190}{100} - \left(\frac{1760}{100}\right)^2$$

$$= 351.9 - 309.76$$

$$= \underline{\underline{42.14}}$$

$$= \sqrt{42.14}$$

$$= \underline{\underline{6.49}}$$

$$= \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$$

$$= \sqrt{\frac{304\ .2}{10}}$$

$$= \underline{\underline{17\ .44}}$$

- N :B : Emphasize the fact that the basic formula for evaluating the standard deviation of a frequency distribution is.

$$S = \sqrt{\frac{\Sigma fi(x_i - \bar{x})^2}{\Sigma f_i}}$$

And the above formula has been developed through expanding this formula to make the work of computing easy.

Activity – 5

Answer the questions related to each of the situations given below.

Situation - I

A and B are two firms involved manufacturing electric bulbs. A wholesale purchaser is in need of coming to know the life time of which kind of bulb is more stable. The summarized measures for the data collected related to the life time of each kind of bulb are as follows.

Measures	A	B
Mean	1500h	1200h
Standard deviation	9h	8h

- State using the standard deviation, of which kind of bulb, is there a greater variation.

  Bulb – A

- Find of which kind of bulb is there a more stable life time, expressing the standard deviation as a percentage relation to the mean. This value is **known** as the co-efficient of variation.

For - A $\frac{9}{1500} \times 100 = 0.6\%$ For - B $\frac{8}{1200} \times 100 = 0.67\%$

- Explain that there is a more stable life time in bulb - A
- Ascertain that the co-efficient of variable is the best measure of relative dispersion.

Situation – 2

Marks received by five students for three subjects are as follows.

Name of the students	Marks for		
	Accounting	Economics	Business Statistics
A	50	65	60
B	52	68	75
C	48	42	42
D	68	75	58
E	72	80	65
Total marks	290	330	300

- Assume that the total number of students who sat for these 3 subjects is only five (population size = 5).
- Derive the answer to the following questions.
- Derive the Mean value of the marks for each subject separately.

  Accounting μ = 58
  Economics μ = 66
  Business Statistics μ = 60

- Find the standard deviation for each subject separately.

  Accounting σ = 9.96
  Economics σ = 13.09
  Business Statistics σ = 10.75

- Subtract the Mean Marks of the respective subject from the raw mark gained by each student individually and divide by the standard deviation. Denote those values derived as the answer by – z

  Use the formula $z = \frac{x_i - \mu}{\sigma}$

Solution

- Z – score of A for each subject

Accounting $Z = \frac{50-58}{9.96} = 0.8032$

Economics $Z = \frac{65-66}{13.09} = -0.0764$

Business Statistics $Z = \frac{60 - 60}{10.75} = 0$

- Z – score of B for each subject

Accounting $Z = \frac{52-58}{9.96} = -0.6024$

Economics $Z = \frac{68-66}{13.09} = 0.1528$

Business Statistics $Z = \frac{75-60}{10.75} = 1.3953$

- Z – score of C for each subject

Accounting $Z = \frac{48-58}{9.96} = -1.0040$

Economics $Z = \frac{42-66}{13.09} = -1.8335$

Business Statistics $Z = \frac{42-60}{10.75} = -1.6744$

- Z – score of D for each subject

Accounting $Z = \frac{68-58}{9.96} = 1.0040$

Economics $Z = \frac{75-66}{13.09} = 0.6875$

Business Statistics $Z = \frac{58-60}{10.75} = -0.1860$

- Z – score of E for each subject

Accounting $Z = \frac{72-58}{9.96} = 1.4056$

Economics $Z = \frac{80-66}{13.09} = 1.0695$

Business Statistics $Z = \frac{65-60}{10.75} = 0.4651$

A guideline to explain the subject matters :

- Scattering of the observation in a data set is known as 'Dispersion'. The extent of the observations being apart from the average value of the series of numbers in depicted by 'dispersion'.
- Some of the uses of the knowledge about 'dispersion' are as follows;
  - Ability to know how data are scattered.
  - Ability to ascertain the reliability of the measures of average.
  - Ability to compare the nature of scattering in data among several distributions.
- Following measures of absolute dispersion can be used to evaluate the 'dispersion'.
  - Range
  - Quartile Deviation
  - Mean Deviation
  - Variance
  - Standard deviation\

Range

- The fdifference between the highest value and the lowest value in a number series (data – set) is known as the 'Range'.

  Range = Highest value – Lowest value

  R = H – L

- Advantages of computing the 'Range'
  - Being a simple measure
  - Being able to compute easily
  - Being able to have a rough idea about the dispersion of data shortly

- Disdvantages of computing the 'Range'.
  - Not being a representative measure, since all the observations are not taken into consideration
  - Considering only the two end values of the distribution

Quartile Deviation

- Half of the Inter Quartile Range $(Q_3 - Q_1)$ is known as Quartile Deviation
- Quartile deviation is a measure derived by considering 50% of the data in the middle range of a distribution, leaving out the end values

  Quartile Deviation or semi-interquartile Range is computed as follows.

$$Q.D. = \frac{(Q_3 - Q_1)}{2}$$

- Advantages of Quartile Deviation
  - Representing many number of data when compared to the Range
  - Since a great deal of data in a distribution are scattered in the middle range, the Quartile Deviation is more appropriate to evaluate the dispersion

- Disadvantages of Quartile Deviation
  - Since all the data are not taken into consideration, it cannot be regarded as a good representative measure of dispersion

Mean Deviation

- The average value of the absolute deviations from the Mean to each and every observation in a data set is known as Mean Deviation.

  Mean Deviation in ungrouped data can be computed using the following formula.

$$MD = \frac{\sum_{i=1}^{n}\left(\left|x_i - \bar{x}\right|\right)}{n}$$

Mean Deviation in grouped frequency distribution can be computed using the following formula.

$$M.D. = \frac{\sum_{i=1}^{n} f_i\left(\left|x_i - \bar{x}\right|\right)f}{\sum_{i-1}^{n} f_i}$$

Advantages of Mean Deviation

- Being a representative measure, since all the observation in the data set are taken into consideration
- Being able to ascertain the eligibility of Mean as a measure of Central Tendency, since the deviation from Mean to each and every observation is evaluated separately

Disadvantages of Mean Deviation

- Not being a good algebraic measure, since only the absolute values of the deviations are considered.

Variance and Standard Deviation

- Variance can be derived by considering the average value of the squares of Mean Deviations.
- Standard Deviation can be computed by deriving the positive square root of variance.
  - Variance of a sample is denoted by S^2 where as.
  - Standard Deviation of a sample by - S -
  - Population variance is denoted by σ^2 where as
  - Population Standard Deviation by σ

- When $x_1, x_2, x_3..........x_n$ is a set of ungrouped data the variance can be computed as follows.

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} \quad \text{or} \quad S^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2$$

- Hence the Standard Deviation is as follows.

$$S = \sqrt{\frac{\sum_{i-1}^{n}(x_i - \overline{x})^2}{n}} \quad \text{or} \quad S = \sqrt{\frac{\sum_{i-1}^{n} x_i^2}{n} - \overline{x}^2}$$

- Variance of a grouped frequency distribution with mid points of k- number of class intervals whose corresponding frequencies are given by f_1, f_2, f_3, …. F_k can be computed using any appropriate formula of the followings.

1. $$S^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i}$$

2. $$S^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} - \left(\frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1} f_i} \right)^2$$

3. When the $d_i = x_i - A$ (A = Assumed Mean)

$$S^2 = \frac{\sum_{i=1}^{n} f_i d_i^{\,2}}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} f_i d_i}{\sum_{i=1}^{n} f_i} \right)^2$$

4. When the width of all the class intervals is the same.

$$S^2 = C^2 \left[\frac{\sum_{i=1}^{k} f_i u_i^{\,2}}{\sum_{i=1}^{k} f_i} - \left(\frac{\sum_{i=1}^{k} f_i u_i}{\sum_{i=1}^{k} f_i} \right)^2 \right]$$

$U_i = \frac{X_i - A}{C}$ (C = Class width)

- The positive square root of the variance is the Standard Deviation.
- Relative advantages and disadvantages of Variance and Standard Deviation are as follows.
  - Since all the data are used to calculate the variance and Standard Deviation, it is good representative measure relative to Range and Quartile Deviation.
  - The extent of each and every observation deviated from Mean is taken into consideration.
  - Reliability of Mean can be evaluated.
  - Taking the square of the real deviations; unlike absolute deviations to avoid the Mean deviation being zero, this measure is mathematically accurate as well.

- Deviations are over-evaluated in variance, but Standard Deviation rules out that weak point.
- Standard Deviation is the best measure of absolute deviation to compare the variations in two or more distributions with the same unit and/or with the similor means.
- Through using these measure of absolute dispersion to compare the variation in two or more distributions, with different units and significant means wrong conclusions may be achieved.
- Measures of Relative Dispersion can be used to compare the variation in such distributions.
- A percentage of its mean.
- That measure is known as the co-efficient of variations and can be computed as follows.

$$C.V. = \frac{S}{\overline{X}} \times 100$$

- Perhaps the mean deviations of a particular observation may required to be stated as how may times of its standard deviation. The values transformed in that manner are known as standardized variables and computed as follows.

$$Z = \frac{X - \overline{X}}{S}$$

Z = standardized variable

μ =Mean

σ = Standard Deviation

X_i = Relevant Observation of the considered variable.

Uses of the measures of Relative Dispersion.

- The type of the distribution in a group of data can be very well comprehended through interpretation of central tendency.
- The most lucrative business opportunity can be chosen by an investor through several alternative business opportunities.
- Being aware of the variance is useful in making optimal decisions.

Assessment and Evaluation :

Divide the students in the class into four groups and involve them involve in each activity mentioned below giving the following data set.

65, 70, 62, 90, 92, 50, 48, 32,60, 71

Group – I

Add 2 for each and every observation in this data set. Find the standard deviation of that new data set.

Group – 2

Subtract per 2 from each and every observation of this data set. Find the standard deviation of that new data set.

Group – 3

Multiply each and every observation of this data set by 2. Find the standard deviation of this new data set.

Group – 4

Divide each and every observation of this data set by 2. Find the standard deviation of this new data set

Given below are some measures computed for the distributions of some different variables.

Distribution	Mean	Standard Deviation
• Weight of grade 12 students	58 kg	6 kg
• Height of grade 12 students	160 cm	12 cm
• Distance of grade 12 students from home to school	1 450 m	80 m
• Weight of grade I students	15 kg	4 kg
• Life time of the type of bulb -A	1 500 hrs	9 hrs
• Life time of the type of bulb - B	1 200 hrs	8 hrs

Competency 3.0 : Analyses Business Data using the Techniques of Descriptive Statistics.

Competency Level 3.5 : Uses the measures of 'Skewness' and Kurtosis to analyses Business Data

No. of Periods : 12

Learning outcomes :

- Interprets 'Skewness' and points out its need.
- Interprets and computes the Pearson's 1st co-efficient of Skewness.
- Explains where the Pearson's 1st co-efficient of Skewness is applicable and inapplicable.
- Interprets and computes the Dearson's 2nd co-efficient of Skewness.
- Explains where the Pearson's 2nd co-efficient of Skewness is applicable and inapplicable.
- Interprets and computes Bowley's co-efficient of Skewness.
- Points out merits and demerits of Bowley's co-efficient of Skewness.
- Interprets and computes Keley's co-efficent of Skewness.
- Points out merits and demerits of Keley's co-efficient of Skewness.
- Examines the nature of the distribution using co-efficient of Skewness.
- Interprets Kutrosis and points out the need of it.
- Explains Leptokurtic, Messokustic and Platykurtic distributions separately.
- Interprets the percentile Kurtosis co-efficient and computes it.
- Examines the type of distributions using the Kurtisis co-efficent.
- Introduces the Box and whisker's plot and explains the need of it.
- Constructs a Box and Whisker's plot for few data sets on the same chart and compares the nature of scattering in those distributions.

Instructions for Lesson Planning :

- Find and produce the histograms in following shapes to the class.
- Hold a discussion highlighting the following facts based on the above histograms.
    - A histogram with appositively skewed distribution No.01)
    - A histogram with a negatively skewed distribution (N0.02)
    - A histogram with a symmetrical distribution (No.03)
    - Distributions of data are in different types.
    - Some of the distributions are symmetrical and some are symmetrical.
    - The first data set above is distributed with a longer tail towards the right and the 2nd data set with a longer tail towards the left.

- The distribution of the data 3rd histogram is symmetrical.
- Asymmetrical distributions are known as Skewed distributions.
- The distributions with a tail towards the right hand side are called right Skewed or positively Skewed distributions.
- The distributions with a tail towards the left hand side are called left Skewed or negatively Skewed distributions.

- Group the students appropriately and provide with the following information and involve them in the Activity.
- Given below are some of the measures computed having analysed the term test marks received by a group of students in a school for four subjects.

  - Economics

    $\overline{X} = 48$

    $Md = 52$

    $S = 10$

    (i) $\dfrac{3(\overline{X} - Md)}{S}$

  - Accounting

    $\overline{X} = 52$

    $M_0 = 45$

    $S = 5$

    (ii) $\dfrac{\overline{X} - Mo}{S}$

  - Business Statistics

    $Q_1 = 42$

    $Q_2 = 50$

    $Q_3 = 62$

    (iii) $\dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

  - English

    $P_{10} = 48$

    $P_{50} = 56$

    $P_{90} = 60$

    (iv) $\dfrac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$

- Lead them to derive a value by substituting the measures received by each group in the formula provided along with.

Solutions :

Economics $= \frac{3(\overline{X} - Md)}{S} = \frac{3(48 - 52)}{10} = \frac{3 \times -4}{10} = \frac{-12}{10} = \underline{\underline{-1.2}}$

Accounting $= \frac{\overline{X} - Mo}{S} = \frac{52 - 45}{5} = \frac{7}{5} = \underline{\underline{1.4}}$

Business Statistics $= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

$= \frac{62 + 42 - 2 \times 50}{62 - 42}$

$= \frac{104 - 100}{20}$

$= \underline{\underline{0.2}}$

English $= \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$

$= \frac{60 + 48 - 2 \times 56}{60 - 48}$

$= \frac{108 - 112}{12}$

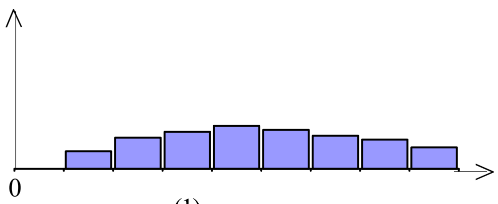
$\underline{\underline{-0.33}}$

- Point out that the marks distributions of Economics and English are negatively Skewed where as the marks distributions of Accounting and Business Statistics are positively Skewed.

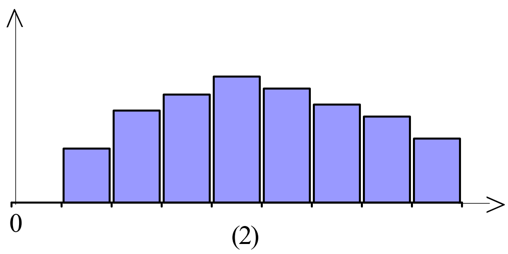
- Ascertain the fact that even though the marks distributions of Economics and English are negatively Skewed, the two values taken by the co-efficient are different to each other.
- Ascertain the fact that even though marks distributions of Accounting and Business Statistics are positively Skewed, the two values taken by the co-efficient are different to each other.
- Make them aware the fact whether the Skeweness higher or lower can be explained in accordance with the value received for the co-efficient Skeweness .

Activity – 02

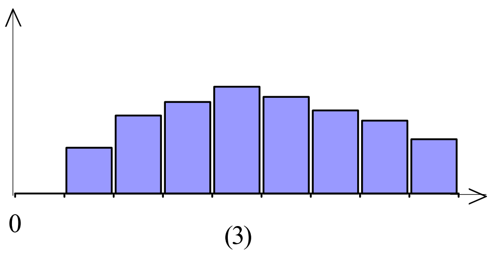
Provide with the following histograms to the class.



(1)



(2)



(3)

- Hold a discussion on the following questions based on the above histograms.

  1. Are these distribution symmetrical or asymmetrical?
  2. How is the formation of each symmetrical distribution different to one another?
     - In accordance with the peakedness of them (curvature)
     - In accordance with how the data are flocked arround or scattered away from the center
  3. How do you explain the nature of scattering the data in each distribution?
     - It is vividly obvious that 2nd distribution consists of a taller peak compared to the peak of the 1st distribution and the 3rd distribution possesses relatively a flatter peak.

- It is also obvious that the data in the 2nd distribution are flocked about the center where as the data are scattered away from the centre to the either side in the 3rd distribution.

- Point out that this formation of data in a distribution is known as "Kurtosis" and further that it is measurable.

Activity 3 :

- Group the students appropriately and involve them in the following activity.

  - Marks received for Mathamatics in three parallel classes are displayed in following frequency distributions.

  - Marks distribution of A – class.

$Q_1 = 15 \qquad P_{10} = 13$

$Q_3 = 40 \qquad P_{90} = 42$

  - Marks distribution of B – class.

$Q_1 = 19 \qquad P_{10} = 15$

$Q_3 = 30 \qquad P_{90} = 36$

  - Marks distribution of C – class.

$Q_1 = 15 \qquad P_{10} = 08$

$Q_3 = 20 \qquad P_{90} = 60$

- Involve in the following activity related to the details received by you.
  - Find the difference between Q_3 and Q_1
  - Find the difference between P_{90} and P_{10}
  - Compare these deviations.
  - Substitute those values in following formula and compute the Kurtosis co-efficient.

$$K = \frac{\frac{1}{2}(Q_3 - Q_1)}{P_{90} - P_{10}}$$

  - Comment of the distribution according to the vlue of co-efficient you have derived.

Solutions :

A class	B class	C class
$K = \dfrac{\frac{1}{2}(Q_3 - Q_1)}{P_{90} - P_{10}}$	$K = \dfrac{\frac{1}{2}(Q_3 - Q_1)}{(P_{90} - P_{10})}$	$K = \dfrac{\frac{1}{2}(Q_3 - Q_1)}{(P_{10} - P_{10})}$
$= \dfrac{\frac{1}{2}(40-15)}{(42-13)}$	$= \dfrac{\frac{1}{2}\times 11}{21}$	$= \dfrac{5}{52}$
$= \dfrac{12.5}{29}$	$= \dfrac{5.5}{21}$	$= \underline{\underline{0.0962}}$
$= \underline{\underline{0.431}}$	$= \underline{\underline{0.2619}}$	

- Marks distribution of A – class reports the highest value for Kurtosis co-efficient.
- Marks distribution of C – class reports the lowest value for Kurtosis co-efficient.
- It can be pointed out that the marks in C-class are scattered away relative to the spread of the marks in A – class

Activity – 04

- Provide with the following stem and leaf diagram to the class and lead them to answer the questions given below.

Stem	Leaf
3	1, 2, 5, 8
4	3, 5, 6, 7
5	1, 3, 6, 8, 8, 9
6	2, 5, 7, 8, 9
7	3, 6, 8
8	9

1. What is the minimum value of the data?
2. What is the maximum value of the data?
3. Compute the first quartile (Q_1)
4. Compute the second quartile (Q_2)
5. Compute the third quartile (Q_3)
6. Draw a number line scaled from 30-90 and plot the above measures on it.

7. Draw a cage (rectangle) containing Q_1 Q_2 and Q_3
8. Divide that cage in to two parts across Q_2
9. Derive Inter Quartile Range (Q3 – Q1)
10. Multiply the value received for (Q3 – Q1) from 1.5
11. Subtract the value derived for the 10th question from Q_1 and add it to Q_3.
12. Check whether there are values less than the lower value received for the 11th question and more than the upper value received.

Solutions :

1. $\underline{\underline{31}}$

2. $\underline{\underline{89}}$

3. $Q_1 = \frac{1}{4}(n+1)$ th term

$= \frac{1}{4} \times 24 = 6$ th term

$Q_1 = \underline{\underline{45}}$

4. $Q_2 = \frac{2}{4}(n+1)$ th term

$= \frac{2}{4} \times 24 = 12$ th term

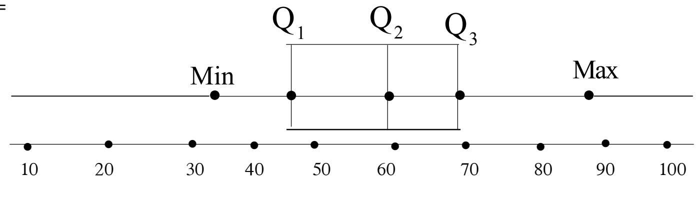
$Q_2 = \underline{\underline{58}}$

5. $Q_3 = \frac{3}{4}(n+1)$ th term

$= \frac{3}{4} \times 24 = 18$ th term

$Q_3 = \underline{\underline{68}}$

6., 7., 8.



9. $Q_3 - Q_1 = 68 - 45 = 23$

10. $23 \times 1.5 = \underline{\underline{34.5}}$

11. $45.0 - 34.5 = 10.5$, $68 + 34.5 = 102.5$

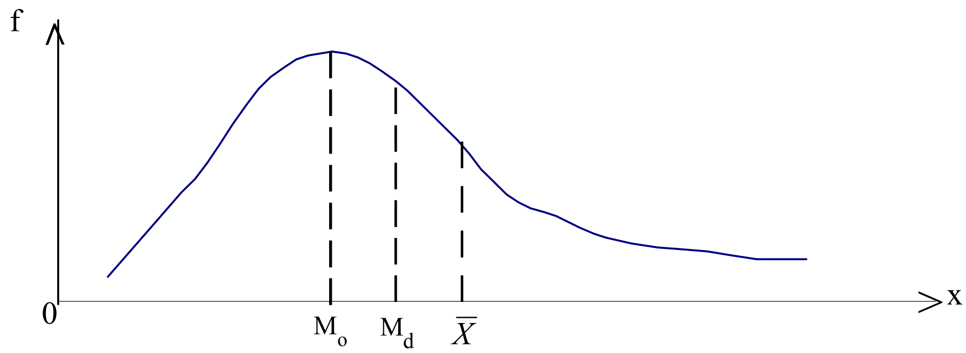
12. (i) There are no any observation less than 10.5.

(ii) There are no any observation exceeding 102.5

- Explain that the above drawn chart is known as the Box Whisker's plot.
- Point out that the scattering of data can be explained using a Box & Whisker's plot as well.

A Guideline to explain the subject matters :

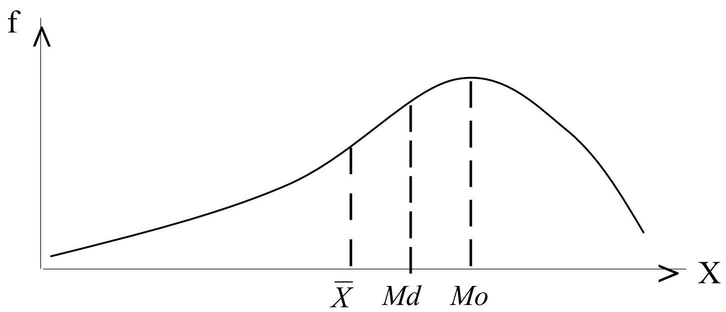
- To which extent a distribution is apart from symmetrical shape or the asymmetrical nature of a distribution is known as 'Skewness'.
- If the smooth frequency curve of a distribution falls with a longer tail towards the right hand side rather than left from its peak (maximum point), then that distribution is right Skewed or positively Skewed.



- In a data set with such a distribution Median is greater than Mode, where as the Mean is greater than Median.

$Mo < Md < \overline{X}$

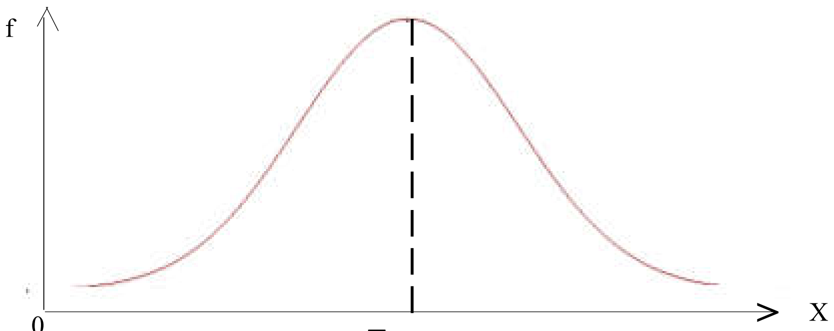
- If the smooth frequency curve of a distribution falls with a longer tail towards the left hand side rather than the right from its peak (maximum point), then that distribution is left Skewed or negatively Skewed.



- In a data set with such a distribution Median is greater than Mean where as the Mode is greater than the Median.

  $\overline{X} < Md < Mo$

- In a symmetrical distribution the Mean, Mode and Median are equal.



- Majority of the observations, in a right skewed distribution are less than the Mean.
- Majority of the observations in a left skewed distribution are grater than the Mean.
- The following relationship exists among the quartiles in a right skewed distribution.

  $Q_2 - Q_1 < Q_3 - Q_2$

- The following relationship exists among the quartiles in a left skewed distribution.

  $Q_2 - Q_1 > Q_3 - Q_2$

- Karl Pearsons' first co-efficient of skewness.

  $$Sk_1 = \frac{\overline{X} - Mo}{S}$$

- Karl Pearsons' second co-efficient of skewness.

  $Sk_2 = \dfrac{3(\overline{X} - Md)}{S}$ can be used to measure the Skewness of a distribution.

- Karl Person's second co-efficient of Skewness can be used when the Mode of a distribution is not identical.
- Quartiles or percentiles are used to evaluate the Skewness of a distribution with extreme values or with open ended classes.

- Quartile Skewness co-efficient.

$$Sk_q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

This is called Bowley's co-efficient of skewness.

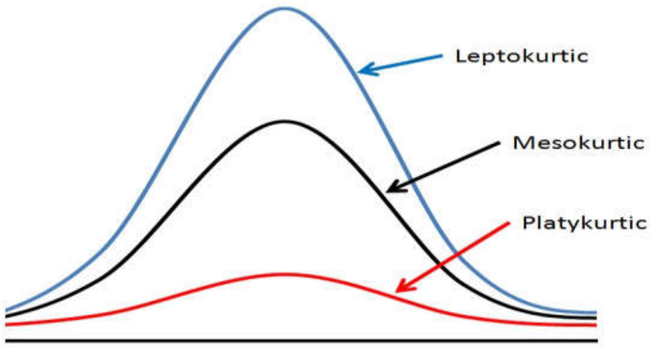
- Percentile Skewness Co-efficent

$$Sk_p = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

This is called Kelley's co-efficient of skewness.

  - If the value of co-efficient of skewness is zero that distribution is symmetrical.

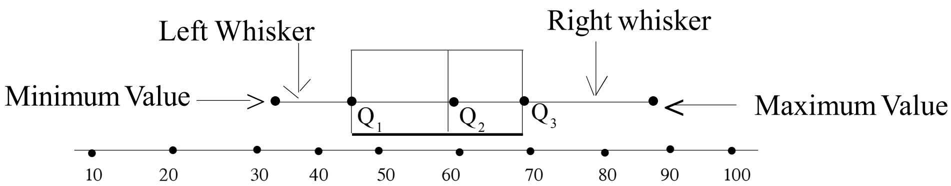
- According to the value of the co-efficeint of skewness the direction as well as the magnitude of skewness can be assessed.
- Karl Pearson's first co-efficebt of skewness (sk_1) falls between -1 and +1.
  - The distributions with $Sk_1 \pm 0.5$ are moderately skewed distributions.
  - The distribution with Sk_1 falling outside ± 0.5 are heavily skewed distributions.
- Karl Pearson's second co-efficient of skewness falls between -3 and +3.
  - The distributions with $-1 < Sk_2 < +1$ are moderately skewned distributions.
  - The distributions with Sk_2 falling outside ± 1 are considered as highly skewed distibutions.
- Skewness of distributions can also be evaluated using Sk_q and Sk_p as well.
- The P Nature of Peakedness (curvature) of a distribution is known as a 'Kurtosis'
- The Kurtosis of a distribution is considered relative to the curvature of the normal distribution which is a symmetrical distribution.
- A distribution relatively with a taller peak is known as a Leptokurtic distribution where as a distribution with a flatter peak relative to a normal distribution is known as a Plattykurtic distribution. A distribution of which the peak is not that tall or not that short is called a messokurstic distribution.



- The percentage kurtosis co-efficient (k) that has been defined using quartile and percentiles is used to evaluate the kurtosis of a distribution.

$$K = \frac{\frac{1}{2}(Q_3 - Q_1)}{P_{90} - P_{10}}$$

- Majority of the data in a Leptokurtic distribution are flocked about the centre of the distribution, so that the distance from P_{90} to P_{10} and Q_3 to Q_1 is not that much. The more these distances are narrower, the value of Kurtosis co-efficient K approaches 0.5.
- Mejortiy of data are scattered away from the centre of a Plattykurtic distribution in such a distribution the difference between Q_1 - Q_3 and P_{10} - P_{90} is greater. The value of K is closer to 0 for a presently flat peaked distribution.
- The value of Kurtosis co-efficient of a messokurtic distribution should be an average between 0 and 0.5, whereas this co-efficient for a normal distribution is 0.263.
- Accordingly if K = 0.263 of a distribution that is known to be a messokurtic distribution, Hence if K <0.263 that is a messokurtic distribution whereas K > 0.273. That is a leptokurtic distribution.
- A chart constructed as follows using minimum value, maximum value, first quartile, second quartile and third quartile of a data set is known as a Box & a whiskers's plot.



- The below mentioned steps can be followed in constructing a Box & a whisker's plot.
  - Drawing a number line.
  - Plotting the minimum value, maximum value Q_1, Q_2 and Q_3 on line drawn parallel to the number line.
  - Drawing the box (rectangle) containing Q_1, Q_2 and Q_3
  - Dividing the box in to two parts at Q_2

- Uses of a Box & a Whisker's plot are as follows :
  - By drawing smooth curve inside the box containing Q_1, Q_2 and Q_3, it can be identified whether the distribution of data in the middle is symmetrical or right skewed or left skewed.
  - Ability to understand the nature of the extreme values in the distribution through the left whisker and the right whisker.
  - If left whisker and right whisker are equal in length, that is a symmetrical distribution.
  - If the left whisker is longer than the right whisker that is a negatively skewed distribution.
  - If the right whisker is longer than the left whisker that is a positively skewed distribution.
  - Ability to find out whether there are extreme values i.e. outliers in the collected data.
- Follow the steps mentioned below to check whether there are outliers in collected data,
  - Find interquartile range $(Q_3 - Q_1)$
  - Multiply the value of $Q_3 - Q_1$ from 1.5
  - Subtract that product from Q_1
  - Add that product to Q_3
  - If there are values less than $Q_1 - (Q_3 - Q_1) \times 1.5$ those are called outliers.
  - If there are values greater than $Q_3 + (Q_3 - Q_1) \times 1.5$ those are called outliers.

Assessment and Evaluation :

Ex (i) Daily sales income of 3 business firms are given in the following table.

Amount of Sales Rs. 000	Number of days		
	Firm A	Firm B	Firm C
10-20	5	26	4
20-30	10	30	5
30-40	20	20	8
40-50	30	10	10
50-60	20	8	23
60-70	10	4	32
70-80	5	2	18
	100	100	100

- Construct the frequency polygons of the three firms on the same co-ordinate plane.
- Compare the distribution of data (using skweness and Kurtosis)

- Find separately for the above three data sets
  - Mean - $(\overline{X})$
  - Median (Md)
  - Mode (M_O)
  - Stand deviation (S)
  - First quartile (Q_1)
  - Third quartile (Q_3)
  - Tenth percentile (P_{10})
  - Ninetieth percentile (P_{90})

- Derive the co-efficient of skewness for three data sets using the following formulae.

$$Sk_1 = \frac{\overline{X} - Mo}{S} \qquad Sk_q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

$$Sk_2 = \frac{(\overline{X} - Md)}{S} \qquad Sk_p = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

- Compare the skewness of each distribution.
- Derive the Percentage kurtosis co-efficient for three data sets above using the following formula.

$$K = \frac{\frac{1}{2}(Q_3 - Q_1)}{P_{90} - P_{10}}$$

- Describe the shape of the distribution of three data sets in accordance with frequency polygon, co-efficient of skewness and co-efficient of kurtosis.

Answer :

Firm	$\overline{X}$	Md	Mo	S	Q_1	Q_3	P_{10}	P_{90}	Sk_1	Sk_2	Sk_q	Sk_p	K
A	45	45	45	14.5	35	55	25	65	0	0	0	0	0.25
B	31.4	28	22.9	15.2	19.6	39.5	13.9	55	0.56	o.67	0.16	0.31	0.24
C	56.1	60	63.9	15.9	48	67.8	31.3	74.4	-0.49	-0.73	-0.21	-0.33	0.23

A $\rightarrow$ is a symmetrical distribution. $\overline{X} = Md = M0$

B $\rightarrow$ is a positively skewed distribution $\overline{X} > Md > M0$

C $\rightarrow$ is a negatively skewed distribution.. $\overline{X} < Md < M0$

Ex (2) :

D Considering the following data set

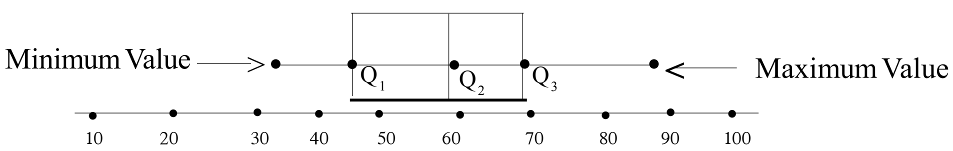
- construct a box & a whisker's plot
- Explain the distribution of data
- Check whether there are outliers.

Stem	leaf
2	1 2 3 3 5 7
3	1 2 6 7 7 7 8 8 8 8 9 9 9
4	0 0 0 0 4 5 6 7 8
5	1 5

Solution :

Minimum value = 22, Maximum value = 55

Q_1 = 31.5 Q_2 = 38 Q_3 = 40



To check outliers.

Lower outliers boundary (Lower inner fence)

$= Q_1 - 1.5 \quad (Q_3 - Q_1)$

$= 31.5 - 1.5 (40 - 31.5)$

$= 31.5 - 1.5 \times 8.5$

$= ~~31.5~~ - 12.75$

$Q_1 = 18.75$

Upper outliers boundary (Upper inner fence)

$= Q_3 + 1.5 \quad (Q_3 - Q_1)$

$= 40 + 1.5 (40 - 31.5)$

$= 40 + 1.5 \times 8.5$

$= 40 + 12.75$

$Q_3 = \underline{\underline{52.72}}$

- There is no any observation less than 18.75, but 55 is an observation greater than 52.72.

  $\therefore$ 55 is an outliers.
- Distribution in the middle of the data set is simply negatively skewed.
- The righ hand side whisker is longer than the left hand side whisker.

Competency 4.0 : Studies the relations between variables and forecasts.
Competency Level 4.1 : Categorizes the variables in accordance with the nature related
No. of Periods : 06

Learning outcomes :

- Interprets 'Variables'
- Names inter-related variables.
- Differentiates between independent variables and Dependent variables.
- Introduces the 'Scatter Diagram'.
- Presents data using scatter diagrams.
- Explains the Linear Relations between variables using scatter diagrams.
- Explains Non-linear Relations between variables using scatter Diagrams.
- Represents a situation where no any Relations exists between variables using a scatter diagram.
- Describes the uses of a scatter diagram.

Guidelines for lesson planning :

- Pay attention of the students for following situations.
  - A manufacturing firm has advertised few products and searching for whether a significant progress in sales of those products have been reported.
  - A meteorologist is involved in a study on the rainfall experienced in a particular area during a selected month and the changes in humidity in the atmosphere observed during the period.
  - An economist is involved in a study about the pattern of changing the consumer demand for a particular product when the market price of the product is gradually increasing.
  - A researcher is studying how the consumer pattern changes once the income level of an individual is changing.
- Hold a discussion with students highlighting the following matters.
  - Point out that the above four facts are related to practical situations with two variables.
  - Further point out that, when those four facts are considered separately the variables included in each are interrelated (correlated).
  - Hence point out that
    - Once few selected products are advertised the sales volume of those products are possible to be increased;

    - Once the rain-fall grows, the humidity in atmosphere also can grow.
    - Once the price of a product increases, the consumer demand for that product may be decreased.
    - Once the income of an individual increases the consumption may be increased and vice versa.
- Accordingly inquire the students about some other variables related to each other.
- Pay the attention of students to the variable in a pair of variables considered, that changes with no any influence of the other variable (Independent variable) and the variable that changes depend on the said variable (Dependent variable).
- Point out that the expenses for advertising is the Independent variable and sales volume which depends on that variable is the Dependent Variable.
- Pay the attention of students to separate the independent variable and the dependent variable of the other 3 pairs of variables. Hence lead them to complete the following table.

Independent variable	Dependent variable
•	•
•	•
•	•
•	•

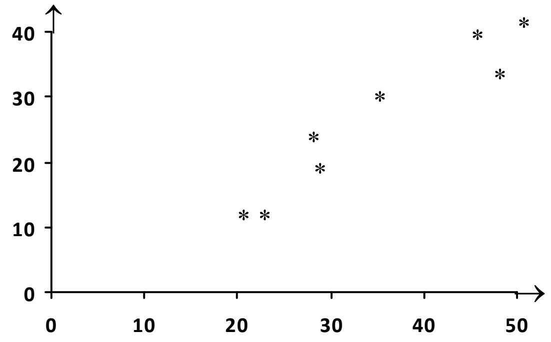
- Point out that the observations related to a pair of variables can be plotted out on a scatter diagram drawn on an appropriate scale.
- Accordingly lead the students to make out the relations between the variables in each pair of variables.
- Involve the students in the following activity.

Activity.

- Given below are the marks received by 10 individuals who sat for an examination held to recruit 10 Sales Promotion Officers and the targeted sales volume assigned on them.

Marks	43	50	22	50	26	34	30	48	40	32
Value of sales in a month (Rs.000)	26	37	15	29	15	21	20	32	29	17

- Guide the students to construct the scatter diagram, having identified the independent variable as X and the dependent variable as Y.

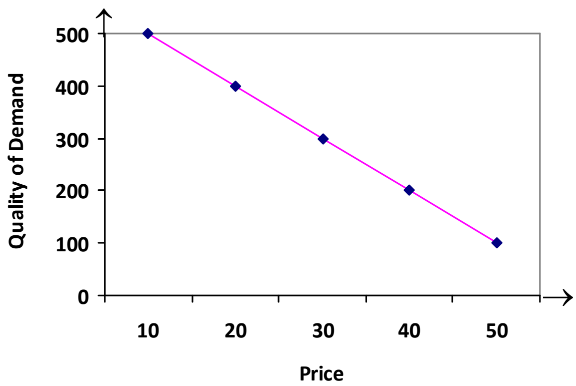


- Involve in a discussion highlighting the following facts.
  - Point out that a rough idea about the relationship between two variables in a pair of variables can be derived in accordance to the scattering pattern of the points on a scatter diagram.
  - Lead the students to check this scatter diagram very carefully and inquire whether a linear relationship which is modled on $Y = a + bX$ structure can be set on this scatter diagram passing through all those points plotted out it.
  - Ascertain that $Y = a + bx$ is the real model of a straight linear relationship and all the points on above scatter diagram do never fall on such a straight line.
  - Anyway, point out the fact that a straight line passing closer to many of those points can be drawn and therefore a linear relationship between these two variables can be interpreted
  - Ascertain the fact that there is a positive relationship between two variables, if the value of one variable is increasing, when the value of the other variable increases, such as usage of fertilizer and yield of corps, sales income and advertising expenses etc.

Lead the students to be involved in the following activity.

- When the price of a particular item is increasing from Rs : 10 to Rs : 50, the movement of demand is given in the following table. Plot out these data in a scatter diagram.

Price (Rs.)	Quantity Demanded (units)
10	500
20	400
30	300
40	200
50	100

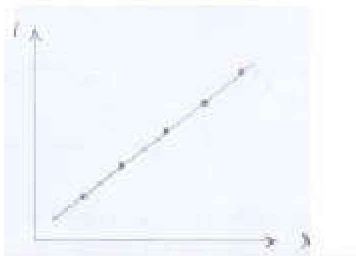


- Point out that when the value of one variable increases the value of the other variable decreases in this diagram.
- Point out further that this is a negative linear relationship and the line falls down from left to right.
- Point out that there are perfectly negative, strongly negative as well as poor negative relationships.
- Point out further that there are variables with non-linear relations as well as no-correlations.
- Enquire the students about the importance of a scatter diagram.

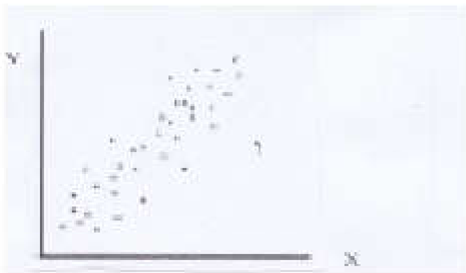
A Guideline to explain the subject matters :

- Mainly the variables are in two types.
    1. Independent variables
    2. Dependent variables

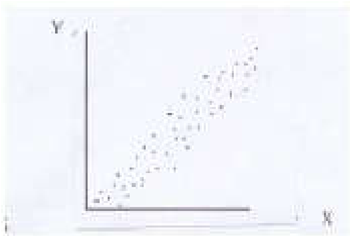
- If a change in one variable does not make any influence for a change in another variable those variables are known as independent variables or else, once two variables are given the variable that changes freely is called the independent variable.
- The variable that changes based on the independent variable, is the 'dependent variable,'
- When the ordered pairs / observed points related to a pair of variables are plotted on a co-ordinate plane, the chart derived is called the scatter diagram.
- In order to construct a scatter diagram the independent variable should be represented on the horizontal axis where the dependent variable on the vertical axis.
- Three types of relations obvious between two variables can be identified by observing a scatter diagram.
  1. Linear (positive/negative) relations
  2. Non-linear relations
  3. Non-related variables.
- If the value of the dependent variable also increases, when the independent variable increases, there is a positive relationship between those variables and then all the points in the scatter diagram may fall on or closer a straight line rising from left to right.



A perfectly positive relationship.

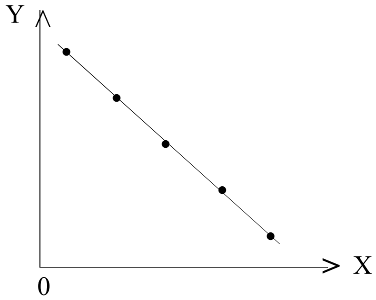


A strong positive relationship.

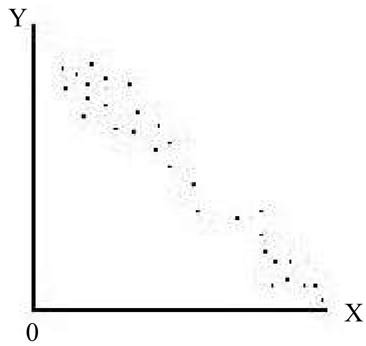


A poor positive relationship.

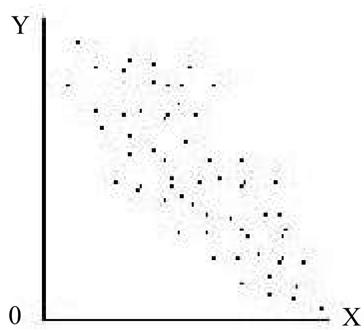
- If the value of the dependent variable decreases, when the value of independent variable decreases as well there is a positive relationship between them.
- If the value of the dependent variable decreases, when the value of the independent variable increases, there is a negative relationship between those variables. There, all the points in the scatter diagram fall on or closer a straight line that falling down from left to right.



A perfectly negative relationship.

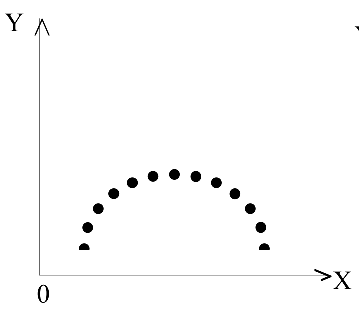


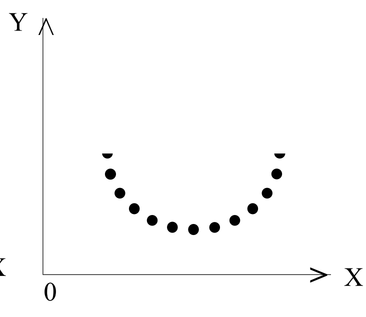
A strong negative relationship.

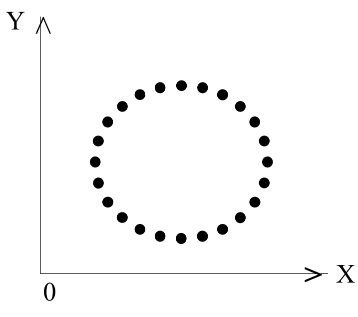


A poor negative relationship.

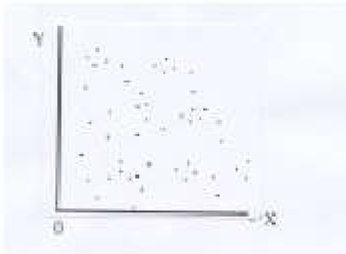
- There are non-linear relations also between the variables.

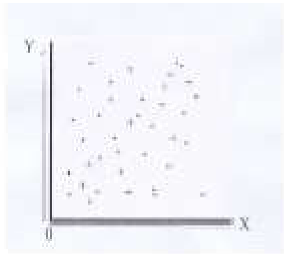






- In the sense, when there is no any relations between two variables, the changes in each variables, occur solely (independently). At such situations a clear linear relationship is not obviously represented on the scatter diagram.





- In this manner whether there is a relationship between two variable or not can be examined by studying a scatter diagram.
- Since a simple idea about the relationship of two variables if any, can be gained, a scatter diagram is very important in the analysis of simple regression.

Competency 4.0 : Studies the relations between variables and forecasts.

Competency Level 4.2 : Studies the concept of Linear Correlation between two variables

No. of Periods : 04

Learning outcomes :

- Interprets the concept of 'correlation'.
- Gives examples for the situations where correlation is applied.
- Describes the uses of the knowledge about correlation obvious between two variables.
- Explains the need of evaluating the size of correlation between two variables,

Gidelines for Lesson Planning :

- Write the following statements on the black/white board.
    - Relationship of the marks received by the students at the exam and the time they spent for studies,
    - Relationship between 'Inflation' and 'production cost.'
    - Relationship between the consumption of fertilizer and the harvest yielded
    - Relationship between the age and weight of individuals.
    - Relationship between the quantity of production in a manufacturing firm and the manufacturing cost.
- Hold a discussion highlighting the following facts.
    - When there are two variables a changes in one variable may affect on the other variable.
    - Once the number of hours studied for the exam is increased by the students, they may gain higher scores.
    - Inclining of the inflation may cause for increasing the production cost.
    - The harvest can be increased by promoting the usage of fertilizer, insecticide and weedicide etc.
    - A growth of the weight of a child can be obvious once he is aging.
    - Once the production volume is increased the total cost also can be increased.
    - Hence, point out that it is important to examine whether there is a relationship between two variables in decision making not only in the business field, but also in many other practical fields.
    - Accordingly point out that the relationship existing between two variables is quantified using the correlative co-efficient.

- Point out that the value of correlative co-efficient falls in the range $-1 \le r \le +1$ and further that whether the correlation is strong or weak can also be reviewed.
- Discuss with the students the importance of examining the correlation between two variables.

A Guideline to explain the subject matters :

- If there is a relationship between two variables those two variables are known to be a pair of correlated variables.
- When the value of one variable increases, if the value of another variable be also increases, those are positively correlated variables.

  Ex : Once the age of machine increases, increasing the maintenance cost.
- When the value of one variable increases, if the value of another valuable decreases those are negatively correlated variables.

  Ex :1. Once the price of a product increases, decreasing the demand for that product.

  2. Decreasing the sales of ice-cream on rainy days.

- If any changes in one variable does not have any influence on another variable, those variables are not co-related.

  Ex : Number of child births delivered in a day in Colombo National Hospital and the number of road accidents take place in Kandy city.
- A measure of linear relationship between two variables is provided with the correlative co-efficient.
- The value of correlative co-efficient falls in the range $-1 \le r \le +1$ The size of correlation is determined on the value or 'r' as follows.

If r = 1	A perfectly positive correlation.
If r is closer to +1	A strong positive correlation
If 0.5 < r < 0.75	there is a moderately positive correlation.
If r = -1	A perfectly negative correlation.
If r is closer to -1,	A strong negative correlation.
If -0.75 < r < -0.5	there is a moderately negative correlation.

If r takes a negative value closer to zero, there is a poor negative correlation.

If r = 0 There is no any linear correlation.

Importance of Correlation.

- In making decisions at various stages in the business filed, most probably a study is launched to check whether there is a relationship between two variables. The concept of "correlation" is more importance in that context.

  Ex : If the sales of a particular product depends on the cost of advertising, it may be comprehended that the sales could be promoted by allocating a considerable amount of money on advertising.

- Since the dependent variable can be estimated by expressing the correlation between two variables in a mathematical function, the concept of "correlation" is more useful in making production plans and estimations.

- Correlative co-efficient is computed in two types.

  1. Karl Pearson's Product Moment Correlative co-efficient.
  2. Spearman's Rank Correlative Co-efficient.

Competency 4.0 : Studies the Relations between variables and forecasts.

Competency Level 4.3 : Quantifies the Product Moment Correlative Co-efficient

No. of Periods : 06

Learning outcomes :

- Interprets the Product Moment Correlative Co-efficient.
- Gives instances for situations where the Product Moment Correlative co-efficient is applicable.
- Computes the Product Moment Correlative Co-efficient to examine the relationship between two variables given.
- Points out the properties of Product Moment Correlative Co-efficient.
- Explains the strength and the direction of correlation between the two variables using the correlative co-efficient.

Guidelines for Lesson Planning :

- Involve the students in following activity to explain the Product Movement correlative co-efficient.

Activity 01 :

- Provide the students with the following data set containing the tail-length of five (5) mice examined and their weight.

Talk length (cm)	9	8	10	6	7
Weight (g)	23	26	30	22	24

- Ask from the students whether these are quantitative variables or qualitative variables.
- Ask from students whether there is a relationship between these variables.
- Guide the students to compute the followings to evaluate the linear relationship of these two variables using the product Moment Correlative Co-efficient.

  (i) $\overline{X}$

  (ii) $\overline{Y}$

  (iii) Values for $\left(X_i - \overline{X}\right)$

(iv) Values for $(Y_i - \overline{Y})$

(v) Values of $(X_1 - \overline{X})$ $(Y_1 - \overline{Y})$ and the sum of those values

$\Sigma[(X_i - \overline{X})(Y_i - \overline{Y})]$

(vi) Values of $(X_i - \overline{X})^2$ and the sum of those values $\Sigma(X_i - \overline{X})^2$

(vii) Values of $(Y_i - \overline{Y})^2$ and the sum of those values $\Sigma(Y_i - \overline{Y})^2$

(viii) Multiply the sum of $(X_i - \overline{X})^2$ by the sum of $(Y_i - \overline{Y})^2$

$\Sigma(X_i - \overline{X})^2 \, \Sigma(Y_i - \overline{Y})^2$

(ix) Derive the square root of the calculative in (viii) above.

(x) Divide the result derived at step – V above by the result derived at step ix above.

Solution of Activity – I

(i) $\overline{X} = \frac{\sum x_i}{n} = \frac{40}{5} = 8$

(ii) $\overline{Y} = \frac{\sum y_i}{n} = \frac{125}{5} = 25$

(iii) $(X_i - \overline{X})$

(9-8) = 1

(8-8) = 0

(10-8) = 2

(6-8) = - 2

(7-8) = -1

(iv) $(Y_i - \overline{Y})$

(23 - 25) = -2

(26 - 25) = 1

(30 - 25) = 5

(22 - 25) = -3

(24 - 25) = -1

(v) $(X_1 - \overline{X})$ $(Y_1 - \overline{Y})$

1	×	-2	=	-2
0	×	1	=	0
2	×	5	=	10
-2	×	-3	=	6
-1	×	-1	=	1
			=	15

(vi) $(X_i - \overline{X})^2$

$1^2 = 1$

$0^2 = 0$

$2^2 = 4$

$-2^2 = 4$

$-1 = 1$

10

(vii) $\left(Y_i - \overline{Y}\right)^2$

-2^2	=	4
1^2	=	1
5^2	=	25
-3^2	=	9
-1^2	=	1
		40

(viii) $\Sigma\left(X_i - \overline{X}\right)^2 \sum\left(Y_i - \overline{Y}\right)^2$

$10 \times 40 = \underline{\underline{400}}$

(ix) $\sqrt{400} = \underline{20}$

(x) $\frac{15}{20} = \underline{\underline{0.75}}$

- Mention that the final answer received at the end of following the above steps is the Product Moment Co-relative co-efficient of the above two variables. Then explain the formula which is used to compute the correlative co-efficient.
- Explain the properties of the Product Moment correlative co-efficient.
- Involve the students in following activity to explain that the correlative co-efficient does not change through considering the linear transformations of the two variables.

Activity – 2

- Substract 5 from each value of the tail-length of mice and denote the new values received by x_i
- Substract 20 from each value of the weight of mice and denote the new value received by y_i
- Compute $\overline{X}$ and $\overline{Y}$
- Complete the following table.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_1 - \bar{x})(y_i - \bar{y})$	$(x_1 - \bar{x})^2$	$(y_i - \bar{y})^2$
Total						

- Subsitute the received summations to the following formula and compute the correlative co-efficient.

$$r = \frac{\sum\left[(x_i - \bar{x})(y_i - \bar{y})\right]}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Activity – 2 Solution.

$$\bar{X} = \frac{\sum x_i}{n} = \frac{15}{5} = 3, \qquad \bar{Y} = \frac{\sum y_i}{n} = \frac{25}{5} = 5$$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_1 - \bar{x})\ (y_i - \bar{y})$	$(x_1 - \bar{x})^2$	$(y_i - \bar{y})^2$
4	3	1	-2	-2	1	4
3	6	0	1	0	0	1
5	10	2	5	10	4	25
1	2	-2	-3	6	4	9
2	4	-1	-1	1	1	1
15	25			15	10	40

$$r = \frac{\Sigma\left[\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right]}{\sqrt{\Sigma\left(x_i - \bar{x}\right)^2 \Sigma\left(y_i - \bar{y}\right)^2}}$$

$$= \frac{15}{\sqrt{10 \times 40}}$$

$$= \frac{15}{20} \quad = \underline{\underline{0.75}}$$

Activity – 3

- Involve the students in following activity.
- Given below are the data related to the cost of advertising incurred annually by the business firms A B C D E and F engaged in the same industry. (in Rs. 000) and the annual profit earned (Rs. Millions)

Firm	A	B	C	D	E	F
Advertising Expenses (Rs.000)	31	33	28	31	35	34
Profit (Rs.000 000)	6	7	5	5	9	8

- Compute the correlative co-efficient using the following formula and interpret.

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum y_i^2 - (\sum y_i)^2\right]}}$$

x_i	y_i	xy	x^2	y^2
31	6	186	961	36
33	7	231	1 089	49
28	5	140	784	25
31	5	155	961	25
35	9	315	1 225	81
34	8	272	1 156	64
192	40	1 299	6 176	280

$$r = \frac{(6\times1299)-(192\times40)}{\sqrt{\left[(6\times6176)-192^2\right]\left[(6\times280)-40^2\right]}}$$

$$= \frac{7794-7680}{\sqrt{[(37056)-36864]\times(1680-1600)}}$$

$$= \frac{114}{\sqrt{192\times80}}$$

$$= \frac{114}{\sqrt{15360}}$$

$$= \frac{114}{123.9} \qquad = 0.92$$

- The value of the Product Moment Correlative Co-efficient being 0.92, it can be declared that there is a strong positive correlation between the cost of advertising and the profit of the business firm.

A Guideline to explain the subject matters :

- The linear correlation between two quantitative variables is evaluated using the Product Moment correlative co-efficient.
- Product Moment Correlative Co-efficient 'r' can be interpreted in following formula.

$$r = \frac{\Sigma[(x_i - \bar{x}).(y_i - \bar{y})]}{\sqrt{\Sigma(x_i - \bar{x})^2 . \Sigma(y_i - \bar{y})^2}}$$

- According to the above mentioned interpretation the Product Moment Correlative Co-efficient possesses the following properties.
  - A co-efficient, which is computed based on the extent of values of the two variables being deviated from their true mean values.
  - The value of Product Moment Correlative Co-efficient falls between – 1 and + 1.
  - Even though it is calculated using a linear transformation for each value of the two variables, the result does not change.
- When the 'sum' denoted by "S", $\left(X_i - \bar{X}\right)$ denoted by 'x' and $\left(Y_i - \bar{Y}\right)$ denoted by 'y'
- The above formula can be simply derived as follows.

$$r = \frac{Sxy}{Sxx.Syy}$$

- Following formulae can be derived by simplifying the above formula to compute the Product Moment correlative co-efficient.

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{\left(n\Sigma x^2 - (\Sigma x)^2\right)\left(n\Sigma y - (\Sigma y)^2\right)}}$$

n= number of pairs in the two variables.

$$r = \frac{\Sigma x_i y_i - n\bar{x}.\bar{y}}{\sqrt{\left(\Sigma x_1^2 - n\bar{x}^2\right)\left(\Sigma y_i^2 - n\bar{y}^2\right)}}$$

- According to the Product Moment Correlative Co-efficient the linear correlation can be identified in two directions as 'plus' and 'minus'.
- When the independent variable increases, increasing the dependent variable or when the independent variable decreases, decreasing the dependent variable is known as the 'plus' (positive) correlation.
- When the independent variable increases, decreasing the dependent variable or when the independent variable decreases, increasing the dependent variable is known as 'minus' (negative) correlation.
- If the Product Moment Correlative Co-efficient takes a value 0.75 or more, there is a strong linear correlation between the two variables.

- If the Product Moment Correlative Co-efficient takes a value 0.25 or less there is a very poor linear correlation between the two variables.
- Even though there is a linear correlation between two variables, a change in one variable can not be said causes for a change in the other variable. There are few reasons for the inability.

  1. When there is a meaningless relationship.

  Ex : Although the number of child births take place daily in Colombo National Hospital and the number of road accidents take place daily in Kandy city may seem to be positively correlated. Those two variables are not actually correlated.

  2. When the changes are occurred on a third factor rather than the considered variables.

  Ex : Since an increasing in both the variables that 'human life expectation' and 'usage of mobile phones' depend on the advancement of modern technology, the decisions made if the life expectations on usage of mobile phones are inappropriately.

  3. Once the changes of Y do not depend only on X.

  Ex : When considered the change in demand for a particular product based on the price of it, not only the price but also some other factors cause for the change in demand.

- In making business decisions these situations also should be taken in to consideration.
- Making decisions on Product Moment Correlative Co-efficient is not suitable at above situations.

Assessment and Evaluation :

- A businessman is in the intention that more the sales of news paper, grater the sales of sweets . He noted down the number of kilograms of news papers sold daily (x_i) and the number of kilograms of sweets sold as (y_i) during a week. The following results have been derived.

  $\Sigma x_i = 266 \qquad \Sigma y_i = 378 \qquad \Sigma x_i^2 = 10136 \qquad \Sigma y_i^2 = 21258 \qquad \Sigma x_i y_i = 14292$

(i) Compute the Product Moment Correlative co-efficient to examine the relationship between the sales of news papers and the sales of sweets.

(ii) Explain whether the intention of the businessman is ascertained or not with sufficient evidendce.

Competency 4.0 : Studies the Relations between variables and forecasts.

Competency Level 4.4 : Quantifies the Rank Correlative Co-efficient

No. of Periods : 04

Learning outcomes :

- Ranks the variables those are not quantitative.
- Interprets the Rank Correlative Co-efficients.
- Provides with instances where the Rank Correlative Co-efficient is applicable.
- Computes the Rank Correlative co-efficient between two non-quantitative variables.
- Explains the properties of Rank Correlative co-efficient.
- Explains the type of agreement or disagreement of the two variables using the Rank Correlative Co-efficient.
- Ranks the quantitative variables.
- Examines the agreement or disagreement between two quantitative variables ranked.

Guidelines for Lesson Planning :

- Involve the students in following activity in order to explain the Rank Correlative Co-efficient.

Activity 01

- Lead the students to Rank the five subjects they study on individual preference as I for the most favourite subject and 2 for the next … etc.
  - Accounting
  - Economics
  - Business Statistics
  - General English
  - General Information Technology

  N : B : Listed ubjects can be changed as needed.

- Display the following table on the board and call upon two students as one after the other and let them make their individual preference for each subject.

Subject	Preference of students	
	1	2
• Accounting • Economics • Business Statistics • General English • General Information Technology		

- Explain that the number values given by each student independently are called the 'ranks'.
- Ask from the students whether there is any relationship between the ranks given by the two students for the five subjects they learn.
- State that the below mentioned co-efficient can be used to examine the relationship of the ranks given by the two students in the above table and display the formula on the board.

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

State that, d_i - difference of corresponding ranks.
n - number of pairs of ranks.

- Lead the students to compute the Rank Correlative Co-efficient, using ranks given by the two students.
- Give an opportunity to the students to comment on the result they have received regarding the correlation (agreement) between the rankings.
- Discuss with the students regarding the situations in business field, where the Rank Correlative Co-efficient is applicable.
- Involve the students in following activity to explain how to rank the quantitative variable and compute the Rank Correlative Co-efficient.

Activity – 2

Marks received by 10 students for Statistics and History at a term test are given below.

Student	A	B	C	D	E	F	G	H	I	J
Statistics	30	50	25	30	60	70	80	65	75	85
History	50	60	30	40	70	50	90	60	40	80

- Rank the marks for Statistics as I for the highest and 2 for the 2nd highst … etc.
- Rank the marks for History also as I for the highest and two for the 2nd highst.. etc.
- Using those rankings and the formula compute the Rank Correlative Co-efficient and interpret.

Solution – 2

Student	Statistics		History		
	Marks X	Ranks RX	Marks Y	Ranks RY	$(X_i - \bar{X})^2$
A	30	8.5	50	6.5	4
B	50	7	60	4.5	6.25
C	25	10	30	10	0
D	30	8.5	40	8.5	0
E	60	6	70	3	9
F	70	4	50	6.5	6.25
G	80	2	90	1	1
H	65	5	60	4.5	0.25
I	75	3	40	8.5	30.25
J	85	1	80	2	1
Total					58

Rank Correlative co-efficient

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

$$= 1 - \frac{6 \times 58}{10(10^2 - 1)}$$

$$= 1 - \frac{348}{990}$$

$$= \underline{\underline{0.65}}$$

- Since the co-efficient takes a plus value it can be concluded that those students who clever for Statistics are also clever for History.

A Guideline to explain the subject matters :

- Relationship between ranked variables is evaluated using the Rank Correlative co-efficient.
- The relationship between non quantitative variables can be measured using the Rank Correlative Co-efficient.
- The Rank Correlative Co-efficient is interpreted as ;

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

  d_i = deviations of corresponding ranks

  n = number of pairs of observations

- In order to evaluate the correlation between the qualitative variables such as intelligence honesty. Beauty etc. . Rank Correlative Co-efficient is used.
- Few instances in the business field where Rank correlative co-efficient is applicable are as follows.
  - To study the agreement of the consumer attitude towards a product introduced to the market under various brand names.
  - To check whether a fair judgment has been given by two independent judges who have ranked a group of competitors/players/singers/artists etc….
  - To study whether an apprenticed officer has got experiences to the extent of a trained officer.
- Rank Correlative Co-efficient possess the following properties.
  - It shows the linear relationship between ranks.
  - If the rankings are perfectly agreed the co-efficient will be +1, where as it will be -1 when the rankings are perfectly disagreed.
  - The Rank Correlative Co-efficient can be a value between -1 and +1.
- Once only the ranks of two variables are available, the one and only technique to be applied in evaluating the correlation between those variables is the Rank Correlative Co-efficient.
- When the number of pairs of observations is exceeding 30, when all the considerable information between the variables haven't been utilized and when the data are available in grouped frequency distributions applying the Rank correlative co-efficient to evaluate the correlation between such variables is not practical.

Assessment and Evaluation :

The preferential ranks assigned by 2 consumers independent by for 6 television sets available for sale under 6 brand names are as follows.

Brand Name	A	B	C	D	E	F
Consumer-1	3	5	4	2	1	6
Consumer-2	4	5	3	1	2	6

- Explain whether there is any linear relationship between the two rankings using the relevant correlative co-efficient.

Competency 4.0 : Studies the Relations between variables and forecasts.
Competency Level 4.5 : Studies the Concept of "Regression".
No. of Periods : 06

Learning outcomes :

- Diffentiates between deterministic models and probabilistic models.
- Sates the dependent variable with respect to the independent variable in an equation (model).
- Interprets the "Regression"
- Differentiates simple Regression from Multiple Regression.
- Provide with instances for the situations where simple Regression and Multiple Regression are used.
- States the Population Regression Model.
- Introduces the variables, co-efficient and error – term of the Population Regression Model.
- States the estimated Regression Model.
- Introduces the variables and co-efficients of the estimated regression model.
- Describes the uses of Regression.

Instructions for Lesson Planning :

- Provide with few correlated variables to the students and hold a discussion as follows.
  - The change in price of a certain commodity and the demand for that commodity.
  - Various factors cause for the demand of a commodity.
    - Price of the considered commodity -Px
    - Consumer taste – T
    - Consumer income – Y
    - Possibility of changing the price of that commodity in future - Ex

      Point out that the demand function can be stated as follows.

      If the quantity of demand is Qd.

      The Demand Function is Qd = f (Px. T. Y. Ex)
- Rain-fall values and the paddy harvest.
- Following factors cause for the paddy harvest.
  - Rainfall – R
  - Nature of the crop – C
  - Consumption of fertilizer – F

- Soli – S
- If the paddy harvest is – H

  point out that H = f (R.C.F.S.)

- Cost of advertising incurred by a firm and the sales income
  - Following factors cause for the sales income of a firm.
    - Advertising Expenses – A
    - Price of the commodity – Px
    - Consumer taste - T

  If the sales income is y, point out that

  Y = f (A.Px.T)

- Explain that there are instances where only one independent variable affects for the dependent variable and such instances are very rare.
- Point out that there are more instances with more than one independent variable affects for the dependent variable in practice as mentioned above.
- Group the students appropriately and involve them in following activity.

Activity – I

(i) Consider that

Employee wage - W

Employee Training - T

Employee education level – E

(ii) Consider that

Height of students - H

Weight of students – W

Gender - S

(iii) Consider that

Quantity of units provided in a machine – Q

Activating time of the machine - T

- State an appropriate functional relationship using an equation or a formula.
- Explain whether there is a deterministic relationship or a probabilistic relationship among the variables at each instance.
- In an instance where there are few correlated independent variables while all the other independent variables except the considering independent variable are held constant, explain that the function of dependent variable (y) on the considering independent variable (x) is $y = f(x) + u$

- Involve the students in following activity to explain the fact that the relationship between two or among few correlated variables can be a straight linear model or a curved model.

Activity – 2

- Provide with the following equations to the students.

(i) $Qd = 10 - 2p$

Qd = Quantity of demand

P = Price

(ii) $Qs = -50 + 5p$

Qs = Quantity of supply

P = price

(iii) $y = (x - 20)^2$

y = Average manufacturing cost.

x = Number of units manufactured

Complete the following tables relevant to each instance mentioned above.

(i)

P	2p	10-2p	Qd
2			
6			
8			
10			
14			

(ii)

P	Sp	- 50 + 5P	Qs
2			
6			
8			
10			
14			

(iii)

P	$(x-20)$	$(x-20)^2$	y
100			
120			
150			
165			
175			
200			

- Construct the graphs using the details in the completed table representing independent variable on horizontal axis and dependent variable on vertical axis.
  - Point out that only the straight linear relations are considered at Advance Level.
  - Explain to the students what 'Regression' is.
  - Explain that Regression is in two types as simple Regression and Multiple Regression with appropriate instances for each type.
  - Explain that the algebraic relationship between two variables related to population details can be represented in a Regression model as the Population Regression Model which is not practical; so that a Regression Model can be built using a sample and such a relationship is known as an estimated Regression Model.
  - Discuss with the students about the uses of fitting a Regression Line.

A Guideline to Explain the subject matters :

- In an instance where only two correlated variables are existing if the value of the dependent variable is perfectly determined on the value assigned to the independent variable, such a relationship is known as a Determination Relationship.
- In an instance where few correlated variables (more than two) are existing, if there are few independent variables affecting towards a change in the dependent variable, such a relationship is known as a probabilistic relationship.
- In an instance where few correlated variables are existing the function of Dependent variable related to one of the independent variables can be expressed in a model assuming that the other factors are held constant.
- That can be either a straight linear model or a curved model.

  Ex : If the demand of a commodity is influenced by the factors such as;
  - Price of the considering commodity - Px
  - Consumer income (-y)

- Consumer Taste – (T)
- Price of other commodities (Py)

  Qd = f (Px. Y. T. Py)

- If it is assumed that the other factors except the price of considering commodity are held constant, the change in the quantity of demand related to the price of considering commodity can be represented in the following model.

  $Qd = f(Px) + (y.T.Py)$

- When considered all the other factors except price of considering commodity as **U** the demand function can be stated as follows;

  $Qd = f(Px) + U$

- Expressing the relationship existing between two variables or among few variables in an algebraic equation is called "Regression" it is in two types.
  - Simple Regression
  - Multiple Regression
- Expressing the relationship existing between two variables in an algebraic equation is known as Simple Regression.
- Some instances where simple Regression is applicable are ;
  - To represent the relationship between the age of vehicles and the cost of maintenance.
  - To represent the relationship between the experience of the employees and their salary scales.
  - To represent the relationship between the educational level of employees and their salary scales.
  - To represent the relationship between sales income and cost of advertising of a business firm.
- Expressing the relationship existing among more than two correlated variables in an algebraic equation is known as "Multiple Regression."
- Some instances where multiple regression is applicable are;
  - To represent the relations existing among consumer expenses, revenue, size of the family, habits and hopes, etc….
  - To represent the cost of maintenance of vehicles, age of vehicles, brand of vehicles and way of usage etc ….
  - To represent the relations existing among the demand for a commodity the price, of that comodity, consumer taste, consumer income and future price expectations etc. .

- Algebraic relationship between two correlated variables can be represented in following regression model.

(i) population regression model

$Y = \beta o + \beta_1 X + U$

Y = dependent variable

X = independent variable

Bo = intercept of the straight line. That means value of y when X = 0

β_1 = Gradiant of the straight line

U = Error. A straight linear relationship does not exist between **X** and **Y** at most of the practical instances. When a linear relationship is fit for such variables an error is emerged.

(ii) Estimated Regression Model

- The estimated regression model constructed on sample data is as follows.

$\hat{y} = \beta_0 + \beta_1$

$\hat{y}$ = Estimated dependent variable

β_0 = intercept of the estimated regression line.

β_1 = Gradient or the slope of the estimated regression line. That means the rate of change in dependent variable.

X = independent variable.

Uses of fitting a Regression Line can be sated as follows.

- Ability of awaring the type of linear relationship existing between two variables. (Whether a proportionate relation or an inverse relation).
- Ability of forecasting the value of a related variable using a details known variable.

Competency 4.0 : Studies the Relations between variables and forecasts.
Competency Level 4.6 : Uses the Free Hand Method to fit a Simple Regression Line.
No. of Periods : 02

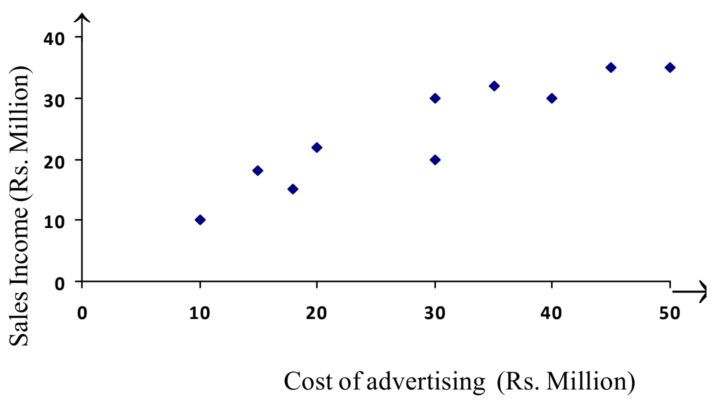
Learning outcomes :

- Introduces the Free Hand Method.
- Fits a Regression line on Free Hand Method.
- Derives a Regression Line on free hand method on a scatter diagram constructed for given data.
- States the advantages and disadvantages of fitting a Regression Line.

Instructions for Lesson Planning :

- Involve students in following activity.
  The cost of advertising and the sales income of a business firm for the last 10 years (Rs. Millions) are as follows.

Cost of advertising (Rs. Million)	10	18	15	20	30	30	35	40	50	45
Sales Income (Rs. Million)	10	15	18	22	20	30	32	30	35	35



- Lead the students to draw the scatter diagram using the above data
- Provide with instructions to draw a straight line which is more appropriate and passing closer to all the points on the scatter diagram.

- Lead the students to derive this straight line in a manner such that a similar number of points falling on either side of it as far as possible.
- Point out that deriving a Regression line easily in this manner without following any mathematical technique is known as the **free hand method.**
- Point out further that the various lines derived by each student as desired are different from one another.
- Guide the students to compute the sales income in next year if an amount of Rs. 55 000 000/= is incurred on advertising.
- Explain that the free hand method is useful, when a regression line is urgently required and also it is easy to have an awareness about the correlation between two variables using this regression line.

A Guideline to Explain the subject matters :

- After plotting the pairs of observations related to a given pair of variables appropriately on a scatter diagram, drawing a straight line passing very closer to all or many of those points as far as possible is known as fitting a regression line on Free Hand Method.
- Here, it's more appropriate to derive this line in such a manner that similar number of points falling on either side of it.

Advantages of Free Hand Method

- Ability to aware the correlation between two variables
- Ability to fit without using any mathematical techniques
- Ability to derive the line very easily

Disadvantages of Free Hand Method

- Getting a biased regression line since it depends on individual desire
- Forecasting not being accurate, since it is not supported by a sound mathematical base

Competency 4.0 : Studies the relations between variables and forecasts.
Competency Level 4.7 : Uses the Least Square Method to fit a Regression Line.
No. of Periods : 04

Learning outcomes :

- Introduces the Least Square Method.
- Derives the regression equation on Least Square Method for given data.
- Describes the Regression Co-efficient.
- Estimates the dependent variable on the independent variable using the Regression Line fit on Least Square Method.
- States the advantages and disadvantages of fitting a regression line on Least Square Method.

Instructions for Lesson Planning :

- Involve the students in following activity.

Activity – I

- Provide with the following data set to the students.

X	2	3	5	6	8	0
Y	6	5	7	8	12	11

- Represent the above data in a scatter diagram.
- Plot an appropriate straight line on the scatter diagram as you like.
- Derive the vertical distance from each point on the scatter diagram to the respective point on that straight line.
- Derive the sum of those vertical deviations.
- Derive the sum of the squares of those vertical deviations.
- Hold a discussion highlighting the following facts in accordance with the above activity.
    - that the straight linear formations are very rare in practice when the related variables are depicted graphically.
    - that the deviations are emerged once a straight line is fit matching to the data set.
    - Explain that those deviation can be computed on two ways.
        - Keeping – x – constant raking the vertical deviations of – y.
        - Keeping – y – constant taking horizontal deviations of x.

- Explain that those deviations are called "errors" (U_i)
  - Point out that $\sum U_i \simeq 0$
  - Point out that the variations can be numerically evaluated through computing $\sum U_i^2$
  - Point out that the best fit straight line for the data set is the line which is fit so as to have the least variance.
  - Explain that the technique of 'calculus' in Mathematics is used to derive the points where the variation is minimized.
  - Acertain that a straight line is fit on the scatter diagram so as to minimize the variations under the least square method.

Activity – 2

Using the data set given for the Activity – I above lead the students to compute the following table.

X	Y	XY	X^2
2	6		
3	5		
5	7		
6	8		
8	12		
9	11		
$\sum X =$	$\sum Y =$	$\sum XY =$	$\sum X^2 =$

- Find the sum of the values on column X - $(\sum X)$
- Find the sum of the values on column Y - $(\sum Y)$
- Find the sum of the values received by multiplying the values on X and Y columns. $(\sum XY)$
- Find the values received by making the values square in column x as $(\sum X_i^2)$
- Substitute those results in the following pair of equations.

$$\sum Y = n\hat{\beta}_0 + \hat{\beta}_1 \sum X \text{.....................(1)}$$

$$\sum XY = \hat{\beta}_0 \sum X + \hat{\beta}_1 \sum X^2 \text{...........(2)}$$

n is the number of pairs of observations.

- Derive the values for $\hat{\beta}_0$ and $\hat{\beta}_1$ solving this pair of simultaneous equations.
- Substitute those values derived for $\hat{\beta}_0$ and $\hat{\beta}_1$ of the following equation.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Solution Activity – 2

X	Y	XY	X^2
2	6	12	4
3	5	15	9
5	7	35	25
6	8	48	36
8	12	96	64
9	11	99	81
33	49	305	219

$$\sum X = 33 \qquad \sum Y = 49 \qquad \sum XY = 305 \qquad \sum X^2 = 219$$

Once these values are substituted in the above pair of equations they are as follows.

$$49 = 6\hat{\beta}_0 + 33\hat{\beta}_1 \text{..............................(1)}$$

$$305 = 33\hat{\beta}_0 + 219\hat{\beta}_1 \text{..........................(2)}$$

When this pair of equations are solved values for $\hat{\beta}_0$ and $\hat{\beta}_1$ are received.

$(1) \times 33 : 1617 = 198\beta_0 + 1089\beta_1$.................(3)

$(2) \times 6 : 1830 = 198\hat{\beta}_0 + 1314\hat{\beta}_1$...............(4)

$(4) - (3) : 213 = 225\beta_1$

$= \underline{\underline{0.95}}$

By substituting $\hat{\beta}_1$ =0.95 In the (i) st equation.

$$49 = 6\hat{\beta}_o + 33\hat{\beta}_1$$

$$49 = 6\hat{\beta}_o + 33 \times 0.95$$

$$49 = 6\hat{\beta}_o + 31.35$$

$$49 - 31.35 = 6\beta_O$$

$$6\hat{\beta}_O = 17.65$$

$$\hat{\beta}_O = \frac{17.65}{6}$$

$$\hat{\beta}_O = 2.94$$

$$\hat{y} = \underline{\underline{2.94 + 0.95x}}$$

Activity – 3

- Involve the students in following Activity.
- Involve in the following exercise according to the regression line derived in Activity – 2 above.
- Derive the rate of changing y, when x changes in I unit.
- Derive the value of y when x = 12.
- Plot the Least Square Regression line on the scatter diagram which was constructed in activity – I
- Point out the advantages and disadvantages of deriving a regression line for a data set.
- Hold a discussion highlighting the facts mentioned below.
- The rate of changing y, when x changes in I unit is 0.95.
- Value of y when the value of x is 12.

$$\hat{y} = 2.94 + 0.95 \times 12$$
$$= 2.94 + 11.40$$
$$= \underline{\underline{14.34}}$$

A Guideline to explain the subject matters :

- The number of straight lines that can be derived in a scatter diagram on free had method is numerous (infinite)
- The method of fitting a straight line that will be most appropriate to a data set so as to minimize the sum of the squares of the vertical deviations from each and every point on the scatter diagram to the straight line drawn is known as the Least Square Method.
- The regression equation on Least Square Method for a given data set is derived as follows.
  - Since there is not a straight linear relation between X and Y most probability in practice the relationship between them cannot be expressed in the model $Y = \beta_0 + \beta_1 X$

    Therefore it will be more appropriate to use an error term denoted by 'U' to represent the relationship between X and Y.
  - Then the relationship between x and y can be mentioned as

    $Y = \beta_0 + \beta_1 X + U$

    $U = \left(Y - \hat{Y}\right)$

    $\sum U^2 = \sum\left(Y - \hat{Y}\right)^2$

Since $Y = \beta_0 + \beta_1 X$ it can be substituted in the above equation as follows.

$$\sum u^2 = \sum(y - \beta_O - \beta_1 x)^2$$

- Applying differentiation in Mathematics to minimize the sum of the equation of errors the following pair of normal equations can be derived.

$$\sum y = n\beta_O + \beta_1 \sum x$$
$$\sum xy = \beta_O \sum x + \beta_1 \sum x^2$$

- Solving this pair of equations for β_0 and β_1, $\hat{\beta}_0$ and $\hat{\beta}_1$ and can be derived as follows.

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$\hat{\beta}_O = \bar{y} - \beta_1 \bar{x}$$

- $\hat{\beta}_1$ is knowm as the regression co-efficient.
- The rate of changing the dependent variable when the independent variable changes in I unit is expressed by the regression co-efficient.
- Once any value of independent variable is substituted in the regression equation the corresponding value of dependent variable can be estimated. That is known as forcasting.
- Advantages and disadvantages of fitting a regression line on Least Square Method are as follows.

Advantages.

- Not being subjective.
- Ability of deriving a straight line minimizing the possible errors.
- Ability of understanding the rate of changing the dependent variable on a change of independent variable through Regression co-efficient.
- Ability to aware whether the linear correlation between two variables proportionate or inverse.

Disadvantages.

- Being expensive when considered relative to the Free Hand Method.
- Spending much time.
- Not being flexible, even though a straight line is inappropriate to the data set, a straight linear trend line is necessarily derived through Least Square Method.
- An advanced mathematical knowledge being needed.
- Being a method that can be comprehended only by intelligentia.

Competency 4.0 : Studies the Relations between variables and forecasts.
Competency Level 4.8 : Examines the Goodness of Fit of a Regression Line
No. of Periods : 08

Learning outcomes :

- Interprets the co-efficient of determination.
- Computes the co-efficient of determination using a Regression Line fit.
- Explains the goodness of fit of the regression line using the co-efficient of determination computed.
- Estimates the corresponding value of 'Y', the dependent variable using the estimated regression line at a given value for independent variable – 'X'.

Instructions for Lesson Planning :

- Involve the students in following activity.

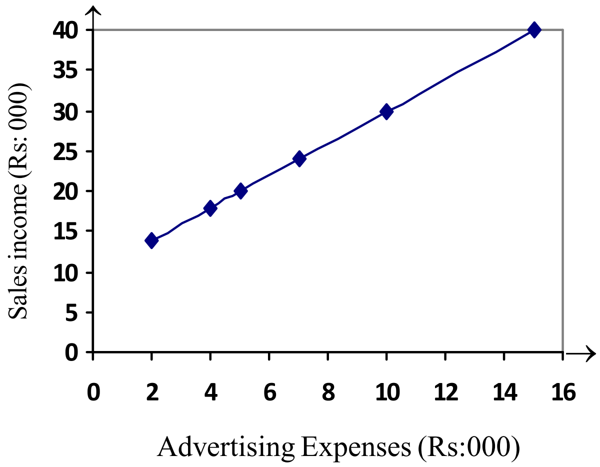
Activity – I

- Provide with following data set to the students.

Advertising Expenses – X (Rs. 000)	Sales income - Y (Rs. 000)
2	14
4	18
5	20
7	24
10	30
15	40

- Plot the above mentioned data in a scatter diagram.
- Fit the regression line for the scatter diagram on least square method.
  ($\sum x = 43, \ \sum y = 146 \quad \sum x^2 = 419, \ \sum xy = 1268$)
- Plot that regression line on the scatter diagram.
- Point out whether the regression line fit for this scatter diagram is appropriate to that scatter diagram.

Solution - Activity – 1



$$\hat{b}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$= \frac{(6\times 1268)-(43\times 146)}{(6\times 419)-43^2}$$

$$= \frac{7608-6278}{2514-1849}$$

$$= \underline{\underline{2}}$$

$$\hat{a} = \overline{y} - \hat{b}\overline{x}$$

$$= 24.33 - (2\times 7.167)$$

$$= 10$$

$$\hat{y} = \underline{\underline{10+2x}}$$

- Discuss with students about these facts.
  - There is no any error in the above situation.
  - It is a perfect straight line'
  - There are hardly any such relations found in practical field.

Activity – 02

Provide with the following data set to the students.

Advertising Expenses – X (Rs. 000) X	Sales income - Y (Rs. 000) Y
02	10
04	15
05	20
06	25
10	35
15	45

- Plot these data in a scatter diagram
- Consider the following values.

$$\left(\sum X = 42 \quad \sum Y = 150 \quad \sum XY = 1355 \quad \sum X^2 = 406\right)$$

- Plot the regression line using the least Square Method on the scatter diagram.
- Point the vertical distance (deviations) from each point to the regression line.
- Discuss with the students to which extent is that regression line derived suitable for the scatter diagram.
- Compute corresponding $\hat{Y}$ values substituting each value of X in the regression line derived.
- Derive the column $(\hat{y} - \bar{y})$
- Derive the column $(\hat{y} - \bar{y})^2$
- Derive the column $(y_1 - \bar{y})$
- Derive the column $(y_i - \bar{y})^2$
- Derive the co-efficient of determination (R^2) using the following formula.

$$R^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Solution Activity – 2

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{(6 \times 1355) - (42 \times 150)}{(6 \times 406) - 42^2}$$

$$= \underline{\underline{2.72}}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$= 25 - 19.04$$

$$= 5.96$$

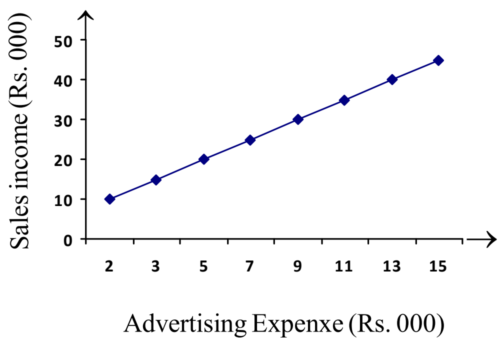
When X = 5

$$\hat{y} = 5.96 + (2.72 \times 5)$$

$$\hat{y} = 19.56$$

$$\hat{y} = 5.96 + (2.72 \times 15)$$

$$\underline{\underline{\hat{y} = 46.76}}$$



x	y	$\hat{y}$	$(\hat{y}-\bar{y})$	$(\hat{y}-\bar{y})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
2	10	11.4	-13.6	184.96	-15	225
4	15	16.84	-8.16	66.58	-10	100
5	20	19.56	-5.44	29.59	-5	25
6	25	22.28	-2.72	7.40	0	0
10	35	33.16	8.16	66.58	10	100
15	45	46.76	21.76	473.50	20	400
42	150			828.61		850

$$R^2 = \frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y_i-\bar{y})^2}$$

$$= \frac{828.61}{850}$$

$$R^2 = \underline{\underline{0.97}}$$

Activity – 3

- Lead the students to compute the co-efficient of determination of the data set considered in the activity I – above.

Activity – 3 - Solutions

x	y	$\hat{y}$	$(\hat{y}-\bar{y})$	$(\hat{y}-\bar{y})^2$	$(y-\bar{y})^2$
2	14	14	-10.33	106.71	106.71
4	18	18	-6.33	40.07	40.07
5	20	20	-4.33	18.75	18.75
7	24	24	-0.33	0.11	0.11
10	30	30	5.67	32.15	32.15
15	40	40	15.67	245.55	245.55
	146			443.34	443.34

$$R^2 = \frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y_i-\bar{y})^2}$$

$$= \frac{443.34}{443.34}$$

$$= \underline{\underline{1}}$$

Activity – 4

- Lead the students to compute the co-efficient of determination using the following formula as well using the data in activity 2- above.

Solution – Activity – 04

x		y	x^2	y^2
2	10	4	100	
4	15	16	225	
5	20	25	400	
6	25	36	625	
10	35	100	1225	
15	45	225	2025	
42	150	406	4600	

$$R^2 = \hat{b}^2\left[\frac{n\Sigma x^2-(\Sigma x)^2}{n\Sigma y^2-(\Sigma y)^2}\right]$$

$$= 2.72^2\left[\frac{6\times 406-42^2}{6\times 4600-150^2}\right]$$

$$= 7.3984\left[\frac{2436-1764}{27600-22500}\right]$$

$$= 7.3984\ \frac{[672]}{5100}$$

$$= \underline{\underline{0.9748}}$$

- Explain the fact that there is no any error when all the points on the scatter diagram fall on the regression line and that line is the best regression model, so that the value of co-efficient of determination is 1.

- Explain that the regression line which is fit, when there is no any linear relationship between the two variables is not suitable at all for the scatter diagram and further point out that the co-efficient of determination is zero at such situations.
- Hence, explain that when the value of co-efficient of determination is closer to +1, the regression line is more suitable for the sctter diagram and when it is closer to zero (o) the suitability of regression line for the scatter diagram is very poor.

Activity – 5

- Pay the attention of students to the estimated regression line based on the data set considered in activity – 2

  $\hat{y} = 5.96 + 2.72x$
- Assign them to estimates the sales income when the cost of advertising is Rs. 16 000
- Discuss with the students regarding the reliability of that estimated sales income.
- Lead them to estimate the sales income when the cost of adverting is Rs. 100 000/-
- Discuss with the students regarding the reliability of that estimated sales income.

Activity 5 – Solution

$$\hat{y} = 5.96 + 2.72x$$
$$\hat{y} = 5.96 + (2.72 \times 16)$$
$$\hat{y} = \underline{\underline{49.48}}$$

- It can be declared that the sales income will be Rs. 49 480/- at 97% level of confidence. The confidence level 97% is determined on the value of co-efficient of determination.

$$\hat{y} = 5.96 + 2.72x$$
$$= 5.96 + (2.72 \times 100)$$
$$= \underline{\underline{277.96}}$$

- Even though the sales income is estimated to be Rs. 277 960/- at Rs. 100 000/- cost of advertising according to the regression line, this value is not meaningful, because the cost of advertising is not the one and only factor causes for the sales income. Even if the cost of advertising is increased with a very large amount a proportionate growth in sales can not be expected practically in the real business field.

A Guideline to explain the subject matters :

- When all the points on the scatter diagram do not fall on the regression line, it is very important to investigate whether the regression line is appropriate to the scatter diagram.
- The co-efficient of determination is used for this purpose.
- The co-efficient of determination is a common measure which is used to evaluate the goodness of fit of the regression line.
- It can be reveled the proportion (percentage) of the total variation of the dependent variable y which is described through the independent variable – x, using the co-efficient of determination.
- The value of co-efficient of determination can be derived using the following formula.

$$R^2 = \frac{\Sigma(\hat{y} - \bar{y})^2}{\Sigma(y_i - \bar{y})^2}$$

- The following optional formula also can be used for calculating the value of co-efficient of determination.

$$R^2 = \hat{b}\left[\frac{n\sum x^2 - (\sum x)^2}{n\sum y^2 - (\sum y)^2}\right]$$

- The value of the co-efficient of variation falls between 0 and +1.

$$O \leq R^2 \leq +1$$

- The value of dependent variable can be estimated by substituting the value of independent variable in the regression equation.
- The value derived for dependent variable by substituting extra ordinary values for independent variable relative to the given data can not be expected as a practical estimation.

Competency 5.0 **:** Demonstrates the preparedness to face business risk

Competency Level 5.1 : Manipulates 'probability' concept using business uncertainties.

No. of Periods **:** 02

Learning outcomes :

- Interprets probability as a statistical technique of measuring uncertainty.
- Lists the business incidents.
- Explains the certain events involved in business.
- Explains the uncertain events involved in business.
- Explains the events that will never happen (impossible events).

Instructions for Lesson Planning :

- Produce few uncertainties we face in our – day-to-day life to the class.
  - Life expectation of a person
  - Getting through in an examination
  - Victory of an election
  - Possibility of rains tomorrow
- Explain that the outcomes of many occurrences take place around us in real life situations cannot be expressed certainly, no any events can be expressed with a perfect certainty and further that decision making is very convenient, if the outcome of each and every occurrence is certain.
- Point out with sufficient instances that there are uncertainties in business field as well as in occurrences come across in real life situations.
- Involving the students in following activity.
- Enquire into the following situations.
  - Possible demand for a newly introduced product to the market
  - Strategic movements of competitors in a business
  - Receiving the ordered batch of items in due time
  - Selling all the items produced to the market during the forthcoming quarter
- In this manner pay attention to the occurrences such as profit/loss, selling/not selling receiving the orders/not receiving orders, increasing the demand / decreasing the demand, supply being limited/ supply not being limited etc., affiliated to the business world.
- Explain the possible problems to be arisen in making decisions at each situation mentioned above.

- Mention the measures to be followed in making optimal decisions before uncertainties.
- Group the students appropriately and provide with the following topics.
  - Certain events
  - Uncertain events
  - Impossible events
- Lead the students to disclose the followings :
- Explain in brief about the topic received by the group.
- Provide with appropriate instances for the event received.
- Assign a quantitative (numerical) value for each event if the possibility of occurring is measureble.

A Guideline to explain the subject matters :

- Most of the occurrences we experience in our day-to-day life are involved with uncertainty.

  Ex : • getting through in an examination
  - victory of an election
- There are uncertainties in making decisions related to various business events.

  Ex : • The batch of items manufactured being entirty sold
  - A batch of raw materials ordered being received on due date
- Measuring the uncertainty in terms of a numerical value is essential in order to make optimal decisions.
- The technique of measuring the uncertainty in terms of a numerical value is 'Probability'.
- Probability is the quantitative measure of evaluating the possibility of occurring or not occurring a particular event.
- Probability of any event can be a value between 0 and +1.
- If the possibility of an event is certain the probability of that event is 'I'
- The probability of an impossible event is '0'.
- Hence the probability of an uncertain event is a value between 0 and I.
- Falling down any object thrown up, falling either head or tail of a coin tossed are some of the certain events.
- A person living for ever, falling 7 on the uppermost surface of a thrown up die numbered from 1 to 6 are some examples for impossible events.
- Getting a winning chance of a lottery ticket, getting rains in next month etc,. are some examples for uncertain events.

Competency 5.0 : Demonstrates the preparedness to Face business risk

Competency Level 5.2 : Separates Random Experiments.

No. of Periods : 02

Learning outcomes :

- Explains the difference between deterministic experiments and random experiments.
- Provide with suitable examples for deterministic experiments and random experiments separately.
- Highlights the instances for random experiments from the business field.
- Interprets the sample space.
- Represents the sample space through Venn diagrams, tree diagrams, point graphs and picture graphs etc.
- Explains what 'a trial' is.

Instructions for Lesson Planning :

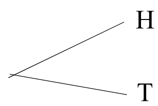
- Produce the following experiments to the class.
    - throwing up a coin
    - choosing an item from a production process
    - burning a piece of magnesium
    - Throwing up a die
- Hold a discussion highlighting the following facts.
    - Once a coin is tossed the fact whether the head will fall or the tail will fall cannot be exactly said.
    - Before an item is drawn out from a production process and checked the fact whether it is defective or non defective cannot be exactly said.
    - A piece of magnesium is burnt to show that an ash colour stuff is remained at the end of combustion.
    - A balloon is blown and made a press on it to show that its shape gets changed since there is a volume of air contained in it.
    - Once a die is rolled the value received cannot be exactly said.
- Lead the students to separate the experiments of which the outcomes cannot be exactly said prior to the experiment is conducted and experiments with an exact outcome among the experiments mentioned above.
- Explain that the experiments of which the outcomes cannot be exactly said prior to conduct the experiment are known as random experiments, where as the experiments

of which the outcomes are known prior to conduct the experiment and even if it is repeated the same outcome is derived are known as deterministic experiments.

- Inform the students that 'probability' is based on random experiments.
- Engage with students in following activities to interpret and represent the sample space related to a random experiment.
- Involve in following activity related to the experiment of tossing a coin.

1. Ask from the students about the possible outcomes of tossing a coin, before it is tossed by the teacher. Then write down all possible outcomes of that experiment on the board asking from students. Represent those outcomes using the following techniques and introduce it as the sample space.

    - Using sets S = {H, T}
    - Using a tree diagram

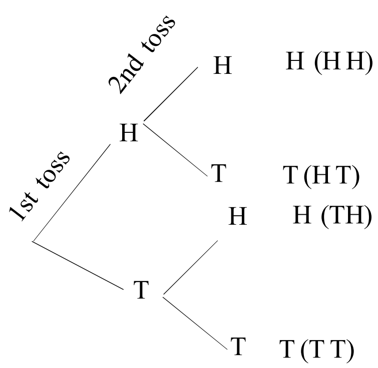


    - Using a point graph

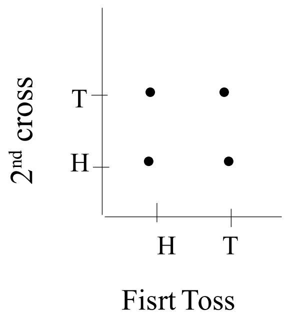
    - Using a picture graph

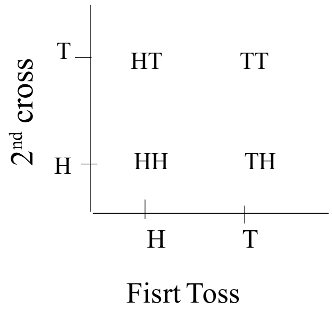
2. Ask from the students about the outcomes of the experiment of tossing a coin twice and represent the sample space using each technique mentioned above.

    - Using sets S = {(H, H) (H, T) (T, H) (T, T)
    - Using tree diagrams



- Using a point graph
- Using a picture graph





- Involve the students in following activity providing with the given experiments.

Activity I.

Seven bulbs including three defectives are packed in a box.

- Testing any bulb drawn in random.
- Testing two consecutive bulbs drawn in random after replacing the first drawn bulb.
- Testing three consecutive bulbs drawn in random without replacing the previous bulb drawn.

Activity II.

- Three red beads, two yellow beads and one blue bead with the same size are contained in a box.
  - Drawing a bead and noticing the colour.
  - Testing two consecutive beads and noticing the colour, replacing the first bead drawn out.
  - Testing three consecutive beads drawn in random without replacing the first drawn beads.
- Lead the students to represent all possible outcomes of each random experiment in following techniques.
  - Using sets
  - Using tree diagrams
  - Using point graphs
  - Using picture graphs
- Discuss with the students highlighting the following problems.
  - What is/are the most appropriate technique/s to be used to represent the sample space of any random experiment?

- Can a particular random experiment be repeated under identical circumstances?
- What are the random experiments that can be repeated among the situations (experiments) given.
- When a random experiment is repeated, by which name is a single experimenting turn introduced?

Activity –I (solution)

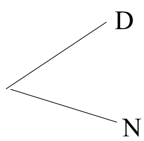
- Testing a bulb drawn in random from a box containing seven bulbs including three defectives.

D – Defective bulbs

N – Non – defective bulbs

- Sample space by sets   S = {N, D}

- By tree diagram



- By point graphs

- By picture graphs

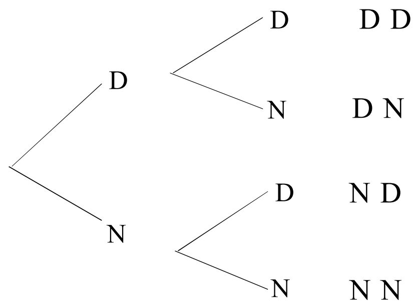
- Testing two consecutive bulbs drawn in random replacing the first drawn bulb.

- Sample space by sets

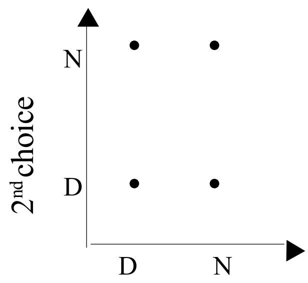
  S = {(D D) (D N) (N D) (N N}

- Sample space by a tree diagrams

1st choice 2nd choice Sample pionts

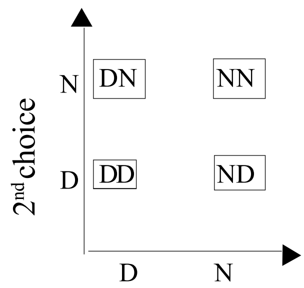


- Sample space by a point graph



1st choice

- Sample space by picture graphs



1st choice

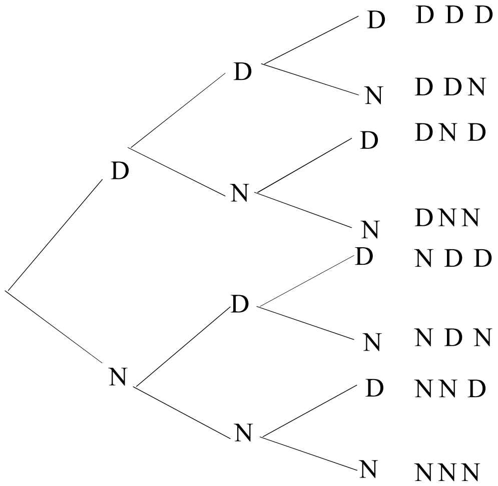
- Testing three consecutive bulbs without replacing the previously drawn bulb.

  - By sets.

    S = {(DDD)(DDN)(DND)(DNN)(NDD)(NDN)(NND)(NNN)}

- By a tree diagrams

1st draw 2nd draw 3rd draw sample points



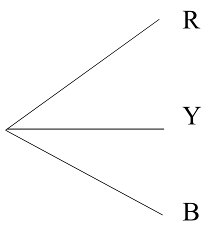
- This sample space cannot be represented by using a point graph or a picture graph, because three dimensional situations cannot be displayed in a Cartesian co-ordinate plane.

Solution for Activity – 2

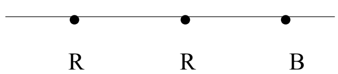
Checking the colour of the bead drawing out a bead from a box containing three red beads two yellow beads and one blue bead.

- By sets. S = {R, Y, B}

- By a tree diagram.

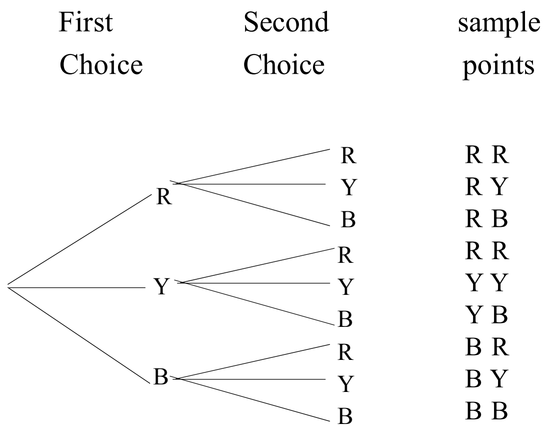


- By a point graph

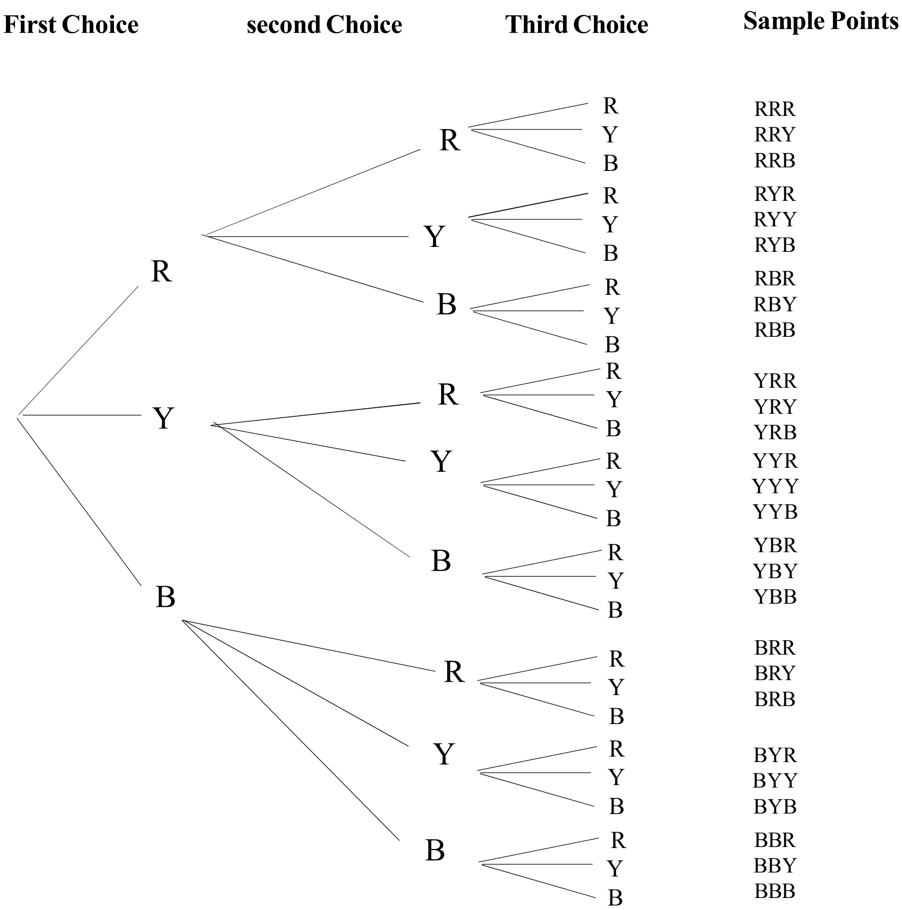


- Sample space of the random experiment of drawing two consecutive beads with replacing the first bead drawn before the second draw.
  - By sets S = {(RR)(RY)(RB)(YR)(YY)(YB)(BR)(BY)(BB)}

- By a tree diagram



- Sample space of the Random experiment of drawing three consecutive beads from a bag containing three Red beads two Yellow beads and one Blue bead, without replacing each bead drawn, using tree diagram.



- By sets :

$$S = \left\{ \begin{matrix} (RRR)\ (RRY)\ (RRB)\ (RYR)\ (RYY)\ (RYB)\ (RBR)\ (RBY)\ (RBB)\ (YRR)\ (YRY)\ (YRB)\ (YYR)\ (YYY) \\ (YYB)\ (YBR)\ (YBY)\ (YBB)\ (BRR)\ (BRY)\ (BRB)\ (BYR)\ (BYY)\ (BYB)\ (BBR)\ (BBY)\ (BBB) \end{matrix} \right\}$$

- Once the number of experimenting times is exceeding the sample space can not be represented in a Cartesian co-ordinate plane or in a picture graph.

A guideline to explain the subject matters :

- The experiments of which the receivable outcome cannot be exactly said prior to the experiment are called random experiments (but, the set of possible outcomes can be stated).

  1. When a balanced die is thrown, the value appeared on the uppermost surface of it
  2. An item produced in a machine being defective or non-defective

- The experiments of which the possible outcomes can be exactly said even without the experiment is conducted are called deterministic experiments

  Ex : 1. Once a tree leaf is covered with polythene to check whether it contains water, observing water vapour inside the polythene cover

  2. To check whether an ash colour stuff will be remained, once a piece of magniszium is burnt

- Probability is based on random experiments.
- The set of all possible outcomes of a random experiment is called the 'sample Space'.
- The sample space can be represented on any of the following methods.
  - Using sets (venn diagrams)
  - Using a grid (dotted graph)
  - Using tree diagrams
  - Using picture graphs
- Representing the sample space by using sets.
  - All possible outcomes of a random experiment which is conducted only once can be represented in sets as follows. The sample space is indicated by S.

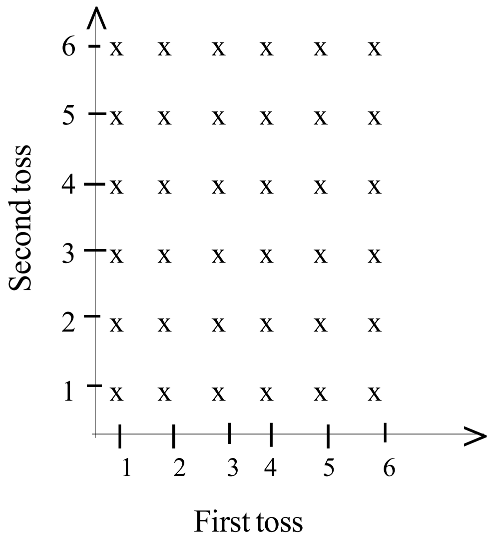
  Ex : (i) Sample space containing all possible outcomes when a balanced die is thrown only once

  S = {1,2,3,4,5,6}

  (ii) Sample space containing all possible outcomes of throwing a balanced coin only once

  S = {H,T}

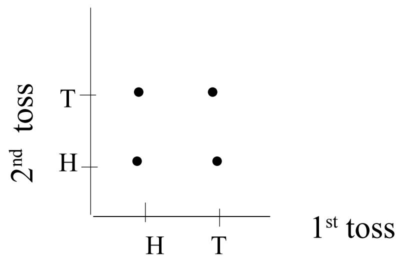
- Representing the sample space using point graphs.
    - All the possible outcomes of a random experiment conducted only for two times can be represented using a grid as follows.
    - Possible outcomes of the experiment of tossing a die twice.



All these outcomes can be represented using sets as well.

S = { (1,1) (1,2),(1,3)(1,4)(1,5)(1,6)(2,1)(2,2)(2,3) (2,4) (2,5) (2,6)
(3,1)(3,2)(3,3) (3,4) (3,5) (3,6) (4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
5,1) (5,2) (5,3) (5,4) (5,5) (5,6), (6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

Ex : All the possible outcomes of the experiment of tossing a coin twice can be represented as follows.



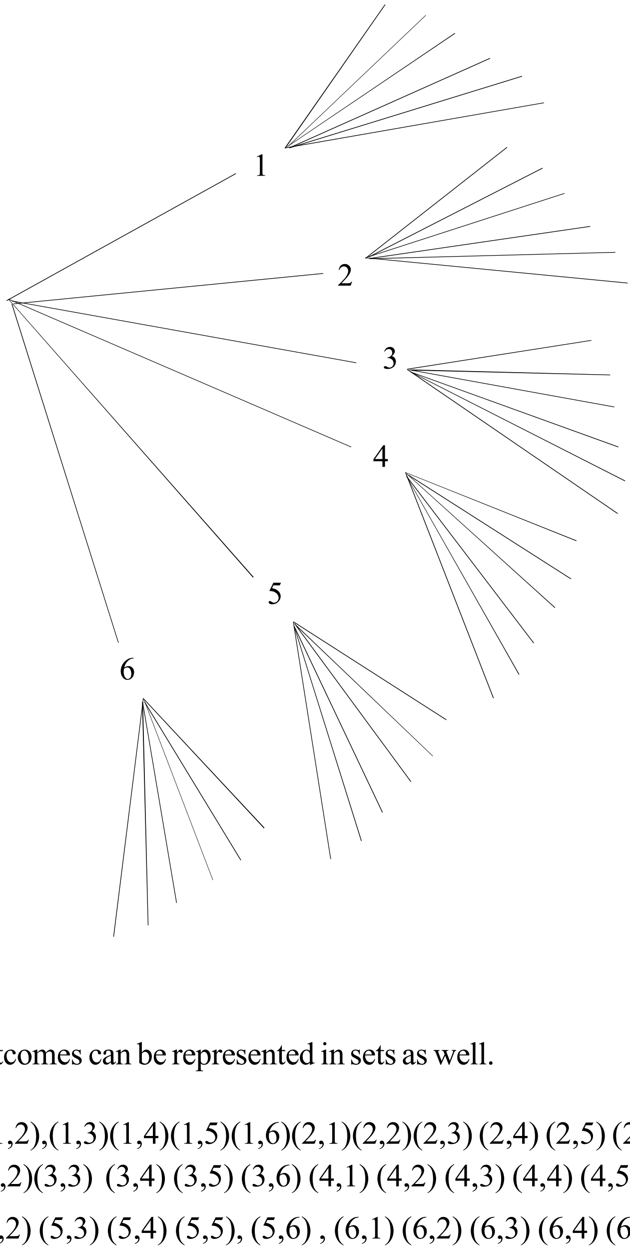
2. All these outcomes can be represented in a set as follows.

S = {(H,H) (H,T) (T, H) (T,T)}

- Representing the sample space in tree diagrams.

  - All the possible outcomes of a random experiment that it is conducted for twice or more than twice can be represented using tree diagrams.

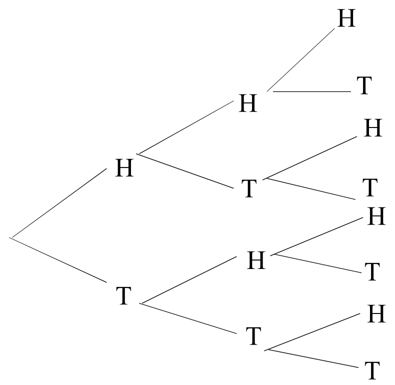
    Ex : Outcomes of the experiment of tossing a die twice are explained in a tree diagram as follows.



  - All of these outcomes can be represented in sets as well.

$$S = \left\{ \begin{array}{l} (1,1)\ (1,2),(1,3)(1,4)(1,5)(1,6)(2,1)(2,2)(2,3)\ (2,4)\ (2,5)\ (2,6) \\ (3,1)(3,2)(3,3)\ \ (3,4)\ (3,5)\ (3,6)\ (4,1)\ (4,2)\ (4,3)\ (4,4)\ (4,5)\ (4,6) \\ 5,1)\ (5,2)\ (5,3)\ (5,4)\ (5,5),\ (5,6)\ ,\ (6,1)\ (6,2)\ (6,3)\ (6,4)\ (6,5)\ (6,6) \end{array} \right\}$$

All the possible outcomes of tossing a coin for three times can be represented as follows.



- All of these outcomes can be represented in sets as well.

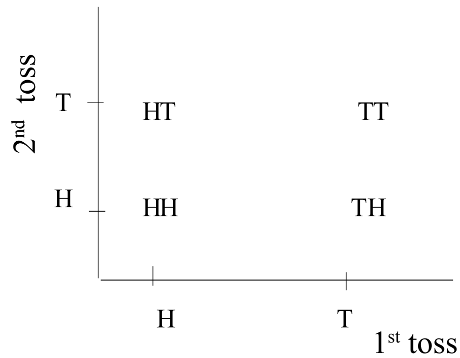
S = { (H, H, H) (H, H,T) (H,T, H) (H,T,T) (T,H,H) (T,H,T) (T,T,H) (T,T,T)}

- Representing the sample space using picture graphs

Once a random experience is conducted only once or only twice, all the possible outcomes can be represented using picture graphs.

Ex : 1. Outcomes of the random experiment of tossing a coin once

2. Outcomes of the random experiment of tossing a coin twice.



- Number of times that a random experiment is conducted is known as 'trials'.
- If the random experiment is conducted only once, the number of trials is 1.
- If the random experiment is conducted twice the number of trials is 2.
- By representing the sample space using a point (doted) graph or a tree diagram the probability based problems can be solved having received a vivid understanding about the random experiment.
- The difference between random experiments and deterministic experiments can be stated as follows.

Random experiments	Deterministic experiments
• Possible outcomes can not be exactly said before the experiment is conducted. • The set of all the possible outcome is known as the sample space. • Based on probability.	• The possible outcome can be exactly said, before the experiment is conducted. • One and only one outcome is received. The experiment is conduced to receive it. • Not based on probability.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.3 : Uses set notations to combine the events

No. of Periods : 04

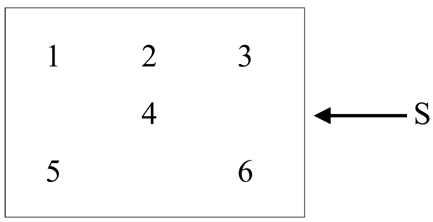
Learning outcomes :

- Interprets the 'events'.
- Separates the regions in the sample space belonged to each event.
- Interprets simple events.
- Interprets composite events.
- Explains that a composite event consists of few simple events.
- Combines the events using **Union** and **Intersection.**
- Expresses the complement of an event using Veen Diagrams and standard symbols.
- Expresses the difference of two events using Venn Diagrams and standard symbols.
- Interprets the **Event space**

Instructions for Lesson Planning :

- Summon a student before the class and make him write the sample space related to the experiment of throwing a balanced die once using two set notation methods.
- If the sample space is S

  S = {1, 2, 3, 4, 5, 6}



- Summon another student before the class and let him mark the following events and name them as follows.
  - Falling 2 as A

    Falling 4 as B

    Falling 6 as C

    Falling an even number as D

- Point out that the first 3 events mentioned above are simple events.
- Emphasise the fact that there is only a single sample point for each of such events.
- Point out that the event of falling an even number is a composite event and it consists of few simple events.
- Discuss with the students that there are more than one sample point contained in a composite event.
- Engage the students in following activity.
- Express each event mentioned below related to the random experiment of throwing a balanced die numbered from 1 to 6 and a balanced coin together using standard symbols and Ven diagrams.

(i) Event that falling head of the coin – A.

(ii) Event that falling 5 of the die – B

(iii) Event that falling head of the coin or the number 5 on the die.

(iv) Event that falling head of the coin and the number 5 on the die.

(v) Event that not falling head of the coin related to this experiment.

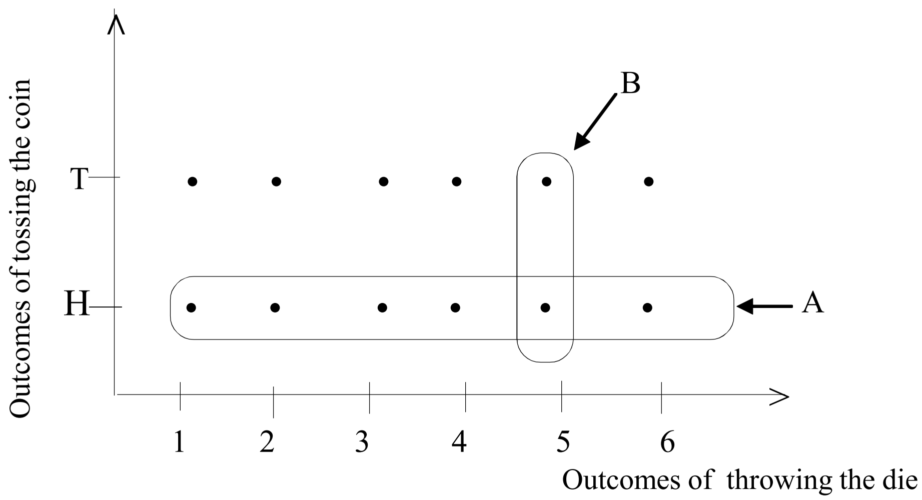
(vi) Event that falling neither the head of the coin nor the number 5 on the die.

(vii) Not occurring both these events simultaneously.

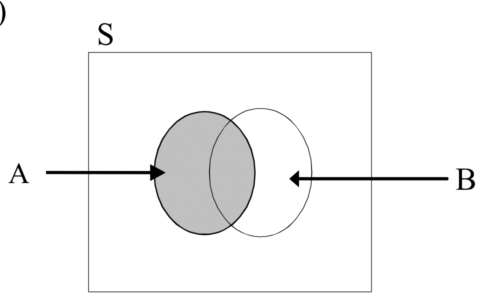
(viii) Event that falling head of the coin, but not the number 5 on the die.

(ix) Event that falling the number 5 on the die, but not the head of the coin.

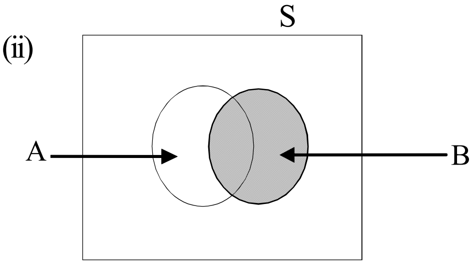
Activity – I : Solutions



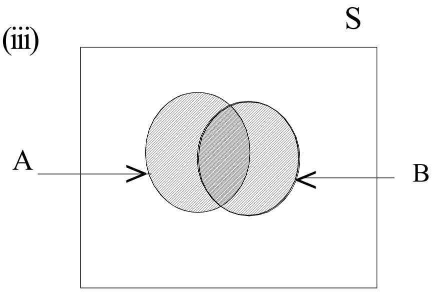
(i)



(ii)

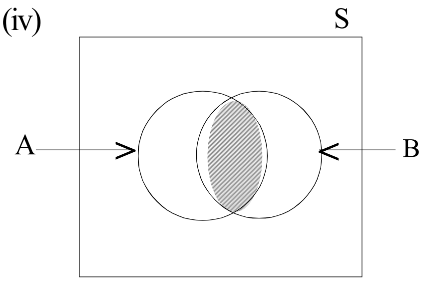


(iii)



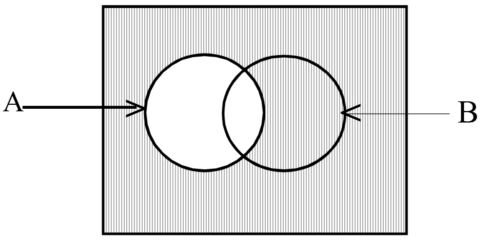
$A \cup B$

(iv)



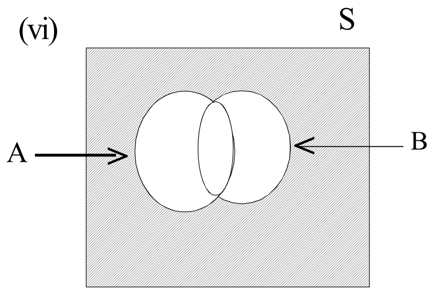
$A \cap B$

(v)



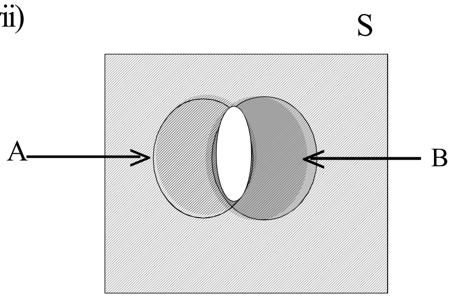
A'

(vi)



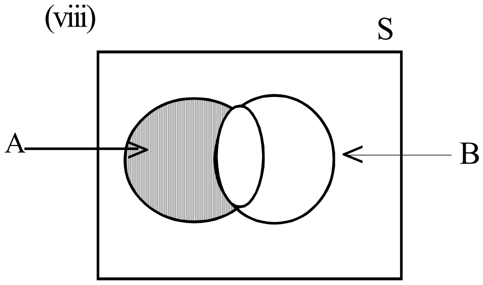
$(A \cup B)'$

(vii)



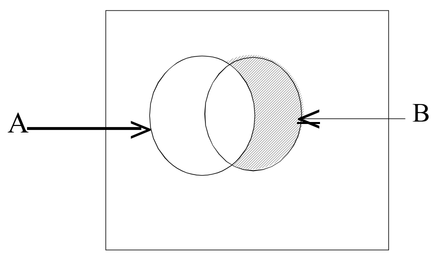
$(A \cap B)^1$

(viii)



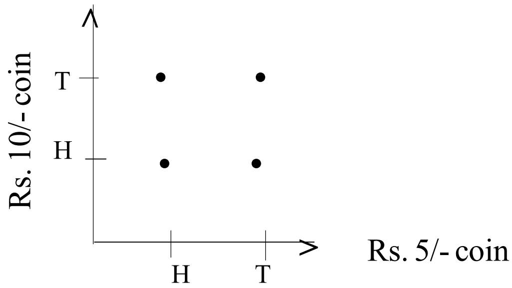
$(A \cap B^1) = A - B$

(ix)

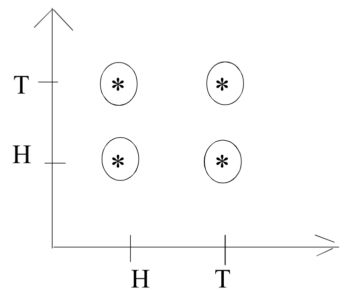


$(A^I \cap B) = B - A$

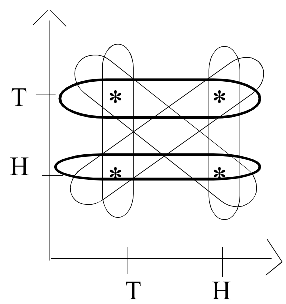
- Hold a discussion with the students to explain the total number of possible events that can be interpreted on the number of sample points available in the sample space.
- Pay attention of the students to the sample space of the random experiment of throwing a Rs 5/= coin and a Rs. 10/= coin together.



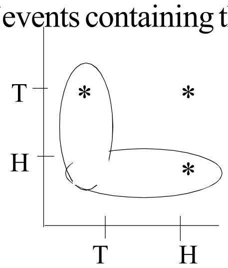
- Note down the number of possible events that can be interpreted under each criterion as follows.
  - No. of events containing only one sample point = 4



- No of events containing two sample points = 6



- No. of events containing three sample points = 4

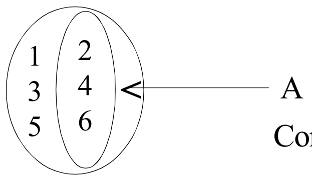


- No of events containing four sample points = 1
- No of events containing no sample points (null events) = 1
- The total number of events that can be derived = 16

A Guideline to explain the subject matters :

- Any subset which is defined on the outcomes of a sample space is interpreted as an event. An event may consist of a single sample point or few sample points.
- Any such single sample point is known as a simple event.
- If there are more than one element in favour of a particular event, that event is known as a composite event, so that if an event defined in a sample space can be further dissociated such an event is known as a composite event.

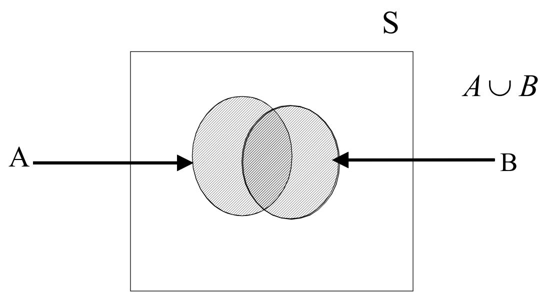
  Ex : The event that receiving an even number, when a die is thrown.



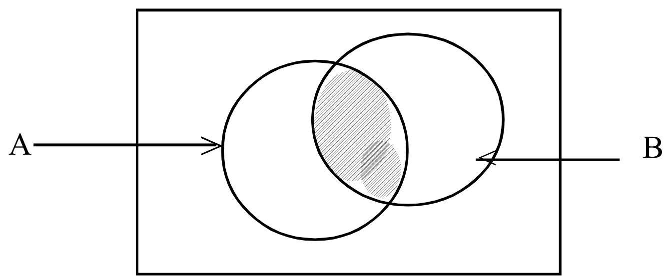
Composite evetnts

$A = \{2,4,6\}$

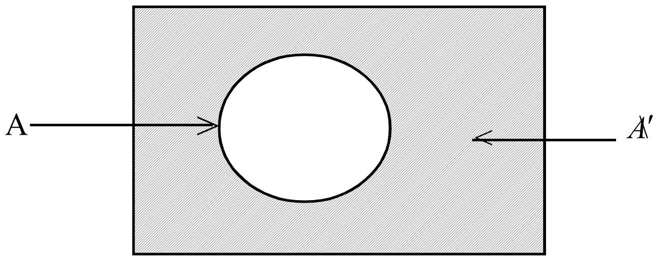
- When A and B are any two events defined in a sample space – S –the region of the sample space in which sample points related to occur at least one of those events, is known as the **union** of those events. That is symbolized as $A \cup B$.



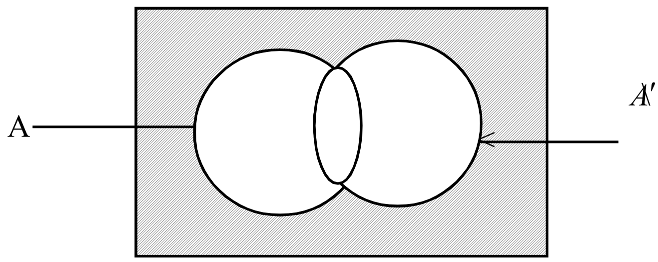
- Further when any number of events as A ,B ,C ……….. have been defined in a sample space, the region in which the sample elements related to either one of those events also can be symbolized as $A \cup B \cup C$ ……
- When A and B are any two events defined in a sample space, the region in which the sample elements related to occur both the elements is known as **intersection** of those two events and symbolized as $A \cap B$



- When **A** is any event defined in the sample space **S**, the region in which the sample elements those are not belonged to A is called the complementary event of A and symbolized as A'.

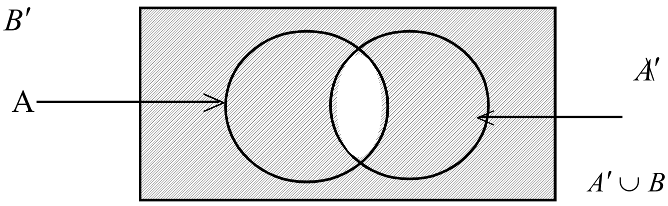


- When A and B are any two events defined in the sample space S the region in which the sample elements those are belonged to neither A nor B is symbolized as $(A \cup B)'$
- This is also symbolized as $A' \cap B'$



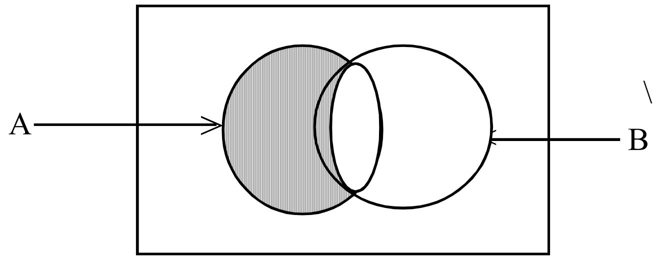
$$(A \cup B)' = A' \cap B'$$

- The region in which the ample elements which are not belonged to the occurance of both the events A and B simultaneously is symbolized as $(A \cap B)'$ and the same idea is given by $A' \cup B'$



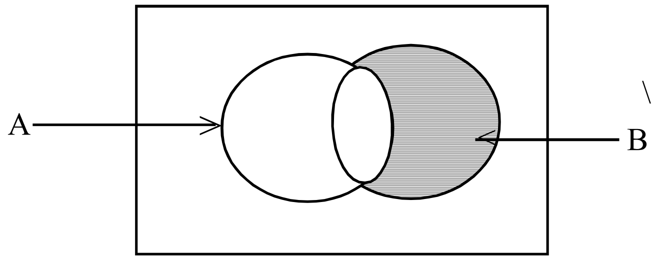
$$A' \cup B' = (A \cap B)'$$

- When A and B are any two events defined in a sample space, the region in which the sample elements belonged to the event A, but not to the event B is denoted as A – B.



A-B

- This can also be symbolised as $A \cap B'$
- In the same way the region in which the sample elements belonged to the event B, but not to the event A is denoted as B – A.



B - A

- This also can be symbolized as $A' \cap B$
- The total number of events that can be defined on all the possible outcomes of a random experiment represented on the sample space is known as the event space.
- Hence the total number of events belonged to the event space that can be defined on sample space containing 'n' number of sample points is given by 2^n.
- The event space cannot be depicted graphically or diagramically like the sample space.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.4 : Arranges and chooses a set of materials.

No. of Periods : 04

Learning outcomes :

- Shows the number of ways through which a set of different materials to each other can be arranged in order.
- Defines the permutations and combinations.
- Writes the formulae to derive permutations and combinations separately.
- Differentiates 'permutations' from 'combinations.
- Solves the problems correctly using relevant formulae.
- Derives the sample space using tree diagrams for random experiments.
- Solves problems related to random experiments using tree diagrams.

Instructions For Lesson Planning :

Produce the following poster before the class.

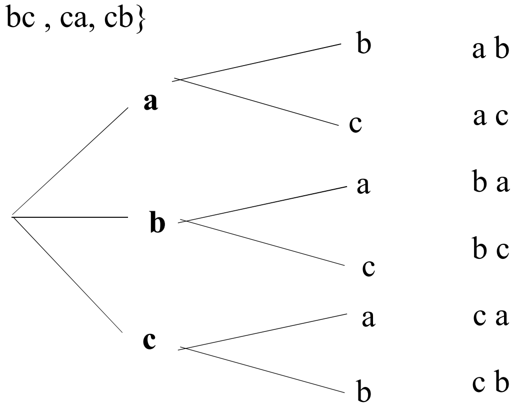
<u>**a**</u> <u>**b**</u> <u>**c**</u> <u>**d**</u> <u>**e**</u>

Let's write a word in five different letters using these five letters.

- Raise the following questions to the students.
  - How many ways are there to choose the first letter of the word?
  - How many ways are there to choose the second letter of the word, after choosing the first letter?
  - How many ways are there to choose the third letter of the word, after choosing the first and second letters?
  - How many ways are there to choose the fourth letter of the word, after choosing the first three letters of the word.
  - How many ways are there to choose the last letter of the word, after choosing the first four letters ?
  - How many words including five letters different to each other can be written ?

- Hold a discussion highlighting the following facts.
  - that the first letter of the word can be chosen in five optional ways.
  - that the second letter of the word can be chosen in four optional ways.
  - that the third letter of the word can be chosen in three optional ways.
  - that the fourth letter of the word can be chosen in two optional ways.
  - that there is only one ways to chose the last letter of the word.
- Explain that there are 120 optional ways to write a word using five different letters from any five different letters given.
  - Point out that the number of words that can be chosen in this manner can be derived as 5 x 4 x 3 x 2 x 1.
  - Explain that 5 x 4 x 3 x 2 x 1 is in factorial notation as 5!
  - Hence explain that 3 ! = 3 x 2 x 1 and 6 ! = 6 x 5 x 4 x 3 x 2 x 1.
  - Engage the students in following activity.
  - Lead the students to write all the possible words using two different letters from the three letters **a, b** and **c.**
- Explain it as follows.

  {ab, ac, ba, bc , ca, cb}



- Lead the students to write all the ways in which two students can be selected from three children Amila, Bindau and Chamila.

  (i) Amila, Bindu

  (ii) Amila, Chamila

  (iii) Bindu, Chamila

- Point out that writing words with two different letters from the letters **a b** and **c** is an 'arrangement and drawing two children from the three children is a 'selection'.

Activity – I

- How many two different digit numbers can be written from the digits 1, 2, 3, 4, 5, 6.
- If one of the numbers written in that way is drawn in random what is the probability that it will be an odd number?

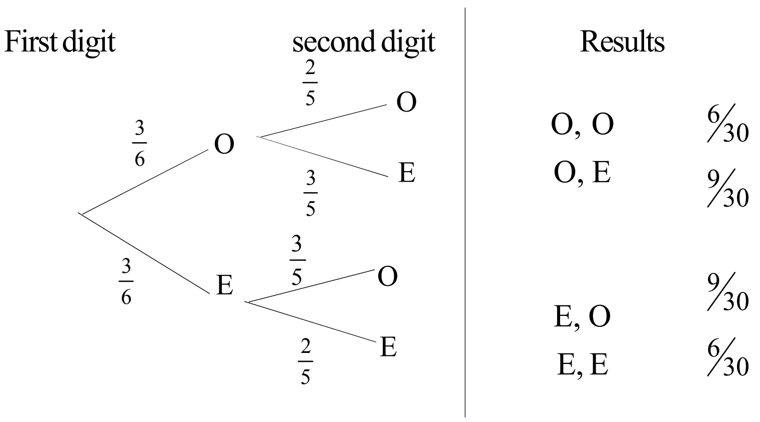
Activity – 1 : Solution.

$n=6, r=2$

$$^{n}p_{r} = \frac{ni}{(n-r)!} \qquad ^{6}P_{2} = \frac{6i}{(6-2)!} = \frac{6\times5\times4i}{4!} = \underline{\underline{30}}$$

30 numbers can be prepared.

- The probability of a random selected number being an odd number can be found using a tree diagram as follows.



- According to the above tree diagram there are two ways to get an odd number as (0, 0) and (E, 0)

$\therefore$ Probability of receiving an odd number $= \frac{6}{30} + \frac{9}{30}$

$$= \frac{15}{30} = \underline{\underline{0.5}}$$

- This problems also can be solved using permutations as follows.
- Number of ways in which both the digits in the number being

Odd numbers $= 3p_1 \times 2p_1 = 6$

- Number of ways in which the first digit of the number being even and the second being odd $= 3p_1 \times 3p_1 = 9$
- Total number of sample points $= 6p_2 = 30$

$\therefore$ The probability of getting an odd number $= \frac{6+9}{30} = \frac{15}{30}$

$= \underline{\underline{0.50}}$

Activity – 2

(i) How many ways are there to select two children from a group consists of four boys as B_1, B_2, B_3, B_4, and there girls as G_1., G_2, G_3.

(ii) Find the probability that both the children being boys.

(iii) Find the probability that both the children being girls.

(iv) Find the probability that at least one of the boys being selected.

Activity – 2 : Solution

(i) Irrespective of the gender the number of ways in which two children out of seven, can be selected , is computed using combinations.

Then $n = 7$ $r = 2$

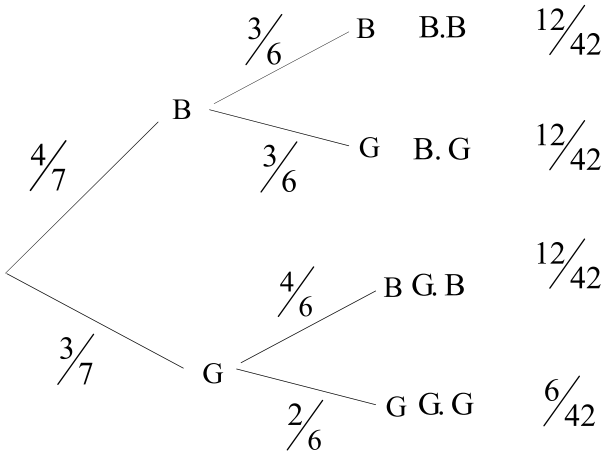
$$\therefore n_{C_r} = \frac{n!}{r!(n-r)i}$$

$$^7C_2 = \frac{7!}{2!(7-2)i}$$

$$= \frac{7 \times 6 \times 5!}{2 \times 1 \times 5!}$$

$$= \underline{\underline{21}}$$

- Lets first use a three diagram to find the probability values.



Probability that both being boys $= \frac{12}{42} = \underline{\underline{\frac{2}{7}}}$

Probability that both being girls $= \frac{6}{42} = \underline{\underline{\frac{1}{7}}}$

Probability that one of the boys being selected $= \frac{12}{42} + \frac{12}{42} + \frac{12}{42}$

$= \frac{36}{42} = \frac{6}{7}$

- That can also be found as follows.
  - Probability that at least one of the boys being selected

$= 1 - p(G.G)$

$= 1 - \frac{6}{42}$

$= \frac{42-6}{42}$

$= \frac{6}{42}$

$= \underline{\underline{\frac{1}{7}}}$

- This can be further be solved using permutations as well.

(i) Probability that both being boys

$$= \frac{4p_1 \times 3p_1}{7p_2}$$

$$= \frac{4 \times 3}{42}$$

$$= \frac{12}{42} = \underline{\underline{\frac{2}{7}}}$$

(ii) Probability that both being girls

$$= \frac{3p_1 \times {}^2P_1}{{}^7p_2}$$

$$= \frac{3 \times 2}{42}$$

$$= \frac{6}{42} = \underline{\underline{\frac{1}{7}}}$$

(iii) Probability that at least one of the boys being selected

$$= \frac{(4p_1 \times 3p_1) + (4p_1 \times 3p_1) + (3p_1 \times 4p_1)}{7p_2}$$

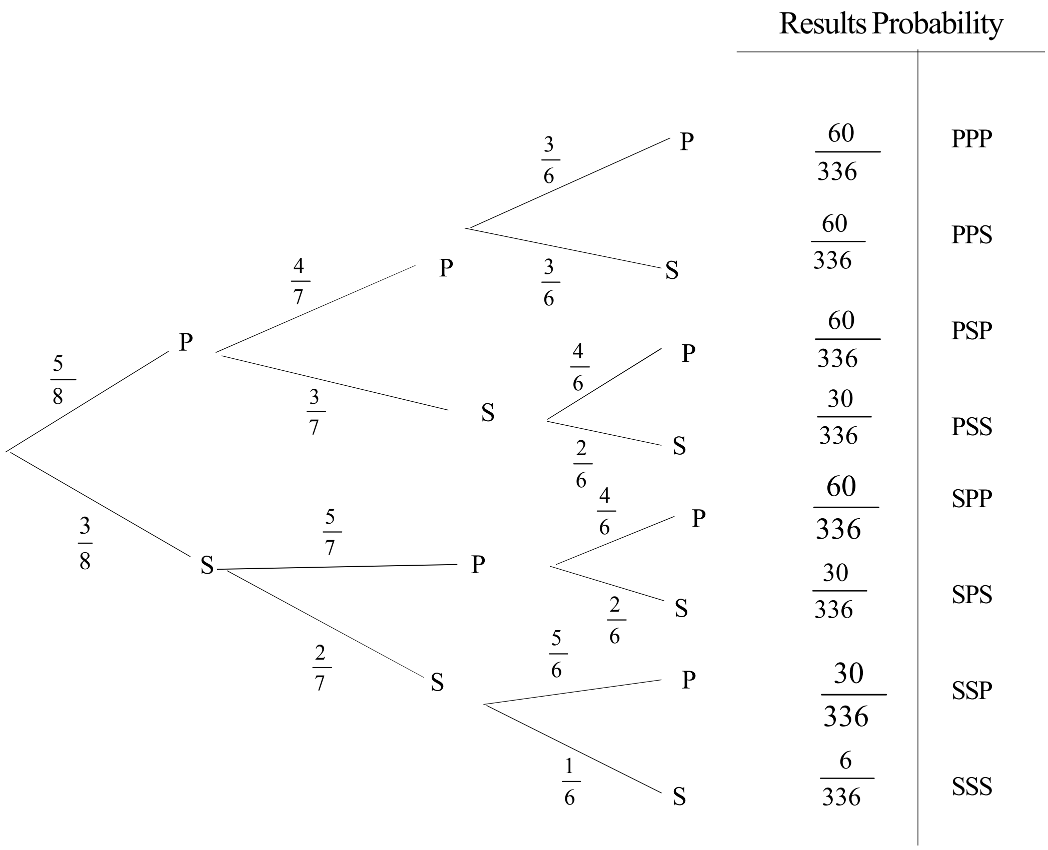
$$= \frac{(4 \times 3) + (4 \times 3) + (3 \times 4)}{42}$$

$$= \frac{36}{42} = \underline{\underline{\frac{6}{7}}}$$

Activity – 3

Five production managers and three sales managers are working in a company. Any three managers of them should have been appointed to the direct board. Find the probability that

(i) getting two production mangers appointed.
(ii) getting two sales managers appointed.
(iii) not getting any sales manager appointed.
(iv) getting at least one of the production managers appointed to the director board of the company.



- Probability of two production mangers being selected can be derived using the tree diagram adding the probability values at the points pps, psp, and spp

$$= \frac{60+60+60}{336} = \frac{180}{336} = \underline{\underline{\frac{15}{28}}}$$

Using combinations $= \frac{{}^5C_2 \times {}^3C_1}{{}^8C_3} = \frac{10 \times 3}{56} = \frac{30}{56} = \underline{\underline{\frac{15}{28}}}$

- Probability of two sales managers being selected can be derived using the tree diagram, adding the probabilities of the points PSS, SPS, and SSP.

$$= \frac{30+30+30}{336} = \frac{90}{336} = \underline{\underline{\frac{15}{56}}}$$

Using combinations $= \frac{{}^5C_1 \times {}^3C_2}{{}^8C_3} = \frac{5 \times 3}{56} = \underline{\underline{\frac{15}{56}}}$

(iii) The probability of a sales manager not being appointed.

Using the tree diagram $= \frac{60}{336} = \frac{5}{28}$

Usisng combinations $= \frac{{}^5C_3 \times {}^3C_0}{{}^8C_3} = \frac{10 \times 1}{56} = \frac{10}{56} = \frac{5}{28}$

(iv) Probability of at least one of the production managers being appointed.

$= 1 - \frac{5}{28} = \frac{28-5}{28} = \frac{23}{28}$

A guideline to explain the subject matters :

- Once set of different materials are ordered in n_1 ways and then in n_2 ways the total number of ordered terms can be computed as n_1 x n_2
- When '**n**' number of different materials are given, the first item can be selected in n number of ways, and then the second item can be selected in (**n** – 1) ways and the third item in (**n** - 2) ways.
- The last item in that set of materials can be selected in one, and only one way and that can be denoted as (n – n) + 1.
- Hence $n(n-1).(n-2).(n-3)............(n-n)+1$

  Can be symbolically stated as n ! (n factorial)
- Here it is generally accepted that 1 ! =1 and 0! = 1
- The number of arrangements that can be made using '**r**' different materials from '**n**' different materials is called a permutation of '**r**'
- Hence the number of permutations that can be derived drawing '**r**' number of different materials from '**n**' number of different materials, is denoted symbolically as ${}^nP_r$ and that number of permutations can be computed as follows.

$${}^nP_r = \frac{n!}{(n-r)!}$$

- A selection made drawing ' r' different materials from 'n' different materials is known as a combination of ' r'.
- Hence the number of combinations that can be served selecting 'r' number of different materials from ' n' numbr of different materials is denoted symbolically as ${}^nC_r$ and the number of combinations can be computed using the following formula.

$${}^nC_r = \frac{n!}{r!(n-r)!}$$

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.5 : Expands a binomial expression

No. of Periods : 04

Learning outcomes :

- Expresses a binomial expression.
- Expands a binomial expression.
- Expands a binomial expression with any power.
- Uses the binomial theorem for expansion of a binomial expression.

Instructions for Lesson Planning :

- Produce the following expression to the class.

$$(a+b)^2$$

- Let the students to expand this expression as they have learnt in grade 10 & 11 classes.
- Hold a discussion highlighting the following facts.
  - An algebraic expression with two terms is called a binomial expression in Mathematics.
  - Such an expression can be produced with any power
  - As (a + b), $(a + b)^2$, $(a + b)^3$ $(a + b)^4$ …….. $(a + b)^n$
- Explain that $(a + b)^2 = (a + b) \times (a + b)$
- Point out that the result derived in the product of these two terms is $a^2 + 2ab + b^2$
- The pattern hidden in this is

$$(a+b)^2 = \begin{bmatrix}\text{Square of the} \\ \text{first term}\end{bmatrix} + 2\begin{bmatrix}\text{Product of} \\ \text{two terms}\end{bmatrix} + \begin{bmatrix}\text{Square of the} \\ \text{second term}\end{bmatrix}$$

- Explain that it is complicated the expansion of binomial expression once its power is greater.
- Hence point out that, it would be easy to get accustomed to a set pattern to derive the expansion of a binomial expression easily.
- Lead the students to derive the expasson of $(a+b)^3$
- Explain that it would be easy to multiply $a^2 + 2ab + b^2$ by (a + b) again.
- Explain that the results is $a^3 + 3a^2b + 3ab^2 + b^3$
- Lead the students to indentify the relationship between the number of terms of the expressions $(a + b)^2$ $(a + b)^3$ and the value of the index of the power.

- Assign the students to derive the expansion of each expression given below according to the identified pattern.

(i)	$(x + 4)^2$	(iii)	$(x + 4)^3$
(ii)	$(a + 5)^2$	(iv)	$(x + 3)^2$

- Write the expressions $(a + b)$, $(a+ b)^2$ and $(a +b)^3$ on the board as follows and point out the pattern of building the co-efficient of each term.

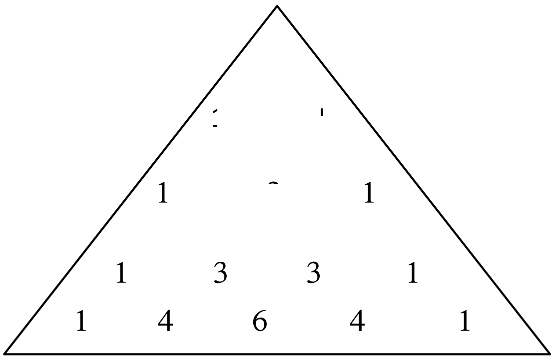
$$(a+b) = \quad 1a + 1b$$

$$\downarrow$$

$$(a+b)^2 = 1a^2 + \ldots b + 1b^2$$

$$(a+b)^3 = 1a \quad 3a^2b \quad ab^2 + 1b^3$$

- Inquire the students about the changing pattern of the index of the unknown terms '**a**' and '**b**'.
- Lead the students to derive the expansion of $(a + b)^4$
- Explain the Pascal's triangle through which the co-efficient of each term of the binomial expansion with any power can be found conveniently.



- Guide the students to derive the expansion of binomial expressions given below.
  - $(2x + y)^4$
  - $(x + 2y)^5$
  - $(x + y)^8$
- Point out that in order to derive the expansion of $(x + y)^8$, the Pascal triangle should be extended up to the 8^{th} step.
- Point out that another method is needed to be followed in order to derive the expansion of any binomial expression with any power.
- Note down only the co-efficients of expansions of the expressions $(a + b)^2$, $(a + b)^3$, $(a +b)^4$ on the board and then find the same values using combinations as follows.

$(a + b)^2$ = 1 2 3

$^2C_0$ $^2C_1$ $^2C_2$

$(a + b)^3$ = 1 3 3 1

$^3C_0$ $^3C_1$ $^3C_2$ $^3C_3$

$(a + b)^4$ = 1 4 6 4 1

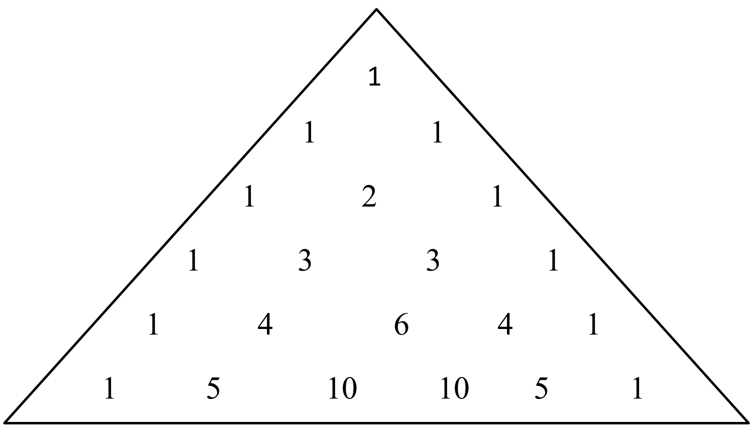
$^4C_0$ $^4C_1$ $^4C_2$ $^4C_3$ $^4C_4$

- Hence point out that the co-efficients of any binomial expansion with any power can be derived using combinations.
- Accordingly build up the expansion of $(a + b)^n$ discussing with the students.

$$(a+b)^n = {}^nC_o a^n b^0 + {}^nc_1 a^{n-1} b^1 + {}^nC_2 a^{n-2} b^2 + {}^nC_3 a^{n-3} b^3 + ... + {}^nC_n a^{(n-n)} b^n$$

A Guideline to explain the subject matters :

- An algebraic expression with two algebraic terms is a binomial expression.
- The product of few binomial expressions with the same opperation is stated as a power of the first expression.
- Ex : if $(a + b)(a + b) = (a + b)^2$ and
  $(a + b)(a + b)(a + b) = (a + b)^3$ and etc.
- The expression derived by multiplying and simplifying such expressions with two algebraic terms is called an expansion of a binomial expression.
- The pattern of co-efficient of a binomial expression is triangular.
- That pattern is known as Pascal triangle.



- The number of terms of the expansion of a binomial expression in the form of $(a + b)^n$ is $(n+1)$
- In the expansion of a binomial expression the index of the first term is gradually decreased from **n** to **0** and the index of the second term is gradually increased from 0 to **n**.
- The co-efficient of the terms of bi-nominal expression can also be derived using combinations.
- The bi-nomial expansion can be generally stated as follows and known as the bi-nomial theorem.

  $(a+b)^n = {}^nC_0 a^n b^0 + {}^nC_1 a^{n-1} b^1 + {}^nC_2 a^{n-2} b^2 + ... + {}^nC_n a^{n-n} b^n$

Assessment and Evaluation :

- Assign the students to write the expansion of the following expressions.

  $(x + 3)^7$ $(2a + b)^4$ $(3x + y)^5$ $(P + 4)^6$

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.6 : Uses the classical approach as an approach to probability

No. of Periods : 02

Learning outcomes :

- Interprets the lassical Approach.
- Points out the situations where the probability can be assessed in accordance with the Classical Approach.
- Computes the probability of an event in accordance with the Classical Approach.
- Points out the weak points of Classical Approach.

Instructions for Lesson Planning :

- Present the following statement to the class.

  " The chance of coin at the first inning of the forthcoming Asian Cricket Competition series will be gained by Sri Lankan Team."

- Hold a discussion highlighting the following facts.
  - That there is an equal chance to fall head or tail, when a balanced coin is tossed.
  - That the outcomes defined in a sample space associated with equal chance to occur are known as equally – liked events.
- Pay attention of the students towards the sample space related to the experiment of throwing a fair die numbered from 1 to 6.
- Point out that there is only one chance to fall number 1 on the upper most surface of the die.
- Point out that the probability of falling the number '1' on the upper most surface of the die is 1/6, since there are 6 possible outcomes of the sample space.
- Ask the number of chances to fall an odd number on the upper most surface of the die and point out that the probability of falling an odd number is 3/6.
- Inquire whether there is any possibility to happen another event in addition to the outcomes defined in the sample space related to the random experiment of throwing a balanced die.
- Highlight through the discussion that there is a possibility of standing the die on its edge or vertex without falling horizontally showing a clear number value.

Activity – 1

If a bulb is drawn out in random from a bag containing four defective bulbs and six non – defective bulbs, find the probability that the bulb being a defective one.

Activity – 1 : Solution

- Let the event that receiving a defective bulb be D.

  N (D) = 4

  N (S) = 10 since the total number of bulbs in the bag is 10'

$$\therefore P(D) = \frac{n(D)}{n(S)}$$

$$= \frac{4}{10} \qquad = \frac{2}{5}$$

Activity – 2

- Find the probability of Upul's birthday falling in week-end.

Activity – 2 : Solution

Number of all possible outcomes

(number of days in the week) = 7

Number of days in weekend

(possible number of days on which Upui's birthday falling) = 2

Hence' if considered the event that his birthday falling in week end as E

$$P(E) = \frac{n(E)}{n(S)}$$

$$= \frac{2}{7}$$

A Guideline to explain the subject matters :

- The basic assumption in Classical Approach is that all possible outcomes of a random experiments are equally – likely.
- An event defined on one or few of those equally likely outcomes is known as an equally likely events.
- The ratio of the possible number of outcomes in favour of a particular event defined on a sample space consists of equally likely outcomes and the total number of possible outcomes in the sample space is known as the probability in the Classical Approach.
- Accordingly when **A** is an event defined on the sample space **S** the probability of occurring the event **A** is mentioned as.

$$P(A) = \frac{n(A)}{n(S)}$$

$P(A)$ = Probability of occurring A

$n(A)$ = Possible number of outcomes in favour of A

$n(S)$ = Total number of outcomes defined in the sample space.

- The Classical Approach is useful to find the probability of an event defined in a sample space containing equally likely outcomes of a random experiment such as tossing a fair coin or throwing a balanced die etc.
- Further this approach is also applicable in practical situations like finding the probability of a randomly drawn unit from a production process containing a certain defective proportions (definite).
- The fact that not considering regarding the probability of an event which is very rare, but possible, like standing a coin on its edge, can be pointed out as a weak point in the Classical Approach.
- This approach also cannot be applied once it is unable to determine whether the outcomes of a particular experiment are equally likely or not.
- Once the total number of possible outcomes of an experiment is unknown this approach cannot be used.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.7 : Uses the Relative Frequency approach as an approach to Probability.

No. of Periods : 04

Learning outcomes :

- Explains 'Relative Frequency' accurately.
- Depicts the relative frequency of occurring the considered event with respect to the each number of experimenting times (number of trials) graphically.
- Interprets the probability of a considered event sing the naps based on the change of relative frequency, once the number of trials is increasing.
- States the situations where the relative frequency approach is applicable in interpretation of ' probability '.

Instructions for Lesson Planning :

- Preset the following clauses to the class.
  - Probability of receiving an odd number at the random experiment of throwing a balanced die numbered from 1 to 6.
  - Probability of receiving a defective unit in a sample of five units drawn in random from a process of production, with the purpose of considering the proportion of defectives.
- Hold a discussion highlighting the following facts. Which are.
  - The events derived from a sample space containing the outcomes of the random experiment of throwing a die, are equally likely.
  - But the events which are derived from the random experiment of checking a sample of products are not equally likely.
  - The classical approach is not applicable to interpret the probability of the events related to the random experiments of which the outcomes are not equally likely.
  - In order to determine the probability of a random selected unit of an item produced in a production process being defective, a large number of samples with the size, five in each should be repeatedly drawn and the defective number contained in each sample should be noted down.
  - Then the relative frequency of a unit being defective can be derived.
  - Divide the students into two groups and engage in following activity.
    1. The event of falling 'head' of the random experiment of throwing a fair coin
    2. The event of appearing an odd number on the upper most surface of a die which is thrown once

- Complete the following table in accordance with the outcomes derived when the experiment is repeated under identical circumstances.

Number of Trials	Number of times that received outcomes in favour of the relevent charateristic	Relative Frequency
5		
10		
15		
30		
50		
100		
150		

Round off the value of relative frequency to the nearest first decimal place.

- Plot the data given in the table in a graph. Represent the number of trials that the experiment was conducted on horizontal (x) axis and the relative frequency on vertical (y) axis. Plot the relative frequencies against each number of trials and join those points.
- Lead the students to explain the movement of relative frequency with respect to the number of trials.
- When the number of trials is increased, the relative frequency of the considered event comes to a particular constant value and explain that, that constant value is considered as the probability of that event.
- Guide the students to provide with instances for practical situations where the Relative Frequency approach is applicable to find the probability.

A Guideline to Explain the Subject Matters :

- The probability of an event defined on the sample space of a random experiment of which the outcomes are not equally likely, can be found using Relative Frequency Approach.
- When the experiment is repeated under identical circumstances the relative frequency of the considered event comes to a particular constant value. That constant value is the probability of that event.
- The probability of an event related to the Relative Frequency Approach is found using the following formula.

$$\text{Probability} = \frac{\text{No. of times received outcomes in favour of the event}}{\text{Total number of trials}}$$

- Relative frequency approach can not be used in following situations.
  - When the experiment ca not be repeated under identical circumstances needed
  - When the statistical data are not available
- Given below are some instances where the probability can be computed on Relative Frequency Approach.
  - To check the quality of a product
  - Inquiring the consumer preference for a particular product
  - To check the popularity of a political party

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.8 : Uses the Subjective Approach as an Aapproach to probability.

No. of Periods : 02

Learning outcomes :

- Explains the Subjective Approach.
- Points out the instances where the probability is expressed using Subjective Approach.
- Points out the weak points in this approach as a method of expressing the probability.

Instructions for Lesson Planning :

- Display the following poster before the class.

> Confidence of winning **X** in forthcoming presidential election is 80%.

- Hold a discussion highlighting the follwing facts.
    - This is an uncertain quantitative statement.
    - It has been predicted based on the previous experience of a particular person.
- Present the following situations to the students and take steps to broaden their awareness regarding the value assignment to probability based on previous experience and personal belief of the people.

    1. An expert specialist doctor selects 10 persons whom are suspected to have symptoms of particular disease observing a gathering of about 200 participants for a medical camp.

       Based on the knowledge and long term experience of the specialist doctor, the people in the selected sample are suspected with a high confidence level to have the symptoms of that particular disease.

    2. A batch of students in a class are preparing to sit for the G.C.E.(A/L) examination in this year. A particular subject teacher states that five students among them will get A passes.

       The relevant teacher has claimed this statement based on his long term experience about the skills of those students and the level of performance they have achieved so far.

    3. A businessman who is aware that a particular machine should be repaired once in every six months, has allocated cash provisions for repairing a machine which was repaired before four months.

Since the particular businessman has understood with his long term experiences that the machine should be repaired once in every six months and allocated money before two months as a pre-preparation without facing financial crisis.

A Guideline for explaining subject Matters :

- Assigning a probability value for a particular uncertain event or an expression based on the knowledge, experience, belief and rational thinking ability in addition to the details available regarding that event or expression is known as subjective approach to probability
- There are many instances where the subjective approach is very important in the business world.
  - Ability to make business decisions quickly
  - Ability to use the experience of the people involved in the business field during a long period
  - Stating the probability of an event on reliability of the person being very important
- But the weak-point in this approach is that it is not applicable in developing other statistical techniques since it is biased from person to person.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.9 : Uses the axiometric approach as an approach to probability.

No. of Periods : 04

Leaning outcomes :

- States the axioms related to probability.
- Writes expressions for probability of various events using the Axioms.
- Interprets the additive law in probability.
- Interprets the mutually exclusive events.
- Expresses the additive law for the union of mutually exclusive events.
- Expresses the additive law for any two events.
- Solves probability based questions using the additive law.
- Uses Venn diagrams and theorems to compute the probability of various events.
- Approaches at rational decisions computing the possibilities of various events.

Instructions for Lesson Planning :

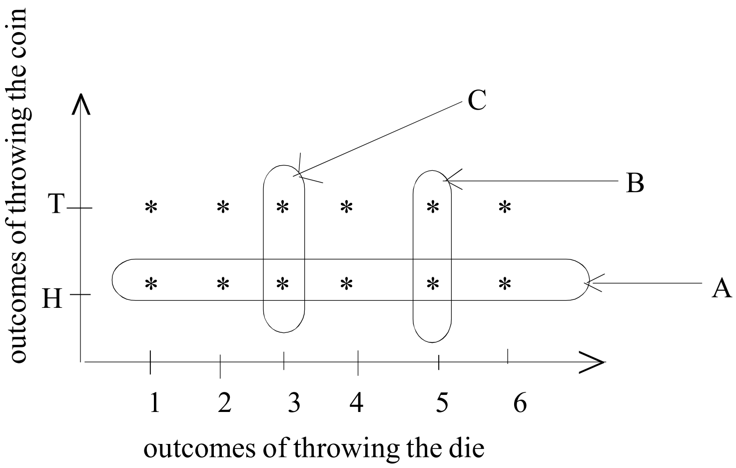
- Pay the attention of the students for following situations.

  "The team of children who made the Vesak lantern are eagerly waiting for illuminating it until the dusk"(darkness falls)

  "Getting something from him is just like to ask for borrowing feather from a tortoise"
- Hold a discussion highlighting the following facts.
  - Falling darkness this evening is certain
  - Illuminating the Vesak lantern is uncertain.
  - Getting feathers from a tortoise is impossible.
- Accordingly explain that the occurrences take place around us can be categorized as certain, uncertain and impossible.
- Point out that the probability of an uncertain event take a value between 0 and 1. Since the probability of a certain event is I (100%) and the probability of an impossible event is 0.
- Explain further that the probability of any event cannot be a minus value.
- Emphasize the fact that it should be examined regarding the generally accepted axioms related to the probability of a particular event.
- Pay the attention of the students to the sample space related to the random experiment of throwing a fair die numbered from 1 to 6 and a balanced coin together.

- Make them mark on that sample space as.
    - The event falling head of the coin as A
    - The event falling 5 on the die as B
    - The event falling 3 on the die as C

- Guide the students to find the follwing probabilities.
    - falling head of the coin
    - falling 5 on the die
    - falling 3 on the die
    - falling 3 or 5 on the die
    - falling 5 on the die and head of the coin
    - falling 5 on the die or head of the coin



1. $P(A) = \frac{n(A)}{n(S)} = \frac{6}{12} = \underline{\underline{\frac{1}{2}}}$

2. $P(B) = \frac{n(B)}{n(S)} = \frac{2}{12} = \underline{\underline{\frac{1}{6}}}$

3. $P(C) = \frac{n(C)}{n(S)} \quad = \frac{2}{12} \quad = \underline{\underline{\frac{1}{6}}}$

4. $P(B \cup C) = P(B) + P(C) \quad = \frac{1}{6} + \frac{1}{6} \quad = \frac{2}{6} = \underline{\underline{\frac{1}{3}}}$

5. $P(A \cap B) = \frac{n(A \cap B)}{n(S)} \quad = \underline{\underline{\frac{1}{12}}}$

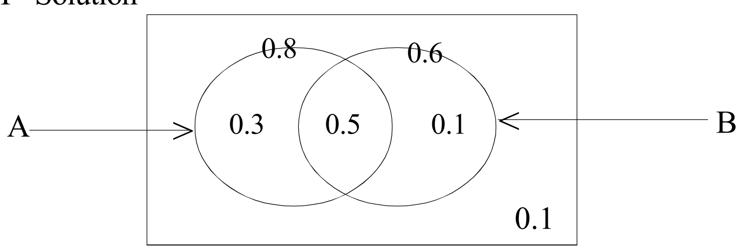
6. $P(A \cup B) = PA + P(B) - P(A \cap B)$

$$= \frac{1}{2} + \frac{1}{6} - \frac{1}{12}$$
$$= \frac{6+2-1}{12} = \underline{\underline{\frac{7}{12}}}$$

Activity I.

- A contractor has submitted two tenders for two construction projects A and B. He is 80% confident of the contract A getting approved and 60% confident of the contract B getting approved. The confidence of both projects getting approved is 50%.
  - (a) Depict these outcomes in a Venn diagram
  - (b) Find the probability of each event mentioned below
    - At least one of the contracts getting approved
    - Neither the two contracts getting approved
    - Not getting both the contracts approved simultaneously
      - Getting only A contract approved
      - Getting only B contract approved
      - Getting only one of these contracts approved

Activity I – Solution
(a)



(1) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 0.8 + 0.6 - 0.5$

$= \underline{\underline{0.9}}$

(2) $P(A \cup B)' = 1 - P(A \cup B)$

$= 1 - 0.9$

$= \underline{\underline{0.1}}$

(3) $P(A \cap B)' = 1 - P(A \cap B)$

$= 1 - 0.5$

$= 0.5$

(4) $P(A - B) = P(A \cap B')$

$= P(A) - P(A \cap B)$

$= 0.8 - 0.5$

$= \underline{\underline{0.3}}$

(5) $P(B - A) = P(A' \cap B)$

$= P(B) - P(A \cap B)$

$= 0.6 - 0.5$

$= \underline{\underline{0.1}}$

(6) $P(A \cap B') \cup (A' \cap B) = 0.3 + 0.1$

$= \underline{\underline{0.4}}$

A Guideline to explain the subject matters :

- Assigning the value for the probability of a particular event should be committed in accordance with generally accepted standards or principles.
- There are axioms that have been created with this regard associated with social phenomenon of our surrounding.
- When X is any event defined in the sample space S, the probability of the event X should be as
- Accordingly the probability of a particular event should not be a minus value.
- Once x_1, x_2 .... x_n are a set of collectively exhaustive events the probability of occurring any event of them is certain $\sum P(x) = 1$
- If A and B are two events free from common elements to both, such events are called Mutually Exclusive events.
- Since those events do not occur simultaneously probability of occurring A and B or $P(A \cap B) = 0$
- When A and B are two mutually exclusive events defined in a sample space S the probability of occurring A or B is $P(A \cup B) = P(A) + P(B)$
- When A, B, C.... are a set of mutually exclusive events defined in a sample space the probability of occurring at least one of those events is $P(A \cup B \cup C \cup -------) = P(A) + P(B) + P(C) + ----$
- Once A and B are any two events defined in a sample space the probability of occurring A or B is $P(A \cup B) = P(A) + P[B] - P(A \cap B)$

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.10 : Uses conditional probability techniques to solve probability based problems.

No. of Periods : 04

Learning outcomes :

- Interprets conditional probability.
- Solves problems related to conditional probability, using the accurate formulae.
- Expresses the multiplicative rule using conditional probability formula.
- Demonstrates the skills of arriving at rational decisions in the business field using the concept of conditional probability.

Instructions for Lesson Planning :

- Present the following statements to the class.
  1. If it rains during Yala season, the farmers will cultivate paddy in large scale.
  2. If an advertisement is telecast, there will be a greater demand for the product.
  3. If he passes in his Advanced Level examination, most probably he will get a job.
  4. If the semi final round is won, the team will get the chance to compete in the final round.
- Engage in a discussion highlighting the following facts.
  - Getting rains during Yala season influences on paddy cultivation.
  - The demand for a product is determined on advertisements.
  - Advanced Level result influences on getting an employment opportunity.
  - A player or a team can compete in the final round on the success of semi final round.
- Accordingly explain the fact that the possibility of occuring most of the events take place in practice are affected by many other events.
- Hence, explain further that the probability of occuring a second event based on the influence of the first event is defined in Statistics as the Conditional Probability.
- Provide with the following table to the students. In a survey conduct using 100 persons the following details have been collected related to their academic level and gender.

	Female	Male	Total
Passed G.C.E. (A/L)	24	20	44
Graduated or higer	26	30	56
Total	50	50	100

- Once an individual is drawn in random from the above mentioned group.

i. Probability of that person being male $= \frac{50}{100}$

ii. Probability of that person being female $= \frac{50}{100}$

iii. Probability of that person having passed in Advanced Level exam $= \frac{44}{100}$

iv. Probability of that person possessing a degree of any higher qualification $= \frac{56}{100}$

v. Probability of that person being a woman bearing a degree or higher qualifications $= \frac{26}{100}$

vi. Given that a woman has been selected, the probability that she is a degree or a higher qualification holder. $= \frac{26}{50}$

- Point out that the denominator of the answer for question number (vi) above has become small, when compared with the answers for the questions from (i) to (v). Explain further that the sample space has become narrow.
- Hence ensure that the probability of the event has increased when compared with another event.
- Point out that once the details are significantly known, increasing the probability value is expected in conditional probability.
- Explain that the probability of the $(vi)^{th}$ question can be derived as follows.

    - Probability of being a woman $= \frac{50}{100}$
    - Probability of being a woman possessing a degree or any higher qualifications $= \frac{26}{100}$

- Given that she is a woman, the probability that she is possessing a degree or any higher qualification

$$= \frac{\frac{26}{100}}{\frac{50}{100}}$$

$$= \frac{26}{50}$$

- According a formula can be derived for solving the problems related to conditional probability as fallows.

$$\text{conditional probability} = \frac{\text{Probability of occurring both events simultaneously}}{\text{Probability of the event given first}}$$

- Lead the students to find the probability of following events using the formula or with reference to the table
  - (i) Given that the random selected person is a male, the probability that he is a degree or any higher qualification holder.
  - (ii) Given that the person selected in random has got through in Advanced Level, he is a male.
  - (iii) Given that the person selected in random is holding a degree or any higher qualification , she is a female.

Answers :

1. $\frac{30}{50}$ 2. $\frac{20}{44}$ 3. $\frac{26}{56}$

- How to get the answer using the formula

1. $\frac{30}{100} \Big/ \frac{50}{100} = \frac{30}{50}$ 2. $\frac{20}{100} \Big/ \frac{44}{100} = \frac{20}{44}$ 3. $\frac{26}{100} \Big/ \frac{56}{100} = \frac{26}{56}$

- With the cross multiplication of the answer derived by using the formula if the probability of person being a male holding a degree or any higher qualification is **x,** it can be derived as follows.

$$\frac{30}{50} = \frac{x}{\frac{50}{100}}$$

$$\frac{50}{100} \times 30 = 50x$$

$$\frac{50}{100} \times 30 \times \frac{1}{50} = x$$

$$\therefore x = \frac{30}{100}$$

2. If the probability that the person being a male got through in Advanced Level is **y**, it can be derived as follows.

$$\frac{20}{44} = \frac{y}{\frac{44}{100}}$$

$$y = \frac{44}{100} \times 20 \times \frac{1}{44}$$

$$44y = \frac{44}{100} \times 20$$

$$\therefore y = \frac{20}{100}$$

3. If the probability of that person being a female holding a degree on any higher qualification is **Z** it can be derived as follows.

$$\frac{26}{50} = \frac{Z}{\frac{50}{100}}$$

$$Z = \frac{50}{100} \times 26 \times \frac{1}{50}$$

$$50Z = \frac{50}{100} \times 26$$

$$\therefore Z = \frac{26}{100}$$

- Accordingly explain that the multiplicative law can be derived using the conditional probability law.
- Inquire the students about the situations where the conditional probability is applicable in business field as far as possible. Few examples are as follows.
  1. Possibility of supplying the undertaken order, based on receiving a batch of raw materials in due time
  2. Possibility of finishing the production in due time caused by a strike being launched
  3. Possibility of changing the productivity of employees based on the training provided with them

A Guideline to explain the subject matters :

- Based on the occurrence of a given event related to a random experiment, the possibility of occurring another event is known as "Conditional Probability".

  Ex : Probability of issuing an undertaken order in the due time, based on receiving an ordered batch of raw materials in due time.

- Following formula can be used to solve the problems related to conditional probability in an easier access.

  - The conditional probability of occurring the event B given that the event A has already occurred is denoted as P(B/A) and

    $$P\left(B/A\right)=\frac{P(A\cap B)}{P(A)} \quad \text{When} \quad P(A)\neq 0$$

  - The probability of occurring the event A given that the event B has already occurred is denoted by $P\left(A/B\right)$ and

    $$P(A/B)=\frac{P(A\cap B)}{P(B)} \quad \text{When } P(B)\neq 0$$

- Applying the cross multiplication of this conditional probability expression the multiplication law of probability can be derived.

$$\frac{P\left(B/A\right)}{1}=\frac{P(A\cap B)}{P(A)}$$

$$\boxed{P(A\cap B)=P(A).P\left(B/A\right)}$$

$$\frac{P\left(A/B\right)}{1}=\frac{P(A\cap B)}{P(B)}$$

$$\boxed{\therefore P(A\cap B)=P(B).P\left(A/B\right)}$$

- Accordingly if A and B are only two events related to a random experiment, the probability of occurring both the events A and B simultaneously can be derived as

$$P(A\cap B)=P(A).P(B/A)$$

or

$$P(A\cap B)=P(B).P(A/B)$$

Assessment and Evaluation :

1. A producer stated that the probability of receiving a batch of raw materials in time is 80% on his long term experience. The probability of the batch of raw materials receiving in time and ability of issuing the ordered batch of items in due time is 60%'. Find the probability of issuing the ordered batch of product in time, given that the batch of raw materials has been received in due time.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.11 : Uses the probability laws of independency in problem solving.

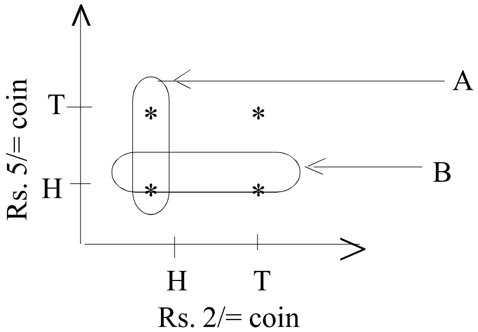
No. of Periods : 04

Learning outcomes :

- Interprets 'Independency'.
- Separates independent events among various types of evens.
- Uses the probability law of independency in making decisions related to business affairs.
- Finds the probability of occurring two independent events simultaneously.
- Finds the probability of occurring several independent events simultaneously.

Instructions for Lesson Planning :

- Pay attention of the students towards the random experiment of tossing a coin twice.
- Ask what the probability of falling head at the first time is.
- Ask what the probability of falling head of that coin at the second time is.
- Point out that falling head of the coin at the first time does not affect on the probability of falling head at the second time.
- Further point out that falling head or not falling head at the first time does not make any influence on the outcomes of its second trial.
- Discuss with the students about some situations similar to followings that can be considered as examples for independent events.
  - Once a utensil contains four red beads and three blue beads taking out two successive beads in random with replacement.
  - In a grinding mill where two machines are used to grind chilies, where one machine is functioning, the other machine is functioning or is not functioning.
  - In a machinery where kerosene is used as fuel, machines are functioning while a power cut is being implemented.
- Pay the attention of students towards the sample space of the random experiment of tossing a Rs. 2/= coin and a Rs. 5/= coin together at the same time.



- Name the event that falling head of the Rs. 2/= coin as A.
- Name the event that falling head of the Rs. 5/= coin as B.
- Point out the fact that head can be expected to fall on both the coins together.
- Point out that the probability of falling head on both the coins is ¼.
- Hence point out that the probability of occurring both the events A and B simultaneously is equal to the product of probability of occurring A and the probability of occurring B.

Since $P(A)=\frac{1}{2}$ හා $P(B)=\frac{1}{2}$ and $P(A\cap B)=\frac{1}{4}$

$$P(A\cap B) = P(A)\bullet P(B)$$
$$= \frac{1}{2}\times\frac{1}{2}$$
$$= \frac{1}{4}$$

- Engage in a discussion to solve the problem that if two events are independent, whether their complementary events are also independent.
- Raise the following questions related to the experiment of taking out two successive toffees from a bag containing three lime flavoured toffees and four pineapple flavoured toffees respectively,
- If the first drawn toffee was lime taste can the second drawn toffee also be lime taste?
- If the first drawn toffee was not lime taste can the second toffee also not being lime taste?
- Accordingly if A and B are two independent events prove that;
  - A' and B' are independent.
  - A and B' are independent.
  - A' and B are independent.

  as follows :

  If A and B are independent.

  $P(A\cap B)=P(A)\bullet P(B)$

  Accordingly if A' and B' are independent it should be proved that $P(A'\cap B')=P(A')\bullet P(B')$

$$
\begin{aligned}
P(A' \cap B') &= P(A \cup B)' \\
&= 1 - P(A \cup B) \\
&= 1 - \{P(A) + P(B) - P(A \cap B)\} \\
&= 1 - \{P(A) + P(B) - P(A) \bullet P(B)\} \\
&= 1 - P(A) - P(B) + P(A) \bullet P(B) \\
&= [1 - P(A)] - P(B)\ [1 - P(A)] \\
&= [1 - P(A)]\ [1 - P(B)]
\end{aligned}
$$

$$\underline{\underline{P(A' \cap B') = P(A') \bullet P(B')}}$$

- If A and $\mathbf{B'}$ are independent it should be proved that $P(A \cap B') = P(A) \bullet P(B')$

$$
\begin{aligned}
P(A \cap B') &= P(A) - P(A \cap B) \\
&= P(A) - P(A) \bullet P(B) \\
&= P(A)(1 - P(B)
\end{aligned}
$$

$$\underline{\underline{P(A \cap B') = P(A) \bullet P(B')}}$$

Activity – 1 :

- Separate the independent events from each pair if events given below;
  1. Taking out two toffees in random as one after the other without replacement from a bag containing three lime taste toffees.
  2. Father being a doctor and his son being a teacher.
  3. Jayamini getting selected to the university and Jayalath geeting selected to the university.
  4. A Sri Lankan student being selected to the medical faculty and law faculty in the same academic year.
  5. A quality control officer rejecting the first checked batch and the second checked batch on a particular day.
  6. Rejecting a batch of raw materials which is needed to be purchased for fulfilling an undertaken order and the order being completed.

Activity – 1 : Solution

1. Dependent
2. Independent
3. Independent
4. Dependent
5. Independent
6. dependent

Activity – 2

- when A and B are any two events defined in a sample space S such that $P(A) = 0.6$ $P(B) = 0.5$ and $P(A \cup B) = 0.8$

(i) find $P(A \cap B')$

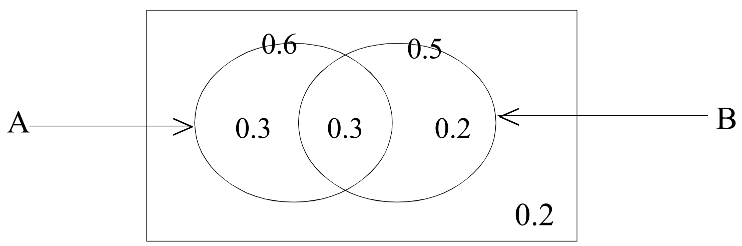
(ii) Examine whether the events A and B are independent.

(iii) Examine whether the events A' and B' are independent.

(iv) Examine whether the events A and B' are independent.

(v) Examine whether the events A' and B are independent.

Activity – 2 : Solution



(i) $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$0.8 = 0.6 + 0.5 - P(A \cap B)$$

$$\therefore P(A \cap B) = 0.6 + 0.5 - 0.8$$

$$= \underline{\underline{0.3}}$$

(ii) $$P(A) \cdot P(B) = 0.6 \times 0.5$$

$$= \underline{\underline{0.30}}$$

$\therefore$ A and B are independent events

(iii) $P(A' \cap B') = P(A \cup B)'$

$= 1 - 0.8$

$= \underline{\underline{0.2}}$

Hence

$P(A').P(B') = 0.4 \times 0.5$

$= \underline{\underline{0.20}}$

$\therefore$ A' and B' are independent events.

(iv) $P(A \cap B') = P(A) - P(A \cap B)$

$= 0.6 - 0.3$

$= \underline{\underline{0.3}}$

Hence

$P(A).P(B') = 0.6 \times 0.5$

$= \underline{\underline{0.30}}$

$\therefore$ A and B' are independent events.

(v) $P(A' \cap B) = P(B) - P(A \cap B)$

$= 0.5 - 0.3$

$= \underline{\underline{0.20}}$

Hence $P(A') \cdot P(B) = 0.4 \times 0.5$

$= \underline{\underline{0.2}}$

A' and B are independent events.

A Guideline to explain the subject Matters :

- The events in a manner such that the occurrence or not occurrence of one event does not affect on the occurrence or not occurrence of another event are known as independent event.
- Once A and B are two independent events the probability of occurring A when B is given can be stated as $P(A/B) = P(A)$ and the probability of occurring B when A is given can be stated as $P(B/A) = P(B)$ (since A and B are independent events early expected probability to occur the event A cannot be changed at all through the occurrence of the event B).
- By substituting the above result the multiplicative law of probability.

$$P(A \cap B) = P(A) \cdot P(B/A)$$

$$\therefore P(A \cap B) = P(A) \cdot P(B) \qquad \text{Since } P(B/A) = P(B)$$

- If three events A, B and C are independent $\therefore P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ and this can be substituted for the independency of any number of events.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.12 : Partitions the sample space accurately for usage of Total Probability Theorem and Beye's Theorem.

No. of Periods : 08

Learning outcomes:

- Explain mutually exclusive and collectively exhaustive events.
- Explains the events that provide the basis for total probability theorem using the sample space.
- Interprets the Total Probability Theorem.
- Provide with examples for the situations where the total probability theorem is applicable.
- Solves probability based problems using total probability theorem.
- Interprets the Beye's Theorem.
- Provides with examples for the situations where the Beye's theorem is applicable.
- Solves problems using the Beye's theorem.
- Solves probability based problems using tree diagrams.

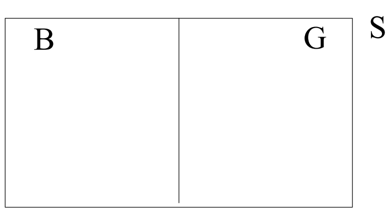
Instructions for Lesson Planning :

- Present the following sets to the class and get them represented in a Venn diagram calling upon a student before the class.

  S = { Children in a class of a mixed school}

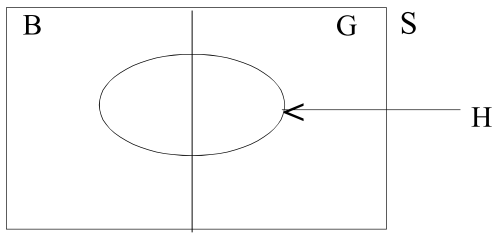
  B = {Boys in that class}

  G = {Girls in that class}
- Point out that this is the best diagram, through various diagrams that may be represented by the student.

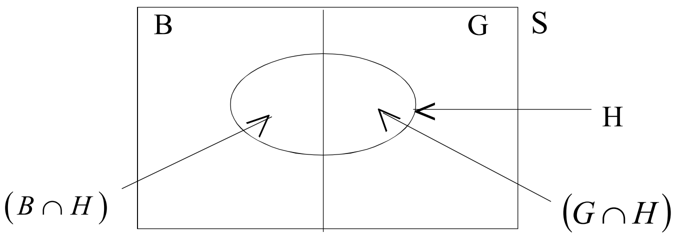


- Point out that, once the students are separated as boys and girls all the students in the class are included in that partition (separation).
- Point out that the probability of a random selected child from the class being a boy or a girl should be 1.

- Point out the events that a random selected child being a boy and being a girl are mutually exclusive.
- Point out that those events are known as collectively exhaustive events, since the entire sample space is covered by the union of those events.
- Call upon another student before the class and get the set of students who are more than five feet tall in that class as H represented on the same Venn diagram. (Assume that there are boys and girls who are less than five feet tall as well as boys and girls greater than five feet tall in that class).
- Point out that then the Venn diagram should be as follows and further that H is an event which is common to the events B and G mutually exclusive and collectively exhaustive events.



- State that the probability of the event H can be interpreted using the total probability theorem as follows.



- Recalling the subject matters learnt relative to conditional probability and the multiplicative law of probability ensure that

  $P(B \cap H) = P(B).P(H/B)$ and

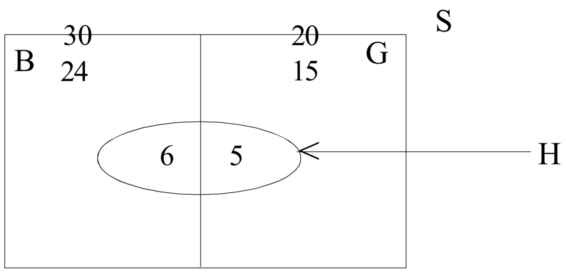
  $P(G \cap H) = P(G).P(H/G)$
- Hence point out that

  $P(H) = P(B).P(H/B) + P(G).P(H/G)$

  and explain the total probability theorem in that manner.

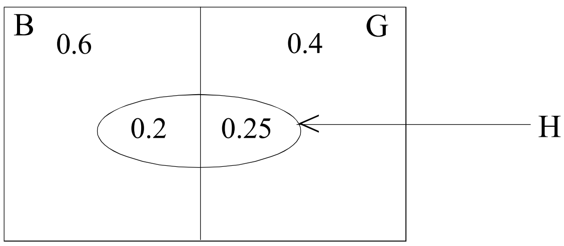
Activity – 1 :

If there are 30 boys and 20 girls in a class including 6 boys and 5 girls who are exceeding 5 feet in height, lead the students to represent these details in a Venn diagram.



- Guide the students to find the probability of a random selected student from this class
  (i) Being a boy
  (ii) Being a girl
  (iii) Being taller than five feet given that a boy has been selected.
  (iv) Being taller than five feet given that a girl has been selected.

- Lead them to re-build the Venn diagram with those probability values.



- Accordingly, lead the students to apply the Total Probability Theorem to find the probability that a random selected student being taller than five feet.

$$
\begin{aligned}
P(H) &= P(B \cap H) + P(G \cap H) \\
&= P(B) \cdot P(H/B) + P(G) \cdot P(H/G) \\
&= (0.6 \times 0.2) + (0.4 \times 0.25) \\
&= 0.120 + 0.100 \\
&= \underline{\underline{0.220}}
\end{aligned}
$$

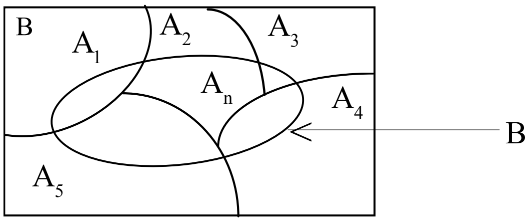
- Point out the fact that the formula of confidential probability can be applied in Beye's theorem to find the probability of a random selected student being a boy given that the student is taller than five feet.

$$
\begin{aligned}
P(B/H) &= \frac{P(B \cap H)}{P(H)} \\
&= \frac{P(B) \cdot P(H/B)}{P(H)} \\
&= \frac{0.6 \times 0.2}{0.22} = \underline{\underline{\frac{6}{11}}}
\end{aligned}
$$

- Discuss with the students regarding few practical instances where the total probability thermo and Beyes theorem are applied.

A Guideline to explain the subject matters :

- If the entire sample space is covered by the union of a set of mutually exclusive events those events are called mutually exclusive and collectively exhaustive events.
- If ,A_1, A_2,A_3 ....A_n are a set of mutually exclusive and collectively exhaustive events defined in a sample space, and if B is another event occurred depend on all those events the probability of occurring, B can be stated as follows.



$$P(B) = P(A_1) \bullet P\left(\frac{B}{A_1}\right) + P\left(A_2\right) \bullet P\left(\frac{B}{A_2}\right) + \ldots\ldots + \bullet P(A_n) \bullet P\left(\frac{B}{A_n}\right)$$

This statement can be summarized using sigma notation as.

$$P(B) = \sum_{i=1}^{n} P(A_i) \bullet P(B/A_i)$$

This relation is interpreted as the Total Probability Theorem.

Given that an event B which is common to A_1, A_2 A_3 .... A_n set of mutually exclusive and collectively exhaustive events has already occurred the probability of occurring an event indicated by A_i in that sample space can be computed as follows.

$$P(A_i/B) = \frac{P(A_i) \bullet P(B/A_i)}{P(B)}$$

- This relation is known as the Beye's Theorem.

Assessment and Evaluation :

(I)

1. A producer purchases a particular type of raw material item only from three registered suppliers named as A, B and C. 50%, 35% and 15% of the total requirement of this raw material item is supplied by A, B and C respectively. 1%' 3%' and 4% of the supply of each person respectively are known to be defective according to the long term experiences. If one unit of this raw materials is chosen in random from the store.

   (i) Find the probability that it will be a defective unit.

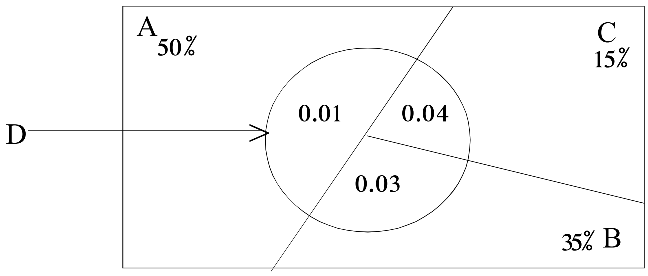
   (ii) Given that it is a defective unit, find the probability that it has been supplied by A.

   (iii) Given that it is a defective unit, find the probability that it has been supplied by B.

   (iv) Given that it is a defective unit, find the probability that it has been supplied by C.

   (v) Given that it is a defective unit, prove that it has been supplied by either A or B or C is certain.

Solution :



(i)

$$P(D) = P(A \cap D) + P(B \cap D) + P(C \cap D)$$

$$= P(A).P(D/A) + P(B).P(D/B) + P(C).P(D/C)$$
$$= (0.5 . 0.01) + (0.35 \times 0.03) + (0.15 \times 0.04)$$
$$= 0.0050 + 0.0105 + 0.0060$$
$$= \underline{\underline{0.0215}}$$

(ii)

$$P(A/D) = \frac{P(A \cap D)}{P(D)} = \frac{0.5 \times 0.01}{0.0215} = \frac{0.005}{0.0215} = \underline{\underline{\frac{10}{43}}}$$

(iii) $P(B/D) = \dfrac{P(B \cap D)}{P(D)} = \dfrac{0.35 \times 0.03}{0.0215} = \dfrac{0.0105}{0.0215} = \underline{\underline{\dfrac{21}{43}}}$

(iv) $P(C/D) = \dfrac{P(C \cap D)}{P(D)} = \dfrac{0.15 \times 0.04}{0.0215} = \dfrac{0.006}{0.0215} = \underline{\underline{\dfrac{12}{43}}}$

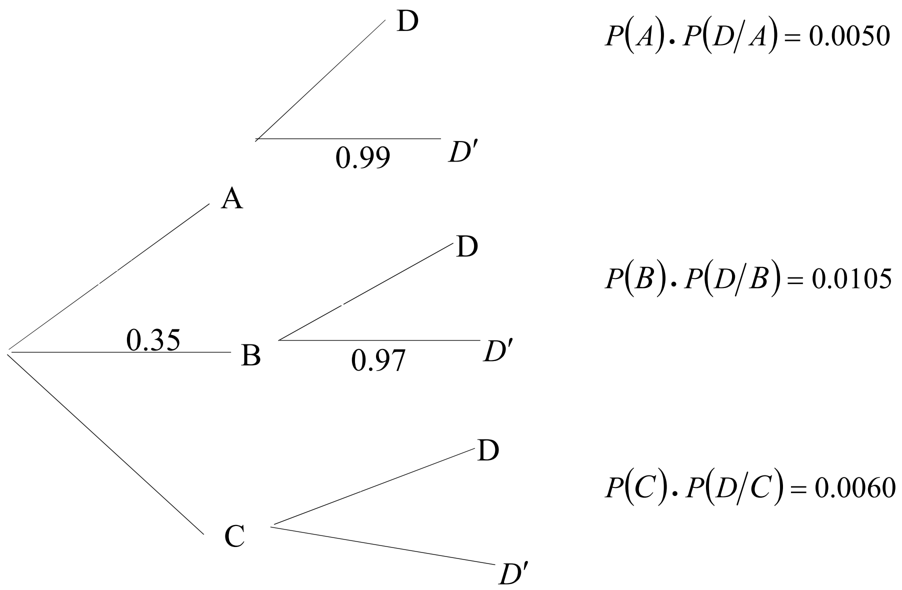
(v) $P(A \cup B \cup C/D) = \dfrac{P(A \cap D) + P(B \cap D) + P(C \cap D)}{P(D)}$

$$= \frac{0.0050 + 0.0105 + 0.0060}{0.0215}$$

$$= \frac{0.0215}{0.0215}$$

$$= \underline{\underline{1}}$$

- According, if a random selected unit of raw material is given that defective the fact that it should have been sent by either A or B or C is certain.
- Lead the students to solve this problem using a tree diagram as well.



$P(D) = P(A) \cdot P(D/A) + P(B) \cdot P(D/B) + P(C) \cdot P(D/C) = \underline{\underline{0.0215}}$

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.13 : Constructs the probability distributions interpreting random variables.

No. of Periods : 08

Learning outcomes:

- Introduces Random Variables.
- Categories Random Variables.
- Gives instances for discrete random variables.
- Gives instances for continuous random variables.
- Introduces the probability distribution.
- Explains the conditions that should be satisfied by a probability distribution.
- Explains the Expected Value and Variance of the probability distribution of a random variable.
- Constructs the probability distribution of a discrete random variable related to a random experiment.
- Computes the expected value and variance of a discrete probability distribution.
- Makes business decisions using the probability distribution of a discrete random variable.

Instructions for Lesson Planning :

Activity – 1

- Present the following situations to the students.
    1. When a coin is thrown up 6 times, the number of times that head is receivable.
    2. Number of months in the year.
    3. Number of times that a school student – player can win in a competition which is conducted at zonal, provincial and national level.
    4. Number of days that a student may come to school in a week.
    5. Number of books available in the school library.
    6. Possible mass of a loaf of bread sold in a school canteen during a week.
    7. Possible life time of bulbs used in the class rooms.
- Discuss the possible outcomes receiveble at each situation mentioned above.
- Lead the students to categorize the above situations as follows.

• Situations with certainly predictable outcomes	• Situations without certainly predictable outcomes
• • •	• • •

- Engage in a discussion highlighting the following facts.
  - Once a coin is thrown 6 times the number of times falling head cannot be exactly said, but the number of times can be 0 or 1 or 2 or 3 or 4 or 5 or 6.
  - The number of months in the year is certain .
  - Number of times that a student-player can win in a competition held zonal, provincial and National level cannot be exactly said, but the number of times he may win should be either 0 or 1 or 2 or 3.
  - Number of days that a student coming to school in a prospective week cannot be pre-determined exactly but can be 0 or 1 or 2 or 3 or 4 or 5.
  - Number of books available in the school library at present can be exactly said.
  - The mass of loaves of bread sold in a canteen during a week cannot be exactly mentioned. Those values can be dispersed in a range.
  - The life time of the bulbs in classrooms cannot be exactly said. It may fall in a range.
- Explain that the variables determined on random experiments of which the receivable outcomes cannot be exactly foresaid are known as Random Variables.

Activity – 2

- Provide with the following situations to the students. Guide them to name the random variable related to each situation and mention whether each event can be assigned a definite value or the values moving in a range.

Situation	variable	Can be assigned a definite value/ values moving?
1. Numner of defectives that may be produced in a machine per day. 2. Possible mass of the products that can be manufactured in an assembling line per day. 3. Possible prices that a product may be sold during an year. 4. Number of telephone calls receivable by a firm in a day. 5. The collar size of the shirts that may be sold in a day.		

Engage in a discussion highlighting the following facts.

- The random variable related to the first situation is 'the number of defective units'.
- Since the number of defective units can be stated as 0, 1, 2 ….., it is a discrete variable.
- The random variable related to the 2nd situation is ' the mass of the products. That variable does not rake a definite value. It may fall in a range of values, so that it is a continuous variable.
- The random variable related to the 3rd situation is 'the price of goods'. Though the 'price is a continuous variable, since it takes a definite value as Rs, 100, Rs, 101, Rs, 101.50' etc. it can be considered as a discrete variable as well.
- The random variable related to the 4th situation is 'the number of telephone calls. Since it can take a definite value as 0, 1, 2, 3 .., it's a discrete random variable.
- The random variable related to the 5th situation is the 'collar size of a shirt', Since it can be a definite value as 15, 15 ½, 16 ……. It's also a discrete random variable.

Activity -3

Involve the students in activity related to the following situation of introduction of probability distributions.

If a fair coin is thrown three times, considering the number of heads receivable as X.

1. Ask each student separately what the possible values that x can take are.
2. Out of those vales uttered by each student write down the correct values on the board. If something is wrong, point out the fault in it.
3. Guide the students to compute the probability of each accurate value mentioned on the board using multiplicative law or a tree diagram.

4. Hence get this table completed by the students.

No of heads receivable x_i	Probability $P(x_i)$

- Hold a discussion highlighting the following facts using the table completed above.
  - Explain that there are two elements in the above distribution as the possible values taken by a random variable (x) and the probability of x as p(x).
  - Explain further that no any value in p(x) column is less than 0 and the sum of the values in that column is equal to 1.
  - Point out that the distribution of a random variable which satisfies these two conditions is known as a probability distribution.

Activity – 4

- Provide with the following description to the students to explain the expected value and variable of a probability distribution.
  - It has been revealed in a survey that 5% of the boxes of matches said to be having 50 matches, are containing 48 matches, 10% of the boxes containing per 49 matches 60% of the boxes containing per 50 matches and in the rest 25% of the boxes per 51 matches.
- Involve the students in Activity giving following instructions.
  1. Considering the number of matches that may contain in a box X = x, write down the probability distribution of X.
  2. Multiply the values in X column by the values in p(x) column and derive the sum of those products as $\sum x.p(x)$
- Point out that the value received at the step 2 above is the number of matches that can be generally expected in a box of matches and that is the expected value of this probability distribution. Guide the students further to find the number of matches needed if 10 000 boxes of matches are to be made.
- Explain that the variance of a probability distribution also can be computed as well as its expected value and using the relevant formula for that task will be more convenient.

Activity – 4 (Solution)

Let's consider the number of matches in a box as 'x'

x	$P(x)$	$x \cdot P(x)$	$x^2 \cdot P(x)$
48	0.05	2.40	115.20
49	0.10	4.90	240.10
50	0.60	30.0	1500
51	0.25	12.75	650.25
		50.05	2 505.55

$$E(x) = \sum x \cdot P(x)$$
$$= \underline{\underline{50.05}}$$

Expected numbr of matches in a box = 50.05

No. of matches required for 10 000 boxes = 50.05 x 10 000 = <u>500 500</u>

$$\text{var}(x) = \sum x^2 \cdot P(x) - [\mathrm{E}(x)]^2$$
$$= 2505.55 - (50.05)^2$$
$$= \underline{\underline{0.55}}$$

A Guideline to explain the subject matters :

- A random variable is a function of rational values (numerical) determined by a random experiment. In other words it is a function of rational values (numerical) defined in a sample space.
- A random variable is denoted by a capital English letter like X ,Y … and the values taken by such a random variable is denoted by a simple English letter like x, y, ….
- Examples for random variables.
  1. An audit officer examines 5 books of Accounts for their accuracy. If the number of non- defective books is denoted by X, the values taken by x can be
     X = 0, 1, 2, 3, 4, 5
  2. When the number of customers coming to a supermarket for buying goods during an hour is denoted by X
     X = 0, 1, 2, 3, 4, 5 …………..
- Random variables can be categorized as follows in accordance with the nature of the values that can be assigned on.
  - **Discrete Random Variables**

    The random variables defined in sample spaces consist of finite points or countable infinite points are known as discrete random variables. Those values can be stated in definite values as zero or minus/plus whole numbers or fractions.

- Examples for discrete random variables are as follows.
    - number of students who may come to the class tomorrow.
    - Number of insured motor vehicles met with accidents,
    - Number of errorness words that may appear in a text of 1000 words in a book.
    - Number of customers that may come to a bank during the lunch hour.
- **Continuous Random Variables.**

The random variables defined in sample spaces consist of infinite number of sample points with respect to the points in a linear interval are known as continuous random variables.

- Those values are moving in a range of values containing minus/ plus values or zero. Few examples for continuous random variables are as follows.
    1. Volume of liquid containable in a bottle of cool drinks mentioned as 300 ml.
    2. The possible temperature that may be experienced in city 'A' during the next hour.
    3. Life time of a battery manufactured in a factory.
- A table or a function or an equation consists of all the possible values taken by a discrete or a continuous random variable associated with corresponding probabilities is known as a probability distribution. A probability distribution of a discrete random variable means that a table consists of the values taken by a discrete random variable associated with the corresponding probability values. Tree diagrams sets or permutations / combinations can be used for calculating the relevant probability values.

X	$P(X)$
x_1	$P(x_1)$
x_2	$P(x_2)$
.	.
.	.
.	.
x_n	$P(x_n)$

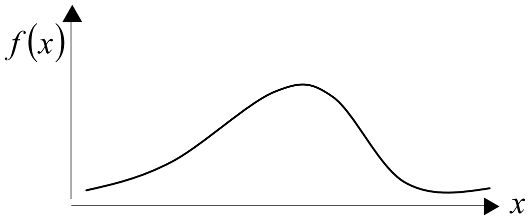
The probability distribution can be constructed in this manner and that distribution is called the probability mass function or the probability function.

Since $p(x_i)$ indicates probability values those values should fall between o and +1.

- The sum of these probability values for all the possible values taken by x should be 1.
- Accordingly following two conditions should be satisfied for representing a probability distribution of a discrete random variable.

(i) $P(xi) \geq 0$ (ii) $\sum_{i=1}^{n} P(x_i) = 1$

- The probability distribution of a continuous random variable is;
  - Suppose that the random variable X takes the values on the rational number line and then that the below mentioned frequency curve is received.



  - If the probability density in a manner such that the area under this curve is equal to 1, can be denoted by the mathematical functions $f(x)$ that is known as the probability distribution of the random variable X or as the probability density function. The total area under $f(x)$ should be equal to1 and no any probability value can be minus.
  - Accordingly the following conditions should be satisfied related to a continues random variable.

(i) $f(x) \geq 0$ (ii) $f(x)dx = 1$

- The expected value of a Discrete Random Varaible.

Once a random experiment is repeated continuously for a long period the average value receivable for the random variable related to that experiment is known as the expected value of that random variable.

The expected value of a particular random variable is its Mean.

If the values taken by a discrete random variable X are as $x_1, x_2, ...x_n$ and the corresponding probability values are as $P(x_1), P(x_2)...............P(x_n)$ when the expected value or the mean is denoted by

$$E(x)$$

$$\text{E}X = \sum_{i=1}^{n} x_i \bullet P(x_i)$$

The chance (possibility) of the expected value of a random variable to be dispersed is computed as its variance.

Once the values of a discrete random variable are $x_1, x_2, ...x_n$ and the corresponding probabilities $P(x_1), P(x_2)...............P(x_n)$ the variance of the probability distribution of a discrete random variables.

Var (X) can be computed as follows.

$$Var(X) = x_1^2 p(x_1) + x_2^2 p(x_2) + ... + x_n{}^n p(x_n) - [E(x)]^2$$

$$Var(X) = \sum_{i-1}^{n} X_i{}^2 p(x_i) - (E(x))^2$$

-

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.14 : Studies the Standard Probability Models

No. of Periods : 04

Learning outcomes:

- Explains the requirement of the standard probability Models.
- Lists the probability models related to a discrete random variable.
- Names the probability models related to a continuous random variable.

Instructions for Lesson Planning :

- Present the following problem to the students to explain the need of probability models.
- Guide the students to compute the expected value and the variance of the number of heads receivable at tossing 50 fair coins together.
- Hold a discussion highlighting the following facts, after the students have responded that it is complicated having tried their best.
- Point out that, the probability models are needed to solve such complicated probability based problems.
- State that there are discrete probability models to solve complicated discrete probabilistic problems, where as continues probability models to solve complicated continuous probabilistic problems.
- Pay the attention of students for solving problems mentioned below.
  1. Finding the probability that there will be at least three defectives in a batch of 10 000 units produced in a machine.
  2. Finding the probability that two lorries coming to a filling station in an hour.
  3. Finding the probability that the volume of a 300 ml bottle of soft drinks being less that 295 ml.
- State that there are probability models to solve such complicated probabilistic problems conveniently.

Guidelines to explain the subject matters :

- A probability based formula, table or a graph which is used to solve complicated probability based problems very simply are called probability models.
- Probability models are needed to study significant practical problems.
- Through identification of the characteristics of various probability models those models can be fit when necessary.

- The binominal probability function and poisson probability function can be used as discrete probability models.
- Since a discrete probability model is a formula that represents the probability related to each value taken by the discrete random variable, it is called a probability mass function.
- The normal probability function (Normal Distribution) can be used as a continuous probability model.
- Since a continuous probability model is a formula that represents probabilities related to all the values of the random variable fallen in a linear range, that is known as a probability density function.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.15 : Solves probability related problems using the Binomial Model

No. of Periods : 10

Learning outcomes:

- Describes the Bernoil Trial.
- Interpretes the binomial distribution stating the relevant conditions.
- Provides with examples for binomial random variables using the binomial theorem.
- Interprets the probability mass function of the binominal distribution.
- Solves probability related problems using the function of binomial distribution.
- Solves problems conveniently using the binomial distribution tables,
- Interprets and computes the Mean and Variance of a binomial distribution.
- Describes the properties of a binomial distribution.

Instructions for Lesson Planning :

Display the following note before the class.

<table>
<tr><td>Experiment</td><td colspan="2">Outcomes</td></tr>
<tr><td>Throwing a coin once</td><td></td><td></td></tr>
<tr><td rowspan="2">Throwing a coin twice</td><td rowspan="2">2nd trial</td><td>1st trial</td></tr>
<tr><td></td></tr>
</table>

- Call upon a student before the class and let him complete the blanks in this table.
- Hold a discussion highlighting the following facts.
  - The experiment of tossing a coin once consists of two outcomes.
  - Conducting an experiment with two outcomes once in this manner is a bernoulli trial.
- There are four possible outcomes in the experiment of tossing a coin twice.
- Once an experiment with two outcomes is conducted for more than once, that is known as a binomial experiment.
- Once the coin is thrown twice that experiment consists of two binomial trials.
  (A trial means that one time that an experiment is conducted)

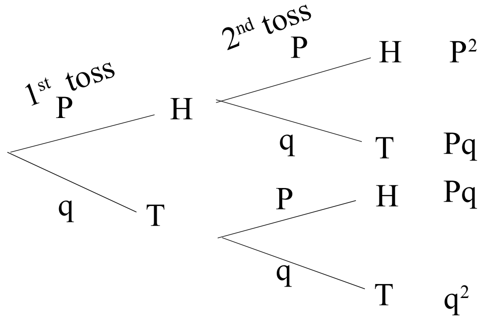
- If falling head of the coin is expected the event that falling head is known as the 'success' (S) and the event that not falling head as the failure (F)
- Point out the fact that the probability of falling head of the coin thrown once (getting the success once) is p = ½.
- Point out that the probability of falling head, when it is tossed for the second time is also ½.
- Explain that falling or not falling head of the coin when it is tossed for the first time does not make any influence on falling or not falling head of the coin when it is tossed for the second time.
- Lead the students to build the probability distribution of X, if x is the random variable that denotes the number of heads receivable at the random experiment of tossing a coin twice.
- Give instructions to use P for the probability of getting success (probability of falling head of the coin) and q for the probability of getting failure (probability of not falling head of the coin)
- Encourage the students to complete the following table.

X	P(X)	P(X) in terms of p and q
0	$\frac{1}{4}$	q^2
1	$\frac{2}{4}$	2Pq
2	$\frac{1}{4}$	P^2

- Point out that the last column of this table can be completed easily using the following tree diagram.

  Event of falling head - H

  Event of falling tail - T



- Hence point out that probability of the variable X taking each value is $P(x) > O$ and also the sum of these probabilities is equal to I.

  1. $p(x) \geq 0$ 2. $\sum p(x) = 1$

  Since $\frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1$

  Ensure that. $\sum_{i=0}^{2} P(x) = q^2 + 2Pq + P^2 = 1$

  Point out that this is the expansion of $(p+q)^2$

- Accordingly give instructions to build the probability distribution of X considering the number of heads receivable at the random experiment of tossing a coin 3 times.

X	P(X)	P(X) in terms of P and q
0	$\frac{1}{8}$	q^3
1	$\frac{3}{8}$	$3pq^2$
2	$\frac{3}{8}$	$3p^2q$
3	$\frac{1}{8}$	p^3

- Point out using this distribution also that the probability of the random variable X taking each value

  $P(x) \geq 0$ and

  $$\sum_{i=0}^{3} P(x) = q^3 + 3pq^2 + 3p^2q + p^3 = 1$$

  Further explain that this is the expansion of $((q+p)^3$

- Point out that the co-efficient of each term of this expansion that 1, 3, 3, 1 can be derived as ${}^3C_o$ ${}^3C_1$ ${}^3C_2$ ${}^3C_3$ Using the technique of combinations.

- Hence highlight through the discussion that the expansion of $((q+p)^3$ can be expressed as follows.

$$\sum_{i=0}^{3} P(x) = {}^3C_0P^0q^3 + {}^3C_1P^1q^{3-1} + {}^3C_2P^2q^{3-2} + {}^3C_3P^3q^{3-3}$$

- Based on these observations, guide the students to build an expression in terms of p, q, n and X_1 for the probability of the binomial random variable X_n which defines a binomial distribution, taking the value denoted by X.

Activity – I

- If a student answered 5 questions in an MCQ paper completely in random find the probability that.
  (i) No any answered being correct.
  (ii) Only one answer being correct
  (iii) Only two answer being correct
  (iv) At most two answers being correct
  (v) At least two answers being correct

Solution

No. 0f multiple choice questions n = 5

Probability that one question being correctly answered P = 0.2

Q = 0.8

if the number of correct answers is the random variable X

$$X \sim Bi(n, p)$$

$$X \sim Bi\ (5, 0.2)$$

$$P(x=x) = {}^nC_x \ p^x q^{n-x}$$

(i) $P(x=o) = {}^5C_0 \times 0.2^0 \times 0.8 \times^{5-0}$

$= 1 \times 1 \times 0.3277$

$= \underline{\underline{0.3277}}$

(ii) $P(x=1) = {}^5C_1 \times 0.2^1 \times 0.8 \times^{5-1}$

$= 5 \times 0.2 \times 0.8^4$

$= 5 \times 0.2 \times 0.4096$

$= \underline{\underline{0.4096}}$

(iii) $P\left(x=2\right) = {}^5C_2 \times 0.2^2 \times 0.8^3$

$= 10 \times 0.04 \times 0.512$

$= \underline{\underline{0.2048}}$

(iv) $P(x \leq 2) = p(x=0) + p(x=1) + p(x=2)$

$= 0.3277 + 0.4096 + 0.2048$

$= \underline{\underline{0.9421}}$

(v) $P(x \geq 2) = 1 - P(x = o) + P(x = 1)$

$= 1 - 0.3277 + 0.4096$

$= 1 - 0.7373$

$= \underline{\underline{0.2627}}$

- Point out the way of deriving the above answers easily with reference to the p=0.2 column in the binomial distribution table of n=5.
- Inquire from the students regarding the possible number of correct answers receivable out of 50 questions in a paper associated with five optional choices for each.
- Here in this context, since the probability of one question being correctly answered is 0.2, point out that 20% of 50 answers can be expected to be correct. Hence, further point out that 10 correct answers can be expected to be correct.
- In this sense, since the concept of 'mean' is known as 'expected value' in probability point out that it is also the mean of the binomial distribution.

A Guideline to explain the subject matters :

- A single trial of random experiment consists if only two outcomes is known as a Bernoulli trial.
- If a random experiment consists of only two outcomes is conducted for more than once, one turn of such an experiment is known as a binomial trial.
- In order that a random variable X which is defined in connection with a binomial experiment, the following four conditions, should be satisfied by X to have a bi normal distribution.

- the experiment should consist of an exact number of trials (n).
- each trial should consist of only two outcomes as success (S) and failure (F).
- the probability of getting success at each trial should be equal.
- each trial should be independent from all the other trials.

- The probability that the random variable X – which satisfies those four conditions, taking a value of X is defined as
- $P(X = x) = {}^{n}C_{x} P^{x} . q^{n-x}$ $\quad x = 0,1,2..........n$

  This is called the probability mass function of a binomial distribution,
- The probability based problems related to random experiments covered by a pattern of a binomial distribution, can be solved using this formula.
- Probability values of X under various probability of getting success (p) for few situations such as $n = 5$" $n = 10$" $n = 15$ have been tabulated and calculations performed with reference to those tables will be convenient.
- The mean and the variable of a binomial distribution are considered as $\mu = np$ and $\sigma^2 = npq$ respectively.
- Since q always takes a value less than 1 (q = 1-p) the variance of a binomial distribution is always less than its mean.
- A binomial distribution whose p=05 is symmetrical where as a binomial distribution with p < 0.5 is right skewed and a binomial distribution with p>0.5 is left skewed.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.16 : Solves Probability based problems using the Poisson Model

No. of Periods **:** 12

Learning outcomes:

- Writes the assumptions on which the Poisson random variable has been built.
- Interprets the 'Poisson distribution'.
- Provides with instances for Poisson random variable.
- Writes the probability mass function of the Poisson distribution.
- Explains the mean and variance of a Poisson random variable.
- Describes the characteristics of a Poisson distribution.
- Solves problems using the probability mass function and the tables.
- Lists the properties of a Poisson distribution.
- States the conditions needed to approximate a binominal distribution using a Poisson distribution.
- Applies the Poisson distribution to solve the problems related to the binomial distribution when the relevant conditions are satisfied.

Instructions for Lesson Planning :

- Present the following statements to the class.
  - "At least three students in every week should be sent home or hospital bearing the transport cost."
  - An acident is reported from anywhere daily.
  - A blood test has revealed a deficiency in growing white blood cells (WBC) contained in blood of a patient coming to a particular clinic.
  - " At least 20 telephone calls are received by the school office in a day."

Hold a discussion highlighting the following facts.

- The variables derived from all the above mentioned statements are discrete random variables.
- The variables derived from 1st and 2nd statements are disperssed through out time and space.
- The variable derived from the 3rd statement is dispersed throughout the space.
- The variable derived from the 4rd statement is dispersed throughout the time.

- Point out that the probability based problems of random experiments related to such random variables disperssed throughout the time or space have been modeled using the Poisson distribution.
- Explain the assumptions on which the Poisson distribution has been modeled using the above mentioned statements.
- Point out in accordance with the 1st statement that the number of students sent home/ hospital for being sick in one week does not affect on that number in other week.
- Point out in accordance with the 3rd example mentioned above that the number of WBC in a blood sample of one patient does not affect on the possible number of WBC of another patient.
- In the number of calls received the office in a day is 20. Point out that it is assumed that the number of calls received in two days is 40 and in half a day is 10.
- Point out that the possibility of receiving more than one call during a short time interval like five seconds is ignorable.
- Further point out that the possibility of occurring more than one event at a particular place at a time is also ignorable.
- Enquiring the students regarding some other variables disperssed on time and / or space justifies the above assumptions.
- Explain that the Poisson distribution can be used to solve the problems of discrete random variables dispersed through time or /and space. Write the probability mass function of a Poisson distribution on the board as follows.
- $P(x = x) = \frac{e^{-2}\lambda^{x}}{x_i}$ $\quad x = 0,1,2........11$

- Explain that λ is the mean of the Poisson distribution and the value of e = 2.7183 which is a constant.

Activity – I

- Engage the students in following Activity 1 :
  - Suppose the past reports have revealed that the average number of child births take place in a hospital on a day is **three** find the probability of that.

    (i) No any child birth

    (ii) Exactly one child birth

    (iii) At least one child births

    (iv) At most two child births will take place in this hospital tomorrow

Solution :

(i) $\lambda = 3 \qquad x = 0$

$$P(x=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$P(x=0) = \frac{2.7183^{-3} \times 3^{0}}{0!}$$

$$= \frac{1}{2.7183^{3}}$$

$$= \underline{\underline{0.0498}}$$

(ii) $$P(x=1) = \frac{2.7183^{-3} \times 3^{1}}{1!}$$

$$= \frac{3}{2.7183^{3}}$$

$$= \underline{\underline{0.1494}}$$

(iii) $$P(x \geq 1) = 1 - P(x=0)$$

$$= 1 - 0.0498$$

$$= 0.9502$$

(iv) $$P(x \leq 2) = P(x=0) + P(x=1) + P(x=2)$$

$$= \frac{2.7183^{-3} \times 3^{0}}{0!} + \frac{2.7183^{-3} \times 3}{1!} + \frac{2.7183^{-3} \times 3^{2}}{2!}$$

$$= 0.0498 + 0.1494 + 0.2240$$

$$= \underline{\underline{0.4232}}$$

N; B : Lead the students to derive these values also using the poisson distribution table with $\lambda = 3$

- Engage the students in following activity to make them aware that the probability of occurring an invents in a time interval/in an area of space is proportionate to the length of time interval / area of the space .

Activity – 2

- The average number of customers coming to a bank in an hour has been experienced as 24. It takes 10 minutes time for one customer to get his work done in the bank. Find the probability that.
    - (i) no any customer coming
    - (ii) two customers coming
    - (iii) at least three customers coming
    - (iv) at most four customers coming to the bank.

Solution – Activity – 2

Average number of customers coming in an hour $\lambda = 24$

$\therefore$ Average number of customers coming in 10 minutes time $\lambda = \frac{24}{6} = 4$

Since $P(x = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

(i) $P(x = 0) = \frac{2.7183^{-4} \times 4^{0}}{0!}$

$= \underline{\underline{0.0183}}$

(ii) $P(x = 2) = \frac{2.7183^{-4} \times 4^{2}}{2!}$

$= \underline{\underline{0.1465}}$

(iii) $P(x \geq 3) = 1 - \{P(x = 0) + P(x = 1) + P(x = 2)\}$

$= 1 - (0.0183 + 0.0733 + 0.1465)$

$= \underline{\underline{0.7619}}$

(iv) $P(x \leq 4) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)$

$= 0.0183 + 0.0733 + 0.1465 + 0.1954 + 0.1954$

$= \underline{\underline{0.6289}}$

- Explain that Mean and Variance of a poisson distribution are equal and therefore mean is λ and variance is also denoted by λ
- Explain the following characteristics of a poisson distribution.
  - Poisson distribution is applied to solve the probability based problems related to discrete random variables.
  - The mean of a poisson distribution is equal to the variance of it.
  - When the probability of getting success is closer to zero and number of trials is large, a binomial distribution can be approximated using a poisson distribution.
- Explain that following condtions should be satisfied by a binomial distribution to be approximated using a poisson distribution.
  - Number of trial should be large $(n \geq 50)$
  - Probability of getting success should be closer to zero $(p \leq 0.1)$
  - Mean shoud less than 5, $\mu = np < 5$

- Explain the students the fact that the problems related to binomial distribution can be easily solved using the poisson distribution, considering the mean of the binomial distribution np for mean of the poisson distribution $(p < 0.1)$
- Give instructions to list the characteristics of the poisson distribution using the facts learnt fo far.

A Guideline to explain the subject matters :

- The theoritical probability model that has been developed to solve the probability based problems related to discrete random variables which are distributed throughout time and/or space is known as the 'Poisson distribution'.
- $P(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$ $\quad\quad x = 0,1,2,..........n$

  is known as the probability mass function of a poisson distribution.
- λ is the mean of the poisson distribution and **e** is a constant $(e = 2.7183....)$
- Mean and variance of a poisson distribution are equal.
- The above probability mass function of the poisson distribution has been developed based on following assumptions.

  (i) The events occurred in a particular time interval / in an area of space are independent from the events occurred in another time interval/ in an area of space which are not intersected on one another.

  (ii) Probability of occurring the events in a particular time interval or in an area of space is proportionate to the length of that time interval or the area of the space.

  (iii) The probability of occurring two or more events in a too tiny interval or in a very small space is ignorably small.
- The charactersistics of a poisson distribution can be listed as follows.
  - Poisson distribution is a probability model related to discrete random variables.
  - The 'mean' of a poisson distribution is equal to the variance of it.
  - When the probability of receiving 'success' is too small $p \leq 0.1$. And the number of trials is large $(n \geq 50)$ of a binomial distribution, the poisson distribution with the mean $(\lambda = np)$ can be used as an approximation.
  - The above two requirements can be combined as $(np < 5)$

Assessment and Evalaution :

- Engage the student in following activity.

  Suppose that 1% of the total production of a particular item is known to be defective with the long term experience. Once a random sample of 400 units from this batch of production is drawn and checked, find the probability that.

(i) no any defective units

(ii) two defective units

(iii) at most two defective units

(iv) at least two defective units being contained in that sample

Solution :

- Lead the students to solve this problem using the function of binomial distribution pointing out that the probability of receiving success is 0.01 and the number of trials $n = 400$

  When considered the number of defective units as X.

$$x \sim \text{bin}\ (n, p)$$

$$x \sim \text{bin}\left(400,\ \ 0.01\right)$$

$$P(X = \text{x}) = {}^{n}C_{x}.p^{x}q^{n-x} \qquad\qquad X\ =\ 0, 1, 2\ldots\ldots$$

$$P(X = 0) = \overset{400}{\underset{0}{C}} \times 0.01^{0} \times 0.99^{400-0}$$

- Let the students to experience themselves that the calculation in this way is too complicated. Point out that this probability value can be approximated using the poisson distribution with $\lambda = np$ When n is large and p is closer to zero.

$$\lambda = np$$

$$\lambda = 400 \times 0.01$$

$$\lambda = \underline{\underline{4}}$$

Since $P \to 0$ and $\to \propto$

$$x \approx Po\ \ (\lambda = np)$$

$$x \approx Po\ \ (\lambda = 4)$$

- Explain the students to use the poisson distribution table with $\lambda = 4$

(i) $p(X = 0) = \underline{\underline{0.0183}}$

(ii) $P(x) = 2 = \underline{\underline{0.1465}}$

(iii) $P(x \leq 2) = p(x=0) + p(x=1) + p(x=2)$

$= 0.0183 + 0.0733 + 0.1465$

$= \underline{\underline{0.2381}}$

(iv) $P(x \geq 2) = 1 - P(x=0) + P(x=1)$

$= 1 - (0.0183 + 0.0733)$

$= 1 - 0.0916$

$= \underline{\underline{0.9084}}$

Competency 5.0 : Demonstrates the preparedness to race business risk
Competency Level 5.17 : Studies the normal distribution as a probablity model.
No. of Periods : 12

Learning outcomes:

- Interprets the normal distribution stating its characteristics.
- Interprets the probability density function of the normal distribution.
- States the parameters of the normal distribution.
- Gives instances for variables that distribute normally.

Instructions for Lesson Planning :

Present the following marks distribution to the class.

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of students	02	04	10	20	28	20	10	04	02

- Guide the students to construct the histogram and the frequency polygon on it.
- Highlight that the distribution of data is symmetrical in accordance with the diagram.
- Lead the students to compute Mean, Median, Mode and Standard Deviation of this distribution.
- Assure that Mean = Median = Mode = 55 of this distribution with the students' responses.
- Pointing out that the standard deviation of this distribution is 16 and lead them to mark the standard deviations of 16 from mean to either side.
- When the standard deviations +16 and -16 are marked on either side of the mean 55 point out that its range is $39 \leq \bar{x} \leq 71$. Further point out that the area in that range between the normal curve and the horizontal axis is very closer to 68.27%.
- Point out that when two standard deviations from either side of the mean 55 are plotted as + 32 and – 32, its range is $23 \leq \bar{x} \leq 87$. The area in that range between the normal curve and the horizontal axis is very closer to 95. 45% of the total area under the normal curve.
- Point out the fact that when three standard deviations from the mean 55 are plotted as +48 and -48, its range is $7 \leq \bar{x} \leq 103$. The area under that range between the normal curve and the horizontal axis is very closer to 99.73% of the total area under the Normal Curve.
- Ask the students "what are the two measures initiated in this analysis?

- Inquire the students about the practical instances for variables come across in day-to-day life experiences , that take a pattern of a normal distribution.
- Inquire the students about the variables come across in the business field, that take a pattern of the normal distribution.

A Guideline to Explain the Subject Matters :

- A frequency distribution that makes a frequency curve with the shape of a symmetrical bell is known as a normal distribution.
- The Mean = Median = Mode of a Normal distribution.
- More than 99.73% of the total area under the normal curve is distributed in the range of + and – three standard deviations to either side of the mean.
- The area under the normal curve in the range of one standard deviation to left $(\mu - \sigma)$ and one standard deviation to right $(\mu + \sigma)$ is 68.27% of the total area under the normal curve.
- The area under the normal distribution in the range of two standard deviations to the left $(\mu - 2\sigma)$ and two standard deviations to the right $(\mu + 2\sigma)$ from the mean is 95.45% of the total area under the Normal distribution.
- The area under the normal curve in the range of three standard deviation left $(\mu - 3\sigma)$ and three standard deviations to the right $(\mu + 3\sigma)$ is 99.73% of the total area under the normal curve.
- The Normal distribution which satisfies these characteristics can be used as a probability model related to continuous random variables.
  - The probability density function of a normal distribution is

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

  - Since the distribution of area under the normal curve is determined on the mean and the standard deviation, the mean μ and variance are considered as the two parameters on which a normal distribution is determined.
  - Some of the variables come across in day-to-day life such as the marks of an examination, height of the individuals, weight, life expectation etc… can be expected to take a pattern of a normal distribution.
  - The business variables such as daily sales income, monthly overhead cost, daily output, employee productivity etc… can be expected to take a normal distribution pattern.

Competency 5.0 : Demonstrates the preparedness to face business risk

Competency Level 5.18 : Uses the Standared normal distribution to solve the probability based problems.

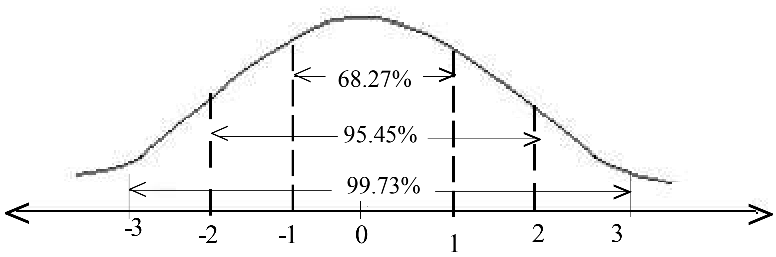
No. of Periods : 12

Learning outcomes :

- Introduces the standard normal distribution.
- States the probability density function of the standard normal distribution
- Transforms normal distribution to standard normal distributions.
- Lists the characteristics of the standard Normal distributions.
- Differentiates the standard normal distribution from the normal distribution.
- Solves problems using the standard normal distribution table.
- Solves problems related to binomial distribution using the normal approximation.
- Solves problems related to poisson distribution using standard normal distribution.
- Explains the importance of the standard normal distribution.

Instructions for Lesson Planning :

- Display the following diagram before the class.



- Inquire the students about the similaries and assmilarites between this diagram and frequency curve constructed in the previous lesson.
- Hold a discussion highlighting the following facts.

  99.73% of the total area under the normal curve is dispersed in the range of the 3 standard deviations to either side from the mean.
- The two ends of the normal curve do not touch the horizontal axis since there is a probability of 0.135% for the considered normal variable to take any value up to plus or minus infinity beyond each limit of + or – 3 standard deviations.
- Emphasis the fact that the relevant variable has been indicated as ‘Z’ and its mean as “0” in this diagram.

- Point out that the Normal Distribution that has been considered in this manner such that mean $\mu = 0$ and variance $\sigma^2 = 1$ is known as the Standard Normal distribution.
- Explain that various variables that take a pattern of a normal distribution (x) learnt in the above lesson can be transformed to theoretical distribution that takes the pattern shown in this diagram.
- Discus with the students how to convert a value taken by any continuous variable denoted by X which is in the pattern of a normal distribution, in to a value of the standard normal variable (z) .
- Introduce the Probability density function of the standard normal distribution.
- Point out that the area of one tail (0.5) lying on either side from the mean under normal curve has been tabulated broadly with sub divisions of the area even too tiny intervals.
- Train the students to transform (convert) any value taken by X to the corresponding Z value of it, using the standard normal distribution table.
- Engage the students in following activity.
- The mean and standard deviation of the marks distribution in an examination are 55 and 16 respectively. If a particular candidate who sat for this exam is drawn in random find the probability that his marks will be.
    1. Between 55 and 79
    2. Exceeding 79
    3. Less than 79
    4. Between 35 and 55
    5. Less than 35
    6. More than 35
    7. Between 40 and 60
    8. Find the minimum marks received by a candidate belongs to the highest 10% marks interval of this examination.

Solution :

Let's suppose that the marks of the examination is denoted by X

$\mu = 55$ $\sigma = 16$

$$X \sim N(\mu, \sigma^2)$$

$$\bar{X} \sim N(55, 16^2)$$

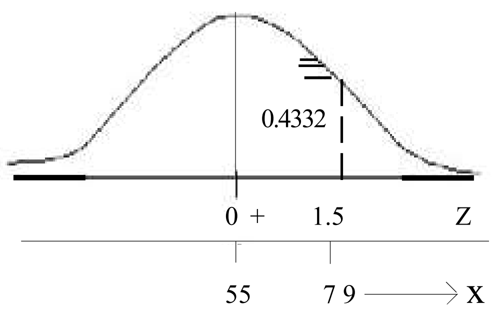
$Z = \frac{x-\mu}{\sigma} \sim N(0,1)$

$= \Pr\left(\frac{55-55}{16} \le Z \le \frac{79-55}{16}\right)$

$= \Pr(0 \le Z \le 1.5)$

$= \underline{\underline{0.4332}}$

$\Pr(55 \le x \le 79)$



- The probability that the marks received by the random chosen candidate falling between 55 and 79 = 0.4332

$\Pr(x \ge 79)$

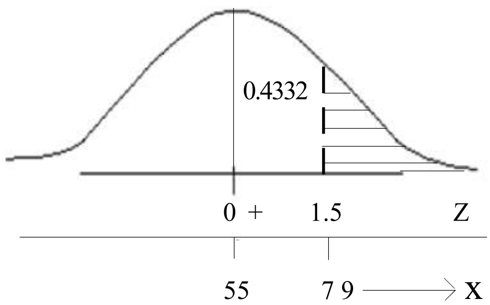
(2)

$= \Pr\left(Z \ge \frac{79-55}{16}\right)$

$= \Pr(Z \ge 1.5)$

$= 00.5000 - 0.4332$

$= \underline{\underline{0.0668}}$



The probability that the marks of the random selected candidates being greater than 79 = 0.0678

$\Pr(x \le 79)$

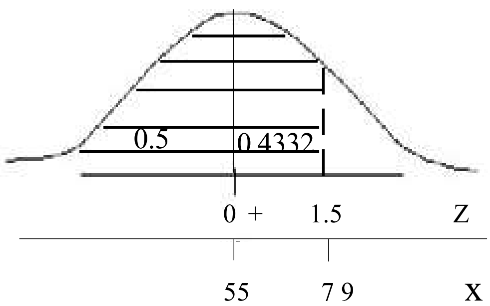
(3)

$= \Pr\left(Z \le \frac{79-55}{16}\right)$

$= \Pr(Z \le 1.5)$

$= 0.5000 + 0.4332$

$= \underline{0.9332}$



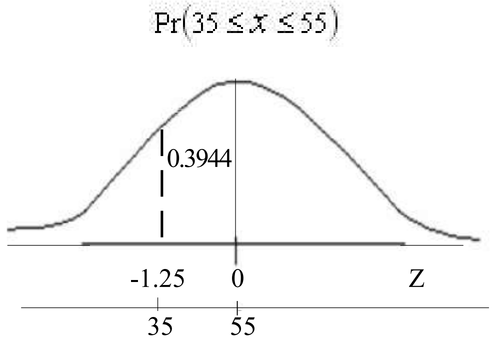
The probability that the marks received by the random selected candidate being less than 79 = 0.9332

(4)

$$= \Pr\left(\frac{35-55}{16} \le Z \le \frac{55-55}{16}\right)$$

$$= \Pr(-1.25 \le Z \le 0)$$

$$= \underline{\underline{0.3944}}$$



Probability that the marks of the random selected candidate falling between 35 and 55 = 0.3944

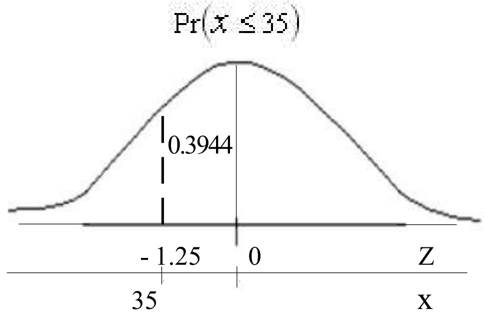
(5)

$$= \Pr\left(Z \le \frac{35-35}{16}\right)$$

$$= \Pr(Z \le -1.25)$$

$$= 0.5000 - 0.3944$$

$$= \underline{\underline{0.1056}}$$



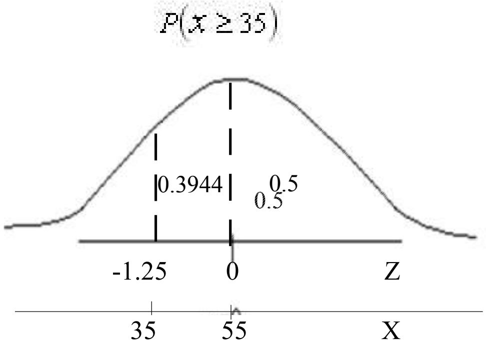
The probability that the marks received by the random selected candidate being less than 35 = 0.1056

(6) $= \Pr\left(Z \ge \frac{35-35}{16}\right)$

$$= \Pr(Z \ge -1.25)$$

$$= 0.5000 + 0.3944$$

$$= \underline{\underline{0.8944}}$$



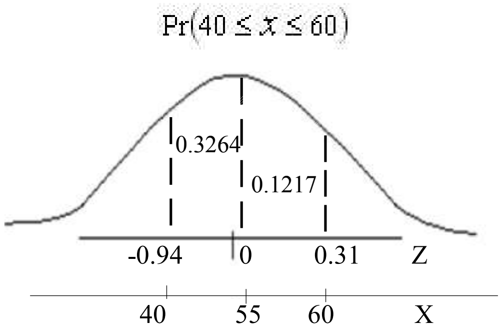
The probability that the marks received by the random selected candidate being exceeding 35 = 0.8944

(7) $= \Pr\left(\frac{40-55}{16} \le Z \le \frac{60-55}{16}\right)$

$$= \Pr(-0.94 \le Z \le 0.31)$$

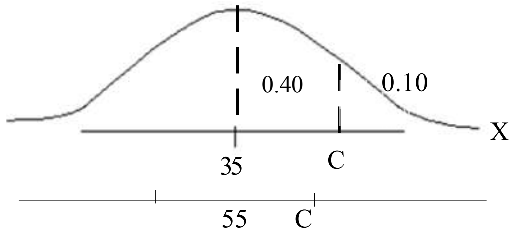
$$= 0.3264 + 0.1217$$

$$= \underline{0.4481}$$



The probability that the marks received by the random selected candidate falling between 40 and 60 = 0.4481

(8)



If the minimum mark in the interval of 10% highest marks is C, the respective Z value for C is 1.28.

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

$$\frac{1.28}{1} = \frac{x-55}{16}$$

$$x-55 = 1.28 \times 16$$

$$x-55 = 20.48$$

$$x = 20.48 + 55$$

$$= \underline{\underline{75.48}}$$

$\therefore$ The minimum marks received by a candidate belongs to the group of highest marks earned is 75 marks.

- Solve the following discussing with the students.

1. A researcher is interested to know the average life time and standard deviation of an electric bulb manufactured in a firm. It is known that 0.62% of the bulbs are expired after minimum life time of 650 hours. Further the percentage of bulbs expired before 575 hours is 11%. Considering that the life time of the bulbs distributes normally, find the mean and standard deviation of the life time of a bulb.
2. Following details have been revealed in a study launched to consider the number of customers coming to a supermarket and the amount of money they spend there.
   The amount of money spent by a customer falls normally with mean Rs. 1500/= and the standard deviation Rs. 1000/=. The number of customers who spend an amount between Rs. 1000/= and Rs. 1800/= is 62. Find,
   (i) The total number of customers expected on that day
   (ii) Number of customers that can be expected to afford more than Rs. 2000/=
3. 40% of the workers employed in a factory are women. If a random sample of 150 workers is drawn, find the probability that there will be at least 50 women in it.
4. The average number of customers coming to a bank in the last working hour is 25. Find the probability that at most 30 customers coming to the bank in the last hour of the next working day.

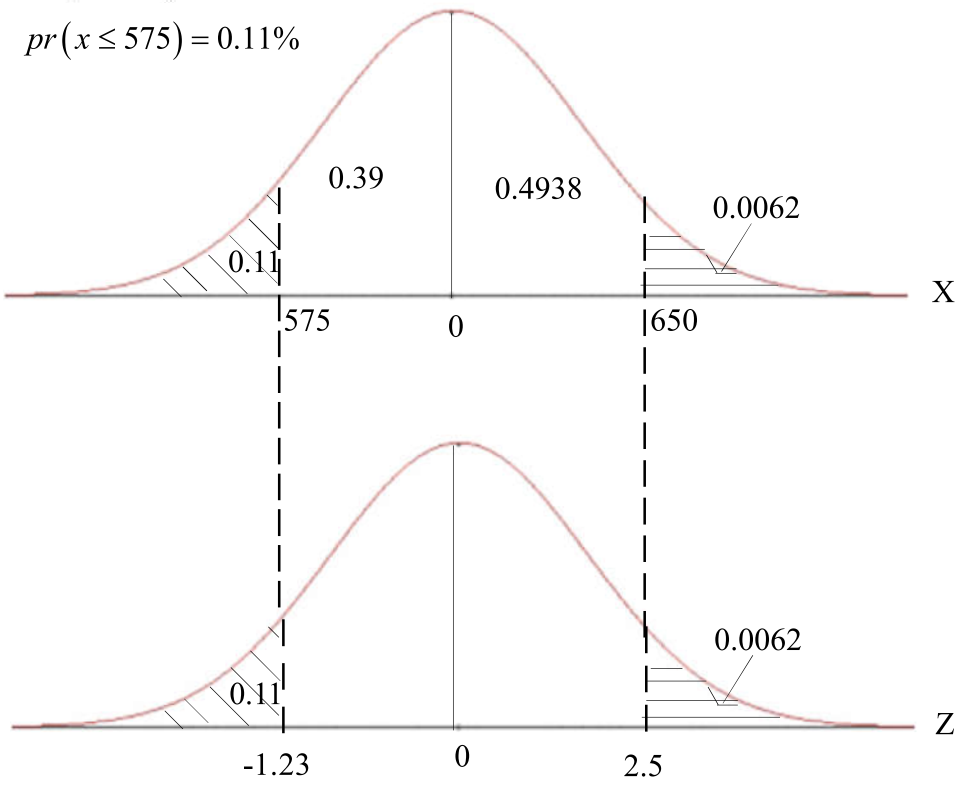
Solution :

Lets suppose that the life time of an electric bulb as X.

$x \sim N(\mu, \sigma^2)$

$\Pr(x \geq 650) = 0.62\%$

$pr(x \leq 575) = 0.11\%$



Z value for 650 hours = 2.5

Z value for 575 hours = -1.23

Using the formula

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

$$2.5 = \frac{650 - \mu}{\sigma}$$

$$2.5\sigma + \mu = 650 \quad \text{———} \quad (1)$$

$$, -1.23 = \frac{575 - \mu}{\sigma}$$

$$-1.23\sigma + \mu = 575 \quad \text{———} \quad (2)$$

$$(1) - (2) \; 3.73\sigma = 75$$

$$\sigma = \frac{75}{3.73}$$

$$\sigma = \underline{\underline{20}}$$

$\sigma = 20$ (1) ට ආදේශ කර

$$2.5 \times 20 + \mu = 650$$

$$\mu = 650 - 50$$

$$\mu = \underline{\underline{600}}$$

$$\therefore \; \underline{\underline{\mu = 600}} \qquad \underline{\underline{\sigma = 20}}$$

- Let's suppose that the amount of money spent as X.

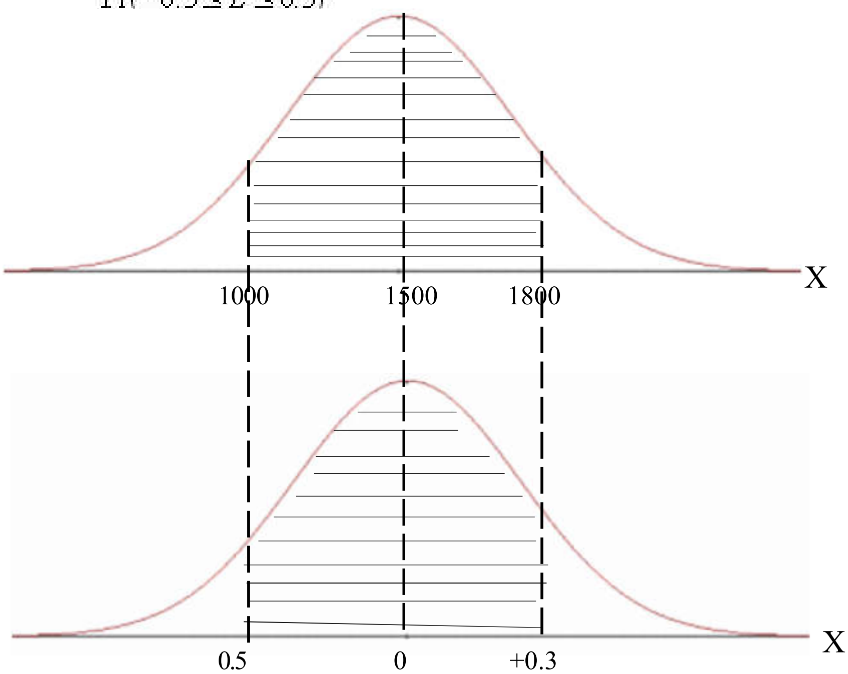
$$X \sim (\mu, \sigma^2)$$

$$x \sim N(1500, 1000^2)$$

$$\Pr(1000 \le X \le 1800)$$

$$\Pr\left(\frac{1000-1500}{1000} \le Z \le \frac{1800-1500}{1000}\right)$$

$$\Pr(-0.5 \le Z \le 0.3)$$



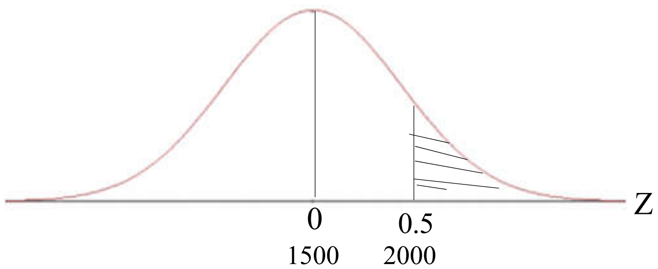
$$0.1915 + 0.1179 = \underline{0.3094}$$

- The total number of customers with respect to the probability of 0.3094 is 62. Hence, the total number of customers

$$= \frac{62}{0.3094} = \underline{\underline{200}}$$

$$= \Pr(x \geq 2000) \qquad = \Pr\left(Z \geq \frac{2000 - 1500}{1000}\right)$$

$$= Pr(Z \leq 0.5) = 0.3085$$



$$= 0.3085 \times 200 = 62$$

A Guideline to explain the subject matters :

- The normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is known as the standard normal distribution.
- The probability density function of the standard normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}}{}^{-\frac{1}{2}(z^2)}$$

- The standard normal curve does not touch the horizontal axis and assumed that it trends up to minus infinity and + infinity from either end.
- 99.73% of the area under the standard normal curve is contained in the range $-3 \leq Z \leq +3$
- The following formula is used to convert the value of X that indicates any continuous variable falls normally to the standard normal variable value Z.

$$Z = \frac{x - \mu}{\sigma} \sim N(0,1)$$

- The total area under the normal curve is considered as 1 square units.
- This is based on the concept that the sum of the relative frequencies of each and every class interval of a frequency distribution being 1.

- The area of one tail to left or right of the total area under the normal curve is 0.5 and that area has been sub divided in to too tiny intervals and tabulated broadly in the table of standard Normal Distribution.
- The probability that a continuous variable taking a particular value or being contained in a range of values can be derived using this table.
- The problems related to binomial distribution can be solved using the normal approximation whose **n** is too large and
  - $P = 0.5$ or
  - $np \geq 5$ or
  - $nq \geq 5$

  taking the parameters as $np = \mu$ and $npq = \sigma^2$
- The problems related Poisson distribution also can be solved using normal approximation when $\lambda \geq 20$ taking the parameters as $\lambda = \mu$ and $\lambda = \sigma^2$

- In order to convert the values in discrete variables to continuous variable when the problems related to binomial distribution and Poisson distribution are solved using the normal approximation a values of ± 0.5 is adjusted. This is called the continuity correction factor.
- Importance of a normal distribution.
  - Since most of the problems found in practice are based on continuous random variables, the probability based problems related to the events can be solved easily using the knowledge of Normal distribution.
  - Even the problems modeled in binomial poisson distributions can be solved easily using the normal distribution subject to satisfying the required conditions.
  - The normal distribution is applied in high frequencies in coming to conclusions in statistical inferences.

Competency 6.0 : Uses appropriate sampling techniques for collecting data required to, make 'Business Decisions.'

Competency level 6.1 : Plans a sample survey for statistical Inference.

No. of Periods : 06

Learning outcomes :

- Introduces 'statistical Inference'.
- Differentiates between 'population' and 'sample'.
- Differentiates between census (complete enumeration) and sample survey.
- Differentiates between 'statistics' and 'parameters'.
- Explains 'Sampling'.
- Describes sampling frame and sampling unit.
- Differentiates between 'sampling with replacement' and 'sampling without replacement'.
- Explains the need of a sample survey.
- Explains the advantages of sample survey relative to 'census'.
- Proposes the appropriate sampling frame for drawing various samples.
- Mentions the situations where the sample surveys should not be used.
- Lists out the steps of sample survey.
- Introduces sampling errors.
- Introduces non-sampling errors and reasonifies for those errors.

Instructions for Lesson Planning :

- Pay attention of the students to following situations.
    - Think of an occasion where you taste a small piece of 'Kakudodol' before buying.
    - Think of an occasion, where a small piece of mango is offered you to taste by the fruit seller before buying mangoes.
    - Think of an occasion where a little rice is taken to the palm and checked, before you buy the rice from the market.
    - Think of your mother tasting the salt and sour of a curry cocked at home.
- Hold a discussion highlighting the following facts.
    - Inquire the view of students about the way we face these situations in our day to day life.
    - Before buying 'Kaludodol' having tasted a small slice we may decide whether to buy or not.

- Before buying mangoes, having tasted the small piece of mango offered by the fruit seller, we may decide whether to buy or not.
- Before buying dry rice from the market, the consumer takes a little rice to his palm and makes a decision whether to buy or not, having carefully checked the colour, odour, and composition etc.
- Mother decides the taste of the curry having tasted a little bit taken to her palm.
- Hence emphasize the fact that a decision is made regarding the entirety having studied a small portion drawn out from the entirety (population) and that is the statistical inference.
- Present the following situations to the students and hold a discussion.
  - Examining the mass of all the packets of tea in a tea factory made during a week in order to check the standard of the mass of their products.

    Examine the mass of 500 packets of tea in the same tea factory during two days selected in that particular week.
- Receiving the points of view of all the consumers coming to a super market during a period of one month, in order to study about the market demand for a particular type of soya meat introduced to the market recently.

  Receiving the points of view regarding that type of soya meat from the consumers coming to that super market in a day.
- In order to launch a particular project in the schools conducting G.C.E (A/l) classes in the Western Province, implementing the project in 10 selected schools from that province.
- Lead the students to categorize the above examples under 'population' and 'sample' after explaining that the set of all the elements that should be undergone to the study is 'population' in a 'portion' of elements selected representing the population is known as 'sample'.

Population	Sample
1. Mass of all the packets of tea products during a week.	1. Mass of 500 packets of tea produced during two days
2. Number of consumers came to the super market during a week	2. Number of consumers came to the super market in a day
3. All the schools where A/L classes are conducted in Western province	3. The 10 selected schools having A/L classes in Western province

- Explain that the process of drawing a small portion from the population is known as 'sampling'.

- Further point out that all the above mentioned populations are finite populations, since there is a finite number of elements contained each.
- Pointed out that when the number of population elements is numerous then it is known as an infinite population.

i.e. :
- Seeds of rice contained in a large batch of rice packs.
- Number of water fountains available in the world.

- Note down the following two statements on the board.
  1. A great deal of information in connection with the population are concealed through the population census conducted by the department of census and statistics once in every 10 years.
  2. In order to estimate the defective percentage of the bulbs manufactured in a large scale factory, 10 bulbs drawn in random in a day are carefully examined.
- Discuss with the students what revealed from these two statements.
- According to the first statement data are collected visiting from door to door in order to conceal important information inconnection to the population census.
- Hence explain the fact that, if all the elements in the population are evaluated individually that process is known as a census.
- Further point out that it would be a difficult task to test each and every bulb manufactured in the factory since a tremendous time, labour and other resources are discharged in large scales and also that the bulb may be destroyed in testing, so that it would be highly recommended to test only a sample of 10 bulbs drawn out in random daily.
- Hence explain that examining all the elements individually in great details in a sample representing the entire population is also practically implemented.
- Explain that a great deal of time, cost and labour are spent for the population census conducted by the department of census and statistics. Further point out that data should be collected covering each and every house in the country since all the people in Sri Lanka should be represented trained invigilators should be employed spending considerable cost to finalise the census and it takes much time to issue the rersults.
- Point out further that it is difficult to collect accurate and reliable data in a country experiencing a civil war. In connection with a manufacturing firm conducting a census may be harmful, since the population elements can be damaged or destroyed through testing.
- Point out that a sample survey can be conducted to understand the defective proportion of bulbs in the bulb manufacturing factory by testing a sample of 10 bulbs in a day related to the 2nd statement mentioned above.
- Point out that the time labour and cost spent for this is less than for a census. Results of a sample survey can be derived quickly. All the sample elements can be carefully

checked in great details in every aspect. The loss of destroying the elements is also minized.

- Note the following two statements on the board to explain parameters and statistics.
  1. Finding the average monthly salary of the people living in Colombo city.
  2. Finding the average monthly salary of the people living a selected lane in Colombo city.
- Point out that the first statement is a fact in connection with the population; a measure in connection to population is required to be computed and that should be the population mean μ
  - Population mean is a constant
  - Population mean is unknown
  - The measures computed in connection with the population are known as 'Parameters'
- Point out that the 2nd statement is related to a sample.

  Sample mean ($\overline{X}$) should be computed and any measure calculated for a sample is known as a 'statistic'.

Note down the following four documents on the board.

- The voters' Registry of Colombo district
- The student Admission Registry in the school
- The Employee Enrolment Registry in a firm
- The patient Admission Registry in a hospital

- Hold a discussing highlighting the following facts
  - The Voters' Registry is the source containing all the individuals entitled with universal franchise in Colombo district.
  - The Students Admission Registry is the source containing all the information of the students admitted to the school.
  - The Employee Enrolment Registry is the source containing all the details about the employees recruited to the firm.
  - The Patient Admission Registry is the source containing all the details about the patients admitted to the hospital.
- Hence point out that a list completed with the purpose of identifying all the members of the population is known as a sampling frame. Further point out that all the details should be included these sampling frames accurately and no any false details should be included. No any name should be repeated and list should be perfectly finalized.
- Pay attention of the students to each of the following phrases.
  - Each voter included in the voter's registry in Colombo district

- Each student in the school in a study in connection to the school
- Each inward patient in connection with a study about the patients in a particular hospital

- Explain that each element included in a population is known as a sampling element as mentioned in the above phrases.
- Engage the students in following activity to explain the sampling with replacement and sampling without replacement.
  - Derive all the possible samples of two,
    1. Without replacement
    2. With replacement

    From the population of number 2, 7, 9

Answer

- Point out that the number of all the possible samples without replacement per 'n' in size from N population elements can be computed ${}^{N}C_{n}$ as follows.

$${}^{N}C_{n} = \frac{N!}{N!(N-n)!}$$

$$\therefore {}^{3}C_{2} = \frac{3}{2!}$$

$$= \frac{3 \times 2!}{2!}$$

$$= \underline{\underline{3}}$$

- Samples without replacement (2,7) (2,9) (7,9)
- Point out that the number of all the possible samples with replacement per n in size from N population elements is N^{n}

$$\therefore N^{n} = 3^{2} = \underline{\underline{9}}$$

Samples with replacement.	(2,7) (2,9) (7,9)
	(7,2) (9,2) (9,7)
	(2,2) (7,7) (9,9)

- Point out that the process of drawing a sample from 2, 7, 9 of population without inserting the first drawn elements to the population before drawing the second elements is called sampling without replacement and all the possible number of samples that can be drawn in that manner can be found as ${}^{N}C_{n}$. Hence the probability of a population element being selected to a sample should be $\frac{1}{{}^{N}C_{n}}$

Accordingly point out that the probability of an element among 2, 7, 9 being selected to the sample is 1/3.

- Point out that when a sample is drawn out from 2, 7, 9 population if the first drawn element is inserted (replaced) to the population before the second element is drawn out that is known as sampling with replacement.
- Hence point out that all the possible number of samples with size n that can be drawn with replacement from a population containing N elements can be found as N^n and the probability of a population element being selected to the sample is $\frac{1}{N^n}$

According to the number of all possible samples with size **n** that can be drawn out from the population 2, 7, 9 is 3^2. And the probability of an element being elected to the sample should be 1/9.

Lead the students to read the following text.

Let's consider a study launched in search of the average monthly income of a family living in a particular city. Suppose that there are 10 000 families in this city and the average monthly income of a family has been computed as Rs. 15 000. Thereafter 1 000 families were drawn and computed the average monthly income of a family as Rs. 10 000/-

Answer the following questions with reference to the text.

1. How do you name the average income of Rs. 15 000/-
2. How do you name the average income of Rs. 10 000/-
3. What is the difference between these two averages?
4. What are the reasons caused for this difference?

Answer :

1. Population Mean
2. Sample Mean
3. Rs. 15 000/- Rs, 10 000 = Rs. 5 000
4. Sample mean depends on the sample elements where as population mean is computed using all the elements in the population.

- Hold a discussion highlighting the following facts.
- The set of all the families living in this city is considered as the population.
- Population Mean μ has been derived by dividing the sum of the salaries of 10 000 families by total number of families (10 000)
- Having drawn out a sample of 1 000 families from the population the total salary of those 1 000 families has been divided by number of sample members and the sample mean $\bar{X}$ has been derived.
- The difference between the population mean and the sample mean is known as the 'Sampling error"
- Various factors causes for generating a sampling error.
- Sample mean depends only on sample elements, but population mean depends all the population elements.
- Pay attention of the students to the below mentioned facts.
  1. It was later revealed that majority of the families in this city has cealed their true income in fear of getting caught by the department of Inland Revenue.
  2. It was also revealed that the invigilators deployed in ample survey have falsely reported the income of the sample members.

Hold a discussion highlighting the following facts.

- People provides with false details regarding the economic factors like income with or without knowing.
- Errors can be occurred in reporting, tabulating or copying of data and those errors are known as non sampling errors.
- These non sampling errors are difficult to be controlled for many reasons.

A Guideline to explain the subject matters :

- After a small portion (sample) is drawn out from the entirely (population) coming to conclusions about the entirety, having studied that small portion in great details is called "Statistical Inference"
- In other words statistical Inference is come to statistical conclusions regarding the population based on the information revealed through a sample survey.
- When considered all the elements that should be studied as a whole that is known to be a 'population'
- Populations are in two types as finite populations and infinite populations.
- A selected portion from the entirety so as to represent all the elements in it is known as a sample. Even though it would be easy to draw out samples from the population when all the elements are homogeneous, it would be quite different and complicated to draw out representative samples, when the population elements are heterogeneous.

- The process of drawing a sample from the population is known as sampling.
- If each and every population element is individually evaluated, that process is called a census or a complete ennumeration.
- Having drawn out a sample from the population, evaluating each and every sample elements individually is known as a sample census or a sample survey.
- Differences between a census and a sample survey are as follows :

Census	Sample Survey
1. Time Consuming	1. Time saving
2. Trained labour is highly required	2. Less labour consuming
3. Very expensive	3. Less cost is to be incurred
4. It takes much time for deriving the results of the study	4. Results oriented in a short period
5. Units being destroyed when each population element being studied individually	5. Destroying the elements is at a minimum level
6. Population elements being undergone to a micro and a broad analysis is very difficult	6. Sample elements can be undergone to a micro and a broad analysis

Steps of a sample survey are as follows :

1. Clarification of objectives of the study
2. Selecting the relevant sampling frame
3. Determining the sample size
4. Choosing the most appropriate sampling method
5. Identification of sample elements
6. Collecting data from the sample that was drawn out
7. Synthesis of data
8. Analysis of data
9. Interpretation of the results and coming to conclusion

- The statistical measures through which a population is interpreted are called 'parameters' Accordingly the population mean μ, population standard deviation σ etc. are parameters.

- The statistical measures by which a sample is interpreted are called 'statistics'. Accordingly the sample mean $\bar{X}$ sample standard deviation 'S" etc are considered to be sample statistics.

Parameters and statistics can be differentiated as follows.

Parameters	Statistics
1. A population characteristic	1. A Sample characteristic
2. Parameters are constants	2. Statistics are variables
3. Value is unknown	3. Value is known
4. Value is estimated	4. Value is computed

- A list of all the elements expected to be undergone to the study is defined as a sampling frame or else a list of sampling elements which is used to draw out a sample required to get information through a study is known as a sampling frame.

  Ex : Voters' Register. List of chief occupants, Telephone Directory etc.
- A good sampling frame should be fulfilled with following properties.
  - The sampling frame should be perfect (complete).
  - The sampling frame should be up to date.
  - The sampling frame should be accurate.
  - The elements should \not be repeated in the sampling frame.
  - When a particular population is defined each element included in that population is known as a sampling element.

    Ex : each patient, each voter, each school student etc….
- After drawing the first element from a particular population to the sample, if the next element is drawn out before the first drawn elements is put back in it, that process of drawing the sample is known as sampling without replacement. The number of samples that can be drawn without replacement with size 'n' from a population consists of 'N' number of elements is calculable by ${}^{N}C_{n}$ The probability of one population element being selected to the sample is $\frac{1}{{}^{N}C_{n}}$
- After one elements is drawn out from the population, if that element is put back in to it before the next element is drawn, that process of drawing the samples is known as sampling with replacement.

The number of samples with size ‘n’ that can be drawn out with replacement from a population with the size ‘N’ is calculable as N^n. Hence the probability of a population element being selected to the sample can be found as $\frac{1}{N^n}$

- The difference between the results of a population study and the result of the interference for the entire population through a sample survey is called “sampling error”.

Factors caused for a sampling error are mentioned below.

- Not using an appropriate sampling method
- Not using a complete and accurate sampling frame
- Not employing trained invigilators
- Population being consists of a greater variation

- Possible errors occurred in the process of data collecting, recording, tabulating, copying and computerizing are known as non sampling errors. These errors can be occurred in a census as well. Even though the sampling errors can be minimized applying various techniques, these non sampling errors are very difficult to be controlled.
- Factors caused for non sampling errors are as follows :
    - Not planning the experiment or sample survey accurately
    - Not employing the trained invigilators
    - Getting false data
    - Possible errors occurred in analysis of data
    - Possible errors in data processing

Assessment & Evaluation :

Activity – I

State whether the census or sample survey is appropriated for each occasion mentioned below.

1. Collecting data related to sanitary facilities in a particular rural area .
2. To examine the maximum weight that can be born by the iron rods manufactured in a factory.
3. To aware of the price level of goods and services in an area experiencing a civil war.
4. To collect data related to the composition of population in a country.
5. To estimate the life time of a type made in a factory manufacturing types.
6. To have a knowledge about the number of children suffering from mal-nutrition among the school children.

7. To come to know about the responses of spectators for a particular T.V. programme.
8. To search for regarding the sales of a newly introduced product to the market.

Answers :

1. Population	5. Sample survey
2. Sample survey	6. Population
3. Sample survey	7. Sample survey
4. Population	8. Sample survey

Activity 2:

Note down the following situation on the board and lead the students to propose appropriate sampling frames for each situation.

1. To search for the details of leaves obtained by teachers in a school.
2. To analyse the salaries of top grade officers employed in a factory.
3. To search for the details of residents in a Grama Niladari division, who are elder than 18 years during the last two years.
4. To choose few houses having land phone connections in an area.
5. To search for the number of text books received by a school in last year.

Answers :

1. Teachers Leave Register
2. Pay sheet
3. Voters Register in the relevant Grama Niladari Division.
4. Telephone Directory
5. School Text Book Order List

Competency 6.0 : Uses appropriate sampling techniques for collecting data required to, make 'Business Decisions.'

Competency level 6.2 : Uses probabilistic sampling methods in the process of sampling.

No. of Periods : 16

Learning outcomes :

- Explains probabilistic sampling.
- States the situations where probabilistic sampling is applicable.
- Lists out the advantages and disadvantages of probabilistic sampling.
- Interprets simple random sampling.
- Draws outs simple random samples from a finite population.
- Gives examples for situations where simple random sampling method is applicable.
- Lists out advantages and disadvantages of simple random sampling method.
- Interprets the stratified probabilistic sampling method.
- Draws out stratified probabilistic samples from a finite population.
- Gives examples for situations where stratified sampling is applicable.
- Lists out advantages and disadvantages of stratified sampling method.
- Interprets 'Cluster sampling'.
- Describes the concepts related to cluster sampling.
- Gives out examples for situations where cluster sampling is applicable.
- States the advantages and disadvantages of cluster sampling.
- Interprets cluster sampling with one stagd clusters, two staged clusters and multi staged clusters.
- Describes how to draw out a sample on one staged, two staged and multi staged cluster sampling techniques.
- Interprets 'Systematic sampling'.
- Lists out the steps followed in drawing out a sample on systematic sampling method.
- Gives examples for situations where systematic sampling method is applicable.
- States the relative advantages and disadvantages of each method describing the relationship among systemetic sampling, stratified sampling and cluster sampling.

Instructions for Lesson Planning :

Present the below mentioned case to the students to explain ' Probabilistic Sampling'.

Activity I :

Five students among 40 in a grade 12 class should have been selected to be sent to take part in an extracurricular event held in a neighboring school. Following optional suggestions have been forwarded to select these five students.

Suggestion I : Selecting five students on the desire of the class teacher

Suggestion II : Selecting five students through a reftle draw from all the 40 students

Suggestion III : Selecting three students through a raffle draw from 30 students learning Business Studies and another two students through a raffle draw from 10 students learning Business Statistics

- Focus the following questions to the attention of students.
    - By which suggestion/s will the students be selected to the sample subjectively?
    - By which suggestion/s will the students be selected to the sample in random?
    - Can the probability of the student named X in this class being selected to the sample be found in accordance with suggestion-I? If possible, what is the probability of it? If impossible, why?
    - If this sample is drawn in accordance with the suggestion-2. What is the probability that the student X in the class being selected to the sample?
    - If this sample is drawn in accordance with the suggestion -3 what is the probability of X who is studying statistics and Y who is studying Business Studies being selected to the sample?

Hold a discussion highlighting the following facts.

- The probability of sample elements being admitted to the sample cannot be expressed, when the samples are drawn subjectively.
- When the samples are drawn in random like in lottery system the probability of sampling elements being admitted to the sample can be expressed.
- Hence, in accordance with the suggestion-2 the probability that X being admitted to the sample is 5/40 = 0.125
- In accordance with suggestion-3 the probability that X being admitted to the sample is 2/10 = 0.2 and the probability that Y being admitted to the sample is 3/30 = 0.1
- Drawing samples so as to be able to find the probability of sampling elements to be admitted to the sample is known as probabilistic sampling.
- Suggestion-2 and suggestion-3 are two probabilistic sampling methods.

- Probabilistic samples can be drawn by grouping the population or by not grouping the population. According to the suggestion-2 the sample is drawn without grouping the population and in accordance to suggestion 3 the sample is drawn by groping the population as the student learning Business Studies and students learning Business Statistics.
- Since the sample is drawn under simple random sampling without grouping the population, there is an equal chance (probability) for each and every sampling element to be admitted to the sample.
- Among various methods that can be followed in grouping the population for drawing out a probabilistic sample separating to strata, clusters or class intervals are very popular.
- Engage the students in Activity 2 and Activity 3 for introducing simple random sampling techniques.

Activity 2 :

- Lead the monitor or monitress of the class to write down the serial number of each student in order of attendance register on same size coupons and to mix then up very well.
- Get another student to draw out five coupons in random from the monitor/monitress.
- Announce the number mentioned on each of those five coupons aloud before the class.
- Call upon those five students before the class referring to the attendance register.
- Explain that drawing a simple random sample in this manner is known as drawing a sample on lottery system.

Activity 3 :

Display the following random number table before the class.

03	37	43	07	50
24	16	35	12	46
38	10	22	02	40
17	44	05	28	34
33	21	11	42	13
01	32	08	27	20
29	15	39	06	09
36	30	26	14	04
45	18	19	31	47
48	25	41	23	49

- Name a student in the class and guide him/her to read five successive numbers on cross ward or down ward beginning from any point. (if any serial number beyond the numbers in the register is come across in reading that manner let it be cancelled and read the next successive number)
- Select the names of the five students with respect to the numbers drawn out from the random number table and call them up on before the class.
- Explain that drawing a simple random sample in this manner is known as using random number table technique.
- In addition to these two techniques, further point out that a simple random sample of five students can be drawn out giving instructions to the computer if the list of names in the class has been computerized
- Provide with following details to the students regarding the employee population in a firm.

Employee category	Number of employees	Monthly salary range Rs.
Managers	20	95 000 – 100 000
Management Assistants		
Grade 1	100	40 000 - 41 000
Grade II	280	30 000 - 30 500
Labourers	600	25 000 - 25 800

- Inform the students that a sample of 50 employees should have been drawn out to launch a study regarding the anormalities in salaries of a firm.
  - Hold a discussion highlighting the following facts :
    - This is a heterogeneous population
    - Salary anomalities between employee categories are significant
    - Salary anomalities within each employee category are ignorable
    - Population consists of strata in accordance with salaries
    - Sample of 50 employees should have been drawn so as to represent all the strata
    - Selecting employees from each stratum should not be done subjectively and employees should be drawn from each stratum to the sample based on simple random sampling techniques.
- Inform the student that following two suggestions have been forwarded.

Suggestion I : Selection of sample members as 2 managers, 4 Grade I Management Assistants, 14 Grade II Management Assistants and 30 laboures.

Suggestion II : Selection of 50 employees to the sample on the ratio of salary strata.

- Lead the students to find the following probabilities if the sample is drawn in accordance with the suggestion I.
  - Probability of a manager being entered the sample
  - Probability of Grade I management assistant being entered the sample
  - Probability of Grade II management assistant being entered the sample
  - Probability of a labourer being entered the sample
- Lead the students to compute the probability of an employee in each category being entered the sample, if the sample is drawn in accordance with the suggestion 2 above.
- According to the results derived make the students aware that a sample element from each stratum, can be selected to the sample with equal probability values or unequal probability values.

Solutions :

When the sample is drawn on suggestion I :

- Probability that a manager being entered the sample $= \frac{2}{20} = 0.1$
- Probability that a grade I Management Assistant being entered the sample $= \frac{4}{100} = 0.04$
- Probability that a grade II Management Assistant being entered the sample

  $= \frac{14}{280} = 0.05$
- Probability that a Labourer being entered the sample $= \frac{30}{600} = 0.05$

- When the sample is drawn on suggestion 2, since the number of managers that can be admitted to the sample is $\left(50 \times \frac{20}{1000}\right) = 1$
- $\therefore$ Probability that a manager being entered the sample $\frac{1}{20} = \underline{\underline{0.05}}$
- Since the number of grade I management assistant that can be

  admitted to the sample is $= 50 \times \frac{100}{1000} = 5$

- Probability that a grade I management assistant being admitted to the sample $\frac{5}{100} = \underline{\underline{0.05}}$

- Number of grade II management assistants to be admitted to the sample is 14

$$= 50 \times \frac{280}{1000} = 14$$

- Probability that a grade II management assistant being admitted the sample $\frac{14}{280} = 0.05$

- Number of labourers to be admitted to the sample is $= 50 \times \frac{600}{1000} = \underline{\underline{30}}$

- Probability that a labourer being admitted the sample $= \frac{30}{600} = 0.05$

- In order to explain cluster sampling pay attention of the students to following situations separately and let them come up with their points of view about the way of a sample can be drawn
- In a study about the earning of individuals, a sample of citizens in a provincial council consists of 10 Pradesheeya Saba divisions, should have been drawn
- In a study to know about the purpose of commuters traveling buses from Colombo to Kandy, a sample of commuters should have been drawn
- Hold a discussion highlighting the following facts on the ideas given by students :
    - There are citizens in every Pradeshiya Sabha, earning income by means of various tasks, and in each Pradeseeya Saba division the citizens earning on all those options can be expected, so that selecting sample members from every Pradeseeya Saba is not necessary and selecting one of those 10 Pradeseeya Saba divisions on simple Random sampling method is more convenient.
- When a sample is drawn in this manner the cost, labour and money to be spent is relatively less.
- In this context each Pradessya Saba can be defined as a cluster.
- Selecting a cluster at the first phase it self in that manner is known as one staged cluster sampling.
- After separating the selected Pradeseeya Sabha division in to villages as clusters, having selected few villages on simple Random Sampling techniques, all the citizens living in those selected villages can be taken as the sample.
- Then that sampling is called two staged cluster sampling.
- After dividing a Pradeesiya sabha in to villages and villages in to families, then all the family members in selected families from random selected villages also can be included in the sample.

- Then the sampling is known as malt staged cluster sampling.
- The commuters traveling from Colombo to Kandy in bus, are going for various purposes. When all the buses are considered, the commuters most probability travelling for those purposes are in almost all the buses.
- Then drawing sample elements from all the buses is not necessary. Selecting few buses in random all the commuters traveling in them can be taken to the sample.
- If necessary the buses can be categorize as S.L.C.T.B. buses and private buses, and again as normal buses, luxury buses and smi-luxury buses. Then few buses can be drawn in random from each category and all the passengers in those selected buses can be taken to the sample.
- In order to explain the systematic sampling divide the students in the class in to two groups and provide with the attendance register to one group and the teachers' attendance register to the other group. Lead them to draw out a sample of 10 elements under the following guide lines.

  1. Considering the register received by your group as the sampling frame and divide the number of elements included in it by 10.
  2. Round off your result to the nearest whole number denoted by (K).
  3. Separate the list of names in the register in to K size intervals.
  4. Provide with K number of raffle coupons numbered as 1, 2, 3 … in each to the groups and ask a student to draw out a raffle coupon in random. Announce the number written in that selected coupon aloud.
  5. Select the student/teacher bearing that number in your sampling frame to the sample as the first element.
  6. Since that selected serial number in the list draw out every K^{th} element to the sample.

A Guideline to explain the subject matters :

- Drawing a sample in accordance with the principles of probability is known as probabilistic sampling.
- Sampling elements are admitted to the sample with a known probability value is probabilistic sampling.
- Few advantages in probabilistic sampling are as follows :
  - Most probably the sample is free from being subjective
  - Getting a distinct probability value for each and every population elements to be represented in the sample
  - The population being very well represented by probabilistic samples

- Being the sampling method that must be followed in drawing samples for making inferences for the population
- Sampling error being able to be computed

- Few disadvantages of probabilistic sampling methods are as follows :
  - The inveiglilater not getting a chance to select the elements which are most suitable for the purpose of the study
  - Once the sample size is small, the population not being well represented by the sample
  - Once there are significant variations among sampling elements, this method being not that suitable
- There are four main probabilistic sampling methods.
  1. Simple random sampling
  2. Stratified random sampling
  3. Cluster sampling
  4. Systematic sampling
- Drawing a sample from the population in a manner such that each and every population element is getting an equal probability to be represented in the sample is known as simple Random Sampling method.
- A simple Random Sample can be drawn out using lottery system, or random number tables or computer programmes.
- Simple Random Sampling method is highly appropriate when a sample should have been drawn from a homogeneous population of which there is a hardly any variation.

Ex :

- Drawing a sample from a large batch of electric circuits which are manufactured in a simple production line
- Drawing a sample of grade – I management assistants for a study in connection with their salaries

- Few advantages of simple Random sampling are as follows.
  - An unbiased sample can be derived
  - The more the sample size, better the sample will represent the population
  - Sampling error can be computed

- Few disadvantages of simple Random sampling are as follows.
  - A sampling frame is essential for deriving the sample
  - This sampling method is not suitable for drawing a sample from a heterogeneous population

- When the sample size is small the population will not be very well represented by the sample
- When the sample elements are highly significant to each other this method can not be used.

- After the population is sub divided (separately identified) in few strata, drawing the number of elements determined from each stratum to the sample in random is known as Stratified Sampling Method.

Ex :

- Having stratified a population of employees in accordance of their salary scales, drawing few elements from each employee category to the sample in random
- Having stratified the school children in accordance with the grades selecting few students from every grade in random

- Few advantages of stratified sampling method are mentioned below.
  - The population is very well represented by the sample
  - Once the population is highly skewed, stratified sampling method can be applied for deriving a representative sample
  - A representative sample can be drawn out from a heterogeneous population using the stratified sampling method.
  - Parameters for each stratum can also be derived separately
  - Administration of affairs related to the sampling survey is very convenient

- Few disadvantages of stratified sampling method are also mentioned below.
  - Inability to derive a sample in the absence of a complete sampling frame
  - Being a method through which the time, money and labour are highly spent
  - Once the strata are intersected (coincided) inability of deriving a representative sample
- Cluster sampling means that, after the population is separated in to groups as clusters, choosing few clusters in random and considering all the elements in those selected clusters as the sample members.
- Grouping the population as cluster means that separating population elements in to groups in a manner such that there being greater variations among the elements inside the group and a very small variance (ignorable) between the groups.
- Cluster sampling can be practiced as one staged or two staged or multi staged cluster sampling.

- Few advantages of cluster sampling are given below.
  - being considerebly a flexible sampling method
  - being an appropriate method for a wider investigation under a lower cost
  - a sampling frame being not essential
  - being a convenient sampling method when the population naturally consists of clusters
- Few disadvantages of cluster sampling are as follows :
  - being a less accurate sampling method compared with the other sampling methods
  - determining the fact such that how many clusters should be identified in the population and what are they being assigned on the invigilator him/herself

- After interpreting 'K' as $\dfrac{\text{Population size (N)}}{\text{Sample size (n)}}$

  and dividing the entire population in to 'K' intervals the method of drawing per one element from each and every interval to the sample is known as systematic sampling.
- Hence, after the first sample element being drawn in random from the first interval and since then every K^{th} element from each interval is added to the sample ahead.
- A complete sampling frame arranged in random is required for launching the systematic sampling method.
- When the sampling frame has been arranged as strata the similar results to stratified sampling can be expected through systematic sampling.
- When the sampling frame has been arranged as clusters the similar results to cluster sampling can be expected through systematic sampling.
- When the sampling frame has been arranged in random the similar results to simple random sampling can be expected through systematic sampling.
- Few advantages of systematic sampling are as follows :
  - This is a very simple and easy sampling method
  - The time and labour spent for drawing the sample are relatively small
  - Ability to find the standard error of the estimators

Few disadvantages of systematic sampling method.

- Sometimes a representative sample may not be derived since the entire sample depends on the first sample element
- Sample may be biased due to cyclical errors appeared in the sampling frame
- Sample cannot be derived without a complete sampling frame

Competency 6.0 : Uses appropriate sampling techniques for collecting data required to make 'Business Decisions.'

Competency level 6.3 : Uses non probabilistic sampling methods for sampling

No. of Periods : 06

Learning outcomes :

- Distinguishes between probabilistic sampling methods and non probability sampling methods.
- Names non probabilistic sampling methods.
- Introduces quota sampling.
- Draws out a sample from a given population under quota sampling method.
- States the advantages and disadvantages of quota sampling.
- Interprets judgment sampling.
- Names the situations where judgment sampling method is applicable.
- Draws out a sample from a given population under judgment sampling method.
- Lists the advantages and disadvantages of judgment sampling.
- Explains the convenient sampling.
- Draws out a sample from a given population under convenient sampling method.
- Lists advantages and disadvantages of convenient sampling method.
- Explains the purposive sampling.
- Gives examples for the situations where purposive sampling is applied.
- Lists advantages and disadvantages of purposive sampling.

Instructions for Lesson Planning :

- Present each of the following problems to the students.
  - In a sample survey with the usage of sim cards, a sample of 50 men and 50 women consists of the groups such that A/L students, individuals in the age less than 20 years, 20-30 years, 30-50 years and over 50 years is intended to be drawn.
  - A study is required to be launched regarding availablity of drinking water in a particular area, a sample of 10 water sources that are being used by the residents in the area are intended to be drawn out.
  - A large batch of readymade garments have been stored. A sample of 1 000 suits among these garments should be drawn out to check the quality level of them.

- An educationist is in need of drawing a sample of 50 Economics teachers in order to study the teaching-learning process of the teachers who are involved in teaching Economics in G.C.E. (A/L) classes.

Hold a discussion highlighting the following facts.

- Explain that application of probabilistic sampling methods in connection with selecting a sample at the above mentioned situations is difficult because of
  - Hardships in finding the sample frames
  - Extreme heterogeneity found among sample elements
  - Difficulty in drawing the sample elements
  - Requirement of a sound knowledge about the sample elements for selecting the sample
- Explain the fact that the invigilator can draw a sample on his personal desire without following probabilistic theories.
- Explain that in connection with drawing a sample for the survey regarding with SIM cards a sample of 50 men and 50 women can be selected representing each category on his/her own personal desire. This method is known as **quota sampling**.
- Highlight relative advantages and disadvantages of quota sampling compared to the probabilistic sampling methods.
- Let the students to name some other situations where quota sampling is applicable.
- Explain that for drawing a sample of 10 water sources in connection with the study of the nature of drinking water, the guidance of a specialist in the field should be followed.
- Explain that the method of drawing a sample under the guidance of a specialist in the relative field is known as **judgement sampling.**
- Let the students to state some other occasions where judgement sampling is applicable.
- List out the advantages and disadvantages in this method derived from the students.
- Discuss with the students about how to draw a sample of suits in order to check the quality level of readymade garments.
- When a sample of 1 000 suits is drawn from a large batch of readymade garments, the investigator can draw the sample elements available at easy access. The method of drawing a sample taking elements at easy access to the invigilator is known as **convenient sampling.**
- Point out that this is a non probabilistic sampling method and lead the students to list the advantages and disadvantages of this method.
- Pay the attention of students towards the study of learning-teaching process.
- Point out that the relevant 50 teachers can be selected on the way that is justified by the educationist.

- Point out that in this context, the sample is drawn by the educationist based on a special reasoning based on his knowledge and experience and that is also known as **purposive sampling** and further that the data can be collected for a detailed study.
- Point out that most probability this is a method of collecting qualitative data.
- Lead the students to list the differences between probabilistic and non probabilistic sampling methods.

A Guideline to explain the subject matters :

- The process of drawing a sample from a population as desired by the investigator with no any probability base is known as non random /non probabilistic sampling.
- When it is difficult to derive a sample in random, when a sample should be drawn immediately, when the population is dispersed in a vast geographical area and when a sample focused at the very purpose of the study should be drawn this method is applied.
- Given below are differences between probabilistic and non probabilistic sampling methods.

Probabilistic Sampling	Non- probabilistic Sampling
1. Probability techniques are applied. 2. The cost and time spent is relatively at a greater level 3. The subjectivity affected on sampling is at a lower level 4. The accuracy of conclusions arrived can be assured	1. Probabilistic techniques are not applied 2. Cost and time spent is at a lower level 3. Subjectivity is at a greater level 4. The accuracy of coming to conclusion is difficult to be assured

- Given below are non-probabilistic sampling methods.
    1. Quota sampling
    2. Convenient sampling
    3. Judgement Sampling
    4. Purposive Sampling

Quota sampling

- After the population is categorised in accordance with some characteristics, the process of drawing a determined number of sampling elements from each and every category on the personal desire of the invigilator is known as quota sampling.

- Quota sampling is applied most probably for surveys that can be completed in a short time incurring a lower cost such as attitude surveys launched for deriving openions in connection to a particular event, marketing research surveys etc.

Advantages of quota sampling

- The time and cost to be invested being minimized, since data are collected from a pre determined group; unlike in probabilistic sampling methods
- Administrative and supervising affairs being convenient
- Unlike in probabilistic sampling methods, it won't be that difficult to draw out sampling elements
- A sampling frame being not essential
- Ability to hope for a better sampling on the experiences of invigilator
- Ability to derive a more representative sample, when the population has been categorized on many more aspects

Disadvantages of quota sampling

- Not receiving a representative sample since it is subjective.
- Since a probability basis required for making inferences is not applied in this method, inability of achieving at statistical conclusions
- Difficulty in understanding the basic criteria affected for quota sampling like the social group
- Inability of evaluating the reliability of sampling outcomes

Judgment Sampling

- The method of drawing a sample with the knowledge of a specialist/consultant involved in a particular field is known as judgment sampling.
- Judgment based samples are used in researches launched in connection with specific fields like gemming or jewellery products etc. and also regarding the patients or diseases etc.

Advantages of Judgment Sampling

- Most probably the sample drawn may be fully representative with the knowledge and experience of the person who draws out the sample
- A sampling frame being not essential
- Cost time and labour being at a minimum level

Disadvantages of Judgment Sampling

- Inability to assure the reliability of each sample selected
- Inability of being applied for making statistical inferences

- Nature of the sample being changeable according to the invigilator

Convenient Sampling

The method of drawing a sample by selecting the sampling elements available at easy access in the population is known as convenient sampling.

Ex : (i) When rice is bought from the market checking a fist of rice.

(ii) Coming to a conclusion inquiring the view of few persons seated on the front row in an assembly hall

Advantages of convenient sampling :

- Ability to use for attitude surveys or marketing surveys which consists of short time objectives
- Ability to draw out a sample keeping time, cost and labour at a minimum level

Disadvantages of convenient sampling :

- Possibility of occurring sampling bias extremly
- Not being applicable for studies with long term objectives

Purposive sampling

The method of drawing a specialised persons in the field of study as sample members by the invigilator in order to collect data focusing a specific objective, is known as purposive dampling.

Advantages of purposive sampling :

- Ability to use this sampling method when qualitative data are required to be collected
- Since a small group directly relevant to the objective is selected ability of collecting data in great details can be assured
- Ability of minimizing time and labour
- Since a specific group focusing the objective is selected, most probably a representative sample being received

Disadvantages of purposive sampling :

- Inability of applying in statistical inferences, since it is a non-probabilistic sampling method
- Ability of being biased in drawing the sample