QUANTITATIVE TECHNIQUES FOR BUSINESS
(BCM4C04)

STUDY MATERIAL

COMPLEMENTARY COURSE

IV SEMESTER

B.Com.
(2019 Admission)

UNIVERSITY OF CALICUT
SCHOOL OF DISTANCE EDUCATION
CALICUT UNIVERSITY P.O.
MALAPPURAM - 673 635, KERALA

19612

School of Distance Education
University of Calicut

Study Material : IV Semester
B.Com. (2019 Admission)

Complementary Course : BCM4C04
QUANTITATIVE TECHNIQUES FOR BUSINESS

Prepared by:

Modules I-IV
Sri. VINEETHAN T.
Assistant Professor, Dept. of Commerce
Govt. College, Madappally.

Module V
Sri. RAJAN P..
Assistant Professor of Commerce
School of Distance Education, University of Calicut.

Scrutinized by:

Modules I-IV
Dr. K. VENUGOPALAN
Associate Professor, Dept. of Commerce
Govt. College, Madappally.

Module V
Sri. MUHAMMED FAISAL T.
Asst. Professor of Commerce
EMEA College of Arts & Science, Kondotty.

CONTENTS

Module I

QUANTITATIVE TECHNIQUES

Meaning and Definition

Quantitative techniques may be defined as those techniques which provide the decision makes a systematic and powerful means of analysis, based on quantitative data. It is a scientific method employed for problem solving and decision making by the management. With the help of quantitative techniques, the decision maker is able to explore policies for attaining the predetermined objectives. In short, quantitative techniques are inevitable in decision-making process.

Classification of Quantitative Techniques

There are different types of quantitative techniques. We can classify them into three categories. They are:

1. Mathematical Quantitative Techniques
2. Statistical Quantitative Techniques
3. Programming Quantitative Techniques

Mathematical Quantitative Techniques

A technique in which quantitative data are used along with the principles of mathematics is known as mathematical quantitative techniques. Mathematical quantitative techniques involve:

1. Permutations and Combinations

Permutation means arrangement of objects in a definite order. The number of arrangements depends upon the total number of objects and the number of objects taken at a time for arrangement. The number of permutations or arrangements is calculated by using the following formula:-

$$n_{P_r} = \frac{n!}{(n-r)!}$$

Combination means selection or grouping objects without considering their order. The number of combinations is calculated by using the following formula:-

$$n_{C_r} = \frac{n!}{(n-r)!}$$

2. Set Theory

Set theory is a modern mathematical device which solves various types of critical problems.

3. Matrix Algebra

Matrix is an orderly arrangement of certain given numbers or symbols in rows and columns. It is a mathematical device of finding out the results of different types of algebraic operations on the basis of the relevant matrices.

4. Determinants:

It is a powerful device developed over the matrix algebra. This device is used for finding out values of different variables connected with a number of simultaneous equations.

5. Differentiation

It is a mathematical process of finding our changes in the dependent variable with reference to a small change in the independent variable.

6. Integration

Integration is the reverse process of differentiation.

7. Differential Equation

It is a mathematical equation which involves the differential coefficients of the dependent variables.

Statistical Quantitative Techniques

Statistical techniques are those techniques which are used in conducting the statistical enquiry concerning to certain Phenomenon. They include all the statistical methods beginning from the collection of data till interpretation of those collected data.

Statistical techniques involve:

1. Collection of data

One of the important statistical methods is collection of data. There are different methods for collecting primary and secondary data.

2. Measures of Central tendency, dispersion, skewness and Kurtosis

Measures of Central tendency is a method used for finding he average of a series while measures of dispersion used for finding out the variability in a series. Measures of Skewness measures asymmetry of a distribution while measures of Kurtosis measures the flatness of peakedness in a distribution.

3. Correlation and Regression Analysis

Correlation is used to study the degree of relationship among two or more variables. On the other hand, regression technique is used to estimate the value of one variable for a given value of another.

4. Index Numbers

Index numbers measure the fluctuations in various Phenomena like price, production etc over a period of time. They are described as economic barometres.

5. Time Series Analysis

Analysis of time series helps us to know the effect of factors which are responsible for changes:

6. Interpolation and Extrapolation

Interpolation is the statistical technique of estimating under certain assumptions, the missing figures which may fall within the range of given figures. Extrapolation provides estimated figures outside the range of given data.

7. Statistical Quality Control

Statistical quality control is used for ensuring the quality of items manufactured. The variations in quality because of assignable causes and chance causes can be known with the help of this tool. Different control charts are used in controlling the quality of products.

8. Ratio Analysis

Ratio analysis is used for analyzing financial statements of any business or industrial concerns which help to take appropriate decisions.

9. Probability Theory

Theory of probability provides numerical values of the likely hood of the occurrence of events.

10. Testing of Hypothesis

Testing of hypothesis is an important statistical tool to judge the reliability of inferences drawn on the basis of sample studies.

Programming Techniques

Programming techniques are also called operations research techniques. Programming techniques are model building techniques used by decision makers in modern times.

Programming techniques involve:

1. Linear Programming

Linear programming technique is used in finding a solution for optimizing a given objective under certain constraints.

2. Queuing Theory

Queuing theory deals with mathematical study of queues. It aims at minimizing cost of both servicing and waiting.

3. Game Theory

Game theory is used to determine the optimum strategy in a competitive situation.

4. Decision Theory

This is concerned with making sound decisions under conditions of certainty, risk and uncertainty.

5. Inventory Theory

Inventory theory helps for optimizing the inventory levels. It focuses on minimizing cost associated with holding of inventories.

6. Net work programming

It is a technique of planning, scheduling, controlling, monitoring and co-ordinating large and complex projects comprising of a number of activities and events. It serves as an instrument in resource allocation and adjustment of time and cost up to the optimum level. It includes CPM, PERT etc.

7. Simulation

It is a technique of testing a model which resembles a real life situations.

8. Replacement Theory

It is concerned with the problems of replacement of machines, etc due to their deteriorating efficiency or breakdown. It helps to determine the most economic replacement policy.

9. Non Linear Programming

It is a programing technique which involves finding an optimum solution to a problem in which some or all variables are non-linear.

10. Sequencing

Sequencing tool is used to determine a sequence in which given jobs should be performed by minimising the total efforts.

11. Quadratic Programming

Quadratic programming technique is designed to solve certain problems, the objective function of which takes the form of a quadratic equation.

12. Branch and Bound Technique

It is a recently developed technique. This is designed to solve the combinational problems of decision making where are large number of feasible solutions. Problems of plant location, problems of determining minimum cost of production etc. are examples of combinational problems.

Functions of Quantitative Techniques

The following are the important functions of quantitative techniques.

1. To facilitate the decision-making process
2. To provide tools for scientific research
3. To help in choosing an optimal strategy
4. To enable in proper deployment of resources
5. To help in minimizing costs
6. To help in minimizing the total processing time required for performing a set of jobs.

USES OF QUANTITATE TECHNIQUES

Business and Industry

Quantitative techniques render valuable services in the field of business and industry. Today, all decisions in business and industry are made with the help of quantitative techniques.

Some important uses of quantitative techniques in the field of business and industry are given below:

1. Quantitative techniques of linear programming is used for optimal allocation of scarce resources in the problem of determining product mix.
2. Inventory control techniques are useful in dividing when and how much items are to be purchase so as to maintain a balance between the cost of holding and cost of ordering the inventory.
3. Quantitative techniques of CPM, and PERT helps in determining the earliest and the latest times for the events and activities of a project. This helps the management in proper deployment of resources.
4. Decision tree analysis and simulation technique help the management in taking the best possible course of action under the conditions of risks and uncertainty.
5. Queuing theory is used to minimize the cost of waiting and servicing of the customers in queues.
6. Replacement theory helps the management of determining the most economic replacement policy regarding replacement of an equipment.

Limitations of Quantitative Techniques

Even though the quantitative techniques are inevitable in decision-making process, they are not free from short comings. The following are the important limitations of quantitative techniques.

1. Quantitative techniques involves mathematical models, equations and other mathematical expressions.

2. Quantitative techniques are based on number of assumptions. Therefore, due care must be ensured while using quantitative techniques, otherwise it will lead to wrong conclusions.
3. Quantitative techniques are very expensive.
4. Quantitative techniques do not take into consideration intangible facts like skill, attitude etc.
5. Quantitative techniques are only tools for analysis and decision-making. They are not decisions itself.

Module 2

CORRELATION ANALYSIS

Introduction

In practice, we may come across with lot of situations which need statistical analysis of either one or more variables. The data concerned with one variable only is called univariate data. For Example: Price, income, demand, production, weight, height marks etc are concerned with one variable only. The analysis of such data is called univariate analysis.

The data concerned with two variables are called bivariate data. For example: rainfall and agriculture; price and demand; height and weight etc. The analysis of these two sets of data is called bivariate analysis.

The date concerned with three or more variables are called multivariate data. For example; agricultural production is influenced by rainfall, quality of soil, fertilizer etc.

The statistical technique which can be used to study the relationship between two or more variables is called correlation analysis.

Definition

Two or more variables are said to be correlated if the change in one variable results in a corresponding change in the other variable.

According to Simpson and Kafka, "Correlation analysis deals with the association between two or more variables".

Lun Chou defines, "Correlation analysis attempts to determine the degree of relationship between variables".

Boddington states that "Whenever some definite connection exists between two or more groups or classes of series of data, there is said to be correlation".

In nut shell, correlation analysis is an analysis which helps to determine the degree of relationship exists between two or more variables.

Correlation Coefficient

Correlation analysis is actually an attempt to find a numerical value to express the extent of relationship exists between two or more variables. The numerical measurement showing the degree of correlation coefficient. Correlation coefficient ranges between-1 and +1.

SIGNIFICANCE OF CORRELATION ANALYSIS

Correlation analysis is of immense use in practical life because of the following reasons:

1. Correlation analysis helps us to find a single figure to measure the degree of relationship exists between the variables.
2. Correlation analysis helps to understand the economic behavior.
3. Correlation analysis enables the business executives to estimate cost, price and other variables.
4. Correlation analysis can be used as a basis for the study of regression. Once we know that two variables are closely related, we can estimate the value of one variable if the value of other is known.
5. Correlation analysis helps to reduce the range of uncertainty associated with decision making. The prediction based on correlation analysis is always near to reality.
6. It helps to know whether the correlation is significant or not. This is possible by comparing the correlation

co-efficient with 6PE. It 'r' is more than 6 PE, the correlation is significant.

Classification of Correlation

Correlation can be classified in different ways. The following are the most important classifications.

1. Positive and Negative correlation
2. Simple, partial and multiple correlation
3. Linear and Non-linear correlation

Positive and Negative Correlation

Positive Correlation

When the variables are varying in the same direction, it is called positive correlation. In other words, if an increase in the value of one variable is accompanied by an increase in the value of other variable or if a decrease in the value of one variable is accompanied by a decree se in the value of other variable, it is called positive correlation.

Eg:

1)	A:	10	20	30	40	50
	B:	80	100	150	170	200
2)	X:	78	60	52	46	38
	Y:	20	18	14	10	5

Negative Correlation

When the variables are moving in opposite direction, it is called negative correlation. In other words, if an increase in the value of one variable is accompanied by a decrease in the value of other variable of if a decrease in the value of one variable is accompanied by an increase in the value of other variable, it is called negative correlation.

Eg: 1)

A:	5	10	15	20	25
B:	16	10	8	6	2

2)

X:	40	32	25	20	10
Y:	2	3	5	8	12

Simple, Partial and Multiple Correlation

Simple Correlation

In a correlation analysis, if only two variables are studied it is called simple correlation. Eg. the study of the relationship between price & demand, of a product or price and supply of a product is a problem of simple correlation.

Multiple correlation

In a correlation analysis, if three or more variables are studied simultaneously, it is called multiple correlation. For example, when we study the relationship between the yield of rice with both rainfall and fertilizer together, it is a problem of multiple correlation.

Partial correlation

In a correlation analysis, we recognize more than two variable, but consider one dependent variable and one independent variable and keeping the other Independent variables as constant. For example yield of rice is influenced by the amount of rainfall and the amount of fertilizer used. But if we study the correlation between yield of rice and the amount of rainfall by keeping the amount of fertilizers used as constant, it is a problem of partial correlation.

Linear and Non-linear correlation

Linear Correlation

In a correlation analysis, if the ratio of change between the two sets of variables is same, then it is called linear correlation.

For examples when 10% increase in one variable is accompanied by 10% increase in the other variable, it is the problem of linear correlation.

X:	10	15	30	60
Y:	50	75	150	300

Here the ratio of change between X and Y is the same. When we plot the data in graph paper, all the plotted points would fall on a straight line.

Non-linear correlation

In a correlation analysis if the amount of change in one variable does not bring the same ratio of change in the other variable, it is called non linear correlation.

X:	2	4	6	10	15
Y:	8	10	18	22	26

Here the change in the value of X does not being the same proportionate change in the value of Y.

This is the problem of non-linear correlation, when we plot the data on a graph paper, the plotted points would not fall on a straight line.

Degrees of Correlation

Correlation exists in various degrees

1. Perfect positive correlation

If an increase in the value of one variable is followed by the same proportion of increase in other related variable or if a decrease in the value of one variable is followed by the same proportion of decrease in other related variable, it is perfect positive correlation. eg: if 10% rise in price of a commodity results in 10% rise in its supply, the correlation is perfectly positive. Similarly, if 5% full in price results in 5% fall in supply, the correlation is perfectly positive.

2. Perfect Negative correlation

If an increase in the value of one variable is followed by the same proportion of decrease in other related variable or if a decrease in the value of one variable is followed by the same proportion of increase in other related variably it is Perfect Negative Correlation. For example if 10% rise in price results in 10% fall in its demand the correlation is perfectly negative. Similarly if 5% fall in price results in 5% increase in demand, the correlation is perfectly negative.

3. Limited Degree of Positive Correlation

When an increase in the value of one variable is followed by a non-proportional increase in other related variable, or when a decrease in the value of one variable is followed by a non-proportional decrease in other related variable, it is called limited degree of positive correlation.

For example, if 10% rise in price of a commodity results in 5% rise in its supply, it is limited degree of positive correlation. Similarly if 10% fall in price of a commodity results in 5% fall in its supply, it is limited degree of positive correlation.

4. Limited degree of Negative correlation

When an increase in the value of one variable is followed by a non-proportional decrease in other related variable, or when a decrease in the value of one variable is followed by a non-proportional increase in other related variable, it is called limited degree of negative correlation.

For example, if 10% rise in price results in 5% fall in its demand, it is limited degree of negative correlation. Similarly, if 5% fall in price results in 10% increase in demand, it is limited degree of negative correlation.

5. Zero Correlation (Zero Degree Correlation)

If there is no correlation between variables it is called zero correlation. In other words, if the values of one variable cannot be associated with the values of the other variable, it is zero correlation.

Methods of measuring correlation

Correlation between 2 variables can be measured by graphic methods and algebraic methods.

I Graphic Methods

1. Scatter Diagram
2. Correlation graph

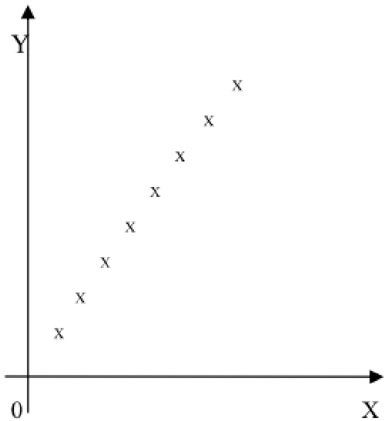
II Algebraic methods (Mathematical Methods or statistical methods or Co-efficient of correlation methods)

1. Karl Pearson's Co-efficient of correlation
2. Spear mans Rank correlation method
3. Concurrent deviation method

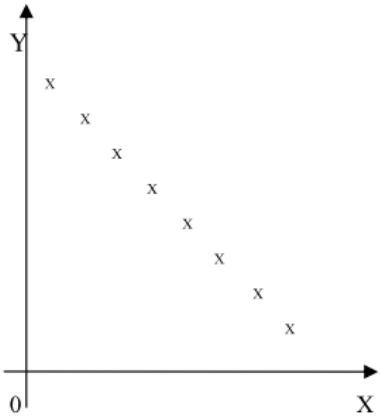
Scatter Diagram

This is the simplest method for ascertaining the correlation between variables. Under this method all the values of the two variables are plotted in a chart in the form of dots. Therefore, it is also known as dot chart. By observing the scatter of the various dots, we can form an idea that whether the variables are related or not.

A scatter diagram indicates the direction of correlation and tells us how closely the two variables under study are related. The greater the scatter of the dots, the lower is the relationship



Perfect Positive Correlation



Perfect Negative Correlation

Merits of Scatter Diagram method

1. It is a simple method of studying correlation between variables.
2. It is a non-mathematical method of studying correlation between the variables. It does not require any mathematical calculations.
3. It is very easy to understand. It gives an idea about the correlation between variables even to a layman.
4. It is not influenced by the size of extreme items.
5. Making a scatter diagram is, usually, the first step in investigating the relationship between two variables.

Demerits of Scatter diagram method

1. It gives only a rough idea about the correlation between variables.
2. The numerical measurement of correlation co-efficient cannot be calculated under this method.

3. It is not possible to establish the exact degree of relationship between the variables.

Correlation graph Method

Under correlation graph method the individual values of the two variables are plotted on a graph paper. Then dots relating to these variables are jointed separately so as to get two curves. By examining the direction and closeness of the two curves, we can infer whether the variables are related or not. If both the curves are moving in the same direction (either upward or downward) correlation is said to be positive. If the curves are moving in the opposite directions, correlation is said to be negative.

Merits of Correlation Graph Method

1. This is a simple method of studying relationship between the variable.
2. This does not require mathematical calculations.
3. This method is very easy to understand

Demerits of correlation graph method

1. A numerical value of correlation cannot be calculated.
2. It is only a pictorial presentation of the relationship between variables.
3. It is not possible to establish the exact degree of relationship between the variables.

Karl Pearson's Co-efficient of Correlation

Karl Pearsons' Co-efficient of Correlation is the most popular method among the algebraic methods for measuring correlation. This method was developed by Prof. Karl Pearson in 1896. It is also called product moment correlation coefficient.

Pearson's co-efficient of correlation is defined as the ratio of the covariance between X and Y to the product of their standard deviations. This is denoted by 'r' or r_{xy}

$$r = \frac{\text{Co variance of X and Y}}{(\text{SD of X}) x (\text{SD of Y})}$$

Interpretation of Co-efficient of Correlation

Pearson's Co-efficient of correlation always lies between +1 and -1. The following general rules will help to interpret the Co-efficient of correlation:

1. When r-+1, it means there is perfect positive relationship between variables.
2. When r = -1, it means there is perfect negative relationship between variables.
3. When r = 0, it means there is no relationship between the variables.
4. When 'r' is closer to +1, it means there is high degree of positive correlation between variables.
5. When 'r' is closer to -1, it means there is high degree of negative correlation between variables.
6. When 't' is closer to 'O', it means there is less relationship between variables.

Properties of Pearson's Co-efficient of Correlation

1. If there is correlation between variables, the Co-efficient of correlation lies between +1 and -1.
2. If there is no correlation, the coefficient of correlation is denoted by zero (ie r=0)
3. It measures the degree and direction of change
4. If simply measures the correlation and does not help to predict cansation.

5. It is the geometric mean of two regression co-efficient.

i.e $r = \sqrt{b_{x_y} - b_{y_x}}$

Computation of Pearsons's Co-efficient of correlation

Pearson's correlation co-efficient can be computed in different ways. They are:

a. Arithmetic mean method

b. Assumed mean method

c. Direct method

Arithmetic mean method

Under arithmetic mean method, co-efficient of correlation is calculated by taking actual mean.

$$r = \frac{\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{\sqrt{\sum\left(x-\bar{x}\right)^2\sum\left(y-\bar{y}\right)^2}}$$

or

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \in y^2}} \text{ whereas } x - x - \bar{x} \text{ and } y = y - \bar{y}$$

Calculate Pearson's co-efficient of correlation between age and playing habits of students.

Age	20	21	22	23	24	25
No. of students	500	400	300	240	200	160
Regular players	400	300	180	96	60	24

Let X = Age and Y = Percentage of regular players

Percentage of regular players can be calculated as follows:

$$\frac{400}{500}x100 = 80; \frac{300}{400}x100 = 75; \frac{180}{300}x100 = 60; \frac{96}{240}x100 = 40,$$

$$\frac{60}{20}x100 = 30; \text{and } \frac{24}{160}x100 = 15$$

Pearson's Coefficient of Correlation (r) $= \dfrac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2(y-\bar{y})^2}}$

Computation of Pearson's Coefficient of correlation						
Age x	% of Regular Player y	x-x (x-22.5)	$(y-\bar{y})$ (y-50)	$(x-x)(y-\bar{y})$	$(x-x)^2$	$(y-\bar{y})^2$
20	80	-2.5	30	-75.0	6.25	900
21	75	-1.5	25	-37.5	2.25	625
22	60	-0.5	10	-5.0	0.25	100
23	40	0.5	-10	-.5.0	0.25	100
24	30	1.5	-20	-30.0	2.25	400
25	15	2.5	-35	-87.5	6.25	1225
135	300			-240	17.50	3350

$$\bar{x} = \frac{\sum x}{N} = \frac{135}{6} = 22.5$$

$$\underline{y} = \frac{\sum x}{N} = \frac{300}{6} = 50$$

$$r = \frac{240}{\sqrt{17.5x3350}} - \frac{-240}{\sqrt{58,625}} = \frac{-240}{\sqrt{242.126}} = 0.9912$$

Assumed mean method

Under assumed mean method, correlation coefficient is calculated by taking assumed mean only.

$$r = \frac{N\sum dxdy - \left(\sum dx\right)\left(\sum dy\right)}{\sqrt{N\sum dx^{2-}\left(\sum dx\right)^2} x \sqrt{N\sum dy^{2-}\left(\sum dy\right)^2}}$$

Where dx = deviations of X from its assumed mean; dy = deviations of y from its assumed mean

Find out coefficient of correlation between size and defect in quality of shoes:

Size	:	15-16	16-17	17-18	18-19	29-20	20-21
No.of shoes produced	:	200	270	340	260	400	300
No.of defectives	:	150	162	170	180	180	114

Let x = size (ie mid-values)

y = percentage of defectives

∴ x values are 15.5, 16.5, 17.5, 18.5, 19.5 and 20.5

y values are 75 60 50 50 45 and 38

Take assumed mean : x = 17.5 and y = 50

Computation of Pearson's Coefficient of Correlation						
x	y	dx	dy	dxdy	dx^2	dy^2
15.5	75	-2	25	-50	4	625
16.5	60	-1	10	-10	1	100
17.5	50	0	0	0	0	0
18.5	50	1	0	0	1	0
19.5	45	2	-5	-10	4	25
20.5	38	3	-12	-36	9	144
		Σdx=3	Σdy=18	Σdxdy=-106	Σdx^2=19	Σdy^2=894

$$r = \frac{N\sum dxdy - \left(\sum dx\right)\left(\sum dy\right)}{\sqrt{N\sum dx^{2-}\left(\sum dx\right)^{2x}}\sqrt{N\sum dy^{2-}\left(\sum dy\right)^{2}}}$$

$$r = \frac{(6x-106)-(3x18)}{\sqrt{(6x10)-3^{2}}\,x\sqrt{(6x894)-18^{2}}}$$

$$= \frac{-636-54}{\sqrt{114-9}\,x\sqrt{5364-324}}$$

$$= \frac{-690}{\sqrt{105}\,x\sqrt{5040}} = \frac{-690}{727.46} = -0.9485$$

Direct Method

Under direct method, coefficient of correlation is calculated without taking actual mean or assumed mean

$$r = \frac{N\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{N\sum x^{2} - \left(\sum x\right)^{2}}\,x\sqrt{N\sum y^{2-}\left(\sum y\right)^{2}}}$$

From the following data, compute Pearson's correlation coefficient

Price :	10	12	14	15	19
Demand (Qty	40	41	48	60	50

Let us take price = x and demand = y

Price (x)	Demand (y)	xy	x^2	y^2
10	40	400	100	1600
12	41	492	144	1681
14	48	672	196	2304
15	60	900	225	3600
19	50	950	361	2500
$\Sigma x=70$	$\Sigma y=239$	$\Sigma xy=3414$	$\Sigma x^2=1026$	$\Sigma y^2=11685$

$$r = \frac{N\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{N\sum x^2 - \left(\sum x\right)^2} x \sqrt{N\sum y^{2-} \left(\sum y\right)^2}}$$

$$r = \frac{(5x3424) - (70x239)}{\sqrt{(5x1026)^{-} 70^2} x \sqrt{(5x11685) - 239^2}}$$

$$r = \frac{-17,070 - 16,7300}{\sqrt{230} x \sqrt{1304}} = \frac{340}{547.65} = +0.621$$

Probable Error and Coefficient of Correlation

Probable error (PE) of the Co-efficient of correlation is a statistical device which measures the reliability and dependability of the value of co-efficient of correlation.

Probable Error = $\frac{2}{3}$ standard error

= 0.6745 x standard error

Standard Error (SE) = $\frac{1 - r^2}{\sqrt{n}}$

$\therefore$ PE = 0.6745 x $\frac{1 - r^2}{\sqrt{n}}$

If the value of coefficient of correlation (r) is less than the PE, then there is no evidence of correlation.

If the value of 'r' is more than 6 times of PE, the correlation is certain and significant.

By adding and submitting PE from coefficient of correlation, we can find out the upper and lower limits within which the population coefficient of correlation may be expected to lie.

Uses of PE

1. PE is used to determine the limits within which the population coefficient of correlation may be expected to lie.
2. It can be used to test whether the value of correlation coefficient of a sample is significant with that of the population.

If r = 0.6 and N = 64, find out the PE and SE of the correlation coefficient. Also determine the limits of population correlation coefficient.

Sol: r = 60

N = 64

PE = 0.6745 x SE

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

$$= \frac{1-(0.6)^2}{\sqrt{64}} = \frac{1-0.36}{8} = \frac{0.64}{8} = 0.08$$

PE = 0.6745 x 0.08

= 0.05396

Limits of population Correlation coefficient = r ± PE

= 06 ± 0.05396

= 0.54604 to 0.6540

Qn. 2 r and PE have values 0.9 and 0.04 for two series. Find n.

Sol: PE = 0.04

$$0.6745 \text{ x } \frac{1-r^2}{\sqrt{n}} = 0.04$$

$$\frac{1-0.9^2}{\sqrt{n}} = \frac{0.04}{0.6745}$$

$$\frac{1-0.81}{\sqrt{n}} = 0.0593$$

$$\frac{0.19}{\sqrt{n}} = 0.0593$$

$$0.0593x\sqrt{n} = 0.19$$

$$\sqrt{n} = \frac{0.19}{0.0593}$$

$$\sqrt{n} = 3.2$$

$N = 3.2^2 = 10.266$

N = 10

Coefficient of Determination

One very convenient and useful way of interpreting the value of coefficient of correlation is the use of the square of coefficient of correlation. The square of coefficient of correlation is called coefficient of determination.

Coefficient of determination = r^2

Coefficient of determination is the ratio of the explained variance to the total variance.

For example, suppose the value of r = 0.9, then r^2 =0.81 = 81%

This means that 81% of the variation in the dependent variable has been explained by (determined by) the independent variable. Here 19% of the variation in the dependent variable has not been explained by the independent variable. Therefore, this 19% is called coefficient of non-determination.

Coefficient of non-determination (K^2) = 1 - r^2

K^2 1- coefficient of determination

Qn: Calculate coefficient of determination and non-determination if coefficient of correlation is 0.8.

Sol: r = 0.8

Coefficient of determination $= r^2$

$= 0.8^2 = 0.64 = 64\%$

Co efficient of non-determination $= 1 - r^2$

$= 1 - 0.64$

$= 0.36$

$= 36\%$

Merits of Pearson's Coefficient of Correlation

1. This is the most widely used algebraic method to measure coefficient of correlation.
2. It gives a numerical value to express the relationship between variables.
3. It gives both direction and degree of relationship between variables.
4. It can be used for further algebraic treatment such as coefficient of determination coefficient of non-determination etc.
5. It gives a single figure to explain the accurate degree of correlation between two variables

Demerits of Pearson's Coefficient of correlation

1. It is very difficult to compute the value of coefficient of correlation.
2. It is very difficult to understand

3. It requires complicated mathematical calculations
4. It takes more time
5. It is unduly affected by extreme items
6. It assumes a linear relationship between the variables. But in real life situation, it may not be so.

Spearman's Rank Correlation Method

Pearson's coefficient of correlation method is applicable when variables are measured in quantitative form. But there were many cases where measurement is not possible because of the qualitative nature of the variable. For example, we cannot measure the beauty, morality, intelligence, honesty etc. in quantitative terms. However it is possible to rank these qualitative characteristics in some order.

The correlation coefficient obtained from ranks of the variables instead of their quantitative measurement is called rank correlation. This was developed by Charles Edward Spearman in 1904.

Spearman's coefficient correlation $(R) = 1 - \frac{6\sum D^2}{N^{3-N}}$

Where D = difference of ranks between the two variables

N = number of pairs

Qn: Find the rank correlation coefficient between poverty and overcrowding from the information given below:

Town	A	B	C	D	E	F	G	H	I	J
Poverty	17	13	15	16	6	11	14	9	7	12
Over crowing	36	46	35	24	12	18	27	22	2	8

Sol: Here ranks are not given. Hence we have to assign ranks

$$R = 1 - \frac{6\sum D^2}{N^{3-N}}$$

N = 10

Computation of Rank Correlation Co-efficient						
Town	Poverty	Over crowding	R_1	R_2	D	D^2
A	17	36	1	2	1	1
B	13	46	5	1	4	16
C	15	35	3	3	0	0
D	16	24	2	5	3	9
E	6	12	10	8	2	4
F	11	18	7	7	0	0
G	14	27	4	4	0	0
H	9	22	8	6	2	4
I	7	2	9	10	1	1
J	12	8	6	9	3	9
					ΣD^2	44

$$R = 1 - \frac{6x44}{10^3 - 10}$$

$$= 1 - \frac{264}{990}$$

= 1-0.2667

= $^{+}$0.7333

Qn: Following were the ranks given by three judges in a beauty context. Determine which pair of judges has the nearest approach to Common tastes in beauty.

Judge I: 1 6 5 10 3 2 4 9 7 8

Judge I:	3	5	8	4	7	10	2	1	6	9
Judge I:	6	4	9	8	1	2	3	10	5	7

$$R = 1 = \frac{6\sum D^2}{N^3 - N}$$

N = 10

Computation of Spearman's Rank Correlation Coefficient								
Judge I (R_1)	Judge II (R_2)	Judge III (R_3)	R_1-R_2 (D_1)	R_2-R_3 (D_2)	R_1-R_3(D_3)	D_1^2	D_2^2	D_3^2
1	3	6	2	3	5	4	9	25
6	5	4	1	1	2	1	1	4
5	8	9	3	1	4	9	1	16
10	4	8	6	4	2	36	16	4
3	7	1	4	6	2	16	36	4
2	10	2	8	8	0	64	64	0
4	2	3	2	1	1	4	1	1
9	1	10	8	9	1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	1	2	1	1	4	1
					ΣD^2	200	214	60

$$R = 1 - \frac{6\sum D^2}{N}$$

Rank correlation coefficient between I & II = $\frac{6x200}{10^2 - 10}$

$$= 1 - \frac{1200}{990}$$

= 1-1.2121

= -0.2121

Rank correlation Coefficient between II & III judges=1- $\frac{6x214}{10^3 - 10}$

$$= 1 - \frac{1284}{990}$$

= -0.297

Rank correlation coefficient between I & II Judges =

$$= 1 - \frac{6x60}{10^3 - 10}$$

$$= 1 - \frac{300}{990}$$

= 1-0.364

= +0.636

The rank correlation coefficient in case of I & III judges is greater than the other two pairs. Therefore, judges I & III have highest similarity of thought and have the nearest approach to common taste in beauty.

Qn: The co-efficient of rank correlation of the marks obtained by 10 students in statistics & English was 0.2. It was later discovered that the difference in ranks of one of the students was wrongly takes as 7 instead of 9 Find the corr ect result.

R = 0.2

$$R = 1 - \frac{6\sum D^2}{N3 - N} = 0.2$$

$$\frac{1 - 0.2}{1} = \frac{6\sum D^2}{10^3 - 10}$$

$$\frac{0.8}{1} = \frac{6\sum D^2}{990}$$

$6\Sigma D^2$ = 90x0.8 = 792

$$\text{Correct} \sum D^2 = \frac{792}{6} = 132 - 7^2 + 9^2$$

$$=164$$

$$\text{Correct R} \quad 1 = \frac{6\sum D^2}{N^3 - N}$$

$$= 1 - \frac{6x164}{10^3 - 10}$$

$$= 1 - \frac{984}{990}$$

$$= 1\text{-}0.9939$$

$$= 0.0061$$

Qn: The coefficient of rank correlation between marks in English and maths obtained by a group students is 0.8. If the sum of the squares of the difference in ranks is given to be 33, find the number of students in the group.

Sol: $$R = 1 - \frac{6\sum D^2}{N^3 - N} = 0.8$$

$$\text{ie}, 1 - \frac{6x33}{N^3 - N} = 0.8$$

$$= 1 - 08 = \frac{6x33}{N^3 - N}$$

$$= 0.2 \text{ x } (N^3 - N) = 198$$

$$N_3 \text{ -N} = \frac{198}{0.2} = 990$$

$$N = 10$$

Computation of Rank Correlation Coefficient when Ranks are Equal

There may be chances of obtaining same rank for two or more items. In such a situation, it is required to give average rank for all. Such items. For example, if two observations for 4th rank, each of those observations should be given the rank 4.5 (i.e $\frac{4+5}{2} = 4.5$)

Suppose 4 observations got 6th rank, here we have to assign the rank, 7.5 $\left(\text{ie.}\frac{6+7+8+9}{4}\right)$ to each of the 4 observations.

When there is equal ranks, we have to apply the following formula to compute rank correlation coefficient:

$$R = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}\left(m^3 - m\right) + \frac{1}{12}\left(m^3 - m\right) + \ldots\ldots\ldots\right]}{N^3 - N}$$

Where D- Difference of rank in the two series

N - Total number of pairs

m - Number of times each rank repeats

Qn: Obtain rank correlation co-efficient for the data

X:	68	64	75	50	64	80	75	40	55	64
Y:	62	58	68	45	81	60	68	48	50	70

Here, ranks are not given we have to assign ranks Further, this is the case of equal ranks.

$$\therefore R = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}\left(m^3 - m\right) + \ldots\ldots\ldots\right]}{N^3 - N}$$

$$R = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}\left(m^3 - m\right) + \frac{1}{12}\left(m^3 - m\right) + \frac{1}{12}\left(m^3 - m\right) \ldots\ldots\ldots\right]}{N^3 - N}$$

Computation of rank correlation coefficient					
x	y	R_1	R_2	$D(R_1-R_2)$	D^2
68	62	4	5	1	1
64	58	6	7	1	1
75	68	6	7	1	1
50	45	9	10	1	1
54	81	6	1	5	25
80	60	1	6	5	25
75	68	2.5	3.5	1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
				ΣD^2	72

$$R = 1 - \frac{6\left[72 + \frac{1}{12}\left(2^3 - 2\right) + \frac{1}{12}\left(3^3 - 3\right) + \frac{1}{12}\left(2^3 - 2\right)\right]}{N^3 - N}$$

$$= 1 - \frac{6\left[72 + \frac{1}{12} + 2 + \frac{1}{12}\right]}{10^3 - 10}$$

$$= 1 - \frac{6x\left[72 + 3\right]}{990}$$

$$= 1 - \frac{6x75}{990}$$

$$= 1 - \frac{450}{990} = 1 - 0.4545$$

$$= 0.5455$$

Merits of Rank Correlation method

1. Rank correlation coefficient is only an approximate measure as the actual values are not used for calculations.
2. It is very simple to understand the method.
3. It can be applied to any type of data, ie quantitative and qualitative.
4. It is the only way of studying correlation between qualitative data such as honesty, beauty etc.
5. As the sum of rank differences of two qualitative data is always equal to zero, this method facilitates a cross check on the calculation.

Demerits of Rank Correlation method

1. Rank correlation coefficient is only an approximate measure as the actual values are not used for calculations.
2. It is not convenient when number of pairs (ie. N) is large
3. Further algebraic treatment is not possible.
4. Combined correlation coefficient of different series cannot be obtained as in the case of mean and standard deviation. In case of mean and standard deviation, it is possible to compute combine arithematic mean and combined standard deviation.

Concurrent Deviation Method

Concurrent deviation method is a very simple method of measuring correlation. Under this method, we consider only the directions of deviations. The magnitudes of the values are completely ignored. Therefore, this method is useful when we are interested in studying correlation between two variables in a casual manner and not interested in degree (or precision).

Under this method, the nature of correlation is known from the direction of deviation in the values of variables. If deviations of 2 variables are concurrent, then they move in the same direction, otherwise in the opposite direction.

The formula for computing the coefficient of concurrent deviation is:

$$r = \pm\sqrt{\pm\frac{(2c - N)}{N}}$$

Where N = No. of pairs of symbol

C = No. of concurrent deviations (ie, No.of + signs in 'dx dy' column)

Steps:

1. Every value of 'X' series is compared with its proceeding value. Increase is shown by '+' symbol and decrease is shown by '-'
2. The above step is repeated for 'Y' series and we get 'dy'
3. Multiply 'dx' by 'dy' and the product is shown in the next column. The column heading is 'dxdy'.
4. Take the total number of '+' signs in 'dxdy' column. '+' signs in 'dxdy' column donotes the concurrent deviations, and it is indicated by 'C'.
5. Apply the formula:

$$r = \pm\sqrt{\pm\left(\frac{2c - N}{N}\right)}$$

If 2c > N, *then r = + ve and if 2c < N, then r = - ve.*

Qn: Calculate coefficient if correlation by concurrent deviation method:

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011
Supply	160	164	172	182	166	170	178	192	186
Price	292	280	260	234	266	254	230	190	200

Sol: Computation of coefficient of concurrent

Deviation

Supply (x)	Price (y)	dx	dy	dxdy
160	292	+	-	-
164	280	+	-	-
172	260	+	-	-
182	234	+	-	-
166	266	-	+	-
170	254	+	-	-
178	230	+	-	-
192	190	+	-	-
186	200	-	+	-

				C=0
				====

$$r = ^{\pm}\sqrt{\pm\frac{(2C-N)}{N}}$$

$$= ^{\pm}\sqrt{\pm\frac{(2x0)-8}{8}}$$

$$= ^{\pm}\sqrt{\frac{0-8}{8}} == \pm\sqrt{\frac{-8}{8}} = -1$$

Merits of concurrent deviation method

1. It is very easy to calculate coefficient of correlation
2. It is very simple understand the method
3. When the number of items is very large, this method may be used to form quick idea about the degree of relationship
4. This method is more suitable, when we want to know the type of correlation (ie, whether positive or negative)

Demerits of concurrent deviation method

1. The method ignores the magnitude of changes. ie. Equal weight is give for small and big changes.
2. The result obtained by this method is only a rough indicator of the presence or absence of correlation.
3. Further algebraic treatment is not possible.
4. Combined coefficient of concurrent deviation of different series cannot be found as in the case of arithmetic mean and standard deviation.

Module 3

REGRESSION ANALYSIS

Introduction

Correlation analysis analyses whether two variables are correlated or not. After having established the fact that two variables are closely related, we may be interested in estimating the value of one variable, given the value of another. Hence, regression analysis means to analyse the average relationship between two variables and thereby provides a mechanism for estimation or predication or forecasting.

The term 'Regression' was firstly used by Sir Francis Galton in 1877. The dictionary meaning of the term 'regression' is "stepping back" to the average.

Definition

"Regression is the measure of the average relationship between two or more variables in terms of the original units of the date".

"Regression analysis is an attempt to establish the nature of the relationship between variables-that is to study the functional relationship between the variables and thereby provides a mechanism for prediction or forecasting".

It is clear from the above definitions that Regression Analysis is statistical device with the help of which we are able to estimate the unknown values of one variable from known values of another variable. The variable which is used to predict the another variable is called independent variable (explanatory variable) and, the variable we are trying to predict is called dependent variable (explained variable).

The dependent variable is denoted by X and the independent variable is denoted by Y.

The analysis used in regression is called simply linear regression analysis. It is called simple because three is only one predictor (independent variable). It is called linear because, it is assumed that there is linear relationship between independent variable and dependent variable.

Types of Regression

There are two types of regression. They are linear regression and multiple regression.

Linear Regression

It is a type of regression which uses one independent variable to explain and/or predict the dependent variable.

Multiple Regression

It is a type of regression which uses two or more independent variable to explain and/or predict the dependent variable.

Regression Lines

Regression line is a graphic technique to show the functional relationship between the two variable X and Y. It is a line which shows the average relationship between two variables X and Y.

If there is perfect positive correlation between 2 variables, then the two regression lines are winding each other and to give one line. There would be two regression lines when there is no perfect correlation between two variables. The nearer the two regression lines to each other, the higher is the degree of correlation and the farther the regression lines from each other, the lesser is the degree of correlation.

Properties and Regression lines

1. The two regression lines cut each other at the point of average of X and average of Y (i.e X and Y)
2. When r = 1, the two regression lines coincide each other and give on line.
3. When r = 0, the two regression lines are mutually perpendicular.

Regression Equations (Estimating Equations)

Regression equations are algebraic expressions of the regression lines. Since there are two regression lines, therefore two regression equations. They are:

1. Regression Equation of X on Y: This is used to describe the variations in the values of X for given changes in Y.
2. Regression Equation of Y on X: This is used to describe the variations in the value of Y for given changes in X.

Least Square Method of Computing Regression Equation

The method of least square is an objective method of determining the best relationship between the two variables constituting a bivariate data. To find out best relationship between the two variables. This can be done by the principle of least squares.

The principle of least squares says that the sum of the squares of the deviations between the observed values and estimated values should be the least. In other words. $\Sigma(y-y_c)^2$ will be the minimum.

With a little algebra and differential calculators we can develop some equations (2 equations in case of a linear relationship) called normal equations. By solving these normal equations, we can find out the best values of the constants.

Regression Equation of Y on X

$$Y = a + bx$$

The normal equations to compute 'a' and 'b' are:

$$\Sigma y = Na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Regression Equation of X on Y

$$X = a + by$$

The normal equations to compute 'a' and 'b' are:

$$\Sigma x = Na + n\Sigma y$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2$$

Qn: Find regression equations x and y and y on x from the following.

A:	25	30	35	40	45	50	55
Y:	18	24	30	36	42	48	54

Sol: Regression equation x on y is:

$$x = a + by$$

Normal equations are:

$$\Sigma x = Na + b\Sigma y$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2$$

Computation of Regression Equations

x	y	x^2	y^2	xy
25	18	625	324	450
30	24	900	576	720
35	30	1225	900	1050

40 45 50 55	36 42 48 54	1600 2025 2500 3025	1296 1764 2304 2916	1440 1890 2400 2970
Σx=280	Σy=252	Σx^2 = 11900	Σy^2 = 10080	Σxy=10920

280 = 7a + 252b ....(1)

20929 = 25a + 10080 b ....(2)

Eq. 1 x 36 ⇒ 10080 = 252a + 9072b ....(3)

10920 = 252a + 10080b...........(2)

(2) x (3) ⇒ 840 = 0 + 1008 b

1008 b = 840

$$b = \frac{840}{1008} = 0.83$$

Substitute b = 0.83 in equation (1)

280 = 7a+(252 x 0.83)

280 = 7x + 209.16

7a+209.116 = 280

7a = 280-209.160

$$a = \frac{70.84}{7} = 10.12$$

Substitute a = 10.12 and b = 0.83 in regression equation

X = 10.12 + 0.83 y

Regression equation Y on X is:

y = a + bx

Normal Equations are:

$\Sigma y = Na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

252 = 7a + 280 b ....(1)

10920 = 280 a + 11900 b ....(2)

(1) x 40 ⇒ 10080 = 280a + 11200 b ....(3)

10920 = 280a + 11900 b ....(2)

(2) - (3) ⇒ 840 - 0 + 700 B

700 b = 840

$$b = \frac{840}{700} = 12$$

Substitute b = 1.2 in equation (1)

252 - 7a + (280x1.2)

252 - 7a + 336

7a + 336 - 252

7a - 252-336 = -84

$$a = \frac{-84}{7} = \text{-}12$$

Substitute a = -12 and b = 1.2 in regression equation

y = -12 + 1.2x

Qn: From the following bivariate data, you are required to:

(a) Fit the regression line Y on X and predict Y if x =20

(b) Fit the regression line X on Y and predict X if y = 10

X :	4	12	8	6	4	4	16	8
Y :	14	4	2	2	4	6	4	12

Computation of Regression Equations				
x	y	x^2	y^2	xy
4 12 8 6 4 4 16	14 4 2 2 4 6 4	16 144 64 36 16 16 256	196 16 4 4 16 36 16	56 48 16 12 16 24 64
8	12	64	144	
Σx=62	Σy=48	Σx^2 = 612	Σy^2 = 432	Σxy=332

Regression equation y on x

$$y = a + bx$$

Normal equations are:

$$\Sigma y = Na + b\Sigma y$$

$$\Sigma xy = a\Sigma \text{ x } b\Sigma x^2$$

48 = 8a + 62b ....(1)

332-62a + 612b..................(2)

eq. (1) x 62 ⇒ 2,976 = 496a + 3844 b ......(3)

eq. (2) x 8 ⇒ 2,976 = 496a + 4896 b ....(4)

eq. (3) x eq. (4) ⇒ 320 = 0 + -1052b

-1052 b = 320

$$b = \frac{320}{-1052}$$

Substitute b = 0-304 in eq (1)

48 = 8a + (62 x -0.304)

48 = 8a + -18.86

48 + 18.86 = 8a

a = 66.86

$$a = \frac{66.86}{8} = 8.36$$

Substitute a =8.36 and b= -0.304 in regression equation y on x:

y = 8.36 + -0.3042 x

y = 8.36 - 0.3042 x

==============

If x = 20, then,

y = 8.36 - (0.3042x20)

= 8.36 - 6.084

= 2.276

=======

(b) Regression equation X on Y:

X = a + by

Normal equations:

$\Sigma x = Na + b\Sigma y$

$\Sigma xy = z\Sigma y + b\Sigma y^2$

62 = 8a + 48b...............(1)

<u>332 = 48a + 432b...........(2)</u>

eq (1) x 6 ⇒ 372 = 48a + 288b .....(3)

eq (2) - (3) ⇒ -40 =0 + 144b

144b = -40

$$b = \frac{-40}{144} = -0.2778$$

Substitute b = -0.2778 in equation (1)

62 = 8a+(48 x 0.2778)

62 = 8a + -13.334

62 + 13.3344 = 8a

8a = 75.334

$$a = \frac{75.3344}{8} = 9.4168$$

Substitute a = 9.4168 and b = 0.2778 in regression equation

x = 9.4168+ -0.2778 y

x = 9.4168 + -0.2778 y

If y = 10, then

x =0.4168 - (0.2778 x 10)

x = 9.4168 - 2.778

x = 6.6388

Regression Coefficient method of computing Regression Equations

Regression equations can also be computed by the use of regression coefficients. Regression coefficient X on Y is denoted as b_{xy} and that of Y on X is denoted as by_{yx}.

Regression Equation x on y

$$x - \overline{x} = b_{xy}\left(y - \overline{y}\right)$$

$$\text{i.e } x - \overline{x} = r.\frac{\sigma_x}{\sigma_y}\left(y - \overline{y}\right)$$

Regression Equation y on x:

$$y - \bar{y} = b_{xy}\left(x - \bar{x}\right)$$

$$\text{i.e } x - \bar{y} = r.\frac{\sigma_y}{\sigma_r}\left(x - x\right)$$

Properties of Regression Coefficient

1. There are two regression coefficients. They are by_{xy} and b_{yx}
2. Both the regression coefficients must have the same signs. If one is +ve, the other will also be a +ve value.
3. The geometric mean of regression coefficients will be the coefficient of correlation. $r = \sqrt{b_{xy} b_{yx}}$
4. If x x and y are the same, then the regression coefficient and correlation coefficient will be the same.

Computation of Regression Co-efficients

Regression co-efficients can calculated in 3 different ways:

1. Actual mean method
2. Assumed mean method
3. Direct method

Actual mean method

Regression coefficient x on y (b_{xy}) = $\frac{\sum xy}{\sum y^2}$

Regression coefficient y on x (b_{yx}) = $\frac{\sum xy}{\sum x^2}$

Where $x = x - \bar{x}$

$y = y - \bar{y}$

Assumed mean method

Regression coefficient x on y (b_{xy}) $\dfrac{\sum dxdy - (\sum dx).(\sum dy)}{\sum dy^2 - (\sum dy)^2}$

Regression coefficient y on x (b_{yx}) $\dfrac{\sum dxdy - (\sum dx).(\sum dy)}{\sum dy^2 - (\sum dy)^2}$

Where dx = deviation from assumed mean of X

dy = deviation from assumed mean of Y

Direct method

Regression coefficient x on y (b_{xy}) } $\dfrac{N\sum xy - \sum x \sum y}{N\sum y^2 - (\sum x)^2}$

Regression coefficient y on x (b_{xy}) } $\dfrac{N\sum xy - \sum x \sum y}{N\sum x^2 - (\sum x)^2}$

Qn: Following information is obtained from the records of a business organization

Sales (in '000)	91	53	45	76	89	95	80	65
Advertisement Expense (Rs.in '000)	15	8	7	12	17	25	20	13

You are required to:

1. Compute regression coefficients under 3 methods
2. Obtain the two regression equations and

3. Estimate the advertisement expenditure for a sale of Rs. 1,20,000

Let x = sales

y = Advertisement expenditure

Computation of regression Coefficients under actual mean method						
x	**y**	$x-\bar{x}$	$y-\bar{y}$	**xy**	x^2	y^2
91	15	16.75	0.375	6.28	280.56	0.14
53	8	-21.65	-6.625	140.78	451.56	43.89
45	7	-29.25	-7.625	223.03	855.56	58.14
76	12	1.75	-2.625	-4.59	3.06	6.89
89	17	14.75	-2.375	35.03	217.56	5.64
95	25	20.75	10.375	215.28	430.56	107.64
80	20	5.75	5.375	30.91	33.06	28.89
65	13	-9.25	-1.625	15.03	85.56	2.64
Σx=594	Σy = 117			Σxy=661.75	Σx^2=2357.48	Σy=253.87

$$\bar{X} = \frac{\sum X}{N} = \frac{294}{8} = 74.25$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{117}{8} = 14.625$$

Regression coefficient x on y (b_{xy}) $\Big\}$ $\frac{\sum xy}{\sum y^2}$

$$= \frac{661.75}{253.87} = 2.61$$

Regression coefficient Y on X` (b_{yx}) $\Big\}$ $\frac{\sum xy}{\sum x^2}$

$$= \frac{661.75}{2357.48} = 0.28$$

Computation of Regression Coefficients under assured mean method						
x	y	x-70 (dx)	y-15 (dy)	dxdy	dx^2	dy^2
91	15	21	0	0	441	0
53	8	-17	-7	+119	289	49
45	7	-25	-8	+200	625	64
76	12	6	-3	-18	36	9
89	17	19	2	+38	361	4
95	25	25	10	+250	625	100
80	20	10	5	+50	100	25
65	13	-5	-2	+10	25	4
		Σdx=34	Σdy= -1	Σdxdy=649	Σdx^2 = 2502	Σdx^2=255

Regression coefficient x on y (b_{xy}) $= \dfrac{N\sum dxdy - \sum dx.\sum dy}{N\sum dy^2 - \left(\sum dy\right)^2}$

$$= \frac{(8x649) - (34x3)}{(8x255) - (-3)^2}$$

Regression coefficient y on x (b_{yx}) $= \dfrac{N\sum dxdy - \sum dx.\sum dy}{N\sum dx^2 - \left(\sum dx\right)^2}$

$$= \frac{(8x649) - (34x - 3)}{(8x2502) - (34)^2}$$

$$= \frac{5192 - -102}{2040 - 9}$$

$$= \frac{5294}{18860}$$

$$= 0.28$$

Computation of Regression Coefficient under direct method				
x	y	xy	x^2	y^2
91	15	1365	8281	225
53	8	424	2809	64
45	7	315	2025	49
76	12	912	5776	144
89	17	1513	7921	289
95	25	2375	9025	625
80	20	1600	6400	400
65	13	845	4225	169
Σx=594	Σy-=117	Σxy=9349594	Σx^2 =46462	Σy^2 = 1965

Regression coefficient x on y (b_{xy})

$$\frac{N\sum xy - \sum x.\sum y}{N\sum y)^2 - \left(\sum y\right)^2}$$

$$= \frac{(8x9349) - (594x117)}{(8x1965) - 117^2}$$

$$= \frac{74792 - 69498}{15720 - 13689} \quad \frac{5294}{2031}$$

$$= \quad 2.61$$

Regression coefficient y on x (b_{yx})

$$\frac{N\sum xy - \sum x\sum y}{N\sum x^2 - \left(\sum x\right)^2}$$

$$= \frac{(8x9349) - (594 - 117)}{(8x46462 - 594)^2}$$

$$= \frac{74792 - 69498}{371696 - 352836}$$

$$= \frac{5294}{18860} = 0.28$$

3) a) Regression equation X on Y

$$\left(x - \bar{x}\right) = b_{xy}\left(y - \bar{y}\right)$$

$$\left(x - 74.25\right) = 2.61\left(\bar{y} - 14.625\right)$$

(x-74.25) = 2.61 y - 38.17

x = 74.25 - 38.17 + 2.61y

x = 36.08 + 2.61 y

b) Regression equation y pm x

$$\left(y - y\right) = b_{yx}\left(x - \bar{x}\right)$$

(y-14.625) = 0.28 (x - 74.25)

y = 14.625 = 0.28 x -20.79

y = 14.625 - 20.79 + 0.28 x

y = -6.165 + 0. 28 x

y = 0.28 x -6. 165

4) If sales (x) is Rs. 1,20,000, then

Estimated advertisement Exp (y) = (0.28x120)-6.165

= (33.6 - 6.165)

= 27.435

i.e Rs. 27.435

Qn: In a correlation study, the following values are obtained

	x	y
Mean	65	67
Standard deviation	2.5	3.5

Coefficient of correlation 0.8

Find the regression equations

Slo: Regression equation x on y is :

$$x - \bar{x} = b_{xy}\left(y - \bar{y}\right)$$

$$x - \bar{x} = r\,\frac{\sigma_x}{\sigma_y}\left(y - \bar{y}\right)$$

$$x - 65 = 0.8x\,\frac{2.5}{3.5}\left(y - 67\right)$$

x - 65 = 0.5714y-38.2838

x - 65-38.2838 + 0.5714y

x = 26.72 + 0.5714y

<u>x = 26.72 + 0.5714y</u>

Regression equation y on x is:

$$x - \bar{y} = b_{yx}\left(x - \bar{x}\right)$$

$$y - \bar{y} = r.\frac{\sigma_y}{\sigma_x}\left(x - x\right)$$

$$x - 67 = 0.8x\,\frac{3.5}{2.5}\left(x - 65\right)$$

y - 67 = 1.12 (x-65)

y = 67 - (1.12 x 65) = 1.12 x

y = 67.72.8 + 1.12 x

y = -5.8 + 1.12x

<u>y = 1.12x - 5.8</u>

Qn: Two variables gave the following data

$\bar{x} = 20, \quad \sigma_x = 4, \quad r = 0.7$

$\bar{y} = 15, \quad \sigma_y = 3$

Obtain regression lines and find the most likely value of y when x = 24

Sol: Regression Equation x on y is

$$\left(x - \bar{x}\right) = b_{xy}\left(y - \bar{y}\right)$$

$$x - \bar{x} = r.\frac{\sigma_x}{\sigma_y}\left(y - \bar{y}\right)$$

$$\left(x - 20\right) = 0.7x\,\frac{4}{3}\left(y - 15\right)$$

$$\left(x - 20\right) = \frac{2.8}{3}\left(y - 15\right)$$

(x -20) = 0.93(y-15)

x-20 = 0.93y - 13.95

x = 20 - 13.95 + 0.93y

<u>x = 6.05 + 0.93 y</u>

Regression Equation y on x is

$$\left(y - \bar{y}\right) = b_{yx}\left(x - \bar{x}\right)$$

$$y - \bar{y} = r.\frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right)$$

$$y - \overline{y} = 0.7x\,\frac{3}{4}(x - 20)$$

(y-15) = 0.525 (x - 20)

y - 15 = 0.525 x -10.5

y = 15-10.5 + 0.515x

<u>y = 4.5 + 0.525 x</u>

If X = 24, then

y = 4.5 + (0.525 x 24)

y = 4.5 + 12.6

<u>y = 17.1</u>

Qn: For a given set of bivariate data, the following results were obtained:

Sol: Regression Equation y on x is:

$$\left(y - \overline{y}\right) = b_{yx}\left(x - \overline{x}\right)$$

(y - 27.9) = -1.5 (x -53.2)

(y - 27.9) = -1.5x + 79.8

y = 79.8 + 27.9 - 1.5 x

<u>y = 107.7 - 1.5 x</u>

If x = 6, then

y = 107.7 - (1.5 x 60)

= 107.7.90

= <u>17.7</u>

r $= \sqrt{b_{xy} x\, b_{yx}}$

$$= \sqrt{1.5 \text{ x } 0.2} = -\sqrt{30} = -0.5477$$

	Correlation	Regression
1	It studies degree of relationship between variables	It studies the nature of relationship between variables
2	It is not used for prediction purposes	It is basically used for prediction purposes
3	It is basically used as a tool for determining the degree of relationship	It is basically used as a tool for studying cause and effect relationship
4	There may be nonsense correlation between two variables	There is no such nonsense regression
5	There is no question of dependent and independent variables	There must be dependent and independent variables

Module 4

THEORY OF PROBABILITY

INTRODUCTION

Probability refers to the chance of happening or not happening of an event. In our day today conversations, we may make statements like "probably he may get the selection". "possibly the Chief Minister may attend the function" etc. Both the statements contain an element of uncertainly about the happening of the even. Any problem which contains uncertainty about the happening of the event is the problem of probability.

Definition of Probability

The probability of given event may be defined as the numerical value given to the likely hood of the occurrence of that event. It is a number lying between '0' and '1' '0' denotes the even which cannot occur, and '1' denotes the event which is certain to occur. For example, when we toss on a coin, we can enumerate all the possible outcomes (head and tail), but we cannot say which one will happen. Hence, the probability of getting a head is neither 0 or 1 but between 0 and 1. It is 50% or ½.

Terms use in Probability

A random experiment is an experiment that has two or more outcomes which vary in an unpredictable manner from trial to trail when conducted under uniform conditions.

In a random experiment, all the possible outcomes are known in advance but none of the outcomes can be predicted with certainty. For example, tossing of a coin is a random experiment because it had two outcomes (head and tail), but we cannot predict any of them which certainty.

Sample point

Every indecomposable outcome of a random experiment is called a sample point. It is also called simple event or elementary outcome.

Eg: When a die is thrown, getting '3' is a sample point.

Sample Space

Sample space of a random experiment is the set containing all the sample points of that random experiment.

Eg: When a coin is tossed, the sample space is (Head, Tail)

Event

An event is the result of a random experiment. It is a subset of the sample space of a random experiment.

Sure Event (Certain Event)

An event whose occurrence is inevitable is called sure even.

Eg: Getting w white ball from a box containing all while balls.

Impossible Events

An event whose occurrence is impossible, is called impossible event. Eg:- Getting a white ball from a box containing all red balls.

Uncertain Events

An event whose occurrence is neither sure nor impossible is called uncertain event.

Eg: Getting a white ball from a box containing white balls and black balls.

Equally likely Events

Two events are said to be equally likely if anyone of them cannot be expected to occur in preference to other. For example, getting herd and getting tail when a coin is tossed are equally likely events.

Mutually exclusive events

A set of events are said to be mutually exclusive of the occurrence of one of them excludes the possibility of the occurrence of the others.

Exhaustive Events:

A group of events is said to be exhaustive when it includes all possible outcomes of the random experiment under consideration.

Dependent Events:

Two or more events are said to be dependent if the happening of one of them affects the happening of the other.

PERMUTATIONS

Permutation means arrangement of objects in a definite order. The number of arrangements (permutations) depends upon the total number of objects and the number of objects taken at a time for arrangement.

The number of permutations is calculated by using the following formula:

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

! = Factorial

n = Total number of objects

r = Number of objects taken at a time for arrangement

If whole the objects are taken at a time for arrangement, then number of permutations is calculated by using the formula:

$$^{n}P_{r} = {}^{n}P_{n}$$

$$= \frac{\text{n!}}{(\text{n - n})!}\frac{\text{n!}}{\text{o!}} = \frac{\text{n!}}{1} = \text{n!}$$

$$^{n}P_{n} = n!$$

Question:-

A factory manager purchased 3 new machines, A, B and C. How many number of times he can arrange the machines?

Solution :

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

$n = 3$

$r = 3$

Here *‘n’* and *‘r’* are same

$$\therefore {}^{\text{n}}P_{r} = n!$$

$$= 3! = 3 \text{ x } 2 \text{ x } 1 = 6 \text{ !}$$

Question :

In how many ways 3 people be seated on a bench if only two seats are available.

Solution

$$^{n}P_{r} = \frac{n!}{(\text{n - r})!}$$

$n = 3$

$r = 2$

$$\therefore 3P_2 \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{3\times 2\times 1}{1} = 6\ ways$$

Computation of Permutation when objects are alike:-

Sometimes, some of the objects of a group are alike. In such a situation number of permutations is calculated as:-

$$^{n}P_{r} = \frac{n!}{n_1!n_2!n_3!......n_n!}$$

n_1 = number of alike objects in first category

n_2 = number of alike objects in second category

If all items are alike, you know, they can be arranged in only one order

ie, $^{n}P_{r} = \frac{n!}{n_1!}$

Question:

Find the number of permutations of letters in the word 'COMMUNICATION'

Solution

$$^{n}P_{r} = \frac{n!}{n_1!n_2!n_3!......}$$

C = 2, 0 = 2; M = 2; U = 1; N= 2;

1 = 2; A = 1; R = 1

$$\therefore {}^{n}P_{r} = \frac{13!}{2!2!2!1!2!2!1!1!}$$

$$= \frac{13\times12\times11\times10\times9\times8\times7\times6\times5\times4\times3\times2\times1}{2\times1\times2\times1\times2\times1\times1\times2\times1\times2\times1\times1\times1}$$

= **194594400 times**

COMBINATIONS

Combination means selection or grouping of objects without considering their order. The number of combinations is calculated by using the following formula:

$$^{n}C_{r} = \frac{n!}{(n-r)!\,r!}$$

Question:

A basket contains 10 mangoes. In how many ways 4 mangoes from the basket can be selected?

Solution

$$^{n}C_{r} = \frac{n!}{(n-r)!\,r!}$$

n = 10

r = 4

$$^{10}C_{4} = \frac{10!}{(10-4)!\,4!} = \frac{10!}{6!\,4!} = \frac{10\times9\times8\times7}{4\times3\times2\times1}$$

= **210 ways**

Question:

How many different sets of 5 students can be chosen out of 20 qualified students to represent a school in an essay context?

Solution

$$^{n}C_{r} = \frac{n!}{(n-r)!\,r!}$$

n = 20

r = 5

$$^{20}C_5 = \frac{20!}{(20-5)!5!} = \frac{20!}{15!5!} = \frac{20\times19\times18\times17\times16}{4\times3\times2\times1}$$

= **15504 Sets**

DIFFERENT SCHOOLS OF THOUGHT ON PROBABILITY

Different Approaches/Definitions of Probability

There are 4 important schools of thought on probability:-

1. Classical or Priori Approach
2. Relative frequency or Empirical Approach

} Objective Probability Approach (1 and 2)

3. Subjective or Personalistic Approach
4. Modern or Axiomatic Approach

1. Classical or Priori Approach

If out of 'n' exhaustive, mutually exclusive and equally likely outcomes of an experiment; 'm' are favourable to the occurrence of an event 'A', then the probability of 'A' is defined as to be $\frac{m}{n}$

$$P(A) = \frac{m}{n}$$

According to Laplace, a Fench Mathematician, "the probability is the ratios of the number of favourable cases to the total number of equally likely cases."

$$P(A) = \frac{\text{Number of favourable cases}}{\text{Total number of equality likely cases}}$$

Question:

What is the chance of getting a head when a coin is tossed?

Solution

Total number of cases = 2

No. of favourable cases = 1

Probability of getting head - $\frac{1}{2}$

Question:

A die is thrown. Find the probability of getting

(1) A '4'

(2) an even number

(3) '3' or '5'

(4) less than '3'

Solution

Sample space is (1, 2, 3, 4, 5, 6)

(1) Probability (getting '4) = $\frac{1}{6}$

(2) Probability (getting an even number) = $\frac{3}{6} = \frac{1}{6}$

(3) Probability (getting 3 or 5) = $\frac{2}{6} = \frac{1}{3}$

(4) Probability (getting less than '3') = $\frac{2}{6} = \frac{1}{3}$

Question:

A ball is drawn from a bag containing 4 white, 6 black and 5 yellow balls. Find the probability that a ball drawn is :-

(1) White (2) Yellow (3) Black (4) Not yellow
(5) Yellow or white

Solution

(1) P (drawing a white ball) = $\frac{4}{15}$

(2) P (drawing a yellow ball) = $\frac{5}{15} = \frac{1}{3}$

(3) P (drawing a black ball) = $\frac{6}{15} = \frac{2}{5}$

(4) P (drawing not a yellow ball) = $\frac{10}{15} = \frac{2}{3}$

(5) P (drawing a yellow or white ball) = $\frac{9}{15} = \frac{3}{5}$

Question:

There are 19 cards numbered 1 to 19 in a box. If a person drawn one at random, what is the probability that the number printed on the card be an even number greater than 10?

Solution

$$\left.\begin{array}{l}\text{The even numers greater than 10 are } 12, 14, 16 \text{ and } 18 \\ \therefore \text{P(drawing a card with an even number greater than 10)}\end{array}\right\} = \frac{4}{9}$$

Question:

Two unbiased dice are thrown. Find the probability that:-

(a) Both the dice show the same number

(b) One die shows 6

(c) First die shows 3

(d) Total of the numbers on the dice is 9

(e) Total of the numbers on the dice is greater than 8

(f) A sum of 11

Solution

When 2 dice are thrown the sample space consists of the following outcomes:-

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,3)	(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)

(a) P (that both the dice shows the same number)= $\frac{6}{36} = \frac{1}{6}$

(b) P (that one die shows 6)= $\frac{10}{36} = \frac{5}{18}$

(c) P (that first die shows 3)= $\frac{6}{36} = \frac{1}{6}$

(d) P (that total of the number on the dice is 9) = $\frac{4}{36} = \frac{1}{9}$

(e) P (that total of the number is greater than 8)= $\frac{10}{36} = \frac{5}{18}$

(f) P (that a sum of 11) = $\frac{2}{36} = \frac{1}{18}$

Problems based on combination results

Question :

A box contains 6 white balls and 4 green balls. What is the probability of drawing a green ball?

Solution

Probable number of cases = ${}^{4}C_{1}$

Total number of cases = ${}^{10}C_1$

P (drawing a green ball) $= \frac{{}^{4}C_1}{{}^{10}C}$

$$= \frac{\frac{4!}{(4-1)!\times 1!}}{\frac{10!}{(10-1)!\times 1!}}$$

$$= \frac{\frac{4!}{3!\times 1!}}{\frac{10!}{9!\times 1!}}$$

$$= \frac{\frac{4\times 3\times 2\times 1}{3\times 2\times 1\times 1}}{9\times 8\times 7\times 6\times 5\times 4\times 3\times 2\times 1\times 1}$$

$$= \frac{4}{10} = \frac{2}{5}$$

Question

What is the probability of getting 3 red balls in a draw of 36 balls from a bag containing 5 red balls and 46 black balls?

Solution

Favourable number of cases = ${}^{5}C_3$

Total number of cases = ${}^{9}C_3$

P (getting 3 white balls) = $\frac{{}^{5}C_3}{{}^{9}C_3}$

$$= \frac{\frac{5!}{(5-3)!3!}}{\frac{9!}{(9-3)!3!)}} = \frac{\frac{5!}{2!3!}}{\frac{9!}{6!3!}}$$

$$= \frac{\frac{5\times 4}{2}}{\frac{9\times 8\times 7}{3\times 2\times 1}} = \frac{\frac{20}{2}}{3\times 4\times 7} = \frac{10}{84}$$

$$= \frac{5}{42}$$

Question

A committee is to be constituted by selecting three people at random from a group consisting of 5 Economists and 4 Statisticians. Find the probability that the committee will consist of :

(a) 3 Economists
(b) 3 Statisticians
(c) 3 Economists and 1 Statistician
(d) 1 Economist and 2 Statistician

(a) P (Selecting 3 Economists) $= \frac{^5C_3}{^9C_3}$

$$= \frac{\frac{5!}{(5-3)!3!}}{\frac{9!}{(9-3)!3!}} = \frac{\frac{5!}{2!3!}}{\frac{9!}{6!3!}}$$

$$\frac{\frac{5\times 4}{2\times 1}}{\frac{9\times 8\times 7}{3\times 2\times 1}} = \frac{10}{84} = \frac{5}{42}$$

(b) P (Selecting 3 Statisticians) = $\frac{^4C_3}{^9C_3}$

$$\frac{\frac{4!}{(4-3)!3!}}{\frac{9!}{(9-3)!3!}} = \frac{\frac{4!}{1!3!}}{\frac{9!}{6!3!}}$$

$$\frac{4}{84} = \frac{1}{21}$$

(c) P (Selecting 2 Economists and 1 Statistician) $= \dfrac{{}^5C_2 \times {}^4C_1}{{}^9C_3}$

$$= \frac{\dfrac{5!}{(5-2)!2!} \times \dfrac{4!}{(4-1)!1!}}{\dfrac{9!}{(9-3)!3!}}$$

$$= \frac{\dfrac{5!}{3!2!} \times \dfrac{4!}{3!1!}}{6!3!}$$

$$= \frac{\dfrac{5\times4}{2\times1} \times \dfrac{4}{1}}{\dfrac{9\times8\times7}{3\times2\times1}} = \frac{10\times4}{84}$$

$$= \frac{40}{84} = \frac{10}{21}$$

(d) P (Selecting 1 Economist and 2 Statistician $= \dfrac{{}^5C_1 \times {}^4C_2}{{}^9C_3}$

$$= \frac{\dfrac{5!}{(5-1)!1!} \times \dfrac{4!}{(4-2)!2!}}{\dfrac{9!}{(9-3)!3!}}$$

$$= \frac{5\times6}{84} = \frac{30}{84} = \frac{5}{14}$$

Questions:

A committee of 5 is to be formed from a group of 8 boys and 7 girls. Find the probability that the committee consists of at least one girl.

Solution

P (that committee consists of at least one girl) = P(one girl & 4 boys) or P(2 girls & 3 boys) or P (3 girls & 2 boys) or P (4 girls & 1 boy) or P(5 girls)

$$= \frac{\left({}^7C_1 \times {}^8C_4\right) + \left({}^7C_2 \times {}^8C_3\right) + \left({}^7C_3 \times {}^8C_2\right) + \left({}^7C_4 \times {}^8C_1\right) + {}^7C_5}{{}^{15}C_5}$$

$$= \frac{\left(\frac{7!}{6!1!} \times \frac{8!}{4!4!}\right) + \left(\frac{7!}{5!2!} \times \frac{8!}{5!3!}\right) + \left(\frac{7!}{4!3!} \times \frac{8!}{6!2!}\right) + \left(\frac{7!}{3!1!} \times \frac{8!}{7!1!}\right) + \frac{7!}{2!5!}}{\frac{15!}{10!5!}}$$

$$= \frac{\left(7 \times \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2}\right) + \left(\frac{7 \times 6}{2} \times \frac{8 \times 7 \times 6}{3 \times 2}\right) + \left(\frac{7 \times 6 \times 5!}{3 \times 2} \times \frac{8 \times 7}{2}\right) + \left(\frac{7 \times 6 \times 5}{3 \times 2} \times 8\right) + \frac{7 \times 6}{2}}{\frac{15 \times 14 \times 13 \times 12 \times 11}{5 \times 4 \times 3 \times 2}}$$

$$= \frac{(7 \times 70) + (21 \times 56) + (35 \times 28) + (35 \times 8) + 21}{3003}$$

$$= \frac{490 + 1176 + 980 + 280 + 21}{3003} = \frac{2947}{3003}$$

This problem can be solved in the following method also.

P(that the committee consists of at least one = 1 – P(that the committee consists of all boys girl)

$$= 1 - \left(\frac{{}^{8}C_{5}}{{}^{15}C_{5}} \right)$$

$$= 1 - \frac{\frac{8!}{3!5!}}{\frac{15!}{10!5!}} = 1 - \frac{\frac{8 \times 7 \times 6}{3 \times 2}}{\frac{15 \times 14 \times 13 \times 12 \times 11}{5 \times 4 \times 3 \times 2}}$$

$$= 1 - \frac{56}{3003}$$

$$= \frac{3003 - 56}{3003} = \frac{2947}{3003}$$

Limitations of Classical Definition:

1. Classical definition has only limited application in coin-tossing die throwing etc. It fails to answer question like "What is the probability that a female will die before the age of 64?"
2. Classical definition cannot be applied when the possible outcomes are not equally likely. How can we apply classical definition to find the probability of rains? Here, two possibilities are "rain" or "no rain". But at any given time these two possibilities are not equally likely.
3. Classical definition does not consider the outcomes of actual experimentations.

Relative Frequency Definition or Empirical Approach

According to Relative Frequency definition, the probability of an event can be defined as the relative frequency with which it occurs in an indefinitely large number of trials.

If an even 'A' occurs 'f' number of trials when a random experiment is repeated for 'n' number of times, then

P(A) = $\underset{n \to \alpha}{Lt} \dfrac{f}{n}$

For practical convenience, the above equation may be written as $P(A) = \dfrac{f}{n}$

Here, probability has between 0 and 1,

i.e. $0 \le P(A) \le 1$

Question

The compensation received by 1000 workers in a factory are given in the following table:-

Wages:	80-100	100-120	120-140	140-160	160-180	180-200
No.of workers	10	100	400	250	200	40

Find the probability that a workers selected has

(1) Wages under Rs.100/-

(2) Wages above Rs.140/-

(3) Wages between Rs.120/- and Rs.180/-

Solution

P(that a worker selected has wages under Rs.140/-)

$$= \frac{10+100+400}{1000} = \frac{510}{1000}$$

P(that a worker selected has wages above Rs.10/-

$$= \frac{250+200+40}{1000} = \frac{490}{1000}$$

P(that a worker selected has wages between 1 20 and 180

$$= \frac{400+250+200}{1000} = \frac{850}{1000}$$

Subjective (Personalistie) Approach to Probability

The exponents of personalistie approach defines probability as a measure of personal confidence or belief based on whatever evidence is available. For example, if a teacher wants to find out the probability that Mr. X topping in M.Com examination, he may assign a value between zero and one according to his degree of belief for possible occurrence. He may take into account such factors as the past academic performance in terminal examinations etc. and arrive at a probability figure. The probability figure arrived under this method may vary from person to person. Hence it is called subjective method of probability.

Axiomatic Approach (Modern Approach) to Probability

Let 'S' be the sample space of a random experiment, and 'A' be an event of the random experiment, so that 'A' is the subset of 'S'. Then we can associate a real number to the event 'A'. This number will be called probability of 'A' if it satisfies the following three axioms or postulates:-

(1) The probability of an event ranges from 0 and 1.

If the event is certain, its probability shall be 1.

If the event cannot take place, its probability shall be zero.

(2) The sum of probabilities of all sample points of the sample space is equal to 1, i.e., P(S) = 1.

(3) If A and B are mutually exclusive (disjoint) events, then the probability of occurrence of either A or B shall be:

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

THEOREMS OF PROBABILITY

There are two important theorems of probability. They are:

1. Additional Theorem 2. Multiplication Theorem.

Addition Theorem

Here, there are 2 situations.

(a) Events are mutually exclusive.

(b) Events are not mutually exclusive.

(a) Addition theorem (Mutually Exclusive Events)

If two events, 'A' and 'B', are mutually exclusive the probability of the occurrence of either 'A' or 'B' is the sum of the individual probability of A and B.

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

$$\text{i.e., } P(A \cup B) = P(A) + P(B)$$

(b) Addition theorem (Not mutually exclusive events)

If two events, A and B are not mutually exclusive the probability of the occurrence of either A or B is the sum of their individual probability minus probability for both to happen.

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

$$\text{i.e., } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Question

What is the probability of picking a card that was red or black?

Solution

Here the events are mutually exclusive

$$P\text{ (picking a red card)} = \frac{26}{52}$$

$$P\text{ (picking a black card)} = \frac{26}{52}$$

$$\therefore P\text{ (picking a red or black card)} = \frac{26}{52} + \frac{26}{52} = 1.$$

Question

The probability that a contractor will get a plumping contract is 2/3 and the probability that he will not get an electric 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?

Solution:

P (getting plumbing contract) = $\frac{2}{3}$

P (not getting electric contract) = $\frac{5}{9}$

P (getting electric contract) = $1 - \frac{5}{9} = \frac{4}{9}$

P (getting at least one contract) = P (getting electric contract) + P (getting plumbing contract) – P (getting both)

i.e., $\frac{4}{5} = \frac{4}{9} + \frac{2}{3} -$ P (getting both)

$\therefore$ P (getting both contracts) $= \frac{4}{9} + \frac{2}{3} + \frac{4}{5}$

$= \frac{20 + 30 - 36}{45} = \frac{14}{45}$

Question

If P(A) = 0.5(B) = 0.6, P(A $\cap$ B) = 0.2, find:-

(a) P(A $\cup$ B)

(b) P(A')

(c) P(A $\cap$ B)

(d) P(A' $\cap$ B')

Solution

Here the events are not mutually exclusive:-

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

$= 0.5 + 0.6 – 0.2$

$= 0.9$

(b) $P(A') = 1 - P(A)$

$= 1 – 0.5$

$= 0.5$

(c) $P(A \cap B')$

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

$= 0.5 – 0.2$

$= 0.3$

(d) $P(A' \cap B') = 1 - P(A \cap B)$

$= 1 – [P(A) + P(B) - P(A \cap B)$

$= 1 – (0.5 + 0.6 – 0.2)$

$= 1 – 0.9$

$= 0.1$

MULTIPLICATION THEOREM

Here there are two situations:

(a) Events are independent

(b) Events are dependent

(a) Multiplication theorem (independent events)

If two events are independent, then the probability of occurring both will be the product of the individual probability.

$P(A \text{ and } B) = P(A). P(B)$

i.e., $P(A \cap B) = P(A). P(B)$

Question

A bag contains 5 white balls and 8 black balls. One ball is drawn at random from the bag and is then replaced.

Again another one is drawn. Find the probability that both the balls are white.

Solution

Here the events are independent

P(drawing white ball in I draw) = $\frac{5}{13}$

P(drawing white ball in II draw) = $\frac{5}{13}$

$\therefore$ P (drawing white ball in both draw) = $\frac{5}{13} \times \frac{5}{13}$

$= \frac{25}{169}$

Question

Single coin is tossed for three tones. What is the probability of getting head in all the 3 times?

Solution

P(getting head in all the 3 times) = P(getting H in 1st toss) $\times$ P(getting Head in 2nd toss) $\times$ P(getting H in 3rd toss)

$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$

$= \frac{1}{8}$

(b) Multiplication theorem (Dependent Events):-

If two events, A and B are dependent, the probability of occurring 2nd event will be affected by the outcome of the first.

P(A $\cap$ B) = P(A). P(B/A)

Question

A bag contains 5 white balls and 8 black balls. One ball is drawn at random from the bag. Again, another one is drawn without replacing the first ball. Find the probability that both the balls drawn are white.

Solution

P(drawing a white ball in 1st draw) = $\frac{5}{13}$

P(drawing a white ball in IInd draw = $\frac{4}{12}$

$= \frac{20}{156}$

Question

The Probability that 'A' solves a problem in Maths is $\frac{2}{5}$ and the probability that 'B' solves it is $\frac{3}{8}$. If they try independently find the probability that:-

(i) Both solve the problem

(ii) at last one solve the problem

(iii) none solve the problem.

Solution

(i) P(that both solve the problem = P(that A

(i) P(that both solve the problem) } =P(that A solve the problem $\times$P(that B solves the problem)

$= \frac{2}{5} \times \frac{3}{8} = \frac{6}{40} = \frac{3}{20}$

(ii) P(that at least one solve the problem = P(that A or B solves the problem)

= P(A solve the problem

+ P(B solve the problem – P(A and B solve the problem)

$= \frac{2}{5}+\frac{3}{8}-\left(\frac{2}{5}\times\frac{3}{8}\right)$

$=\frac{2}{5}+\frac{3}{8}-\frac{6}{40}$

$= \frac{16+15-6}{40}$

$= \frac{25}{40}=\frac{5}{8}$

(iii) P(that none solve the problem) = 1-P(at least one solve the problem)

= 1-P(A or B solve the problem)

= 1-[P(A sole then Problem)+P(B solve the problem) – P(A & B solve the problem)]

$= 1-\left[\frac{2}{5}+\frac{3}{8}-\left(\frac{2}{5}\times\frac{3}{8}\right)\right]$

$= 1-\left(\frac{2}{5}+\frac{3}{8}-\frac{6}{40}\right)$

$= 1-\left(\frac{16+15-6}{40}\right)$

$= 1-\frac{25}{40}=\frac{15}{40}=\frac{3}{8}$

Question

A university has to select an examiner from a list of 50 persons. 20 of them are women and 30 men. 10 of them know Hindi and 40 do not. 15 of them are teachers and remaining are not. What is the probability that the university selecting a Hindi knowing woman teacher?

Solution

Here the events are independent.

P(selecting Hindi knowing woman teacher } = P(selecting Hindi knowing person, woman and teacher

P(selecting Hindi knowing persons $= \frac{10}{50}$

P(selecting woman) $= \frac{20}{50}$

(P(selecting teacher) $= \frac{15}{50}$

P(selection Hindi knowing woman teacher $= \frac{10}{50} \times \frac{20}{50} \times \frac{15}{50}$

$= \frac{2}{10} \times \frac{4}{10} \times \frac{3}{10} = \frac{24}{1000}$

$= \frac{3}{125}$

Question

'A' speaks truth in 70% cases and 'B' in 85% cases. In what percentage of cases they likely to contradict each other in stating the same fact?

Let P(A) = Probability that 'A' speaks truth

P(A') = Probability that 'A' does not speak truth

P(B) = Probability that 'B' speaks truth

P(B') = Probability that 'B' does not speak truth

P(A) = 70% = 0.7

P(A') = 30% = 0.3

P(B) = 85% = 0.85

P(B') = 15% = 0.15

$\therefore$ P (A and B contradict each other) = P('A' speaks truth and 'B' does not OR, A does not speak truth & B speaks

$= P(A\&B') \cup (A'\&B)$

$= (0.7 \times 0.15) + (0.3 \times 0.85)$

$= 0.105 + 0.255$

$= 0.360$

$\therefore$ Percentage of cases in which A and B contradict each other $= 0.360 \times 100$

= **36%**

Question

20% of students in a university are graduates and 80% are undergraduates. The probability that graduate student is married is 0.50 and the probability that an undergraduate student is married is 0.10. If one student is selected at random, what is the probability that the student selected is married?

P(selecting a married student) = P(selecting a graduate married student or selecting an undergraduate married student)

= P(selecting a graduate &

married OR selecting un undergraduate & married)

$= (20\% \times 0.50)+(80\% \times 0.10)$

$= \left(\frac{20}{100} \times 0.50\right) + \left(\frac{80}{100} \times 0.10\right)$

$= (0.2 \times 0.50)+(0.8 \times 0.1)$

$= 0.1+0.08$

$=$ **<u>0.18</u>**

Question

Two sets of candidates are competing for the position on the board of directors of a company. The probability that the first and second sets will win are 0.6 and 0.4 respectively. If the first set wins, the probability of introducing a new product is 0.8 and the corresponding probability if the second set wins is 0.3, What is the probability that the new product will be introduced?

Solution

P(that new product will be introduced } = P(that new product is introduced by first set OR the new product is introduced by second set)

= P(Ist set wins & Ist introduced the new produced OR IInd set wins the new product)

$= (0.6 \times 0.8)+(0.4 \times 0.3)$

$= 0.48+0.12$

$=$ **<u>0.60</u>**

Question:

A certain player say Mr. X is known to win with possibility 0.3

if the truck is fast and 0.4 if the track is slow. For Monday there is a 0.7 probability of a fast track and 0.3 probability of a slow track. What is the probability that Mr. X will win on Monday?

Solution:

P(X will won on Monday) = P(to win in fast track OR to win in slow track)

= P (to get fast track & to win OR to get slow track & to win)

= (0.7x0.3) + (0.3 x 0.4)

= 0.21 + 0.12

= 0.33

CONDITIONAL PROBABILITY

Multiplication theorem states that if two events, A and B are dependent events then, the probability of happening both will be the product of P(A) and P(B/A).

i.e., P(A and B) or $P(A \cap B)$ = P(A.P(B/A)

Here, P(B/A) is called Conditional probability

$$P(A \cap B) = P(A).P\left(B/A\right)$$

$$\text{i.e., } P(A).P\left(B/A\right) = P(A \cap B)$$

$$\therefore P\left(B/A\right) = \frac{P(A \cap B)}{P(A)}$$

$$\text{Similarly, } P\left(A/B\right) = \frac{P(A \cap B)}{P(B)}$$

If 3 event, A, B and C and dependent events, then the probability of happening A, B and C is:-

$$P(A \cap B \cap C) = P(A).P\left(\tfrac{B}{A}\right)P\left(\tfrac{C}{AB}\right)$$

i.e., $$P(A).P\left(\tfrac{B}{A}\right)P\left(\tfrac{C}{AB}\right) = P(A \cap B \cap C)$$

$$P\left(\tfrac{C}{AB}\right) = \frac{P(A \cap B \cap C)}{P(A).P\left(\tfrac{B}{A}\right)}$$

Question

If P(A)= $\frac{1}{13,}$ P(B)= $\frac{1}{4,}$ and P(A $\cap$ B = $\frac{1}{52,}$ find:-

(a) P(A/B)

(b) P(B/A)

Solution

Here we know the events are dependent

(a) $$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{52}}{\frac{1}{4}} = \frac{1}{52} \times \frac{4}{1} = \frac{4}{52} = \frac{1}{13}$$

(b) $$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{52}}{\frac{1}{13}} = \frac{1}{52} \times \frac{13}{1} = \frac{13}{52} = \frac{1}{4}$$

Inverse Probability

If an event has happened as a result of several causes, then we may be interested to find out the probability of a particular cause of happening that events. This type of problem is called inverse probability.

Baye's theorem is based upon inverse probability.

BAYE'S THEOREM:

Baye's theorem is based on the proposition that probabilities should revised on the basis of all the available information. The revision of probabilities based on available information will help to reduce the risk involved in decision-making. The probabilities before revision is called priori probabilities and the probabilities after revision is called posterior probabilities.

According to Baye's theorem, the posterior probability of event (A) for a particular result of an investigation (B) may be found from the following formula:-

$$P(A/B) = \frac{P(A).P(B)}{P(A).P(B) + P(NotA).P(\frac{B}{Not}A)}$$

Steps in computation

1. Find the prior probability
2. Find the conditional probability
3. Find the joint probability by multiplying step 1 and step 2
4. Find posterior probability as percentage of total joint probability

Question

A manufacturing firm produces units of products in 4 plants, A, B, C and D. From the past records of the proportions of defectives produced at each plant, the following conditional probabilities are set:-

A: 0.5; b: 0.10; C: 0.15 and D: 0.02

The first plant produces 30% of the units of the output, the second plant produces 25%, third 40% and the forth 5%

A unit of the products made at one of these plants is tested and is found to be defective. What is the probability that the unit was produced in Plant C.

Solution

Computation of Posterior Probabilities				
Machine	**Priori Probability**	**Conditional Probabilities**	**Joint Probability**	**Posterior Probability**
A	0.30	0.05	0.015	$\frac{0.015}{0.101} = 0.1485$
B	0.25	0.10	0.025	$\frac{0.025}{0.101} = 0.2475$
C	0.40	0.15	0.060	$\frac{0.060}{0.101} = 0.5941$
D	0.05	0.02	0.001	$\frac{0.001}{0.101} = 0.0099$
			0.101 ======	1.0000 =======

Probability that defective unit was produced in Machine C=0.5941

Question

In a bolt manufacturing company machine I, II and III manufacture respectively 25%, 35% and 40%. Of the total of their output, 5%, 4% and 2% are defective bolts. A bolt is drawn at random from the products and is found to be defective. What are the probability that it was manufactured by:-

(a) Machine I

(b) Machine II

(c) Machine III

Solution

Computation of Posterior Probabilities				
Machine	Priori Probability	Conditional Probabilities	Joint Probability	Posterior Probability
I	0.25	0.05	0.0125	0.362
II	0.35	0.04	0.0140	0.406
III	0.40	0.02	0.0080	0.232
			0.0345	1.000

P(that the bolt was manufactured by Machine I) = 0.362

P(that the bolt was manufactured by Machine II) = 0.406

P(that the bolt was manufactured by Machine III) = 0.232

Question

The probability that a doctor will diagnose a particular disease correctly is 0.6. The probability that a patient will die by his treatment after correct diagnosis is 0.4 and the probability of death by wrong diagnosis is 0.7. A patient of the doctor who had the disease died. What is the probability that his disease was not correctly diagnosed?

Solution

Computation of Posterior Probabilities				
Nature of Diagnosis	Priori Probability	Conditional Probabilities	Joint Probability	Posterior Probability
Correct	0.6	0.4	0.24	$\frac{0.24}{0.52} = 0.1462$
Not correct	0.4	0.7	0.28	$\frac{0.28}{0.52} = 0.538$
			0.52	1.000

Probability that the disease was not correctly diagnosed= 0.538

Question

There are two Urns, one containing 5 white balls and 4 black balls; and the other containing 6 white balls and 5 black balls. One Urn is chosen and one ball is drawn. If it is white, what is the probability that the Urn selected is the first?

Solution

Computation of Posterior Probabilities				
No. of Urn	**Probability of drawing white ball (Priori Probability)**	**Condi-tional Probabi-lities**	**Joint Probability**	**Posterior Probability**
Correct	$\frac{5}{9}$	$\frac{1}{2}$	$\frac{5}{18} = 0.2778$	$\frac{0.2778}{0.5505} = 0.5046$
Not correct	$\frac{6}{11}$	$\frac{1}{2}$	$\frac{6}{22} = 0.2727$ --------- 0.5505	$\frac{0.2727}{0.5505} = 0.4954$ ---------- 1.000

P(that the white bals drawn is from Urnn 1 = 0.5046

======

CHAPTER 5
PROBABILITY DISTRIBUTION
(THEORETICAL DISTRIBUTION)

DEFINITION

Probability distribution (Theoretical Distribution) can be defined as a distribution obtained for a random variable on the basis of a mathematical model. It is obtained not on the basis of actual observation or experiments, but on the basis of probability law.

Random Variable

Random variable is a variable who value is determined by the outcome of a random experiment. Random variable is also called chance variable or stochastic variable.

For example, suppose we toss a coin. Obtaining of head in this random experiment is a random variable. Here the random variable of "obtaining heads" can take the numerical values.

Now, we can prepare a table showing the values of the random variable and corresponding probabilities. This is called probability distributions or theoretical distribution.

In the above, example probability distribution is:-

Obtaining of heads X	Probability of obtaining heads P(X)
0	$\frac{1}{2}$
1	$\frac{1}{2}$
	$\sum P(X) = 1$

Properties of Probability Distributions:

1. Every value of probability of random variable will be greater than or equal to zero.

   i.e., $P(X) \geq 0$

   i.e., $P(X) \neq$ Negative value

2. Sum of all the probability values will be 1

   $\sum P(X) = 1$

Question

A distribution is given below. State whether this distribution is a probability distribution.

X:	0	1	2	3	4
P(X)	0.01	0.10	0.50	0.30	0.90

Solution

Here all values of P(X) are more that zero; and sum of all (PX) value is equal to 1

Since two conditions, namely P(X) ≥ 0 and $\sum P(X) = 1$, are satisfied, the given distribution is a probability distribution.

MATHEMATICAL EXPECTATION

(EXPECTED VALUE)

If X is a random variable assuming values x_1, x_2, x_3, …………, x_n with corresponding probabilities P_1, P_2, P_3,……., P_n, then the operation of X is defined as $X_1P_1 + X_2P_2 + X_3P_3 +$ ……….+ X_nP_n.

$$E(X) = \sum [X.P(X)]$$

Question

A petrol pump proprietor sells on an averae Rs.80.000/-

worth of petrol on rainy days and an average of Rs.95,000 on clear days. Statistics from the meteorological department show that the probability is 0.76 for clear weather and 0.24 for rainy weather on coming Wednesday. Find the expected value of petrol sale on coming Wednesday.

$$\left.\begin{array}{r}\text{Expected Value}\\ E(X)\end{array}\right\} = \sum\left[X.P(X)\right]$$

$$= (80.000 \times 0.24)+(95000 \times 0.76)$$

$$= 19.200 + 72.200$$

$$= \text{Rs.}91{,}400$$

Question

There are three alternative proposal before a business man to start a new project:-

Proposal I :	Profit of Rs.5 lakhs with a probability of 0.6 or a loss of Rs.80,000 with a probability of 0.4
Proposal II:	Profit of Rs.10 lakhs with a probability of 0.4 or a loss of Rs.2 lakhs with a probability of 0.6
Proposal III:	Profit of Rs.4.5 lakhs with a probability of 0.8 or a loss of Rs.50,000 with a probability of 0.2

If he wants to maximize profit and minimize the loss, which proposal he should prefer?

Solution

Here, we should calculate the mathematical expectation of each proposal.

Expected value (E(X) = $\sum\left[X.P(X)\right]$

$$\left.\begin{array}{l}\text{Expected value of}\\ \text{Proposal I}\end{array}\right\} = (5{,}00.000\times0.6+(80.000\times0.4)$$

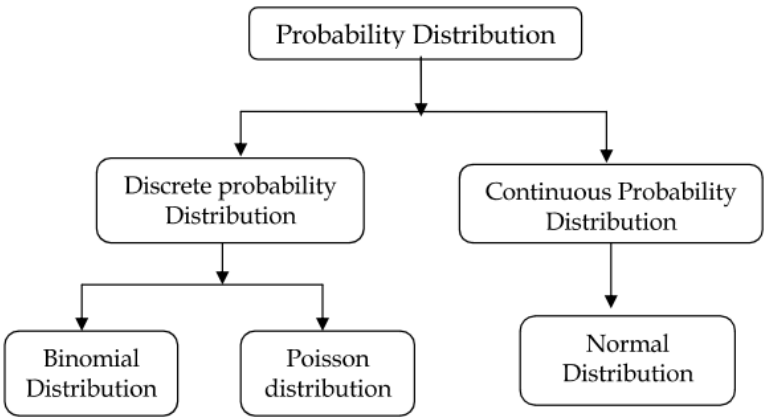
= 3,00.000-32,000
= Rs.2,68,000
=======

Expected value of Proposal II } = (10,00.000×0.4+(12,00.000×0.6)
= 4,00.000-1,20.000
= Rs.2,80,000
=======

Expected value of Proposal III } = (4,50.000×0.8+(-50,000×0.2)
= 3,60,000-10,000
= Rs.3,50,000
=======

Since expected value is highest in case of proposal III, the businessman should prefer the proposal

Classification of Probability Distribution



Discrete Probability Distribution

If the random variable of a probability distribution assumes specific values only, it is called discrete probability distributions. Binomial distribution and poisson distribution are discrete probability distributions.

Continuous Probability Distributions:-

If the random variable of a probability distribution assumes any value in a given interval, then it is called continuous probability distributions. Normal distributions is a continuous probability distribution.

CHAPTER 6
BINOMIAL DISTRIBUTION

Meaning & Definition:

Binomial Distribution is associated with James Bernoulli, a Swiss Mathematician. Therefore, it is also called Bernoulli distribution. Binomial distribution is the probability distribution expressing the probability of one set of dichotomous alternatives, i.e., success or failure. In other words, it is used to determine the probability of success in experiments on which there are only two mutually exclusive outcomes. Binomial distribution is discrete probability distribution.

Binomial Distribution can be defined as follows: "A random variable r is said to follow Binomial Distribution with parameters n and p if its probability function is:

$$P(r) = {}^{N}c_{r}p^{r}q^{n-r}$$

Where, P = probability of success in a single trial

q = 1 – p
n = number of trials
r = number of success n 'n' trials

Assumption of Binomial Distribution OR
(Situations where Binomial Distribution can be applied)

Binomial distribution can be applied when:-

1. The random experiment has two outcomes i.e., success and failure.
2. The probability of success in a single trial remains constant from trial to trial of the experiment.
3. The experiment is repeated for finite number of times.
4. The trials are independent.

Properties (features) of Binomial Distribution:

1. It is discrete probability distribution
2. The shape and location of Binomial distribution changes as 'p' changes for a given 'n'.
3. The mode of the Binomial distribution is equal to the value of 'r' which has the largest probability.
4. Mean of the Binomial distribution increases as 'n' increases with 'p' remaining constant.
5. The mean of Binomial distribution is np.
6. The Standard deviation of Binomial distribution is $\sqrt{npq}$
7. If 'n' is large and if neither 'p' nor 'q' is too close zero, Binomial distribution may be approximated to Normal Distribution.
8. If two independent random variables follow Binomial distribution, their sum also follow Binomial distribution.

Qn: Six coins are tossed simultaneously. What is the probability of obtained 4 heads?

Sol: $P(r) = {}^{n}C_{r}p^{r}q^{n-r}$

$r = 4$

$n = 6$

$p = ½$

$q= 1 – p =1-½ = ½$

$\therefore p(r=4) = {}^{6}C_{4}(½)^{4}(½)^{6-4}$

$$= \frac{6!}{(6-4)!4!} \times \left(\frac{1}{2}\right)^{4+2}$$

$= \frac{6!}{2!4!} \times \left(\frac{1}{2}\right)^6$

$= \frac{6 \times 5}{2 \times 1} \times \frac{1}{64}$

$= \frac{30}{128}$

$= \underline{0.234}$

Qn: The probability that Sachin Tendulkar scores a century in a cricket match is $\frac{1}{3}$. What is the probability that our of 5 matches, he may score century in:-

(1) Exactly 2 matches

(2) No match

Sol: Here $p = \frac{1}{3}$

$Q = 1 - \frac{1}{3} = \frac{2}{3}$

$P(r) = {}^nC_r p^r q^{n-r}$

Probability that Sachin scores centuary in exactly 2 matches is :

$P(r=2) = {}^5C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3$

$= \frac{5!}{(5-2)2!} \times \frac{1}{9} \times \frac{8}{27}$

$= \frac{5 \times 4}{2 \times 1} \times \frac{1}{9} \times \frac{8}{27}$

$= \frac{160}{486}$

$= \frac{80}{243} = 0.329$

(1) Probability that Sachin scores centuary in no matches is:

$P(r=0) = {}^5C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{5-0}$

$= \frac{5!}{(5-0)!0!} \times 1 \quad \times \left(\frac{2}{3}\right)^5$

$= \frac{5!}{5! \times 0!} \times 1 \times \frac{32}{243}$

$= \frac{32}{243}$

$= \underline{0.132}$

Qn: Consider families with 4 children each. What percentage of families would you expect to have:-

(a) Two boys and two girls
(b) At least one boy
(c) No girls
(d) At the most two girls

Sol: p (having a boy = $\frac{1}{2}$

P (having a girl) = $\frac{1}{2}$

n = 4

a) P (getting 2 boys & 2 girls) = p (getting 2 boys) = p(r=2)

$P(r) = {}^nC_r p^r q^{n-r}$

$P(r=2) = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2}$

$$= \frac{4!}{(4-2)!2!} \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^2$$

$$= \frac{4!}{2!2!} \times \left(\frac{1}{2}\right)^{2+2}$$

$$= \frac{4\times3}{642\times1} \times \left(\frac{1}{2}\right)^4$$

$$= \frac{12}{2} \times \frac{1}{16}$$

$$= \frac{12}{32} = \frac{3}{8}$$

$\therefore$% of families with 2 boys & 2 girls = $\frac{3}{8}\times100$

= 37.5%

(b) Probability of having at least one boy

= p (having one boy or having 2 boys or having 3 boys or having 4 boys)

= p (having one boy) + p (having 2 boys) + p (having 3 boys) + p (having 4 boys)

= p(r=1) + p(r=2) + p(r=3) + p(r=4)

P(r) = $^nC_r p^r q^{n-r}$

$$p(r=1)={}^4C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{4-1}$$

$$= 4\times\left(\frac{1}{2}\right)^4 = 4\times\frac{1}{16} = \frac{4}{16}$$

$$p(r=2)={}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{4-2}$$

$$= 6\times\frac{1}{16}=\frac{6}{16}$$

$$p(r=3) = {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{4-3}$$

$$= 4\times\left(\frac{1}{2}\right)^4 = 4\times\frac{1}{16}=\frac{4}{16}$$

$$p(r=4)={}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{4-4}$$

$$= {}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0$$

$$= 1\times\left(\frac{1}{2}\right)^{4+10} = 1\times\left(\frac{1}{2}\right)^4$$

$$= 1\times\frac{1}{16}=\frac{1}{16}$$

$\therefore$ p(r=1 or r=2 or r=3 or r=4)= $\frac{4}{16}+\frac{6}{16}+\frac{4}{16}+\frac{1}{16}=\frac{15}{16}$

$\therefore$ % of families with at least no girls = $\frac{1}{16}\times 100 = 6.25\%$

c) Probability of having no girls = p (having 4 boys)

$$p\ (r = 4) = 4\ C_4 \left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{4-4}$$

$$= \frac{4!}{(4-4)!4!}\,x\left(\frac{1}{2}\right)4+0$$

$$= \frac{4!}{0!4!}\,x\,\frac{1}{16} = \frac{1}{16}$$

$\therefore$ % of families with at least no girls = $\frac{1}{16}$x100 = 6.25%

d) Probability of having at the most 2 girls

= p (having 2 girls or having 1 girl or having no girl)

= p (having 2 boys or having 3 boys or having 4 boys)

= p (r=2) + p(r=3) + p(r=4)

$$p(r=2)={}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{4-2}$$

$$= \frac{6}{16}$$

$$p(r=3)={}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{4-3}$$

$$= \frac{4}{16}$$

$$p(r=4)={}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{4-4}$$

$$= {}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0$$

$$= 1\times\left(\frac{1}{2}\right)^4 \quad = \frac{1}{16}$$

$\therefore$ p (having at the most 2 girls) = $\frac{6}{16}+\frac{4}{16}+\frac{1}{16} \quad = \frac{11}{16}$

$\therefore$ % of families with at the most 2 girls = $\frac{11}{16}\times 100 = 68.75\%$

Mean and Standard Deviation of Binomial Distribution

Mean of Binomial Distribution = np

Standard Deviation of Binomial Distribution = $\sqrt{npq}$

Qn: For a Binomial Distribution, mean = 4 and variance = $\frac{12}{9}$ Find n.

Sol: Mean = np = 4

Standard Deviation - $\sqrt{npq}$

$\therefore$ Variance = Standard Deviation^2

$= \left(\sqrt{npq}\right)$

= npq

npq $= \frac{12}{9}$

Divide npq by np to get the value of q

i.e., $\frac{npq}{np} = q$

$$q = \frac{npq}{NP} = \frac{\frac{12}{9}}{\frac{4}{1}}$$

$$= \frac{12}{9} \times \frac{1}{4} = \frac{3}{9} = \frac{1}{3}$$

q = $\frac{1}{3}$

p = 1 – q

$= 1 - \frac{1}{3} = \frac{2}{3}$

np = 4

$n \times \frac{2}{3} = 4$

n = $4 \div \frac{2}{3}$

n $= 4 \times \frac{3}{2} = 6$

<u>Qn</u>: For a Binomial Distribution, mean is 6 and Standard Deviation is $\sqrt{2}$. Find the parameters.

<u>Sol</u>: Mean (np) = 6

Standard Deviation $\left(\sqrt{npq}\right) = \sqrt{2}$

$\therefore$ npq = 2

$$\frac{npq}{np} = \frac{2}{6}$$

$$q = \frac{1}{3}$$

$\therefore$ p = 1-q

$$= 1 - \frac{1}{3} = \frac{2}{3}$$

np = 6

$$n \times \frac{2}{3} = 6$$

$$\therefore n = \frac{6}{\frac{2}{3}} = 6 \times \frac{3}{2} = 9$$

Value of parameters:

$$p = \frac{2}{3} \quad q = \frac{1}{3} \qquad n = 9$$

<u>Qn</u>: In a Binomial Distribution consisting 5 independent trials, probability for 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the parameter p.

<u>Sol</u>: As there are 5 trials, the terms of the Binomial Distribution are:-

p(r=0); p (r=1); p(r=2); p(r=3); p(r=4) and p(r=5)

p(r=1) = 0.4096

p (r=2) = 0.2048

$p(r=1) = {}^nC_r p^r q^{n-r} = {}^5C_1 p^1 q^{5-1} = 5pq^4 = 0.4096$

$p(r=2) = {}^nC_r p^r q^{n-r} = {}^5C_1 p^2 q^{5-2} = {}^{10p2q3} = 0.2048$

Divide the first term by the second term

$$\frac{5pq^4}{10p^2q^3} = \frac{0.4096}{0.2048}$$

i.e., $\frac{q}{2p} = \frac{2}{1}$

4p = q

4p = 1 – p

4p + p = 1

5p = 1

$P = \frac{1}{5}$

Fitting a Binomial Distribution

Steps: Divide the first term by the second term

1. Find the value of n, p and q
2. Substitute the values of n, p and q in the Binomial Distribution function of ${}^nC_r p^r q^{n-r}$
3. Put r = 0, 1, 2, ……….. in the function ${}^nC_r p^r q^{n-r}$
4. Multiply each such terms by total frequency (N) to obtain the expected frequency.

Qn: Eight coins were tossed together for 256 times. Fit a Binomial Distribution of getting heads. Also find mean and standard deviation.

Sol: P (getting head) = $p = \frac{1}{2}$

$$q = 1 - \frac{1}{2} = \frac{1}{2}$$

N = 8

Binomial Distribution function is $p(r) = {}^nC_r p^r q^{n-r}$

Put r = 0, 1, 2, 3 ………..8, then are get the terms of the Binomial Distribution.

No. of heads i.e. r	P(r)	Expected Frequency P(r) x N N=256
0	${}^8C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^8 = 1\times1\times\frac{1}{256} = \frac{1}{256}$	1
1	${}^8C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7 = 8\times\frac{1}{256} = \frac{8}{256}$	8
2	${}^8C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 = 28\times\frac{1}{256} = \frac{28}{256}$	28
3	${}^8C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5 = 56\times\frac{1}{256} = \frac{56}{256}$	56
4	${}^8C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^4 = 70\times\frac{1}{256} = \frac{70}{256}$	70
5	${}^8C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3 = 56\times\frac{1}{256} = \frac{56}{256}$	56
6	${}^8C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2 = 28\times\frac{1}{256} = \frac{28}{256}$	28
7	${}^8C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1 = 8\times\frac{1}{256} = \frac{8}{256}$	8
8	${}^8C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0 = 1\times\frac{1}{256} = \frac{1}{256}$	1

Mean = np

$$= 8 \times \frac{1}{2} = 4$$

Standard Deviation = $\sqrt{npq}$

$$= \sqrt{8x\frac{1}{2}x\frac{1}{2}} = \sqrt{2} = 1.4142$$

CHAPTER 7
POISSON DISTRIBUTION

Meaning and Definition:

Poisson Distribution is a limiting form of Binomial Distribution. In Binomial Distribution, the total number of trials are known previously. But in certain real life situations, it may be impossible to count the total number of times a particular event occurs or does not occur. In such cases Poisson Distribution is more suitable.

Poison Distribution is a discrete probability distribution. It was originated by Simeon Denis Poisson.

The Poisson Distribution is defined as:-

$$p(r) = \frac{e^{-m}.m^r}{r!}$$

Where r = random variable (i.e., number of success in 'n' trials.

e = 2.7183

m = mean of poisson distribution

Properties of Poisson Distribution

1. Poisson Distribution is a discrete probability distribution
2. Poisson Distribution has a single parameter 'm'. When 'm' is known all the terms can be found out.
3. It is a positively skewed distribution.
4. Mean and Varriance of Poisson Distribution are equal to 'm'.
5. In Poisson Distribution, the number of success is relatively small.
6. The standard deviation of Poisson Distribution is $\sqrt{m}$.

<u>Practical situations where Poisson Distribution can be used</u>

1. To count the number of telephone calls arising at a telephone switch board in a unit of time.
2. To count the number of customers arising at the super market in a unit of time.
3. To count the number of defects in Statistical Quality Control.
4. To count the number of bacterias per unit.
5. To count the number of defectives in a park of manufactured goods.
6. To count the number of persons dying due to heart attack in a year.
7. To count the number of accidents taking place in a day on a busy road.

Qn: A fruit seller, from his past experience, knows that 3% of apples in each basket will be defectives. What is the probability that exactly 4 apples will be defective in a given basket?

Sol: $p(r) = \frac{e^{-m}.m^r}{r!}$

m = 3

$\therefore p(r=4) = \frac{e^{-3}.3^4}{4!} = \frac{0.04979 \times 81}{4 \times 3 \times 2 \times 1}$

$= \frac{0.04979 \times 81}{24}$

= <u>0.16804</u>

<u>Qn</u>: It is known from the past experience that in a certain plant, there are on an average four industrial accidents per year. Find the probability that in a given year there

will be less than four accidents. Assume poisson distribution.

Sol: $p(\text{r} < 4) = \text{p}(\text{r} = 0 \text{ or } 1 \text{ or } 2 \text{ or } 3)$

$= \text{p(r=0) + p(r=1) + p(r=2) + p(r=3)}$

$$p(r) = \frac{e^{-m}.m^r}{r!}$$

m = 4

$$\therefore p(r=0) = \frac{e^{-4}.4^0}{0!} = \frac{0.01832 \times 1}{1} = 0.01832$$

$$p(r=1) = \frac{e^{-4}.4^1}{1!} = \frac{0.01832 \times 4}{1} = 0.07328$$

$$p(r=2) = \frac{e^{-4}.4^2}{2!} = \frac{0.01832 \times 16}{2 \times 1} = 0.14656$$

$$p(r=3) = \frac{e^{-4}.4^3}{3!} = \frac{0.01832 \times 64}{3 \times 2 \times 1} = 0.19541$$

$\therefore$ p(r<4) = 0.01832+0.07328+0.14656+0.19541

= 0.43357

Qn: Out of 500 items selected for inspection, 0.2% are found to be defective. Find how many lots will contain exactly no defective if there are 1000 lots.

Sol: p = 0.2% = 0.002

n = 500

m = np = 500 × 0.002 = 1

$$p(r) = \frac{e^{-m}.m^r}{r!}$$

$$p(r=0) = \frac{e^{-1}.1^0}{0!} = \frac{0.36788 \times 1}{1} = 0.36788$$

$\therefore$ Number of lots containing no defective if there are 1000 lots

= 0.36788×1000 = 367.88 = 368

Qn: In a factory manufacturing optical lenses, there is a small chance of $\frac{1}{1500}$ for any one lense to be defective. The lenses are supplied in packets of 10. Use Poisson Distribution to calculate the approximate number of packets containing (1) one defective (2) no defective in a consignment of 20,000 packets. You are given that $e^{-0.02} = 0.9802$.

Sol: n = 10

p = probability of manufacturing defective lense = $\frac{1}{1500} = 0.002$

m = np = 10×0.002 = 0.02

$$p(r) = \frac{e^{-m}.m^r}{r!}$$

$$p(r=1) = \frac{e^{-0.02} \times 0.02^1}{0!} = \frac{0.9802 \times 0.02}{1} = 0.019604$$

∴No.of packets containing one defective lense

=0.019604 × 20,000

= 392

$$p(r=0) = \frac{e^{-0.02} \times 0.02^0}{0!} = \frac{0.9802 \times 1}{0} = 0.9802$$

∴N0.of packets containing no defective lense

=0.9892 × 20,000

= 19604

Qn: A Systematic sample of 100 pages was taken from a dictionary and the observed frequency distribution of foreign words per page was found to be as follows:

No. of foreign words per page(x)	:	0	1	2	3	4	5	6
Frequency (f)	:	48	27	12	7	4	1	1

Calculate the expected frequencies using Poisson Distribution.

Sol: $p(r) = \frac{e^{-m}.m^r}{r!}$

Here first we have to find out 'm'.

Computation of mean (m)		
x	f	fx
0	48	0
1	27	27
2	12	24
3	7	21
4	4	16
5	1	5
6	1	6
	N = 100	Σfx = 99

$$\overline{X} = \frac{\sum fx}{N} = \frac{99}{100} = 0.99$$

$\therefore$ m = 0.99

$\therefore$ Poisson Distribution = $\frac{e^{-0.99} \times (0.99)^r}{r!}$

Computation of expected frequencies		
X	**P(x)**	**Expected frequency Nx p(x)**
0	$\frac{e^{-0.99} \times (0.99)^0}{0!} = 0.3716$	100×0.3716=37.2

1.	$\frac{e^{-0.99}\times(0.99)^1}{1!}=0.3679$	100×0.3679=36.8
2.	$\frac{e^{-0.99}\times(0.99)^2}{2!}=0.0601$	100×0.1821=18.21
3.	$\frac{e^{-0.99}\times(0.99)^3}{3!}=0.0601$	100×0.0601=6
4.	$\frac{e^{-0.99}\times(0.99)^4}{4!}=0.0149$	100×0.0149=1.5
5.	$\frac{e^{-0.99}\times(0.99)^5}{5!}=0.0029$	100×0.0029=0.3
6.	$\frac{e^{-0.99}\times(0.99)^6}{6!}=0.0005$	100×0.0005=0.1

Hence, the expected frequencies of this Poisson distribution are:-

No. of foreign words per page:	0	1	2	3	4	5	6
Expected frequencies (Rounded):	37	37	18	6	2	0	0

CHAPTER 8
NORMAL DISTRIBUTION

The normal distribution is a continuous probability distribution. It was first developed by De-Moivre in 1733 as limiting form of binomial distribution. Fundamental importance of normal distribution is that many populations seem to follow approximately a pattern of distribution as described by normal distribution. Numerous phenomena such as the age distribution of any species, height of adult persons, intelligent test scores of students, etc. are considered to be normally distributed.

Definition of Normal Distribution

A continuous random variable 'X' is said to follow Normal Distribution if its probability function is:

$$P(X) = \frac{1}{\sqrt{2\pi\sigma}} \times e \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

π = 3.146

e = 2.71828

μ = mean of the distribution

σ = standard deviation of the distribution

Properties of Normal Distribution (Normal Curve)

1. Normal distribution is a continuous distribution.
2. Normal curve is symmetrical about the mean.
3. Both sides of normal curve coincide exactly.
4. Normal curve is a bell shaped curve.
5. Mean, Median and Mode coincide at the centre of the curve.

6. Quantities are equi-distant from median.
   $Q_3 - Q_2 = Q_2 - Q_1$
7. Normal curve is asymptotic to the base line.
8. Total area under a normal curve is 100%
9. The ordinate at the mean divide the whole area under a normal curve into two equal parts. (50% on either side).
10. The height of normal curve is at its maximum at the mean.
11. The normal curve is unimodel, i.e., it has only one mode.
12. Normal curve is mesokurtic.
13. No portion of normal curve lies below the x-axis.
14. Theoretically, the range of normal curve is - α to + α. But practically the range is $\mu - 3\sigma$ to $\mu + 3\sigma$

$\mu \pm 1\sigma$ covers 68.27% area

$\mu \pm 2\sigma$ covers 95.45% area

$\mu \pm 3\sigma$ covers 98.73% area

Importance (or uses) of Normal Distribution

The normal distribution is of central importance in statistical analysis because of the following reasons:-

1. The discrete probability distributions such as Binomial Distribution and Poisson Distribution tend to normal distribution as ‘n’ becomes large.
2. Almost all sampling distributions conform to the normal distribution for large values of ‘n’.
3. Many tests of significance are based on the assumption that the parent population from which samples are drawn follows normal distribution.
4. The normal distribution has numerous mathematical properties which make it popular and comparactively easy to manipulate.

5. Normal distribution finds applications in Statistical Quality Control.
6. Many distributions in social and economic data are approximately normal. For example, birth, death, etc. are normal. For example, birth, death, etc. are normally distributed.

Area under Standard Normal Curve

In case of normal distribution, probability is determined on the basis of area. But to find out the area we have to calculate the ordinate of 2 – scale.

The scale to which the standard deviation is attached is called 2-scale.

Z = ------

Qn: Find p(z>1.8)

Sol: z> 1.8 means the area above 1.8; i.e., the area to the right of 1.8

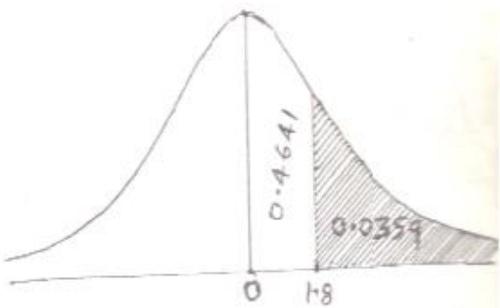
Area upto 1.8 (Table value of 1.8) = 0.4641

Total area on the right side = 0.5

∴ Area to the right of 1.8=0.5-0.4641

= 0.0359

∴ p(z>1.8)= 0.359



Qn: Find p(z<-1.5)

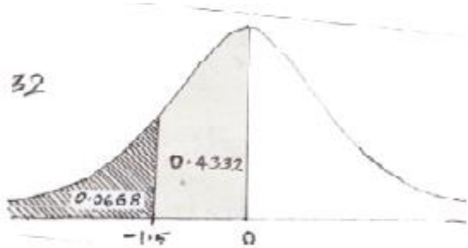
Sol: z<-1.5 means the area to the left of – 1.5

Area between 0 and -1.5 (Table value of 1.5) = 0.4332

Total area on the left side = 0.5

∴Area to the left – 1.5=0.5 – 0.4332

= 0.0668



Qn: Find p(z<-1.96)

Sol: z<-1.96 means the entire area to the left of +1.96

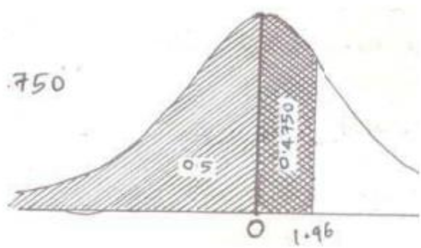
Table value of 1.96 (Area between 0 and 1.96)= 0.4750

Total area on the left side of normal curve = 0.5

∴Area to the left of 1.96=0.4750+0.5

= 0.9750

∴ p(z < 1.96) = 0.9750



Qn: Find p(-1.78 < z < 1.78)

Sol: -1.78 < z < 1.78 means the entire are between-1.78 and +1.78

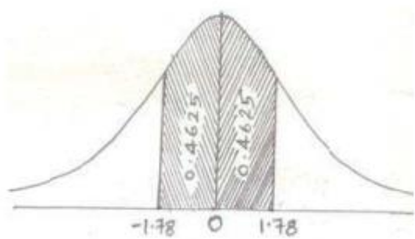
Table value of 1.78 (Area between 0 and 1.78)=0.4625

Total area on the left side = 0.5

∴ Area between – 1.78 and +1.78=0.4625+0.4625

= 0.9250

∴ p(-1.78 < z < 1.78) = 0.9250



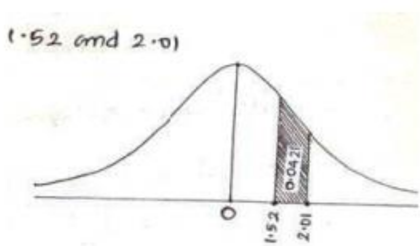
Qn: Find p(1.52 < z < 2.01)

Sol: 1.52< z < 2.01 means area between 1.52 and 2.01

Table value of 1.52 (Area between 0 and 1.52)= 0.4357

Table value of 2.01 (Area between 0 and 2.01)= 0.4778

∴ Area between 2.01 and 1.52=0.4778-0.4357=0.04778



Qn: Find p(-1.52 < z < 0.75)

Sol: -1.52< z < -0.75 means area between -1.52 and -0.75

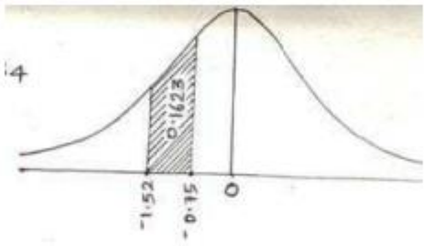
Table value of 0.75 (Area between 0 and -0.75)=0.2734

Table value of 1.52 (Area between 0 and -1.52)=0.4357

∴ Area between -0.75 and -1.52=0.4357-0.2734

= 0.1623

p(-1.52 < z < -0.75) = 0.1623



Qn: Assume the mean height of soldiers to be 68.22 inches with a variance of 10.8 inches. How many soldiers of a regiment of 1000 would you expect to be over six feet tall?

Sol: 6 feet = 12 × 6 = 72 inches

Given μ = 68.22 inches

Varriance = 10.8 inches

∴ σ = ----

X = 72 inches

Z = ------

= -------- = -------- = 1.15023

Table value of 1.15 = 0.3749

Area above 1.15 (above 6 feet) = 0.5 – 0.3749

= 0.125

∴ Number of soldiers who have over 6 feet tall out of

1000 = 0.125 × 1000 = 125

Qn: An aptitude test was conducted for selecting officers in 4 bank from 1000 students. The average score is 42 and the Standard Deviation is 24. Assume normal distribution for scores and find:-

(a) The number of candidates whose score exceed 58.

(b) The number of candidates whose score lie between 30 and 66.

Sol: (a) Given N = 1000

$\mu = 42$

$\sigma = 24$

X = 58

Z = -----

= -------- = -- = 0.667

Table value of 0.667 (Area between 0 and 0.667)= 0.2486

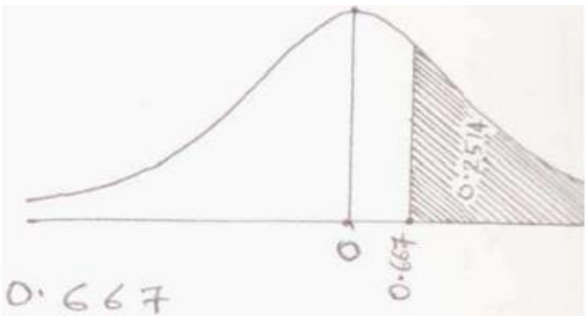
$\therefore$ Area above 0.667 = 0.5 – 0.2486

= 0.2514

$\therefore$ Number of students whose score exceed 58=0.2514×1000

= 251.4

= 251 students



(b) Given N = 1000

= 42

σ =24

X1 = 30

X2 = 66

Z = -----

Z1 = ------- = --------- = ---- = -0.5

Z2 = ------- = -------- = ---- = 1

Table value of 0.5 (Area between 0 and -0.5) = 0.1915

Table value of 1 (Area between 0 and +1) = 0.3413

∴ Area between -0.5 and +1 = 0.1915 + 0.3413

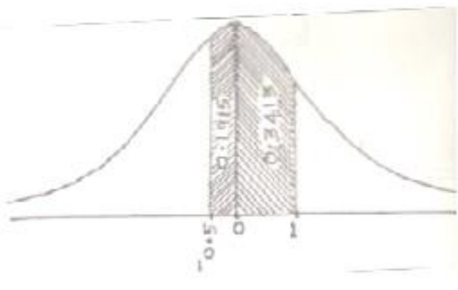
= 0.5328

∴ Number of students whose score lie between 30 and 66

= 0.5328 ×1000

= 532.8

= 533 students



Fitting of a Normal Distribution

Procedure :

1. Find the mean and standard deviation of the given distribution (i.e.,μ and σ)

2. Take the lower limit of each class.
3. Find Z value for each of the lower limit.

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

4. Find the area for z values from the table. The first and the last values are taken as 0.5.
5. Find the area for each class. Take difference between 2 adjacent values if same signs and take total of adjacent values if opposite signs.
6. Find the expected frequency by multiplying area for each class by N.

Qn: Fit a normal distribution of the following data:

Marks :	10-20	20-30	30-40	40-50	50-6	60-70	70-80
No.of Students :	4	22	48	66	40	16	4

Sol:

Computation of mean and standard deviation							
Marks (x)	Mid(m) Point	No.of students(f)	d (m-35)	d’	fd’	d'^2	fd'^2
10-20	15	4	-20	-2	-8	4	16
20-30	25	22	-10	-1	-22	1	22
30-40	35	48	0	0	0	0	0
40-50	45	66	10	1	66	1	66
50-60	55	40	20	2	80	4	160
60-70	65	16	30	3	48	9	144
70-80	75	4	40	4	16	16	64
		N=200			Σfd’=180		$\Sigma fd'^2$=472

$$\bar{x} = \text{Assumed mean} + \frac{\Sigma fd'}{N} + C$$

$$= 35 + \frac{180}{200} \times 100$$

= 35 + 9

= <u>44</u>

$$\sim = \sqrt{\frac{\Sigma fd!^2}{N} + \left(\frac{\Sigma fd!}{N}\right)^2} \times C$$

$$= \sqrt{\frac{472}{200} + \left(\frac{180}{200}\right)^2} \times 10$$

$$= \sqrt{2.36 - 0.9^2} \times 10$$

$$= \sqrt{2.36 - 0.81} \times 10$$

$$= \sqrt{1.55} \times 10$$

= 1.245 ×10 = <u>12.45</u>

Computation of expected frequencies				
Lower class limit	$Z = \frac{x-\mu}{\sigma}$	Area under normal curve	Are for each class	Expected frequency (4)×200
1	2	3	4	5
10	-2.73	0.5000	-	-
20	-1.93	0.4732	0.0268	5
30	-1.12	0.3686	0.1046	21
40	-0.32	0.1255	0.2431	49
50	+0.48	0.1844	0.3099	62
60	+1.29	0.4015	0.2171	43
70	+2.09	0.4817	0.0802	16
80	+2.89	0.5000	0.0183	4
			Total	**200**

Module V

QUANTITATIVE APPROACH TO DECISION MAKING

Similar to operations management, project management employs an array of quantitative techniques while performing planning, scheduling, forecasting, and monitoring tasks. The main purpose of the quantitative approach is to make an optimal decision by using mathematical and statistical models in a situation when the probability of all outcomes is uncertain.

Quantitative approach to decision-making produces the best results when the problem is clearly defined, several alternatives exist, and decision outcomes are easily measurable. However, in the case that many external factors are outside of the decision-maker's control and their probability is unknown, the quantitative methods can become unreliable.

Quantitative approach techniques, especially the ones relying on statistical software, have the advantage of suggesting the best solution to the problem without even identifying all possible alternatives. This feature is quite useful in problems where the number of possible alternatives is very large though only a few are worth considering for selection. Once the problem and conditions are defined, the decision-making process becomes quick.

Decision making

Decision making is a human process; inasmuch as they are made under conditions of uncertainty, decisions require human judgment. Sometimes, that judgment can be based upon our "gut feeling" which ideally arises on the basis of learning from past experience. For most decisions that are simple, this "gut feeling" is adequate. However, with increasing uncertainty and/or a growing number of independent variables, decisions become more complex and our intuitive judgments become less reliable. At that point, we require reliable methods and tools to help us make wiser choices between alternate courses of action.

Various quantitative techniques for decision making are:-

1. Mathematical Programming
2. Cost Analysis (Break-Even Analysis)
3. Cost-Benefit Analysis
4. Linear Programming
5. Capital Budgeting
6. Inventory Management
7. Expected Value
8. Decision Tree
9. Simulation
10. Queuing or Waiting Line Theory
11. Game Theory
12. Information Theory
13. Preference Theory/Utility Theory and Few Others.

Steps in Decision Making

Step 1: Identify the decision

The first step in making the right decision is to identify the real problem or opportunity and also to decide how to address it. Evaluate why and how this decision will make a difference in the outcome at the end.

Step 2: Gather relevant information

In order to make a good decision, the decision maker collects some pertinent information. He will decide what information is needed to take the decision, which are the best source of information, and how to get such information, and how to get such information. The collection of information is related to both internal and external. Internal information is related to the availability of materials, availability of skilled labor, production capacity of the firm etc. external information is related to the market condition, government policy, technological changes etc.

Step 3: identify the alternatives:

A particular problem can be solved in many ways. The decision maker will think about several possible paths of action, or alternatives for solving the problem. He may use his imagination and judgement to construct new alternatives rather than thinking about the conventional method of solving the problem. In this step, he will list all the possible and desirable alternatives.

Step 4: Weigh the evidence

In this step, he will need to "evaluate for feasibility, acceptability and desirability" to know which alternative is best. All the alternatives may not be suited to the organization. It depends upon the existing practices, employees attitude, policy of the management, legal implications etc. Considering the goal and preference to the alternatives considered, a priority list of alternatives is prepared for final selection.

Step 5: Choose among alternatives

At this stage, the decision maker is ready to select the alternative which is seemed to be the best. The final selection of alternative may be a combination more than one alternative also which is best suited to the situation.

Step 6: Take action (implementation of the decision)

At this stage, the finalized decision is implemented as planned by gaining the support from employees and stakeholders. At the time of implementation, the management should be prepared to address any question or concerns which may arise.

Decision Tree Analysis

A Decision Tree Analysis is a graphic representation of various alternative solutions that are available to solve a problem. The manner of illustrating often proves to be decisive when making a choice. A Decision Tree Analysis is created by answering a number of questions that are continued after each affirmative or negative answer until a final choice can be made.

Decision making process

A Decision Tree Analysis is a scientific model and is often used in the decision making process of organizations. When making a decision, the management already envisages alternative ideas and solutions. By using a decision tree, the alternative solutions and possible choices are illustrated graphically as a result of which it becomes easier to make a well-informed choice. This graphic representation is characterized by a tree-like structure in which the problems in decision making can be seen in the form of a flowchart, each with branches for alternative choices.

Definition

Decision tree analysis is a powerful decision-making tool which initiates a structured nonparametric approach for problem-solving. It facilitates the evaluation and comparison of the various options and their results, as shown in a decision tree. It helps to choose the most competitive alternative.

It is a widely used technique for taking crucial decisions like project selection, cost management, operations management, production method, and to deal with various other strategic issues in an organization.

Terminologies Used

Let us understand some of the relevant concepts and terms used in the decision tree:

- **Root Node**: A root node compiles the whole sample, it is then divided into multiple sets which comprise of homogeneous variables.
- **Decision Node**: That sub-node which diverges into further possibilities, can be denoted as a decision node.
- **Terminal Node**: The final node showing the outcome which cannot be categorized any further, is termed as a value or terminal node.
- **Branch**: A branch denotes the various alternatives available with the decision tree maker.
- **Splitting**: The division of the available option (depicted by a node or sub-node) into multiple sub-nodes is termed as splitting.
- **Pruning**: It is just the reverse of splitting, where the decision tree maker can eliminate one or more sub-nodes from a particular decision node.

Steps in Decision Tree Analysis

Following steps simplify the interpretation process of a decision tree:

1. The *first step* is understanding and specifying the problem area for which decision making is required.
2. The *second step* is interpreting and chalking out all possible solutions to the particular issue as well as their consequences.
3. The *third step* is presenting the variables on a decision tree along with its respective probability values.
4. The *fourth step* is finding out the outcomes of all the variables and specifying it in the decision tree.

5. The *last step* is highly crucial and backs the overall analysis of this process. It involves calculating the EMV values for all the chance nodes or options, to figure out the solution which provides the highest expected value.

Example 1

ABC Ltd. is a company manufacturing skincare products. It was found that the business is at the maturity stage, demanding some change. After rigorous research, management came up with the following alternatives.

First alternative – Expansion of Business Unit:

- 40% possibility that the market share will hike, increasing the overall profitability of the company by ₹2500000;
- 60% possibility that the competitors would take over the market share and the company may incur a loss of ₹800000.

Second alternative - New Product Line of Shower Gel:

- 50% chances are that the project would be successful and yield ₹1800000 as profit;
- 50% possibility of failure persists, leading to a loss of ₹800000.

Third alternative Do Nothing:

- 40% chances are there that yet, the organization can attract new customers, generating a profit of ₹1000000;
- 60% chances of failure are there due to the new competitors, incurring a loss of ₹400000.

Solution

Given below is the evaluation of each of these alternatives:

Expansion of Business Unit:

If the company invests in the development of its business unit, there can be two possibilities, i.e.:

- 40% possibility that the market share will hike, increasing the overall profitability of the company by ₹2500000;
- 60% possibility that the competitors would take over the market share and the company may incur a loss of ₹800000.

To find out the viability of this option, let us compute its EMV (Expected Monetary Value):

$$EMV = \left(\frac{40}{100} X\ 250000\right) + \left(\frac{60}{100} X - 800000\right)$$

EMV = 1000000 - 480000

= ₹520000.

New Product Line of Shower Gel:

If the organization go for new product development, there can be following two possibilities:

- 50% chances are that the project would be successful and yield ₹1800000 as profit;
- 50% possibility of failure persists, leading to a loss of ₹800000.

To determine the profitability of this idea, let us evaluate its EMV:

$$EMV = \left(\frac{50}{100} X\ 1800000\right) + \left(\frac{50}{100} X - 600000\right)$$

EMV = 900000 - 600000

=₹300000.

Do Nothing:

If the company does not take any step, still there can be two outcomes, discussed below:

- 40% chances are there that yet, the organization can attract new customers, generating a profit of ₹1000000;
- 60% chances of failure are there due to the new competitors, incurring a loss of ₹400000.

Given below is the EMV in such circumstances:

$$\text{EMV}= \left(\frac{40}{100} X\ 1000000\right) + \left(\frac{60}{100} X - 400000\right)$$

EMV = 400000 - 240000

= ₹160000.

From the above evaluation, we can easily make out that the option of a new product line has the highest EMV. Therefore, we can say that the company can avail this opportunity to make the highest gain by ensuring the best possible use of its resources.

Advantages of Decision Tree Analysis

Business organizations need to consider various parameters during decision making. A decision tree analysis is one of the prominent ways of finding out the right solution to any problem. Let us now understand its various benefits below:

- **Depicts Most Suitable Project/Solution**: It is an effective means of picking out the most appropriate project or solution after examining all the possibilities.
- **Easy Data Interpretation and Classification**: Not being rocket science, decision tree eases out the process of segregation of the acquired data into different classes.
- **Assist Multiple Decision-Making Tools**: It also benefits the decision-maker by providing input for other analytical methods like nature's tree.

- **Considers Both, Categorial and Numerical Data**: This technique takes into consideration the quantitative as well as the qualitative variables for better results.
- **Initiates Variable Analysis**: Its structured phenomena also facilitates the investigation and filtration of the relevant data.

Disadvantages of Decision Tree Analysis

Decision tree analysis has multidimensional applicability. However, its usage becomes limited due to its following shortcomings:

- **Inappropriate for Excessive Data**: Since it is a non-parametric technique, it is not suitable for the situations where the data for classification is vast.
- **Difficult to Handle Numerous Outcomes**: If there are multiple possible results of every decision, it becomes tedious to compile all these on a decision tree.
- **Chances of Classification Errors**: A less experienced decision tree maker usually makes a mistake while putting the variables into different classes.
- **Impact of Variance**: Making even a slightest of change becomes problematic since it results in a completely different decision tree.
- **Unsuitable for Continuous Variables**: Incorporating many open-ended numerical variables increases the possibility of errors.
- **Sensitive towards Biasness**: A decision tree maker may lay more emphasis on preferable variables which may divert the direction of analysis.
- **Expensive Process**: Collection of sufficient data, its classification and analysis demand high expense, being a resource-intensive process.

Example

There is 40% chance that a patient admitted to the hospital, is suffering from cancer. A doctor has to decide whether a serious operation should be performed or not. If the patient is suffering from cancer and the serious operation is performed, the chance that he will recover is 70%, otherwise it is 35%. On the other hand, if the patient is not suffering from cancer and the serious operation is performed the chance that he will recover is 20%, otherwise it is 100%. Assume that recovery and death are the only possible results. Construct an appropriate decision tree. What decision should the doctor take?

Solution

Chance of suffering from cancer = 40%

Chances of not suffering from cancer = 60%

Patients in suffering from cancer, the patient will recover after the serious operation = 70%

The patient will not recover after operation = 30%

Probability that the patient will recover after serious operation = 40×70/100 = 0.28

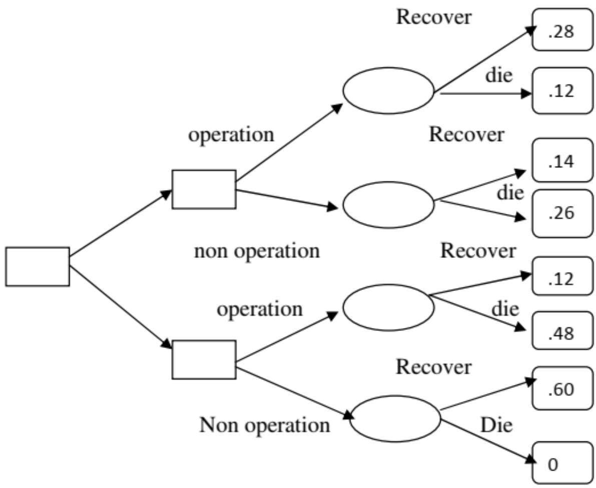
Probability that the patient will not recover after serious operation = 40×30/100 = 0.12 Probability that the patient will recover without serious operation = 40 ×35/100 = 0.14 Probability that the patient will not recover without serious operation = 40×65/100 = 0.26

Patient is not suffering from cancer.

Probability that the patient will recover with serious operation = 60×20/100 = 0.12

Probability that the patient will not recover with serious operation = 60 ×80/100 = 0.48 Probability that the patient will recover without operation = 60× 100/100 = 0.60

Probability that the patient will not recover without operation = .0



Model Building

A model is an abstraction of reality or a representation of a real object or situation. In other words, a model presents a simplified version of something. It may be as simple as a drawing of house plans, or as complicated as a miniature but functional representation of a complex piece of machinery. A model airplane may be assembled and glued together from a kit by a child, or it actually may contain an engine and a rotating propeller that allows it to fly like a real airplane.

A more useable concept of a model is that of an abstraction, from the real problem, of key variables and relationships. These are abstracted in order to simplify the problem itself. Modeling allows the user to better understand the problem and presents a means for manipulating the situation in order to analyze the results of various inputs ("what if" analysis) by subjecting it to a changing set of assumptions

MODEL CLASSIFICATIONS

Some models are replicas of the physical properties (relative shape, form, and weight) of the object they represent. Others are physical models but do not have the same physical appearance as the object of their representation. A third type of model deals with symbols and numerical relationships and expressions. Each of these fits within an overall classification of four main categories: physical models, schematic models, verbal models, and mathematical models.

1. Physical models.

Physical models are the ones that look like the finished object they represent. Iconic models are exact or extremely similar replicas of the object being modeled. Model airplanes, cars, ships, and even models of comic book super-heroes look exactly like their counterpart but in a much smaller scale. Scale models of municipal buildings, shopping centers, and property developments such as subdivisions, homes, and office complexes all hopefully look exactly as the "real thing" will look when it is built. The advantage here is the models' correspondence with the reality of appearance. In other words, the model user can tell exactly what the proposed object will look like, in three dimensions, before making a major investment.

In addition to looking like the object they represent, some models perform as their counterparts would. This allows experiments to be conducted on the model to see how it might perform under actual operating conditions. Scale models of airplanes can be tested in wind tunnels to determine aerodynamic properties and the effects of air turbulence on their outer surfaces. Model automobiles can be exposed to similar tests to evaluate how wind resistance affects such variables as handling and gas mileage. Models of bridges and dams can be subjected to multiple levels of stress from wind, heat, cold, and other sources in order to test such variables as endurance and safety. A scale model that behaves in a manner that is similar to the "real thing" is far less expensive to create and test than its actual counterpart. These types of models often are referred to as prototypes.

Additionally, some physical models may not look exactly like their object of representation but are close enough to provide some utility. Many modern art statues represent some object of reality, but are so different that many people cannot clearly distinguish the object they represent. These are known as analog models. An example is the use of cardboard cutouts to represent the machinery being utilized within a manufacturing facility. This allows planners to move the shapes around enough to determine an optimal plant layout.

2. Schematic models.

Schematic models are more abstract than physical models. While they do have some visual correspondence with reality, they look much less like the physical reality they represent. Graphs and charts are schematic models that provide pictorial representations of mathematical relationships. Plotting a line on a graph indicates a mathematical linear relationship between two variables. Two such lines can meet at one exact location on a graph to indicate the break-even point, for instance. Pie charts, bar charts, and histograms can all model some real situation, but really bear no physical resemblance to anything.

Diagrams, drawings, and blueprints also are versions of schematic models. These are pictorial representations of conceptual relationships. This means that the model depicts a concept such as chronology or sequence. A flow chart describing a computer program is a good example. The precedence diagrams used in project management or in assembly-line balancing show the sequence of activities that must be maintained in order to achieve a desired result.

3. Verbal models.

Verbal models use words to represent some object or situation that exists, or could exist, in reality. Verbal models may range from a simple word presentation of scenery described in a book to a complex business decision problem (described in words and numbers). A firm's mission statement is a model of its beliefs about what business it is in and sets the stage for the firm's determination of goals and objectives.

Verbal models frequently provide the scenario necessary to indicate that a problem is present and provide all the relevant and necessary information to solve the problem, make recommendations, or at least determine feasible alternatives. Even the cases presented in management textbooks are really verbal models that represent the workings of a business without having to take the student to the firm's actual premises. Oftentimes, these verbal models provide enough information to later depict this problem in mathematical form. In other words, verbal models frequently are converted into mathematical models so that an optimal, or at least functional, solution may be found utilizing some mathematical technique. A look in any mathematics book, operations management book, or management science text generally provides some problems that appear in word form. The job of the student is to convert the word problem into a mathematical problem and seek a solution.

4. Mathematical models.

Mathematical models are perhaps the most abstract of the four classifications. These models do not look like their real-life counterparts at all. Mathematical models are built using numbers and symbols that can be transformed into functions, equations, and formulas. They also can be used to build much more complex models such as matrices or linear programming models. The user can then solve the mathematical model (seek an optimal solution) by utilizing simple techniques such as multiplication and addition or more complex techniques such as matrix algebra or Gaussian elimination. Since mathematical models frequently are easy to manipulate, they are appropriate for use with calculators and computer programs. Mathematical models can be classified according to use (description or optimization), degree of randomness (deterministic and stochastic), and degree of specificity (specific or general). Following is a more detailed discussion of mathematical model types.

Types Of Mathematical Models

a. Descriptive Models.

Descriptive models are used to merely describe something mathematically. Common statistical models in this category include the mean, median, mode, range, and standard deviation. Consequently, these phrases are called "descriptive statistics." Balance sheets, income statements, and financial ratios also are descriptive in nature.

b. Optimization Models.

Optimization models are used to find an optimal solution. The linear programming models are mathematical representations of constrained optimization problems. These models share certain common characteristics. Knowledge of these characteristics enables us to recognize problems that can be solved using linear programming.

c. Deterministic models.

Deterministic models are those for which the value of their variables is known with certainty. In a previous example, the manager knew profit margins and constraint values with certainty. This makes the linear programming model a deterministic optimization model.

d. Specific Models.

Specific models apply to only one situation or model one unique reality. The previous examples of profit function (descriptive), objective function (optimization), and payoff matrix (probabilistic) are all specific models. In other words, the values established in the model are relevant for that one unique situation. Linear programming models can be said to be deterministic specific, while decision trees can be called probabilistic specific models.

e. General Models.

General models can be utilized in more than one situation. For example, the question of how much to order is determined by using an economic order quantity (EOQ) model. EOQ models identify the optimal order quantity by minimizing the sum of certain annual costs that vary with order size. On the other hand, the question of how much should be ordered for the next (fixed) interval is determined by the fixed order interval (FOI) model, which is used when orders must be placed at fixed time intervals (weekly, twice, etc.).

Model building steps

1. Define the problem, decision, situation, or scenario and the factors that influence it.
2. Select criteria to guide the decision, and establish objectives. A perfect example of this is the use of heuristics in assembly-line balancing to guide the decision and the criteria of maximizing efficiency /minimizing idle time as an objective.
3. Formulate a model that helps management to understand the relationships between the influential factors and the objectives the firm is trying to achieve.

4. Collect relevant data while trying to avoid the incorporation of superfluous information into the model.
5. Identify and evaluate alternatives. Once again, the example of assembly-line balancing is appropriate. The user can manipulate the model by changing the heuristics and comparing the final results, ultimately finding an optimal solution through trial-and-error. However, the production of alternatives may not be necessary if the model in use initially finds an optimal solution.
6. Select the best alternative
7. Implement the alternative or reevaluate

Linear Programming Models

INTRODUCTION

A model, which is used for optimum allocation of scarce or limited resources to competing products or activities under such assumptions as certainty, linearity, fixed technology, and constant profit per unit, is linear programming.

Linear Programming is one of the most versatile, powerful and useful techniques for making managerial decisions. Linear programming technique may be used for solving broad range of problems arising in business, government, industry, hospitals, libraries, etc. Whenever we want to allocate the available limited resources for various competing activities for achieving our desired objective, the technique that helps us is LINEAR PROGRAMMING. As a decision making tool, it has demonstrated its value in various fields such as production, finance, marketing, research and development and personnel management.

Linear programming is a simple technique where we depict complex relationships through linear functions and then find the optimum points. The important word in the previous sentence is depicted. The real relationships might be much more complex – but we can simplify them to linear relationships.

Applications of linear programming are everywhere around you. You use linear programming at personal and professional fronts. You are using linear programming when you are driving from home to work and want to take the shortest route. Or when you have a project delivery you make strategies to make your team work efficiently for on-time delivery.

In Mathematics, linear programming is a method of optimising operations with some constraints. The main objective of linear programming is to maximize or minimize the numerical value. It consists of linear functions which are subjected to the constraints in the form of linear equations or in the form of inequalities.

Linear programming is considered as an important technique which is used to find the optimum resource utilisation. The term "linear programming" consists of two words such as linear and programming. The word "linear" defines the relationship between multiple variables with degree one. The word "programming" defines the process of selecting the best solution from various alternatives. Linear Programming is widely used in Mathematics and some other field such as economics, business, telecommunication, and manufacturing fields.

Definition

Linear programming (LP) or Linear Optimisation may be defined as the problem of maximizing or minimizing a linear function which is subjected to linear constraints. The constraints may be equalities or inequalities.

Characteristics of Linear Programming

The following are the five characteristics of the linear programming problem:

1. **Constraints** – The limitations should be expressed in the mathematical form, regarding the resource.
2. **Objective Function** – In a problem, the objective function should be specified in a quantitative way.
3. **Linearity** – The relationship between two or more variables in the function must be linear. It means that the degree of the variable is one.
4. **Finiteness** – There should be finite and infinite input and output numbers. In case, if the function has infinite factors, the optimal solution is not feasible.
5. **Non-negativity** – The variable value should be positive or zero. It should not be a negative value.
6. **Decision Variables** – The decision variable will decide the output. It gives the ultimate solution of the problem. For any problem, the first step is to identify the decision variables.

Advantages of Linear programming

The advantages of linear programming are:

1. Linear programming provides insights to the business problems
2. It helps to solve multi-dimensional problems
3. According to the condition change, LP helps in making the adjustments
4. By calculating the cost and profit of various things, LP helps to take the best optimal solution

Properties of linear programming model

Any linear programming model must have the following properties

a. The relationship between variables and constraints must be linear
b. The model must have an objective function
c. The model must have structural constraints
d. The model must have non negativity constraints

Maximizing models

Example:- A retail store stocks two types of shirts A and B. These are packed in attractive cardboard boxes. During a week the store can sell a maximum of 400 shirts of type A and a maximum of 300 shirts of type B. The storage capacity, however, is limited to a maximum of 600 of both types combined. Type A shirt fetches a profit of Rs. 2/- per unit and type B a profit of Rs. 5/- per unit. How many of each type the store should stock per week to maximize the total profit? Formulate a mathematical model of the problem.

Solution:

Here shirts A and B are problem variables. Let the store stock 'a' units of A and 'b' units of B. As the profit contribution of A and B are Rs.2/- and Rs.5/- respectively, objective function is:

Maximize $Z = 2a + 5b$ subjected to condition (s.t.)

Structural constraints are, stores can sell 400 units of shirt A and 300 units of shirt B and the storage capacity of both put together is 600 units. Hence the structural constraints are:

$la + 0b \geq 400$ and $0a + 1b \leq 300$ for sales capacity and $la + 1b \leq 600$ for storage capacity.

And non-negativity constraint is both a and b are ≥ 0. Hence the model is:

Maximize $Z= 2a + 5b$ s.t.

$$la + 0b \leq 400$$

$$0a + lb \leq 300$$

$$la + 1b \leq 600 \text{ and}$$

Both a and b are ≥ 0.

Example

A ship has three cargo holds, forward, aft and center. The capacity limits are:

Forward 2000 tons, 100,000 cubic meters

Center 3000 tons, 135,000 cubic meters

Aft 1500 tons, 30,000 cubic meters.

The following cargoes are offered, the ship owners may accept all or any part of each commodity:

Commodity	Amount in tons	Volume/ton in cubic meters	Profit/ton in cubic meters
A	6000	60	60
B	4000	50	80
C	2000	25	50

In order to preserve the trim of the ship the weight in each hold must be proportional to the capacity in tons. How should the cargo be distributed so as to maximize profit? Formulate this as linear programming problem.

Solution:

Problem variables are commodities, A, B, and C. Let the shipping company ship 'a' units of A and 'b'units of B and 'c' units of C. Then Objective function is:

Maximize Z= 60a + 80b + 50c s.t.

Constraints are:

Weight constraint: 6000a + 4000b +2000c ≤ 6,500 (= 2000 + 3000 +1500)

The tonnage of commodity is 6000 and each ton occupies 60 cubic meters, hence there are 100 cubic meters capacity is available.

Similarly, availability of commodities B and C, which are having 80 cubic meter capacities each. Hence capacity inequality will be:

$100a + 80b + 80c \leq 2{,}65{,}000$ (= 100,000 + 135,000 + 30,000). Hence the l.p.p. Model is:

Maximise $Z = 60a+80b+50c$ s.t. $100a = 6000/60 = 100$

$6000a + 4000b + 2000c \leq 5\ 6{,}500$ $80b = 4000/50 = 80$

$100a+80b+80c\ 5\ 2{,}65{,}000$ and $80c = 2000/25 = 80$ etc.

a,b,c all ≥ 0

Minimization models

Example

A patient consult a doctor to check up his ill health. Doctor examines him and advises him that he is having deficiency of two vitamins, vitamin A and vitamin D. Doctor advises him to consume vitamin A and D regularly for a period of time so that he can regain his health. Doctor prescribes tonic X and tonic Y, which are having vitamin A, and D in certain proportion. Also advises the patient to consume at least 40 units of vitamin A and 50 units of vitamin Daily. The cost of tonics X and Y and the proportion of vitamin A and D that present in X and Y are given in the table below. Formulate l.p.p. to minimize the cost of tonics.

Vitamins	Tonic		Daily requirements in units
	X	Y	
A	2	4	40
D	3	2	50
Cost in Rs per unit	5	3	

Solution:

Let patient purchase x units of X and y units of Y.

Objective function: Minimize Z= 5x + 3y

Inequality for vitamin A is $2x + 4y \geq 40$ (Here at least word indicates that the patient can consume more than 40 units but not less than 40 units of vitamin A daily).

Similarly the inequality for vitamin D is $3x + 2y \geq 50$.

For non-negativity constraint the patient cannot consume negative units. Hence both x and y must be ≥ 0.

Now the l.p.p. model for the problem is:

Minimize Z= 5x + 3y s.t.

$2x + 4y \geq 40$

$3x + 2y \geq 50$ and

Both x and y are ≥ 0.

Graphical method

A linear program can be solved by multiple methods. In this section, we are going to look at the Graphical method for solving a linear program. This method is used to solve a two-variable linear program. If you have only two decision variables, you should use the graphical method to find the optimal solution.

A graphical method involves formulating a set of linear inequalities subject to the constraints. Then the inequalities are plotted on an X-Y plane. Once we have plotted all the inequalities on a graph the intersecting region gives us a feasible region. The feasible region explains what all values our model can take. And it also gives us the optimal solution.

In graphical method, the inequalities (structural constraints) are considered to be equations. This is because;

one cannot draw a graph for inequality. Only two variable problems are considered, because we can draw straight lines in two-dimensional plane (X- axis and Y-axis). More over as we have non- negativity constraint in the problem that is all the decision variables must have positive values always the solution to the problem lies in first quadrant of the graph. Some times the value of variables may fall in quadrants other than the first quadrant. In such cases, the line joining the values of the variables must be extended in to the first quadrant. The procedure of the method will be explained in detail while solving a numerical problem.

The characteristics of Graphical method are:

(i) Generally the method is used to solve the problem, when it involves two decision variables.

(ii) For three or more decision variables, the graph deals with planes and requires high imagination to identify the solution area.

(iii) Always, the solution to the problem lies in first quadrant.

(iv) This method provides a basis for understanding the other methods of solution.

Example

A company manufactures two products, X and Y by using three machines A, B, and C. Machine A has 4 hours of capacity available during the coming week. Similarly, the available capacity of product X requires one hour of Machine A, 3 hours of machine B and 10 hours of machine C. Similarly one unit of product Y requires 1 hour, 8 hour and 7 hours of machine A, B and C respectively. When one unit of X is sold in the market, it yields a profit of Rs. 5/- per product and that of Y is Rs. 7/- per unit. Solve the problem by using graphical method to find the optimal product mix.

Solution:

The details given in the problem is given in the table below:

Machines	Products (time required in hours)		Available capacity in hours
	X	Y	
A	1	1	4
B	3	8	24
C	10	7	35
Profit per units in Rs	5	7	

Let the company manufactures x units of X and y units of Y, and then the L.P. model is:

Maximize Z = 5x + 7y s.t.

$1x + ly \leq 4$

$3x + 8y \leq 24$

$10x + 7y \leq 35$ and

Both x and y are ≥ 0.

As we cannot draw graph for inequalities, let us consider them as equations.

Maximize Z= 5x + 7y s.t

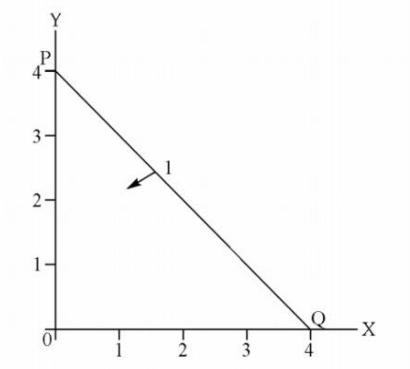
$1x + ly=4$

$3x + 8y = 24$

$10x + 7y = 35$ and both x and y are ≥ 0

Let us take machine A. and find the boundary conditions. If x = 0, machine A can manufacture

4/1= 4 units of y.



Graph for machine A

Similarly, if y = 0, machine A can manufacture 4/1 = 4 units of x. For other machines:

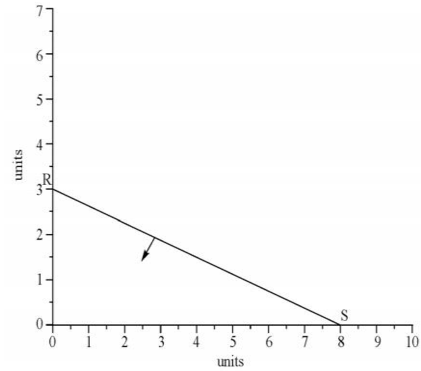
Machine B When x = 24/8 = 3 and when y = 0 x = 24/3 = 8

Machine C When x = 0, y = 35/10 = 3.5 and when y=0, x= 35 / 7 = 5.

These values we can plot on a graph, taking product X on x-axis and product Y on y-axis.

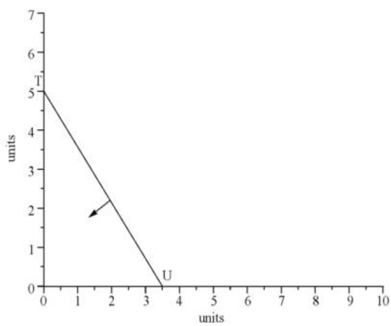
First let us draw the graph for machine A. In above figure we get line which represents 1x + ly = 4. The point P on Y axis shows that the company can manufacture 4 units of Y only when does not want to manufacture X. Similarly the point Q on X axis shows that the company can manufacture 4 units of X, when does not want to manufacture Y. In fact triangle POQ is the capacity of machine A and the line PQ is the boundary line for capacity of machine A.

Similarly figure below show the Capacity line RS for machine B. and the triangle ROS shows the capacity of machine B i.e., the machine B can manufacture 3 units of product Y alone or 8 units of product X alone.

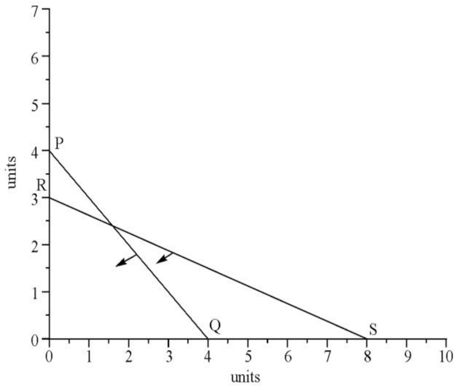


Graph for machine B

The graph below shows that the machine C has a capacity to manufacture 5 units of Y alone or 3.5 units of X alone. Line TU is the boundary line and the triangle TOU is the capacity of machine C. The graph is the combined graph for machine A and machine B. Lines PQ and RS intersect at M. The area covered by both the lines indicates the products (X and Y) that can be manufactured by using both machines. This area is the feasible area, which satisfies the conditions of inequalities of machine A and machine B. As X and Y are processed on A and B the number of units that can be manufactured will vary and the there will be some idle capacities on both machines.

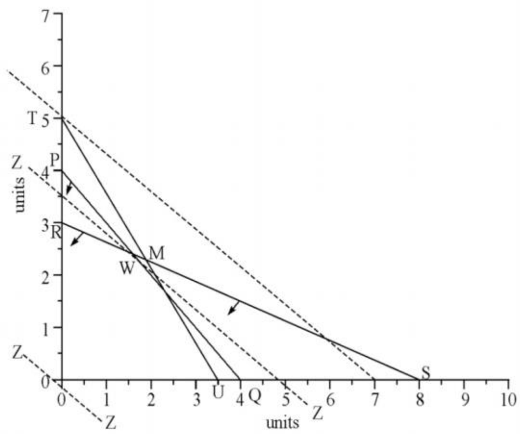


Graph for machine C



Graph of Machine A and B

Figure below shows the feasible area for all the three machines combined. This is the fact because products X and Y are complete when they are processed on machine A, B, and C. The area covered by all the three lines PQ. RS, and TU form a closed polygon ROUVW. This polygon is the feasible area for the three machines. This means that all the points on the lines of polygon and any point within the polygon satisfies the inequality conditions of all the three machines.



Graph for Machine A, B and C combined

Example

A manufacturing company is engaged in producing three types of product M, N and O. the production department produces each day, components sufficient to make 100 units of M, 50 units of N and 60 units of O. the management is confronted with the problem of optimizing the daily production of products in the assembly department, where any 200 man hours are available daily for assembling the products. The following additional; information is available.

Type of product Profit contribution Assembly per product (hrs)

time Per unit of product (Rs

M	24	1.6
N	40	3.4
O	90	5

The company has a daily order for 40 units of product M and total of 30 units of product N and O. Formulate this problem as linear programming problem so as to maximize total profit.

Solution

Let X_1 = number of units of product M

X_2 = number of units of product N

X_3 = number of units of product O

Profit contribution per unit of products M, N and O are 24, 40 and 90 respectively.

So the objective function is

$24X_1 + 40 X_2 + 90X_3$

The objective function is based on certain constraints. They are

Assembly time per product in hours = 1.6. 3.4 and 5 for product M, N and O respectively,.

The constraint can be written as

$1.6X_1 + 3.4X_2 + 5X_3 \leq 200$

The second constraint is the maximum units of production.

M = 200 units, N = 50 units. O = 60 units

It can be written as

$X_1 \leq 100$, $X_2 \leq 50$, $X_3 \leq 60$

The next constraint is order commitment.

40 units of product M, 30 units of product N and O

That is $X_1 \geq 40$, $X_2 + X_3 \geq 30$

The problem can written as

Maximize Z = 24x1 + 40 x2 + 90x3

Subject to

$1.6X_1 + 3.4X_2 + 5X_3 \leq 200$

$X_1 \leq 100$,

$X_2 \leq 50$,

$X_3 \leq 60$

is $X_1 \geq 40$,

$X_2 + X_3 \geq 30$

Non negative restrictions are

$X_1 \geq 0$. $X_2 \geq 0$, $X_3 \geq 0$