Dr. J.PRAHANNAYAKI, Ph.D.,
Assistant Professor of Commerce,
K. N. Government Arts College for Women (A),
Thanjavur

Reference material

1. **Statistical Methods. S.P. Gupta**, Sultan Chand & Sons, New Delhi, 2016.
2. **Operations Research, V.K. Kapoor**, Sultan Chand & Sons, New Delhi, 2017

UNIT I
QUANTITATIVE TECHNIQUES

INTRODUCTION

Quantitative techniques are the means for problem solving and decision making.

Quantitative techniques may be defined as those techniques which provide the decision maker with a systematic and powerful means of analysis and help based on quantitative data, in exploring policies for achieving pre determined goals.

Quantitative techniques are those statistical and operations research or programming techniques which help in the decision making process especially concerning business and industry.

Quantitative techniques involve the introduction of the element of quantities. i.e., they involve the use of numbers, symbols and other mathematical expressions.

CLASSIFICATION OF QUANTITATIVE TECHNIQUES

Quantitative techniques are divided into two groups

1. Statistical techniques
2. Operations Research or Programming Techniques.

STATISTICAL TECHNIQUES

Statistical techniques are used in conducting the statistical inquiry concerning a certain phenomenon. They are including all the statistical methods beginning from the collection of data till the task of interpretation of the collected data. Includes all statistical method used for collection of data to its interpretation. It includes the statistical techniques like:

i. Methods of collection of statistical data
ii. Techniques of classification and tabulation of data collected.
iii. Calculation of various statistical measures such as Arithmetic Mean, Standard Deviation, Co-efficient of Correlation etc.,
iv. Techniques of analysis and interpretation.
v. Deriving inferences and judging their reliability

STATISTICAL METHODS AND MEASURES

(i) Methods of collecting data
(ii) Classification and tabulation of collected data
(iii) Probability theory and Sampling Analysis
(iv) Correlation and regression
(v) Index number
(vi) Time series Analysis
(vii) Interpolation and Extrapolation
(viii) Survey Techniques and Methodology
(ix) Ratio Analysis
(x) Statistical quality control
(xi) Analysis of Variance
(xii) Statistical Inferences and Interpretation
(xiii) Theory of Attributes

PROGRAMMING TECHNIQUES

It can be defined as operational research or simply (O.R.) are the model building techniques used by decision maker in modern times. They include wide variety of techniques such as linear programming, theory of games, simulation, network analysis, queuing theory and many other similar techniques.

(i) Linear Programming
(ii) Decision Theory
(iii) Theory of Games
(iv) Simulation: a. Monte Carlo Techniques b. System Simulation
(v) Waiting Line (queuing) Theory
(vi) Inventory Planning
(vii) Integrated Production Models
(viii) Network Analysis/ PERT
(ix) Others - a. Non- Linear Programming
b. Dynamic Programming
c. Search Theory
d. Integer Programming
e. Quadratic Programming
f. Parametric Programming
g. The Theory of Replacement etc.

ROLE OF QUANTITATIVE TECHNIQUES

These techniques are especially increasing since World War II in the technology of business administration. These techniques help in solving complex and intricate problems of business and industry. Quantitative techniques for decision making are, in fact, examples of the use of scientific method of management. Their role can be well understood under the following heads

(i) Provide a tool for scientific analysis
(ii) Provide solution for various business problems
(iii) Enable proper deployment of resources
(iv) Helps in minimizing waiting and servicing costs

(v) Assists in choosing an optimum strategy
(vi) They render great help in optimum resource allocation
(vii) Enable the management to decide when to buy and how much to buy
(viii) They facilitate the process of decision making
(ix) Through various quantitative techniques management can know the reactions of the integrated business systems

LIMITATIONS

Quantitative techniques though are a great aid to management but still they cannot be substitute for decision making. The choice of criterion as to what is actually best for the business enterprise is still that of an executive who has to fall back upon his experience and judgement. This is so because of the several limitations of quantitative techniques. Important limitations of these techniques are as given below:

(i) The inherent limitation concerning mathematical expressions
(ii) High costs are involved in the use of quantitative techniques
(iii) Quantitative techniques do not take into consideration the intangible factors i.e., non measurable human factors
(iv) Quantitative techniques are just the tools of analysis and not the complete decision making

STEPS INVOLVED IN APPLICATION OF PROGRAMMING TECHNIQUES.

All quantifiable factors are defined in mathematical language viz.,

Variable => Factors which are controllable parameters.

Co-efficient => Factors which are not controllable.

Appropriate mathematical expressions [equations] also known as mathematical model is a formula which describes the interrelations of all variables and parameters.

An optimum solution is determined on the basic of various equations of the model which maximize profit or reduce cost or fulfil some other goal.

The solution value of the model obtained is tested against actual observations and the model modified as per requirements till satisfactory result is obtained.

Finally the solution is put to work programming techniques involve:-

Building up to the mathematical model i.e., sets of equations and inequations.

Which give solutions to problems in terms of the values of the variables involved.

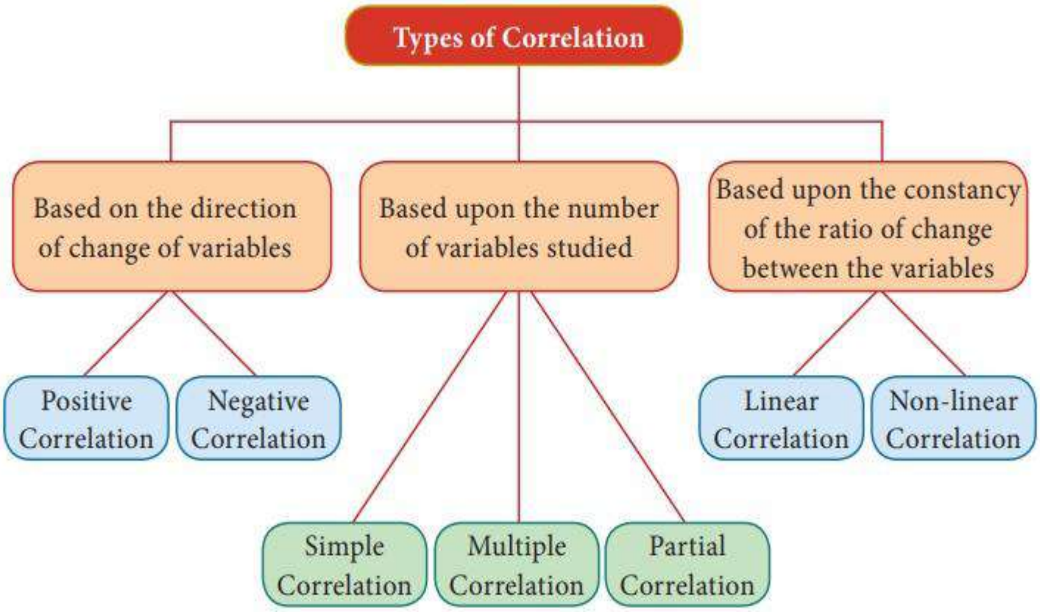
CORRELATION

Correlation may be defined as a measure of relation between two variables. In other words, to describe and understand the association between two continuous variables (interval or ratio data), we compute Correlation. Thus two variables are said to be correlated if an increase (or decrease) in one variable is accompanied by an increase or decrease (decrease or increase) in the other variable.

TYPES OF CORRELATION

Correlation is classified in several different ways:

1. Positive or Negative correlation
2. Linear and Non-linear correlation
3. Simple, Multiple and Partial correlation
4. Spurious correlation



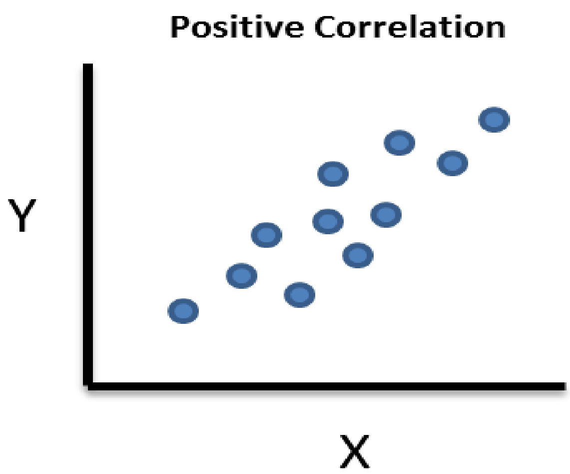
POSITIVE CORRELATION

Two variables are said to be positively correlated, if an increase in the value of one is accompanied by an increase in the value of other or a decrease in the value of one is accompanied by a decrease in the value of other i.e. the value of two variables deviate in the same direction. For example, paired variables like supply and demand of commodities, household income and expenditure, price and supply of commodities is positively correlated.

Examples:

X :	10	12	14	16	18
Y :	15	20	25	30	35

X :	80	75	70	60	50
Y :	6	5	4	2	1



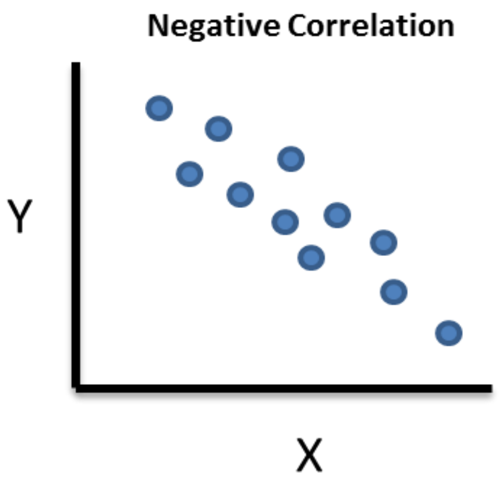
NEGATIVE CORRELATION

Two variables are said to be negatively correlated, if an increase in the value of one is accompanied by a decrease in the value of other or a decrease in the value of one is accompanied by an increase in the value of other. For example, paired variables like pressure and volume of a gas, current and resistance (voltage being constant), demand and price of commodities are negatively correlated.

Examples:

X :	10	12	14	16	18
Y :	15	12	10	8	6

X :	80	75	70	60	50
Y :	60	70	75	80	90



LINEAR CORRELATION

The correlation between two variables is said to be linear, if there exists a relationship of the form:

$$y = a + bx$$

where, a and b are real numbers.

In a linear correlation the amount of change in one variable tends to bear constant ratio to the amount of change in the other variable. The graph of linear correlation is a straight line.

Example:

X :	10	12	14	16	18
Y :	15	20	25	30	35

NON LINEAR CORRELATION

The correlation between two variables is said to be non - linear, if there exists a relationship of the form of polynomial of order more than one, for example:

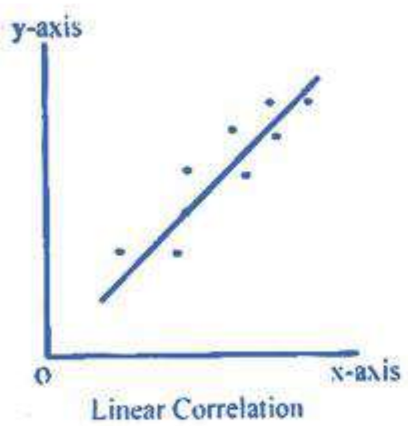
$y = a + bx + cx2$

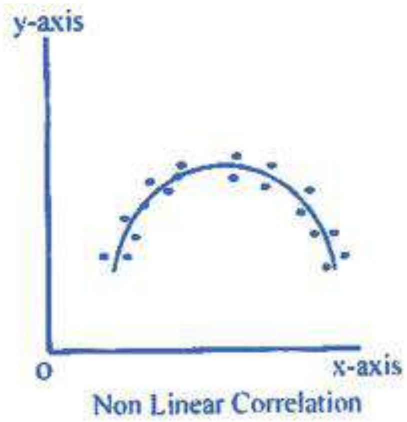
where, a , b and c are real numbers.

In the non - linear correlation, a change of one unit in one variable does not correspond to same amount of change in the other variable. The graph of non - linear correlation is not a straight line.

Example:

X :	10	12	14	16	18
Y :	15	20	22	30	40





SIMPLE CORRELATION

In simple correlation the number of variables studied/ included is two. All the above mentioned correlations are simple; including positive, negative, linear and non-linear.

MULTIPLE CORRELATION

When three or more variables are studied simultaneously it is a case of multiple correlations. For example, when we study the relationship between the yield of rice per acre, amount of rainfall and the amount of fertilizers used, it is a problem of multiple correlations.

PARTIAL CORRELATION

In partial correlation we recognize more than two variables, but consider only two variables to be influencing each other the effect of other influencing variables being kept constant. For example, in the rice problem taken above if we limit our correlation analysis of yield and rainfall to periods when a certain average daily temperature exist it becomes a problem relating to partial correlation only.

SPURIOUS CORRELATION

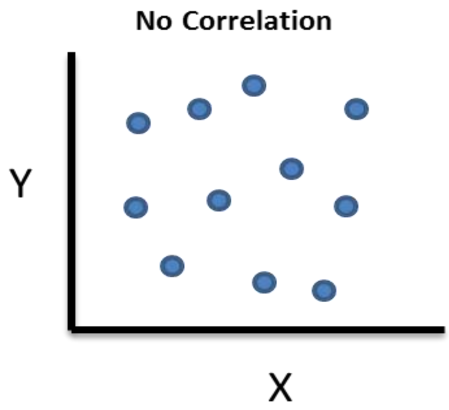
Given data regarding any two variables, it is possible that on calculation of r, one may say that statistically the two variables are correlated, but if there is no justifiable, logical explanation for correlation to exists, then such statistical correlation is termed as spurious correlation. In a spurious correlation two variables have no direct causal connection, yet it may be wrongly inferred that they do, due to either coincidence or the presence of a certain third, unseen factor (referred to as a “confounding factor” or “lurking variable”). Suppose there is found to be a correlation between A and B. Aside from coincidence, there are three possible relationships:

A causes B,
B causes A,
OR
C causes both A and B.

In the last case there is a spurious correlation between A and B. In a regression model where A is regressed on B but C is actually the true casual factor for A, this misleading choice of independent variable (B instead of C) is called specification error.

NO CORRELATION.

No correlation means there is no relationship among the variables.



CORRELATION COEFFICIENT - DIRECT METHOD

The Coefficient of Correlation ‘r’ is a measure of the degree of linear relationship between two variables, say, x and y i.e. it measures the degree of association between the two values of related variables given in the data set.

(i) The coefficient of correlation ‘r’ (also known as Karl Pearson’s Coefficient of Correlation) is given by:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

Where

$$X = X - \bar{X}$$

$$\bar{X} = \frac{\Sigma X}{N}$$

$$Y = Y - \bar{Y}$$

$$\bar{Y} = \frac{\Sigma Y}{N}$$

N = Number of observations

The value of coefficient of correlation 'r' lies between - 1 and + 1. If for a given two sets or data:

a) r = +1, the two sets or data are said to be perfect positively correlated.
b) r = - 1, the two sets or data are said to be perfect negatively correlated.
c) r = 0, the two sets or data are said to be uncorrelated, that is, there is absence of any linear relationship between the variables. However, there may exist some other form of relationship between them
d) r = 0.8, the two sets or data are said to be strongly correlated.
e) r = 0.2, the two sets or data are said to be weakly correlated.

Properties of Coefficient of Correlation 'r'

a) The value of 'r' is independent of the origin of reference and the scale of reference i.e. 'r' is not affected by addition or subtraction of a constant to the values of either or both variables. It is also, unaffected by the multiplication or division of the values of either or both variables, by a constant.

b) The value of 'r' is free from any of the units. In other words, 'r' will have a definite meaning, whether the units of x and y variables are comparable or not.

c) The coefficient of correlation is symmetrical in two variables.

d) The value of 'r' lies between + 1 and - 1.

Illustration

1. Calculate the coefficient of correlation between the height of father and son from the following data:

Height of Father (in Inches) (x) : 64 65 66 67 68 69 70
Height of Son (in Inches) (y) : 66 67 65 68 70 68 72

Solution:

Height of Father (in inches) x	Height of Son (in inches) y	X= (X – 67)	Y=(Y– 68)	X^2	Y^2	XY
64	66	-3	-2	9	4	6
65	67	-2	-1	4	1	2
66	65	-1	-3	1	9	3
67	68	0	0	0	0	0
68	70	1	2	1	4	2
69	68	2	0	4	0	0
70	72	3	4	9	16	12
Σx=469	Σy=476	ΣX=0	ΣY=0	Σx^2=28	Σy^2=34	Σxy=25

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$\bar{X}$ = ΣX/N = 469/7 = 67
$\bar{Y}$= ΣY/N = 476/7 = 68
Σxy=25, Σx^2=28, Σy^2=34

Substituting the values in the formula,

$$= \frac{25}{\sqrt{28x34}}$$

$$= \frac{25}{\sqrt{952}}$$

$$= \frac{25}{30.85}$$

= 0.81

r = 0.81

Height of the father and height of the son is highly correlated.

ASSUMED MEAN METHOD

The above method of calculating correlation is based on Arithmetic Mean and the correlation can also be calculated using assumed mean, the following formula is used.

$$\mathbf{r} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

where dx = X - A
dy = Y - A
A = Assumed Mean
N = Number of Observations.
Σdx = Total of dx
Σdy = Total of dy
Σdxdy = Total of dxdy
C = Common factor (can be used wherever / whenever necessary).

2. Find the co-efficient of Correlation from the following data

X: 300	350	400	450	500	550	600	650	700
Y: 800	900	1000	1100	1200	1300	1400	1500	1600

SOLUTION

X	**Y**	dx=$\frac{X-A}{c}$	dy = $\frac{Y-A}{c}$	**dx²**	**dy²**	**dxdy**
300	800	-2	-4	4	16	8
350	900	-1.5	-3	2.25	9	4.5
400	1000	-1	-2	1	4	2

450	1100	-0.5	-1	0.25	1	0.5
500	1200	0	0	0	0	0
550	1300	0.5	1	0.25	1	0.5
600	1400	1	2	1	4	2
650	1500	1.5	3	2.25	9	4.5
700	1600	2	4	4	16	8
		Σdx=0	Σdy=0	$\Sigma dx^2=15$	$\Sigma dy^2=60$	Σdxdy=30

$$\mathbf{r} = \frac{N\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

where dx = X - A
dy = Y - A
A = Assumed Mean
N = Number of Observations.
Σdx = Total of dx
Σdy = Total of dy
Σdxdy = Total of dxdy
C = Common factor (can be used wherever necessary).

A (X) = 500; A(Y) = 1200; C = 100; Σdx=0; Σdy=0; $\Sigma dx^2=15$; $\Sigma dy^2=60$; Σdxdy=30; N=9

Substituting the values in the formula

$$= \frac{9(30) - (0)\,(0)}{\sqrt{9(15) - (0)^2}\sqrt{9(60) - (0)^2}}$$

$$= \frac{270}{\sqrt{135}\sqrt{540}}$$

$$= \frac{270}{11.62\ x\ 23.24}$$

$$= \frac{270}{270.0488}$$

r = 0.9998

r = 1

There is perfect positive relationship between variables.

PARTIAL CORRELATION

To measure the correlation between a dependent variable and one particular independent variable when all other variables involved are kept constant i.e., other things being equal, can be obtained by calculating co-efficient of partial correlation.

By $\mathbf{r_{12.3}}$, the co-efficient of partial correlation between **X_1 and X_2 keeping X_3 constant,** we find that

$$r_{12.3} = \frac{r_{12} - r_{13}\ r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}}$$

Where $\mathbf{r_{13\cdot2}}$ is the co-efficient of partial correlation between **X_1 and X_3 keeping X_2 constant.**

$$r_{23.1} = \frac{r_{23} - r_{12}\ r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}}$$

Where $\mathbf{r_{23.1}}$ is the co-efficient of partial correlation between **X_2 and X_3 keeping X_1 constant.**

ZERO ORDER, FIRST ORDER AND SECOND ORDER CO-EFFICIENTS

Partial Co-efficients such as $r_{12.3}$, $r_{13.2}$ are often referred to as ***first order co-efficient*** since one variable has been hold constant.

Simple Co-efficients (correlation between two variables only) are called ***Zero order Co-efficients***, since no variables are held constant.

$r_{12.34}$, $r_{13.24}$ etc., are called ***second order Co-efficients*** since two variable are kept constant.

SECOND ORDER PARTIAL CORRELATION CO-EFFICIENTS.

If $r_{12.34}$ is the co-efficient of partial correlation between X_1 and X_2 keeping X_3 and X_4 constant, then.

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} \cdot r_{24.3}}{\sqrt{1-r_{14.3}^2}\ \sqrt{1-r_{24.3}^2}}$$

$$r_{13.24} = \frac{r_{13.2} - r_{14.2} \cdot r_{34.2}}{\sqrt{1-r_{14.2}^2}\sqrt{1-r_{34.2}^2}}$$

$$r_{14.23} = \frac{r_{14.2} - r_{13.2} \cdot r_{34.2}}{\sqrt{1-r_{12.3}^2}\sqrt{1-r_{34.1}^2}}$$

Illustration

1. On the basis of the following information. Compute (i) $r_{23.1}$ (ii) $r_{13.2}$ (iii) $r_{12.3}$ where

$r_{12} = 0.70$ $r_{13} = 0.61$ $r_{23} = 0.40$

$$r_{23.1} = \frac{r_{23} - r_{12}\ r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}}$$

Substituting the values in the formula

$$r_{23.1} = \frac{0.40-0.70\ .0.61}{\sqrt{1-0.70^2}\ \sqrt{1-0.61^2}}$$

$$= \frac{0.40-0.43}{\sqrt{1-0.70^2}\ \sqrt{1-0.61^2}}$$

$$= \frac{-0.03}{\sqrt{1-0.49}\ \sqrt{1-0.37}}$$

$$= \frac{-0.03}{\sqrt{0.51}\ \sqrt{0.63}} = \frac{-0.03}{0.71\ x\ 0.79}$$

$$= \frac{-0.03}{0.56}$$

$\mathbf{r_{23.1}}$ **= - 0.05357**

$$\mathbf{r_{13.2}} = \frac{\boldsymbol{r_{13} - r_{12}\, r_{23}}}{\sqrt{\mathbf{1}-\boldsymbol{r_{12}^2}}\sqrt{\mathbf{1}-\boldsymbol{r_{23}^2}}}$$

$$= \frac{0.61-0.70x0.40}{\sqrt{1-0,70^2}\ \sqrt{1-0.40^2}}$$

$$= \frac{0.61-0.28}{\sqrt{1-0.49}\,\sqrt{1-0.16}}$$

$$= \frac{0.33}{\sqrt{0.51}\,\sqrt{0.84}}$$

$$= \frac{0.33}{0.71\ x\ 0.92}$$

$$= \frac{0.33}{0.6532}$$

$$\mathbf{r_{13.2} = 0.5052}$$

$$\mathbf{r_{12.3}} = \frac{\boldsymbol{r_{12} - r_{13}\ r_{23}}}{\sqrt{\boldsymbol{1-r_{13}^2}}\,\sqrt{\boldsymbol{1-r_{23}^2}}}$$

$$= \frac{0.70-0.61x0.40}{\sqrt{1-0.61^2}\,\sqrt{1-0.40^2}}$$

$$= \frac{0.70\ -\ 0.2444}{\sqrt{1-0.37}\,\sqrt{1-0.16}}$$

$$= \frac{0.456}{\sqrt{0.63}\,\sqrt{0.84}}$$

$$= \frac{0.456}{0793x0.916}$$

$$= \frac{0.456}{0.726}$$

$$\mathbf{r_{12.3} = 0.6281}$$

MULTIPLE CORRELATION

The co-efficient of multiple linear correlations is represented by R_1 and it is common to add subscripts designating the variables involved. Thus $R_{1.234}$ would represent the co-efficient of multiple linear correlation between X_1 on the one hand and X_2, X_3 and X_4 on the other. The subscript of the dependent variable is always to the left of the point.

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2\,r_{12}r_{13}r_{23}}{1-r_{23}^2}}$$

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}\,r_{13}\,r_{23}}{1-r_{13}^2}}$$

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2\,r_{12}\,r_{13}\,r_{23}}{1-r_{12}^2}}$$

It should be noted that $R_{1.23}$ is the same as $R_{1.32.}$

A co-efficient of multiple correlation such as R1.23 lies between 0 and 1. The closer it is to 1, the better is the linear relationship between variable. The closer it is 0, the worse is the linear relationship. If the co-efficient of multiple correlations is 1, the correlation is called perfect. Although a correlation co-efficient of 0 indicates no linear relationship between the variables, it is possible that a non-linear relationship may exist.

It should be noted that whereas the correlation co-efficients range from +1 to 0 to -1, **the co-efficients of multiple correlations are always positive in sign and range from + 1 to 0.**

The variables whose value is estimated are called dependent variable and the other variables on which estimates are based are known as independent variable.

ILLUSTRATION

1. Calculate (i) $R_{1.23}$ (ii) $R_{3.12}$ (iii) $R_{2.13}$ for the following data, r_{12} = 0.6, r_{13} = 0.7 and r_{23} = 0.65.

$$\mathbf{R_{1.23}} = \sqrt{\frac{\boldsymbol{r_{12}^2 + r_{13}^2 - 2\,r_{12}r_{13}r_{23}}}{\boldsymbol{1-r_{23}^2}}}$$

$$R_{1.23} = \sqrt{\frac{0.6^2 + 0.7^2 - 2(0.6)(0.7)(0.65)}{1-0.65^2}}$$

$$R_{1.23} = \sqrt{\frac{0.36+0.49-0.546}{1-0.4225}}$$

$$R_{1.23} = \sqrt{\frac{0.85-0.546}{0.5775}}$$

$$R_{1.23} = \sqrt{\frac{0.304}{0.5775}}$$

$$R_{1.23} = \sqrt{0.5264}$$

$$R_{1.23} = 0.7255$$

$$\mathbf{R_{2.13}} = \sqrt{\frac{\boldsymbol{r_{12}^2 + r_{23}^2 - 2r_{12}\,r_{13}\,r_{23}}}{\boldsymbol{1-r_{13}^2}}}$$

$$R_{2.13} = \sqrt{\frac{(0.6)^2 + (0.65)^2 - 2(0.6)(0.7)(0.65)}{1-(0.7)^2}}$$

$$R_{2.13} = \sqrt{\frac{0.36 + 0.4225 - 0.546}{1 - 0.49}}$$

$$R_{2.13} = \sqrt{\frac{0.7825 - 0.546}{0.51}}$$

$$R_{2.13} = \sqrt{\frac{0.2365}{0.51}}$$

$$R_{2.13} = \sqrt{0.4637}$$

$$\mathbf{R_{2.13} = 0.6809}$$

$$\mathbf{R_{3.12}} = \sqrt{\frac{\boldsymbol{r_{13}^2 + r_{23}^2 - 2\, r_{12}\, r_{13}\, r_{23}}}{\boldsymbol{1 - r_{12}^2}}}$$

$$R_{3.12} = \sqrt{\frac{(0.7)^2 + (0.65)^2 - 2\,(0.6)(0.7)(0.65)}{1 - (0.6)^2}}$$

$$R_{3.12} = \sqrt{\frac{0.49 + 0.4225 - 0.546}{1 - 0.36}}$$

$$R_{3.12} = \sqrt{\frac{0,9125 - 0.546}{0.64}}$$

$$R_{3.12} = \sqrt{\frac{0.3665}{0.65}}$$

$$R_{3.12} = \sqrt{0.5638}$$

$$\mathbf{R_{3.12} = 0.7508}$$

THEORYQUESTIONS
SHORT ANSWER QUESTIONS

1. Define quantitative techniques/Correlation
2. What is meant by statistical techniques/Programming techniques/ partial correlation /Multiple correlation/Positive correlation/Negative Correlation/Linear Correlation/ Non-linear Correlation/Spurious Correlation/ First order / Second order /Zero order correlation Co-efficients?

ESSAY TYPE QUESTIONS

1. Elucidate the various types of Correlation with diagrams and examples.
2. Enumerate the classification of quantitative techniques

3. Explain the role of quantitative techniques in business.
4. List out the Statistical techniques/Programming techniques
5. Discuss the properties of Correlation Co-efficients.
6. Explain the limitations of quantitative techniques.
7. State the steps involved in the programming techniques.

UNIT II
TESTING OF HYPOTHESES

STATISTICAL INFERENCE

Statistical inference is that branch of statistics which is concerned with using probability concept to deal with uncertainty in decision making. It refers to the process of selecting and using a sample statistic to draw inference about a population parameter based on a subset of it - the sample drawn from the population. Statistical inference treats two different classes of problems:

1.Hypothesis testing i.e., to test some hypothesis about parent population from which the sample is drawn.

2. Estimation, i.e., to use the "statistics" obtained from the sample as estimate of the unknown "parameter" of the population from which the sample is drawn.

HYPOTHESIS TESTING

Hypothesis begins with an assumption, called hypothesis, which we make about a population parameter.

DEFINITION - Prof. MORRIS HAMBURG.

"A hypothesis in statistics is simply a quantitative statement about the population ".

PROCEDURE FOR TESTING HYPOTHESIS

i. Set up a hypothesis
ii. Set up a suitable significance level
iii. Setting a test criterion
iv. Doing computations
v. Making decisions.

TYPES OF HYPOTHESES

The two hypotheses in a statistical test are normally referred to as

i. Null hypothesis, and
ii. Alternative hypothesis.

NULL HYPOTHESIS

It asserts that there is no real difference in the sample and the population in the particular matter under consideration and it is denoted by Ho.

ALTERNATIVE HYPOTHESIS

Any hypothesis which is complementary to the null hypothesis (which the researcher believes to hold true) is called Alternative hypothesis and is denoted by Hα (H_1)

LEVEL OF SIGNIFICANCE

The maximum probability with which the researcher would be willing to take risk is called level of significance and is normally expressed in percentage.

TWO TYPES OF ERRORS IN TESTING OF HYPOTHESIS.

When a statistical hypothesis is tested there are four possibilities

a) The hypothesis is true but our test rejects it. (Type I error)
b) The hypothesis is false but our test accepts it (Type II error)
c) The hypothesis is true and our test accepts it (Correct decision)
d) The hypothesis is false and our test rejects it (Correct decision)

In a statistical hypothesis testing experiment, a Type I error is committed by rejecting the null hypothesis when it is true. The probability of committing a Type I error is denoted by a α, where

α = prob. (Type I error)

= prob (Rejecting Ho/H_1 is true)

On the other hand, a Type II error is committed by not rejecting (i,e., accepting) the null hypothesis when it is false. The probability of committing a Type II error is denoted by β, where

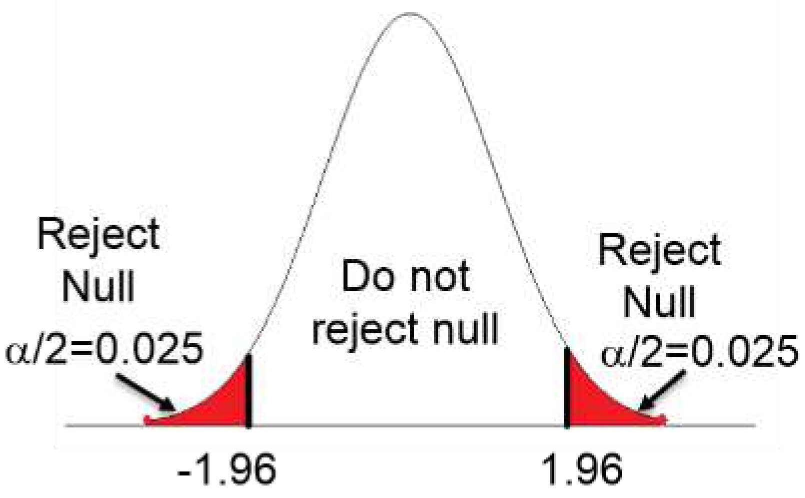
β = prob (Type II error)

= prob (not rejecting or accepting Ho/H_1 is false)

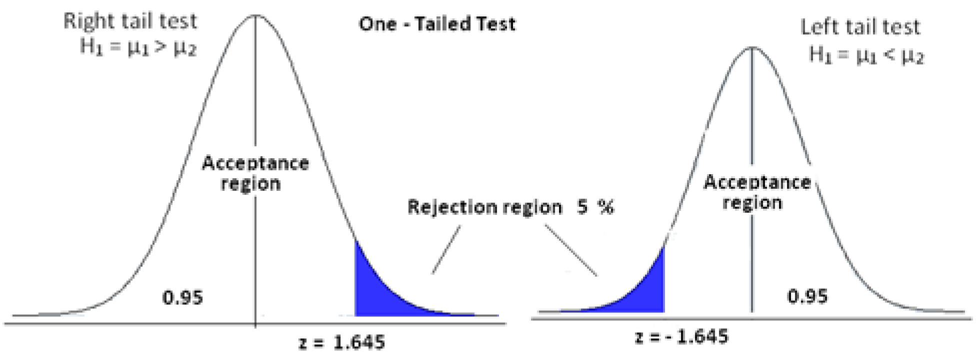
	Accept Ho	***Reject H_1***
Ho is true	Correct decision	Type I error
Ho is false	Type II error	Correct decision

TWO TAILED AND ONE TAILED TESTS OF HYPOTHESIS

A two tailed test of hypothesis will reject the null hypothesis, if the sample statistic is significantly higher than or lower than the hypothesised population parameter. Thus in a two tail test the rejection region is located in both the tails.



One tailed test is so called because the rejection region will be located in only one tail which may be either left or right depending upon the alternative hypothesis formulated.



χ^2 TEST AND GOODNESS OF FIT

χ^2 test (pronounced as Chi-square test) describes the magnitude of the discrepancy between theory and observation. It is defined as:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where O = Observed frequency
E = Expected frequency

STEPS

1. Calculate the expected frequencies $E = \frac{RT \ x \ CT}{N}$

Where E = Expected frequency
RT = the row total for the row containing the cell
CT = The column total for the column containing the cell
N = Number of observations.

2. Take the difference between observed and expected frequencies and obtain the squares of these differences i.e., obtain the values of $(O-E)^2$

3. Divide the values of $(O-E)^2$ obtained in step (ii) by the respective expected frequency and obtain the total $\sum \frac{(O-E)^2}{E}$. This gives the value of χ^2 which can range from O to which can range from O to ∞. The calculated value of χ^2 is compared with the table value of χ^2 for given degrees of freedom at a certain specified level of significance.

DEGREES OF FREEDOM

It means the number of classes to which the values can be assigned arbitrarily or at will without violating the restrictions or limitations placed. Symbolically υ (pronounced as Nu) or d.f, obtained by

$$\upsilon = n - k$$

where n = Number of observations

k = Number of independent constraints

In a contingency table the degrees of freedom are calculated in a slightly different manner.

$$\upsilon = (r-1)(c-1)$$

where r = number of rows
c = number of columns.

χ^2 TEST AS A TEST OF GOODNESS OF FIT.

χ^2 test is very popularly known as test of goodness of fit for the reason that it enables us to ascertain how appropriately the theoretical distributions such as Poison, Normal, etc., fit empirical distributions i.e., those obtained from sample data.

i. A null and alternative hypothesis is established and a significance level is selected for rejection of the null hypothesis.
ii. A random sample of observations is drawn from a relevant statistical population.
iii. A set of expected or theoretical frequencies is derived under the assumption that the null hypothesis is true.
iv. The observed frequencies are compared with the expected or theoretical frequencies
v. If the **calculated value of χ^2 is less than the table value** at a certain level of significance and for certain degrees of freedom the fit is considered to be good **(accepted).** On the other hand, if **the calculated value of χ^2 is greater than the table value**, the fit is considered to be poor **(rejected).**

ILLUSTRATION

1. In an anti malarial campaign in a certain area, quinine was administered to 1,624 persons out of a total population of 6496. The number of fever cases is shown below:

Treatment	Fever	No fever	**Total**
Quinine	40	1,584	1,624
No Quinine	440	4,432	4,872
Total	**480**	**6,016**	**6,496**

Discuss the usefulness of quinine in checking malaria.

SOLUTION

Let us frame hypotheses.

Ho = Quinine is not effective in checking malaria
H_1 = Quinine is effective in checking malaria.

Applying χ^2 test

$$E = \frac{RT\ x\ \ CT}{N} \quad (40) = \frac{480\ x\ 1624}{6496} = 120$$

$$(440) = \frac{480\ x\ 4,872}{6496} = 360$$

$$(1584) = \frac{6016\ x\ 1624}{6496} = 1504$$

(4432) $= \frac{6016 \, x \, 4872}{6496} = 4512$

The table of expected frequencies shall be

120	**1504**	**1624**
360	4512	4872
480	6016	6496

O	E	(O-E)	$(O - E)^2$	$\frac{(O-E)^2}{E}$
40	120	-80	6400	53.333
440	360	80	6400	17.778
1584	1504	80	6400	4.255
4432	4512	-80	6400	1.419
				76.785

$\sum\frac{(O-E)^2}{E} = 76.785$

$$\chi^2 = \sum\frac{(O-E)^2}{E} = 76.\,785$$

$$\upsilon = (r-1)(c-1) = (2-1)(2-1) = 1$$

for $\upsilon = 1, \quad \chi^2_{0.05} = 3.84$

Calculated value > table value

76.785 > 3.84

CONCLUSION

The calculated value of χ^2 is greater than the table value. Hence the null hypothesis is rejected. It can be concluded that quinine is useful in checking in malaria.

2. 200 digits are chosen at random from a set of tables. The frequencies of the digits are as follows:

Digit:	0	1	2	3	4	5	6	7	8	9
Frequency:	18	19	23	21	16	25	22	20	21	15

Use χ^2 test to assess the correctness of the hypothesis that the digits were distributed in equal numbers in the tables from which they are chosen.

SOLUTION

STEP 1 Framing of hypothesis

Ho = The digits were distributed in equal numbers.

H_1 = The digits were not distributed in equal numbers.

Calculation of Expected frequencies.

$$E = \frac{200}{10} = 20.$$

Applying the χ^2 test

O	E	(O-E)	$(O-E)^2$	$(O-E)^2 / E$
18	20	2	4	0.20
19	20	1	1	0.05
23	20	3	9	0.45
21	20	1	1	0.05
16	20	4	16	0.80
25	20	5	25	1.25
22	20	2	4	0.20
20	20	0	0	0
21	20	1	1	0.05
15	20	5	25	1.25
ΣO =200	**ΣE=200**			**$\Sigma(O-E)^2 /E=4.3$**

$$\upsilon = n - k$$
$$= 10-1 = 9$$
$$\upsilon = 9 \qquad \chi^2_{0.05} = 16.22$$

CALCULATED VALUE < TABLE VALUE

4.3 < 16.22

Hence the calculated value of χ^2 is less than the table value the null hypothesis is accepted.

Therefore it is concluded that the digits were distributed in equal numbers.

THE F TEST OR THE VARIANCE RATIO TEST.

The object of the F test is to find out whether the two independent estimates of population variance differ significantly or whether the two samples may be regarded as drawn from the normal populations having same variance.

$$F = \frac{S_1^2}{S_2^2}$$

$$S_1^2 = \frac{\Sigma(X_1-\overline{X_1})^2}{n_1-1}$$

$$S_2^2 = \frac{\Sigma(X_2-\bar{X}_2)^2}{n_2-1}$$

It should be noted that S_1^2 is always the larger estimate of variance i.e., $S_1^2 > S_2^2$.

$$F = \frac{Larger\ estimate\ of\ variance}{Smaller\ estimate\ of\ variance}$$

$\upsilon_1 = n_1 - 1$ and $\upsilon_2 = n_2 - 1$

υ_1 = degrees of freedom for sample having larger variance

υ_2 = degrees of freedom for sample having smaller variance.

Calculated value of F < Table value of F => Null hypothesis accepted
Calculated value of F > Table value of F => Null hypothesis rejected.

ILLUSTRATION

1. Two random samples were drawn from two normal populations and their values are

A: 66 67 75 76 82 84 88 90 92

B: 64 66 74 78 82 85 87 92 93 95 97.

Test whether the two populations have the same variance at the 5% level of significance. (F = 3.36) at 5% level for $\upsilon_1 = 10$ and $\upsilon_2 = 8$.

SOLUTION

STEP 1 Framing of hypothesis

Ho = Two populations have same variance
H_1 = Two populations does not same variance

STEP 2 Applying F Test
Applying F test

$$F = \frac{S_1^2}{S_2^2}$$

A - X_1	$(X_1 - \overline{X_1})$	$(X_1 - \overline{X_1})^2$	**B - X_2**	$(X_2 - \bar{X}_2)$	$(X_2 - \bar{X}_2)^2$
66	-14	196	64	-19	361
67	-13	169	66	-17	289
75	-5	25	74	-9	81
76	-4	16	78	-5	25
82	2	4	82	-1	1
84	4	16	85	2	4
88	8	64	87	4	16
90	10	100	92	9	81
92	12	144	93	10	100
			95	12	144
			97	14	196
720	**0**	**734**	**913**	**0**	**1298**

$\Sigma X_1 = 720$; $\Sigma(X_1 - \overline{X_1}) = 0$; $\Sigma(X_1 - \overline{X_1})^2 = 734$; $\Sigma X_2 = 913$; $\Sigma(X_2 - \bar{X}_2) = 0$; $\Sigma(X_2 - \bar{X}_2)^2$ =1298

$$\bar{\bar{X}}_1 = \frac{\Sigma X_1}{N_1} = \frac{720}{9} = 80$$

$$\bar{\bar{X}}_2 = \frac{\Sigma X_2}{N_2} = \frac{913}{11} = 83$$

$$S_1^2 = \frac{\Sigma(X_1 - \overline{X_1})^2}{n_1 - 1} = \frac{734}{9-1} = 91.75$$

$$S_2^2 = \frac{\Sigma(X_2 - \bar{X}_2)^2}{n_2 - 1} = \frac{1298}{11-1} = 129.8$$

$$F = \frac{S_1^2}{S_2^2} = \frac{91.75}{129.8} = 0.707$$

Calculation of degrees of freedom.

For $\upsilon_1 = 8(9-1)$ $\upsilon_2 = 10(11-1)$ and F value = 3.36

Comparing calculated value and table value

Calculated value < Table value
0.707 < 3.36

As the calculated value is less than the table value null hypothesis is accepted. Hence it is concluded that the two populations has same variance.

ANALYSIS OF VARIANCE

The analysis of variance referred to by the contraction ANOVA is designed to test whether the means of more than two quantitative populations are equal. Its purpose is to test for the significance of the differences among sample means.

TECHNIQUE OF ANALYSIS OF VARIANCE

The technique of analysis of variance has been discussed separately for

i. One way classification
ii. Two way classification.

ONE WAY CLASSIFICATION

In one way classification the data are classified according to only one criterion. The null hypothesis is:

Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots\ldots\ldots = \mu_k$

H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \ldots\ldots\ldots\ldots\ldots \neq \mu_k$

All the means are not equal.
i.e., the arithmetic means of populations from which the K samples were randomly drawn are equal to one another.

$$F = \frac{Between\ column\ variance}{Within\ column\ variance}$$

$$F = \frac{S_1^2}{S_2^2}$$

The specimen of **ANOVA table** is given below:

Sources of variation	**SS(Sum of Squares)**	**V(degrees of freedom)**	**MS (Mean Square)**	**Variance of Ratio of F**
Between samples	SSC	υ_1 = C-1	MSC = SSC/C-1	$F = \frac{MSC}{MSE}$
Within samples	SSE	υ_2=n-c	MSE= SSE/n-c	
Total	SST	n-1		

SST = Total of sum of squares of variation
SSC = Sum of squares between samples (columns)
SSE = Sum of squares within samples (rows)
MSC = Mean sum of squares between samples
MSE = Mean sum of squares within samples.

NOTE:
The same procedure for analysis of variance is applicable for both the equal and unequal sample sizes.

CORRECTION FACTOR

A factor that is multiplied with the result of an equation to correct for a known amount of systematic error.

Sum of all items of various samples = $\Sigma x_1+\Sigma x_2+\Sigma x_3+........\Sigma Kn$

Correction factor (CF) = $\frac{T^2}{N}$

T = Sum of all items of various samples
N = No of observations

Total sum of squares (SST) = $\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + ... - CF$

Sum of squares between the samples (SSC) (Columns) =

$$= \frac{(\Sigma X_!)^2}{N} + \frac{(\Sigma X_2)^2}{N} + \frac{(\Sigma X_3)^2}{N} \ldots - \text{CF}$$

Sum of squares within samples (SSE) = Total sum of squares - Sum of squares between samples (SST - SSC)

ILLUSTRATION

1. The three samples below have been obtained from normal populations with equal variance. Test the hypothesis that the sample means are equal:

8	7	12
10	5	9
7	10	13
14	9	12
11	9	14

The table value of F at 5% level of significance for $\upsilon_1 = 2$ and $\upsilon_2 = 12$ is 3.88.

SOLUTION

STEP 1: Setting of hypothesis
Ho = Sample means are not equal
H_1 = Sample means are equal.

STEP 2: CORRECTION FACTOR.

Correction factor (CF) = $\frac{T^2}{N}$

	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
	8	64	7	49	12	144
	10	100	5	25	9	81
	7	49	10	100	13	169
	14	196	9	81	12	144
	11	121	9	81	14	196
TOTAL	**50**	**530**	**40**	**336**	**60**	**734**

Correction factor (CF) = $\frac{T^2}{N} = \frac{150^2}{15} = \frac{42500}{15}$ = 1500

STEP 3: TOTAL SUM OF SQUARES

Total sum of squares (SST) = $\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \ldots\ldots$ - CF

= 530 +336+734 - 1500
= 1600 - 1500
SST = 100

STEP 4: SUM OF SQUARES

Sum of squares between the samples (SSC)(Columns) =

$$= \frac{(\Sigma X_!)^2}{N} + \frac{(\Sigma X_2)^2}{N} + \frac{(\Sigma X_3)^2}{N} \ldots - \text{CF}$$

$$= \frac{(50)^2}{5} + \frac{(40)^2}{5} + \frac{(60)^2}{5} - 1500$$

= 500 + 320 + 720 - 1500

= 1540 - 1500

SSC = 40

STEP 5 SUM OF SQUARES WITHIN SAMPLES

Sum of squares within samples (SSE) = Total sum of squares - Sum of squares between samples (SST - SSC)

= 100 - 40

SSE = 60

ANOVA Table

Sources of variation	**SS(Sum of Squares)**	**V(degrees of freedom)**	**MS (Mean Square)**	**Variance of Ratio of F**
Between samples	SSC = 40	V_1 = C-1 3-1 =2	MSC = SSC/C-1 = 40/2 = 20	F = $\frac{MSC}{MSE}$ = 20/5=4
Within samples	SSE = 60	V_2=n-c=15-3=12	MSE= SSE/n-c = 60/12=5	
Total	SST = 100	n-1 =15-1=14		

STEP 6 Comparing Calculated value and table value

For $\upsilon_1 = 2$ and $\upsilon_2 = 12$ $F_{0.05} = 3.88$

Calculated value > Table value

4 > 3.88

As the calculated value is greater than table value null hypothesis is rejected. Hence it is concluded that samples are equal.

ANALYSIS OF VARIANCE IN TWO WAY CLASSIFICATION MODEL

In a two-way classification the data are classified according to two different criteria or factors.

ANOVA Table: Two way classification Model.

Sources of variation	SS (Sum of Squares)	V(Degrees of freedom)	MS(Mean sum of squares)	Ratio of F
Between samples	SSC	(c-1)	MSC = $\frac{SSC}{(c-1)}$	MSC / MSE
Between rows	SSR	(r-1)	MSR = $\frac{SSR}{(r-1)}$	MSR/MSE

Residual or error	SSE	(c-1)(r-1)	MSE = $\frac{SSE}{(c-1)(r-1)}$	
Total	SST	n-1		

SSC = Sum of squares between columns
SSR = Sum of squares between rows
SSE = Sum of squares due to error
SST = Total sum of squares
SSE = SST - [SSC + SSR]

$F(\upsilon_1, \upsilon_2) = \frac{MSC}{MSE}$ $\quad\quad$ $F(\upsilon_1, \upsilon_2) = \frac{MSR}{MSE}$

Where $\upsilon_1 = (c-1)$, $\upsilon_2 = (c-1)(r-1)$ $\quad\quad$ $\upsilon_1 = (r-1)$, $\upsilon_2 = (c-1)(r-1)$

Sum of all items of various samples = $\Sigma x_1 + \Sigma x_2 + \Sigma x_3 + \ldots\ldots.. \Sigma Kn$

Correction factor (CF) = $\frac{T^2}{N}$

T = Sum of all items of various samples
N = No of observations

Total sum of squares (SST) = $\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \ldots\ldots.. - CF$

Sum of squares between the samples (Columns)

$$SSC = \frac{(\Sigma X_1)^2}{N} + \frac{(\Sigma X_2)^2}{N} + \frac{(\Sigma X_3)^2}{N} \ldots CF$$

Sum of squares between the samples (rows)

$$SSR = \frac{(\Sigma X_1)^2}{N} + \frac{(\Sigma X_2)^2}{N} + \frac{(\Sigma X_3)^2}{N} \ldots.. - CF$$

Sum of squares within samples (SSE) = Total sum of squares - Sum of squares between Samples - Sum of squares within samples [SST - SSC - SSR]

Calculated value of F > Table value of F, Ho is rejected otherwise accepted.

ILLUSTRATION

1. A tea company appoints four salesmen, A, B, C and D and observes their sales in three seasons - summer, winter and monsoon. The figures (in Lakhs0 are given in the following table

Seasons	Salesman				Season's Total
	A	B	C	D	
Summer	36	36	21	35	128
Winter	28	29	31	32	120
Monsoon	26	28	29	29	112

Total	90	93	81	96	360

I. Do the salesmen significantly differ in performance?
II. Is there significant difference between the seasons?

SOLUTION

STEP 1: Framing of hypothesis

HYPOTHESIS 1
Ho = Salesmen does not differ in performance
H_1 = Salesman differ in performance.

HYPOTHESIS 2
Ho = There is no significant differences between the seasons
H_1 = There is significant differences between the seasons.
.

STEP 2: CORRECTION FACTOR

Correction factor (CF) = $\frac{T^2}{N}$

	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
	36	1296	36	1276	21	441	35	1225
	28	784	29	841	31	961	32	1024
	26	676	28	784	29	841	29	841
TOTAL	**90**	**2756**	**93**	**2921**	**81**	**2243**	**96**	**3090**

Sum of all items of various samples = $\Sigma x_1+\Sigma x_2+\Sigma x_3+.\Sigma x_4$
= 90 + 93 + 81 + 96
= 360

$$= \frac{T^2}{N} = \frac{(360)^2}{12} = \frac{12900}{12}$$

CF = 10800

STEP 3 TOTAL SUM OF SQUARES

Total sum of squares (SST) = $\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \ldots\ldots\ldots$ - CF
= 2756 + 2921 + 2243 + 3090 - 10800
= 11010 - 10800
SST = 210

STEP 4
Sum of squares between the samples (SSC) (Columns)

$$= \frac{(\Sigma X_!)^2}{N} + \frac{(\Sigma X_2)^2}{N} + \frac{(\Sigma X_3)^2}{N} \ldots \text{ CF}$$

$$= \frac{(90)^2}{3} + \frac{(93)^2}{3} + \frac{(81)^2}{3} + \frac{(96)^2}{3} - 10800$$

= 2700 + 2883 + 2187 + 3072 -10800
= 10842 - 10800
SSC = 42

STEP 5 SUM OF SQUARES BETWEEN SAMPLES

Sum of squares between the samples (rows) (SSR)

$$\textbf{SSR} = \frac{(\Sigma X)^2}{N} + \frac{(\Sigma X_2)^2}{N} + \frac{(\Sigma X_3)^2}{N} \ldots\ldots - CF$$

$$= \frac{(128)^2}{4} + \frac{(120)^2}{4} + \frac{(112)^2}{4} - 10800$$

= 4096 + 3600 + 3136 - 10800
= 10832 - 10800
SSR = 32

STEP 6 SUM OF SQUARES WITHIN SAMPLES

Sum of squares within samples

(SSE) = Total sum of squares - Sum of squares between columns - Sum of squares between rows (SST - SSC - SSR)

= 210 - 42 - 32
SSE = 136

Sources of variation	**SS (Sum of Squares)**	**V(Degrees of freedom)**	**MS(Mean sum of squares)**	**Ratio of F**
Between samples	SSC =42	(c-1) = 4-1=3	MSC = $\frac{SSC}{(c-1)} = \frac{42}{3}$ = 14	MSC / MSE = $\frac{14}{22.66}$ = 0.62
Between rows	SSR = 32	(r-1)=3-1=2	MSR = $\frac{SSR}{(r-1)} = \frac{32}{2}$ = 16	MSR/MSE = $\frac{16}{22.66}$ = 0.70
Residual or error	SSE = 136	(c-1)(r-1) =3x2 =6	MSE = $\frac{SSE}{(c-1)(r-1)} = \frac{136}{6}$ = 22.66	
Total	SST = 210	n-1 = 11		

$\upsilon_1(3,6)$ $F_{0.05}$= 8.94
Calculated value < Table value
0.62 < 8.94

As the calculated value is less than the table value null hypothesis is accepted. Hence it can be concluded that the salesman does not significantly differ in performance.

$\upsilon_2(2,6)$ $F_{0.05}$ = 19.33

Calculated value < Table value

0.70 < 19.33

As the calculated value is less than the table value null hypothesis is accepted. Hence it can be concluded that there is no significant differences between seasons.

Students *t*-distribution

t-distribution is commonly called Student's t-distribution or simply student's distribution. The t-distribution is used when sample size is 30 or less and the population standard deviation is unknown,

The t-statistic is defined as

$$t = \frac{\bar{X} - \mu}{S} \; x \sqrt{n}$$

where $S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}}$

$\bar{X}$ = The mean of the sample

μ = The actual or hypothetical mean of the population

n = The sample size/ Number of Observations

S = The Standard deviation of the sample.

Degrees of freedom υ = n -1

TESTING DIFFERENCE BETWEEN MEANS OF TWO SAMPLES (INDEPENDENT SAMPLES)

Given two independent random samples of size n_1 and n_2 with means $\bar{X}_!$ and $\bar{X}_2$ and standard deviation S_1 and S_2 testing the hypothesis that the samples come from the same normal population. To carry out the test,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \; x \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

where $\bar{X}_!$ = Mean of the first sample

$\bar{X}_2$ = Mean of the second sample

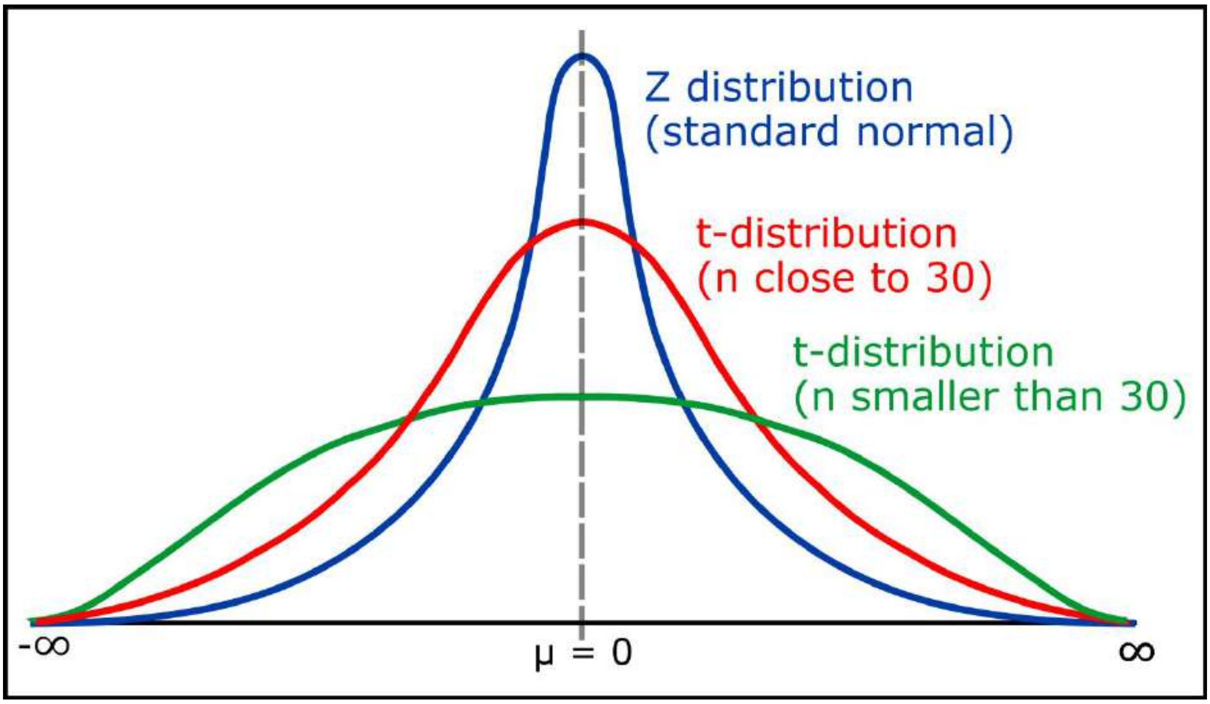
n_1 = Number of observations in the first sample

n_2 = Number of observations in the second sample

S = Combined Standard Deviation.

$$S = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2 + \Sigma(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

$$\upsilon = n_1 + n_2 - 2$$



ILLUSTRATION

1. The manufacturer of a certain make of electric bulbs claims that his bulbs have a mean life of 25 months with a standard deviation of 5 months. A random sample of 6 such bulbs gave the following values:

Life of bulbs: 24 26 30 20 20 18
(in months)

Can you regard the producer's claim to be valid at 1% level of significance? (Given that the table values of the appropriate test statistics at the said level are 4.035, 3.707 and 3.499 for 5, 6 and 7 degrees of freedom respectively.

SOLUTION

STEP 1: FRAMING OF HYPOTHESIS

Ho = The bulbs does not have the mean life of 25 months
H_1 = The bulbs have the mean life of 25 months.

STEP 2; APPLYING t-distribution

$$t = \frac{\bar{X} - \mu}{S} \; x \sqrt{n}$$

CALCULATION OF MEAN

$$\bar{X} = \frac{\Sigma X}{N} = \frac{138}{6} = 23$$

$$\boldsymbol{\bar{X} = 23}$$

X	$\boldsymbol{X - \bar{X}}$	$\boldsymbol{(X - \bar{X})^2}$
24	1	1
26	3	9
30	7	49
20	-3	9
20	-3	9
18	-5	25
ΣX = 138		$\boldsymbol{\Sigma(X - \bar{X})^2}$ **= 102**

CALCULATION OF STANDARD DEVIATION

$$S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}$$

$$S = \sqrt{\frac{102}{6-1}} = \sqrt{20.4}$$

S = 4.516

$\bar{X}$ = 23; μ = 25; S = 4.516; n = 6

Applying t-distribution

$$t = \frac{23-25}{4.516}\ x\ \sqrt{6} = \frac{-2}{4.516}\ x\ 2.44 = -0.44 \times 2.44$$

t = -1.08

Calculation of degrees of freedom

υ = n - 1 = 6 - 1 = 5

for 5 degrees of freedom with 1% level of significance is 4.032

STEP 5 COMPARING THE CALCULATED VALUE AND TABLE VALUE

Calculated value	<	Table value
-1.08	<	4.032

Since the calculated value is less than the table value null hypothesis is accepted. Hence it is concluded that the manufacturer's claim of that his bulbs has an average life of 25 months becomes invalid.

t-distribution for TWO SAMPLES

1. Two types of drugs were used for reducing weight on 5 and 7 people with drug A and drug B respectively. Drug A was imported and Drug B was indigenous. Decrease in the weight after using the drugs.

Drug A:	10	12	13	11	14		
Drug B:	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of the two drugs?

t value for υ = 10 with 5% level of significance is 2.228.

SOLUTION

STEP 1: FRAMING OF HYPOTHESIS

Ho = There is no significant differences in the efficacy of two drugs

H_1 = There is significant differences in the efficacy of two drugs.

X_1	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	X_2	$X_2 - \bar{X}_2$	$(X_2 - X_2)^2$
10	-2	4	8	-3	9
12	0	0	9	-2	4
13	1	1	12	1	1
11	-1	1	14	3	9
14	2	4	15	4	16
			10	-1	1
			9	-2	4

$\Sigma X_1 = 60$; $\Sigma(X_1 - \bar{X}_1)^2 = 10$; $\Sigma X_2 = 77$; $\Sigma(X_2 - X_2)^2 = 44$

CALCULATION OF MEAN

$$\bar{X}_1 = \frac{\Sigma X_1}{N_1} = \frac{60}{5} = 12$$

$$\bar{X}_2 = \frac{\Sigma X_2}{N_2} = \frac{77}{7} = 11$$

CALCULATION OF STANDARD DEVIATION

$$S = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2 + \Sigma(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

$$S = \sqrt{\frac{10+44}{5+7-2}} = \sqrt{\frac{54}{10}} = \sqrt{5.4} = 2.323$$

APPLYING t-distribution

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \; x \; \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$t = \frac{12-11}{2.323} \; x \; \sqrt{\frac{5 \, x \, 7}{5+7}} \; = \frac{1}{2.323} \; x \; \sqrt{\frac{35}{12}} \; = 0.430 \; x \; \sqrt{2.916}$$

$= 0.430 \text{ x } 1.707$

t = 0.734

Calculation of degrees of freedom

$$\upsilon = n_1 + n_2 - 2$$

$= 5 + 7 - 2$

$\upsilon = 10$

STEP 3 COMPARING CALCULATED VALUE AND TABLE VALUE

For 10 degrees of freedom with 5% level of significance is 2.228.

Calculated value < Table Value

0.734 < 2.228

As the calculated value is less than the table value null hypothesis is accepted. Hence it is concluded that there is no significant differences between imported drugs and indigenous drugs

Z Test

Z test is a concept of statistics which compares means of two populations. Z test assumes normal distribution under null hypothesis. Z test uses an assumed value which is generally within the limits of given data to calculate Z score. This value is known as Standardised random variable. It is used to determine whether two samples means are different when variances are known and sample is large ($n \geq 30$)

$$\text{Z score} = \left| \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right|$$

$\bar{X}$ = Sample mean

μ = Population mean

σ = Standard deviation of population

n = Number of observations

ILLUSTRATION

1. A principal at a certain school claims that the students in his school are above average intelligence. A random sample of 30 students IQ scores have a mean score of 112. Is there

sufficient evidence to support the principal's claim? The mean population IQ is 100 with a standard deviation of 15.   Z score for 5% level of significance = 1.645

SOLUTION

STEP 1: Framing of hypothesis

Ho = The students are not above average intelligence
H_1 = The students are above average intelligence

STEP 2 CALCULATION OF Z SCORE

$$\text{Z score} = \left|\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}\right|$$

$\bar{X}$ = 112; μ = 100; σ = 15; n = 30

$$= \left|\frac{112-100}{\frac{15}{\sqrt{30}}}\right| = \left|\frac{12}{\frac{15}{5.48}}\right| = \left|\frac{12}{2.74}\right|$$

$$Z = 4.38$$

STEP 3 COMPARING CALCULATED VALUE AND CRITICAL VALUE

Calculated value > Table value
4.38 > 1.645

As the calculated value is greater than z critical score for 5% level of significance the null hypothesis is rejected. Thus it is concluded that the Principal's claim of students are above average becomes true.

THEORY QUESTIONS
SHORT ANSWER QUESTIONS

1. Define hypothesis
2. What is meant by statistical inference/Type I error /Type II error / Null hypothesis / Alternative hypothesis /calculated value / critical value / One way classification of ANOVA / Two way classification of ANOVA / Degrees of freedom / Level of significance / Arithmetic Mean / Standard Deviation / Correction factor / Variance Ratio test / One tailed test / Two tailed test ?
3. List the purpose solved by Chi-square test/ F test / *t* distribution / Z test / ANOVA .

ESSAY TYPE QUESTIONS

1. Enumerate the procedure for testing hypothesis
2. List out the types of error that occurs while testing hypothesis

UNIT III
LINEAR PROGRAMMING

Linear Programming problem deals with determining optimal allocations of limited resources to meet given objectives. The resources may be in the form of men, raw materials, machines, money etc. The objectives is usually maximising profit, minimising total cost, maximising utility etc.

Linear programming problem deals with optimisation, maximisation and minimisation of a function of decision variables [The variables whose values determine the solution of a problem are called decision variables of the problems] known as objective function subject to a set of simultaneous linear equation [or inequaliser] known as constraints. The term linear means that all the variable occurring in the objective functions and constraints are of first degree of the problem under consideration and the term programming means the process of determining a particular course of action.

REQUIREMENT FOR EMPLOYING LPP TECHNIQUES.

- There must be a well defined objective function.
- There must be alternative courses of action to choose.
- Atleast some of the resources must be limited [ex., supply, which give rise to constraints]
- Both objective and constraints must be linear equation or inequalities.
-

MATHEMATICAL FORMULATION OF LPP
PROCEDURE FOR FORMING A L.P.P. MODEL.

STEP 1

Identify the unknown decision variables to be determined and assign symbols to them.

STEP 2

Identify all the restrictions or constraints [or influencing factors] in the problem and express them as linear equations or inequations of decision variables.

STEP 3

Identify the objective or aim and represent it also as a linear function of decision variables.

STEP 4

Express the complete formulation of LPP as a general mathematical model.

1. A company is making a chart to decide the minimum amount of constituents like proteins, vitamins, carbohydrates, fats etc. which a man needs on daily basis to fulfil his requirement for medical awareness. The choice is to be made from different type of foods (4 types). The yields per unit for different types of foods are explained below in the chart. Formulation of the linear programming model is required for this p problem.

Type of food	**Yield per unit**			**Cost per unit**
	Proteins	**Fats**	**Vitamins**	
A	6	6	12	135
B	8	6	8	120
C	16	21	14	170
D	12	15	8	130

Minimum required	1600	400	1400	

Solution.

Formulation of Linear Programming Model

Let X_1, X_2, X_3, X_4 be the no. of units of food of type 1, 2, 3 & 4 respectively.

Objective is to minimize the cost i.e.

Minimize Z = Rs. $(135X_1 + 120X_2 + 170X_3 + 130X_4)$

Constraints are on the fulfilment of daily requirements of the various constituents
i.e. Proteins requirement $6X_1 + 8X_2 + 16X_3 + 12X_4 \geq 1600$

Fats $6X_1 + 6X_2 + 21X_3 + 15X_4 \geq 400$

Vitamins $12X_1 + 8X_2 + 14X_3 + 8X_4 \geq 1400$

Where each $X_1, X_2, X_3, X_4 \geq 0$

2. A farmer has 100 acre of farm for harvesting. potatoes, carrots & beans produced by him are sold by him at the rates Rs. 2.00, Rs. 1.50 and Rs. 4.00 resp. The average productivity per acre of potatoes is 4 tonnes, carrots are 6 tonnes and 2 tonnes of beans is produced per acre. Fertilizer is available at rate of Rs. 1.0 per kg. and amount of fertilizer required for potatoes and carrots is 200 kg per acre and 100 kg for beans. Labour required for sowing, cultivating and harvesting per acre is 5 man–days for potatoes and carrots, and 6 man-days for beans. A total of 400 man-days of labour are available at Rs. 60 per man day.

Formulate using Linear Programming to maximise farmer's total profit.

Solution.

How much area should be given to each type of crop for cultivation so that profit from the production can be maximized has to be decided.

Let X_1, X_2, & X_3 be the acres of land for different crops (potatoes, carrots, and beans resp.).

Therefore the crops production will be
$4000X_1$ of potatoes, $6000X_2$ of carrots & $2000X_3$ of beans

Total sales = $[2 \times 4000X_1 + 1.5 \times 6000X_2 + 4 \times 2000X_3] = 8000X_1 + 9000X_2 + 8000X_3$
Total expenses on fertilizers = $[1 \text{ X } \{200\text{X}(X_1 + X_2) + 100X_3\}] = 200X_1 + 200X_2 + 100X_3$

Labour expenses = $[60 \text{ X } \{5\text{X}(X_1 + X_2) + 6X_3\}] = 300X_1 + 300X_2 + 360X_3$

Net profit (P) = Total sales – Total expenses

$= 8000X_1 + 9000X_2 + 8000X_3 - 500X_1 - 500X_2 - 460X_3 = 7500X_1 + 8500X_2 + 7540X_3$

Since total area is restricted to 100 acres of land, $X_1 + X_2 + X_3 \leq 100$

Find X_1 & X_2 such that the profit $P = 3X_1 + 5X_2$ is maximum, subject to the conditions

$X_1 + 2X_2 \leq 2000$
$X_1 + X_2 \leq 1500$
$X_2 \leq 600$
$X_1 \geq 0$ & $X_2 \geq 0$

Also, the total man-days labour is restricted to 400 man–days, therefore

$$5X_1 + 5X_2 + 6X_3 \leq 400$$

Hence the farmer's allocation problem can be finally put in the form

SIMPLEX METHOD

Method of locating the optimal vertex is the simplex method or simplex technique or algorithm which was developed by G. Dantzig. Since the number of vertex is finite the method leads to an optimum vertex in a finite number of steps or indicates the existence of an unbounded solution.

Given m linear equations with n variables (m≥n). The solution obtained by setting (n-m=0) variables is called basic solution (m→basic variables n-m→non basic variables.).

Basic solution is non-degenerate if none of the basic variable is zero

A feasible solution which is also basic is called a basic feasible solution (x>0 feasible).

Ex., Max $\geq = 4x_1 + 2x_2 + 6x_3$
$2x_1 + 3x_2 + 2x_2 \geq 6$
$3x_1 + 4x_2 = 8$
$6x_1 - 4x_2 + x_3 \leq 10$
$c = 4, 2, 6 \quad x = x_1, x_2, x_3$
$b = 6, 8, 10$

SIMPLEX ALGORITHM - STEPS

1. Check whether the objective function is to maximise or minimise. If minimise convert into a maximisation

Minimise z = 1 maximize (-z)

2. Check whether all bi are positive. If any of the bi is negative multiply both sides of that constraint by -1 so as to make bi positive.

3. **Conversion of inequalities to equalities**

a. Introduce slack variable for constraints of type ≤ representing the unused quantity of resource.

e.g., $5x_1 + 3x_2 = 27$
$5x_1 + 3x_2 + S_1 = 27$
$S_1 \geq 0$

b. Adding surplus variables for constraints of type ≥ representing the excess resource.
e.g., $3x_1 + x_2 \geq 5$
$3x_1 + x_2 - S_1 = 3$
$S_2 \geq 0$

4. **Simplex Table**

	Cj	C1	C2	C3	0	0	0	
CB X/B	**XB**	**X_1**	**X_2**	**X_3**	**S_1**	**S_2**	**S_3**	**Minimum rates**
0 S_1	*	*	*	*	1	0	0	
0 S_2	*	*	*	*	0	1	0	
0 S_3	*	*	*	*	0	0	1	
(Zj - Cj)	Zo	Z_1-C_1	Z_2-C_2					

Cj row → co-efficient of variables in objective function.
CB column → Co-efficient of basic variables in objective function.
YB column → basic variables
XB column → values of Basic variables
Zj-Cj → net evaluation.

5. Compute net evaluation

Zj-Cj = Cbxaj-Cj examine the sign of Zj-Cj.
a. If all Zj-Cj ≥ 0 → current basic feasible solution xb is optimal.
b. If atleast one (Zj-Cj) < 0 then the current basic feasible solution is not optimal.

6. Find the entering (also called incoming variables) variables (optimal column or key column or pivot column) entering the variables → most negative of all (Zj - Cj) (in case of b time anyone can be selected) column corresponding in the entering variable is the pivot column.

7. Find the leaving variable.

Compute ratio ϴ = min $\left\{\frac{xBi}{air},\ air > 0\right\}$

in order to decide the leaving variable select the least positive ratio quantity column values their corresponding optimal column values, (i.e.) the ratio between the solution column and the entering variables column by considering only the positive denominators.

a). If all (air ≤ 0) indicates unbounded solution.
b). If atleast one (air ≤ 0) then leaving variables is the basic variables corresponding to the key row or Pivot row or Pivot equation and the element at the intersection of the Pivot column and pivot row or pivot equation and the element at the intersection of the pivot column and pivot row is called the pivot element or key element or leading element.

8. Drop the leaving variables and introduce the entering variable with its associated vales under cb column.

i). New Pivot equation $= \dfrac{Old\ pivot\ equation}{pivot\ element}$

calculate the values for the remaining row.

ii). New equation = Old equation - $\begin{bmatrix} corresponding \\ pivot\ column \\ co-efficients \end{bmatrix}$ X $\begin{bmatrix} pivot \\ equation \end{bmatrix}$

Go to step V and repeat the procedure until either on optimum solution or unbound solution is reached.

NOTE 1

For Maximisation problem.

a) If all Zj-Cj $\geq$ 0 then the current basic feasible solution is optimal.
b) If atleast one (Zj-Cj) < 0, then the current basic feasible solution is not optimal.
c) The entering variable is the non-basic variable corresponding to the most negative value of (Zj-Cj).

NOTE 2

a) If all (Zj-Cj) $\leq$ 0, then the current basic feasible solution is optimal.
b) If all (Zj-Cj) > 0, then the current basic feasible solution is not optimal.

The entering variable is non-basic variable corresponding to the most positive value of (Zj-Cj)

NOTE 3

For both maximisation and minimisation problems that leaving variables is the basic variable corresponding minimum ratios Θ.

ILLUSTRATION 1

1. Using Simplex method, solve the LPP

Maximize $z = 4x_1 + 10x_2$ subject to

$$2x_1 + x_2 \leq 50$$
$$2x_1 + 5x_2 \leq 100$$
$$2x_1 + 3x_2 \leq 90$$
$$x_1, x_2 \geq 0.$$

SOLUTION

STEP 1

Objective function maximise

STEP 2

Check whether all bi are positive

STEP 3

Standard form → All constraints are $\leq$ type so slack variable S_1, S_2, S_3

Max ≥ $4x_1 + 10x_2 + 0S_1 + 0S_2 + 0S_3$
Subject to

$2x_1 + x_2 + S_1 + 0S_2 + 0S_3 = 50$
$2x_1 + 5x_2 + 0S_1 + 0S_2 + 0S_3 = 100$
$2x_1 + 3x_2 + 0S_1 + 0S_2 + S_3 = 90$

STEP 4 SIMPLEX TABLE

	C_J	4	10	0	0	0	
CB X/B	**XB**	**X_1**	**X_2**	**S_1**	**S_2**	**S_3**	**Minimum rates**
0 S_1	50	2	1	1	0	0	50
0 S_2	100	2	5	0	1	0	20 *
0 S_3	90	2	3	0	0	1	30
(Zj - Cj)	0	-4	-10	0	0	0	

X_2 → entering variable
0 S_2 → leaving variable

→ pivot row

→ pivot column

5 → pivot element

Zj-Cj

Cj = 0 x 50 + 0 x 100 + 0 x 90 = 0
Zj = CB x ai
= 0 x 2 + 0 x 2 + 0 x 2 = 0 - 4 = - 4

STEP 5 NET EVALUATION

All Zj - Cj are not positive i.e., ≥ 0 so the current basic feasible solution is not optimal

STEP 6 ENTERING VARIABLE / PIVOT COLUMN

Most negative of Zj-Cj = - 10. ∴ entering variable is x_2 and it is the pivot column

STEP 7 FINDING THE LEAVING VARIABLE / PIVOT ROW / PIVOT ELEMENT

$$\Theta = \min \left[\frac{50}{1}, \frac{100}{5}, \frac{90}{3} \right]$$

= min [50, 20, 30]

$\therefore$ 20

Corresponding row $\rightarrow$ pivot row

Leaving variable $\rightarrow$ S2

Pivot element $\rightarrow$ intersection point of pivot row and pivot column = 5

STEP 8 NEW SIMPLEX TABLE

Drop the new variable and introduce entering variable with its associated values under cb column.

	C_J	4	10	0	0	0
CB X/B	**XB**	**X_1**	**X_2**	**S_1**	**S_2**	**S_3**
0 S_1	30	8/5	0	1	-1/5	0
0 S_2	20	2/5	1	0	1/5	0
0 S_3	30	4/5	0	0	-3/5	1
(Zj - Cj)	200	0	0	0	2	0

New Pivot equation = $\frac{Old\ pivot\ equation}{pivot\ element}$

$$= \left[\frac{100}{5}, \frac{2}{5}, \frac{5}{5}, \frac{0}{5}, \frac{1}{5}, \frac{0}{5}\right]$$

= 20, 2/5, 1, 0, 1/5, 0

New S1 equation = Old equation - $\begin{bmatrix} corresponding \\ pivot\ column \\ co-efficients \end{bmatrix}$ X $\begin{bmatrix} pivot \\ equation \end{bmatrix}$

S_1 Old	=	50	2	1	1	0	0
New	=	20	2/5	1	0	1/5	0
		30	8/5	0	1	-1/5	0
S_2 Old	=	90	2	3	0	0	1
New	=	60	6/5	3	0	3/5	0
		30	4/5	3	0	-3/5	1

Net equation Zj-Cj ≥ 0 so the current basic feasible solution is optimal.

Max z = 200

$x_1 = 0$

$x_2 = 20$

2. Find the Non-Negative values x_1 , x_2 and x_3 which maximise $Z = 3x_1 + 2x_2 + 5x_3$

Subject to $x_1 + 4x_2 \leq 420$

$3x_1 + \ 2x_3 \leq 460$

$x_1 \ + 2x_2 + x_3 \ \leq 430$

SOLUTION

STEP 1

Objective function → maximize

STEP 2

Check whether all bi are positive

STEP 3

Standard form → All constraints are $\leq$ type so slack variable S_1, S_2, S_3

Max Z = $3x_1 + 2x_2 + 5x3 + 0S1 + 0S2 + 0S_3$

Subject to $x_1 + 4x_2 + 0S_3 + S_1 + 0S_2 + 0S_3 \ = \ 420$

$3x_1 + \ 0x_2 + 2x_3 \ + 0S_1 + S_2 + 0S_3 \ = 460$

$x_1 + 2x_2 + x_3 + 0S_1 + 0S_2 + S_3 = 430$

STEP 4 SIMPLEX METHOD

	C_J	3	2	0	0	0		
CB X/B	**XB**	**X_1**	**X_2**	**X_3**	**S_1**	**S_2**	**S_3**	**Minimum rates**
0 S_1	420	1	4	0	1	0	0	0
0 S_2	460	3	0	2	0	1	0	230
0 S_3	430	1	2	1	0	0	1	430
(Zj - Cj)		-3	-2	-5	0	0	0	

X_3 → entering variable

0 S_2 → leaving variable

→ pivot row

→ pivot column

2 → pivot element

Zj - Cj

Cj = 0 x 420 + 0 x 460 + 0 x 430 = 0

Zj = CB x aj = 0x1 + 0x3 + 0 x 1 = 0-3 = -3

STEP 5 NET EVALUATION

All Zj - Cj are not positive i.e., ≥ 0 so the current basic feasible solution is not optimal.

STEP 6 ENTERING VARIABLE / PIVOT COLUMN

Most negative of Zj-Cj = -5. ∴ entering variables is X_3 and it is the pivot column

STEP 7 FIND THE LEAVING VARIABLE / PIVOT ROW / PIVOT ELEMENT.

$$\Theta = \min \left[\frac{420}{0}, \frac{230}{2}, \frac{430}{1}\right]$$

$$= \min [\ 0,\ 230,\ 430\ ]$$

∴ 230

Corresponding row → pivot row leaving variable → S_2 pivot element → intersection point of pivot row and pivot column = 230

STEP 8 NEW SIMPLEX TABLE.

Drop the row variable and introduce entering variable with its associated values under cb column.

	C_J	3	2	0	0	0		
CB X/B	**XB**	**X_1**	**X_2**	**X3**	**S_1**	**S_2**	**S_3**	**Minimum rates**
0 S_1	420	1	4	0	1	1	0	105
0 S_2	230	3/2	0	1	0	1/2	0	0
0 S_3	200	-1/2	2	0	0	-1/2	1	100
(Zj - Cj)	1150	9/2	-2	0	0	5/2	0	

$$\Theta = \min \left[\frac{420}{4}, \frac{230}{0}, \frac{200}{2}\right]$$ = 105, 0, 100 = -2 most negative

New Pivot equation $= \dfrac{\textit{Old pivot equation}}{\textit{pivot element}}$

$$= \left[\frac{460}{2}, \frac{3}{2}, \frac{0}{2}, \frac{2}{2}, \frac{0}{2}, \frac{1}{2}, \frac{0}{2}\right]$$

= 230, 3/2, 0, 1, 0, 1/2,0

New S_1 equation = Old equation - $\begin{bmatrix} corresponding \\ pivot\ column \\ co-efficients \end{bmatrix}$ X $\begin{bmatrix} pivot \\ equation \end{bmatrix}$

S_1 Old	=	420	1	4	0	1	0	0
New	=	0	0	0	0	0	0	0

		420	1	4	0	1	0	0
S_2 Old	=	430	1	2	1	0	0	1
New	=	230	3/2	0	1	0	1/2	0
		200	-1/2	2	0	0	-1/2	1

STEP 9 SECOND SIMPLEX TABLE

	C_J	3	2	5	0	0	0
CB X/B	**XB**	**X_1**	**X_2**	**X3**	**S_1**	**S_2**	**S_3**
0 S_1	20	2	0	0	1	1	-2
0 S_2	230	3/2	0	1	0	1/2	0
0 S_3	100	-1/4	1	0	0	-1/4	1/2
(Zj - Cj)	1350	4	0	0	0	2	1

New Pivot equation $= \dfrac{Old\ pivot\ equation}{pivot\ element}$

$$= \left[\frac{200}{2}, \frac{-1}{2}, \frac{2}{2}, \frac{0}{2}, \frac{0}{2}, \frac{-1}{2}, \frac{1}{2}\right]$$

= 100, -1/4, 1, 0, 0, -1/4, -1/2

Old	420	1	4	0	1	1	0
New	400	1/4	4	0	0	1/4	2
	20	2	0	0	1	1	-2

Net equation All Zj - Cj ≥ 0 so the current basic feasible solution is optimal.

Maz Z = 1350

$x_1 = 0$

$x_2 = 100$

$x_3 = 230$

QUANTITATIVE TECHNIQUES - LPP - GRAPHICAL METHOD Unit III

BY
DR.J.J.JEYAKUMARI., M.Com., M.phil., Ph.D., MBA., PGDCA.,
RESEARCH ADVISOR & ASSISTANT PROFESSOR OF COMMERCE,
K.N.GOVT. ARTS COLLEGE FOR WOMEN, AUTONOMOUS, THANJAVUR -7

Reference : Operations Research, Sundaresan V, Ganapathy Subramanian K. S, Ganesan K.

GRAPHICAL METHOD

Linear Programming problems (LPP) involving only two variables can be effectively solved by graphical method which provides a pictorial representation of the problem and its solutions.

- This method is simple and easy to understand.
- Redundant constraints are automatically eliminated from the system.

Problem 1:

Solve the following LPP by the graphical method

Max $Z = 3x_1+2x_2$

Subject to $-2x_1 + x_2 \le 1$

$x_1 \le 2$

$x_1 + x_2 \le 3$

and $x_1, x_2 \ge 0$

Solution - Problem 1

Step 1 : Draw x_1 and x_2 axis.

Step 2: First convert the inequality constraints as equalities.

$$-2x_1 + x_2 = 1$$
$$x_1 = 2$$
$$x_1 + x_2 = 3$$
$$x_1 = 0,$$
$$x_2 = 0$$

Step 3: Find the axis intercept

Constraint 1: $-2x_1 + x_2 = 1$

For x_1 axis intercept, $x_2 = 0$

$$-2x_1 + 0 = 1$$
$$-2x_1 = 1$$
$$x_1 = 1/-2 = -0.5$$

For x_2 axis intercept, $x_1 = 0$

$$-2 \times 0 + x_2 = 1$$
$$x_2 = 1$$

The vertices are (-0.5, 0) and (0, 1)

Constraint 2 : $x_1 = 2$

The vertices are (2, 0) i.e $x_1 = 2$, $x_2 = 0$

Constraint 3: $x_1 + x_2 = 3$

For x_1 axis intercept, $x_2 = 0$

$$x_1 + 0 = 3$$
$$x_1 = 3$$

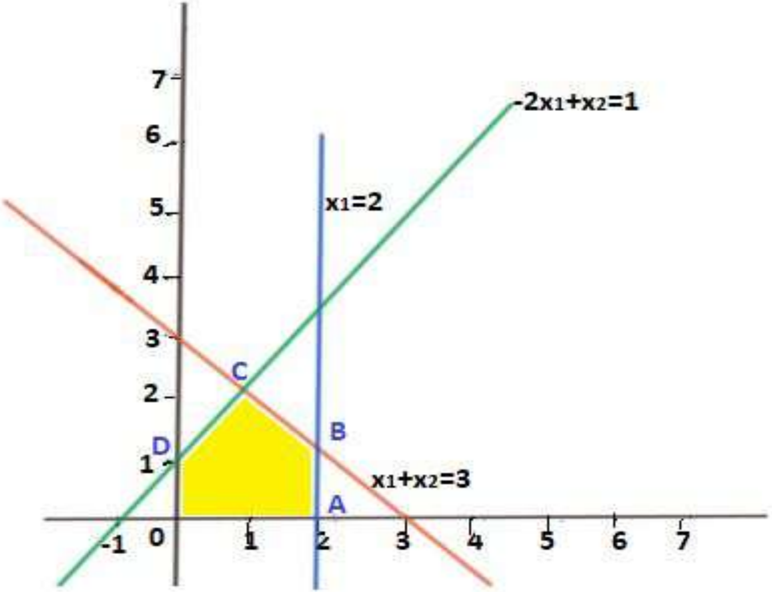
For x_2 axis intercept, $x_1 = 0$

$$0 + x_2 = 3$$
$$x_2 = 3$$

The vertices are (3, 0) and (0, 3)

Constraint 4 : $x_1 = 0$ (0,0)

Constraint 5 : $x_2 = 0$ (0,0)



Step 4:

Now the feasible region is OABCD.

Every point within this area satisfies all the constraints.

We have to find the optimal solution.

Step 5: Find the vertices of the solution space.

The vertices of solution space are:

1) O - (0,0)
2) A – (2,0)
3) B - ?

B is the point of intersection of equations,

$x_1 = 2$ ---- eqn 1 and

$x_1 + x_2 = 3$ ---- eqn 2

Solving these two equations

$x_1 = 2$ and $x_2 = 1$

The vertex of B is (2,1)

4) C - ?

C is the point of intersection of

$-2x_1 + x_2 = 1$ ---- eqn 1 and

$x_1 + x_2 = 3$ ---- eqn 2 ---Solving 1 & 2

$-3x_1 = -2$

$x_1 = 2/3$ -----Substituting in eqn 2

$2/3 + x_2 = 3$

$x_2 = 3 - 2/3 = (9-2)/3 = 7/3$

The vertex of C is (2/3, 7/3)

5) D – (0,1)

Step 6: Find the value of Z at these vertices.

VERTEX	VALUE OF Z ($Z = 3x_1 + 2x_2$)
O (0, 0)	0
A (2, 0)	6
B (2,1)	**8 (OPTIMUM)**
C (2/3, 7/3)	20/3
D (0, 1)	2

OUTPUT

The optimum solution LPP is

$Z = 8$, $x_1 = 2$ and $x_2 = 1$.

Problem 2

Solve the following LPP by the graphical method

Max $Z = 2x_1 + 3x_2$

Subject to

$x_1 + x_2 = 400$

$2x_1 + x_2 = 600$

and $x_1, x_2 = 0$

Problem 3

Solve the following LPP by the graphical method

Min $Z = 3x_1+5x_2$

Subject to $-3x_1+4x_2 = 12$

$x_1 = 4$

$2x_1 - x_2 = -2$

$x_2 = 2$

$2x_1+3x_2 = 12$

and $x_1, x_2 = 0$

Solution - Problem 3

Step 1 : Draw x_1 and x_2 axis.

Step 2: First convert the inequality constraints as equalities.

$-3x_1+4x_2 = 12$

$x_1 = 4$

$2x_1 - x_2 = -2$

$x_2 = 2$

$2x_1+3x_2 = 12$

$x_1 = 0, x_2 = 0$

Step 3: Find the axis intercept

Constraint 1: $-3x_1+4x_2 = 12$

For x_1 axis intercept, $x_2 = 0$

$-3x_1+0 = 12$

$-2x_1 = 12$

$x_1 = 12/-3 = -4$

For x_2 axis intercept, $x_1 = 0$

$0 - x_2 = -2$

$-x_2 = -2$

$x_2 = 2$

The vertices are (-1,0) and (0, 2)

Constraint 4 : $x_1 = 2$

The vertices are (0, 2) i.e $x_1 = 0, x_2 = 2$

Constraint 5: $2x_1+3x_2 = 12$

For x_1 axis intercept, $x_2 = 0$

$2x_1+0 = 12$

$2x_1 = 12$

$x_1 = 12/2 = 6$

For x_2 axis intercept, $x_1 = 0$

$0 + 3x_2 = 12$

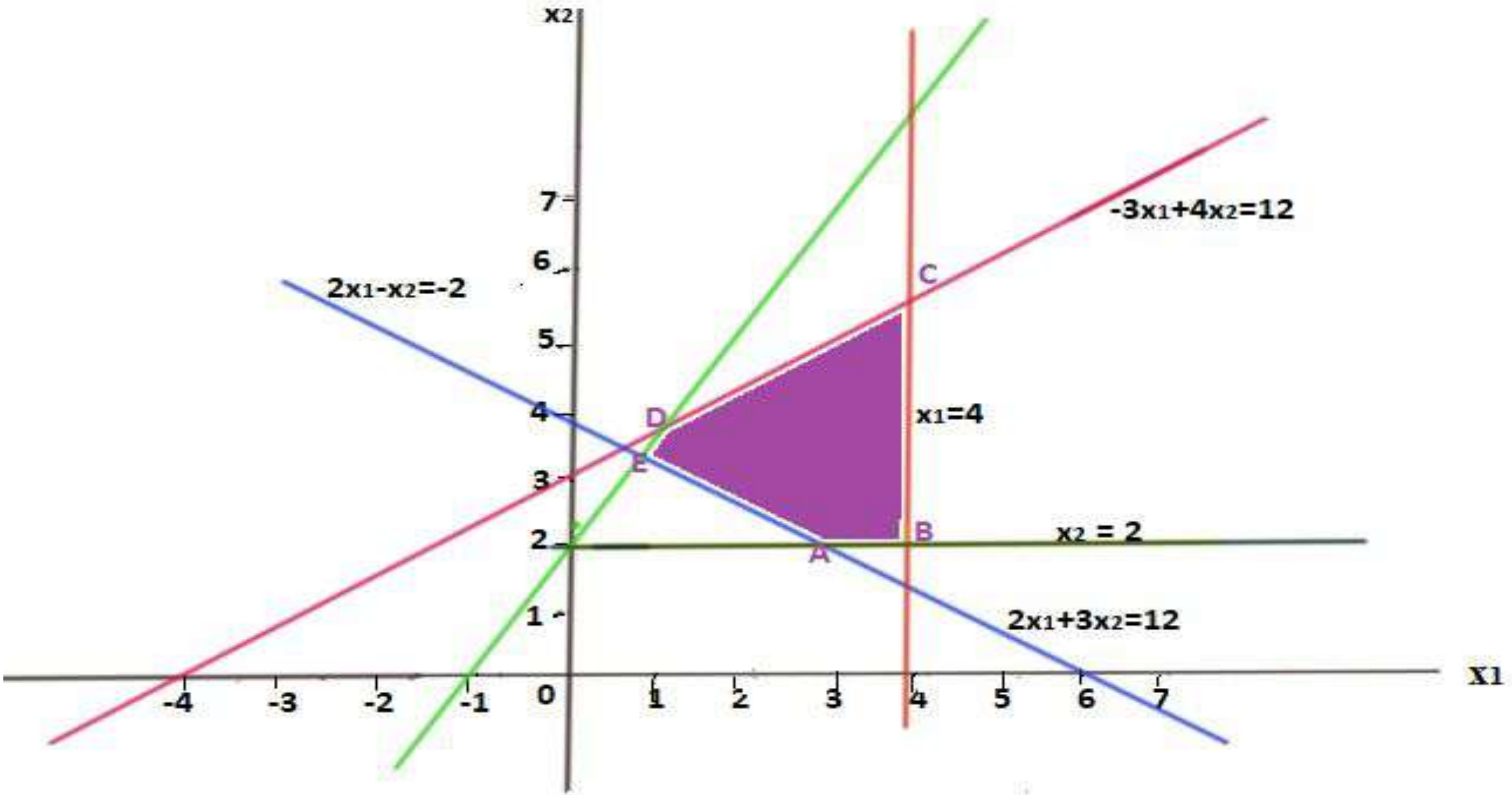
$3x_2 = 12$

$x_2 = 12/3 = 4$

The vertices are (6, 0) and (0, 4)

Constraint 6 : $x_1 = 0$ (0,0)

Constraint 7 : $x_2 = 0$ (0,0)



x2
7
6
5
4
3
2
1
0
-4
-3
-2
-1
1
2
3
4
5
6
7
X1
-3x1+4x2=12
2x1-x2=-2
C
D
E
x1=4
B
x2 = 2
A
2x1+3x2=12

Step 4:

Now the feasible region is ABCDE.

Every point within this area satisfies all the constraints.

We have to find the optimal solution.

Step 5: Find the vertices of the solution space.

The vertices of solution space are:

1) A - ?

A is the point of intersection of equations,

$X_2 = 2$ ---- eqn 1 and

$2x_1 + 3x_2 = 12$ ---- eqn 2

Solving these two equations

$2x_1 + 6 = 12$

$2x_1 = 12 - 6$

$2x_1 = 6$

$x_1 = 6/2 = 3$

The vertex of A is (3, 2)

2) B - ?

B is the point of intersection of equations,

$x_1 = 4$ ---- eqn 1 and

$x_2 = 2$ ---- eqn 2

The vertex of B is (4, 2)

3) C - ?

C is the point of intersection of

$x_1 = 4$ ---- eqn 1 and

$-3x_1 + 4x_2 = 12$ ---- eqn 2

Solving these two equations

$-12 + 4x_2 = 12$

$4x_2 = 12 + 12$

$4x_2 = 24$

$x_2 = 24/4 = 6$

The vertex of C is (4, 6)

4) D - ?

D is the point of intersection of equations,

$-3x_1 + 4x_2 = 12$ ---- eqn 1 and

$2x_1 - x_2 = -2$ ---- eqn 2 ---Solving 1 & 2

Eqn 2 multiplied by 4

$-3x_1 + 4x_2 = 12$

$8x_1 - 4x_2 = -8$

$5x_1 = 4$

$x_1 = 4/5$ -----Substituting in eqn 2

$(2 \times 4/5) - x_2 = -2$

$8/5 - x_2 = -2$

$-x_2 = -2 - 8/5$

$-x_2 = (-10 - 8)/5$

$-x_2 = -18/5$

$x_2 = 18/5$

The vertex of D is (4/5, 18/5)

5) E - ?

$2x_1 + 3x_2 = 12$ ---- eqn 1

$2x_1 - x_2 = -2$ ---- eqn 2 ---Solving 1 & 2

Eqn 2 multiplied by 3

$2x_1 + 3x_2 = 12$

$6x_1 - 3x_2 = -6$

$8x_1 = 6$

$x_1 = 6/8 = 3/4$ Substituting in eqn 2

$(2 \times 3/4) - x_2 = -2$

$6/4 - x_2 = -2$

$-x_2 = -2 - 6/4$

$-x_2 = (-8 - 6)/4$

$-x_2 = -14/4$

$x_2 = 14/4 = 7/2$

The vertex of E is (3/4, 7/2)

Step 6: Find the value of Z at these vertices.

VERTEX	**VALUE OF Z ($Z = 3x_1 + 5x_2$)**
A (3, 2)	**19 Min (optimum)**
B (4, 2)	22
C (4, 6)	42
D (4/5, 18/5)	102/5 = 20.4
E (3/4, 7/2)	79/4 = 19.75

OUTPUT

The optimum solution LPP is

$Z = 8$, $x_1 = 2$ and $x_2 = 1$.

Working procedure for graphical method

Step 1 : Draw x_1 and x_2 axis.

Step 2: First convert the inequality constraints as equalities.

Step 3: Find the axis intercept of all the constraints.

Step 4: Find the feasible region. Every point which lies in the feasible region satisfies all the constraints.

Step 5: Find the value of Z at all vertices of the feasible region.

Step 6: i) For maximisation problem, choose the vertex for which z is maximum. ii) For minimisation problem choose the vertex for which z is minimum.

THEORY QUESTIONS

1. What is meant by LPP?
2. Point out the methods for solving a linear programming problem.
3. When can graphical method used to solve linear programming problem?
4. What is an optimal solution?
5. What for is simplex technique used?
6. What is a pivot element/
7. How will you identify a leaving variable?
8. What is the difference between slack and surplus variable.
9. What are slack and surplus variables?
10. Define objective function and optimum solution.
11. State the steps involved in mathematical formulation of a Linear programming Problem.
12. Explain the procedure to solve a LPP using simplex method?
13. Describe the procedure to solve a LPP using graphical method.

QUANTITATIVE TECHNIQUES - Unit IV

BY
DR.J.J.JEYAKUMARI., M.Com., M.phil., Ph.D., MBA., PGDCA.,
RESEARCH ADVISOR & ASSISTANT PROFESSOR OF COMMERCE,
K.N.GOVT. ARTS COLLEGE FOR WOMEN, AUTONOMOUS, THANJAVUR -7

Reference : Operations Research, Sundaresan V, Ganapathy Subramanian K. S, Ganesan K.

LINEAR PROGRAMMING TECHNIQUES (LPP)

Linear programming deals with determining the Optimum allocation of Limited Resources to meet the given objectives.

The objective is usually

- maximizing profit,
- minimising total cost,
- maximizing utility etc....

Linear programming problem deal with **Optimisation** (**maximisation/minimisation**) of a function of decision variable known as objective function subject to a set of simultaneous linear equation known as **constraints.**

The variable whose values determine the solution of a problem are called **decision variables.**

Linear - all variables and constants are the first degree in the problem under consideration.

Programming - process of determining a particular course of action

Requirements of employing LPP technique:

1. A well defined objective function
2. Alternative course of action to choose.
3. Some resources must be Limited in supply which gives rise to constraints
4. Both objective function and constraints must be linear equation or inequalities

Transportation problem

- Objective is to determine the amount of to be shifted from each source to each destination so that the total transportation cost is minimum.
- Transportation deal with the transportation of a commodity (single product) from ‘m’ sources to ‘n’ destination.

Assumption

- Level of supply at each source and the amount of demand at each destination is known.
- The unit transportation cost from each source to each destination are known.
- Cost of the transportation is linear.

STANDARD TRANSPORTATION TABLE

m rows, n columns

C_{11} = unit transportation cost

destination ---------------- source	D1	D2	D3	
S1	C_{11}	C_{12}	C_{13}	a1
S2	C_{21}	C_{22}		a2
S3			C_{33}	a3
	b1	b2	b3	

a,b = Constraints/Rim requirements.

We have to find x which is the units to be transported.

Initial Basic Feasible solution = m+n-1

Methods

1. North-West Corner Rule
2. Least Cost Method
3. Vogels' Approximation method

PROBLEMS - 2 TYPES

1. Balanced Method - Row total & Column Total Equal
2. Unbalanced Method - Row total & Column total not equal.

Problem 1

Find the initial basic feasible solution for the following transportation problem using i) NWCR ii) LCM & iii) VAM

SINK

		A	B	C	D	E	
	P	2	11	10	3	7	4
ORIGIN	Q	1	4	7	2	1	8
	R	3	9	4	8	12	9
		3	3	4	5	6	

1. NORTH WEST CORNER RULE

1. In the north-west corner C11 cell, verify the constraints supply and demand, take which is low and put it as X and cancel the row or column.
2. If min(a,b)=a,put x=a and decrease b by a. Cross row 1. Move to next Nort-west corner cell C21.
   (or)
   If min(a,b)=b, put x=b and decrease a by b. Cross column 1. Move to next north-west corner cell C12.
   (or)
   If a=b, put x=a=b. Cross row1 and column 1. Move to next north-west corner cell C22.

NWCR

SINK

ORIGIN

	A	B	C	D	E	Supply
P	[3] 2	[1] 11	10	3	7	4 1
Q	1	[2] 4	[4] 7	[2] 2	1	8 6 2
R	3	9	4	[3] 8	[6] 12	9 6
Demand	3	3 2	4	5 3	6	21

$a_{ij} = b_{ij} = 21 \rightarrow$ Balanced.

Initial Basic Feasible solution

m+n-1

Where m is number of rows and n is number of columns

3+5-1 = 7

Transportation Cost:

3 x Rs2 + 1 x Rs 11 + 2 x Rs 4 + 4 x Rs 7 + 2 X Rs 2 + 3 x Rs 8 + 6 x Rs 12 = 6+11+8+28+4+24+72 = Rs 153

2.LEAST COST METHOD

1. Select the cell with the least amount of cost, verify the constraints supply and demand, take which is low and put it as x and cancel the row or column.
2. If min(a,b)=a,put x=a and decrease b by a. Cross row of least cost cell. Move to next least cost cell.
   (or)
   If min(a,b)=b, put x=b and decrease a by b. Cross column of the least cost cell. Move to next least cost cell.
   (or)
   If a=b, put x=a=b. Cross column and row of the least cost cell. Move to the next least cost cell.
3. If there is a tie i.e more than 1 cell containing the least cost, then the selection is left to your preference. But for uniformity let us select the first one first.

LCM

Sink.

Origin	A	B	C	D	E	
P	2	11	10	[4] 3	7	~~4~~
Q	[3] 1	4	7	2	[5] 1	~~8~~ ~~5~~
R	3	[3] 9	[4] 4	[1] 8	[1] 12	~~9~~ ~~5~~ ~~2~~ 1
	~~3~~	~~3~~	~~4~~	~~5~~ ~~1~~	~~6~~ 1	21

$a_{ij} = b_{ij} = 21 \rightarrow$ Balanced.

Initial Basic Feasible solution

m+n-1

Where m is number of rows and n is number of columns

3+5-1 = 7

Transportation Cost:

4 x Rs3 + 3 x Rs 1 + 5 x Rs 1 + 3 x Rs 9 + 4 x Rs 4 + 1 x Rs 8 + 1 x Rs 12 = 12+3+5+27+16+8+12 = Rs 83

3. VOGEL'S APPROXIMATION METHOD (VAM)/ UNIT COST PENALTY METHOD(UCPM)

1. Find the difference(**PENALTY**) between the smallest and next smallest cost element in each row and column and write them in brackets against the corresponding row.
2. **Identify the row/column with the largest penalty**. Choose the **cell with the smallest cost element** and allocate the maximum possible amount and cross the row/column as in the previous methods.
3. Again compute the row/column penalties for the remaining rows/columns and follow step 2.

VAM

Sink.

Origin	A	B	C	D	E	Supply	Penalties
P	2	11	10	[4] 3	7	~~4~~	(1) (1) (1) (1)
Q	1	[2] 4	7	2	[6] 1	~~8~~ ~~2~~	(0) (1)
R	[3] 3	[1] 9	[4] 4	[1] 8	12	~~9~~ ~~5~~ ~~2~~ 1	(1) (1) (1) (5) (5)* (1)*
Demand	3	~~3~~ 1	~~4~~	~~7~~ x	~~6~~	21	
Penalties	(1) (1) (1) (1)	(5 (5)* (2) (2)	(3) (3) (6)*	(1) (1) (5) (5)*	(6)*		

(1)* → Not necessary

$a_{ij} = b_{ij} = 21 \rightarrow$ balanced.

Initial Basic Feasible solution

m+n-1

Where m is number of rows and n is number of columns

3+5-1 = 7

Transportation Cost:

4 x Rs3 + 4 x Rs 2 + 6 x Rs 1 + 3 x Rs 3 + 1 x Rs 9 + 4 x Rs 4 + 1 x Rs 8 = 12+8+6+9+9+16+8 = Rs 68

2. Find the initial basic feasible solution for the following transportation problem using 1) NWCR 2) LCM and 3) VAM

	TO				Supply
	1	2	1	4	30
From	3	3	2	1	50
	4	2	5	9	20
Demand	20	40	30	10	

3. Find the initial basic feasible solution for the following transportation problem using

1) NWCR 2) LCM and 3) VAM

Distribution centre / Availability

		D1	D2	D3	D4	Availability
ORIGIN	S1	11	13	17	14	250
	S2	16	18	14	10	300
	S3	21	24	13	10	400
Requirements		200	225	275	250	

UNBALANCED TRANSPORTATION PROBLEM

When a_{ij} **not equal to** b_{ij} i.e **when row total is not equal to column total then the transportation problem is said to be unbalanced.**

How to solve an unbalanced transportation problem?

- First the transportation problem has to be balanced.
- Introduce a **dummy row if row total** (i.e. a_i) **is less.**
- Introduce a **dummy column if column total** (i.e. b_j) **is less.**

4. Find the optimal solution for the following transportation problem using 1) NWCR 2) LCM and 3) VAM

Destination

Source		A	B	C	D	supply
	1	11	20	7	8	50
	2	21	16	20	12	40
	3	8	12	18	9	70
demand		30	25	35	40	

Σa_i not equal to Σb_j - so unbalanced transportation problem.

Σa_i = 160 i.e. supply - **row total greater** than column total (demand) i.e. **Σb_j = 130**
Hence to convert it into a balanced transportation problem a **dummy column E with 30 units** (160-130) **is introduced.**

Degeneracy in transportation problem

In transportation problem, when the number of non-negative allocations is less than m+n-1, then the basic solution is said to be a degenerate one.

To resolve degeneracy, we allocate an extremely small amount (close to Zero) to one or more empty cells of the transportation table, (generally to the minimum cost cells if possible), so that the total number of occupied cells is m+n-1.

We done this small amount by **ϵ (epsilon). $\epsilon = 0$.**

Problem 4: Solution-1 –North West Corner Rule (NWCR)

Destination

Source	A	B	C	D	E	supply
1	[30] 11	[20] 20	7	8	0	~~50~~ ~~20~~
2	21	[5] 16	[35] 20	12	0	~~40~~ ~~35~~
3	8	[ε] 12	18	[40] 9	[30] 0	~~70~~ 30
demand	~~30~~	~~25~~ ~~5~~	~~35~~	~~40~~	30	160

$m+n-1 = 3+5-1 = 7$ but allotment is 6. So this is degenerate solution.

To resolve degeneracy a small quantity ϵ is allocated to (3, 2). $\epsilon \to 0$.

$30 \times Rs\ 11 + 20 \times Rs\ 20 + 5 \times Rs\ 16 + 35 \times Rs\ 20 + \underset{(0)}{\epsilon} \times Rs\ 12 + 40 \times Rs\ 9 + 30 \times Rs\ 0 =$ **Rs 1870**

Problem 4: Solution-2 – Least Cost Method (LCM)

Destination

Source		A	B	C	D	E	supply
	1	11	20	(20) 7	8	(30) 0	~~50~~ 20
	2	21	(25) 16	(15) 20	12	0	~~40~~ 15
	3	(30) 8	12	(ε) 18	(40) 9	0	~~70~~ 40
demand		30	25	~~35~~ 15	40	30	160

m + n − 1 = 3 + 5 − 1 = 7. Since there is 6 allotments the basic solution is degenerate. To resolve degeneracy a small ε is allocated to (3, 3). ε → 0

Optimal Transportation cost =
20 × Rs 7 + 30 × Rs 0 + 25 × Rs 16 + 15 × Rs 20 + 30 × Rs 8 + ε (0) × Rs 18 + 40 × Rs 9

= **Rs 1440**

Problem 4: Solution-3 –Vogel's Approximation Method (VAM)

Destination

Source	A	B	C	D	E	supply	Penalties
1	11	20	[35] 7	[15] 8	0	~~50~~ 15	(7) (4) (3) (3) (3) (6)
2	21	16	20	[10] 12	[30] 0	~~40~~ 10	(12)* (4) (4) (9)*
3	[30] 8	[25] 12	18	[15] 9	0	~~70~~ ~~45~~ 15	(8) (1) (1) (1) (1)
demand	~~30~~ 10	25	35	~~40~~ ~~30~~ 15	30	160	
Penalties	(3) (3) (3) (3) (3)*	(4) (4) (4)*	(1) (1)*	(1) (1) (1) (1) (1) (1) (1) not necessary.	(0) (0)		

m+n−1 =
3+5−1 = 7.

Since there are 7 allotments the solution is non-degenerate basic feasible.

Optimal transportation cost = 35×Rs 7 + 15×Rs 8 + 10×Rs 12 + 30×Rs 0 + 30×Rs 8 + 25×Rs 12 + 15×Rs 9 = **Rs 1160/-**

MAXIMISATION CASE IN TRANSPORTATION PROBLEM

To solve the maximisation problem first it has to be converted into a minimisation problem.

How to convert maximisation problem into a minimisation problem?

Subtract all the values in transportation table from the highest value (OR)

Multiply all the values in transportation problem by -1.

5. Solve the transportation problem to maximise profit using 1)NWCR 2)LCM and 3) VAM

		Destination				Supply
		A	B	C	D	
Source	1	40	25	22	33	100
	2	44	35	30	30	30
	3	38	38	28	30	70
Demand		40	20	60	30	

Since the given problem is maximisation type - To convert it into minimisation problem **all the values in transportation table is subtracted from 44.** The problem now becomes:

Destination

Source		A	B	C	D	Supply
	1	40	25	22	33	100
	2	44	35	30	30	30
	3	38	38	28	30	70
Demand		40	20	60	30	

Conversion to minimisation

	A	B	C	D	
1	4	19	22	11	100
2	0	9	14	14	30
3	6	6	16	14	70
	40	20	60	30	

Now carry on the usual procedure to solve the transportation problem.

ASSIGNMENT PROBLEM

The assignment problem is a particular case of the transportation problem in which the objective is to assign a number of tasks (jobs or origin or sources) **to an equal number of facilities** (machines or persons or destinations) at a minimum cost or maximum profit.

- Assign 'n' Jobs to 'm' machines.
- Assign jobs at minimum cost or maximum profit
- On the assumption that each machine can perform each job but with varying degree of efficiency.
- Assignment problem is stated in the form of m x n matrix (c_{ij}) called the cost matrix or effectiveness matrix.

Difference between the transportation and the assignment problem

S.NO	TRANSPORTATION PROBLEM	ASSIGNMENT PROBLEM
1	Supply at any source may be any positive quantity a_i	Supply at any source will be 1. i.e a_i=1
2	Demand at any destination may be any positive quantity b_j	Demand any destination will be 1. i.e b_j=1
3	One or more source to any number of destination.	One source to only one destination

STEPS IN SOLVING AN ASSIGNMENT PROBLEM

Step 1: Select the **smallest cost element in each row** and **subtract** this from all the elements of the corresponding row.
Step 2: Select the **smallest cost element in each column** and **subtract** this from all the elements of the corresponding column.
Step 3: Assigning Zero's.
Now each row and column would have at least one zero.
Examine each row successively until a row with exactly one '0' is found. Circle it and cross all other 0's in the said **column**. Repeat this for all rows.
Same way examine each column successfully until a column with exactly one '0' is found. Circle it and cross all other 0's in the said **row**. Repeat this for all columns.
Step 4: Applying Optimal Test

- **If each row and each column contain one encircled zero, the current assignment is optimal.** Compute the answer.
- If at least one row/column is without an assignment i.e without an encircled zero then the current assignment is not optimal. Go to the next step.

Step 5: Cover all the zeros by drawing a minimum number of straight lines

A. **Mark the rows that do not have an assignment.**
B. **Mark the columns** (not already marked) **that have zeros in the marked row.**
C. **Mark the rows** (not already marked) **that have assignments in the marked columns.**
D. Repeat (B) and (C) until no marking is required.
E. **Draw lines on all unmarked rows and marked columns.**

Step 6:

1. Find the smallest cost element not covered by the straight lines.
2. Subtract this smallest cost element from all the uncovered cost elements.
3. Add this smallest cost element to all the cost elements lying in the intersection point of the straight lines.
4. Do not change the remaining elements which lie on the straight line.

Step 7: Repeat from step 3 until an optimum assignment is made.

Assignment Problem 1:

Consider the problem of assigning five jobs to five persons. The assignment costs are given as follows.

	1	2	3	4	5
A	8	4	2	6	1
B	0	9	5	5	4
C	3	8	9	2	6
D	4	3	1	0	3
E	9	5	8	9	5

Determine the optimal Assignment Schedule.

Solution: - Assignment Problem 1

Step 1: Select the **smallest cost element in each row** and **subtract** this from all the elements of the corresponding row.

	1	2	3	4	5
A	7	3	1	5	0
B	0	9	5	5	4
C	1	6	7	0	4
D	4	3	1	0	3
E	4	0	3	4	0

Step 2: Select the **smallest cost element in each column** and **subtract** this from all the elements of the corresponding column.

	1	2	3	4	5
A	7	3	0	5	0
B	0	9	4	5	4
C	1	6	6	0	4
D	4	3	0	0	3
E	4	0	2	4	0

Step 3: Assigning Zero's.

*Now each row and column would have at least one zero.

*Examine each row successively until a row with exactly one '0' is found. Circle it and cross all other 0's in the said **column**. Repeat this for all rows.

*Same way examine each column successfully until a column with exactly one '0' is found. Circle it and cross all other 0's in the said **row**. Repeat this for all columns.

	1	2	3	4	5
A	7	3	~~0~~	5	[0]
B	[0]	9	4	5	4
C	1	6	6	[0]	3
D	4	3	[0]	~~0~~	3
E	4	[0]	3	4	~~0~~

Step 4: Applying Optimal Test

Each row and each column contain one encircled zero, the current assignment is optimal.

Optimum Assignment Schedule:

A – 5, B – 1, C – 4, D – 3, E – 2

Optimum Assignment Cost:

1 + 0 + 2 + 1 + 5 = Rs 9.

AP - PROBLEM 2

The Processing time in hours for the jobs when allocated to the different machines is indicated below. Assign machines for the jobs so that the total processing time is minimum.

MACHINES

JOBS	M1	M2	M3	M4	M5
J1	9	22	58	11	19
J2	43	78	72	50	63
J3	41	28	91	37	45
J4	74	42	27	49	39
J5	36	11	57	22	25

Step 1: Select the smallest cost element in each row and subtract this from all the elements of the corresponding row.

	M1	M2	M3	M4	M5
J1	0	13	49	2	10
J2	0	35	29	7	20
J3	13	0	63	9	17
J4	47	15	0	22	12
J5	25	0	46	11	14

Step 2: Select the smallest cost element in each column and subtract this from all the elements of the corresponding column.

	M1	M2	M3	M4	M5
J1	0	13	49	0	0
J2	0	35	29	5	10
J3	13	0	63	7	7
J4	47	15	0	20	2
J5	25	0	46	9	4

Step 3: Assigning Zeros

	M1	M2	M3	M4	M5
J1	~~0~~	13	49	[0]	~~0~~
J2	[0]	35	29	5	10
J3	13	[0]	63	7	7
J4	47	15	[0]	20	2
J5	25	~~0~~	46	9	4

Step 4: Applying Optimal Test

Since row 5 and column 5 does not have any assignment, the above assignment is not optimal.

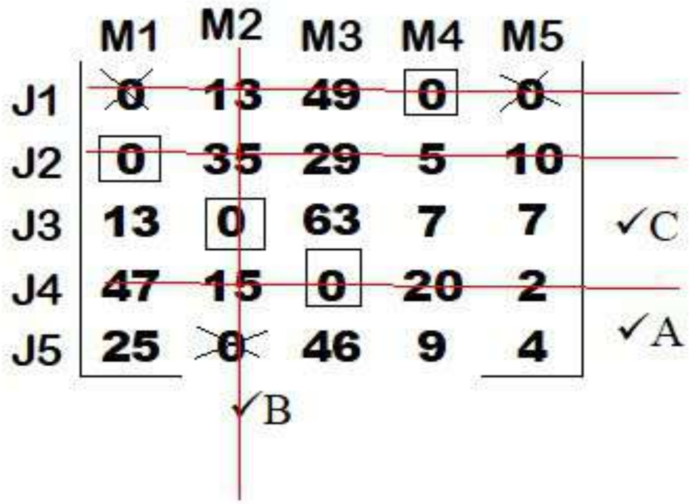
Step 5:

A. **Mark the rows that do not have an assignment.**

B. **Mark the columns** (not already marked) **that have zeros in the marked row.**

C. **Mark the rows** (not already marked) **that have assignments in the marked columns.**

D. Repeat (B) and (C) until no marking is required. E. **Draw lines on all unmarked rows and marked columns.**



Step 6:

1. **Find the smallest cost element not covered by the straight lines.Smallest is 4.**
2. **Subtract this smallest cost element - 4 from all the uncovered cost elements.**
3. **Add this smallest cost element - 4 to all the cost elements lying in the intersection point of the straight lines.**
4. **Do not change the remaining elements which lie on the straight line.**

	M1	M2	M3	M4	M5
J1	0	17	49	0	0
J2	0	39	29	5	10
J3	9	0	59	3	3
J4	47	19	0	20	2
J5	21	0	42	5	0

Step 7: Assigning zeros

Since all rows and columns have at least one zero, repeat step 3- assigning zeros

	M1	M2	M3	M4	M5
J1	~~0~~	17	49	[0]	~~0~~
J2	[0]	39	29	5	10
J3	9	[0]	59	3	3
J4	47	19	[0]	20	2
J5	21	~~0~~	42	5	[0]

Step 8: Applying Optimal test

Since each row and column has one encircled zero, the **current assignment is optimal.**

OUTPUT

The optimal assignment schedule is **J1 - M4, J2 - M1, J3 - M2, J4 - M3, J5 - M5**

= 11+43+28+27+25 = 134 hours

UNBALANCED ASSIGNMENT PROBLEM

If in the given matrix the **number of rows is not equal to number of columns then the problem is said to be unbalanced.**

To solve the unbalanced assignment problem, it has to be **converted into a balanced assignment problem by adding dummy rows or dummy columns with zero cost elements.**

If rows are lesser than columns, **m < n, introduce dummy row**.

If columns are lesser than rows, **m>n, introduce dummy column.**

Assignment problem 3:

Assign four trucks to vacant spaces A,B,C,D,E & F so that the distance travelled is minimised. The following matrix shows the distance.

	1	2	3	4
A	4	7	3	7
B	8	2	5	5
C	4	9	6	9
D	7	5	4	8
E	6	3	5	4
F	6	8	7	3

Solution - Problem 2

Step 1: Since there are 6 rows and 4 columns the problem is unbalanced. We introduce 2 dummy columns to balance it.

	1	2	3	4	5	6
A	4	7	3	7	0	0
B	8	2	5	5	0	0
C	4	9	6	9	0	0
D	7	5	4	8	0	0
E	6	3	5	4	0	0
F	6	8	7	3	0	0

Step 2: Select the **smallest cost element in each row** and **subtract** this from all the elements of the corresponding row.

Step 3: Select the **smallest cost element in each column** and **subtract** this from all the elements of the corresponding column.

	1	2	3	4	5	6
A	0	5	0	4	0	0
B	4	0	2	2	0	0
C	0	7	3	6	0	0
D	3	3	1	5	0	0
E	2	1	2	1	0	0
F	2	6	4	0	0	0

Step 4: Assigning Zeros.

	1	2	3	4	5	6
A	~~0~~	5	[0]	4	~~0~~	~~0~~
B	4	[0]	2	2	~~0~~	~~0~~
C	[0]	7	3	6	~~0~~	~~0~~
D	3	3	1	5	[0]	~~0~~
E	2	1	2	1	~~0~~	[0]
F	2	6	4	[0]	~~0~~	~~0~~

Step 5: Optimal test
Since each row and column has one assignment the current assignment is optimal.
OUTPUT:
The optimal assignment schedule is
A - 3, B - 2, C - 1, D - 5, E - 6, F - 4
= 3+2+4+0+0+3 = 12 units of distance.

(Note: since in problem Km or miles is not specified, it has been taken as units of distance.)

MAXIMISATION CASE IN ASSIGNMENT PROBLEM

To solve the maximisation problem first it has to be converted into a minimisation problem.

How to convert maximisation problem into a minimisation problem?

Subtract all the values in Assignment Matrix from the highest value (OR)

Multiply all the values in Assignment problem by -1.

Assignment Problem 4:

A company has a team of four salesmen and there are four districts where the company wants to start its business. After taking into account the capabilities of salesmen and the nature of districts, the company estimates that the profits per day in rupees for each salesman in each district as follows:

	1	2	3	4
A	16	10	14	11
B	14	11	15	15
C	15	15	13	12
D	13	12	14	15

Find the assignment of salesmen to various districts which will yield maximum profit.

Solution - problem 4 - step 1 - Since the problem is maximisation of profit, first convert it into minimisation problem by **subtracting all the values in matrix from the highest value - 16**

	1	2	3	4
A	0	6	2	5
B	2	5	1	1
C	1	1	3	4
D	3	4	2	1

Now carry on the usual procedure to solve the assignment problem.

Questions - Theory

1. What is a transportation problem? State the assumptions based on which transportation problem is solved.
2. What are different methods of solving a transportation problem?
3. What is a basic feasible solution?
4. When is a solution said to be non-degenerate basic feasible?
5. Give a brief note on degeneracy in transportation problem.
6. Give the meaning of balanced and unbalanced transportation problem.
7. How would you solve a transportation problem using North west corner rule.?
8. How would you solve a transportation problem using Least Cost method?
9. How would you solve a transportation problem using Value added method?
10. What is unit cost penalty method?
11. How is penalty computed in solving a transportation problem?
12. How would you solve a maximisation transportation problem?
13. What do mean by assignment problem?
14. Give the meaning of balanced and unbalanced assignment problem.
15. Differentiate transportation problem from assignment problem.
16. Explain the algorithm to solve transportation problem.
17. Enumerate the procedure to solve an assignment problem.
18. Explain the Hungarian method of solving an assignment problem.

QUANTITATIVE TECHNIQUES – GAME THEORY
Unit V

BY
DR.J.J.JEYAKUMARI., M.Com., M.phil., Ph.D., MBA., PGDCA.,
RESEARCH ADVISOR & ASSISTANT PROFESSOR OF COMMERCE,
K.N.GOVT. ARTS COLLEGE FOR WOMEN, AUTONOMOUS, THANJAVUR -7

Reference: Research Management Techniques, Sundaresan V, Ganapathy Subramaniam S, Ganesan K

COMPETITIVE SITUATION

A situation in which there are two or more opposite parties with conflicting interest and the action of one depends upon the action which the opponent takes. The outcome of a situation is controlled by the decisions of all the parties involved in this situation and is called a competitive situation.

GAMES

The term games refers to the general situation of conflict and competition in which two or more competitors are involved in decision making activities in anticipation of certain outcomes over a period of time. The competitors are referred to as players.

GAME

A **competitive situation is a game** if it has the following properties:

1. There is a **finite number of competitors** called players.
2. A list of **finite or infinite number of course of action is available to each player.** A play is played when each player chooses one of his course of action. The course of actions are simultaneous and unknown to others until decided.
3. **Every play** (Combination of course of action) is **associated with an outcome, known as the payoff** (generally money) which determine a set of gains, one to each player. Loss is considered as a negative gain.

Two players - A and B - Payoff matrix indicates gains to A (in each possible outcome) and negative entries denotes payment from A to B.

Assumptions

- **Player A - Maximising player,**
- **Player B - Minimising player.**

Other Assumptions - 1. All players act rationally, 2. Each player optimise his gain or loss, 3. Complete relevant information is known to each player, 4. Each player makes individual decisions without communication, and 5. A game involving 'n' players is called n-person game.

Two types of strategy → **Pure startegy** / **Mixed Strategy**

1. Pure Strategy : Each player knows in advance all the available strategies out of which he selects one, regardless of the other players strategy and the objective is to maximize gain or minimise loss.
2. Mixed strategy : Courses of action that are to be selected on a particular occasion with some fixed probability are called mixed strategy. Choice to be made among pure strategies with fixed probabilities.

Main characteristics of Game theory

1. Number of persons or groups playing the game is finite.

2. Number of activities which may be finite or infinite

3. Alternative course of action may be finite or infinite.

4. The amount of information about past activities - complete, part or none

Two Person-Zero Sum Games / Rectangular Games

Zero-sum games with two players. In this, **gain of one player is exactly equal to the loss of other player.** The gains represented in the matrix form called payoff matrix.

Sum of Gains of one player = Sum of Losses of other player.

So that the **sum of gains and losses is exactly equal to zero**. Then it is a zero sum game.

Otherwise non-zero sum game.

The particular strategy (or complete plan) by which a player optimises his gains or losses knowing the competitors strategy is **optimal strategy.**

The maxi-min and mini-max principle

Saddle point : A saddle point of a payoff matrix is that position or point in payoff matrix where maximin coincides (intersects) with minimax. The payoff at saddle point is called value of the game denoted by **v.**

minimax = $\overline{\mathbf{v}}$

maximin = $\underline{\mathbf{v}}$

$\overline{\mathbf{v}} = \mathbf{v} = \underline{\mathbf{v}}$ **the game is strictly determinable.**

$\underline{\mathbf{v}} = 0 = \overline{\mathbf{v}}$ **the game is fair.**

The Maximin - Minmax Principle

Example: Consider the following game:

	Player A	
Player B	1	1
	4	-3

Step 1: Find row minima & column maxima

	Player A		Row minima
Player B	1	1	1
	4	-3	-3
Column maxima	4	1	

Step 2: Minimum of maximum = Minimax

Maximum of Minimum = Maximin

So minimum of maximum is **1 = minimax =** $\overline{v}$

and maximum of minimum is **1 = maximin =** $\underline{v}$

The selection of strategies by A and B depends upon the maximin-minimax principle.

Step 3: Find Saddle point of a payoff matrix - is the position in payoff matrix where maximin coincides (intersects) with minimax.

So the saddle point is (1,2) and the value at saddle point v = 1

Step 4:

$\overline{v} = \underline{v}$ **So pure strategy**

$\overline{v} = v = \underline{v}$ **and the game is strictly determinable.**

GT - Problem 1 - Pure Strategy

Solve the game whose payoff matrix is given by:

Player B

		B1	B2	B3
PlayerA	A1	1	3	1
	A2	0	-4	-3
	A3	1	5	-1

Solution -1

Step 1: Maximin-Minmax principle

Player B

		B1	B2	B3	Row minima
PlayerA	A1	1	3	1	1
	A2	0	-4	-3	-4
	A3	1	5	-1	-1
Column maxima		1	5	1	

Minimax = 1, Maximin = 1

Step 2: Saddle point - intersection point is (1,1) & (1,3)

v=1 i both places.

S_0 = (A1,B1) OR (A1,B3)

Output:

Minimax = Maximin

Pure strategy

Value of game v = 1.

GT - Problem 2 - Mixed strategy

Solve the game whose payoff matrix is given by:

Vijay

		D	J
Surya	1	6	15
	2	10	5

Solution - Problem 2

Step 1: Check for dominance

Vijay

Surya		D	J	Row minima
	1	6	15	6
	2	10	5	5
Column maxima		10	15	

Minimax (10) is not equal to Maximin(6). **No saddle point. So Mixed Strategy.**

Step 2:

Let **Q** be the number of times Surya plays Strategy 1, then

1-Q is the number of times Surya plays Strategy 2

Let **P** be the number of times Vijay plays Strategy D, then

1-P is the number of times Vijay plays Strategy J

Vijay

Surya		P D	1-P J
	Q1	6	15
	1-Q2	10	5

Step 3:

Vijay's Payoff when Surya plays Strategy 1 = Vijay's Payoff when Surya plays Strategy 1

6P + 15(1 - P) = 10P + 5(1 - P)

6P + 15 - 15P = 10P + 5 - 5P

15 – 9P = 5P + 5

-9P – 5P = 5 – 15

-14P = -10

P = -10/-14 = .714

1-P = 1 - .714 = .286

So probability of Vijay playing strategy D is 71.4 % and Strategy J is 28.6 %.

Step 4:

Surya's Payoff when Vijay plays Strategy D = Surya's Payoff when Vijay plays Strategy J

6Q + 10(1 -QP) = 15Q + 5(1 - Q)

6Q + 10 – 10Q = 15Q + 5 – 5Q

10 – 4Q = 10Q + 5

-4Q – 10Q = 5 – 10

-14Q = -5

Q = -5/-14 = .357

1-Q = 1 - .357 = .643

So probability of Surya playing strategy 1 is 35.7 % and Strategy 2 is 64.3 %.

		Vijay	
		P (.714) D	1-P (.286) J
Surya	Q(.357) 1	6	15
	1-Q(.643) 2	10	5

Step 5:

a) Probability of each payoff (Joint)

P(6) = P(1) x P(D) = (.357) (.714) = .255 (25.5 %)

P(15) = P(1) x P(J) = (.357) (.286) = .102 (10.2%)

P(10) = P(2) x P(D) = (.643) (.714) = .459 (45.9%)

P(5) = P(2) x P(J) = (.643) (.286) = .184 (18.4 %)

Total = 1 (100 %)

b) Joint Probability

		Vijay	
		D	J
Surya	1	.255	.102
	2	.459	.184

Step 6: Expected Value of Game

EV = Σ (Payoff x Joint Probability)

EV1D = 6 X .255 = 1.53

EV2D = 10 X .459 = 4.59

EV1J = 15 X .102 = 1.53

EV2J = 5 X .184 = 0.92

EXPECTED VALUE = 8.57

Answer EV = 8.57

Game Theory – Dominance

Superior Strategy dominates inferior strategy. Two ways to check dominance.

1. Check for dominance.
2. Decision Analysis - Assume risk and play it safe.
3.

Problem 3: Solve the following using dominance property.

Jill

		A	B	C
	1	15	25	35
Jack	2	12	14	16
	3	2	4	7

Solution – Problem 3

Let Jack be the maximizing player and Jill be the minimizing player.

Jill (Min)

		A	B	C
(Max)	1	15	25	35
Jack	2	12	14	16
	3	2	4	7

For Jack (Maximising Player) - Compare to find maximum

Compare Strategy 1 & 2 = 1 > 2, So Strategy 1 dominates 2. – Strike 2.

Compare Strategy 1 & 3 = 1 > 3, So Strategy 1 dominates 3. – Strike 3.

For Bruce Minimising player – Compare to find minimum

Compare Strategy A & B = A < B, So Strategy A dominates B. – Strike B

Compare Strategy A & C = A < C , So strategy A dominates C. – Strike C

Now we could see that there exists a pure strategy at (1,A) = 15.

Answer:

So strategy 1 for Jack and Strategy A for Jill is the pure strategy.

Value of game = 15.

Game Theory - Decision Analysis

Problem 4: Solve the following using decision analysis.

Jill

		A	B	C
	1	15	25	35
Jack	2	12	14	16
	3	2	4	7

Solution - Problem 4

Step 1: Maximin

- **Identify minimum for each strategy** - Eg: S1 – min 15, S2 – min 12, S3 – min 2
- **Select maximum** - 15

Step 2: Minimax

- ★ **Identify maximum for each strategy** - Eg: A – max 15, B – max 25, C – max 35
- **Select minimum** - 15

Jill

		A	B	C
	1	15 ★	25 ★	35 ★
Jack	2	12	14	16
	3	2	4	7

Now we could see that **there exists a pure strategy at (1,A) i.e** ★ **= 15**

Answer:

So strategy 1 for Jack and Strategy A for Jill is the pure strategy.

Value of game = 15.

Game theory – Graphical Method

Used to solve

- 2 x n games – 2 rows & n columns
- m x 2 games – m rows & 2 columns

Problem 5: 2 x n games

Player B

		B1	B2	B3	B4	B5
PlayerA	A1	-4	2	5	-6	6
	A3	3	9	7	4	8

Solution - Problem 5

Step 1: Find saddle point to know the strategy.

Player B

		B1	B2	B3	B4	B5	Row minima
PlayerA	A1	-4	2	5	-6	6	-6
	A3	3	-9	7	4	8	-9
Column maxima		3	2	7	4	8	

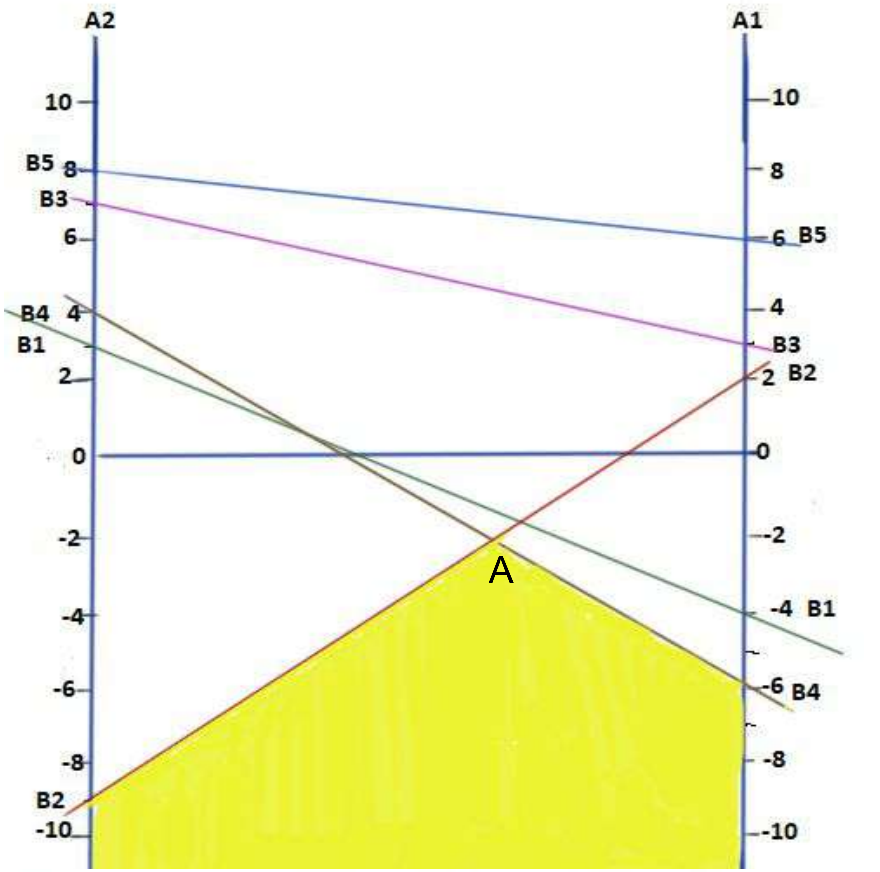
Minimax = 2 **Maximin = -6** **Maximin ? Minimax**

No Saddle point – So mixed startegy

Step 2: Reduce the size of payoff matrix by applying dominance property if it exists.

Step 3 : Graph – Draw A1 & A2 lines - Maximum value 8 , Minimum value -9

Plot A1 & A2 in graph



Step 4: From point of view of A maximising player

Maximin – So shade the minimum (less than everyhing)

Step 5:

B2 & B4 are in the shaded area. The payoff is now reduced to 2 x 2 game.

A \ B	2	4
1	2	-6
2	-9	4

Now usual procedure should be followed to solve the mixed strategy game using probability.

Problem 6 - m x 2 games

PlayerA \ Player B	1	2
1	6	-7
2	1	3
3	3	1
4	5	-1

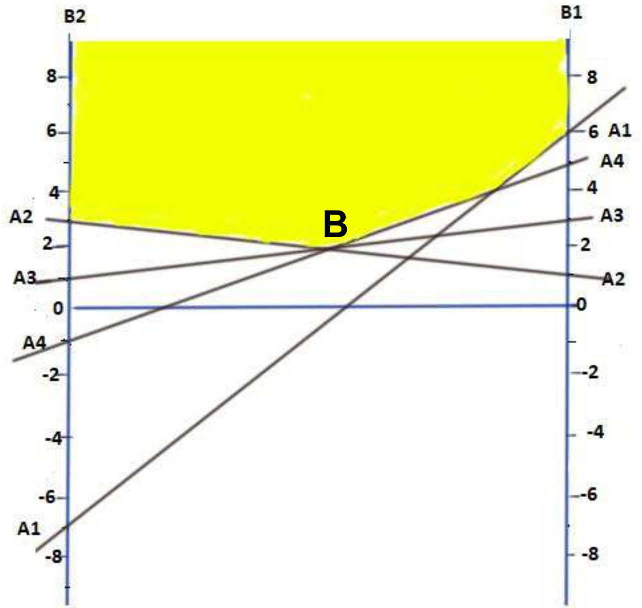
Step 1: No Saddle point. So mixed strategy.

Step 2: Dominance rule not applicable.

Step 3: Graph - Draw B1 & B2 lines.

Max value: 6, Min value: -7.

Plot B1 & B2 values.



Step 4: **From the point of B minimax.** So **shade the lowest intersection point in the upper boundary.**

Minimax point = B

3 lines A2, A3 & A4 passes through the minimax point. Can select any two lines with opposite slopes.

Therefore A2 & A4 is selected. The reduced 2 x 2 game is

		Player B	
		1	2
PlayerA	2	1	3
	4	5	-1

Now usual procedure should be followed to solve the mixed strategy game using probability.

Questions - Theory

1. What is a game?
2. What are the main characteristics of a game?
3. When does a competitive situation arise?
4. List down the assumptions made in solving a game.
5. What are the types of strategy followed in games?
6. What is meant by pure strategy in game theory?
7. What is mixed strategy in game theory?
8. Define the dominance strategy.
9. State the minmax and maximin principal.
10. What is a saddle point? How it is helpful in solving a game?
11. When iis a game said to be strictly determinable?
12. When is a game said to be fair?
13. Briefly expalin the procedure to solve a game.