UNIT –I

Statistics:

Statistics is the science of the collection, organization, and interpretation of data. It deals with all aspects of this, including the planning of data collection in terms of the design of surveys and experiments.It is defined as, the assemblage, inspection and clarification of the numerical values. There are four methods taken for statistical investigation. This is given below,

- Collection of data
- Analysis of data
- Presentation of data
- Interpretation of data

Different Types of Statistics:

Statistics can be divided into two parts. The statistics are used to inspect the comments for all persons of a group or population and suggestion. This is said to be descriptive statistics other wise deductive statistics. Let us see about different types of statistics.

Type 1: Numerical Statistics.

Type 2: Pictorial Statistics.

Numerical statistics:

The numerical statics associated with numbers that is numerical values. Using statics we can find the mean, variance, standard deviation, etc,.

Pictorial statistics:

When the numerical information is defined in the figure of pictures and graphs, is known as pictorial statistics. This statistics provides puzzling and complicated data or information. These informations are simple, effortless and uncomplicated.

Examples for Different Types of Statistics:

Now, we will see some of the examples of the different types of statistics that is numerical and pictorial statistics.

Example:

Find the mean value of the given data 4,6,2,7,9,4.

Solution:

The given data set is 4,6,2,7,9,4.

Mean can be calculated as, **Sum of all number / Total numbers.**

$$= 4+6+2+7+9+4 / 6.$$

= 5.3 approximately 5.

Thus, the mean value is 5.

Example:

Find the mean value of the data. The data is 10,15,45,15,70.

Solution:

The given data is 10,15,45,15,70.

Mean can be calculated as, **Sum of all number / Total numbers.**

= 10+15+45+15+70 / 5.

= 31

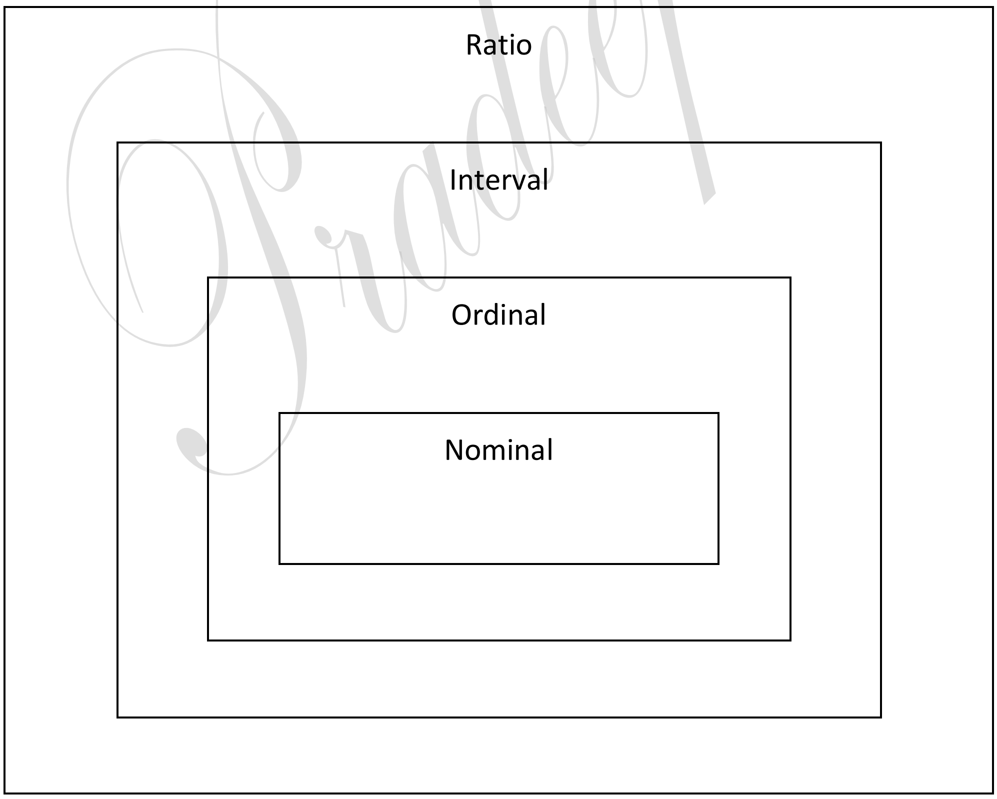
Thus, the mean value is 18.

Example :

Using the below table represents the table values using pictorial form.

Types of data:

There are four types of data that may be gathered in social research, each one adding more to the next. Thus ordinal data is also nominal, and so on.



UNIT –I

Nominal

The name 'Nominal' comes from the Latin *nomen*, meaning 'name' and nominal data are items which are differentiated by a simple naming system.

The only thing a nominal scale does is to say that items being measured have something in common, although this may not be described.

Nominal items may have numbers assigned to them. This may appear ordinal but is not -- these are used to simplify capture and referencing.

Nominal items are usually *categorical*, in that they belong to a definable category, such as 'employees'.

Example

The number pinned on a sports person.

A set of countries.

Ordinal

Items on an ordinal scale are set into some kind of *order* by their position on the scale. This may indicate such as temporal position, superiority, etc.

The order of items is often defined by assigning numbers to them to show their relative position. Letters or other sequential symbols may also be used as appropriate.

Ordinal items are usually categorical, in that they belong to a definable category, such as '1956 marathon runners'.

You cannot do arithmetic with ordinal numbers -- they show sequence only.

Example

The first, third and fifth person in a race.

Pay bands in an organization, as denoted by A, B, C and D.

Interval

Interval data (also sometimes called *integer*) is measured along a scale in which each position is equidistant from one another. This allows for the distance between two pairs to be equivalent in some way.

This is often used in psychological experiments that measure attributes along an arbitrary scale between two extremes.

Interval data cannot be multiplied or divided.

Example

My level of happiness, rated from 1 to 10.

Temperature, in degrees Fahrenheit.

Ratio

In a ratio scale, numbers can be compared as multiples of one another. Thus one person can be twice as tall as another person. Important also, the number zero has meaning. Thus the difference between a person of 35 and a person 38 is the same as the difference between people who are 12 and 15. A person can also have an age of zero.

Ratio data can be multiplied and divided because not only is the difference between 1 and 2 the same as between 3 and 4, but also that 4 is twice as much as 2.

Interval and ratio data measure quantities and hence are *quantitative*. Because they can be measured on a scale, they are also called *scale data*.

Example

A person's weight

The number of pizzas I can eat before fainting

Parametric vs. Non-parametric

Interval and ratio data are *parametric*, and are used with parametric tools in which distributions are predictable (and often Normal).

Nominal and ordinal data are *non-parametric*, and do not assume any particular distribution. They are used with non-parametric tools such as the Histogram.

Continuous and Discrete

Continuous measures are measured along a continuous scale which can be divided into fractions, such as temperature. Continuous variables allow for infinitely fine sub-division, which means if you can measure sufficiently accurately, you can compare two items and determine the difference.

Discrete variables are measured across a set of fixed values, such as age in years (not microseconds). These are commonly used on arbitrary scales, such as scoring your level of happiness, although such scales can also be continuous.

Application of Statistics in Different Fields:

Statistics plays a vital role in every fields of human activity. Statistics has important role in determining the existing position of per capita income, unemployment, population growth rate, housing, schooling medical facilities etc…in a country. Now statistics holds a central position in almost every field like Industry, Commerce, Trade, Physics, Chemistry, Economics, Mathematics, Biology, Botany, Psychology, Astronomy etc…, so application of statistics is very wide. Now we discuss some important fields in which statistics is commonly applied.

1) Business:

Statistics play an important role in business. A successful businessman must be very quick and accurate in decision making. He knows that what his customers wants, he should therefore, know what to produce and sell and in what quantities. Statistics helps businessman to plan production according to the taste of the costumers, the quality of the products can also be checked more efficiently by using statistical methods. So all the activities of the businessman based on statistical information. He can make correct decision about the location of business, marketing of the products, financial resources etc…

2) In Economics:

Statistics play an important role in economics. Economics largely depends upon statistics. National income accounts are multipurpose indicators for the economists and administrators. Statistical methods are used for preparation of these accounts. In economics research statistical methods are used for collecting and analysis the data and testing hypothesis. The relationship between supply and demands is studies by statistical methods, the imports and exports, the inflation rate, the per capita income are the problems which require good knowledge of statistics.

3) In Mathematics:

Statistical plays a central role in almost all natural and social sciences. The methods of natural sciences are most reliable but conclusions draw from them are only probable, because they are based on incomplete evidence. Statistical helps in describing these measurements more precisely. Statistics is branch of applied mathematics. The large number of statistical methods like probability averages, dispersions, estimation etc… is used in mathematics and different techniques of pure mathematics like integration, differentiation and algebra are used in statistics.

4) In Banking:

Statistics play an important role in banking. The banks make use of statistics for a number of purposes. The banks work on the principle that all the people who deposit their money with the banks do not withdraw it at the same time. The bank earns profits out of these deposits by lending to others on interest. The bankers use statistical approaches based on probability to estimate the numbers of depositors and their claims for a certain day.

5) In State Management (Administration):

Statistics is essential for a country. Different policies of the government are based on statistics. Statistical data are now widely used in taking all administrative decisions. Suppose if the government wants to revise the pay scales of employees in view of an increase in the living cost, statistical methods will be used to determine the rise in the cost of living. Preparation of federal and provincial government budgets mainly depends upon statistics because it helps in estimating the expected expenditures and revenue from different sources. So statistics are the eyes of administration of the state.

6) In Accounting and Auditing:

Accounting is impossible without exactness. But for decision making purpose, so much precision is not essential the decision may be taken on the basis of approximation, know as statistics. The correction of the values of current asserts is made on the basis of the purchasing power of money or the current value of it.

In auditing sampling techniques are commonly used. An auditor determines the sample size of the book to be audited on the basis of error.

7) <u>In Natural and Social Sciences</u>:

Statistics plays a vital role in almost all the natural and social sciences. Statistical methods are commonly used for analyzing the experiments results, testing their significance in Biology, Physics, Chemistry, Mathematics, Meteorology, Research chambers of commerce, Sociology, Business, Public Administration, Communication and Information Technology etc…

8) In Astronomy:

Astronomy is one of the oldest branch of statistical study, it deals with the measurement of distance, sizes, masses and densities of heavenly bodies by means of observations. During these measurements errors are unavoidable so most probable measurements are founded by using statistical methods.

Example: This distance of moon from the earth is measured. Since old days the astronomers have been statistical methods like method of least squares for finding the movements of stars.

UNIT –I

Frequency distribution:

A representation, either in a graphical or tabular format, which displays the number of observations within a given interval. The intervals must be mutually exclusive and exhaustive. Frequency distributions are usually used within a statistical context.
The size of the intervals used in a frequency distribution will depend on the data being analyzed and the goals of the analyst. However, the most important factor is that the intervals used must be non-overlapping and contain all of the possible observations.
For example, a frequency distribution in a tabular format for weekly stock returns may look like:

	Interval			
	<-3%	-3% to <0%	0% to <3%	>3%
Frequency	2	4	5	1

Example: Newspapers

These are the numbers of newspapers sold at a local shop over the last 10 days:

22, 20, 18, 23, 20, 25, 22, 20, 18, 20

Let us count how many of each number there is:

Papers Sold	Frequency
18	2
19	0
20	4
21	0
22	2
23	1
24	0
25	1

It is also possible to group the values. Here they are grouped in 5s:

Papers Sold	Frequency
15-19	2
20-24	7
25-29	1

UNIT –I

Diagrammatic / graphical presentations

Most people show lack of interest or have no time to go through facts and figures given in a daily newspaper or a magazine. But if these figures are graphically presented, they become easier to grasp and catch the eye and have a more lasting effect on the reader's mind.

The graphical representation of data makes the reading more interesting, less time-consuming and easily understandable. The disadvantage of graphical presentation is that it lacks details and is less accurate.

Bar chart:

A bar chart or bar graph is a chart with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.

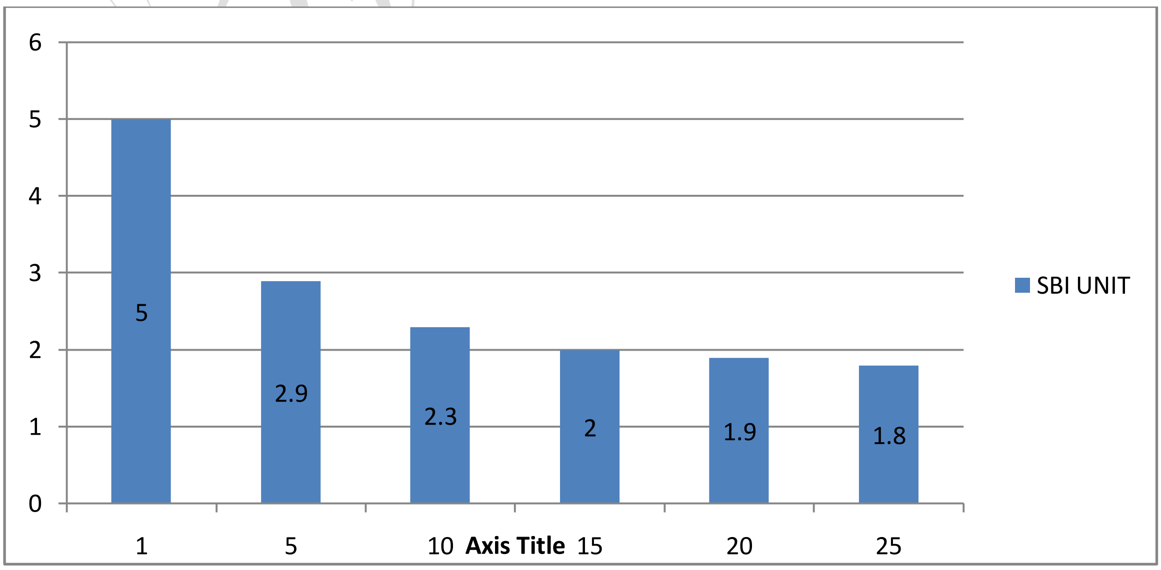
Bar charts are used for plotting discrete (or 'discontinuous') data i.e. data which has discrete values and is not continuous. Some examples of discontinuous data include 'shoe size' or 'eye colour', for which you would use a bar chart. In contrast, some examples of continuous data would be 'height' or 'weight'. A bar chart is very useful if you are trying to record certain information whether it is continuous or not continuous data.

Example:

Draw the bar chart for SBI unit regular insurance

years	1	5	10	15	20	25
SBI units	5	2.9	2.3	2	1.9	1.8

Solution:



UNIT –I

Multiple bar chart:

A multiple bar graph contains comparisons of two or more categories or bars. One axis represents a quantity and the other axis identifies a specific feature about the categories. Reading a multiple bar graph includes looking at extremes (tallest/longest vs. shortest) in each grouping.

Subdivided bar charts:

Subdivided (stacked) bar charts contain bars that are broken into segments to equal the total - the number at the end of and outside the right-most segment. These charts show comparisons between the segment categories as well as between the overall totals of each subdivided bar.

Example:

The following table gives the rate of interest for loans of various money lending institutions calculate the following data which graph to get on idea about the rate of interest for meeting the budgets.

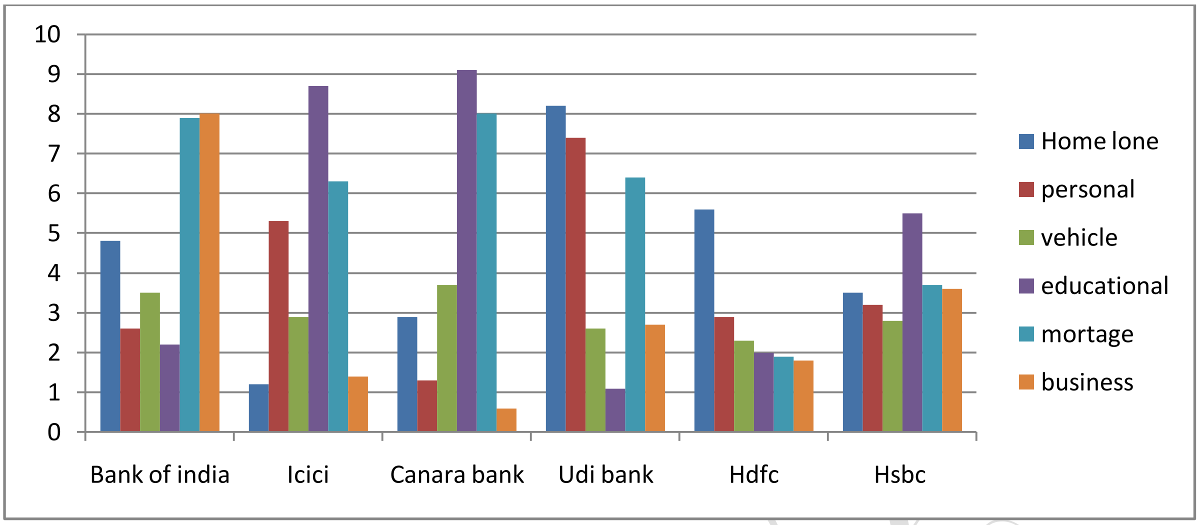
	Bank of india	Icici	Canara bank	Udi bank	Hdfc	Hsbc
Home lone	4.8	1.2	2.9	8.2	5.6	3.5
personal	2.6	5.3	1.3	7.4	2.9	3.2
vehicle	3.5	2.9	3.7	2.6	2.3	2.8
educational	2.2	8.7	9.1	1.1	2	5.5
mortage	7.9	6.3	8	6.4	1.9	3.7
business	8	1.4	0.6	2.7	1.8	3.6

Draw the multiple bar chart, subdivided bar chart.

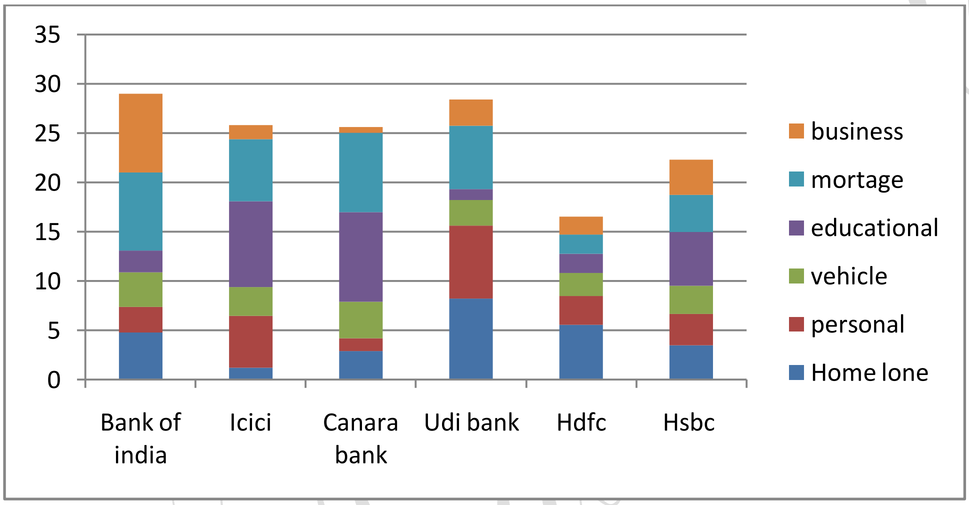
Solution:

UNIT –I

Multiple bar chart:



Subdivided bar chart:



Line graph:

A line chart or line graph is a type of graph, which displays information as a series of data points connected by straight line segments. It is a basic type of chart common in many fields. It is an extension of a scatter graph, and is created by connecting a series of points that represent individual measurements with line segments. A line chart is often used to visualize a trend in data over intervals of time – a time series – thus the line is often drawn chronologically.

Pareto chart:

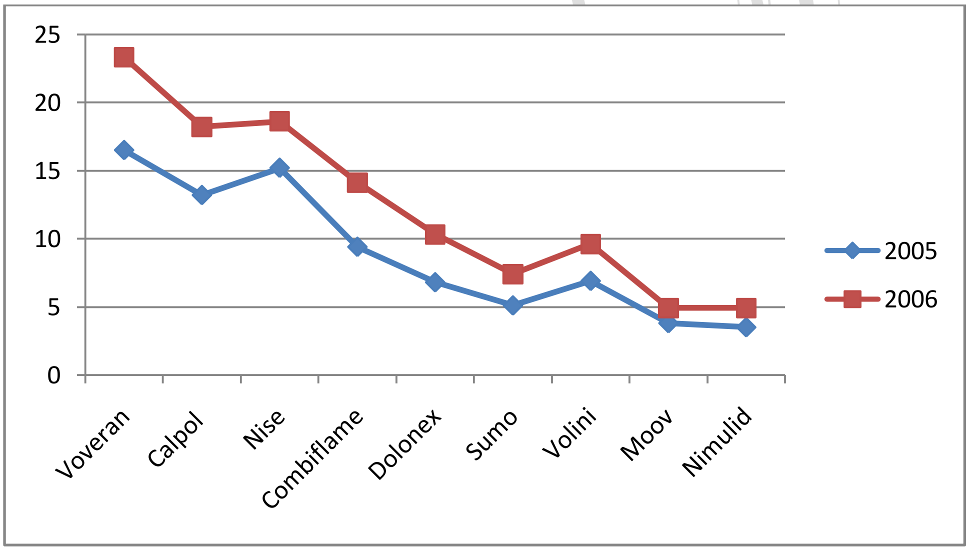
A Pareto chart, named after Vilfredo Pareto, is a type of chart that contains both bars and a line graph, where individual values are represented in descending order by bars, and the cumulative total is represented by the line.

Example:

Draw Line graph, Pareto chart for the following data giving sales of top market brands among pain killers in India. (Rs. in Crores)

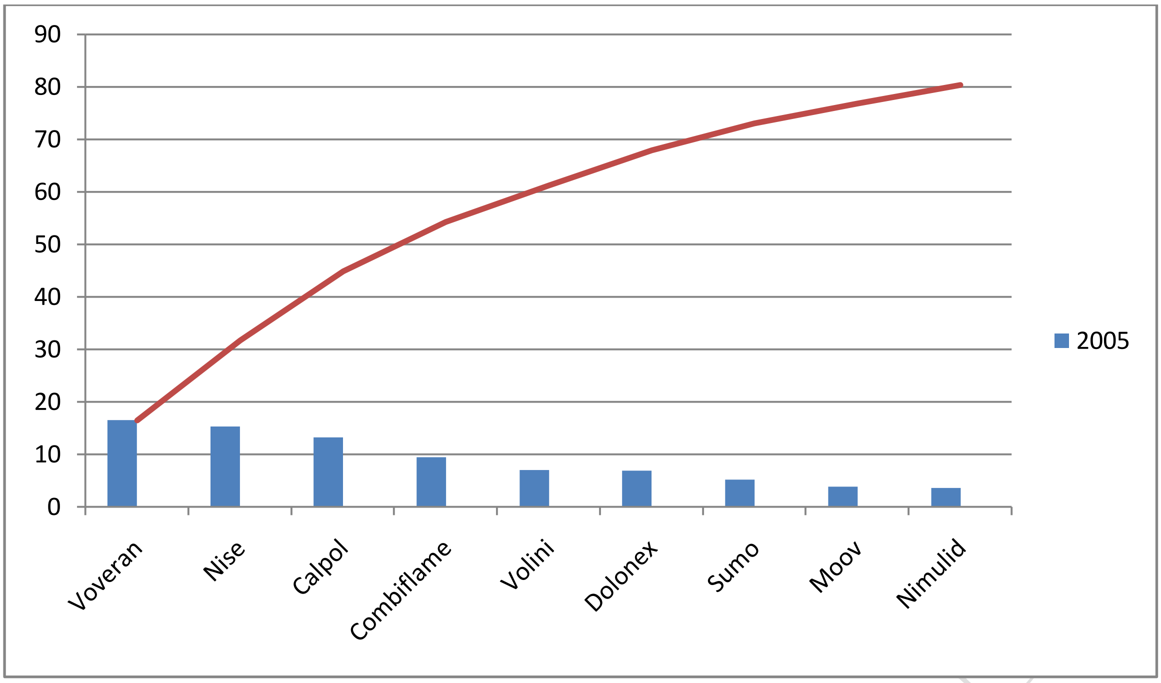
Pain killer	2005	2006
Voveran	16.5	23.3
Calpol	13.2	18.2
Nise	15.2	18.6
Combiflame	9.4	14.1
Dolonex	6.8	10.3
Sumo	5.1	7.4
Volini	6.9	9.6
Moov	3.8	4.9
Nimulid	3.5	4.9

Line graph:

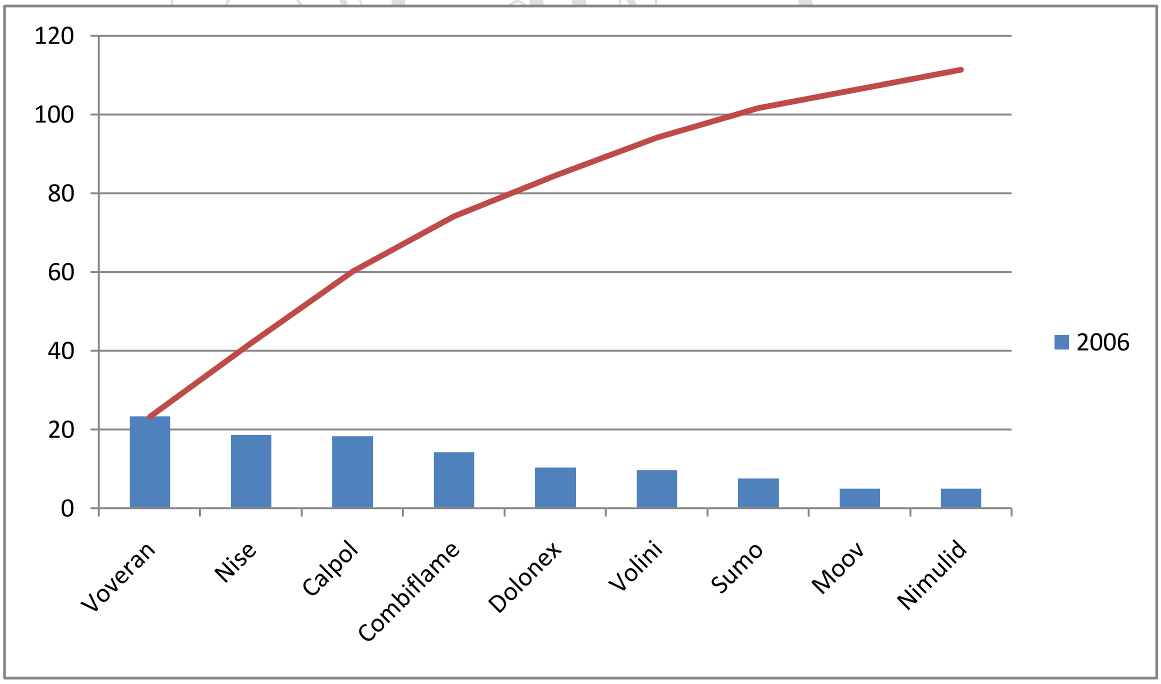


UNIT –I

Pareto chart for the year 2005:



Pareto chart for the year 2006:



UNIT –I

Pie chart:

A pie chart (or a circle graph) is a circular chart divided into sectors, illustrating proportion. In a pie chart, the arc length of each sector (and consequently its central angle and area), is proportional to the quantity it represents. When angles are measured with 1 turn as unit then a number of percent is identified with the same number of centiturns. Together, the sectors create a full disk. It is named for its resemblance to a pie which has been sliced.

Sometimes a circle is used to represent a given data. The various parts of it are proportionally represented by sectors of the circle. Then the graph is called a Pie Graph or Pie Chart.

To find the angle of each sector

Total of data corresponds to 360^o.

Let x^o = the angle at the centre for item A, then

$$x^o = \frac{\text{Value of item A}}{\text{Total value of all the items}} \times 360^o$$

The data given in example 1 can be used to draw a pie graph.

Calculation of Angles

Food:

Angle at centre $= \frac{360}{7200} \times 3000$

$= 150^o$

Rent:

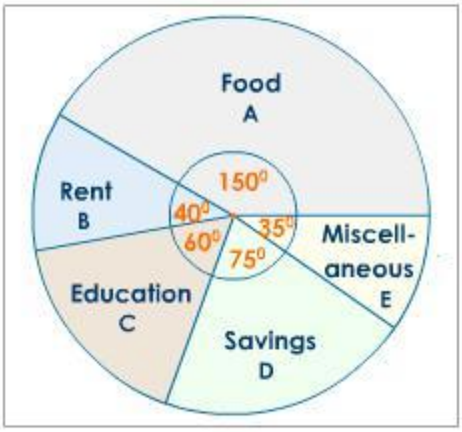
Angle at centre $= \frac{360}{7200} \times 800$

$= 40^o$

Similarly we can calculate the remaining angles, and the total of angles column should always come to 360^o.

UNIT –I

Item	Amount (Rs.)	Angle
Food (A)	3000	150^0
Rent (B)	800	40^0
Education (C)	1200	60^0
Savings (D)	1500	75^0
Miscellaneous	700	35^0
Total	**7200**	**360^0**



Percentage bar chart:

Sub-divided bar chart may be drawn on percentage basis. To draw sub-divided bar chart on percentage basis, we express each component as the percentage of its respective total. In drawing percentage bar chart, bars of length equal to 100 for each class are drawn at first step and sub-divided in the proportion of the percentage of their component in the second step. The diagram so obtained is called percentage component bar chart or percentage staked bar chart. This type of chart is useful to make comparison in components holding the difference of total constant.

Example:

The table below shows the quantity in hundred kgs of Wheat, Barley and Oats produced on a certain form during the years 1991 to 1994.

UNIT –I

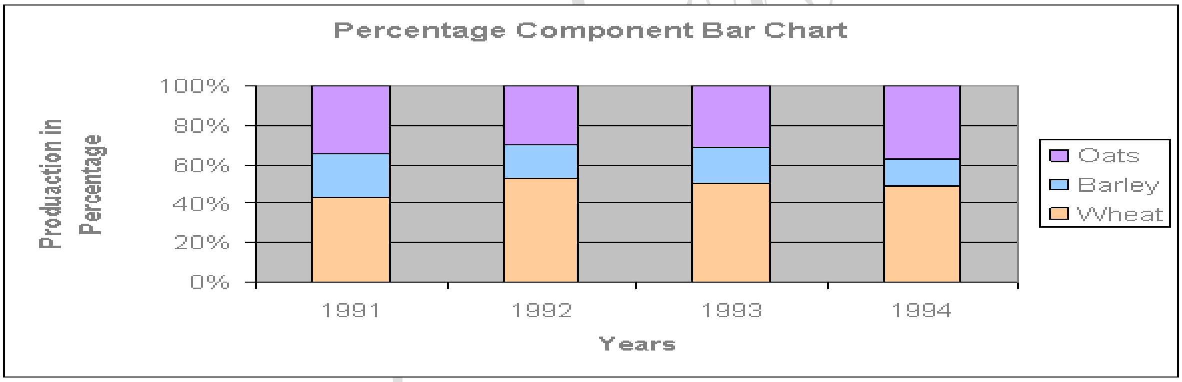
Years	Wheat	Barley	Oats
1991	34	18	27
1992	43	14	24
1993	43	16	27
1994	45	13	34

Construct a percentage component bar chart to illustrate this data.

Solution:

Item	1991		1992		1993		1994	
	%	cum %	%	cum %	%	cum %	%	cum %
Wheat	43.0	43.0	53.1	53.1	50.0	50.0	48.9	48.9
Barley	22.8	65.8	17.3	70.4	18.6	68.6	14.1	63.0
Oats	34.2	100	29.6	100	31.4	100	37.0	100
Total	100		100		100		100	

- % indicates Percentage of each item
- **cum** % indicates the cumulative percentage



Example:

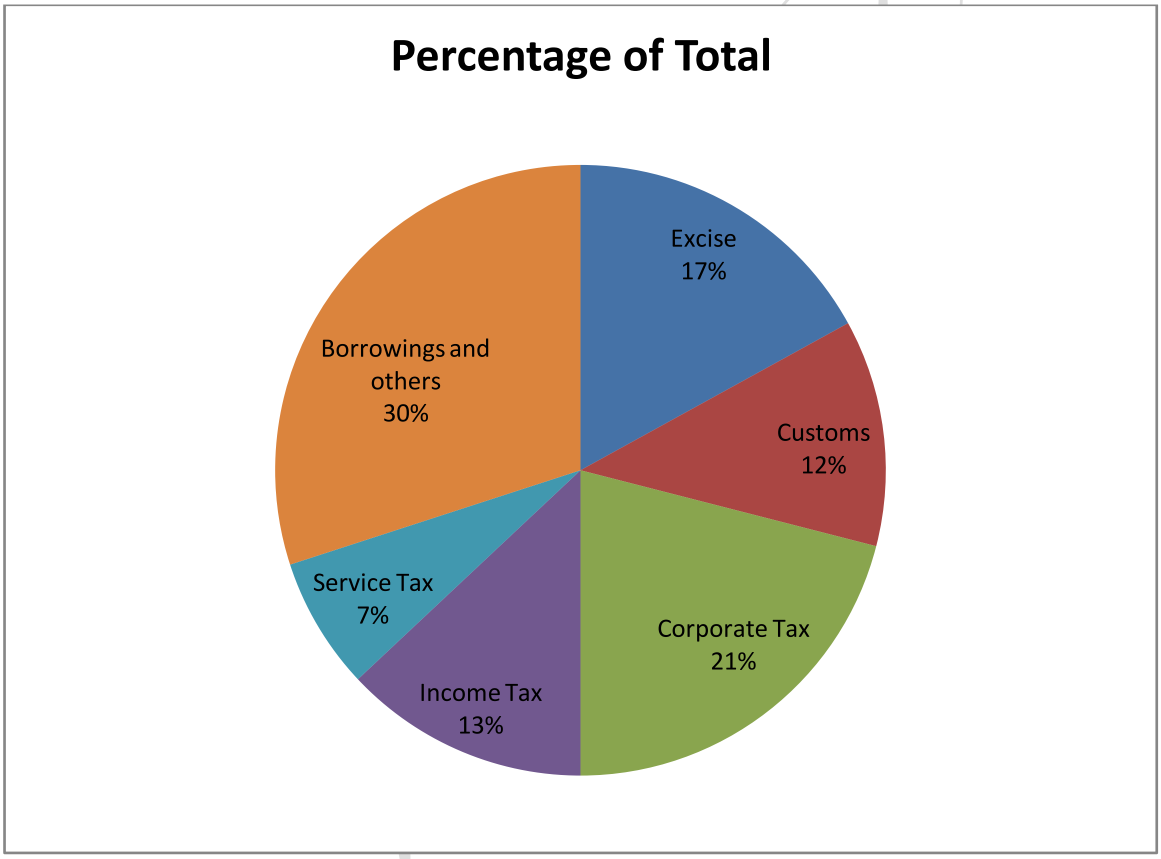
Draw Pie chart and Percentage Bar chart for the following data giving the sources of funds in Government of India's budget for the year 2007-2008.

UNIT –I

Sources of Funds	Percentage of Total
Excise	17
Customs	12
Corporate Tax	21
Income Tax	13
Service Tax	7
Borrowings and others	30

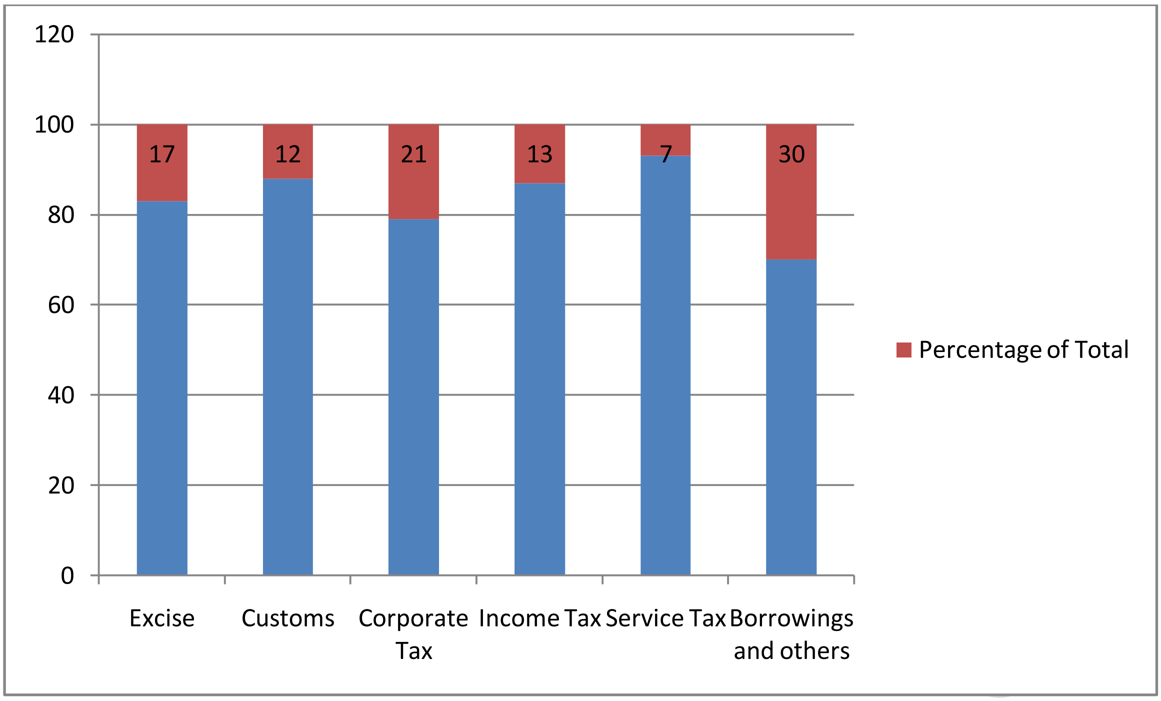
Solution:

Pie Chart



UNIT –I

Percentage Bar chart



Histogram:

A two dimensional frequency density diagram is called a histogram. A histogram is a diagram which represents the class interval and frequency in the form of a rectangle. There will be as many adjoining rectangles as there are class intervals.

To draw a histogram, follow the steps stated below

(1) Mark class intervals on X-axis and frequencies on Y-axis.

(2) The scales for both the axes need not be the same.

(3) Class intervals must be exclusive. If the intervals are in inclusive form, convert them to the exclusive form.

(4) Draw rectangles with class intervals as bases and the corresponding frequencies as heights.

The class limits are marked on the horizontal axis and the frequency is marked on the vertical axis. Thus a rectangle is constructed on each class interval.

If the intervals are equal, then the height of each rectangle is proportional to the corresponding class frequency.

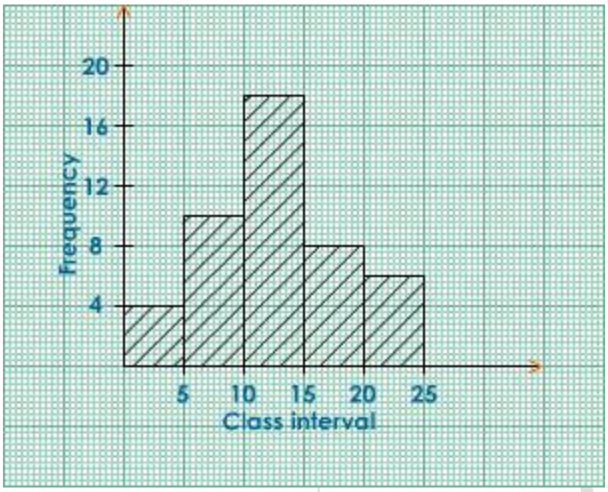
If the intervals are unequal, then the area of each rectangle is proportional to the corresponding class frequency.

Example:

Draw a histogram for the following data:

Class Interval	Frequency
0 - 5	4
5 - 10	10
10 - 15	18
15 - 20	8
20 - 25	6

Suggested answer:



Lorentz curve:

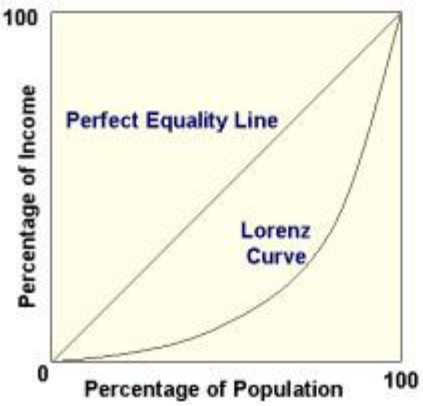
The Lorenz Curve is a graphical representation of the proportionality of a distribution. It represents a probability distribution of statistical values, and is often associated with income distribution calculations and commonly used in the analysis of inequality.

The population in the Lorenz curve is represented as households and plotted on the **x** axis from 0% to 100%. The income is plotted on the y axis and is also from 0% to 100%.

For example, a Lorenz curve can show that the bottom 50% of households bring in 35% of a country's income.

The Lorenz Curve model was developed by economist Max Lorenz in 1905.

To build the Lorenz curve, all the elements of a distribution must be ordered from the most important to the least important. Then, each element is plotted according to their cumulative percentage of X and Y, X being the cumulative percentage of elements and Y being their cumulative importance. For instance, out of a distribution of 10 elements (N), the first element would represent 10% of X and whatever percentage of Y it represents (this percentage must be the highest in the distribution). The second element would cumulatively represent 20% of X (its 10% plus the 10% of the first element) and its percentage of Y plus the percentage of Y of the first element.



Perfect Equality Line:

If income distributions were perfectly equal, 20% of households would generate 20% of income, 50% of households would generate 50% of income, 72% of households would generate 72% of income, and so on. This situation where each element has an equal value in its shares of X and Y, that is a linear relationship, is called the perfect equality line, and would be equal to the Lorenz curve.

The perfect inequality line represents a distribution where one element has the total cumulative percentage of Y while the others have none. For instance, in a distribution of 20 elements, if there is perfect equality, the 10th element would have a cumulative percentage of 50% for X and Y. The perfect equality line forms an angle of 45 degrees with a slope of 100/N.

Frequency polygon:

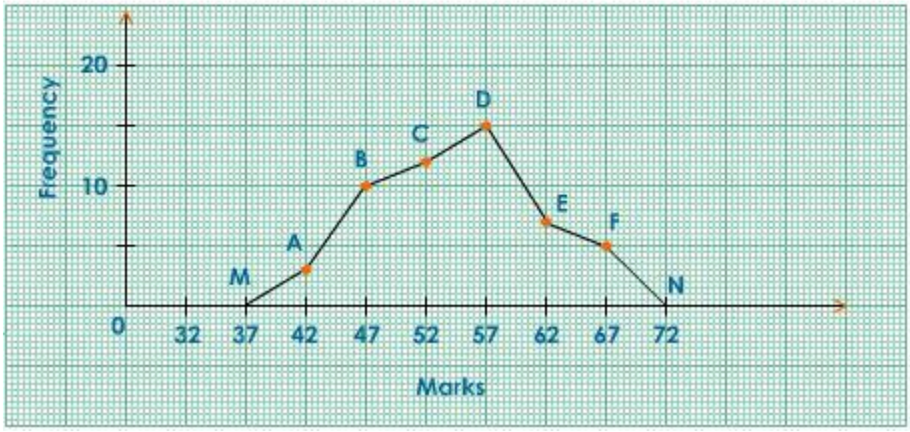
In a frequency distribution, the mid-value of each class is obtained. Then on the graph paper, the frequency is plotted against the corresponding mid-value. These points are joined by straight lines. These straight lines may be extended in both directions to meet the X - axis to form a polygon.

Example:

The weights of 50 students are recorded below. Draw a frequency polygon for this data.

Class	Mid-mark	Frequency
40 - 44	42	3
45 - 49	47	10
50 - 54	52	12
55 - 59	57	15
60 - 64	62	7
65 - 69	67	5

Solution;



Ogive:

The other name for Ogive is cumulative frequency curve.

Cumulative frequency curve or ogive is used to obtain the following information from a set of grouped data.

- median
- lower quartiles
- upper quartile
- inner quartile range

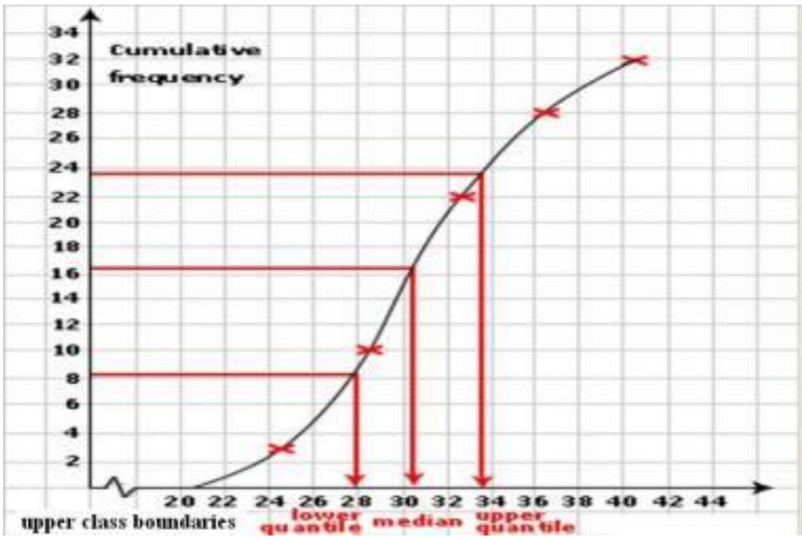
Cumulative frequency table is first step in constructing ogive.

Box and whisker plot can also constructed Using a cumulative frequency curve or ogive.

Steps to learn how to draw ogive

Ogive or cumulative frequency curve looks like this:-

UNIT –I



The cumulative frequency is the basic step in calculating ogive it is obtained by adding up the frequencies as you go along to give a 'running total'.

Before drawing the cumulative frequency curve or ogive, we need to work out on the cumulative frequencies. This is done by adding the frequencies in turn.

Let's see the sample problem.

Use all the steps of how to draw an ogive to solve the problem

Draw the Ogive curve or Cumulative frequency curve for the following data given and also find the following data

1.Median

2.Lower Quartile

3.Upper Quartile

4.Inter Quartile Range.

Age (a years	0≤a<10	10≤a<20	20≤a<30	30≤a<40	40≤a<50	50≤a<60	70≤a<80
Frequency	8	26	32	45	37	16	7

Solution:-

The initial step to be learnt of how to draw an ogive or cumulative frequency curve is calculating cumulative frequency.

To calculate cumulative frequency we have find the running sum of the frequency.

The cumulative frequency for the data set given above is

8, 34, 66, 111, 148, 177, 193, 200.

Now the two necessary things for drawing ogive or cumulative frequency curve is upper class boundary and cumulative frequency.

The following table shows the upper class boundary and cumulative frequency.

Age (a years)	**Cumulative frequency**
a < 10	8
a < 20	34
a < 30	66
a < 40	111
a < 50	148
a < 60	177
a < 70	193
a < 80	200

When we graph the above values in the coordinate plain we get the Ogive curve.

The X-axis is age and y-axis is cumulative frequency,We plot the graph as shown in the example and find the values of median , lower quartile,upper quartile and inter quartile range.

The median value is 37.

The Lower quartile is 25.

The upper quartile is 51

The interquartile range = upper quartile - lower quartile

= 51 - 25.

= 26.

MEASURES OF CENTRAL TENDENCY AND DISPERSION

Ungrouped Data (Raw Data): The information collected systematically regarding a population or a sample survey is called an ungrouped data. It is also called raw data.

Grouped Data (Classified Data): When a frequency distribution is obtained by dividing an ungrouped data in a number of strata according to the value of variate under study, such information is called grouped data or classified data.

Measures of central tendency:
There are several measures of central tendency. Out of these, the following 3 are used more often
1. Mean, 2. Median and 3. Mode

Ungrouped Data

Mean or arithmetic mean of ungrouped data:
Let $x_1, x_2, x_3, \ldots, x_n$ be n observations then mean is obtained by dividing the sum of n observations by n. It is denoted by

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

Example:
Find the mean of 4,6,8,6,7,8
Solution:

$$\bar{x} = \mu = \frac{\sum x_i}{n}$$
$$= \frac{(4 + 6 + 8 + 6 + 7 + 8)}{6}$$
$$= \frac{39}{6}$$
$$\bar{x} = 6.5$$

Geometric mean:
The average of a set of products, the calculation of which is commonly used to determine the performance results of an investment or portfolio.
The geometric mean of a data set $\{x_1, x_2, x_3, \ldots, x_n\}$ is given by:

$$GM = \{x_1\, x_2\, x_3 \ldots x_n\}^{\frac{1}{n}} \; or$$
$$GM = Antilog\left(\frac{1}{n}\sum \log x_i\right) \; or$$
$$GM = 10^g \;\; where\; g = \frac{1}{n}\sum \log x_i$$

Example:
Find the geometric mean of 4, 6, 8, 6, 7, and 8.
Solution: *Here* $\mathrm{n} = 5$

$$GM = 10^g \;\; where\; g = \frac{1}{n}\sum \log x_i$$
$$g = \frac{1}{5}\,(\log 4 + \log 6 + \log 8 + \log 6 + \log 7 + \log 8)$$
$$g = \frac{1}{5}\,(0.6021 + 0.7782 + 0.9031 + 0.7782 + 0.8451 + 0.9031)$$
$$g = \frac{1}{5}\,(4.8098) = 0.962$$
$$GM = 10^{0.962} = 9.1622$$

Harmonic mean:
The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals.

UNIT –I

The harmonic mean H of the positive real numbers $x_1, x_2, x_3, \ldots, x_n > 0$ is defined to be

$$H = \frac{n}{\sum \frac{1}{x_i}}$$

Example:

A gas-powered pump can drain a pool in 4 hours and a battery-powered pump can drain the same pool in 6 hours, then how long it will take both pumps to drain the pool together.

Solution: It is one-half of the harmonic mean of 6 and 4.

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{2}{\frac{1}{4} + \frac{1}{6}} = \frac{2 \times 6 \times 4}{6 + 4}$$

$$H = \frac{48}{10} = 4.8$$

Therefore it will take both pumps $\frac{4.8}{2} = 2.4$ hours, to drain the pool together.

Example:

Find the harmonic mean of 4, 6, 8, 6, 7, and 8.

Solution: *Here* $\mathrm{n} = 5$

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{5}{\frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}} = \frac{5}{0.25 + 0.167 + 0.125 + 0.167 + 0.143 + 0.125}$$

$$H = \frac{5}{0.977} = 5.118$$

Relation between arithmetic, geometric and harmonic means:

i) $\text{Harmonic mean} \leq \text{Geometric mean} \leq \text{Arithmetic mean}$

ii) $\text{Geometric mean} = \sqrt{\text{Arithmetic mean} \times \text{Harmonic mean}}$

Weighted mean:

The weighted mean of a set of numbers $x_1, x_2, x_3, \ldots, x_n$ with corresponding weights $w_1, w_2, w_3, \ldots, w_n$ is computed from the following formula:

$$Weighted\ mean\ WM = \frac{\sum w_i x_i}{\sum w_i}$$

Example:

During a one hour period on a hot Saturday afternoon cabana boy Chris served fifty drinks. He sold five drinks for \$0.50, fifteen for \$0.75, fifteen for \$0.90, and fifteen for \$1.10. Compute the weighted mean of the price of the drinks.

Solution:

$$Weighted\ mean\ WM = \frac{\sum w_i x_i}{\sum w_i} = \frac{5 \times 0.50 + 15 \times 0.75 + 15 \times 0.90 + 15 \times 1.10}{5 + 15 + 15 + 15}$$

$$WM = \frac{44.34}{50}$$

$$WM = 0.89$$

Median of ungrouped data:

If the observations of an ungrouped data are arranged in increasing or decreasing order of their magnitude, a value which divides these ordered observations into two equal parts is called the median of the data. It is denoted by M.

If the number of observations (n) is an **odd** integer, then

$$M = Value\ of\ \frac{(n+1)}{2}^{th}\ observation$$

If the number of observations (n) is an **even** integer, then

$$M = \frac{(Value\ of\ \frac{n}{2}^{th}\ observation + Value\ of\ \frac{(n+1)}{2}^{th}\ observation)}{2}$$

Example:
Find the median of the following observations 4,6,8,6,7,8,8

Solution: Observations in the ascending order are :
4, 6, 6, 7, 8, 8, 8
Here, $n = 7$ is odd.
Median :

$M = Value\ of\ \frac{(n+1)}{2}^{th}\ observation$
$M = Value\ of\ \frac{(7+1)}{2}^{th}\ observation$
$M = Value\ of\ 4^{th}\ observation$
$M = 7$

Note: The **median** divides the data into a lower half and an upper half.

Mode of ungrouped data:
An observation occurring most frequently in the data is called mode of the data. It is denoted by Z.

Example:
Find the median of the following observations
4,6,8,6,7,8,8
Solution: In the given data, the observation 8 occurs maximum number of times (3)
$Mode(Z) = 8$

Relation between Mean, Median and Mode:

$$Mean - Mode = 3(Mean - Median)$$

Range of ungrouped data:
The **range** of a set of data is the difference between the highest and lowest values in the set.

Example:
Cheryl took 7 math tests in one marking period. What is the range of her test scores?
89, 73, 84, 91, 87, 77, 94

Solution: Ordering the test scores from least to greatest, we get: 73, 77, 84, 87, 89, 91, and 94

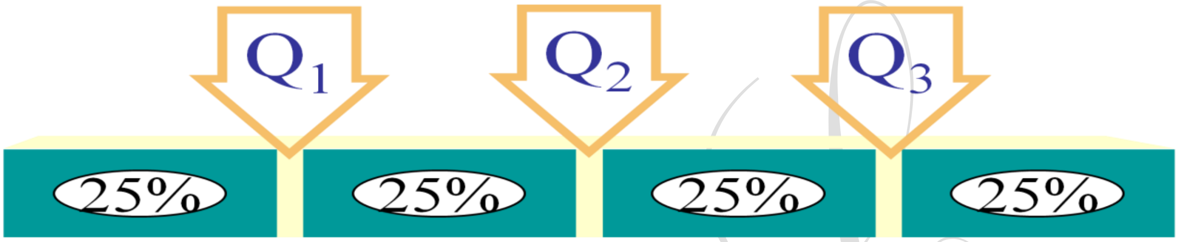
Highest - lowest = 94 - 73 = 21

Therefore the range of these test scores is 21 points.

Quartile of ungrouped data:

Quartile - measures of central tendency that divide a group of data into four subgroups

- Q_1: 25% of the data set is below the first quartile
- Q_2: 50% of the data set is below the second quartile
- Q_3: 75% of the data set is below the third quartile



The **lower quartile** or first quartile is the middle value of the lower half.

The **upper quartile or third quartile** is the middle value of the upper half.

Example:

Find the median, lower quartile and upper quartile of the following numbers.

12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25

Solution:

First, arrange the data in ascending order:

5, 7, 12, 14, 15, 22, 25, 30, 36, 42, 53

Median (middle value) = 22

Lower quartile (middle value of the lower half) = 12

Upper quartile (middle value of the upper half) = 36

If there is an even number of data items, then we need to get the average of the middle numbers.

Interquartile range:

Interquartile range = Upper quartile – lower quartile

Example:

Find the median, lower quartile, upper quartile, interquartile range and range of the following numbers.

12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25, 65

Solution:

First, arrange the data in ascending order:

5, 7, 12, 14, 15, 22, 25, 30, 36, 42, 53, 65

Lower quartile or **first quartile** = $\frac{12+14}{2} = 13$

Median or **second quartile** = $\frac{22+25}{2} = 23.5$

Upper quartile or **third quartile** = $\frac{36+42}{2}$ = 39

Interquartile range = Upper quartile – lower quartile
= 39 – 13 = 26

Range = largest value – smallest value
= 65 – 5 = 60

Variance of an ungrouped data:

Variance is the average of the squared deviations from the arithmetic mean.
The variance of a set of values, which we denote by σ^2, is defined as

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

Where $\bar{x}$ is the mean, n is the number of data values, and x_i stands for data value in ith position.

An alternative, yet equivalent formula, which is often easier to use is

$$\sigma^2 = \frac{\sum(x_i)^2}{n} - \bar{x}^2$$

Example:

Find the variance of 6, 7, 10, 11, 11, 13, 16, 18, and 25.

Solution: Here $n = 9$

Firstly we find the mean,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6+7+\cdots+25}{9} = \frac{117}{9} = 13$$

Method 1:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

It is helpful to show the calculation in a table:

x_i	6	7	10	11	11	13	16	18	25	Total
$x_i - \bar{x}$	-7	-6	-3	-2	-2	0	3	5	12	
$(x_i - \bar{x})^2$	49	36	9	4	4	0	9	25	144	280

$$\sigma^2 = \frac{280}{9} = 31.11$$

Method 2:

$$\sigma^2 = \frac{\sum(x_i)^2}{n} - \bar{x}^2$$

x_i	6	7	10	11	11	13	16	18	25	Total
x_i^2	36	49	100	121	121	169	256	324	625	1801

$$= \frac{\sum(x_i)^2}{n} - \bar{x}^2 = \frac{1801}{9} - 13^2$$

$$= 200.11 - 169 = 31.11$$

Standard Deviation (σ)

Since the variance is measured in terms of x_i^2, we often wish to use the standard deviation where

$$\sigma = \sqrt{Variance}$$

The standard deviation, unlike the variance, will be measured in the same units as the original data.

$$\sigma = \sqrt{31.11} = 5.58$$

Coefficient of Variation:

Coefficient of Variation (CV) is the ratio of the standard deviation to the mean, expressed as a percentage.

$$CV = \frac{\sigma}{\mu} \times 100$$

Example:

Find the Coefficient of Variation of 6, 7, 10, 11, 11, 13, 16, 18, and 25.

Solution: Here $n = 9$

Firstly we find the mean,

$$\bar{x} = \mu = \frac{\sum x_i}{n} = \frac{6 + 7 + \cdots + 25}{9} = \frac{117}{9} = 13$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

It is helpful to show the calculation in a table:

x_i	6	7	10	11	11	13	16	18	25	Total
$x_i - \bar{x}$	-7	-6	-3	-2	-2	0	3	5	12	
$(x_i - \bar{x})^2$	49	36	9	4	4	0	9	25	144	280

$$\sigma^2 = \frac{280}{9} = 31.11$$

$$CV = \frac{\sigma}{\mu} \times 100 = \frac{\sqrt{31.11}}{13} \times 100$$

$$CV = \frac{5.58}{13} \times 100 = 42.92$$

Mean Absolute deviation:

Mean Absolute Deviation is the average of the absolute deviations from the mean.

$$M.A.D = \frac{\sum |x_i - \bar{x}|}{n}$$

Example:

Find the Mean Absolute Deviation of 6, 7, 10, 11, 11, 13, 16, 18, and 25.

Solution: Here $n = 9$

Firstly we find the mean,

$$\bar{x} = \mu = \frac{\sum x_i}{n} = \frac{6 + 7 + \cdots + 25}{9} = \frac{117}{9} = 13$$

It is helpful to show the calculation in a table:

UNIT –I

x_i	6	7	10	11	11	13	16	18	25	Total
$\|x_i - \bar{x}\|$	7	6	3	2	2	0	3	5	12	40

$$M.A.D = \frac{\sum |x_i - \bar{x}|}{n} = \frac{40}{9} = 4.44$$

Grouped Data:

Mean:

For the grouped data the class interval and frequency is given we have to find the midpoint of class in terval and find the mean using the formula

$$\text{Mean } \bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Where x_i is the midpoint of the ith class interval,
f_i is the frequency of the ith class interval.

Median:

The median for the grouped data is also defined as the value corresponding to the $\frac{n+1}{2}^{th}$ observation, and is calculated from the formula:

$$\text{Median} = L_m + \frac{\left(\frac{n}{2} - f_c\right)}{f_m} \times W_m$$

Where L_m is the lower limit of the median class interval i.e. the interval which contains $\frac{n}{2}^{th}$ observation,
f_m is the frequency of the median class interval,
f_c is the cumulative frequency up to the interval just before the median class interval,
W_m is the width of the median class interval, and n is the number of total observations i.e, $n = \sum f_i$.

Mode:

In a grouped data, the mode is calculated by the following formula

$$\text{Mode} = L_m + \frac{f_m - f_0}{2f_m - f_0 - f_2} \times W_m$$

Where L_m is the lower limit of the modal class interval i.e. the class interval having highest frequency,
f_m is the frequency of the modal class interval,
f_0 is the frequency of to the interval just before the modal class interval,
f_2 is the frequency of to the interval just after the modal class interval
W_m is the width of the modal class interval

First Quartile:

$$\text{First Quartile } Q_1 = L_{Q_1} + \frac{\left(\frac{n}{4} - f_c\right)}{f_{Q_1}} \times W_{Q_1}$$

Where L_{Q_1} is the lower limit of the first quartile class interval i.e. the interval which contains $\frac{n}{4}^{th}$ observation,

f_{Q_1} is the frequency of the first quartile class interval,
f_c is the cumulative frequency up to the interval just before the first quartile class interval,
W_{Q_1} is the width of the first quartile class interval, and n is the number of total observations i.e, $n = \sum f_i$.

Third Quartile:

$$\text{Third Quartile } Q_3 = L_{Q_3} + \frac{\left(\frac{3n}{4} - f_c\right)}{f_{Q_3}} \times W_{Q_3}$$

Where L_{Q_3} is the lower limit of the third quartile class interval i.e. the interval which contains $\frac{3n}{4}^{th}$ observation,
f_{Q_3} is the frequency of the third quartile class interval,
f_c is the cumulative frequency up to the interval just before the first quartile class interval,
W_{Q_3} is the width of the third quartile class interval, and n is the number of total observations i.e, $n = \sum f_i$.

Quartile Deviation or Semi Inter-Quartile Range:

$$Quartile\ Deviation = \frac{Q_3 - Q_1}{2}$$

Percentiles:

Percentiles which spilt the data into several parts, expressed in percentages. A Percentile also known as centile, divides the data in such a way that “given percent of the observations are less than it”.

$$P_m = L_{P_m} + \frac{\left(\frac{mn}{100} - f_c\right)}{f_{P_m}} \times W_{P_m}$$

WhereL_{P_m} is the lower limit of the percentile class interval i.e. the interval which contains $\frac{mn}{100}^{th}$ observation,
f_{P_m} is the frequency of the percentile class interval,
f_c is the cumulative frequency up to the interval just before the first quartile class interval,
W_{P_m} is the width of the percentile class interval, and n is the number of total observations i.e, $n = \sum f_i$.

Deciles:

The deciles divide the data into ten parts- first decile (10%), second (20%) and so on**.**

$$\text{Decile } D_m = L_{D_m} + \frac{\left(\frac{mn}{10} - f_c\right)}{f_{D_m}} \times W_{D_m}$$

WhereL_{D_m} is the lower limit of the percentile class interval i.e. the interval which contains $\frac{mn}{10}^{th}$ observation,
f_{D_m} is the frequency of the percentile class interval,
f_c is the cumulative frequency up to the interval just before the first quartile class interval,
W_{D_m} is the width of the percentile class interval, and n is the number of total observations i.e, $n = \sum f_i$

UNIT –I

Mean Absolute Deviation:

$$M.A.D = \frac{\sum(f_i|x_i - \bar{x}|)}{\sum f_i}$$

Where x_i is the midpoint of the i^{th} class interval,
f_i is the frequency of the i^{th} class interval.
$\bar{x}$ is the mean of the given data

Variance:

$$\sigma^2 = \frac{\sum f_i x_i^2 - (\sum f_i)\,\bar{x}^2}{\sum f_i}$$

Where x_i is the midpoint of the i^{th} class interval,
f_i is the frequency of the i^{th} class interval.
$\bar{x}$ is the mean of the given data

Example:

Find Mean, Variance, standard deviation and Coefficient of variation for the following data given below:

Class Interval:	**2000-3000**	**3000-4000**	**4000-5000**	**5000-6000**	**6000-7000**
Frequency:	**2**	**5**	**6**	**4**	**3**

Solution:

Class Interval	Frequency f_i	Midpoint of Class Interval x_i	x_i^2	$f_i x_i$	$f_i x_i^2$
2000-3000	2	2500	6250000	5000	12500000
3000-4000	5	3500	12250000	17500	61250000
4000-5000	6	4500	20250000	27000	121500000
5000-6000	4	5500	30250000	22000	121000000
6000-7000	3	6500	42250000	19500	126750000
SUMS	**20**			**91000**	**443000000**

Mean $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{91000}{20} = 4550$

Variance $\sigma^2 = \frac{\sum f_i x_i^2 - (\sum f_i)\,\bar{x}^2}{\sum f_i} = \frac{443000000 - 20(4550)^2}{20} = 1447500$

Standard deviation $\sigma = \sqrt{1447500} = 1203.12$

Coefficient of Variation $C.V = \frac{\sigma}{\bar{x}} \times 100 = \frac{1203.12}{4550} \times 100 = 26.44\ \%$

Example:

UNIT –I

The distribution of Intelligence Quotient (I.Q.) scores measured for 100 students in a test is as follows

I.Q. :	**40-50**	**50-60**	**60-70**	**70-80**	**80-90**	**90-100**
Number of students :	**10**	**20**	**20**	**15**	**15**	**20**

Find Mean, Median, Mode, first Quartile, third Quartile, 60th Percentile, 9th Decile, and Mean deviation.

Solution:

Class Interval	**Frequency** f_i	**Midpoint of Class Interval** x_i	$f_i x_i$	**Cumulative Frequency** f_c	$\lvert x_i - \bar{x} \rvert = \lvert x_i - 71.5 \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
40-50	10	45	450	10	26.5	265
50-60	20**	55	1100	30***	16.5	330
60-70	20	65	1300	50*	6.5	130
70-80	15	75	1125	65#	3.5	52.5
80-90	15	85	1275	80****	13.5	202.5
90-100	20	95	1900	100 ##	23.5	470
SUMS	**100**		**7150**			**1450**

To find Mean:

$$\text{Mean } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{7150}{100} = 71.5$$

To find Median:

$$n = \sum f_i = 100$$

$$\frac{n}{2} = \frac{100}{2} = 50$$

The Median class interval * is 60-70 $(\,Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{n}{2}^{th}\ f_c\ column)$

$$\text{Median } = L_m + \frac{\left(\frac{n}{2} - f_c\right)}{f_m} \times W_m = 60 + \frac{\left(\frac{100}{2} - 30\right)}{20} \times (70 - 60)$$

$$= 60 + \frac{(50 - 30)}{20} \times 10$$

$$\text{Median} = 70$$

To find Mode:

The Mode class interval ** is 50-60 $(\,Corresponding\ class\ interval\ of\ highest\ value\ in\ f_i\ column)$

$$\text{Mode} = L_m + \frac{f_m - f_0}{2f_m - f_0 - f_2} \times W_m$$
$$= 50 + \frac{20 - 10}{2(20) - 10 - 20} \times (60 - 50)$$
$$= 50 + \frac{10}{10} \times 10$$
$$\text{Mode} = 60.$$

To find First Quartile:

$$n = \sum f_i = 100$$
$$\frac{n}{4} = \frac{100}{4} = 25$$

The First Quartile class interval *** is 50-60 ($Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{n}{4}^{th}\ f_c\ column$)

$$\text{First Quartile } Q_1 = L_{Q_1} + \frac{\left(\frac{n}{4} - f_c\right)}{f_{Q_1}} \times W_{Q_1} = 50 + \frac{\left(\frac{100}{4} - 10\right)}{20} \times (60 - 50)$$

$$= 50 + \frac{(25 - 10)}{20} \times 10$$
$$\text{First Quartile } Q_1 = 57.5$$

To find Third Quartile:

$$n = \sum f_i = 100$$
$$\frac{3n}{4} = \frac{3(100)}{4} = 75$$

The Third Quartile class interval **** is 80-90 ($Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{3n}{4}^{th}\ f_c\ column$)

$$\text{Third Quartile } Q_3 = L_{Q_3} + \frac{\left(\frac{3n}{4} - f_c\right)}{f_{Q_3}} \times W_{Q_3} = 80 + \frac{\left(\frac{3(100)}{4} - 65\right)}{15} \times (90 - 80)$$

$$= 80 + \frac{(75 - 65)}{15} \times 10$$
$$\text{Third Quartile } Q_3 = 86.67$$

To find Percentile :

$$P_m = L_{P_m} + \frac{\left(\frac{mn}{100} - f_c\right)}{f_{P_m}} \times W_{P_m}$$
$$\frac{mn}{100} = \frac{60 \times 100}{100} = 60$$

The 60th Percentile class interval # is 70 - 80

$$(\,Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{mn^{th}}{100}\ f_c\ column)$$

$$60^{th}\text{ Percentile } P_{60} = 70 + \frac{(60-50)}{15} \times (80-70)$$

$$60^{th}\text{ Percentile } P_{60} = 76.67$$

To find Decile:

$$\text{Decile } D_m = L_{D_m} + \frac{\left(\frac{mn}{10} - f_c\right)}{f_{D_m}} \times W_{D_m}$$

$$\frac{mn}{10} = \frac{9 \times 100}{10} = 90$$

The 9th Decile class interval ## is 90 - 100

$$(\,Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{mn^{th}}{10}\ f_c\ column)$$

$$9^{th}\text{ Decile } D_9 = 90 + \frac{(90-80)}{20} \times (100-90)$$

$$9^{th}\text{ Decile } D_9 = 95$$

To find Mean deviation:

$$\text{Mean deviation} = \frac{\sum f_i|x_i - \bar{x}|}{\sum f_i} = \frac{1450}{100} = 14.5$$

Example:

Calculate the mean, median, mode, 70th percentile and 4th decile from the following data:

Class-interval:	**0-4**	**4-8**	**8-12**	**12-16**	**16-20**	**20-24**	**24-28**
Frequency :	**10**	**12**	**18**	**7**	**5**	**8**	**4**

Solution:

Class Interval	**Frequency f_i**	**Midpoint of Class Interval x_i**	**$f_i x_i$**	**Cumulative Frequency f_c**
0 - 4	10	2	20	10
4 - 8	12	6	72	22
8 - 12	18**	10	180	40*##
12 - 16	7	14	98	47#
16 - 20	5	18	90	52
20 - 24	8	22	176	60
24 - 28	4	26	104	64
Sums	**64**		**740**	

To find Mean:

$$\text{Mean } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{740}{64} = 11.56$$

To find Median:

$$n = \sum f_i = 64$$

$$\frac{n}{2} = \frac{64}{2} = 32$$

The Median class interval * is 8 - 12
$(\, Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{n}{2}^{th}\ f_c\ column)$

$$\text{Median} = L_m + \frac{\left(\frac{n}{2} - f_c\right)}{f_m} \times W_m = 8 + \frac{(32 - 22)}{18} \times (12 - 8)$$

$$= 8 + \frac{10}{18} \times 4$$

$$\text{Median} = 10.22$$

To find Mode:
The Mode class interval ** is 8 - 12
$(\, Corresponding\ class\ interval\ of\ highest\ value\ in\ f_i\ column)$

$$\text{Mode} = L_m + \frac{f_m - f_0}{2f_m - f_0 - f_2} \times W_m$$

$$= 8 + \frac{18 - 12}{2(18) - 12 - 7} \times (12 - 8)$$

$$= 8 + \frac{6}{17} \times 4$$

$$\text{Mode} = 9.41$$

To find Percentile :

$$P_m = L_{P_m} + \frac{\left(\frac{mn}{100} - f_c\right)}{f_{P_m}} \times W_{P_m}$$

$$\frac{mn}{100} = \frac{70 \times 64}{100} = 44.8$$

The 70th Percentile class interval # is 12 - 16

$(\, Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{mn}{100}^{th}\ f_c\ column)$

$$70^{th}\ \text{Percentile}\ P_{70} = 12 + \frac{(44.8 - 40)}{7} \times (16 - 12)$$

$$70^{th}\ \text{Percentile}\ P_{70} = 14.74$$

To find Decile:

$$\text{Decile}\ D_m = L_{D_m} + \frac{\left(\frac{mn}{10} - f_c\right)}{f_{D_m}} \times W_{D_m}$$

$$\frac{mn}{10} = \frac{4 \times 64}{10} = 25.6$$

The 4th Decile class interval ## is 8 - 12

$(\, Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{mn}{10}^{th}\ f_c\ column)$

$$4^{th}\ \text{Decile}\ D_4 = 8 + \frac{(25.6 - 22)}{18} \times (12 - 8)$$

4th Decile $D_4 = 8.8$

Example:

Calculate mean, median, mode, first quartile, third quartile,45th percentile and 9th decile, variance and Coefficient of variation from the following data relating to production of a steel mill on 60 days.

Production (in Tons per day) :	**21-22**	**23-24**	**25-26**	**27-28**	**29-30**
Number of days :	**7**	**13**	**22**	**10**	**8**

Solution: Since the data is not continuous in production we make the data continuous by dividing the width of the interval by 2 and subtracting it from lower value or left value of the interval and adding it in higher value or right value of the interval. Here the width is 1 dividing it by 2 we get 0.5 so subtract 0.5 from lower value or left value of the interval and add 0.5 to higher value or right value of the interval.

Class interval	Frequency f_i	Midpoint of Class Interval x_i	$f_i x_i$	x_i^2	$f_i x_i^2$	Cumulative Frequency f_c
20.5- 22.5	7	21.5	150.5	462.25	3235.75	7
22.5-24.5	13	23.5	305.5	552.25	7179.25	20***
24.5-26.5	22**	25.5	561	650.25	14305.5	42*#
26.5-28.5	10	27.5	275	756.25	7562.5	52****
28.5-30.5	8	29.5	236	870.25	6962	60##
Sum	60	127.5	1528	3291.25	39245	

To find Mean:

$$\text{Mean } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1528}{60} = 25.467$$

To find Median:

$$n = \sum f_i = 60$$

$$\frac{n}{2} = \frac{60}{2} = 30$$

The Median class interval * is 24.5 - 26.5 $(\mathit{Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ } \frac{n^{th}}{2}\ f_c \mathit{\ column})$

$$\text{Median} = L_m + \frac{\left(\frac{n}{2} - f_c\right)}{f_m} \times W_m = 24.5 + \frac{\left(\frac{60}{2} - 20\right)}{22} \times (26.5 - 24.5)$$

$$= 24.5 + \frac{(30 - 20)}{22} \times 2$$

$$\text{Median} = 25.41$$

To find Mode:

The Mode class interval ** is 24.5 - 26.5 $(\mathit{Corresponding\ class\ interval\ of\ highest\ value\ in\ } f_i \mathit{\ column})$

$$\text{Mode} = L_m + \frac{f_m - f_0}{2f_m - f_0 - f_2} \times W_m$$

$$= 24.5 + \frac{22 - 13}{2(22) - 13 - 10} \times (26.5 - 24.5)$$

$$= 24.5 + \frac{9}{21} \times 2$$

$$\text{Mode} = 25.36$$

To find First Quartile:

$$n = \sum f_i = 100$$

$$\frac{n}{4} = \frac{60}{4} = 15$$

The First Quartile class interval *** is 22.5 - 24.5

(*Corresponding class interval of the nearest greater than or equal to* $\frac{n}{4}^{th}$ f_c *column*)

$$\text{First Quartile } Q_1 = L_{Q_1} + \frac{\left(\frac{n}{4} - f_c\right)}{f_{Q_1}} \times W_{Q_1} = 22.5 + \frac{\left(\frac{60}{4} - 7\right)}{13} \times (24.5 - 22.5)$$

$$= 22.5 + \frac{(15 - 7)}{13} \times 2$$

$$\text{First Quartile } Q_1 = 23.73$$

To find Third Quartile:

$$n = \sum f_i = 60$$

$$\frac{3n}{4} = \frac{3(60)}{4} = 45$$

The Third Quartile class interval **** is 26.5 - 28.5

(*Corresponding class interval of the nearest greater than or equal to* $\frac{3n}{4}^{th}$ f_c *column*)

$$\text{Third Quartile } Q_3 = L_{Q_3} + \frac{\left(\frac{3n}{4} - f_c\right)}{f_{Q_3}} \times W_{Q_3} = 26.5 + \frac{\left(\frac{3(60)}{4} - 42\right)}{10} \times (28.5 - 26.5)$$

$$= 26.5 + \frac{(45 - 42)}{10} \times 2$$

$$\text{Third Quartile } Q_3 = 27.1$$

To find Percentile :

$$P_m = L_{P_m} + \frac{\left(\frac{mn}{100} - f_c\right)}{f_{P_m}} \times W_{P_m}$$

$$\frac{mn}{100} = \frac{45 \times 60}{100} = 27$$

The 45^{th} Percentile class interval # is 24.5 - 26.5

(*Corresponding class interval of the nearest greater than or equal to* $\frac{mn}{100}^{th}$ f_c *column*)

$$45^{th} \text{ Percentile } P_{45} = 24.5 + \frac{(27 - 20)}{22} \times (26.5 - 24.5)$$

$$45^{th} \text{ Percentile } P_{45} = 25.14$$

UNIT –I

To find Decile:

$$\text{Decile } D_m = L_{D_m} + \frac{\left(\frac{mn}{10} - f_c\right)}{f_{D_m}} \times W_{D_m}$$

$$\frac{mn}{10} = \frac{9 \times 60}{10} = 54$$

The 9th Decile class interval ## is 28.5 – 30.5

$$(\,Corresponding\ class\ interval\ of\ the\ nearest\ greater\ than\ or\ equal\ to\ \frac{mn^{th}}{10}\ f_c\ column)$$

$$9^{\text{th}} \text{ Decile } D_9 = 28.5 + \frac{(54-52)}{8} \times (30.5 - 28.5)$$

$$9^{\text{th}} \text{ Decile } D_9 = 29$$

$$\text{Variance } \sigma^2 = \frac{\sum f_i x_i^2 - (\sum f_i)\,\bar{x}^2}{\sum f_i} = \frac{39245 - 60(25.467)^2}{60} = 5.52$$

$$\text{Standard deviation } \sigma = \sqrt{5.52} = 2.35$$

$$\text{Coefficient of Variation C.V} = \frac{\sigma}{\bar{x}} \times 100 = \frac{2.35}{25.467} \times 100 = 9.23\ \%$$

Probability

Random Experiment: An experiment is said to be a random experiment, if it's out-come can't be predicted with certainty.

Example:

If a coin is tossed, we can't say, whether head or tail will appear. So it is a random experiment.

Sample Space: The set of all possible out-comes of an experiment is called the sample space. It is denoted by $'S'$ and its number of elements are $n(s)$.

Example:

In throwing a dice, the number that appears at top is any one of 1,2,3,4,5,6.

S ={1,2,3,4,5,6} and n(s) = 6

Similarly in the case of a coin, S={Head,Tail} or {H,T} and n(s)=2.

The elements of the sample space are called sample point or event point.

Event: Every subset of a sample space is an event. It is denoted by 'E'.

Example:

UNIT –I

In throwing a dice $S = \{1,2,3,4,5,6\}$, the appearance of an event number will be the event $E = \{2,4,6\}$.

Clearly E is a sub set of S.

Simple event: An event, consisting of a single sample point is called a simple event.

Example:

In throwing a dice, $S = \{1,2,3,4,5,6\}$, so each of $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\ and\ \{6\}$ are simple events.

Compound event: A subset of the sample space, which has more than one element is called a mixed event or compound event.

Example:

In throwing a dice, the event of appearing of odd numbers is a compound event, because $E = \{1,3,5\}$ which has '3' elements.

Equally likely events: The outcomes of an experiment are equally likely to occur when the probability of each outcome is equal.

Example:

1. In the experiment of tossing a coin:

Where A : the event of getting a "HEAD" and

B : the event of getting a "TAIL"

Events "A" and "B" are said to be equally likely events
[Both the events have the same chance of occurrence].

2. In the experiment of throwing a die:

Where A : the event of getting 1

B : the event of getting 2

...

...

F : the event of getting 6

Events "A", "B", "C", "D", "E", "F" are said to be equally likely events
[All these events have the same chance of occurrence.]

3. In the experiment of selecting integers

Where

M : the event of getting an even number

N : the event of getting an odd number

The two compound events "M" and "N" are said to be equally likely.

4. In the experiment of selecting integers
Where

P : the event of getting an odd number {1, 3, 5}

Q : the event of getting 6

The two events "P" and "Q" cannot be said to be equally likely.

Event "P" occurs when any of the elementary events of getting "1", "3" and "5" occur

Event "Q" occurs only when the elementary event of getting "6" occur.

Event "P" is three times more likely to occur than "Q" ⇒ "P" and "Q" are not equally likely.

Exhaustive events: One or more events are said to be exhaustive if all the possible elementary events under the experiment are covered by the event(s) considered together. In other words, the events are said to be exhaustive when they are such that at least one of the events compulsorily occurs.

Exhaustive events may be elementary or compound events. They may be equally likely or not equally likely.

Example

1. In the experiment of tossing a coin:

Where

A : the event of getting a HEAD

B : the event of getting a TAIL

The two events "A" and "B" are called exhaustive events. [When we conduct the experiment, at least one of these will occur.]

2. In the experiment of throwing a die:

Where

A : the event of getting 1

B : the event of getting 2

...

UNIT –I

...

F : the event of getting 6

The six Events "A", "B", "C", "D", "E", "F" together are called exhaustive events. [One of these events will occur whenever the experiment is conducted.]

Classical definition of probability:

If 'S' be the sample space, then the probability of occurrence of an event 'E' is defined as:

$$P(E) = n(E)/N(S) = \frac{\text{number of elements in 'E'}}{\text{number of elements in sample space 'S'}}$$

Example:

Find the probability of getting a tail in tossing of a coin.

Solution:

Sample space $S = \{H, T\}$ and $n(s) = 2$

Event $'E' = \{T\}$ and $n(E) = 1$

$$P(E) = \frac{n(E)}{n(S)} = \frac{1}{2}$$

Sure event: Let 'S' be a sample space. If E is a subset of or equal to S then E is called a sure event.

Example:

In a throw of a dice, S={1,2,3,4,5,6}

Let E_1=Event of getting a number less than '7'.

So 'E_1' is a sure event.

So, we can say, in a sure event n(E) = n(S)

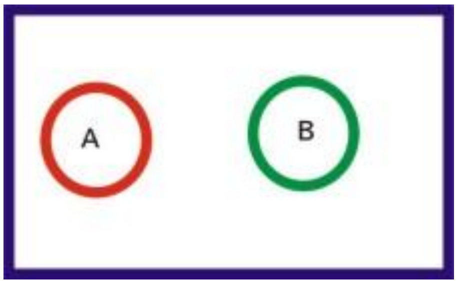
Mutually exclusive or disjoint event: If two or more events can't occur simultaneously, that is no two of them can occur together.

So the event 'A' and 'B' are mutually exclusive if

$$A \cap B = \emptyset \;, \quad P(A \cap B) = 0$$

Pictorial Representation:

UNIT –I



$$A \cap B = \emptyset$$

Example:

When a coin is tossed, the event of occurrence of a head and the event of occurrence of a tail are mutually exclusive events.

Independent or mutually independent events: Two or more events are said to be independent if occurrence or non-occurrence of any of them does not affect the probability of occurrence or non-occurrence of the other event.

If A and B are independent events then $P(A \cap B) = P(A)P(B)$

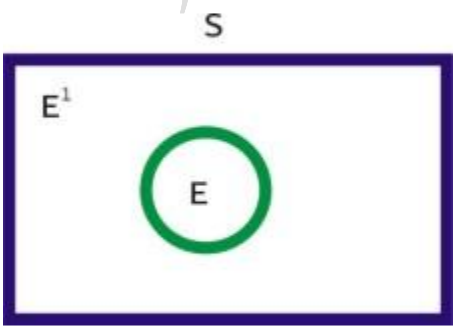
Example:

When a coin is tossed twice, the event of occurrence of head in the first throw and the event of occurrence of head in the second throw are independent events.

Difference between mutually exclusive and mutually independent events: Mutually exclusiveness is used when the events are taken from the same experiment, where as independence is used when the events are taken from different experiments.

Complement of an event: Let 'S' be the sample for random experiment, and 'E' be an event, then complement of 'E' is denoted by 'E' is denoted by E'. Here E' occurs, if and only if E' doesn't occur.

Here E’ occurs, if and only if ‘E’ doesn’t occur.



$$n(E) + n(E') = n(S)$$

UNIT –I

Theorem 1: The probability of an event lies between 'O' and '1'.

i.e. $O \leq P(E) \leq 1$.

Proof: Let 'S' be the sample space and 'E' be the event.

Then

$0 \leq n(E) \leq n(S)$

$$\frac{0}{n(S)} \leq \frac{n(E)}{n(s)} \leq \frac{n(S)}{n(S)}$$

or $O \leq P(E) \leq 1$

The number of elements in 'E' can't be less than 'O' i.e. negative and greater than the number of elements in S.

Theorem 2 : The probability of an impossible event is '0' i.e. $\boldsymbol{P}(\emptyset) = \mathbf{0}$

Proof: Since ∅ has no element, ⇨ $n(\emptyset) = 0$

From definition of Probability:

$$P(\emptyset) = n(\emptyset) / n(S) = 0 / n(S)$$

⇨ $P(\emptyset) = 0$

Theorem 3 : The probability of a sure event is 1. i.e. $P(S) = 1$. where 'S' is the sure event.

Proof : In sure event $n(E) = n(S)$

[Since Number of elements in Event 'E' will be equal to the number of element in sample-space.]

By definition of Probability :

$P(S) = n(S)/ n(S) = 1$

⇨ $P(S) = 1$

Theorem 4: If two events 'A' and 'B' are such that $A \subseteq B$, then $P(A) \leq P(B)$.

Proof: $A \subseteq B$

$\Rightarrow n(A) \leq n(B)$

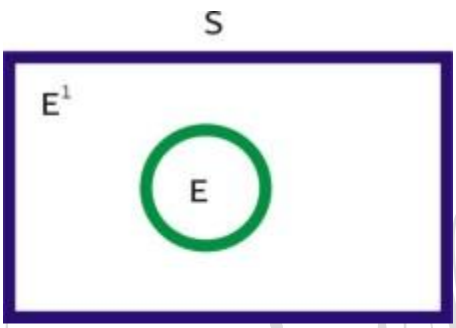
$$n(A)/\, n(S) \leq n(B)\,/\,n(S)$$

$\Rightarrow P(A) \leq P(B)$

Since ‘A’ is the sub-set of ‘B”, so from set theory number of elements in ‘A’ can’t be more than number of element in ‘B’.

Theorem 5 : If ‘E’ is any event and E’ be the complement of event ‘E’, then $P(E) + P(E') = 1$.

Proof:



Let ‘S’ be the sample – space, then

$$n(E) + n(E') = n(S)$$

$$n\,(E)\,/\,n\,(S) \quad + \quad n\,(E')/\,n\,(S) = 1$$

$$P(E) + P(E') = 1$$

Algebra of Events: In a random experiment, let ‘S’ be the sample – space.

Let $A \subseteq S \; and \; B \subseteq S$, where ‘A’ and ‘B’ are events.

Thus we say that :

(i) $A \cup B$, is an event occurs only when at least of ‘A’ and ‘B’ occurs. $\Rightarrow$ $(A \cup B)$ means

$(A \text{ or } B)$.

Ex.: if A = { 2,4,6,} and B = {1, 6}, than the event ‘A’ or ‘B’ occurs, if ‘A’ or ‘B’ or both occur i.e. at least one of ‘A’ and ‘B’ occurs. Clearly ‘A’ or ‘B’ occur, if the out come is any one of the outcomes 1, 2, 4, 6. That is $A \cup B = \{1,2,4,6\}$. (From set – theory).

(ii) $A \cap B$ is an event, that occurs only when each one of ‘A’ and ‘B’ occur ⇒ $(A \cap B)$ means $(A \text{ and } B)$.

Ex.: In the above example, if the out come of an experiment is ‘6’, then events ‘A’ and ‘B’ both occur, because ‘6’ is in both sets. That is $A \cap B = \{6\}$.

(iii) A is an event, that occurs only when ‘A’ doesn’t occur – category of problems related to probability :

(1) Category A – When n(E) and n(S) are determined by writing down the elements of ‘E’ and ‘S’.

(2) Category B – When n(E) and n(S) are calculated by the use of concept of permutation and combination.

(3) Category C – Problems based on $P(E) + P(E^1) = 1$

Q1: A coin is tossed successively three times. Find the probability of getting exactly one head or two heads.

Sol.: Let ‘S’ be the sample – space. Then,

S = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

⇒ n (S) = 8

Let ‘E’ be the event of getting exactly one head or two heads.

Then:

E = { HHT, HTH, THH, TTH, THT, HTT }

⇒ n (E) = 6

Therefore:

$P(E) = n(E) / n(S) = 6/8 = 3/4$

Q2: Three coins are tossed. What is the probability of getting (i) all heads, (ii) two heads, (iii) at least one head, (iv) at least two heads?

Sol.: Let 'S' be the sample – space. Then

S = { HHH, HHT, HTH, THH, HTT, THT, TTH, TTT }

(i) Let 'E_1' = Event of getting all heads.

Then E_1 = { HHH }

$n(E_1) = 1$

⇨ $P(E_1) = n(E_1) / n(S) = 1/8$

(ii) Let E_2 = Event of getting '2' heads.

Then:

E_2 = { HHT, HTH, THH }

$n(E_2) = 3$

⇨ $P(E_2) = 3/8$

(iii) Let E_3 = Event of getting at least one head.

Then:

E_3 = { HHH, HHT, HTH, THH, HTT, THT, TTH }

$n(E_3) = 7$

⇨ $P(E_3) = 7/8$

(iv) Let E_4 = Event of getting at least one head.

Then:

E_4 = { HHH, HHT, HTH, THH, }

$n(E_4) = 4$

⇨ $P(E_4) = 4/8 = 1/2$

Q3: What is the probability, that a number selected from 1, 2, 3, --- 2, 5, is a prime number, when each of the numbers is equally likely to be selected.

Sol.: S = { 1, 2, 3, ---- , 2, 5} ⇨ n(S) = 25

And E = { 2, 3, 5, 7, 11, 13, 17, 19, 23 } ⇨ n(E) = 9

Hence P(E) = n(E) / n(S) = 9 / 25

Q4: Two dice are thrown simultaneously. Find the probability of getting :

(i) The same number on both dice,

(ii) An even number as the sum,

(iii) A prime number as the sum,

(iv) A multiple of '3' as the sum,

(v) A total of at least 0,

(vi) A doublet of even numbers,

(vii) A multiple of '2' on one dice and a multiple of '3' on the other dice.

Sol.: Here:

S = { (1,1), (1,2) ------, (1,6), (2,1), (2,2), ---- (2,6), (3,1), (3,2), -----, (3,6), (4,1), (4,2), -------- (4,6), (5,1), (5,2), ----- (5,6), (6,1,), (6,2), --------------- (6,6) }

n(S) = 6 x 6 = 36

(i) Let E_1 = Event of getting same number on both side:

⇨ E_1 = { (1,1), (2,2), (3,3), (4,4), (5,5), (6,6) }

$n(E_1) = 6$

$P(E_1) = n(E_1)/n(S) = 6/36 = 1/6$

UNIT –I

(ii) Let E_2 = Event of getting an even number as the sum.

E_2 = { (1,1), (1,3), (1,5), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5), (4,2), (4,4),

(4,6), (5,1), (5,5), (6,2), (6,4), (6,6) }

$n(E_2)$ = 18 hence $P(E_2)$ = $n(E_2)/n(S)$ = 18/36 = 1/2

(iii) Let E_3 = Event of getting a prime number as the sum..

E_3 = { (1,1), (1,2), (1,4), (1,6), (2,1), (2,3), (2,5), (3,2), (3,4), (4,1), (4,3),

(5,2), (5,6), (6,1), (6,5),}

$n(E_3)$ = 15

$P(E_2)$ = $n(E_3) / n(S)$ = 15/36 = 5/12

(iv) Let E_4 = Event of getting a multiple of '3' as the sum.

E_4 = { (1,2), (1,5), (2,1), (2,4), (3,3), (3,6), (4,2), (4,5), (5,1), (5,4),

(6,3), (6,6),}

$n(E_4)$ = 12

$P(E_4)$ = $n(E_4)/n(S)$ = 12/36 = 1/3

(v) Let E_5 = Event of getting a total of at least 10.

E_5 = { (4,6), (5,5), (5,6), (6,4), (6,5), (6,6), }

$n(E_5)$ = 6

$P(E_5)$ = $n(E_5)/n(S)$ = 6/36 = 1/6

(vi) Let E_6 = Event of getting a doublet of even numbers.

E_6 = { (2,2), (4,4), (6,6), }

$n(E_6)$ = 3

$P(E_6) = n(E_6)/ n(S) = 3/36 = 1/12$

(vii) Let E_7 = Even of getting a multiple of '2" on one dice and a multiple of '3' on the other dice.

E_7 = { (2,3), (2,6), (4,3), (4,6), (6,3), (3,2), (3,4), (3,6), (6,2), (6,4) }

$n(E_7) = 11$

$P(E_7) = n(E_7) / n(S) = 11/36$

Q5.: What is the probability, that a leap year selected at random will contain 53 Sundays?

Sol.: A leap year has 366 days, therefore 52 weeks i.e. 52 Sunday and 2 days.

The remaining 2 days may be any of the following :

(i) Sunday and Monday

(ii) Monday and Tuesday

(iii) Tuesday and Wednesday

(iv) Wednesday and Thursday

(v) Thursday and Friday

(vi) Friday and Saturday

(vii) Saturday and Sunday

For having 53 Sundays in a year, one of the remaining 2 days must be a Sunday.

$n(S) = 7$

$n(E) = 2$

$P(E) = n(E) / n(S) = 2 / 7$

<u>CATEGORY – B</u>

UNIT –I

Problems based on fundamental principal of counting and permutations and combinations :

Q1. A bag contains '6' red, 4 white and 8 blue balls. If three balls are drawn at random, find the probability, that

(i) '1' is red and '2' are white, (ii) '2' are blue and 1 is red, (iii) none is red.

Sol.: We have to select '3' balls, from 18 balls (6+4+8)

⇨ $n(S) = {}^{18}C_3 = 18! / (3! \times 15!) = (18x17x16) /(3x2x1) = 816$

(i) Let E_1 = Event of getting '1' ball is red and '2' are white

Total number of ways = $n(E_1) = {}^{6}C_1 \times {}^{4}C_2$

= 6 x 4 / 2

= 36

$P(E_1) = n(E_1) / n(S) = 36/816 = 3/68$

(ii) Let E_2 = Event of getting '2' balls are blue and '1' is red.

= Total no. of ways ⇨ $n(E_2) = {}^{8}C_2 \times {}^{6}C_1$

= (8 x 7) / 2 x 6 / 1 = 168

$P(E_2)$ = 168 / 816 = 7/34

(iii) Let E_3 = Event of getting '3' non – red balls. So now we have to choose all the three balls from 4 white and 8 blue balls.

Total number of ways :

$n(E_3) = {}^{12}C_3$ = (12x11x10) / (3x2x1) = 220

$P(E_3) = n(E_3) / n(S)$ = 220 / 816 = 55/204

Q: A box contains 12 bulbs of which '4' are defective. All bulbs took alike. Three bulbs are drawn randomly.

What is the probability that :

(i) all the '3' bulbs are defective?

(ii) At least '2' of the bulbs chosen are defective?

(iii) At most '2' of the bulbs chosen are defective?

Sol.: We have to select '3' bulbs from 12 bulbs.

⇨ $n(S) = {}^{12}C_3 = (12x11x10) / (3x2x1) = 220$

(i) Let E_1 = All the '3' bulbs are defective.

⇨ All bulbs have been chosen, from '4' defective bulbs.

⇨ $n(E_1) = {}^{4}C_3 = 4$

⇨ $P(E_1) = n(E_1) / n(S) = 4 /220 = 1/55$

(ii) Let E_2 = Event drawing at least 2 defective bulbs. So here, we can get '2' defective and 1 non-defective bulbs or 3 defective bulbs.

$n(E_2) = {}^{4}C_2 \times {}^{8}C_1 + {}^{4}C_3$ [Non-defective bulbs = 8]

= 4x3 / 2 x 8/1 + 4/1 = 48+4

$n(E_2) = 52$

⇨ $P(E_2) = n(E_2) / n(S) = 52/220 = 13/55$

(iii) Let E_3 = Event of drawing at most '2' defective bulbs. So here, we can get no defective bulbs or 1 is defective and '2' is non-defective or '2' defective bulbs.

$n(E_3) = {}^{8}C_3 + {}^{4}C_1 \times {}^{8}C_2 + {}^{4}C_2 \times {}^{8}C_1$

= (8x7x6) / (3x2x1) + 4 x (8x7)/2 + (4 x 3) / 2 + 8/1

= 216

$P(E_3) = n(E_3) / n(S) = 216 / 220 = 54 / 55$

Q: In a lottery of 50 tickets numbered from '1' to '50' two tickets are drawn simultaneously. Find the probability that:

(i) Both the tickets drawn have prime number on them,

(ii) None of the tickets drawn have a prime number on it.

Sol.: We want to select '2' tickets from 50 tickets.

⇨ $n(S) = {}^{50}C_2 = (50x49) / 2 = 1225$

(i) Let E_1 = Event that both the tickets have prime numbers Prime numbers between '1' to '50' are :

2,3,5,7,11,13,17,19,23,29,31,37,41,43,47.

Total Numbers = 15.

We have to select '2' numbers from these 15 numbers.

⇨ $n(E_1) = {}^{15}C_2$ = 15? / (2? x 13?) = (15x47) / 2 = 105

⇨ $P(E_1) = n(E_1) / n(S)$ = 105/1225 = 21/245

(ii) Non prime numbers between '1' to '50' = 50-15 = 35

Let E_2 = Event that both the tickets have non-prime numbers.

Now we have to select '2' numbers, from '35' numbers.

⇨ $n(E_2) = {}^{35}C_2$ = 35? / (2? x 33?) = (35x34) / 2 = 595

⇨ $P(E_2) = n(E_2) / n(S)$ = 595 / 1225 = 17/35

Q.: A bag contains 30 tickets, numbered from '1' to '30'. Five tickets are drawn at random and arranged in ascending order. Find the probability that the third number is 20.

Sol.: Total number of ways of selecting '5' tickets from 30 tickets = ${}^{30}C_5$

⇨ $n(S) = {}^{30}C_5$ = 30? / (5? x 25?) = (30 x 29 x 28 x 27 x 26) / (5 x 4 x 3 x 2 x 1)

n(S) = 29 x 27 x 26 x 7

Suppose the '5' tickets are a1, a2,20, a4, a5

They are arranged in ascending order.

⇨ a1, a2 ⊆ {1, 2, 3, ------- , 19} and a4, a5 ⊆ { 21, 22, 23, -----, 30}

We have to select '2' tickets from first '19' tickets and '2' tickets from last 10 tickets.

⇨ n(E) = ${}^{19}C_2 \times {}^{10}C_2$

= 19? / (2? x 17?) = 10? / (2? x 8?) = (19 x 18) / 2 = (10 x 9) / 2

= 19 x 9 x 5 x 9

⇨ P(E) = n(E) / n(S) = (19x9x5x9) / (29x27x26x7) = 285 / 5278

Odds is Favour and Odds against an Event:

UNIT –I

Let ‘S” be the sample space and ‘E’ be an event. Let ‘E’ devotes the complement of event ‘E’, then.

(i) Odds in favour of event ‘E’ = $n(E) / n(E^1)$

(ii) Odds in against of an event ‘E’ = $n(E^1) / n(E)$

Note : Odds in favour of ‘E’ = $n(E) / n(E^1)$

$= [n(E) / n(S)] / [n(E^1) / n(S)] = P(E) / P(E^1)$

Similarly odds in against of ‘E’ = $P(E^1) / P(E)$

Ex.: The odds in favours of an event are 3:5 find the probability of the occurrence of this event.

Sol.: Let ‘E’ be an event.

Then odds in favour of E = $n(E) / n(E^1) = 3 / 5$

⇨ $n(E) = 3$ and $n(E^1) = 5$

Total number of out-comes $n(S) = n(E) + n(E^1) = 3+5 = 8$

$P(E) = n(E) / n(S) = 3 / 8$

Q.: If ‘12’ persons are seated at a round table, what is the probability that two particulars persons sit together?

Sol.: We have to arrange 12 persons along a round table.

So if ‘S” be the sample – space, then n(S) = (12-1)? = 11?

n(S) = 11?

Now we have to arrange the persons in away, such that ‘2’ particulars person sit together.

Regarding that 2 persons as one person, we have to arrange 11 persons.

Total no. of ways = (11-1)? = 10? ways.

That ‘2’ persons can be arranged among themselves in 2? ways.

So, total no. of ways, of arranging 12 persons, along a round table, so that two particular person sit together : = 10? x 2?

⇨n(E) = 10? x 2?

⇨ P(E) = n(E) / n(S) = (10? x 2?) / 11? = 2 / 11

Q.: 6 boys and 6 girls sit in a row randomly, find the probability that all the '6' girls sit together.

Sol.: We have to arrange '6' boys and '6' girls in a row.

⇨ n(S) = 12?

Now, we have to arrange '6' girls in a way, such that all of them should sit together.

Regarding all the 6 girls as one person, we have to arrange 7 person in a row.

⇨ Total no. of ways = 7?

But 6 girls can be arranged among themselves in 6? ways.

⇨ n(E) = 7? x 6?

⇨ P(E) = n(E) / n(S) = (7? x 6?) / 12? = (6x5x4x3x2x1) / (12x11x10x9x8)

P(E) = 1 / 132

Q: If from a pack of '52' playing cards one card is drawn at random, what is the probability that it is either a kind or a queen?

Sol.: n(S) = Total number of ways of selecting 1 card out of 52 cards.

= $^{52}C_1$ = 52

n(E) = Total number of selections of a card, which is either a kind or a queen.

= $^{4}C_1$ + $^{4}C_1$ = 4 + 4 = 8

P(E) = n(E) / n(S) = 8 / 52 = 2 / 3

Q.: From a pack of 52 playing cards, three cards are drawn at random. Find the probability of drawing a king, a queen and a jack.

Sol.: Here n(S) = $^{52}C_3$ = 52? / (3? x 49?) = (52x51x50) / (3x2x1)

= 52x17x25

n(E) = $^{4}C_1$. $^{4}C_1$. $^{4}C_1$

= 4? / (1? x 3?) x 4? / (1? x 3?) = 4? / (1? x 3?)

n(E) = 4 x 4 x 4

⇨ P(E) = n(E) / n(S) = (4x4x4) / (52x17x25) = 16 / 5525

CATEGORY – C

Problems based on finding P(E'), by the use of P(E') = 1 – P(E) :

Note : When an event has a lot of out comes, then we use this concept.

Ex.: What is the probability of getting a total of less than '12' in the throw of two dice?

Sol.: Here n(S) = 6x6 = 36

It is very difficult to find out all the cares, in which we can find the total less then '12'.

So let E = The event, that the sum of numbers is '12'.

Then E = { 6, 6}

n(E) = 1

⇨ P(E) = n(E) / n(S) = 1/36

Required probability, P(E') = 1-P(E)

= 1 – 1/36

P(E1) = 35 /36

Ex.: There are '4' envelopes corresponding to '4' letters. If the letters are placed in the envelopes at random, what is the probability that all the letters are not placed in the right envelopes?

Sol.: We have to place '4' letters in 4 envelopes.

⇨ n(S) = 4!

Now:

Let E = The event, that all the 4 letters are placed in the corresponding envelopes.

So E' = The event that all the '4' letters are not placed in the right envelope.

Here n(E) = 1

P(E) = n(E) / n(S) = 1 / 4! = 1 / 24

Required probability, P(E') = 1- P(E)

= 1 – (1/24)

P(E') = 23 / 24

Some information's about playing cards:

(1) A pack of 52 playing cards has 4 suits :

(a) Spades, (b) Hearts, (c) Diamonds, (d) Clubs.

(2) Spades and clubs are black and Hearts and Diamonds are red faced cards.

(3) The aces, kings, queens, and jacks are called face cards or honours – cards.

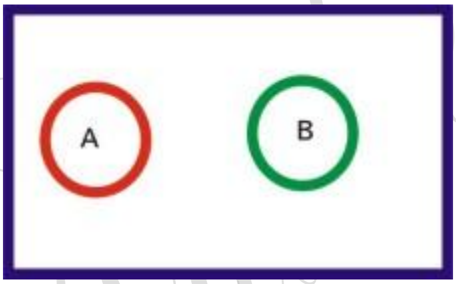
Part – 2 : (Total Probability)

Theorem – 1 : If 'A' and 'B' are mutually exclusive events then $P(A \cap B) = 0$ or P (A and B) = 0

Proof : If 'A' and 'B' are mutually exclusive events then $A \cap B = \emptyset$

⇨ $P(A \cap B) = P(\emptyset)$

$= n(\emptyset) / n(S)$ [By definition of probability]



$= o / n(S)$ [Since the number of elements in a null – set is '0']

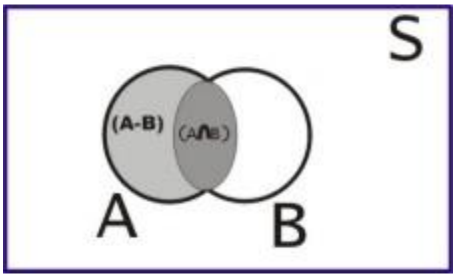
$P(A \cap B) = 0$

(2) Addition Theorem of Probability : If 'A' and 'B' by any two events, then the probability of occurrence of at least one of the events 'A' and 'B' is given by:

P(A or B) = P(A) + P(B) – P (A and B)

P(A ∪ B) = P(A) + P(B) – P (A ∩ B)

UNIT –I



From set theory, we have :

n(A ∪B) = n(A) + n(B) – n(A ∩B)

Dividing both sides by n(S) :

n(A ∪B) / n(S) = n(A) / n(S) + n(B) / n(S) - n(A ∩B) / n(S)

or P(A ∪B) = p(A) + P(B) – P(A ∩B)

Corollary : If 'A' and 'B' are mutually exclusive events,

Then P(A ∩B) = 0. [As we have proved]

In this case :

⇨ P(A ∪B) = p(A) + P(B)

Addition theorem for '3' events 'A', 'B' and 'C' :

P(A ∪B ∪C) = P(A) + P(B) - P(A ∩B) – P(B ∩C) – P(A ∩C) + P(A ∩B ∩C)

Proof : P(A ∪B ∪C) = P[(A ∪B) ∪C]

= P(A ∪b) + P(C) – P[(A ∪B) ∩C] [By addition theorem for two events]

= P(A ∪B) - P(C) – [P(A ∩C) + P(B ∩C) – P(A ∩C ∩B ∩C)]

= P(A) + P(B) – P(A ∩B) + P(C) – P(A ∩C) – P(B ∩C) + P(A ∩B ∩C)

P(A ∪B ∪C) = P(A) + P(B) + P(C) – P(A ∩B) – P(B ∩C) – P(A ∩C) + P(A ∩B ∩C)

Corollary : If 'A', 'B' and 'C' are mutually exclusive events, then P(A ∩B) = 0, P(B ∩C) = 0, P(A ∩C) = 0 and P(A ∩B ∩C) = 0.

In this case :

⇨ P(A ∪B ∪C) = P(A) + P(B) + P(C)

General Form of Addition Theorem of Probability:

UNIT –I

$$P(A_1 \cup A_2 \cup \text{-----} \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j)$$

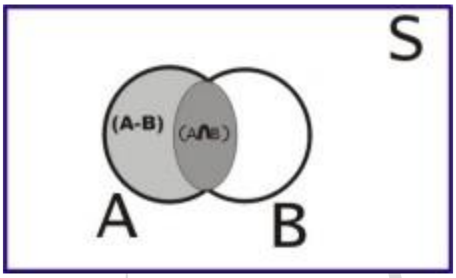
$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \text{-----} (-1)^{n-1} P(A_1 \cap A_2 \cap \text{----} \cap A_n)$$

Corollary : For any number of mutually exclusive events, $A_1, A_2, \text{------}, A_n$:

$P(A_1 \cup A_2 \cup \text{----} \cup A_n) = P(A_1) + P(A_2) + \text{------} + P(A_n)$

Theorem – 3 : For any two events 'A' and 'B'

$P(A\text{-}B) = P(A) – P(A \cap B) = P(A \cap B^1)$



From the figure:

$(A\text{-}B) \cap (A \cap B) =$ ---------------> (i)

and

$(A\text{-}B) \cup (A \cap B) = A$

$P[(A\text{-}B) \cup (A \cap B)] = P(A)$

or $P(A\text{-}B) + P(A \cap B) = P(A)$

[From (i) $(A\text{-}B) \cap (A \cap B) =$ i.e. These events are mutually exclusive]

⇨ $P(A\text{-}B) = P(A) – P(A \cap B)$

or $P(A \cap B) = P(A) – P(A \cap B)$

Similarly $P(A \cap B) = P(B) – P(A \cap B)$

Proof of $P(E) + P(E^1) = 1$, by the addition theorem of probability:

UNIT –I

We know that :

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Putting A = E and B = E^1

$P(E \cup E^1) = P(E) + P(E^1) - P(E \cap E^1)$ ----------------> (1)

From set theory : $E \cup E^1 = S$

And $E \cap E^1 =$

From:

$P(S) = P(E) + P(E^1) - P(\ )$

⇨ $1 = P(E) + P(E^1) - 0$

or $P(E) + P(E^1) = 1$

EXAMPLES

Problems based on addition theorem of probability:

Working rule :

(i) $A \cup B$ denotes the event of occurrence of at least one of the event 'A' or 'B'

(ii) $A \cap B$ denotes the event of occurrence of both the events 'A' and 'B'.

(iii) $P(A \cup B)$ or $P(A+B)$ denotes the probability of occurrence of at least one of the event 'A' or 'B'.

(iv) $P(\cap B)$ or $P(AB)$ denotes the probability of occurrence of both the event 'A' and 'B'.

Ex.: The probability that a contractor will get a contract is '2/3' and the probability that he will get on other contract is 5/9 . If the probability of getting at least one contract is 4/5, what is the probability that he will get both the contracts ?

Sol.: Here $P(A) = 2/3$, $P(B) = 5/9$

$P(A \cup b) = 4/5$, $(P(A \cap B) = ?$

By addition theorem of Probability:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 4/5 = 2/3 + 5/9 - P(A \cap B)$

or $4/5 = 11/9 – P(A \cap B)$

or $P(A \cap B) = 11/9 – 4/5 = (55-36) / 45$

$P(A \cap B) = 19/45$

Ex2.: Two cards are drawn at random. Find the probability that both the cards are of red colour or they are queen.

Sol.: Let S = Sample – space.

A = The event that the two cards drawn are red.

B = The event that the two cards drawn are queen.

⇨ $A \cap B$ = The event that the two cards drawn are queen of red colour.

⇨ $n(S) = {}^{52}C_2$, $n(A) = {}^{26}C_2$, $n(B) = {}^{4}C_2$

$n(A \cap B) = {}^{2}C_2$

⇨ $P(A) = n(A) / n(S) = {}^{26}C_2 / {}^{52}C_2$, $P(B) = n(B) / n(S) = {}^{4}C_2 / {}^{52}C_2$

$P(A \cap B) = n(A \cap B) / n(S) = {}^{2}C_2 / {}^{52}C_2$

$P(A \cup B) = ?$

We have $P(A \cup B) = P(A) + P(B) – P(A \cap B)$

$= {}^{26}C_2 / {}^{52}C_2 + {}^{4}C_2 / {}^{52}C_2 – {}^{2}C_2 / {}^{52}C_2$

$= ({}^{26}C_2 + {}^{4}C_2 – {}^{2}C_2) / {}^{52}C_2$

$= (13X25+2X3-1) / (26X51)$

$P(A \cup B) = 55/221$

Ex.3: A bag contains ‘6’ white and ‘4’ red balls. Two balls are drawn at random. What is the chance, they will be of the same colour?

So.: Let S = Sample space

A = the event of drawing ‘2’ white balls.

B = the event of drawing ‘2’ red balls.

$A \cup B$ = The event of drawing 2 white balls or 2 red balls.

i.e. the event of drawing '2' balls of same colour.

⇨ $n(S) = {}^{10}C_2 = 45$

$n(A) = {}^{6}C_2 = (6 \times 5) / 2 = 15$

$n(B) = {}^{4}C_2 = (4 \times 3) / 2 = 6$

$P(A) = n(A) / n(S) = 15/45 = 1/3$

$P(B) = n(B) / n(S) = 6/45 = 2/15$

⇨ $P(A \cup B) = P(A) + P(B)$

$= 1/3 + 2/15 = (5+2) / 15$

$P(A \cup B) = 7/15$

Ex.: For a post three persons 'A', 'B' and 'C' appear in the interview. The probability of 'A' being selected is twice that of 'B' and the probability of 'B' being selected is thrice that of 'C', what are the individual probability of A, B, C being selected?

Sol.: Let 'E_1', 'E_2', 'E_3' be the events of selections of A, B, and C respectively.

Let $P(E_3) = x$

⇨ $P(E_2) = 3. P(E_3) = 3x$

and $P(E_1) = 2P(E_2) = 2 \times 3x = 6x$

As there are only '3' candidates 'A', 'B' and 'C' we have to select at least one of the candidates A or B or C, surely.

⇨ $P(E_1 \cup E_2 \cup E_3) = 1$

and E_1, E_2, E_3 are mutually exclusive.

⇨ $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3)$

$1 = 6x + 3x + x$

⇨ $10x – 1$ or $x = 1/10$

⇨ $P(E_3) = 1/10$, $P(E_2) = 3/10$ and $P(E_1) = 6/10 = 3/5$

Conditional Probability:

UNIT –I

The conditional probability of an event B in relationship to an event A is the probability that event B occurs given that event A has already occurred. The notation for conditional probability is P(B|A), read as *the probability of B given A*. The formula for conditional probability is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example 1: A jar contains black and white marbles. Two marbles are chosen without replacement. The probability of selecting a black marble and then a white marble is 0.34, and the probability of selecting a black marble on the first draw is 0.47. What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

Solution: $$P(\text{White}|\text{Black}) = \frac{P(\text{Black and White})}{P(\text{Black})} = \frac{0.34}{0.47} = 0.72 = 72\%$$

Example 2: The probability that it is Friday and that a student is absent is 0.03. Since there are 5 school days in a week, the probability that it is Friday is 0.2. What is the probability that a student is absent given that today is Friday?

Solution: $$P(\text{Absent}|\text{Friday}) = \frac{P(\text{Friday and Absent})}{P(\text{Friday})} = \frac{0.03}{0.2} = 0.15 = 15\%$$

Example 3: At Kennedy Middle School, the probability that a student takes Technology and Spanish is 0.087. The probability that a student takes Technology is 0.68. What is the probability that a student takes Spanish given that the student is taking Technology?

Solution: $$P(\text{Spanish}|\text{Technology}) = \frac{P(\text{Technology and Spanish})}{P(\text{Technology})} = \frac{0.087}{0.68} = 0.13 = 13\%$$

Total Probability theorem:

Let $A_1, A_2, \ldots, A_n$ are mutually exclusive events whose probabilities sum to unity and B be any arbitrary event, then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

Example:

One bag contains 5 white and 4 black balls. Another bag contains 7 white and 9 black balls. A ball is transferred from the first bag to the second and then a ball is drawn from second. What will be the probability that the ball is white?

Solution:

Let A_1 denotes the event that the first ball chosen is white, A_2 denotes the event that the first ball chosen is black, and B denotes the event the the ball from the second bag is white.

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) = \frac{5}{9}\frac{8}{17} + \frac{4}{9}\frac{7}{17} = \frac{68}{153}$$

Example: There are three boxes, each containing a different number of light bulbs. The first box has 10 bulbs, of which four are dead, the second has six bulbs, of which one is dead, and the third box has eight bulbs of which three are dead. What is the probability of a dead bulb being selected when a bulb is chosen at random from one of the three boxes?

Solution:

Let A_1, A_2, A_3 denotes the events of selecting bulbs from bags 1,2 and 3 respectively. Let B denotes the event the bulb selected are dead.

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

$$P(B|A_1) = \frac{4}{10}, P(B|A_2) = \frac{1}{6}, \quad P(B|A_3) = \frac{3}{8}$$

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$$

$$= \frac{1}{3}\frac{4}{10} + \frac{1}{3}\frac{1}{6} + \frac{1}{3}\frac{3}{8} = \frac{113}{360}$$

Bayes theorem:

Let $A_1, A_2, \ldots, A_n$ be a set of mutually exclusive events that together form the sample space S. Let B be any event from the same sample space, such that $P(B) > 0$. Then,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Example 1

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

Solution: The sample space is defined by two mutually-exclusive events - it rains or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. Notation for these events appears below.

- Event A_1. It rains on Marie's wedding.
- Event A_2. It does not rain on Marie's wedding
- Event B. The weatherman predicts rain.

In terms of probabilities, we know the following:

- $P(A_1) = 5/365 =0.0136985$ [It rains 5 days out of the year.]
- $P(A_2) = 360/365 = 0.9863014$ [It does not rain 360 days out of the year.]
- $P(B | A_1) = 0.9$ [When it rains, the weatherman predicts rain 90% of the time.]
- $P(B | A_2) = 0.1$ [When it does not rain, the weatherman predicts rain 10% of the time.]

We want to know $P(A_1 | B)$, the probability it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman. The answer can be determined from Bayes' theorem, as shown below.

$$P(A_1 | B) = \frac{P(A_1) P(B | A_1)}{P(A_1) P(B | A_1) + P(A_2) P(B | A_2)}$$

$$P(A_1 | B) = (0.014)(0.9) / [(0.014)(0.9) + (0.986)(0.1)]$$

$$P(A_1 | B) = 0.111$$

Example 2: There are 3 urns A, B and C each containing a total of 10 marbles of which 2, 4 and 8 respectively are red. A pack of cards is cut and a marble is taken from one of the urns depending on the suit shown - a black suit indicating urn A, a diamond urn B, and a heart urn C. What is the probability a red marble is drawn?

Solution:

Let $U_1, U_2\ and\ U_3$ are the events of selecting marbles from urns $A, B\ and\ C$ respectively. Let R be the event of selecting red marble from the urns.

$$P(U_1) = P(U_2) = P(U_3) = \frac{1}{3}$$

$$P(R|U_1) = \frac{2}{10}, P(R|U_2) = \frac{4}{10}, \qquad P(R|U_3) = \frac{8}{10}$$

Probability of drawing a red marble from urn A,

$$P(U_1|R) = \frac{P(U_1)P(R|U_1)}{P(U_1)P(R|U_1) + P(U_2)P(R|U_2) + P(U_3)P(R|U_3)}$$

$$= \frac{\frac{1}{3}\frac{2}{10}}{\frac{1}{3}\frac{2}{10} + \frac{1}{3}\frac{4}{10} + \frac{1}{3}\frac{8}{10}} = \frac{2}{14} = \frac{1}{7}$$

Probability of drawing a red marble from urn B,

$$P(U_2|R) = \frac{P(U_2)P(R|U_2)}{P(U_1)P(R|U_1) + P(U_2)P(R|U_2) + P(U_3)P(R|U_3)}$$

$$= \frac{\frac{1}{3}\frac{4}{10}}{\frac{1}{3}\frac{2}{10} + \frac{1}{3}\frac{4}{10} + \frac{1}{3}\frac{8}{10}} = \frac{4}{14} = \frac{2}{7}$$

Probability of drawing a red marble from urn C,

$$P(U_3|R) = \frac{P(U_3)P(R|U_3)}{P(U_1)P(R|U_1) + P(U_2)P(R|U_2) + P(U_3)P(R|U_3)}$$

$$= \frac{\frac{1}{3}\frac{8}{10}}{\frac{1}{3}\frac{2}{10} + \frac{1}{3}\frac{4}{10} + \frac{1}{3}\frac{8}{10}} = \frac{8}{14} = \frac{4}{7}$$