KARNATAKA STATE OPEN UNIVRSITY

Mukthagangothri . Mysuru 570006

MBA

(Third Semester)

ELECTIVE D : OPERATIONS

OPERATIONS RESEARCH AND ANALYTICS

MB116D/ SC D3.1

BLOCK 1- 4

KARNATAKA STATE OPEN UNIVERSITY

MUKTHAGANGOTHRI, MYSURU- 570 006.

DEPARTMENT OF STUDIES AND RESEARCH IN MANAGEMENT

M.B.A III SEMESTER

COURSE - MBSC - 3.1 D / MB 116 D

OPERATIONS RESEARCH AND ANALYTICS

BLOCK - 1: INTRODUCTION TO OPERATION RESEARCH

UNIT - 1
INTRODUCTION TO OR & LINEAR PROGRAMMING 01-18

UNIT - 2
INTRODUCTION TO TRANSPORTATION PROBLEMS 19-31

UNIT - 3
OPTIMAL SOLUTION 32-48

UNIT - 4
ASSIGNMENT 49-58

BLOCK-2: SEQUENCING AND REPLACEMENT

UNIT - 5
SEQUENCING PROBLEMS OF 2 & 3 MACHINES 59-68

UNIT - 6
SEQUENCING PROBLEMS OF N JOBS AND M MACHINES 69-80

UNIT - 7
REPLACEMENT 81-92

UNIT - 8
TRAVELLING SALESMAN PROBLEM 93-106

BLOCK-3: QUALITY CONTROL AND SIMULATION

UNIT - 9
STATISTICAL QUALITY CONTROL TECHNIQUES 107-133

UNIT - 10
MARKOV CHAIN AND ANALYSIS 134-145

UNIT - 11
SIMULATION 146-157

UNIT - 12
SIMULATION- MONTE CARLO METHOD 158-174

BLOCK-4: BUSINESS ANALYTICS

UNIT - 13
INTRODUCTION TO BUSINESS ANALYTICS 175-188

UNIT - 14
BIG DATA MANAGEMENT 212-221

UNIT - 15
DATA MINING AND ANALYSIS 222-231

UNIT - 16
APPLICATIONS OF BUSINESS ANALYTICS 232-241

CREDIT PAGE

Programme Name : MBA **Year/Semester : 2nd Year, 3rd Semester** **Block No: 1 to 4**

Course Name : Operations Research and Analytics **Credit : 04** **Units No: 1 to 16**

Course Design Expert Committee

Prof. Vidya Shankar Vice-Chancellor Karnataka State Open University Mukthagangothri, Mysore-06	**Chairman**	**Prof. Kamble Ashok** Dean (Academic) Karnataka State Open University Mukthagangothri, Mysore-06	**Member**

Course Designer/Course Co-ordinator

Dr. Rajeshwari H. Assistant Professor, DOS & R in Management KSOU, Mukthagangothri, Mysore-06	**BOS Chairman & Member**	**Dr. P. Savitha** Assistant Professor, DOS & R in Management KSOU, Mukthagangothri, Mysore-06	**Department Chairman & Member Convener**

Editorial Committee

Dr. Rajeshwari H. Assistant Professor DOS & R in Management KSOU, Mukthagangothri, Mysore-06	**BOS Chairman & Member**	**Prof. S. J. Manjunath** BIMS, UoM Mysuru	**External Subject Expert & Member**
Prof. C. Mahadevamurthy Professor DOS & R in Management KSOU, Mukthagangothri, Mysore-06	**Internal Subject Expert & Member**	**Dr. P. Savitha** Assistant Professor, DOS & R in Management KSOU, Mukthagangothri, Mysore-06	**Department Chairman & Member Convener**

Course Writers		**Course Editor**	
Dr. Rajeshwari H DoS & R in Management KSOU, Mukthagangothri Mysore-06.	**Unit 1-8**	**Prof. C. Mahadevamurthy** Professor DOS&R in Management, KSOU, Mukthagangothri, Mysuru–06	**Unit 1-8**
		Dr. Rajeshwari H DoS & R in Management KSOU, Mukthagangothri Mysore-06.	**Unit 9-16**

Prof. Nagesh **Unit 9-12**
SJCE-CMS
Mysore

Dr. Yatishchandra **Unit 13-16**
Associate Professor
Department of Management
VTU- PG Centre,
Mysore

Copy Right

Registrar
Karnataka State Open University
Mukthagangothri, Mysuru - 570006

Developed by the Department of Studies and Research in Management, KSOU, under the guidance of Dean (Academic), KSOU, Mysuru

Karnataka State Open University, January-2021

Further information on the Karnataka State Open University Programmes may obtained from the University's office at Mukthagangothri, Mysuru-570006

Printed and Published on behalf of Karnataka State Open University. Mysuru-570006 by **Registrar (Administration)-2021**

BLOCK -1 INTRODUCTION

BLOCK - 2 INTRODUCTION

BLOCK - 3 INTRODUCTION

BLOCK - 4 INTRODUCTION

BLOCK - I

UNIT -1: INTRODUCTION TO OR & LINEAR PROGRAMMING

Contents:

1.0 Objectives

1.1 Introduction to Operations Research

1.2 Meaning of Operations Research

1.3 Scope of Operations Research

1.4 Introduction to Linear Programming

1.5 Formulation of Linear Programming Models

1.6 Graphical Method of Solving Linear Programs

1.7 Simplex Method of Solving Linear Programs

1.8 Key Words

1.9 Summary

1.10 Self Assessment Questions

1.11 Reference

1.0 OBJECTIVES

After studying this unit you will be able to;

- Explain meaning and scope of OR.
- Formulate L.P.P.
- Solve L.P.P. using graphical & simpler method.

1.1 INTRODUCTION

Operations research is a new branch of mathematics dealing with the optimization problems in real-life situation. It is also a quantitative technique to deal many management problems. In this discipline, we study cost minimization of various inventory problems, the minimization of transportation costs of sending goods from various warehouses to different centers, the profit maximization or cost minimization in linear programming models, the assignment of different persons to different jobs so that the total time taken to perform the job is minimized, the congestion problem in traffic places, airline counters, supermarkets, to find out the waiting time of customers in the queue, the project completion time with limited resources and many more similar problems.

1.2 MEANING OF OPERATIONS RESEARCH

Operation research has a number of quite distinct variations of meaning. To some, OR is that certain body of problems, techniques, and solutions that has been accumulated under the name of OR over the past 30 to 40 years, and we apply OR when we recognize a problem of that certain genre. To other, it is an activity or process something we do rather than know-which by its very nature is applied. How, then, can we define operations research. The Operational Research Society of Great Britain has adopted the following. definition: “Operational research is the application of the methods of science to complex problems arising in the direction n and management of large system of men, machines, materials and money in industry, business, government and defense. The distinctive approach is to develop a scientific, model of the system, incorporating measurements of factors such as chance and risk, with which to predict and compare the outcomes of alternative decisions, strategies or controls. The purpose is to help management determing its policy and actions scientifically”.

The Operation research society of America has offered a shorter, but similar description:

“Operations research is concerned with scientifically deciding how to best design and operate man-machine systems, usually under conditions requiring the allocation of scarce resources”. It is noteworthy that both definition emphasize the motivation for the work: namely, to aid decision makers in dealing with complex real-world problems. A more precise description of the OR methodology would indicate its reliance on “models”.

1.3 SCOPE OF OPERATIONS RESEARCH

The scope of OR is not confined to any specific agency like defense services or an industrial field. The scope is wider. It is useful in every field of human activities, where optimization of resources is required. Some of the industrial/government/business problems which can be analyzed by OR is listed as below;

1.3.1 Finance and Accounting

1.3.1.1 Dividend policies, investment portfolio management, auditing, balance sheet and cash flow analysis.

1.3.1.2 Break-even analysis, capital budgeting, cost allocation and control, financial, planning.

1.3.1.3 Claims and complaint procedure, public accounting.

1.3.2 Marketing

1.3.2.3 Sales effort allocation and assignment.

1.3.2.1 Selection of product-mix, marketing planning, export planning

1.3.2.2 Advertising and media planning

1.3.2.3 Sales effort allocation and assignment.

1.3.3 Purchasing, procurement and exploration.

1.3.3.1. Optimal buying and recording under price quantity discount.

1.3.3.2. Bidding policies.

1.3.3.3. Transportation planning.

1.3.3.4. Vendor analysis.

1.3.3.5. Replacement policies.

1.3.4. Production Management.

1.3.4.1 Facilities Planning

1.3.4.1.1. Location and size of warehouse, distribution centers and retail Outlets

1.3.4.1.2. Logistics, layout, engineering design.

1.3.4.1.3. Transport planning and scheduling.

1.3.4.2. Manufacturing

1.3.4.2.1. Aggregate production planning, assembly line, blending, purchasing, inventory control.

1.3.4.2.2. Employment, training, and quality control.

1.3.4.3. Maintenance and project Scheduling:

1.3.4.3.1. Maintenance policies and preventive maintenance.

1.3.4.3.2. Maintenance crew size and scheduling

1.3.4.3.3. Project scheduling and allocation of resources.

1.3.5. Personnel Management.

1.3.5.1. Manpower planning,

1.3.5.2. Scheduling of training programmes.

1.3.6. Government

1.3.6.1. Economic planning, natural resources, social planning, energy.

1.3.6.2. Military, police, pollution control.

1.3.6.3. Urban and housing problems.

1.4. INTRODUCTION TO LINEAR PROGRAMMING

Linear Programming (LP) is the general technique of optimum allocation of 'scarce' or 'limited' resources, such as labour, material, machine, capital, energy, etc., to several competing activities such as products, services, jobs, new equipments, products, etc. On the basis of a given criterion of optimality. Thus the basic objective of LP is to minimize cost and maximize profit. This is called the "objective Function". The word linear in Linear Programme, describes the proportionate relationship of two or more variables in a model. The word programming is used to specify a sort of planning that involves the economic allocation of limited resources by adopting a particular course of action amongst various alternatives to achieve the desired objectives.

1.5 FORMULATION OF LINEAR PROGRAMMING M ODELS

The process for mathematical formulation of Liner programming model consists of the following steps.

Step – 1: Identify the decision variable of the problem.

Step – 2: Formulate the objective function to be optimized (maximize or minimized) as a linear function of the decision variables.

Step – 3: Formulate the constraints of the problem such resource limitations, market conditions, interrelation between variables and others as linear equation or inequalities in terms of the decision variable.

Step – 4: Add the non-negativity constraints so that negative values of the decisions variable do not have any valid physical interpretation.

Problem 1.5.1:

A firm manufactures two type of products, A and B and sells them at profit of Rs.2 on A and Rs. 3 on type B. Each product is processed on two machines M and N. Type requires 1 minute of processing time on machine M and 2 minutes on machine N. Type requires 1 minute on M and 1 minute on N. The machine M is available for not more than 6 hours 40 minutes while machine N is available for 10hours per day. Formulate the linear programming model.

Solution:

Decision Variable:

Let 'X' be the number of products of type A and 'Y' number of units of type B to be produced. The given in formation is tabulated as shown below:

Machine	Time Products(Min)		Available time (min)
	A	B	
M	1	1	400
N	2	1	600

Table – 1.1

Objective Function:

Since the profit on type A is Rs.2 per unit, 2X will be the profit on selling X units of type A. Similarly 3Y will be the profit on selling Y units of type B. Therefore, total profit Z on selling X units of A and Y units of B is **Z =2X+3Y.** This is the objective function.

Constraints:

Machine M takes 1 minute time on type A and 1 minute of time on type B, the total number of minutes required on machine M is X + Y. But machine M is available for not more than 6 hours and 40 minutes (400 minutes). Therefore, X + Y < = 400. Similarly for machine N the total time available is 10 hours (600 minutes) thus constraint is 2X + Y < = 600.

Non-Negativity Restriction :

Since it is not possible to produce negative quantities we have X>=0 and Y>=0. Hence the linear programme model is,

Maximize $\quad Z = 2X + 3Y$

Subjected to,

$$X + Y <= 400$$

$$2X + Y <= 600$$

And $\quad X, Y >= 0$

Problem 1.5.2 :

A Company has two plants each of which produces and supplies two products: A and B. The plants can each work up to 16 hours a day. In plant 1, it takes 3 hours to prepare and pack 1000 gallons of A and 1 hour to prepare and pack 1 quintal of B, in plant 2 it takes 2 hours to prepare and pack 1000 gallons of A and 1.5 hours to prepare and pack 1 quintal of B. In plant 1, it cost Rs. 15,000 to prepare and pack 1000 gallons of A and Rs. 28,000 to prepare and pack a quintal of B whereas these costs are Rs. 18,000 and Rs. 26,000 respectively in plant 2. The company is obliged to produce daily at least 10,000 gallons of A and 8 quintals of B. Formulate the LP model.

Solution :

X1 be the quantity of product A (in 1000 gallons) to be produced in plant 1.

X2 be the quantity of produce B (in quintal) to be produced in plant 1.

X3 be the quantity of product A (in 1000 gallons) to be produced in plant 2.

X4 be the quantity of product B (in quintal) to be produced in plant 2.

The given information is tabulated as shown below:

Plant	**Time on Products**		**Available time (hours)**
	A	B	
1	3 hrs / 1000 gallon	1 hrs / quintal	16
2	2 hrs / 1000 gallon	1.5 hrs / quintal	16

Table – 1.2

Objective Function :

Since it costs Rs. 15,000 to prepare and pack 1,000 gallons of A and Rs. 28,000 to prepare and pack a quintal of B in plant 1, whereas these costs are Rs. 18,000 and Rs. 26,000 respectively in plant 2. Thus the total cost Z has to be minimized, **Z = 15000 X1 + 28000 X2 + 18000 X3 +26000X4** is the objective function.

Constraints :

In plant 1 it takes 3 hours to manufacture 1000 gallons of product type A and 1 hour of time on type B, the total number of hours required in plant 1 is **3X1 + X2**. Plant 1 works up to 16 hours a day. Therefore 3X1 + X2 < = 16. Similarly for plant 2 thus constraint is **2X3 + 1.5X4 < = 16.** Further it is necessary to manufacture at least 10000 gallons of A and 8 quintals of B. Therefore the other constraints are **X1 + X3>=10** and **X2 + X4 > = 8**, Thus the linear programme model is.

Minimize Z = 15000 X1 + 28000 X2 + 18000X3 + 26000 X4

Subjected to,

$3X_1 + X_2 <= 16$

$2X_3 + 1.5X_4 <= 16$

$X_1 + X_3 >= 10$

$X_2 + X_4 >= 8$

$X_1, X_2, X_3, X_4 >= 0$

1.5 GRAPHICAL, METHOD OF SOLVING LINEAR PROGRAMMES

Following are the steps involved in graphical solution of a linear programme module.

Step – 1 : Formulation of mathematical model as explained in the previous section.

Step – 2 : Convert the constraints given in the form of inequality to that of equality.

Step – 3 : Draw a graph and represent the constraint equation as lines on the graph. For each constraints there will be a line.

Step – 4 : Identify the feasible area of the solution (convex region), which satisfies all constraints.

Step – 5 : The convex region (feasible area) will be a polynomial having several lines. It will also have as many corner points at that of interesting lines. One of the corner points will represents the optimum point.

Step – 6 : In order to identify the optimum point, estimate the value of Z at each point at which the objective function is satisfied. That point represents the optimum point.

Problem 1.6.1

Use the graphical method to solve the following linear programme problem :

Maximize $Z = 15X_1 + 10X_2$.

Subjected to the constraints,

$4X_1 + 6X_2 <= 360$

$3X_1 + 0X_2 <= 180$

$0X_1 + 5X_2 <= 200$

$X_1, X_2 > 0$

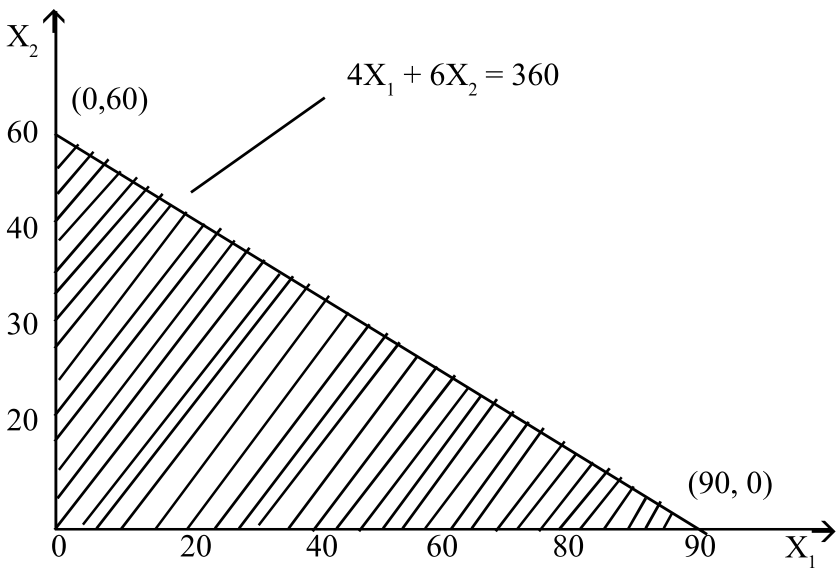
Step – 1 : State the problem in mathematical form. The problem is already to mathematical form.

Step – 2 : Convert inequality form into equality form.

$4X_1 + 6X_2 = 360$; $3X_1 + 0X_2 = 180$; $0X_1 + 5X_2 = 200$; $X_1 = 0$ and $X_2 = 0$

Step – 3 : Draw the graph and represents the equation in step 2. This is done by assigning different values of X1 and finding and corresponding value of X2. Since X1 and X2 represent linear equation, the loci will be a straight line.

FIGURE 1.1



In the equation 4X1 + 6X2 = 360, Let X2 = 0 therefore X1 = 360/4 = 90 SIMILARLY LET X1 = 0, X2 = 360/6 = 60. Similarly the constraints 3X1 = 180 and 5X2 = 200 are also plotted on the graph and indicated by the shaded area as shown in the figure below. The feasible region is indicated the points OABCD.

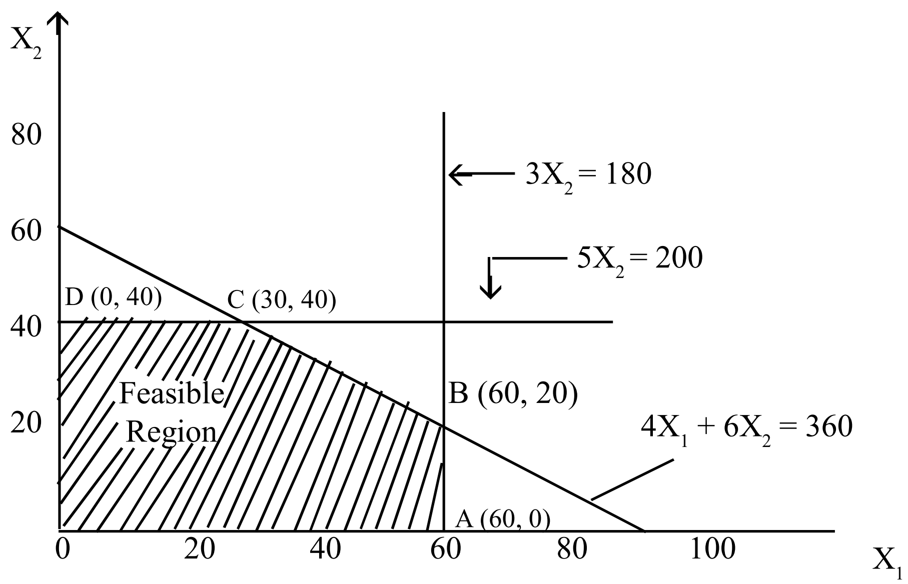


FIGURE 1.2

Step – 4 : Determine the coordinate of extreme points.
O = (0,0), A = (60,0), B =(60,20), C = (30,40) and D = (0,40)

Step – 5 : Evaluate the value of the objective at extreme points

Extreme Point	Coordinates (X1, X2)	Objective function value Z = 15X1 +10X2
O	(0,0)	15 (0) + 10 (0) = 0
A	(60, 0)	15 (60) + 10 (0) = 900
B	(60, 20)	15 (60) + 10 (20) = 1100
C	(30, 40)	15 (30) + 10 (40) = 850
D	(0, 40)	15 (0) + 10 (40) = 400

Table : 1.3

Since the objective function is to maximize, the optimal solution to the problem is at point B (60, 20) therefore X1 = 60 and X2 = 20 and Max Z =1100.

Problem 1.6.2 :

Use the graphical method to solve the following linear programme problem :

Minimize Z = - X1 + 2X2

Subjected to the constraints,

-X1 + 3X2 < = 10

X1 + X2 < = 6

X1 – X2 < = 2

X1, X2 > = 0

Solution :

Step – 1 : State the problem in mathematical form. The problem is already in mathematical form.

Step – 2 : Convert inequality form into equality form
-X1 +3X2 = 10; X1 + X2 = 6; X1 – X2 = 2; X1, X2 = 0

Step – 3 : Draw the graph and represent the equation in step 2. This is done by assigning different values of X1 and finding the corresponding value of X2. Since X1 and X2 represent linear equation, the loci will be a straight line.

In the equation –X1 + 3X2 = 10, let X2 = 0 therefore X1 = -10/1 = -10 similarly let X1=0, X2=10/3 = 3.33. Similarly the constraints X1 + X2 = 6 and X1 – X2 = 2 are also plotted on the graph and indicated by the shaded area as shown in the figure below. The feasible region is indicated the points OABCD.

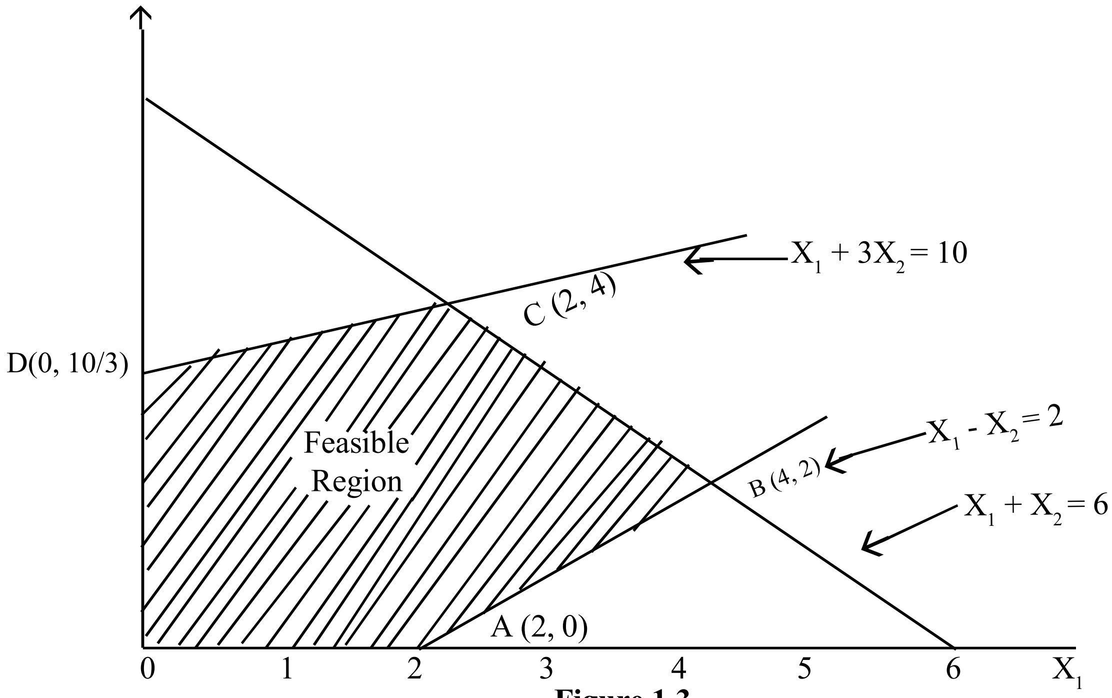


Figure 1.3

Step – 4 : Determine the coordinate of extreme points.

O = (0,0), A = (2,0), B = (4,2), C = (2,4), and D = (0, 10/3)

Step – 5 : Evaluate the value of the objective at extreme points

Extreme Point	**Coordinates (X1, X2)**	**Objective function value Z = X1 + 2X2**
O	(0,0)	-1(0) + 2(0) =0
A	(2, 0)	-1(2) + 2(0) = -2
B	(4, 2)	-1(4) + 2(2) = 0
C	(2, 4)	-1(2) + 2(4) = 6
D	(0, 10/3)	-1(0) + 2(10/3) = 20/3

Table : 1.4

Since the objective function is to maximize, the optimal solution to the problem is at point A (2,0) therefore X1 = 2 and X2 = 0 and Min Z = -2.

1.7 SIMPLEX METHOD OF SOLVING LINEAR PROGRAMS

Following are the steps involved in simplex method of solving a linear programme module.

Step – 1 : Formulation of LPP.

Step – 2 : Convert constraints into equality form.

Step – 3 : Construct the starting simplex table.

Step – 4 : Test optimality by analysis.

Step – 5 : Find “Incoming” and “Outgoing” variables and rewrite the table as per the rules.

Step – 6 : Repeat from step-4 onwards again till the optimum basic feasible solution is reached.

Problem 1.7.1

A manufacturer produces two products A and b each of which gives a profit of Rs. 4 and Rs.3 respectively per unit. These are made either in plant 1 or plant 2. Plant 1 and plant 2 have capacities of 72 and 48 hours per day. Plant 1 takes 2 hours and 1 hour to produce A and B respectively. Similarly plant 2 takes 1 and 2 hours for product A and B respectively. Determine optimum product mix to maximize profit by simplex method.

Solution

Step – 1: Formulation of LPP

Plant	**Time on Products**		**Available time (hours)**
	A	B	
1	2 hrs	1 hrs	72
2	1 hrs	2 hrs	48

Thus,

Maximize Z = 4X1 + 3X2

Subjected to constraints

2X1 + X2 < = 72

X1 + 2X2 < = 48

X1, X2 > = 0

Step – 2: Introducing non-negative slack variables S1 and S2 to convert inequalities to equality.

The LPP becomes,

Maximize Z = 4X1 + 3X2 +0S1 + 0S2

Subjected to constraints 2X1 + X2 + S1 = 72

X1 + 2X2 + S2 = 48

X1, X2, S1, S2,>=0

Step – 3: Construct the simplex table.

C	**Variables**	**Solution**	**4**	**3**	**0**	**0**	**Ratio**
	B	X_B	X1	X2	S1	S2	
0	S1	72	$\boxed{2}$	1	1	0	36 →
0	S2	48	1	2	0	1	48
	Z	0	0	0	0	0	
		C-Z	4 ↑	3	0	0	-

Table -1.5

Explanation:

1. Product mix (B): This is the basic variables. These are variables whose coefficients are unit vectors (Unit Matrices)
2. Product vector (C): Cost vector. This is profit per unit in this problem. When S1 and S2 are produced in the initial solution profit = 0. It denotes corresponding co-efficient of B.
3. Basic Solution (X_B): Basic vector. In the initial solution, $X_B = \begin{bmatrix} 72 \\ 48 \end{bmatrix}$
4. Objective function Z = 72S1 + 48S2
5. C-Z = Net profit addition which results by introducing one of the corresponding variable to product mix. In initial solution co-efficient of objective function.

Step – 4: Iteration 1

Rule -1: Select the maximum value in the C-Z row as the incoming value in the Initial table. This column is indicated by an upward arrow. Calculate the ratio as shown below.

$R=X_B/E$ Where E the Co-efficient element of optimum column. In this case we have

R_{11}=72/2=36

R_{21} = 48/1 = 48

Rule – 2: Select row having the minimum ratio value as the outgoing value. A horizontal arrow is drawn to identify this row. This row is called the replacing row. The intersection element of replacing row and optimum column is called the key element. Hence the key element = 2

Rule - 3: Replace the old elements in the replacing row (the outgoing row) with new values. **New values of replacing row = old values/value of key element**

Old Values	0	S1	72	2	1	1	0
New values	0	X1	72/2=36	2/2=1	1/2	1/2	0

Table – 1.6

Rule – 4: Replace the old element of other rows with new values

New value=old value-(intersection element of old row x corresponding new value of the replacing row) For row 2 the intersecting value is 1

Old Value	0	S2	48	1	2	0	1
New Values	0	S2	48-(1x36) =12	1-(1x1) =0	2-1X1/2) =3/2	0-(1X1/2) =1/2	1-(1x0) =1

Now C Value I row is changed to 4. i.e., coefficient of X1

Similarly do the same if there are more rows. The second table looks as follows:

C	Variables in Basis	Solution Values	4	3	0	0	Ratio	
	B	X_B	X1	X2	S1	S2		
4	X1	36	1	1/2	1/2	0	72	(36 ÷ 1/2)
0	S2	12	0	$\boxed{3/2}$	-1/2	1	8 →	(12 ÷ 3/2)
	Z	4x36=144	4	2	2	0	-	(Multiply row X1 with 4)
	C-Z	-	0	1	-2	0	-	(Deduct the value from original values)

Table – 1.7

1. (C-Z) row has a positive value.
2. Thus solution is not optimal.
3. X2 is the optimum column and becomes the incoming value.
4. S2 is replacing row since the ratio for S2 is minimum.
5. The key element is 3/2

Step – 5: Iteration 2 Based on the rule explained in step 4 rewrite the new table.

C	**Variables in Basis**	**Solution Values**	**4**	**3**	**0**	**0**
	B	X_B	X1	X2	S1	S2
4	X1	32	1	0	$\frac{2}{3}$	$\frac{-1}{3}$
3	X2	8	0	1	$\frac{-1}{3}$	$\frac{2}{3}$
	Z	4x32+3x8=152	4	3	$\frac{5}{3}$	$\frac{2}{3}$
	C-Z		0	0	$\frac{-5}{3}$	$\frac{-2}{3}$

Table – 1.8

Multiply first row with 4 and add 2nd row multiplying with 3

From the above table we find the following:

1. The C-Z row has no positive values.
2. The solution is optimal with X1 = 32 and X2 = 8
3. Maximum Profit Z=4X1 + 3X2 = (4 x 32) + (3 x 8) = Rs.152.
4. Constraints 2X1+X2 = (2x32) + 8 = 72 and

   X1 + 2X2 =32 + (2x8) = 48

Problem 1.7.2:

Solve the following LPP using simplex method.

Minimize Z = X1-3X2+2X3

Subjected to 3X1 – X2 + 3X3<=7

-2X1 + 4X2 <=12

-4X1 + 3X2 + 5X3 <=10

X1, X2, X3 >=0

Solution:

This is the problem of minimization convert the objective function from minimization to maximization by multiplying the minimizing problem by – 1.

Step – 1: Rewrite the problem as,

Maximize - Z = X1+3X2-2X3=Max Z'

Where – Z = Z'

Subjected to 3X1 – X2 + 3X3<=7

-2X1 + 4X2 <=12

-4X1 + 3X2 + 5X3 < = 10

Step – 2: Converting inequality to equality form

Minimize Z = X1+ 3X2 - 2X3 + 0S1 + 0S2 + 0S3

Subjected to 3X1 - X2 + 3X3 + S1 = 7

-2X1 + 4X2 + S2 = 12

-4X1 + 3X2 + 8X3 + S3 = 10

Let X1, X2, X3 be assigned to zero then we have S1=7, S2=12, S3=10

Step – 3 Initial table

C	**Variables in Basis**	**Solution Values**	**-1**	**3**	**-2**	**0**	**0**	**0**	**Ratio**
	B	X_B	X1	X2	X3	S1	S2	S3	
0	S1	7	3	-1	3	1	0	0	$\frac{7}{-1} = -7$
0	S2	12	-2	4	0	0	1	0	$\frac{12}{4} = 3$ →
0	S3	10	-4	3	8	0	0	1	$\frac{10}{3} = 3.33$
0	Z'	0	0	0	0	0	0	0	
		C-Z	-1	3 ↑	-2	0	0	0	

Table - 1.9

1. (C-Z') row has a positive value
2. Thus solution is not optimal
3. X2 is optimal column and becomes the incoming value
4. S2 is the replacing row since the ratio for S2 minimum (neglect sign)
5. The key element is 4
6. X2 replace S2

Step – 4: Iteration 1

C	Variables in Basis	Solution Values	-1	3	-2	0	0	0	Ratio
	B	X_B	X1	X2	X3	S1	S2	S3	
0	S1	10	$\boxed{\frac{5}{2}}$	0	3	1	$\frac{1}{4}$	0	4 →
3	X2	3	$\frac{-1}{2}$	1	0	0	$\frac{3}{4}$	0	-6
0	S3	1	$\frac{-5}{2}$	0	8	0	$\frac{-3}{4}$	1	-25
	Z'	9	-3	$\frac{3}{2}$	0	0	$\frac{3}{4}$	0	
		C-Z	$\frac{1}{2}$ ↑	0	-2	0	$\frac{-3}{4}$	0	

Table -1.10

1. (C-Z') row has a positive value
2. Thus solution is not optimal
3. X1 is the optimum column and becomes the incoming value
4. S1 is the replacing row since the ratio for S1 is minimum
5. The key element is 5/2
6. X1 replace S1

Step – 5: Iteration 2

C	Variables in Basis	Solution Values	-1	3	-2	0	0	0
	B	X_B	X1	X2	X3	S1	S2	S3
-1	X1	4	1	0	$\frac{6}{5}$	$\frac{2}{5}$	$\frac{1}{10}$	0
3	X2	5	0	1	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{3}{10}$	0
0	S3	11	0	0	11	1	$\frac{-1}{2}$	1
	Z'	11	-1	3	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{8}{10}$	0
		C-Z'	0	0	$\frac{-13}{5}$	$\frac{-1}{5}$	$\frac{-8}{10}$	0

Table – 1.11

From the above table we find that,

1. (C – Z') row does not have any positive value
2. Solution is optimum
3. Solution is X1 = 4, X2 = 5 X3 = 0
4. Max Value of Z' =11 Minimum value of Z=-11

1.8 KEY WORDS

Linear Programming : A mathematical technique to help management decide how to make most effective use of an organization's scarce resources.

Objective Function : **A** mathematical statement of the goal of an organization, stated as intent to maximize or to minimize some important quantity such as profits or costs.

Optimal Solution : A point in the feasible region that maximize the objective function.

Feasible Region : It is the collection of feasible solution-the area showing all possible production combination in a linear programming problem.

Simples Method : An algorithm for solving LP problems investigates feasible corner points only, always maintaining or improving the objective function, until an optimal solution is obtained.

Basic Variable : The set of variables that are in the solution and are listed in the product mix column.

Basis : The set of variables that comprises a basic solution to an LP problem.

(C-Z) Row : The row containing variable net profit or loss that will result by bringing one unit of a variable indicated in that column into the solution of LP problem.

1.9 SUMMARY

In this chapter we have introduced linear programming as a powerful technique of Operations Research designed to solve allocation problems. It helps in choosing the best combinations of various activities to meet various criteria in such a way that the solution obtained is optimal. LP may be applicable in case of problems which,

1. Are deterministic in nature
2. Involves multiple variables which are interrelated.
3. Has a set of criteria which must be met and
4. Have a single objective.

All LP problems can be formulated in a common format. The objective function is always to maximize or minimize a linear summation of the variables subjected to a given set of constraints. In this section we have discussed the formulation of a linear problem model, solving the linear problem using a graphical method and using the simplex method to solve the linear problems.

1.10 SELF ASSESSMENT QUESTIONS

Case – 1 : Sam agency wishes to plan an advertising campaign in three different media via television, radio and magazines. The purpose of the advertising programme is to reach as many potential customers as possible. Results of a pilot study is as given below:

Particulars	Television		Radio	Magazines
	Daytime	**Prime Time**		
Cost of an advertising unit	Rs.40,000	Rs.75,000	Rs30,000	Rs.15,000
Number of customer reached/unit	4,00,000	9,00,000	5,00,000	2,00,000
No. of women customer reached/unit	3,00,000	4,00,000	2,00,000	1,00,000

The company does not want to spend more than Rs.8,00,000 on advertising. It further requires that,

1. At least 2 million exposures take place among women.
2. Advertising on television be limited to Rs.5,00,000
3. At least 3 advertising units be brought on daytime television, and two units during prime time.
4. The number of advertising units on radio and magazine should each be between 5 and 10.

Formulate the above problem into a linear programming model.

Case – 2 : Shivin limited is drawing up production plan coming year. Four products are available with the following financial characteristics.

Product	A	B	C	D
Selling Price	Rs.55	Rs.53	Rs.97	Rs.86
Cost of Material	Rs.17	Rs.25	Rs.19	Rs.11
Labour Hours				
Grade A	10	6	-	-
Grade B	-	-	10	20
Grade C	-	-	12	6
Variable Overheds	Rs.6	Rs.7	Rs.5	Rs.6

Fixed overheads of the firm amount to Rs.35,000 per annum. Each grade of labour is paid Rs.1.50 per hour but skill are specific to a grade so that an employee in work grade cannot be used to undertake the work of another grade. The annual supply of each grade is limited to the following maximum: Grade A 9,000 hours; Grade B 14,500 hours; Grade C 12,000 hours.

1. Calculate the product mix which will maximize profits.
2. Solve the problem using both Graphical and Simplex method to obtain the optimal solution.

1.11 REFERENCES

1. Resource Management - N.G. Nair; Vikas Publication.
2. Operations Research - V.K. Kapoor; Sultan Chand Publication
3. Operations Research - J.K. Sharma - Macmillan Publication
4. Operations Research - S. D. Sharma
5. Problems in Operations Research - P.K. Gupta and D.S. Hira

UNIT -2: INTRODUCTION TO TRANSPORTATION PROBLEMS

STRUCTURE

2.0 Objectives

2.1 Introduction

2.2 Transportation Problems

- 2.2.1 Example
- 2.2.2 Features of Transportation Problem
- 2.2.3 Application of Network Optimization

2.3 Mathematical Formulation

2.4 Feasible Solution

- 2.4.1 Basic feasible solution
- 2.4.2 Initial Feasible solution
- 2.4.3 Optimum Feasible Solution

2.5 Solution to transportation problem

- 2.5.1 North West Corner Method
- 2.5.2 Least cost Method
- 2.5.3 Vogel's approximation method

2.6 Summary

2.7 Self Assessment Questions

2.8 References

2.9 Case Study

2.0 OBJECTIVES

After studying this unit you should be able to :

- Formulate Transportation Problem
- Define initial, basic and optimum Feasible solutions
- Derive an initial Basic Feasible Solution for any transportation problem using North West, least cost and VAM method.

2.1 INTRODUCTION

The transportation problems deal with allocations. A company may have 'm' number of origins or the initial locations where the items are stored and 'n' warehouses or destinations where they have to be distributed. Then a question arises that from which origin one has to transport items to which location. Obviously our aim is to reduce the transportation cost. The solutions to such type of problems are given by these transportation problems. There are different methods available to solve these type of problems. The transportation or distribution problems is an early example of network linear optimization and it is now a standard application and industrial firms having several manufacturing plants, ware houses, sales territories & distribution out lets.

2.2 TRANSPORTATION PROBLEM

The problem in transportation is to transport various amounts of a product that are initially stored at different places to different destinations (End Places) in such a way that the total transportation cost is minimum.

Basically transportation problems deal with minimization of transportation costs. If there are more than one place from where the material has to be sent and these have to be sent to more than one place then the transportation problem occurs. These problems can be mathematically formulated provided all the costs of transportation and quantity to be transported is given.

2.2.1 Example

A soft drink manufacturing firm has m plants located in n different cities. The total production is absorbed by a retail shops in a different cities. We want to determine the transportation schedule that minimizes the total cost of transporting soft drinks from various plants to various retail shops. First we will formulate this as a linear programming problem.

Formulation of the problem

Let us consider the m-plant locations (origins) as $O_1, O_2, \ldots \ldots \ldots O_m$ and the n-retail shops (destination) as $D_1, D_2, \ldots \ldots \ldots D_n$ respectively. Let $a_1 \geq 0$, $i = 1,2, \ldots \ldots m$, be the amount available at the i^{th} plant O_i. Let the amount required at the j^{th} shop D_j be $b_j \geq 0$, $j = 1,2, \ldots \ldots n$.

Let the cost of transporting one unit of soft drink from ith origin to jth destination be $C_{ij,}$ $i=1,2, \ldots \ldots m$, $j=1,2, \ldots \ldots n$. If $X_y \geq 0$ be the amount of soft drink to be transported from ith origin to jth destination, the the problem is to determine X_y

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} X_y C_y$$

so as to Minimize

Subject to the constraint

C_y and $X_y \geq 0$, for all i and j.

$$\sum_{i=1}^{m} X_{ij} = b_j, j = 1,2, \ldots \ldots n.$$

$$\sum_{j=1}^{n} X_{ij} = a_i, j = 1,2, \ldots \ldots m.$$

This LPP (Linear programming problem) is called a Transportation Problem.

2.2.3 Application of Network Optimization

Applications of Network Optimization

Applications	Physical analog of nodes	Physical analog of areas	Flow
Communication Systems	Phone exchanges computers, transmission facilities, satellites	Cables, fiber optic links, microwave relay links	Voice messages, Data. Video transmissions
Hydraulic systems	Pumping stations Reservoirs, Lakes	Pipelines	Water, Gas, Oil, Hydraulic fluids
Integrated computer circuits	Gates, registers, processors	Wires	Electrical current
Mechanical systems	Joints	Rods, Beams, Springs	Heat, Energy
Transportation systems	Intersections, Airports, Rail yards	Highways, Airline routes Railbeds	Passengers, freight, vehicles, oprators

2.3 MATHEMATICAL FORMULATION

Ware House → Plants ↓	1	2	J	N	Qty produced
1	C_{11}	C_{12}	C_{1j}	C_{1n}	a_1
2	C_{21}	C_{22}	C_{2j}	C_{2n}	a_2
i	C_{i1}	C_{i2}	C_{ij}	C_{in}	a_i
m	C_{m1}	C_{m2}	C_{mj}	C_{mn}	a_n
Qty to be sent	b_1	b_2	b_j	b_n	

Let there be m origins and n destinations, Let a1be the Quantity products produced in factory 1 similarly a_2, a_3 . . . . , a_i . . . . a_n

Let b_1 be the quantity to be sent to ware house 1. Similarly b_2, b_3. . . . . .b_j. . . . . b_n.

Let Cij be the cost of sending the products from factory i to ware house j & Xij be the quantity to be shipped from factory i to destinaton j.

Answer is may be given as

Ware House	1	2	J	n	Qty products
1	X_{11}				a_1
2		X_{22}			a_2
i				X_{in}	a_i
m		X_{m2}			a_m
Qty to be sent	b_1	b_2	b_j	b_n	

It is assumed that the total quantity produced is equal to toal products to be sent to different destinations.

i.e.,

$$\Sigma_{ai} = \Sigma_{aj}$$

Note : However in some cases these two may not match. Such problems are called as unbalanced transportation to problem which we will come across later.

The transportation problems demands calculation of all X_{ij} (Qty to be sent from each location to each destination) so that

$$\sum_{j=1}^{a} X_{ij} = a_i \text{; for } i = 1,2,. . . . . m \text{ \&}$$

(i.e, $X_{11} + X_{12} + X_{13}$ . . . . $= a_1$)

$$\sum_{j=1}^{m} X_{ij} = b_j \text{; for } j = 1,2,. . . . . n$$

(i.e, $X_{11} + X_{21} + X_{31}$ . . . . $= b_1$)

Minimizing the total cost of transportation

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} C_y$$

2.4 FEASIBLE SOLUTION

A Feasible solution refers to a set of non negative individual allocation $(X_y > 0)$ which simultaneously remove deficiencies. i.e. the solution obtained aftrer allotments are done subjected to supply & demand constraints.

2.4.1 Basic feasible solution

A feasible solution to a m -origin, n destination problem in said to be basic if the No. of positive allocation is m + n-1.

Supply Constraint: The sum of allotments in a row should be equal to that plant capacity.

Demand Constraint: The sum of allotments in a column should be equal to that warehouse requirement.

Suppose take a problem

	X	Y	Z	Capacity
A	8	7	3	60
B	3	8	9	70
C	11	3	5	80
Requirement	50	80	80	

Say the answer to this problem is

		60
50		20
	80	

Then No. of allocation is 4,

Where as m + n = 1 = 3 + 3-1=5

Then this is a problem of degeneracy

2.4.2 Initial Solution

An initial solution refers to the solution obtained using any of the three methods that will be discussed. A initial solution is obtained in the beginning and it is checked for optimality and improved further.

2.4.3 Optimum Feasible Solution

A feasible solution is said to be optimal if it minimize the total transportation cost.

Note: An optimum solution may or may not be Basic feasible solution.

2.5 SOLUTION TO TRANSPORT PROBLEMS

There are 3 methods to determine the initial feasible solution.

1. North West corner rule
2. Least cost method
3. Vogel's appoximation method

North West Corner Method : In this method, allocation are does orbiarly (without considering the cost factor) so that either row maxima or Column maxima constraint in fulfilled. Start from the North West corner; then take next row / column.

Step (1): The first assignment is made in the cell occupying the upper left-hand (North West) corner of the transportation table. The maximum feasible amount is allocated there, i.e.; $X_{11} = \min (a_1, b_1)$.

Step (2): If $b_1 > a_1$, the capacity of origin O_1 is exhausted but the requirement at destination. D_1 is not satisfied. So move down to the second row, and make the second allocation:

$X_{21} = \min (a_2, b_1 - x_{11})$ in the cell (2,1)

If $a_2 > b_1$, allocate $x_{12} = \min (a_2 - x_{11}, b_2)$ in the cell (1,2)

Continue this until all the requirements and supplies are satisfied.

Least Cost Method : In this method allocation is done initial to that cell where the total cost is minimum, Then the next minimum cost cell is considered. The Northwest Corner Method does not utilize shipping costs. It can yield an initial BFS easily but the total shipping cost may be very high. The minimum cost method uses shipping costs in order come up with a BFS that has a lower cost. To begin the minimum cost method, first we find the decision variable with the smallest shipping cost (Cij). Then assign Xij its largest possible value, which is the minimum of a_i and b_j. After that, as in the Northwest Corner Method we should cross out row i and column j and reduce the supply or demand of the noncrossed-out row or column by the value of Xij. Then we will choose the cell with the minimum cost of shipping from the cells that do not lie in a crossed-out row or column and we will repeat the procedure.

Vogel Approximation Method: This method gives a solution which is quite close to the optimal solution. In this method penalty i.e. the difference between least and next least cost are considered for allotment.

Step 1: For each row of the transportation table, identify the smallest and the next to-smallest costs. Determine the difference between them for each row. Display them alongside the transportation table against the respective rows. Similarly compute the differences for each column.

Step 2: Identify the row or column with the largest difference among all the rows and columns. If a tie occurs, use any arbitrary tie breaking choice. Let the greatest difference correspond to i^{th} row and the minimum cost be C_{ij}. Allocate a maximum feasible amount.

$X_{ij} = \min(a_i, b_j)$ in the $(i,j)^{th}$ cell, and cross off the i^{th} row or j^{th} column.

Step 3: Re compute the column and row differences for the reduced transportation table and go to step 2. Repeat the procedure until all the rim requirements are satisfied.

Remark. VAM determines an initial basic feasible solution, which is very close to the optimum solution.

More & clear knowledge of these methods can be obtained using the illustrations.

Problems:

1. Find out an initial feasible solution for the below problem

Ware House	W_1	W_2	W_3	W_4	Factory
F_1	19	30	50	10	7
F_2	70	30	40	60	9
F_3	40	8	70	20	18
	5	8	7	14	34

Step 1 : Check whether

$\sum_{ai} = \sum_{bj}$ 5 + 8 + 7 + 14 = 34

7 + 9 + 18 = 34

Apply any method to solve problem

First method : North West corner rule

	W1	W2	W3	W4	Available
F1					7
F2					9
F3					18
Requirements	5	8	7	14	

Step 3: Insert allocation in each cell such that total in each row and column equates the values shown against each row and column. Start from top-left cell. If row value is more than column value enter the column value and move towards next cell in same row. If row value is less than column value enter the row value and move towards next cell in same column.

i.e., First allot 5 in top left column (X_{11} =5), remaining 2 in next cell (X_{12}= 2). Then move down allot 6. (X_{22}=6), then go right allot 3 (X_{23}=3), go down allot 4 (X_{33}=4) Then again go right allot 14 (X_{34}=4)

Step 3: Write down respective cost in same cell in brackets.

5(19)	2(30)			7
	6(30)	3(40)		9
		4(70)	14(20)	18
5	8	7	14	

Step 4: Calculate total transportation cost

5 X 19 + 2 X 30 + 6 X 30 + 3 X 40 + 4 X70 + 14 X 20 = 1015

95 + 60 + 180 + 120 + 280 + 280 = 1015

The total transportation cost is Rs. 1015.

Second Method : Least Cost Method

Second Method : Least Cost Method

Let us consider the

	W_1	W_2	W_3	W_4	Capacity
F1	19	30	50	10	7
F2	70	30	40	60	9
F3	40	8	70	20	18
Ware house requirement	5	8	7	14	34

Step -1 : Check whether total ware house requirement is equal to

Total factory capacity

$\Sigma_{ai} = 34$ & $\Sigma_{bj} = 34$

Step -2 : Apply least cost method,. Allot in those cells where transportation cost is minimum. i.e., take C_{32} (8). Allot as much as possible to that cell ($X_{32} = 8$). Then C_{14} is minimum (10), Allot as much as possible (X_{14}=7) Next min is 19 (C_{11}). But allocation is not possible as row maximum is satisfied. Next minimum is C_{34} (20). So allot as much as possible ($X_{10} = 7$) as column maximum is satisfied.

19	30	50	(7) 10	7
(2) 70	30	(7) 40	60	9
(3) 40	(8) 8	70	(7) 20	18
5	8	7	14	

Next is 40 (C_{31}, or C_{23}), Take any one say C_{31}. Allot as much as possible ($X_{31} = 3$) has already 15 has been allotted in that row. Lastly allot $X_{23} = 7$

Note : This method is quite logical as it emphasizes on cost factor unlike North west corner rule.

Transportation Cost is
7 X 10 + 2 X 70 + 7 X 40 + 3 X 40 + 8 X 8 + 7 X 20
70 +140 + 280 + 120 + 64 + 120 = 814

You can note the difference between the transportation costs of next two are 1015-814 = 201 Rs.

Never the less it does not takes into consideration the penalty costs. i.e. if the allocation done to a cell where the cost is minimum. It would block all the remaining cells in that row or column. Invariably we may have to allocate to a cell (finally) where transportation is maximum. In the previous example. Final allocation was made to a cell where T.C. was 70 Rs.

III. VAM method or Vogel's approximation method

This method is also known as unit cost penalty method.

Take the same example

	W1	W2	W3	W4	Capacity	Penalty Cost		
F1	5	30	50	10	7	9	9	40
F2	70	30	40	60	9	10	20	20
F3	40	8 (8)	70	20	18	12	20	20
Warehouse	5	8	7	14				
Penalty Cost	21	22	10	10				

Step 1 : First find out the difference between lowest cell & next lowest cell & next lowest cost in each row & column. This is called penalty cost (i.e. if you miss the lowest cell then have to pay that much extra)

Step 2 : Mark the value with highest penalty (↑) (In this case 22)

Step 3 : Allot as much as possible in that row / column and in the cell having min. transportation cost.

Step 4 : Strike of that row / column of which maxima in satisfied. Now recalculate column/ row penalties (in this case the column maxima is fulfilled (8) so strike out that column) After recalculating the penalties you can see 21 is maximum, In that column C11 is minimum (19). So allot as much as possible (5). Strike out that column & recalculate penalties now 50 is maximum.

Step 5 : Repeat this process until all are allocated

Once you strike out a row or column then no cell from that row or column can be chosen for further entries or penalty calculations.

19X5 + 10X2 + 40X7 + 60X2 + 8X8 + 20X10

95 + 20 + 280 +120 + 64 + 200 =779

19 5	30	50	10 2	7	9	9	40
70	30	40 7	60 2	9	10	20	20
40	8 8	70	20 10	18	12	20	50
5	8	7	14				
↑21	↑ 22	10	10				
		10	50 ↑				

The total transportation cost is Rs. 779/-

2.6 SUMMARY

Here the strategic decisions involve selecting transportation routes so at to allocate the production of various plants to several ware houses or terminal points. Thus there are 3 methods which helps one in obtaining an initial feasible solution however this may not be optimal. Optimally test is done on this to generate a optimum feasible solution. North West corner method is an arbitary method Least cost method is trying to reduce the cost. VAM is more logical which tries to give a solution close to optimal solution.

2.7 SELF ASSESSMENT QUESTIONS

1. Explain with an example North West corner rule least cost method & VAM to obtain an IBFS of a transportation problem.
2. Determine the IBFS to be following problems using North West corner rule.

	D1	D2	D3	D4	Supply
O1	6	4	1	5	14
O2	8	9	2	7	16
O3	4	3	6	2	5
	6	10	15	4	35

(128)

	I	II	III	IV	
A	13	11	15	20	2000
B	17	14	12	13	6000
C	18	18	15	12	7000
	3000	3000	4000	5000	

(1,99,000)

3. Obtain an initial feasible solution using least cost method.

21	16	15	13	11
17	18	14	23	13
32	27	18	41	19
6	10	12	15	43

4. Obtain an inital basic feasible solution to the following transportation problem using Vogels approximation method.

	I	II	III	IV	
A	5	1	3	3	34
B	3	3	5	4	15
C	6	4	4	3	12
D	4	-1	4	2	19
	21	25	17	17	

5. Using VAM method calculate an initial feasible solution. If use N-W-C method low much does the transportation cost increases.

	D1	D2	D3	D4	
O1	1	2	1	4	30
O2	3	3	2	1	50
O3	4	2	5	9	20
	20	40	30	10	100

2.8 REFERENCES

1. OR -Sharma
2. OR -Wagner
3. OR -Kanti Swaroop
4. Operation Research Magazine

2.9 CASE STUDY

A CASE IN TRANSPORTATION

South India Soaps Ltd.,

The South India Soaps Ltd., (SISOL) operates three factories from which it shipped soaps to regional warehouses. In 2000, the demand for soaps was 24,000 tonnes which is distributed as follows;

Region	**Demand in 000 tonnes**
Coachin	3
Salem	4
Madhurai	11

One shift production capacity in each of the three factories was as follows:

Chennai	12
Coimbatore	7
Bangalore	7

Estimated transport costs (in hundred rupees per thousand tones) are given below:

Regional warehouses

Factory

	Cochin	Nellore	Salem	Madurai
Chennai	95	105	80	15
Combatore	115	180	40	30
Bangalore	155	180	95	70

SISOL folowed a policy of decentralization under which each of the four regionasl warehouses was under the direct supervision of a regional sales manager and he was resposible for the profitability of operation under his control.

over a period of time, this procedure led to increasing friction in the organization. There were questions costs; also there was no co-ordination. For instance, in 1999, the sales manager of Madurai and Nellore placed their orders with the Chennai Factory which did not have capacity to meet all demands. This led to insufficient and duplicate orders, friction etc., The final pattern to allocation that emerged in 1999 was as follows.

	Cochin	Nellore	Salem	Madurai
Chennai	0	1	2	9
Combatore	3	0	0	2
Bangalore	2	3	2	0

The General Manager of SISOL called meeting of the executives at the Central office. Some executives proposed that all orders should be routed through the central office which would determine the optimal programme. Others protested that his would seriously conflict with the firm's philosophy of decentralization.

You have been hired as a consultant by the General Manager. Prepare a minimum cost distribution schedule for SISOL. Compare this schedule with the present schedule (1999) Which is better ?

UNIT - 3: OPTIMAL SOLUTION

STRUCTURE

3.0 Objective

3.1 Introduction

3.2 Moving Towards Optimal solution

3.3 MODI Method

- 3.3.1 Test For Degeneracy
- 3.3.2 Determination of Net Evaluation
- 3.3.3 Transportation Algorithm
- 3.3.4 Problem

3.4 Stepping Stone Method

- 3.4.1 Problem

3.5 Unbalanced Transportation Problem

3.6 Degeneracy in Transportation Problem

3.7 Maximization Problem

3.8 Summary

3.9 Self Assessment Question

3.10 Reference

3.0 OBJECTIVE

After reading this unit you will be in a position to

* To perform optimality test
* Differentiate approaches to optimize the solution
* Variants in transportation problem such as unbalanced, degeneracy and maximization explain problems.

3.1 INTRODUCTION

After obtaining a basic feasible solution to a given transportation problem, the next step is to derive an optimal solution. i.e., test has to be done to check whether the obtained solution is optimal or not. If the solution is not optimal a new solution is derived through iterations.

3.2 MOVING TOWARDS OPTIMAL SOLUTIONS

An optimum solution refines the original solution in order to reduce the total cost. For this a test for optimality is done for IBFS. The basic techniques is

1. Determine the net evaluation for non basic variables i.e., un occupied cells
2. Choose those net calculations which may improve the current basic feasible solution.
3. Determine the current occupied cells which leaves the basic i.e., becomes un occupied) and repeating (1) through (3) until an optimum solution is obtained.

The optimum solution can be obtained through

1. MODI - (Modified Distributed Method)
2. Stepping Stone Method

3.3 MODI METHOD

Steps to obtain an optimal solution

Step-1 Examine the IFS for degeneracy

Step-2 Determine net evaluation for empty cells

Step-3 Test for optimality

Step-4 If the solution can be improved then select entering and exiting variable

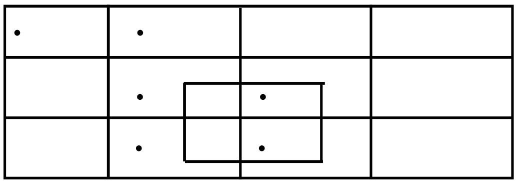
Step-5 Repeat the step 1 -4 till optimum solution is obtained.

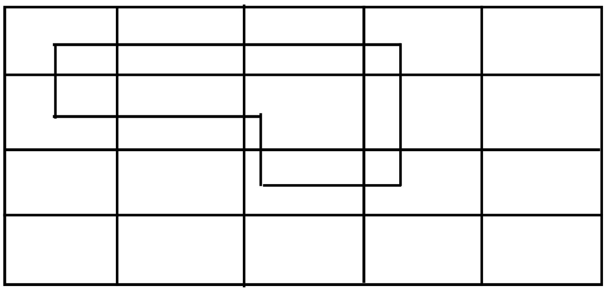
Let us discuss these steps in detail

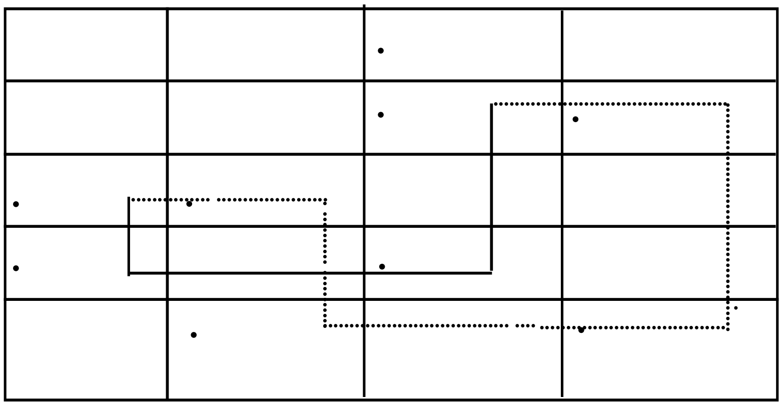
3.3.1 Test for degeneracy

As discussed earlier a basic solution should contain exactly m + n-1 number of individual allocation. These allocations must be independent positions.

Independent position means it should not be possible to get a closed loop. A loop may or not involve all location. But it should drawn using horizontal and vertical line through allocations. Take following examples.

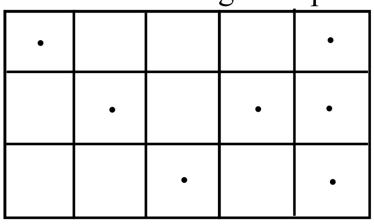






In all these examples the loops can be drawn

Take the following examples



Therefore loop can be made. So here allocations are in independent position.

3.3.2 Determining of net evaluation

Take an example of m origin - n destination transportation problem.

Determine X_{ij} so as to minimize

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} [C_{ij}]$$ Subjected to

$$\sum_{j=1}^{n} X_{ij} = a_1 \quad \text{or} \quad a_1 - \sum_{j=1}^{n} X_{ij} = 0 \quad \text{for } i = 1,2 \ldots\ldots.m$$

$$\sum_{i=1}^{m} X_{ij} = b_j \quad \text{or} \quad b_j - \sum_{i=1}^{m} X_{ij} = 0 \quad \text{for } i = 1,2 \ldots\ldots.m$$

& $X_{ij} \geq$ for all ij

3.3.3 Transportation Algorithm

Step-1 Construct transportation table

Step-2 Find IBFS by VAM or using other method

Step-3 Assume row variables $u_1, u_2 \dots\dots\dots u_n$ & column variable, $(V_1, V_2 \dots\dots\dots V_n)$ such that $u_1 + v_1 = C_{11}$, $u_1 + v_2 = C_{12}$ etc.,

Step-4 Assume a value to a row variable u_n (preferably) or having highest allocation (preferably). Assume the value to be zero (preferably)

Step-5 Find out corresponding column values. Similarly find out all the variables (Refer problem)

Step-6 Compute the cost difference $d_{ij} = C_{ij} - (u_1 - v_1)$ for all the non basic and enter them.

Step-7 Apply optimality test by examining the sign of each d_{ij}

1. If all $d_{ij} > 0$, the current basic feasible solution is an optimum one.
2. If at least one $d_{ij} < 0$ (negative) select the variable X_{rs} (having the most negative dij to enter the basic)

Step-8 Allocate an unknown quantity say θ . Then construct a loop that starts and ends at θ and connects some of the basic cell. The amount θ is added to and subtracted from cells of the loops in such a manner that total remains constant.

Step-9 Assign the largest possible value to θ in such a way that the value of at least one basic variable becomes zero & other basic remain non-negative (>0). The basic cell whose allocation has been made zero will leaves the basic.

Step-10 Repeat steps 3 - 9 till all d_{ij} are positive.

3.3.4 Problem

Consider the same problem for which solution is obtained by VAM

		W1	W2	W3	W4	
U1 = 10	F1	5(19)			2(10)	7
U2 = 40	F2			7(40)	2(60)	9
U3 =0	F3		6(8)		10(20)	15
		5	8	7	14	
		$V_1 = 29$	$V_2=8$	$V_3=0$	$V_4=20$	

Step-1 The IFS has 6 allocations where as m=3 & n=4 & m+n-1 = 6

So first Condition is satisfied.

Step-2 Assume $u_3=0$ (Since that row two allocations)

Then $V_2 + U_3 = 8 => V_2 = 8$

$V_4 + U_3 = 20 = . V_4 = 20$

Then $U_2 + V_4 = 60 => U_2 = 40$

$U_1 + V_4 = 10 => U_1 = 10$

Then $U_1 + V_1 = 19 => V_1 = 29$

$U_2 + V_3 = 40 => V_3 = 0$

Step-3

Now calculate dij for all non-basic cells. $\dfrac{(C_{ij}-(U_i + V_j))}{}$

C_{ij} - $(U_i + V_j)$

-	(30)	(50)	-
(70)	(30)	-	-
(40)	-	(70)	-

	29	8	0	20
10	-	-2	-10	-
40	69	48	-	-
0	29	-	0	-

For example : D_{12} = 30 - (-10+8) = 32

-	32	60	-
1	-18		-
11		70	-

Step-4

All are non-negative except X_{22}. That means solution is not optimum

Now take the original allocation. Include all the cells with highest negative d_{ij} value (in this case only one)

$5_{(19)}$			$2_{(10)}$
	$0 + \theta$	7(40)	$2 - \theta$
	$8 - \theta$		$10 + \theta$

Allocate θ unit to X_{22}. Then construct a loop. The cells included are X_{22}, X_{24}, X_{34}, X_{32}.
Deduct θ from adjacent cells & add θ to their adjacent.
Out of the member of loop deduct the minimum allocation associated with θ (In this case 2 i.e. X_{24})
Then θ =2 Now X_{24} leaves basic & X_{22} enters basic.

5(19)			2(10)
	2(30)	7(40)	
	6(8)		12(20)

Now the total cost is

7 X 40 + 5 X19 + 2 X 10 + 2 X 30 + 6 X 8 + 12 X 20 = 743

Step-5

Now again find out the value of U & V

U_1 = -10	$5_{(19)}$			$2_{(10)}$
U_2 = 22		$2_{(30)}$	$7_{(40)}$	
U_3 = 0		$6_{(8)}$		$10_{(20)}$
	V_1 = 29	V_2 = 8	V_3 = 18	V_4 = 20

Cij -------- Ui + Vj

	30	50	
70			60
40		70	

\-

	-2	8	
51			42
29		18	

Dij

	32	42	
19			18
11		52	

Since all d_{ij} are positive. This solution is optimum

3.4 STEPPING STONE METHOD

Step-1- Find IBFS of the transportation

Step-2- Check the No. of occupied cells. If there are less than m + n -1, there exists degeneracy & we introduce a very small quantity $\varepsilon (\varepsilon \to 0)$ in suitable independent position, so that the No. of occupied cells is exactly equal m + n - 1.

Step-3- Compute improvement index for each of the unoccupied cells. This computed by calculating the opportunity cost of an unoccupied cell. This means that if we shift one unit from a cell containing positive shipment to unoccupied cell that will be the net cost. If all the unoccupied cells have positive improvement index, then the given solution is optimum.

Step-4- If there are several unoccupied cells with negative improvement indices, then we select cell having the largest negative improvement index & shift maximum possible unit to that cell without violating the supply & demand constraint, after it again go to step 3.

3.4.1 Problem :

A company has factories at F_1, F_2 & F_3 which supply warehouses at W_1, W_2 & W_3. Weekly factory capacities are 200, 160 & 90 respectively. Weekly warehouse requirements are 180, 120 & 150 units respectively unit skipping costs (in Rs) are as follows.

Factory	Ware house			Supply
	W1	W2	W3	
F1	16	20	12	200
F2	14	8	18	160
F3	26	24	16	90
Requirement	180	120	150	350

Step 1 : Derive an Initial feasible solution using VAM method :

140 16	20 20	60 12	200	4	4
40 14	120 8	18	160	6	4
26	24	90 16	90	8	0
180	120	150	350		←
2	12 ↑	4			
2		6 ↑			

The total transportation cost is

140 X 16 +60 X 12 + 40 X 14 + 120 X 8 + 90 X 16

2240 + 720 + 560 + 960 + 1440 = 5920

Step 2 :

M + n - 1 = 3 + 3 - 1 = 5
The No. of occupied cells is equal to 5.
So this condition is satisfied & there is no degeneracy.

Step 3 :

Now consider unoccupied cells. Say (1,2). Now shift one unit from other cell (having highest cost) to this. Once it is shipped a loop has to be drawn & adjustments must be made correspondingly so as to not to change supply & demand constructs.

	W1	W2
F1	-θ 140 16	+θ 20
F2	40 +θ 14	120 -θ 8

Change in unit total cost
= 20 X 1 + 16 X (-1) + 14 X 1 + 8 X (-1)
= 20 - 16 + 14 - 8 = 10
This indicates that shift of one unit into (1,2) increases the transportation cost by Rs. 10/-. So this can not be included.
Now take (2, 3)

	W1	W2
F1	140 + θ 16	60 - θ 12
F2	40 -θ 14	+θ 18

Unit total cost change will be

16 x 1 + 12 x (-1) + 14 (-1) + 18 x 1

= 16 -12-14+18 = 8

As this also has positive opportunity cost this can not be taken.

Now consider cell (3,1)

	W1	W2
F1	140 - θ 16	60 + θ 12
F2	+ θ 26	90-θ 16

Change in unit total cost

16 x (-1) + 12 x 1 + 26 x 1 + 16 x (-1)

= - 16 + 12 + 26 - 16

= 6

Now taking the last one i.e., (3,2)

	W1	W2	W3
F1	140 - θ 16	20	60 + θ 12
F2	40 + θ 14	120 - θ 8	18
F3	26	θ 24	90 - θ 16

Now compute total cost

16 x (-1) + 12 (1) + 14 x 1 + 8 x (1) + 24 x 1 + 16 (-1)

-16 + 12 + 14 - 8 + 24 - 16 = 10

Since all unit total costs are positive the original solution is optimum.

3.5 UNBALANCED TRANSPORTATION PROBLEM

In this types of problem the supply & demand do not match. In such cases a dummy allocation is made having zero transportation cost.

Problem :

A company manufacturing air cooler has 2 plants with a weekly capacity of 200 & 400 units. It supplies air coolers to its 4 show rooms have a demand 75,100, 100 & 30 units. The cost of transportation has given below.

					Production
F1	90	90	100	100	200
F2	50	70	130	85	100
	75	100	100	30	

Total production is = 300
Total ware house requirements is = 305
Here production does not match with ware house requirements. Hence this is an unbalanced transportation problem.
So add a dummy plant having production capacity of (305- 300) 5 units having zero unit transportation cost.

	W1	W2	W3	W4	
F1	90	90	100	100	200
F2	50	70	130	85	100
F3	0	0	0	0	5
	75	100	100	30	

Find out initial Basic Feasible solution using VAM

	W1	W2	W3		W4	Capacity	Penalty
F1	90	75 90	95	 100	30 100	200	0 10
F2	75 50	25 70		 130	 85	100	20 15
F3	 0	 0	5	 0	 0	5	0
	75	100	100		30		
	50	70	100		85		
	40	20	30		15		

The initial transportation cost is

75 X 90 + 95 X 100 + 30 X 100 + 75 X 50 + 25 X 70 + 5 X 0 = 24375

Now find out optimal solution using MODI Method

Check for degeneracy

M + n - 1 = 3 + 4 - 1 = 6

No. of allocation = 6

So there no degeneracy

	V_1	V_2	V_3	V_4	
U_1	90	75 / 90	95 / 100	30 / 100	100
U_2	75 / 50	25 / 70	130	85	80
U_3	0	0	5 / 0	0	0
	-30	-10	0	0	

Put $U_3 = 0$ Then $V_3 = 0$

$U_1 + V_3 = 100 \Rightarrow U_1 = 100$

$U_1 + V_2 = 90 \Rightarrow V_2 = - 10$

$U_1 + V_4 = 100 \Rightarrow V_4 = 0$

$V_2 + U_2 = 70 \Rightarrow U_2 = 80$

$U_2 + V_1 = 50 \Rightarrow V_1 = - 30$

Now calcualte Z_{ij} for remaining cells

$C_{(1,1)} = 100 - 30 = 70$

$C_{(2,3)} = 80 + 0 = 80$

$C_{(2,4)} = 0 + 80 = 80$

$C_{(3,1)} = - 30 + 0 = - 30$

$C_{(3,2)} = -30 + 0 = -10$

$C_{(3,4)} = 0 + 0 = 0$

Now calcualting d_{ij}

$C_{11} = 100 - 70 = 30$

$C_{23} = 130 - 80 = 50$

$C_{24} = 85 - 80 = 5$

$C_{31} = 0 + 30 = 30$

$C_{32} = 0 + 10 = 10$

$C_{34} = 0 + 0 = 0$

Since all d_{ij} are non-negative the solution is optimum

3.6 DEGENERACY TRANSPORTATION PROBLEM

This type of problem occurs when No. of allocated cells are less than the No. of rows & columns. Thus it is not possible for us to form a loop for checking for optimality. In such cases a dummy allocation of (ϵ) is allotted to any column after which checking for optimality can be done.

Consider the following transportation problem

Problem

Solve the following transportation problem

	1	2	3	4	5	6	
A	7	5	7	7	5	3	60
B	9	11	6	11	-	5	20
C	11	10	6	2	2	8	90
D	9	10	9	6	9	12	50
Demand	60	20	40	20	40	40	

Also it is not possible to transport any item from factory B to go down 5.

Sol : Let us put the cost as ∞ (infinity for the cell (2,5)

$\sum a_i = 60 + 20 + 90 + 50 = 220$

$\sum b_j = 60 + 20 + 40 + 20 + 40 + 40 = 220$

Find out an initial basic feasible solution using VAM

7	20 5	7	7	5	40 3	60	2	2	2	4	0	
10 9	11	10 6	11		5	20	1	1	1	1	1	1
11	10	30 6	20 2	40 2	8	90	0	0	4	2	2	5
50 9	10	9	6	9	12	50	3	3	0	0	0	0
60	20	40	20	40	40							
2	5	0	4	3	2							
0		0										
0		3										

Since the No. of occupied cells is 8 i.e., less than m + n = 1. There is degeneracy in the initial solution.

To overcome degeneracy, we allocated a small quantity ϵ ($\epsilon \rightarrow 0$) in the cell (1,5) being the un occupied cell having the lowest transportation cost and find out U & V values.

	V_1	V_2	V_3	V_4	V_5	V_6	
U_1	 7	20 5	 7	 7	ε 5	40 3	0
U_2	10 9	 11	10 6	 11	 ∝	 5	-3
U_3	 11	 10	30 6	20 2	40 2	 8	-3
U_4	50 9	 10	 9	 6	 9	 12	-3
	60	20	40	20	40	40	
Vj	12	5	9	5	5	3	

Then find out the dij values

12		9	5		
	2		2	2	0
15	8				0
	2	6	2	2	0

-5		-2	2		
	9		9	∝	5
4	-2				8
	8	3	4	7	0

Since he above table has negative value soultion is not optimal. So introducing

θ

Then find out the dij values

							Ui
	+θ 7	20 5	7	7	ε - θ 5	40 3	0
	10 -θ 9	11	10 +θ 6	11	∞	5	-3
	11	10	30 - 6	20 2	40 +θ 2	8	-3
	50 9	10	9	6	9	12	-3
Vj	60	20	40	20	40	40	
	12	5	9	5	5	3	

After computing you will find that Cell (1,1) has most negative value. So introduce Cell (1,1) & Drop cell (1,5) and again go for U & V values and find out Dij Value.

	V_1	V_2	V_3	V_4	V_5	V_6	
U_1	ε 7	20 5	7	7	5	40 3	0
U_2	10 9	11	10 6	11	∞	5	2
U_3	11	10	30 6	20 2	40 2	8	2
U_4	50 9	10	9	6	9	12	2
Vj	60	20	40	20	40	40	
	7	5	4	0	0	3	

Since all net evaluation are non negative

The solution is optimum

The min transportation cost is

20 X 5 + 40 X 3 + 10 X 9 + 10 X 6 + 30 X 6 + 20 X 2 + 40 X 2 + 50 X 9 = 1120

Calculation dij values

		4	0	0	
	7		2	2	5
4	7				5
	7	6	2	2	5

		3	7	5	
	4		9	9	5
2	3				3
	3	3	4	7	7

3.7 MAXIMIZATION PROBLEM

ABC Enterprises is having 3 plants location at different place having various production cost. Company has 5 sales office at different region of the country having varied sales rice the shipping cost are as below.

Plants No.	Production Cost/Unit	Max. Capacity in No. of units
1	20	150
2	22	200
3	18	125

Shipping costs

	A	B	C	D	E
1	1	1	5	9	4
2	9	7	8	3	6
3	4	5	3	2	7
Demand	80	100	75	45	125
Sales price	30	32	31	34	29

Solution :

Step 1 - Find out the profit matrix profit is calculated as below

Profit = Sales Price - Production cost - Transportation cost

Take cell (1,1)

Profit = 30 - 20 -1 = 9

Similarly calculate for all the cell.

	A	B	C	D	E	
1	9	11	6	5	5	150
2	-1	3	1	9	1	200
3	8	9	10	14	4	125
	80	100	75	45	125	475

Step 2 - Subtract all the element from the largest element in the matrix, Now this is the loss matrix

1	5	3	8	9	9
2	15	11	13	5	13
3	6	5	4	0	10

Now this is a usual problem of minimization which can be solved as usually using VAM method.

5 50	3 100	8	9	9	0	150	3 3 2 2 4 4
15 25	11	13	5	13 125	0 50	200	5 11 2 2 2
6 5	5	4 75	0 45	10	0	125	0 4 1 1 4
80	100	75	45	125	50		
1	2	4	5	1	0		

Optimal Solution

No. of allocation is 8

m + n - 1 = 6 + 3 - 1 = 8

So this problem is non degenerate. So use stepping stone method to find out optimal solution.

For Cell (1,3)

Consider cells (1,1) (3,1) (3,3)

8 X 1 + 5 X (-1) + 6 X 1 + 4 x -1

= 8 - 5 + 6 - 4 = 5

For Cell 1,4, Consider cells (1,1) (3,1) (3,4)

= 9 X 1 + 5 X -1 + 6 X 1 + 0 X -1 = 9 - 5 + 6 + 0 = 11

For Cell (1,5) consider cells (1,1) (2,1) (2,5)

= > 9 X 1 + 5 X -1 + 15 X 1 + 13 X 1 = 9-5 + 15 - 13 = 6

For cell (1,6) consider cells (1,1) (1,6) (2.1) & 2, 6

= > 0 X 1 + 5 X -1 + 15 X 1 + 0 X -1 =

0 - 5 + 15 + 0 = 10

For cell 2,2 consider (1,1) (1,2) (2,1)

= 5 X 1 + 3 X -1 + 15 X -1 + 11 X 1 = > 5 - 3 - 15 + 11 = - 2

For cell (2,3) consider (2,1) (3,1) (3,3)

= 15 X - 1 + 13 X 1 + 4 X - 1 + 6 X 1

= - 15 + 13 - 4 + 6 = 0

For cell (2,4)

Consider (2,1) (3,1) (3,4)

15 X - 1 + 5 X 1 + 0 X - 1 + 6 X 1

= - 15 + 5-0 + 6 = - 4

For cell (3,2) Consider (1,1) (1,2) (3,1)

5 X 1 +3 X (-1) + 6 X - 1 +5 X 1

= 5 - 3 - 6 + 5 = 1

For cell 3, 5 consider (2, 1) (2,5) (3,1)

15 X 1 + 13 X - 1 + 10 X 1 + 6 X - 1

=15 - 13 + 10 - 6 = 6

For cell 3,6 consider (2,1) (2,6) (3,1)

15 X 1 + 0 X -1 + 0 X 1 + 6 X -1

= 15 - 6 = 9

All are non negative except (2,2) (2,3) & (2,4)

Index of 2,3 is zero which means there is no loss no profit by shifting. Out of 2,2 & 2,4 (2,4) is having more negative index. So shifting from 2, 1 to 2, 4 as much as possible i.e., 25.

15	5
25 - 25	25 + θ
5 + θ	45 - θ
6	0

Now θ= 25 and now the allocation becomes

50 5	100 3	8	10 9	2 9	6 0
15	11	13	25 5	125 13	50 0
30 6	5	75 4	20 0	2 10	5 10

Now recalculate index

50 5	100 3	5 8	10 9	2 9	6 0
4 15	2 11	4 13	25 5	125 13	50 0
30 6	1 5	75 4	20 0	2 10	5 10

Since all are non negative the solution is optimum

Total profit

= 50 X 9 + 100 X 11 + 25 X 9 + 125 X 13 + 30 X 8

+ 75 X 10 + 20 X 14 = Rs. 3170

3.8 SUMMARY

The transportation is one of the subclasses of L.P.P. in which the objective is transport various quantity of a single homogeneous commodity that are initial stored at various origins, the solution is obtained in three steps.

1. An initial basic feasible solution is obtain using NWC or Least cost or VAM
2. Optimum testing is done.
3. Optimum solution is obtained through MODI or steeping stone method.

3.9 SELF ASSESSMENT QUESTIONS

1. Find out a optimum solution.

From	D_1	D_2	D_3	D_4	Available
O_1	10	0	20	11	15
O_2	1	7	9	20	25
O_3	12	14	16	18	5
Required	12	8	15	10	45

2. Find fluid manufacturing plants are located at S1, S2, S3 & S4. Which supply the hospitals located at H1, H2, H3, H4 & H5. Daily plant capacities are 80, 120, 125 & 75 respectively. Daily hospital requirements are 30, 70, 150, 50 & 50 units. Unit transportation cost is given below.

	H1	H2	H3	H4	H5
S1	6	8	15	17	9
S2	11	13	7	4	16
S3	13	15	8	6	11
S4	10	5	9	3	6

3.10 REFERENCE

1. Operation Research by S.D. Sharma

2. Operation Research by Wagner

3. Operation Research by Taha

4. Operation Research by Kanthi Swaroop

UNIT - 4: ASSIGNMENT

STRUCTURE

4.0 Objective

4.1 Introduction

4.2 Assignment

4.3 Mathematical formulation of the problem

4.4 Assignment : The Hungarian method

4.5 Assignment Problems

4.6 Assignment having un-assignable jobs

4.7 Maximization Problem

4.8 Summary

4.9 Self Assessment Questions

4.10 Reference

4.0 OBJECTIVE

After studying this unit, you will be in a position to

* Explain meaning of Assignment
* Reach optimal solution
* Solve maximization & special types of problems.

4.1 INTRODUCTION

The best person for the job is an apt description of what the assignment model seeks to accomplish. The situation can be illustrated by assignment of workers to job, where any worker may undertake any job albeit with varying degrees of skill. A job that happens to match a worker's skill costs less than in which operator is not as skillful. The objective of the model is to determine the optimum (least-cost) assignment of workers to jobs.

4.2 ASSIGNMENT

The assignment problem is a special case of transportation in which the objective is to assign a number of origins to equal number of destinations at a minimum. Cost or maximum profit.

The general assignment model with n workers, n jobs is represented in below table

	1	2	j	n	
1	C_{11}	C_{12}	C_{11}	C_{1n}	1
2	C_{21}	C_{22}	C_{2j}	C_{2n}	1
i	C_{i1}	C_{i2}	C_{ij}	C_{in}	1
n	C_{n1}	C_{n2}	C_{nj}	C_{nn}	1
	1	1	1	1	

4.3 MATHEMATICAL FORMULATION OF THE PROBLEM

Consider a problem of assignment of n resources to n activities so as to minimize over all cost.

Resource

C_{11}	C_{12}	. . . . . C_{12}.	C_{1n}
C_{21}	C_{22}	. . . . . C_{23}.	C_{2n}
C_{11}	C_{12}	. . . . . C_{ij}.	C_{in}
C_{n1}	C_{n2}	. . . . . C_{nj}.	C_{nn}

The cost matrix is similar to tranportation problem.

Let X_{ij} denote the assignment of i^{th} resource j^{th} activity such that $X_{ij} = 1$ if assignment is done or $X_{ij} = 0$ of assignment is not done.

Then the mathematical formulation of the assignment problem is Minimize.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \; X_{ij}$$

Subject to constraints

$$\sum_{i=1}^{n} X_{ij} = 1 \ \& \sum_{j=1}^{n} X_{ij} = 1, X_{ij} = 0 \text{ or } 1$$

for all i = 1, 2 . . . . . . .n & j = 1,2, ...n where C_{ij} indicates cost associated with cell (ij)

4.4 ASSIGNMENT : THE HUNGARAIN METHOD

Assignment the Hungarian Method : - The method of solving assignment problems is given by D. Konig which is as follows.

Step 1 : Determine the cost table (if it is not given directly)

Step 2 : Check if the number of rows are equal to number of columns. If the number of rows is more, add dummy columns having unit costs as zero. Similarly if number of columns is more then add dummy rows having the cell values as zero

Step 3 : Locate the smallest element in each row and deduct the same from all the elements of that row. Similarly locate the smallest element in each column & deduct the same from all the elements of that column.

Step 4 : Start allocations : Starting from I row mark assignment (☐) in the rows having only one zero on that row in the cell having value zero. Cross (X) other zeros in that column. If all the rows have one assignment each, then the solution is optimum. If not, follow the next step.

Step 5 : Draw the min. number of horizontal and vertical lines to cover all the zeros. Use the following method.

a. Mark (✓) row is that do not have any unsigned zero.
b. Mark (✓) columns that have zeros in the marked row.
c. Mark (✓) rows that have assigned zeros in marked columns.
d. Draw the line through all the unmarked rows & marked columns.

Step 6 : Find the smallest element of the reduced matrix not covered by any of the lines. Subtract this element from all the uncovered elements & add the same to all the elements lying at the intersection of any two lines.

Step 7 : Go to step 4 repeat the procedure until an optimum solution is attained.

4.5 ASSIGNMENT PROBLEMS

A department head has 4 takes to be performed and 4 subordinates who differ in efficiency. The time estimates are given below in the matrix. How should it allocate tasks one to each man to as to minimize the total man hours?

	MEN			
Tasks	1	2	3	4
I	9	26	15	14
II	13	26	6	20
III	35	20	15	25
IV	18	30	20	10

Step - 1

Check whether no. of columns are equal to No. of rows. In this case it is true. Subtract the smallest elements of each row from all elements of that row.

0	17	6	5
7	20	0	14
20	5	0	10
8	20	10	0

Subtract the smallest elements of each column from all elements of that column

0	12	6	5
7	15	0	14
20	0	0	10
8	15	10	0

Now start assignment start from rows I. Once all the rows are completed start from columns.

$\boxed{0}$	12	6	5
7	15	$\boxed{0}$	14
20	$\boxed{0}$	$\otimes$	10
8	15	10	$\boxed{0}$

Since all the rows have assignment, the solution is optimum. The assignments are

I $\rightarrow$ 1, II $\rightarrow$ 3, III $\rightarrow$ 2, IV $\rightarrow$ 4

The total man hours is equal to

9 + 6 + 20 + 10 = 45

2. A dept. head has 4 subordinates & 4 taks to be performed. The subordinates differ in efficiency & the tasks differ in their intrinsic difficulties. The estimate of the time each man would take to perform each task is given in the matrix below.

	MEN			
Tasks	E	F	G	H
A	18	26	17	11
B	13	28	14	26
C	38	19	18	15
D	19	26	24	10

Sol.

1. Check if the No. column are equal to rows in this problem, both are same.
2. Subtract the smallest element of each row from the corresponding elements of that row.

	E	F	G	H	
A	7	15	6	0	(subtract 11)
B	0	15	1	13	(subtract 13)
C	23	4	3	0	(subtract 15)
D	9	16	14	0	(subtract 10)

3. Subtract the smallest element of each column from the correponding elements of that column.

	E	F	G	H
A	7	11	5	0
B	0	11	0	13
C	23	0	2	0
D	9	12	13	0

4. Starting from row I enrectangle (i.e mark assignment) a single zero if any and an cross all over zero in that column. Similarly start from column 1 afterwards, till all zeros are marked.

	E	F	G	H
A	7	11	5	$\boxed{0}$
B	$\boxed{0}$	11	$\otimes$	13
C	23	$\boxed{0}$	2	$\otimes$
D	9	12	13	$\otimes$

Step 5 : Since row 4 has no assignment mark this row (✓). Since the forth column has zero in forth row, mark that column. Since the forth column has assignment in the first row, mark that row. Draw straight line row all un marked rows and marked column.

	E	F	G	H	
A	7	11	5	$\boxed{0}$	✓
~~B~~	~~$\boxed{0}$~~	~~11~~	~~$\otimes$~~	~~13~~	
~~C~~	~~23~~	~~0~~	~~2~~	~~$\otimes$~~	
D	9	12	13	$\otimes$	✓
				✓	

The smallest elements not covered by the line in 5. Subtract this element from all uncovered elements and add the same to all the elements lying at the intersection of the line we get new matrix.

	E	F	G	H
A	2	6	0	0
B	0	11	0	18
C	23	0	2	5
D	4	7	8	0

Step 7 : Repeat step 4

	E	F	G	H
A	2	6	$\boxed{0}$	$\otimes$
B	$\boxed{0}$	11	$\otimes$	18
C	23	$\boxed{0}$	2	5
D	4	7	8	$\boxed{0}$

Now all rows and columns have assignments so this solution is optimum.
A → G, B → E, C → F, D → H

The total man hour is

17 + 13 + 19 + 10 = 59

3. A company wishes to assign 3 jobs to different machines in such a way that each job is assigned to some machines. He cost of assigning jobs i to machine j is given by the matrix below.

Cost matrix

$$\begin{array}{cccc} 8 & 7 & 6 & 5 \\ 5 & 6 & 8 & 7 \\ 6 & 8 & 7 & 8 \end{array}$$

Step 1

Since the problem is unbalanced. Add a dummy row having all the values as zero.

$$\begin{array}{cccc} 8 & 7 & 6 & 5 \\ 5 & 6 & 8 & 7 \\ 6 & 8 & 7 & 8 \\ 0 & 0 & 0 & 0 \end{array}$$

Deduct smallest element from each row

$$\begin{array}{cccc} 3 & 2 & 1 & 0 \\ 0 & 1 & 3 & 2 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}$$

Since all rows and columns have zeros assignments can be started

$$\begin{array}{cccc} 3 & 2 & 1 & \boxed{0} \\ \boxed{0} & 1 & 3 & 2 \\ \otimes & 2 & 1 & 2 \\ \otimes & \boxed{0} & \otimes & \otimes \end{array}$$

Since row 3 has no assignment mark this row (✓) since the third row has zero in first column, mark that column. Since assignment of that column is in II row mark that row. Draw straight line row all un marked rows and marked column

$$\begin{array}{ccccc} 3 & 2 & 1 & 0 & \\ \otimes & 1 & 3 & 2 & \checkmark \\ 0 & 2 & 1 & 2 & \checkmark \\ 0 & 0 & 0 & 0 & \\ \checkmark & & & & \end{array}$$

Deduct one (1) from uncovered elements 1 add the same to intersection points.

4	2	1	0
0	0	2	1
0	1	0	1
1	0	0	0

Now again start assignments

4	2	1	$\boxed{0}$
$\boxed{0}$	⊗	2	2
⊗	1	$\boxed{0}$	2
1	$\boxed{0}$	⊗	⊗

Now all rows have assignments so the solution is optimum

I ➞ 4 II ➞ 1, III ➞ 3 IV ➞ 2

The total cost of assigning jobs is 5 + 5 + 7 = 17

4.6 ASSIGNMENT HAVING UNASSIGNABLE JOBS

There are 4 jobs A, B, C and D to be done by 4 operators. I - IV. Given the time needed by different operators for different jobs in the matrix below. Suppose if Job A should not be assigned to operator II, than how much extra time would be required.

JOBS	A	B	C	D
I	15	13	14	17
II	11	12	15	13
III	18	12	10	11
IV	15	17	14	14

No. of rows are equal to No. of columns

Sol : Deduct smallest element from each row and column

2	0	1	4
0	1	4	2
8	2	0	1
1	3	0	0

Since all rows and columns have zeros assignment can be started

2	$\boxed{0}$	1	4
$\boxed{0}$	1	4	2
8	2	$\boxed{0}$	1
1	3	⊗	$\boxed{0}$

Since all rows and columns have assignments the solution is optimum

Assignments are as follows :

I ➞ B II ➞ A, III ➞ C IV ➞ D

The total times taken in

13 + 11 + 10 + 14 = 48

Suppose job A can not be assigned to II Then instead of time specified, use infinity

15	13	14	17
∞	12	15	13
18	12	10	11
15	17	14	14

After deducting smallest elements from each row

2	0	1	4
∞	0	3	1
8	2	0	1
1	3	0	0

Now the assignment can be started as each row and column zeros

1	$\boxed{0}$	1	4
∞	$\otimes$	3	1
7	2	$\boxed{0}$	1
$\boxed{0}$	3	$\otimes$	$\otimes$

Now II row has no assignment, so marking has to be done.

1	$\boxed{0}$	1	4 ✓	
∞	$\otimes$	3	1	✓
7	2	$\boxed{0}$	1	
$\boxed{0}$	3	$\otimes$	$\otimes$	
	✓			

Now add smallest elements (1) to inter section and deduct it from uncovered elements

0	0	0	3
∞	1	2	0
7	3	0	1
0	4	0	0

Now again start assignment

$\otimes$	$\boxed{0}$	$\otimes$	3
∞	1	2	$\boxed{0}$
7	3	$\boxed{0}$	1
$\boxed{0}$	3	$\otimes$	$\otimes$

Now all rows and columns have assignment and hence the solution is optimum
The assignment are as below

I $\rightarrow$ B II $\rightarrow$ D III $\rightarrow$ C IV $\rightarrow$ A

The total times taken in

13 + 13 + 10 + 15 = 51

i.e., 3 min (51 - 48) extra time is required if job A should be not done by B

4.7 MAXIMIZATION PROBLEM

4. A manufacturing company has 3 zones A, B, C, & 3 sales engineers P, Q, R, respectively for assignment. Their performance score matrix is given below. Help them by determining the optimal assignment maximizes the total performance score.

1	A	B	C
$P_{1\,(P)}$	20	26	42
$P_{2(Q)}$	24	32	50
$P_{3(R)}$	32	34	44

First of all convert this problem to a minimization problem by deducting all the elements by the biggest element i.e., 50

30	24	8
26	18	0
18	16	6

Subtract all the rows from the least element of each row

22	16	0
26	18	0
12	10	0

Subtract the column from the smallest element of each column.

10	6	0
14	8	0
0	0	0

Now we can start assignment

10	6	$\boxed{0}$
14	8	⊗
$\boxed{0}$	⊗	⊗

Since II row is not having any assignment, mark the un assigned row, column having 0 in that row and row having assignment in that column.

10	6	$\boxed{0}$	✓
14	8	⊗	✓
$\boxed{0}$	⊗	⊗	
		✓	

Deduct smallest element i.e. 6. and restart assignment.

	P	Q	R
A	4	$\boxed{0}$	⊗
B	8	2	$\boxed{0}$
C	$\boxed{0}$	⊗	6

Since all rows & columns have assignment the solution is optimum

A $\rightarrow$ Q B $\rightarrow$ R C $\rightarrow$ P

The total performance score is 26 + 50 + 32 = 98

4.8 SUMMARY

The assignment model is actually a special case of transportation model in which the workers represent the sources and the jobs represent destination. Supply amount at each source and the demand amount at each destination exactly equal to 1, In effect, the assignment model can be solved directly as a regular transportation model. Never the less, the fact that all the supply and demand amounts equal 1 has led to the development of a simple solution. This algorithm is called the Hungerian method.

4.9 SELF ASSESSMENT QUESTIONS

A person has 4 children who have varied interest (measured on scale 1 - 1) in different jobs. It is summarized in the table below. Assuring no two can do the same job. Which job suit whom.

	1	2	3	4
1	1	4	6	3
2	9	7	10	9
3	4	5	11	7
4	8	7	8	5

2. Machineco has four jobs to be completed. Each machine must be assigned to complete one job. The time required to setup each machine for completing each job is shown in the table below. Machinco wants wants to minimize the total setup time needed to complete the four jobs.

	Time (Hours)			
	Job1	Job2	Job3	Job4
Machine 1	14	5	8	7
Machine 2	2	12	6	5
Machine 3	7	8	3	9
Machine 4	2	4	6	10

4.10 REFERENCE

1. Operation Research by S.D. Sharma
2. Operation Research by Kanthi Swaroop
3. Operation Research by Taha
4. www.wikipedia.com

UNIT -5: SEQUENCING PROBLEMS OF 2 & 3 MACHINES

Structure

5.0 Objectives

5.1 Introduction

5.2 Assumptions in Sequencing Problems

5.3 Sequencing of ‘N’ Jobs Through ‘2’ Machines

5.4 Sequencing of ‘N’ Jobs Through ‘3’ Machines

5.5 Summary

5.6 Self Assessment Questions

5.7 Reference

5.0 OBJECTIVES

After studying this module you will be able to;

- Give Meaning and Concept of Sequencing
- Define basic terminologies in Sequencing
- Identify assumptions made in Sequencing Problems and
- Design processing of jobs through Machines

5.1 INTRODUCTION

All organisations would like to utilize its productive systems effectively and efficiently. When many operations are to be performed on various jobs with limited resources in terms of plant & machinery, the manufacturer has to keep in mind two important factors: namely cost and time. The available facilities should be used to minimize the total cost involved and the total time it takes to do the jobs.

A sequence is simply the order in which different jobs are to be processed. When there is a choice that number of jobs can be performed in different sequences/ orders then the problem of sequencing arises.

Sequencing problems have been most commonly encountered in production shops where different products are to be processed over various combinations of Machines, Overhauling of equipments etc.

Some of the important terms commonly used in sequencing are;

- **Job :** This is the task or activity which is to be sequenced. The job conserve, time and resources.
- **Number of Machines:** This is the service facility through which a job must flow before it is completed.
- **Sequencing :** Sequencing of operation refers to a systematic way of determining the order in which a series of jobs will be processed.
- **Processing time:** The time each job requires on each machine.
- **Idle time on a Machine:** This is the time a machine remains idle during total elapsed time.
- **Total Elapsed time:** This is the time between starting the first job and the completion of the last job. This includes the Idle time if any.
- **No passing Rule:** Passing of jobs on Machines is not allowed i.e. the same order of jobs is manitained over each machine.

5.2. ASSUMPTION IN SEQUENCING PROBLEMS

- Only one operation is carried out on a machine at a time.
- Processing times are known and do not change.
- Each operation, once started must be performed till completion.

- A job is an entity, i.e., even though the job represents a lot of individual parts, no lot may be processed by more than one machine at a time.
- There is only one machine of each type.
- All jobs are ready for processing
- An operations must be completed before its succeeding job can start.
- The time required to transfer jobs between machines is negligible.

5.3. SEQUENCING OF 'N' JOBS THROUGH '2' MACHINES

The sequencing rule for this category of problems was developed by Johnson and is called Johnson's Algorithm. In this situation 'n' jobs will be processed through machines M_1 & M_2. The processing time of all the 'n' jobs on M_1 & M_2 is known and it is required to find the sequence, which minimizes the time to complete all the jobs.

Steps in Johnson's Algorithm

1. List operation time for each job on Machine M_1 & M_2.
2. Select the shortest processing time in the above list.
3. If minimum processing time is on M_1, place the corresponding job first in the sequence. However if the minimum processing time is on M_2 place the corresponding job, last in the sequence. In case of a tie in shortest processing time, it can be broken arbitrarily.
4. Eliminate the jobs which have already been sequenced as a result of step 3.
5. Repeat step 2 and step 3 until all the jobs are sequenced.

Note: In the above algorithm, the machine order is assumed to be M_1 M_2 .

Problem - 1:

There are five jobs, each of which must go through two machins A_1 & B_1 in the order. The A_1 & B_1 . The processing time in (hrs) for each job on each machine is given in the following table.

Processing time (hours)

JOB	1	2	3	4	5
Machine A_1	5	1	9	3	10
Machine B_1	2	6	7	8	4

Determine a sequence for the five jobs that will minimize the total elapsed time. Also calculate the Idle time on both the machines.

Solution : Use steps in Johnsons Algorithm.

Step 1: Processing time for each job on Machines A_1 & B_1 is given in the question.

Step 2: The shortest processing time is 1 hour for job 2 on Machine A_1.

Step 3: As the minimum processing time is on machine A_1 job 2 has to be kept first in the sequence.

2				

Step 4: We ignore job 2 and find shortest time for the rest of the jobs. Now the least processing time is 2 hour for job 1 on machine B_1 . Since it is on B_1 it has to be placed last.

2				1

Step 5: The next minimum processing time is 3 hours for job 4 on machine A_1. Place job 4 from the begining or from the front after job 2.

2	4			1

Step 6: The next minimum time is 4 hours for job 5 on Machine B_1. Place it second last.

2	4		5	1

Step 7 : Job 3 has to be sequenced in the gap or vacant space. The complete sequence of jobs is as follows.

2	4	3	5	1

Further, it is also possible to calculate the minimum elapsed time corresponding to the optimal sequence determined using Johnson's Rule, [by using the individual processing time given in the problem]. For every job in the optimal sequence the time in and time out on machines A_1 & B_1 are calculated and the values are tabulated as follows.

Job	**Machine A_1**		**Machine B_1**		
Sequence	**Time in**	**Time out**	**Time in**	**Time out**	**Idle time**
2	0	1	1	7	1
4	1	4	7	15	-
3	4	13	15	22	-
5	13	23	23	27	1
1	23	28	28	$\boxed{30}$	1
					3

Thus, the Total elapsed time to complete the five job according to the sequence 24351 is 30 hours. During this period the Idle time on the two Machines are:

Idle time on Machine A_1 = 30-28 = 2 hrs (Final 'Time Out' on B - Final Time out on A)

Idle time on Machine B_1 = 3hrs

Note :-

1. To calculate 'Time in for Machine B_1 Minimum time between 'Time Out' on A for corrosponding and 'Time Out' on B for previous job is taken.
2. To calculate idle time on Machine B difference between 'Time in' for next job on Machine B and 'Time Out' for previous job on Machine B is taken.

Let us consider another problem

2. In a factory there are 6 jobs to be performed each of which should go through Machines A & B in the order A,B. The processing timings in hours are given. Determine the sequence for performig the jobs that would minimize the total elapsed time T.

JOBS	1	2	3	4	5	6
M1	1	3	8	5	6	3
M2	5	6	3	2	2	10

Step - 1:

Construct the table as below:

[1 row & no of columns are equal to number of jobs].

Find out the smallest processing time in the given problem. If it is on machine A write the job number on the left side, If it is on machine B write it on the right side.

In this case smallest time is 1 for job 1 Machine A. so write it on left side.

1				

Next minimum time is 2 on Machine B for job 4 & 5. If any two timings are equal check the corrosponding time on the other machine. Write the one having the largest time on right most side. The corrosponding times on Machine A are 5 & 6 So write job 5 on right most side & 4 next to that,

1			4	5

Next is 3 hours for job 2 & 6 on machine A & for job 3 on machine B

1	6	2	3	4	5

So determine the optimum total elapsed for this optimum schedule.

Construct a table showing in & out time on each machine. II machine remains idle when first machine processing the first job, out time is determined by in time + processing time.

To calculate in time for II machine, consider both out time of corrosponding job on Machine 1 & out time of previous job on M/c 2 and select whichever is larger.

JOB	Machine A		Machine B		
	IN	OUT	IN	OUT	Idle Time
1	0	1	1	6	1
6	1	4	6	16	-
2	4	7	16	22	-
3	7	15	22	25	-
4	15	20	25	27	-
5	20	26	27	$\boxed{29}$	-
					1

Idle time for Machine A = (29-26) = 3

Idle time for Machine B = 1 (Since Machine B starts 1 minute for after Machine A starts)

Total elapsed time = 29

5.4 SEQUENCING OF 'N' JOBS THROUGH '3' MACHINES

Johnsons Algorithm which is used to determine a sequence for 'n' jobs 2 machines problem can be used to solve 'n' jobs '3' Machines problem by making some modifications. However on the two conditions mentioned below should be satisfied to solve the problem.

Condition- 1 : The minimum time on Machine 'A' should be greater than or equal to the maximum time on Machine 'B'

Condition-2 : The minimum time on Machine 'C' should be greater than or equal to the maximum time on Machine 'B'

The Procedure explained here is to replace the problem with an equivalent problem involving 'n' jobs and two fiction machines denoted by G and H. The Corresponding time Gi and Hi are defined by,

$$Gi = Ai + Bi$$
$$Hi = Bi + Ci$$

When this problem with prescribed ordering GH is solved, the resulting optimal Sequence will also be optimal for the original problem.

Problem - 1 :

There are five jobs each of which must go through Machines A, B & C in the order ABC Processing times are given below.

Processing time in hours

JOB	1	2	3	4	5
Machine A	8	10	6	7	11
Machine B	5	6	2	3	4
Machine C	4	9	8	6	5

Determine a sequence for five jobs that will minimize the total elapsed time.

Solution : Min Ai = 6   Max Bi = 6   Min Ci = 4

Condition :

Min Ai $\geq$ Max Bi 06
Min Ci $\geq$ Max Bi
6 $\geq$ 6 (Satisfied)
4 $\geq$ 6 (Not Satisfied)

Since one of the two conditions is met, the problem can be converted into 'n' jobs 2 machines.

By creating ficticious machines G H is shown below :

JOB	Gi= [Ai + Bi]	Hi= Bi + Ci
1	8 + 5 = 13	5 + 4 = 9
2	10 + 6 = 16	6 + 9 = 15
3	6 + 2 = 8	2 + 8 = 10
4	7 + 3 = 10	3 + 6 = 9
5	11 + 4 = 15	4 + 5 = 9

This new problem can be solved by the procedure described earlier because of ties in the minimum processing time the following optional sequences are obtained.

3	2	1	4	5

3	2	4	1	5

3	2	4	5	1

3	2	5	4	1

3	2	1	5	4

3	2	5	1	4

Any of the six sequences may be used to order the jobs through machines ABC and they will give the same minimum elapsed time.

JOB	M/c A		M/c B			M/c C		
sequence	Time In	Time Out	Time In	Time Out	Idle Time	Time In	Time Out	Idle Time
3	0	6	6	8	6	8	16	8
2	6	16	16	22	8	22	31	-
1	16	24	24	29	2	31	35	-
4	24	31	31	34	2	35	41	-
5	31	42	42	46	8+5	46	51	5
					31			19

The Minimum elapsed time to process all the five jobs according to the sequences 32145 is 51 hrs.

The Idle time on M/c A = 51-42 = 09 hrs

The Idle time on M/c B = 31 hrs

The Idle time on M/c C = 19 hrs

Let us consider another problem

Problem-2

Determine the optimal schedule of jobs that minimizes the total elapsed time based on the following information:

Job	**A**	**B**	**C**	**D**	**E**	**F**	**G**
M1	3	8	7	4	9	8	7
M2	4	3	2	5	1	4	3
M3	6	7	5	11	5	6	12

Solution

The condition is satisfied as the minimum time on M3 is more than maximum time on M2.

Now Assume only two machines Machine G & Machine H such that,

$t_{Gj} = t_{1j} + t_{2j}$ ....t_{k-1j}

$t_{Hj} = t_{2j} + t_{3j}$ ....t_{kj}

Then solve the problem in the usual method. Now assume two machines.

$G = t_{1j} + t_{2j}$

$H = t_{2j} + t_{3j}$

Now the problem is simplified as,

JOB	**A**	**B**	**C**	**D**	**E**	**F**	**G**
G	7	11	9	9	10	12	10
H	10	10	7	16	6	10	15

Sequence is

A	D	G	B	F	C	E

Now calculate total elapsed time

		A	D	G	B	F	C	E
M1	in out	0 3	3 7	7 14	14 22	22 30	30 37	37 46
M2	in out idle	3 7 3	7 12 -	14 17 2	22 25 5	30 34 5	37 39 3	46 47 7
M3	in out idle	7 13 7	13 24 -	24 36 -	36 43 -	43 49 -	49 54 -	54 59 -

Total idle time for Machine A (59-46) = 13

Total idle time for Machine B 3+2+5+5+3+7+(59-47) = 37

Total idle time for Machine C 7

Total elapsed time 59

5.5 SUMMARY

The selection of an appropriate order for series of jobs to be done on definite number of services facilities in some pre assigned order is called sequencing.

Let there be n jobs to be performed once at a time on each of m machine. The sequence of the machine in which jobs should be done is given. The actual time or expected time is also given. The problem is to find the sequence which minimizes the total elapsed time between the starts of job in the I machine and completion of the last job on the last machine.

5.6 SELF ASSESSMENT QUESTIONS

Problem -1 :

Solve the n jobs '2' machines problem given the processing times is shown.

JOB	1	2	3	4
M_1	2	4	5	1
M_2	6	4	2	3

Problem - 2 : Solve the following 'n' jobs '3' machines problem given the processing times is shown on each machine. Calculate the total elapsed time.

Job	M_1	M_2	M_3
1	13	3	8
2	18	8	4
3	8	6	13
4	23	6	8

5.7 REFERENCES

Operations Research - S.D. Sharma

Operations Research - Lalitha Raman

NOTES

UNIT - 6 : SEQUENCING PROBLEMS OF N JOBS AND M MACHINES

Structure

6.0 Objectives

6.1 Introduction

6.2 Sequencing of 'n' jobs through 'm' Machines

6.3 Two jobs and 'm' Machines problems (Graphical Solution)

6.4 Travelling Salesman Problems (TSP)

6.5 Summary

6.6 Self Assessment Questions

6.7 Reference

6.8 Case Study

6.0 OBJECTIVES

After studying this module you will be able to;

- Design Processing of jobs through different number of Machines and
- Solve Travelling Salesman Problems

6.1 INTRODUCTION

There is no general method available by which we can obtain optimal sequences in problems involving processing of n jobs through k machines. They can be handled only be enumaration which is very lengthy & eloborative because a total of (n!) K different sequences would require consideration in such a case. However we do have a method applicable under the condition that no passing of jobs is permissable.

6.2. SEQUENCING 'N' JOBS THROUGH 'M' MACHIINES.

This a problem of sequencing 'n' jobs 1,2,3------n and 'm'machines M_1, M_2, M_3,------ M. If the following conditions are used, then we can replace 'm' machines by an equivalent of two machines problem which allows us to use Johnson's Algorithm.
Machines are arranged in the order M_1, M_2, M_3,------M_n then.
Condition 1: Minimum time on M_1 Maximum time on [M_2, M_3,------M_{n-1}]
Condition 2: Minimum time on M_n Maximum time on [M_2, M_3,------M_{n-1}]

Where M_1 is the first machine and M_n is the last machine.

Two fictitious machines denoted as G & H are created and their corresponding processing times are given by,

$$Gi = M_1i + M_2i + M_3i \text{ --------- } + (M_{n-1})i$$

$$Hi = M_2i + M_3i + M_4i \text{ --------- } + (M_{n-1})i + M_n$$

Problem -1:

Determine the optimal Sequence of performing 5 jobs on 4 machines. The machines are placed in the order M_1, M_2, M_3, M_4. The processing time of the 5 jobs on the 4 machines are given in the following table.

Job	M1	M2	M3	M4
1	8	3	4	7
2	9	2	6	5
3	10	6	6	8
4	12	4	1	9
5	7	5	2	3

Solution : In this problem,

Condition 1: Minimum time on $M_1 \geq$ Max time on $[M_2, M_3]$ OR

Condition 2: Minimum time on $M_4 \geq$ Max time on $[M_2, M_3]$

Min $M_1 = 7$, Max $[M_2, M_3] = [6,6]$

Min $M_4 = 3$

$7 \geq [6,6]$ or

$4 \geq [6,6]$

Since the first condition is met, the above problem can be converted in 'n' jobs '2' machines problem as follows:

Job	M/c G	M/c H
1	8 + 3 + 4 =15	3 + 4 + 7 = 14
2	9 + 2 + 6 =17	2 + 6 + 5 = 13
3	10 + 6 + 6 =22	6 + 6 + 8 = 20
4	12 + 4 + 1 =17	4 + 1 + 9 = 14
5	7 + 5 + 2 =14	5 + 2 + 3 = 10

Using Johnson's Algorithm the optional sequence is,

3	1	4	2	5

The total elapsed time can be calculated on follows:

Job	**M_1**		**M_2**		**M_3**		**M_4**	
Sequence	Time in	Time out	Time in	Time out	Time in	Time out	Time in	Time out
3	0	10	10	16	16	22	22	30
1	10	18	18	21	22	26	30	37
4	18	30	30	34	34	35	37	46
2	30	39	39	41	41	47	47	52
5	39	46	46	51	51	53	53	56

The total elapsed time is 56 hours to process all the five jobs according to the optimal sequence 31425.

Problem -2 :

Four jobs are to be processed on '5' machines A, B, C, D, & E in the order ABCDE. Find the total minimum elapsed time if no passing of jobs is permitted.

Jobs	**A**	**B**	**C**	**D**	**E**
1	7	5	2	3	9
2	6	6	4	5	10
3	5	4	5	6	8
4	8	3	3	2	6

Solution : Minimum time on A = 5

Minimum time on E = 6

Maximum time on [B, C, D] = [6, 5, 6]

Min time on A $\geq$ Max time on [B, C, D]

5 $\geq$ [6, 5, 6] condition not satisfied

Min time on E $\geq$ Max time on [B, C, D]

6 $\geq$ [6, 5, 6] condition satisfied

Hence convert the problem into 'n' jobs 2 machines problem by creating two fictious machines G & H as follows :

Job	**M/c G**	**M/c H**
1	7 + 5 + 2 + 3 =17	5 + 2 + 3 + 9 = 19
2	6 + 6 + 4 + 5 =21	6 + 4 + 5 + 10= 25
3	5 + 4 + 5 + 6 =20	4 + 5 + 6 + 8 = 23
4	8 + 3 + 3 + 2 =16	3 + 3 + 2 + 6 = 14

Using Johnson's Algorithm the optional sequence is ;

1	3	2	4

Minimum elapsed time can be calculated as follows:

Job	**M/c A**		**M/c B**		**M/c C**		**M/c D**		**M/c E**	
Sequence	In	Out	In	Out	In	Out	In	Out	In	Out
1	0	7	7	12	12	14	14	17	17	26
3	7	12	12	16	16	21	21	27	27	35
2	12	18	18	24	24	28	28	33	35	45
4	18	26	26	29	29	32	33	35	45	51

The total time elapsed to completely process the four jobs on five machines is 51 hours

6.3 TWO JOBS AND 'M' MACHIINES PROBLEM (Graphical Solution)

Consider the following situation,

(a) There are 'm' machines denoted by A, B, C--- K.

(b) There are only two jobs to be performed namely JOB 1, and JOB 2

(c) The technological ordering of each of the two jobs through 'm' machines are known.

(d) The actual processing times A_1, B_1, C_1 -----K_1 and A_2, B_2, C_2,------K_2 are known.

When the problem is to minimize the total elapsed time i.e., to minimize the time for the start of the first job to the completion of the last job then such problem can be solved using graphical technique.

The procedure for solving 2 jobs on 'm' machines problem on graph is described below:

Step -1: Draw two axes at right angles to each other. Represent processing time on job 1 and job 2 along the Vertical & Horizontal axes.

Step-2 : Layout the machine times for the two jobs corresponding axes in the given technological order.

Step -3: With respect to both the jobs, construct machine blocks on the graph for all the machines.

Step- 4 : Start moving from the origin till the finish point. While moving move on a 45° line which represents simultaneous work on both jobs. If not, then move on horizontal or vertical lines to reach the finish point. When you move on horizontal axis, then mean the corresponding job is in progress and the job 2 is idle. Similarly when you are mainly on vertical axis the corresponding job marked on vertical axis is in progress and the job 2 is Idle.

Step- 5: Find the optional path. An optimal path is one which takes the least time for completing both the jobs.

Both jobs cannot be processed simultaneously on one machine which mean that diagnal movement or 45° movement through machine blocks are not allowed.

A good path accordingly is selected by inspection.

Solution- 1: Using graphical method, calculate the minimum time needed to process jobs '1' and '2' on five machines A, B, C, D and E. The technological ordering of the machines with respect to both the jobs are given below.

JOB 1 :	Sequence	A	B	C	D	E
	Time (hrs)	1	2	3	5	1
JOB 2 :	Sequence	C	A	D	E	B
	Time (hrs)	3	4	2	1	5

Solution :

Step-1: Draw two axis at right angles to each other. Represent processing time on job 1 along horizontal axis and processing time on job 2 along vertical axis

Step-2 : Layout the machine times for the two jobs on corresponding axis in the given technological order.

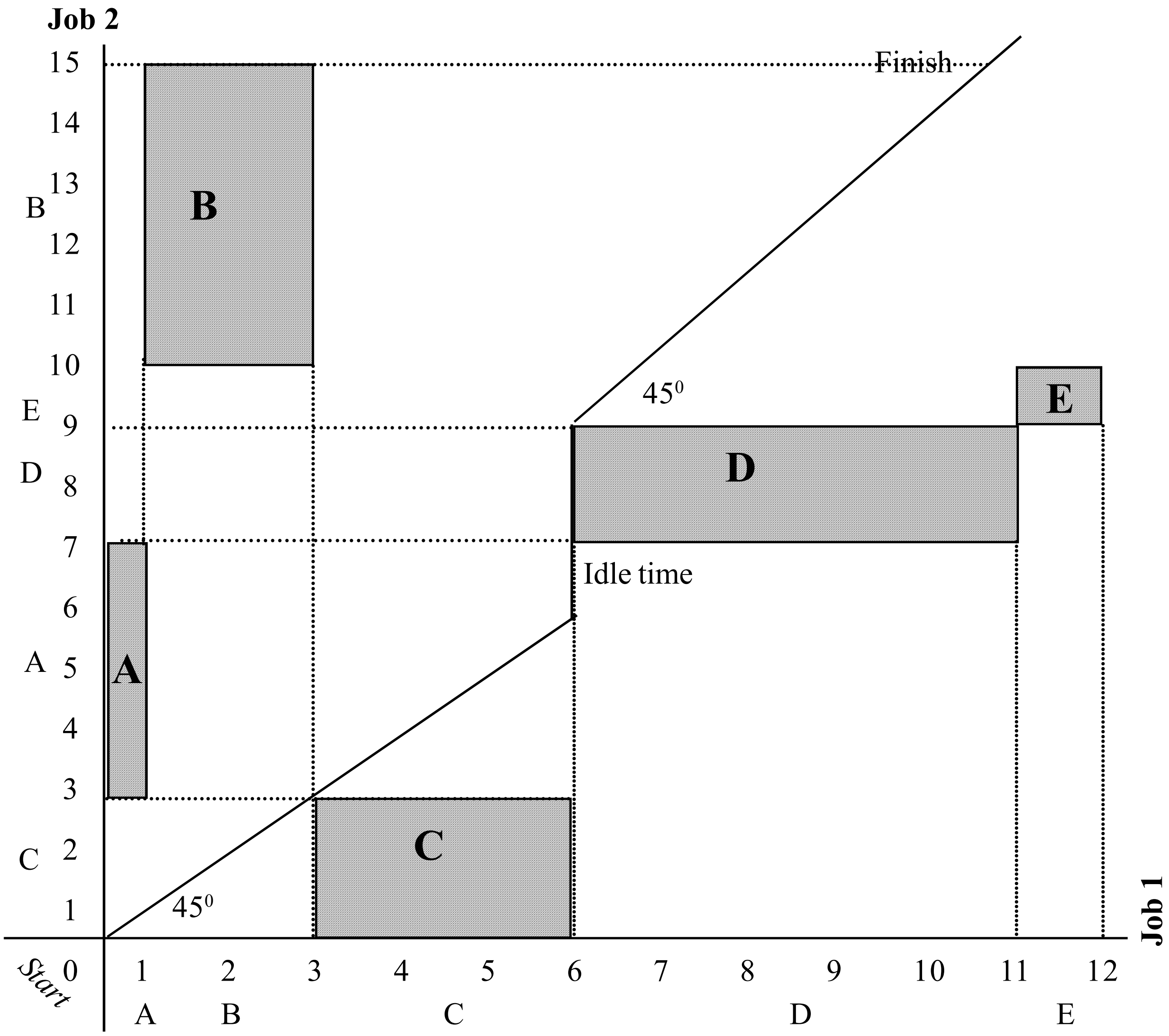
Step-3 : Machine 'A' requires 1 hr for job 1 and 4 hrs for job 2. A rectangle is constructed to represent Machine 'A' similar rectangles an consttructed for machines B, C, D & E.

Step-4 : Develop a path by starting from the origin 'D' and moving through the of completion till the finish is reached. Choose path consisting of only horizontal, vertical and 45^0 lines. A horizontal line represents work on job1 while job 2 remains idle. Similarly a vertical line represents work on job '2' while job 1 remain idle. A 45^0 line indicates simultaneous work on both jobs.

Step -5 : Find the optimal path. The optimal path is one which minimizes the total time required to complete both the jobs.

Step - 6 : Find the elapsed time. It is obtained by adding the idle time for either job to the processing time for that job.

The above steps are illustrated on the graph below:



Total elapsed time = 12+3 = 15 hrs

Job 1 was idle for 3 hrs.

Job 2 was not idle. Processed Continously for 15 hrs.

6.4 TRAVELLING SALESMAN PROBLEM (TSP)

TSP are typically sequencing problems where a salesman must visit certain number of cities and come back to his head quarters by completing his journey in minimum possible time/cost/distance covered.

If the distance/time/cost between every pair of cities is independent of the direction of travel then th problem is a symmetrical TSP. If one or more pairs of cities, the distance/ cost/time varies with the direction, the problem is called asymmetrical TSP.

TSP can be used in the following areas like postal deliveries, inspection, Television relays, assembly lines etc.

TSP is a sequencing problem or a accounting problem but the assignment techniques is used to solve the problem initially. Hence a given TSP is solved for an optimal assignment. If the solution gives suitable route for the salesman then it is accepted as it is otherwise many other techniques can be used to find the sequence.

Problem- 1:

A salesman wants to visit cities A, B, C, D and E. He does not want to visit any city twice before completing his tour of all cities and wishes to return to the point of starting journey. Cost of going from one city to another (in rupees) is shown in the following table. Find the least cost route.

TO CITY

		A	B	C	D	E
	A	0	2	5	7	1
	B	6	0	3	8	2
FROM CITY	C	8	7	0	4	7
	D	12	4	6	0	5
	E	1	3	2	8	0

Solution:

The given travelling salesman problem is first solved as an assignment problem using Hungarian technique. The distances between the same cities is indicated in the table as '0'. However it will be converted to '∞' before starting the assignment problem as shown below:

∞	2	5	7	1
6	∞	3	8	2
8	7	∞	4	7
12	4	6	∞	5
1	3	2	8	∞

The hungarian technique is used to get the optimal assignment. Detailed procedure is given elsewhere in this study material.

Row Minima Operation

	A	B	C	D	E	Row minima
A	∞	1	4	6	0	1
B	4	∞	1	6	0	2
C	4	3	∞	0	3	4
D	8	0	2	∞	1	4
E	0	2	1	7	∞	1
						12
Column Minima	0	0	1	0	0=1	

Lower bound
12+1=13

Column Minima Operation

	A	B	C	D	E
A	∞	1	3	6	$\boxed{0}$
B	4	∞	$\boxed{0}$	6	⊗
C	4	3	∞	$\boxed{0}$	3
D	8	$\boxed{0}$	1	∞	1
E	$\boxed{0}$	2	⊗	7	∞

Solution is optimal. All the rows and column have an assignment. The optimal route for the salesman from the above solution is,

'A - E - A'

As can be seen, the travelling salesman starts from 'A' visits city E and comes back to city 'A'. Hence this is not a solution for a TSP. Many methods are available to solve the TSP. However the iterative penalty method is illustrated for this TSP as shown below. This is also called break and bound method.

Starting with the final table/ optimum table of assignment, calculate the penalty for all zeores in the matrix. Penalty is calculated by adding minimum element of that row and minimum element of that column.

	A	B	C	D	E
A	∞	1	3	6	0
B	4	∞	0^1	6	0^0
C	4	3	∞	0^9	3
D	8	0^2	2	∞	1
E	0^4	2	0^0	7	∞

This penalties are indicated above the Zero in the respective cells. Select the highest penalty in the matrix and route the salesman test. Now the route is $\boxed{\text{C-D}}$ **become the maximum penalty** is 9. Remove the C- row and D column from the matrix and route the matrix

	A	B	C	E
A	∞	1	3	0^1
B	4	∞	0^0	0^0
D	8	0^2	∞	1
E	0^4	2	0	∞

[Note : The reverse route i.e., route D-C is blocked by taking the anti as ∞]

Now the maximum penalty is '4' for route $\boxed{\text{E-A}}$ remove the E- row and 'A' column and the reverse route is taken as ∞ if the possibility exists.

	B	C	E
A	1	3	∞
B	∞	0	0
D	0	∞	1

As there is no zero in the first row perform row minima operation

	B	C	E
A	0^2	2	∞
B	∞	0^2	0^1
D	0^1	∞	1

Select the maximum penalty which is two select either $\boxed{\text{A- B}}$ or B- C as the penalty is same.

	C	E
B	0	0
D	∞	1

Perform row minima operation as there is no zero in the second row.

	C	E
B	0^{∞}	0^0
D	∞	0^{∞}

The maximum penalty is ∞ and either B-C or D-E route can be selected.

	E
D	

Hence the optimal route is,

A- B- C- D- E- A

The corresponding cost is,

2 + 3 + 4 + 5 + 1 = 15/- Rs

Note : In the above, penalty is the summation of row minima and column minima calculated wherever zeros exist.

6.5 SUMMARY

The sequencing problems of multiple machines can be solved as discussed. One of the extension of the sequencing problem is travelling salesman problem where in a person would travel different places once at a time and return back to his own city without visiting any cities twice.

6.6 SELF ASSESSMENT QUESTIONS

Problem- 1 : Solve the n jobs '2' machines problem given the processing times is shown.

JOB	1	2	3	4
M1	2	4	5	1
M2	6	4	2	3

Problem-2 : Solve the following 'n' jobs '3' machines problem given the processing times is shown on each machine. Calculate the total elapsed time.

JOB	M^1	M^2	M^3	
1	13	3	8	
2	18	8	4	
3	8	6	13	
4	23	6	8	

Problem-3 : Solve the following sequencing problem when passing is not allowed.

JOB	M^1	M^2	M^3	M^4
1	15	5	4	15
2	12	2	10	12
3	16	3	5	16
4	17	3	4	17

Problem- 4 : There are two jobs to be processed through four machines A, B, C & D.
The prescribed technological orders are;
JOB 1: A B C D

JOB 2: D B A C

Processing times in hours are given in the following table.

JOB	MACHINES			
	A	B	C	D
1	2	4	5	1
2	6	4	2	3

Find the total elapsed time to complete both the jobs.

Problem- 5: Solve the following salesman problem using penalty method. Cell entries represent time in hours.

To City

From City		A	B	C	D
	A	∞	4	7	3
	B	4	∞	6	3
	C	7	6	∞	7
	D	3	3	7	∞

6.7 REFERENCE

Operations Research - S. D. Sharma

Operations Research - Lalitha Raman

Operations Research - Kanthi Swaroop

Operations Research - Hamdy Taha

6.8 CASE STUDY

A manager has a total of ten employees working on six projects. There are overlaps among the assignments as the following table shows,

	1	2	3	4	5	6
1		X		X	X	
2	X		X		X	
3		X	X	X		X
4			X	X	X	
5	X	X	X			
6	X	X	X	X		X
7	X	X			X	X
8	X	∞	X	X		
9					X	X
10	X	X		X	X	X

∞

The manager must meet all 10 employees once in a week to discuss their progress. Currently the meeting with each employee lasts about 20 minutes- that is, a total of 3 hours and 20 minutes for all 10 employees. A suggestion is made to reduce the total time by holding group meetings, depending on the projects the employees share. The manager wants to schedule the projects in way that will reduce the traffic (number of employees) in and out of the meeting room. How should the projects be scheduled.

NOTES

UNIT-7: REPLACEMENT

7.0 Objectives

7.1 Introduction

7.2 Replacement Problems

7.2.1.1. Basis of replacement problems.

7.3 Replacement of machines that deteriorate with time

7.3.1.1. Replacement policy when the value of money does not changes with time

7.3.1.2. Replacement policy when the value of money changes with time

7.4 Replacement of equipments that fail suddenly.

7.5 Summary

7.6 Self Assessment Questions

7.7 Reference

7.8 Case Study

7.0 OBJECTIVES

After studying this unit you will be able to:

- ♦ Explain replacement theory
- ♦ Solve the replacement problems that deteriorates with time
- ♦ Solve replacement problems that fail suddenly
- ♦ Decide whether individual replacement is better or group replacement is better.

7.1 INTRODUCTION

Repairs and replacement are the parts of our daily life. When ever we purchase any durable item it would attract annual maintenance costs and after some years it is due for replacement. The same is of more significant in nature in industries as number of durable items or machines is more in number. When to replace the machine or what is the optimal time of replacement is a matter of concern of operation research. Lot of studies has been taken place in deciding this optimum time.

7.2 REPLACEMENT PROBLEMS

The replacement problems are concerned with the situations that arise when some items such as machines electric- bulb etc., need replacement due to their decreased efficiency, failure or break down. Such decreased efficiency or complete breakdown may either be gradual or all of a sudden.

When items face the problem of decreased efficiency due to age, costs like operating cost, repair cost, scrap etc., would be increased, In such case replacement of old item with new one is the only alternative to prevent such increased expenses. The problem of replacement is to decide best policy. To determine an age at which the replacement is most economical instead of continuing at increased cost.

7.2.1 Basis of replacement problem

Suppose we have to use an expensive piece of equipment for the next T years. The life span of this piece, L, is much smaller than T so sooner or later we shall have to replace it with a new piece of equipment. Suppose that it cost Rs C(t) to keep this piece of equipment for t years, t=1,2,3,....,L. This cost includes price, maintenance, insurance and salvage value. What is the optimal replacement policy? Namely, what is the replacement policy that will minimize the cost over the next T years? We assume that we commence this process with a brand new piece of equipment.

Different decision that need to be taken are:

1. Whether to wait for complete facture or replace earlier.
2. Whether to replace same item or buy an improved version.

If time is measured in discrete units, then the average annual cost will be minimized by replacing the Machine when the next period costs become greater than the current average cost.

In replacement problems it is assumed that the maintenance is done in the beginning of every year.

A machine would fail in the following ways

- ° Gradual failure
- ° Sudden failure
- ° Progressive factors: Here the probability of failure increases with time.
- ° Retrogressive failure: Certain items would have more probability of failure in the beginning of their life and as the time passes the chance of failure becomes less.
- ° Random failure: Here failure is related to age. But the item may fail any time during its life may be because of shock or accidents.

There are two types of problems that one needs to encounter. They include

- ° Replacement that deteriorate with time
- ° Replacement that fail suddenly.

7.3 REPLACEMENT THAT DETERIORATES WITH TIME

Here the associated costs are

1) Running costs

2) Capital costs of purchasing new item

Here the cost of maintenance of a machine (m/c) is given as a function increasing with time and its scrap value remains constant. The average annual cost will be minimized by replacing m/c average cost to date becomes equal to the current maintenance cost.

These types of problems are handled in two ways.

- ° Without considering the time value of money.
- ° Considering the time value of money.

7.3.1 Without considering the time value of money

Here cost of subsequent years are directly added without taking into consideration the time value of money. The total cost of maintenance is calculated. The total costs are added cumulatively & then average cost based on number of years is calculated.

Problem:

1. Cost of a machine is Rs. 6100 and its scrap value is only Rs. 100. The maintenance cost are found from experience to be

Year	1	2	3	4	5	6	7	8
Maintenance cost	100	250	400	600	900	1250	1600	2000

Step -1

Find an average cost of 1 year during the life of the Machine as following

Total cost of Replacement in I year

= Maintenance cost in I year + loss in purchase price

= 100 + (6100 -100) = Rs. 6100

Total cost of Replacement of II year

= Total maintenance cost in I and II year + loss in purchase price.

100 + 250 + (6100-100) = Rs. 6350

Average cost per year during first two years = Rs. 3175

Similarly the total cost & average cost of all the years are calculated.

Replacement of the end of year 1	Maintenance Cost R_n 2	Total Maint, cost ΣR_n Cumulative 3	Different Price & resale price (C.S.) 6100-100 4	P(n) Total cost 5= 3+ 4	Avg. Cost PO(n)/n 6= 5/1
1	100	100	6000	6100	6100
2	250	350	6000	6350	3175
3	400	750	6000	6750	2250
4	600	1350	6000	7350	1837
5	900	2250	6000	8250	1650
6	1250	3500	6000	9500	1583
7	1600	5100	6000	11100	1586
8	2000	7100	6000	13100	1638

It can be observed that the maintenance cost in the 6^{th} year is minimum and average cost starts increasing from 7^{th} year.

Hence the Machine should be replaced before beginning of 7^{th} or at the end of 6^{th} year.

2. The Machine owner has 3 Machine of purchase price Rs. 6000 each & cost / year of maintaining each Machine as shown below. Two of these are two year old & third one is one year old. He is considering a new Machine of purchase price of Rs. 8000 with 50% more capacity than one of the old one.

Year	1	2	3	4	5	6	7	8
Maintenance of new Machine	1200	1500	1800	2400	3100	4000	5000	6100
Resale Price of new Machine	4000	2000	1000	500	300	300	300	300
Maintenance of old Machine	1000	1200	1400	1800	2300	2800	3400	4000
Resale price of old Machine	3000	1500	750	375	200	200	200	200

Assuming that the loss of flexibility due to fewer Machine is of no imp & he continues to have sufficient work for the old Machine what should his policy be.

Solution:

The Machine owner has to replace 3 old machines for 2 new Machine to have 50% more capacity if he wants to replace

Let us calculate the average cost in both the cases.

a. Average cost of old machine

Year	Maintenance cost	Total Maintenance	Diff. bet. Price & resale price	Total cost	Average Cost
1	1000	1000	3000	4000	4000
2	1200	2200	4500	6700	3350
3	1400	3600	5250	8850	2950
4	1800	5400	5625	11025	2756
5	2300	7700	5800	13500	2700
6	2800	10500	5800	16300	2717

b. Average cost of new machine

Year	Maintenance cost	Total Maintenance	Diff. bet. Price & resale price	Total cost	Average Cost
1	1200	1200	4000	5200	5200
2	1500	2700	6000	8700	4350
3	1800	4500	7000	11500	3833
4	2400	6900	7500	14400	3600
5	3100	10000	7700	17700	3540
6	4000	14000	7700	21700	3616

The old machines have to be replaced at the end of 5^{th} year.

The lowest average cost of Rs 3540 of for new machine is equivalent is 3540 X 2/3 i.e Rs 2360 for smaller machine. As 3 old machines or equal to 2 new machines. Since this value is less than minimum average cost of Rs 2700 of small machine it is appropriate to go for new machine.

3. Machine A costs Rs 9000/- Annual operating cost is Rs 200 and then increase by 2000 every year i.e. the in the fourth year operating cost become 6200. Determine the best age at which replacement has to be done. If the optimal replacement policy is followed what will be the average yearly cost of owning & operating the machine?

Machine B costs Rs 10,000/-. Annual operating cost is Rs 400 and then increase by 800 every year. You have own a machine of type A which is one year old. Should you replace it with B & if so when.

Year	Operating Cost	Total Op. Cost	Total cost of M/c + Op. Cost	Avg. cost
1	200	200	9200	9200
2	2200	2400	11400	5700
3	4200	6600 .	15600	5200
4	6200	12800	21800	5450

So the Machine should be replaced at the end of III year.

	M/c B Op.	Total Op.	Cost of M/c+ Op. Cost	Average cost
1	400	400	10400	10400
2	1200	1600	11600	51800
3	2000	3600	13600	4533
4	2800	6400	16400	4100
5	3600	10000	20000	4000
6	4400	14400	24400	4066

That means Machine has to be replace at the end of 5th year. The optimum operating cost would be Rs. 4000.

So Machine A should be replaced by Machine B When operating cost Machine B is less than that of Machine A.

For I year 2200< 4000

II year 4200> 4000

That Machine A should be replaced by Machine B in the succeeding year i.e. at the end of II year.

4. The operating cost, maintenance cost and the resale value from running a car is a given below. Determine the best time for replacing the car.

Year	1	2	3	4	5
Year end trade in value	1900	1050	600	500	500
Annual operating cost	1500	1800	2100	2400	2700
Manintenance cost	300	400	600	800	1000

Year	Op cost	Maint. cost	Resale Value R. V	Total OP+ Maint	Cumulative Total Cost	Diff. 3000-RV	Total cost	Avg. cost
1	1500	300	1900	1800	1800	1600	3400	3400
2	1800	400	1050	2200	4000	2450	3450	3225
3	2100	600	600	2700	16200	2400	9600	3200
4	2400	800	500	3200	9900	3000	12400	3225
5	2700	1000	500	3700	13600	3000	12600	3320

This table shows that the car should be replaced at the end of third year.

7.3.2 Replacement Policy when Value of Money Changes With Time.

When value of money is taken into consideration, we can not added costs of different years directly. Instead a suitable discount factor is taken.

Here two assumptions are made.

1. The equipment has no salvage value
2. The maintenance costs are incurred in the beginning of different time periods.

To determine a policy for selection of an economically best item amongst the available equipments.

Step-1 : Consider the case of two equipments A & B we will first find the best replacement age for the both the equipments by making use of

$R_n - 1<(1-V)V_n < R_n$

Let the optimum replacement age for A and B comes out to be n1 & n2 respectively.

Step-2 : Compute the fixed annual payment or weighed average cost for each equipment by using the formula

$$W(n) = \quad \alpha\, n= \frac{C-R_0+VR_1+V^2R_2+............V_{nd}R_n}{1+V+........,V^{n+}}$$

Substituting n = n1 for equipment A and n= n_2 for equipment B

Step-3 :

1. If W $(n_1) < W(n_2)$ Choose A
2. If W $(n_1) > W(n_2)$ Choose B
3. If W $(n_1) < W(n_2)$ both the equipments are equally good.

Problem:

Let the value of money be assumed to be 10% per year and suppose that Machine A is replaced after 3 years where as Machine B is replaced after every 6 years. The yearly costs of both the Machine are given below.

Year	1	2	3	4	5	6
A	1000	200	400	1000	200	400
B	1700	100	200	300	400	500

The present worth of the money to be spent over in a period of one year as

$$V= \frac{100}{100+10} = \frac{10}{11} =0.909$$

The total discount cost (present worth) of A for 3 years is

$1000 + 200 \text{ x } 0.9091 + 400 \text{ x } (0.9091)^2 = \text{Rs. } 1512$

Total discounted cost of B for 6 years

$1700+ 100 \text{ x}0.9091+ 200 +(0.9091)^2+300 \text{ x } 0.9091^3 +400 \text{ x } (0.90914)^4 + 500 \text{ x } (0.9091)^5 = \text{Rs } 2765$

Avg. Yearly cost of Machine A = Rs 1512/3 = 504

Avg. Yearly cost of Machine = Rs. 2765/ 6 = Rs. 461

Even though Machine B appears to be good: Suppose we consider 6 years period for M/c A also

The total discounted cost is

$1000 + 200 \text{ x } (0.9091) + 400 \text{ x } (0.9091)^2 + 1000 \text{ x } (0.9091)^3 +$
$200 \text{ x } (0.9091)^4 +400 \text{ x } (0.9091)^5 = 2647$

Avg. yearly cost of Machine A = 2647/6 = 118

Hence Machine A should be purchased.

2. A pipeline is due for repairs, it will cost Rs. 10,000 & lasts for 3 years. Alternatively a new pipe line can be laid at a cost of Rs. 30,000 and last for 10 years. Assuming cost of capital to be 10% which should be chosen.

Solution:

The present worth of money

$$= \frac{100}{100+10} = 0.909$$

Let K_n denote the discount Value of all future costs associated with a policy of replacing the equipment after n years. This if we designate initial out lay by C

$Kn = C + CV^n + CV^{2n} + \ldots\ldots\ldots\ldots \infty$

$= C\,(1+V^n+V^{2n}+\ldots\ldots\ldots+\infty) = C/(1-V^n)$

$$K^3 = \frac{10000}{1-(0.909)^3} = \text{Rs. } 4021$$

For the existing pipe line

$$K_{10} = \frac{30000}{1-(0.909)^{10}} = \frac{30000}{1-0.3855} = \text{Rs. } 48{,}820$$

for new pipe line

Since $K_3 < K_{10}$ the existing pipe line should be continued.

7.4. REPLACEMENT OF EQUIPMENTS THAT FAIL SUDDENLY

It is difficult to predict that a particular equipment will fail at a particular time. This can be overcome by determining the probability distribution of failure.

Here the objectives becomes to find the value of t (i.e. end of period at which failure occurs) which minimizes the total cost involved for the replacement.

Here two types of policies can be implemented.

Individual replacement policy:

Under this policy an item is replaced immediately after its failure

Group replacement policy

Under the policy all the items are replaced before the optimal time.

Mortality tables are used to derive the probability distribution of the life span of an equipment.

Let M(t) = No. of survivors at anytime t

M(t-1) = No. of survivors at any time t-1 and

N = initial No. of equipments

Then the probability of failure during time period t is given by

$P(t) = [M(t-1) - M(t)]/N$

The probability that an equipment survived till age (t-1), will fail during the interval (t-1) to t can be defined as the conditional probability of failure.

It is given by

$P_c(t) = [M(t-1) - M(t)] / M(t-1)$

The probability of survival till age t is given by

$P_s(t) = M(t)/N$

1. The following failure ratios have been observed for a certain type of transistors in a digital computer.

End of the week	1	2	3	4	5	6	7	8
Probability of failure to date	0.05	0.13	0.25	0.43	0.68	0.88	0.96	1

The cost of replacing an individual failed transistor is Rs. 1.25/- The decision is made to replace all these transistors simultaneously at fixed intervals and to replace the individual transistors as they fail in service. If the cost of group replacement is 30Paisa /transistor, what is the best interval between group replacements. At what group replacement price/ transistor would a policy of strictly individual replacement become preferable to the adopted policy.

Solution :

Suppose there are 1000 transistors in use. Let P_i be the probability that a transistor which was new when placed in position for use, fails during the I week of its life. Thus, we have

$P_1 = 0.05$ $P_2 = 0.13 - 0.5 = 0.08$

$P_3 = 0.25 - 0.13 = 0.12$ $P_4 = 0.43 - 0.25 = 0.18$

$P_5 = 0.68 - 0.43 = 0.25$ $P_6 = 0.88 - 0.68 = 0.20$

$P_7 = 0.96 - 0.88 = 0.08$ $P_8 = 1 - 0.96 = 0.04$

Let N_j denote the No. of replacement made at the end of the i^n week. Then we have

N_0 = No. of transistors in the beginning = 1000

$N_1 = N_0P_1 = 1000 \times 0.05 = 50$

$N_2 = N_0P_2 + N_0P_2 = 1000 \times 0.08 + 50 \times 0.05 = 82$

$N_3 = N_0P_3 + N_1P_2 + N_2P_1 = 1000 \times 0.12 + 50 \times 0.08 + 82 \times 0.05 = 128$

$N_4 = N_0P_4 + N_1P_3 + N_2P_2 + N_3P_1 = 199$

$N_5 = N_0P_5 + N_1P_4 + N_2P_3 + N_3P_2 + N_4P_1 = 289$

$N_6 = N_0P_6 + N_1P_5 + N_2P_4 + N_3P_3 + N_4P_2 + N_5P_1 = 272$

$N_7 = N_0P_7 + N_1P_6 + N_2P_5 + N_3P_4 + N_4P_3 + N_5P_2 + N_6P_1 = 194$

$N_8 = N_0P_8 + N_1P_7 + N_2P_6 + N_3P_5 + N_4P_4 + N_5P_3 + N_6P_2 + N_7P_1 = 195$

From the above calculation, we observe that the expected No. of transistors failing each week increases till 5^{th} week and then starts decreasing and later again increasing from 8^{th} week.

The expected life of each transistor is

1x0.05 + 2x0.08 +3x0.12 + 4x0.18 + 5x0.25 + 6x0.2 +7x0.08 + 8x0.04 = 4.62 weeks

Avg. No. of failure/wk. = 1000/4.62 = 216

Cost of individual replacement

216 x 1.25 = 270 / week.

Since the replacement of all the 1000 transistors simultaneously cost 30P per transistor and the replacement of individual transistor on failure cost of Rs 1.25, the average cost of different group replacement policy is as given below.

End of week	Individual Replacement	Total cost Individual + Group	Average cost
1	50	50X1.25 + 1000X0.30= 363	363
2	132	132X1.25 + 1000X0.30= 465	232.50
3	260	260X1.25 + 1000X0.30=625	208.30
4	459	459X1.25 + 1000X0.30=874	218.50

Since the average cost is lowest against week 3, the optimum interval between group replacement is 3 weeks, further since the average cost is less than Rs 270 (for individual replacement), the policy of group replacement is better.

7.5. SUMMARY

The replacement problems are concerned with the situations that arises when some items such as machines electric- bulb etc., need replacement due to their decreased efficiency, failure or break down such decreased efficiency or complete breakdown may be either be gradual or all of a sudden.

The replacement problems are studied under 2 main heading

1. Replacement of the machine that fail suddenly
2. Replacement of the machine that deteriorate with time.

7.6. SELF ASSESSMENT QUESTIONS

1. A firm is considering replacement of a machine whose cost price is Rs 12,200 and the scrap value Rs 200. The running costs are as below

Year	1	2	3	4	5	6	7	8
Running cost	200	500	800	1200	1800	2500	3200	4000

When the machine should be replaced?

2. The value of money be 10% per year & the yearly cost of both the machine is as given below.

Year	1	2	3	4	5
Machine A	1800	1200	1400	1600	1000
Machine B	2800	200	1400	1100	600

Determine which machine should be purchased

3. In a machine shop a particular cutting tool Cost Rs 6/- to replace. If the tool breaks on the job, the production disruption and associated costs amount Rs 30/-. The past life of a tool is given as follows.

Job No	1	2	3	4	5	6	7
Proportion of broken Tools on job	.01	.03	.09	.13	.25	.55	.95

After how many jobs should the shop replace a tool before it breaksdown.

7.7. REFERENCE

1. Operation Research By Kanti Swaroop
2. Operation Research By S.D. Sharma
3. Operation Research By N.D Vohra
4. Theory and Problems of Operation research By Richard Brouer

7.8. CASE STUDY

Murugan leather works is a medium sized leather goods manufacturing company situated near Chennai. The M.D. Mr Pandyan who started his career as an instructor in leather technology in a government institute, originally startedf a small leather goods workshop with ten workers who were all his former students. As most of the workers were commuters he started giving them a monthly conveyance allowance of Rs 40/- as a gesture of goodwill. As years passed the small workshop steadily grew into a fairly big work centre employing nearly 200 men in two shifts. The conveyance allowance was continued and the employee considered it as an integral part of their wages.

As the activities of the company grew steadily, the M.D. appointed a qualified accountant, labour welfare officer, and an office assistant. Soon after his joining Mr Chary, the accountant started working on a number of cost reduction measures. He was much concerned about the conveyance allowances of Rs 96,000/- being paid to workers every year. One day he got a new idea to provide conveyance to workers and to save Rs 96,000/- annually by purchasing buses. These buses can also be used for transporting goods from & to the railway station saving an amount of Rs 15,000 annually. He estimated that the two buses would cost Rs. 4 lakhs & the amount could be easily recovered in five to six years.

The buses could be operated for at leas 10 years without major repairs and with some repair they could be used for a longer period. He estimated that in case buses were replaced after 10 years they could fetch Rs 35,000/- if not more. He also estimated that the fuel and other running expenses for the buses would cost less than 40,000 annually. Further the depreciation would earn tax exemptions, tax rate being 50%

Mr. Chary pointed out to Mr. Pandyan that this scheme would not only solve the conveyance problem and result in savings to the company, but the employees would be happy to travel in company's bus rather than commute in public transport. The labour welfare officer said that though it is a good scheme there would be many difficulties in implementing the scheme. He knew he can earn 15% after taxes in amount invested by the company. He wanted to know what is rate of return on the proposed investment of Rs 4 lakhs and how it would compare with that if his present investment in the company.

UNIT-8 : TRAVELLING SALESMAN PROBLEM

Structure

8.0 Objectives

8.1 Introduction

8.2 Travelling Salesman Problem

8.2.1 Application

8.2.2 Formulation

8.3 Problems

8.4 Transshipment Problem

8.4.1 Characteristic of transshipment problems

8.4.2 Sample Problem

8.5 Summary

8.6 Self Assessment Questions

8.7 Reference

8.0 OBJECTIVE

After studying this unit you will be able to:

- Understand Travelling salesman Problem
- Find out the solution to symmetric & asymmetric types of traveling salesman problems
- Solve transshipment problems

8.1 INTRODUCTION

The traveling salesman problem (TSP) asks for the shortest route to visit a collection of cities and return to the starting point. The traveling sales man problems are unique type of problems in the assignment problem.

Suppose a salesman wants to visit certain number of cities alloted to him, he knows the distance or cost or time of journey between each pair of cities, usually denoted by C_{ij} i.e from city i to city j, the problem is to select such a route that starts from his home city passes through each city once & only once & return to his home city in the shortest possible distance or at least cost or at least time.

8.2 TRAVELLING SALESMAN PROBLEM

The Traveling Salesman Problem (TSP) is a deceptively simple combinatorial problem. It can be stated very simply:

A salesman spends his time visiting n cities (or nodes) cyclically. In one tour he visits each city just once, and finishes up where he started. In what order should he visit them to minimize the distance traveled?

The TSP naturally arises as a sub problem in many transportion and logistics applications, for example the problem of arranging school bus routes to pick up the children in a school district. This bus application is of important historical significance to the TSP, since it provided motivation for Merrill Flood, one of the pioneers of TSP research in the 1940s. A second TSP application from the 1940s involved the transportation of farming equipment from one location to another to test soil, leading to mathematical studies in Bengal by P.C. Mahalanobis and in Iowa by R.J. Jessen. More recent applications involve the scheduling of service calls at cable firms, the delivery of meals to homebound persons, the scheduling of stacker cranes in warehouses, the routing of trucks for parcel post pickup, and a host of others.

The problem may be classified in two forms

Symmetrical : The problem is said to be symmetrical if distance or cost or time between every pair of cities is independent of the direction of his journey.

i.e. $C_{ij} = C_{ji}$

Asymmetrical : The problem is said to be asymmetrical if in one or more pairs of cities the distance or cost or time changes with direction.

$C_{ij} = C_{ji}$

8.2.2 Application of travelling salesman problem

Although transportation applications are the most natural setting for the TSP, the simplicity of the model has led to many interesting applications in other areas. A classic example is the scheduling of a machine to drill holes in a circuit board or other object. In this case the holes to be drilled are the cities, and the cost of travel is the time it takes to move the drill head from one hole to the next. The technology for drilling varies from one industry to another, but whenever the travel time of the drilling device is a significant portion of the overall manufacturing process then the TSP can play a role in reducing costs.

8.2.2 Formulization

The traveling salesman problem can be formularized as assignment problem & it is solved in the same way the assignment problems are solved. Only the last step of checking of completeness of cycle differs.

Suppose C_{ij} is the distance or cost or time from city i to city j & $X_{ij} = 1$ if the sales man goes directly from city i to city j & $X_{ij} = 0$ otherwise.

The assignment table would look as below.

	A_1	A_2	A_3	A_j	A_n
A_1	∞	C_{12}	C_{13}	C_{1i}	C_{1n}
A_2	C_{21}	∞	C_{23}	C_{2i}	C_{2n}
A_3	C_{31}	C_{32}	∞	C_{3j}	C_{3n}
Ai	C_{i1}	C_{i2}	C_{i3}	∞	C_{in}
An	C_{n1}	C_{n2}	C_{n3}	C_{ni}	∞

Min

$$\sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij} \; C_{ij}$$

so that Xij must be so chosen that no city is visited twice. Further we cannot travel from a city to same city.

So put Cii= ∞

8.3 PROBLEM

A person has to travel among 5 cities the distance between which is given below. He can visit each of the cities only once. Determine the optimum sequences he should follow to minimize total distance traveled (Hungarian Method).

	A	B	C	D	E
A	-	4	7	3	4
B	4	-	6	3	4
C	7	6	-	7	5
D	3	3	7	-	7
E	4	4	5	7	-

Reduce the cost matrix first by deducting row minimum.

∞	1	4	0	1
1	∞	3	0	1
2	1	∞	2	0
0	0	4	∞	4
0	0	1	3	∞

Deduct column minimum

∞	1	3	0	1
1	∞	2	0	1
2	1	∞	2	0
0	0	3	∞	4
0	0	0	3	∞

Now start assignment

∞	1	3	$\boxed{0}$	1	✓
1	∞	2	ⓧ	1	✓
2	1	∞	2	$\boxed{0}$	
$\boxed{0}$	ⓧ	3	∞	4	
ⓧ	ⓧ	$\boxed{0}$	3	∞	
			✓		

Add the minimum of uncovered elements (1) to all intersection points and deduct from uncovered elements and again start assignment.

	A	B	C	D	E
A	∞	X	2	$\boxed{0}$	X
B	$\boxed{0}$	∞	1	X	X
C	2	1	∞	3	$\boxed{0}$
D	X	$\boxed{0}$	3	∞	4
E	X	X	$\boxed{0}$	4	∞

Now all the rows have assignment

Now the salesman has to move

A ⟶ D, D ⟶ B, B ⟶ A, C ⟶ E & E ⟶ C

& min cost is 3+ 3 + 4 + 5 + 5 = 20

But this solution is not optimum as without completing C & E, the salesman returns back to A.

II Iteration

Now the next minimum element is 1. Try to switch from 0 to 1 so that the cycle is completed

There are two 1's here

i.e instead of B ⟶ A we can go for B ⟶ C

Instead of C ⟶ E we can go for C ⟶ B

i.e, A ⟶ D, D ⟶ B, B ⟶ C, C ⟶ E, E ⟶ A

4+5+6+3+3 =21

OR

A ⟶ E, E ⟶ C, C ⟶ B, B ⟶ D, D ⟶ A

Now the total cost is 21.

If a machine has to produce 5 different products A_1, A_2, A_3, A_4, A_5, sequencing must be determined so as to which item should be taken next.

2. Given the matrix of set up costs, show how to sequence the production. So as to minimize the set-up cost/ cycle.

	A_1	A_2	A_3	A_4	A_5
A_1	-	2	5	7	1
A_2	6	-	3	8	2
A_3	8	7	-	4	7
A_4	12	4	6	-	5
A_5	1	3	2	8	-

After reducing the matrix by subtracting row min,

∞	1	4	6	0
4	∞	1	6	0
4	3	∞	0	3
8	0	2	∞	1
0	2	1	7	∞

After reducing the matrix by subtracting column min

∞	1	3	6	0
4	∞	0	6	0
4	3	∞	0	3
8	0	1	∞	1
0	2	0	7	∞

Now start assignment

	A	B	C	D	E
A	∞	1	3	6	$\boxed{0}$
B	4	∞	$\boxed{0}$	6	⊗
C	4	3	∞	$\boxed{0}$	4
D	8	$\boxed{0}$	1	∞	1
E	$\boxed{0}$	2	⊗	7	∞

Now each row & each column has got an assignment

The route becomes

A⟶E, E⟶A, B⟶C, C⟶D and D⟶B

And the cost is

1+1+3+4+4= 13

But this is not an optimum assignment as the cycle is not complete.

Now taking 1 instead of 0

Let us the change the path from A⟶E to A⟶ B

Now the path becomes

A⟶B, B⟶C, C ⟶ D and D ⟶ B

Again he has to travel from D$\longrightarrow$B

So change Shift D$\longrightarrow$B to D$\longrightarrow$ E at the cost of 1 Rs extra

The final path is

A $\longrightarrow$B, B $\longrightarrow$C, C$\longrightarrow$D, D $\longrightarrow$E and E $\longrightarrow$A

The final cost is

2+3+4+5+1 = 15

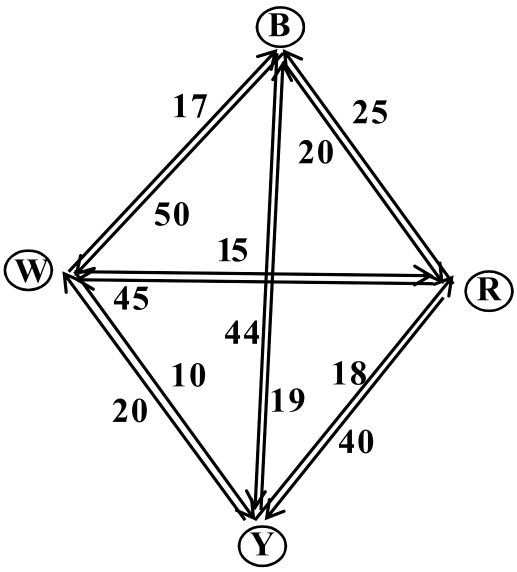
3. The daily production schedule at the Rainbow company includes batches of white (W), Yellow (Y), Red (R), and Black (B) paints. As the company uses same facilities, proper cleaning between batches is necessary. The clean uptime varies as the colour proceeds and succeeds varies which is given in the table below.

Determine the optimal sequencing for all daily production of 4 colours that will minimize the associated total clean uptime.

Clean up minutes given next paint is

Current paint	White	Yellow	Black	Red
White	∞	10	17	15
Yellow	20	-	19	18
Black	50	44	-	25
Red	45	40	20	-

Fig. Summarizes the problem. Each paint is represented by a node and direction arrows represents the clean-up time needed to reach one node from the other. The situation thus reduces to determine shortest loop that starts at one node & passes through each of the remaining three nodes exactly once before returning back to the starting node.



We can solve this problem by exhaustively enumerating the six (4-1) ! = 3! = 6 possible loops of net work. The following table shows all possible combinations.

Production loop	Total clean-uptime
W $\longrightarrow$ Y $\longrightarrow$ B $\longrightarrow$ R $\longrightarrow$ W	10+19+25+45=99
W $\longrightarrow$ Y $\longrightarrow$ R $\longrightarrow$ B $\longrightarrow$ W	10+18+20+50=98
W $\longrightarrow$ B $\longrightarrow$ R $\longrightarrow$ Y $\longrightarrow$ W	17+25+40+20=102
W $\longrightarrow$ B $\longrightarrow$ Y $\longrightarrow$ R $\longrightarrow$ W	17+44+18+45=124
W $\longrightarrow$ R $\longrightarrow$ B $\longrightarrow$ Y $\longrightarrow$ W	15+20+44+20=99
W $\longrightarrow$ R $\longrightarrow$ Y $\longrightarrow$ B $\longrightarrow$ W	15+40+19+50=124

The above table shows 98 is the minimum time if we follow

W $\longrightarrow$ Y $\longrightarrow$ R $\longrightarrow$ B $\longrightarrow$ W Path

4. Solve the following Symmetrical transportation problem given by the following data.

C_{12} =20, C_{13}= 4, C_{14}= 10, C_{23}= 5, C_{34}= 6, C_{25}= 10, C_{35} = 6, C_{45}= 20

Where $C_{ij} = C_{ji}$

& there is no route between cities i & j if the value of for Cij not shown

Sol: First express the given problem in the form of an assignment problem by taking C_{ij}= for i=j

	1	2	3	4	5
1	∞	20	4	10	∞
2	20	∞	5	∞	10
3	4	5	∞	6	6
4	10	∞	6	∞	20
5	∞	10	6	20	∞

Subtract the smallest element in each row.

∞	16	0	6	∞
15	∞	0	∞	3
0	1	∞	2	2
4	∞	0	∞	14
∞	4	0	14	∞

Subtract the smallest element in each column.

∞	15	0	4	∞
15	∞	0	∞	3
0	0	∞	0	0
10	∞	0	∞	20
∞	10	0	20	∞

Now assignments can be made

∞	15	$\boxed{0}$	4	∞	✓
15	∞	X	∞	3	✓
$\boxed{0}$	X	∞	X	X	
10	∞	X	∞	20	✓
∞	10	X	20	∞	✓
		✓			

Now deduct the smallest element i.e. 3 and start reassignment. You can notice that again row 4 doesn't have any assignment. So marking is done accordingly.

∞	12	$\boxed{0}$	1	∞	✓
12	∞	X	∞	$\boxed{0}$	
$\boxed{0}$	X	∞	X	X	
1	∞	X	∞	9	✓
∞	$\boxed{0}$	X	9	∞	
		✓			

Now again deduct least element i.e. 1 and start assignment.

∞	11	X	$\boxed{0}$	∞
12	∞	X	∞	$\boxed{0}$
$\boxed{0}$	X	∞	X	X
1	∞	$\boxed{0}$	∞	8
∞	$\boxed{0}$	1	9	∞

The above solution is in optimum assignment solution having the assignment as below.

(1,4), (4.3), (3,1), (2,5), (5,2)

But this is not optimal solution for a traveling salesman as the cyclic fashion is not followed.

So instead of 0 switch 1

Say take (2,3) (4,5)

Now the possiblef solutions are

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

& the total cost is

10+10+10+5+4 = 49

8.4 TRANSSHIPMENT PROBLEM

A transportation problem is a problem in which available Quantity frequently moves from one source to another source or from one destination to another destination before reaching its actual destination.

i.e. here all the products need not move only from origin to destination.

Suppose there are sources S_1, & S_2 & destination D_1, & D_2, the usual way is

1. S_1-D_1 or / & S_1-D_2
2. S_2-D_1 or / & S_2-D_2

In transshipment problem the possibilities are

1. S_1-D_1 & / or S_1-D_2 &/ or S_1- S_2-D_1 or /&
   S_1-S_2-D_2 ,S_1-D_1-D_2, &/ or S_1- D_2-D_1 or /&
   S_1-S_2-D_1-D_2, S_1-S_2- D_2-D_1, S_1-D_2 -S_2-D_1 etc.
2. S_2-D_1 &/ or S_2-D_2 &/ or S_2-S_1-D_2 &/or S_2-S_1-D_1 etc.

4.4.1 Characteristics of transshipment problems

1. The No. of sources and destinations in the transportation problem are m & n respectively. In transshipment we have m +n sources and destination.

2. In the transshipment problems $S_i \longrightarrow S_j \neq S_j \longrightarrow S_i$

3. In solving the transshipment problems we first formulate the complete matrix find the optimum solution to the transportation problem and neglect whichever is not required.

4. The basic feasible solution contains 2m + 2n - 1 basic variables, if we omit the variables appearing in the m + n diagonal cells, we are left with m + n -1 basic variables.

4.4.2 Sample Problems:

A firm having 2 sources S_1 & S_2 wishes to ship its products to two destination D_1 & D_2. The No. of units available at S_1 & S_2 are 5 & 25 respectively. The firm, instead of shipping directly from sources to destinations, decides to investigate the possibility of transshipment. The unit transportation cost (in Rs). are given in the following table. Find the optimal shipping schedule.

	Source		**Destination**		
	S_1	S_2	D_1	D_2	Available
S_1	0	2	3	4	5
S_2	2	0	2	4	25
D_1	3	2	0	5	
D_2	4	4	1	0	
Demand			20	10	

Solution:

The transportation problem can be extended which includes the supply points & the demand points, since all the demand may concentrate at any one supply & demand points fictious quantity may be assumed for each of these supply (source) points & demand (destination) points. These quantities may be regarded as buffer stocks. Each of these buffer stocks must at least be equal to total supply (or demand) in the original problem. For the given problem the buffer stock will be 30 units. The role of buffer stock in the transshipment problem is only to convert the problem into a standard form and once the optimum solution is obtained, necessary adjustments will be made to uniform to the original problem.

We obtain the following transportation table as given below

	S_1	S_2	D_1	D_2	Capacity
S	0	2	3	4	35
S_2	2	0	2	4	55
D_1	3	2	0	5	30
D_2	4	4	1	0	30
Demand	30	30	50	40	150

Using Vogel's Approximation for an IBFS

	S_1	S_2	D_1	D_2
S	$\boxed{30}$ 0	2	3	$\boxed{5}$ 4
S_2	2	$\boxed{30}$ 0	$\boxed{25}$ 2	4
D_1	3	2	$\boxed{25}$ 0	$\boxed{5}$ 5
D_2	4	4	1	$\boxed{30}$ 0

Use MODI method for optimum solution to the extended transportation problem.

The final allocation is

a. Transport 25 units from S_2 to destination D_1. It increases the capacity of S_2 to 55 units including 30 as buffer stock.

b. From S_1 transport 5 units to D_2

c. Out of 25 units available in D_1 transport 5 units to D_2.

8.5. SUMMARY

Given a collection of cities and the cost of travel between each pair of them, the **traveling salesman problem, or TSP** for short, is to find the cheapest way of visiting all of the cities and returning to your starting point. In the standard version we study, the travel costs are symmetric in the sense that traveling from city X to city Y costs just as much as traveling from Y to X.

The simplicity of the statement of the problem is deceptive. The TSP is one of the most intensely studied problems in computational mathematics and yet no effective solution method is known for the general case. Indeed, the resolution of the TSP would settle the P versus NP problem and fetch a $1,000,000 prize from the Clay Mathematics Institute.

One requirement of the transportation problem is advance knowledge of the method of distributing flow from each source to each destination. This determines the cost. Sometimes, however, the best method of routing the flow is not clear because of the possibility of trans-shipments. Suppose the company decided to look at using common carriers to ship their goods (common carriers are trucking compaines for hire, as opposed to internal trucking divisions). Since no single trucking company serves all of the area, many shipments will need to be transfer to another truck at least once along the way. These transfers can be made at an intermediate cannery or warehouse, or be made at one of other junctions including supply sources.

8.6 SELF ASSESSMENT QUESTIONS

1. Solve the following traveling salesman problem

The ABC Ice cream company has a distribution depot in Greater Kailash Part 1 for distributing Ice creams in South Delhi. There are four vendors located in south Delhi A, B, C, D. The following matrix displays the distance between the depot and four vendors.

	Depot	A	B	C	D
Depot	0	3.5	3	4	2
A	3.5	0	4	2.5	3
B	3	4	0	4.5	3.5
C	4	2.5	4.5	0	4
D	2	3	3.5	4	0

Which route the Van should follow.

2. Consider the following transshipment problem with 2 sources and two destinations. The cost of shipment is given below. Determine the optimum shipping schedule

$S_1 \longrightarrow S_2 1, D_1 6, D_2 2$

$S_2 \longrightarrow S_1 1, D_1 8, D_2 2$

$D_1 \longrightarrow S_1 7$ and $D_2 3$

$D_2 \longrightarrow S_1 1, D_1 1$

Assume all other cost be zero

Supply Capacity $S_1 4$ & $S_2 5$

Demand Capacity $D_1 3$ & $D_2 6$

8.7 REFERENCE

1. Operation Research By Wagner

2. Operation Research By S.D. Sharma

3. Operation Research By Kanthi Swaroop

4. OR Magazine

NOTES

UNIT - 9

STATISTICAL QUALITY CONTROL TECHNIQUES

Structure:

9.0 Objectives

9.1 Introduction

9.2 Basic Statistical Concepts

9.3 Classification of Data

9.4 Application of Scientific / Statistical Techniques in Quality

9.5 Measurement of Quality Characteristics

9.6 Specification of Quality

9.7 Diagrammatic Presentation

9.8 Measures of Dispersion

9.9 Practical Limitation of Control Charts for Variables

9.10 Some More Examples for Practice

9.11 Probability Distribution

9.12 Notes

9.13 Summary

9.14 Keywords

9.15 Answer to Check Your Progress

9.16 Self-Assessment Question

9.17 Reference

9.0 OBJECTIVES

After studying this unit, you should be able to;

- Explain basic statistical Quality concepts.
- Measure and classify quality characteristics
- Emphasize on application of statistical technique to quality management

9.1 INTRODUCTION TO STATISTICAL QUALITY CONCEPT

A quality control system performs inspection, testing and analysis to ensure that the quality of the product / process is as per the laid down quality standards. It is called "statistical quality control" when statistical techniques are employed to control, improve and maintain the quality of the process / product. Building an information system to satisfy the concepts of 'preparation' and control and improving upon the product quality, requires statistical thinking. In the view of growing emphasis on quantitative techniques of analysis and presentations, statistics has found place in many professional course and its applications.

9.2 BASIC STATISTICAL CONCEPTS

9.2.1 The concept of variation

The concept of variation states that no two items will be perfectly identical and variation is a fact of nature. The variations may be small or large, variation exist in all processes.

The variation can be existing in following 3 ways.

- Variations within the object.
- Variations among the objects during the same period
- Variations among objects at different periods and time.

<u>Examples for variations</u>

a) Variation among the object / part / process itself

If blood pressure (BP) of a person is measured four times in a day using the same instrument by same doctor the reading may be different at each time.

b) Variation among the object at same period of time

If a BP of 5 persons are measure at the same time using same instrument and doctor, these are possibilities of getting 5 different readings.

c) Variations among objects at different period of time

If a BP of 5 persons are measured at 2 times in a day using same instrument and doctor, these could be a different reading at each time of measurement.

9.2.3 Causes / Reasons for Variations

a) Variations due to assignable causes

b) Variations due to unassignable causes

9.2.3.1 Variations Due to Assignable Causes

These variations are greater magnitude as compared to those due to chance causes and can be easily traced or detected. The variations due to assignable causes may be because of following factors –

a) Different among machines

b) Difference among persons using machines

c) Difference among materials

d) Difference in each of these factor overtime

e) Difference in their relationship to one other

These variations may also be caused due to change in working conditions, mistake on the part of person operating machine, lack of quality etc.

Example: The temperature of patient measured using thermometer could be varied at each time of measurement. This could be due to different thermometer used or may be due to different viewers of thermometer or interactions of both. One can trace the source of variability and efforts can be initiated to reduce / avoid this.

9.2.3.2 Variations Due to Un assignable Causes

These causes are inevitable in any process or product. They are difficult to trace and difficult to control them even under best conditions. Since, these variations may be due to some inherent characteristic of the process or machine which functions at random.

Example: when a person stands on weight measuring machine, suddenly the needle sometimes moves forward and comes back to show actual weight of person.

9.3 CLASSIFICATION OF STATISTICAL DATA

Statistical data can be classified as

- Variable data
- Attribute data

9.3.1 Variable Data

When the record is made of an actual measured quality characteristic, such as blood pressure, glucose level of person, the quality characteristic is said to be expressed in variables.

For example: BP level of person at an hour measured over 4 days. The recorded values are called variable data.

9.3.2 Attribute Data

When record show acceptable and non-acceptable characteristic to confirm to any specified requirements, it is said to be recorded by attributes. These characteristics may be checked in inspection observations or test.

For example: Segregation of patients as fevered or not based on temperature felt by doctor using hand touch without measuring them. Segregating the people as physically challenged or not be viewing their personality of physical.

9.4 APPLICATION OF SCIENTIFIC / STATISTICAL TECHNIQUES IN QUALITY

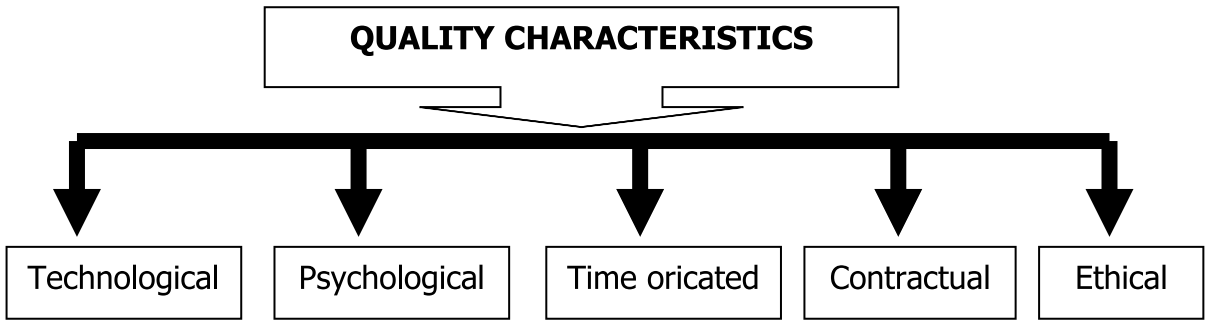
- Setting proper standards and product specifications
- To setup reliable method of quality measurement
- Measuring and reporting the product/process reliability and making needed changes]

9.4.1 Quality Characteristics

A person can be described in terms of his age, height, weight and physical appearance etc. Any one of these factors or some of these factors can be used to define the person and these factors are known as quality characteristics of persons. Similarly, an Apple can be described in terms of its weight, color, taste etc. and each factor describes the quality of Apple and is known as quality characteristics of Apple.

Hence, a physical or chemical or any such property or any other requirements used to define the nature of product / service which contributes to fitness for use is a quality characteristic. Thus, a medicine may be defined by stating the quality characteristic such as, the type like capsules, tablets, or injections, its chemical composition, presentations, strength, weight etc.

9.4.2 Classification of Quality Characteristics



Some of the features like age, height, weight, etc, are used in case of person's quality characteristic. In case of medicine, its type, composition, weight, presentation etc. are technological quality characteristics.

- Psychological quality characteristics are sensory related. Examples, taste of food, odor, feeling etc.
- Time ordinated quality characteristics includes reliability, life etc.
- Contractual quality characteristics involves features like guarantee, safety etc.
- Ethical quality characteristics involves factors like honesty and integrity etc.

9.5 MEASUREMENT OF QUALITY CHARACTERISTICS

Some of the quality characteristics can be measured directly while some of them cannot be measured directly.

A) **Directly Measurable Quality Characteristics**

These quality characteristics can be measured directly using an instrument or with the help of an equipment.

Example: Percentage of gold in ornament, Mail age, efficiency of a car, wattage of a bulb

B) **Non-measurable quality characteristics**

These quality characteristics cannot be measured directly.

Example: looks of car, colour a vehicle, finishing of a Road

9.6 SPECIFICATION OF QUALITY

The demands of application are translated in to the requirements and requirements are quantified. These quantified requirements are called specifications. The specifications house the list of essential / needed characteristics and their tolerance. The matter of quality specification may include:

(a) Material specification (example – ingredients)

(b) Process/test specification etc.

The specifications could be standard, customer and company specifications.

- Standard specifications are formulated by standard bodies. In India, we have Bureau of Indian Standards to formulate this.
- Sometimes customer provides his needs or requirements in the form of specification, which is called customer specification.
- When a company manufactures products to its own specification due to varied constraints and customer accept them, the specifications may be called manufactures / company specification.

9.7 DIAGRAMMATIC PRESENTATION

Diagrams are the visual form for presentation of statistical data. Diagrams refer to various types of bars, circles, maps, pictorials etc. which are strictly speaking not graphic. Diagrams may be bound by two axis or may be more than two axes.

Diagrams helps us to see the pattern and shape of the complex situation. Diagrams helps us to visualize the whole meaning of numerical complex at a single glance.

For Example, model of a house or a machine

9.7.1 Objectives of Diagrammatic Presentation

a) To have attractive presentation

b) To have visual impact in better way

c) To make data simple

d) To depict the characteristics of data

e) To facilitate comparison

9.7.2 Limitations of diagrams

a) Utility of some diagrams is only for experts especially for multi-dimensional one.

b) Diagram shows approximate values and cannot be analyzed further.

c) All diagrams are not simple.

d) Too many details cannot be present through diagram without loss of clarity.

e) Diagrams supplements tabular presentation and not an alternative for that.

9.7.3 Run charts

Run charts sometimes called as line graphs displays process performance over the time upward and downward trends, cycles and large aberrations may be spotted and investigated further. In Run chart events are plotted on X-axis (time period) and values are plotted on Y-axis. Run charts are most useful to know the process trend and can be used to track improvements that have been put into place, checking to determine their success. An average line can be introduced in Run chart to know the spread of the process with reference to mean value of the process. Several variables may be tracked in single chart, with each variable having its own line, then chart becomes multiple Run chart.

Illustration -1

Month	No.of patients infected after surgery
January	4
February	6
March	1
April	2
May	4
June	6

Run Chart

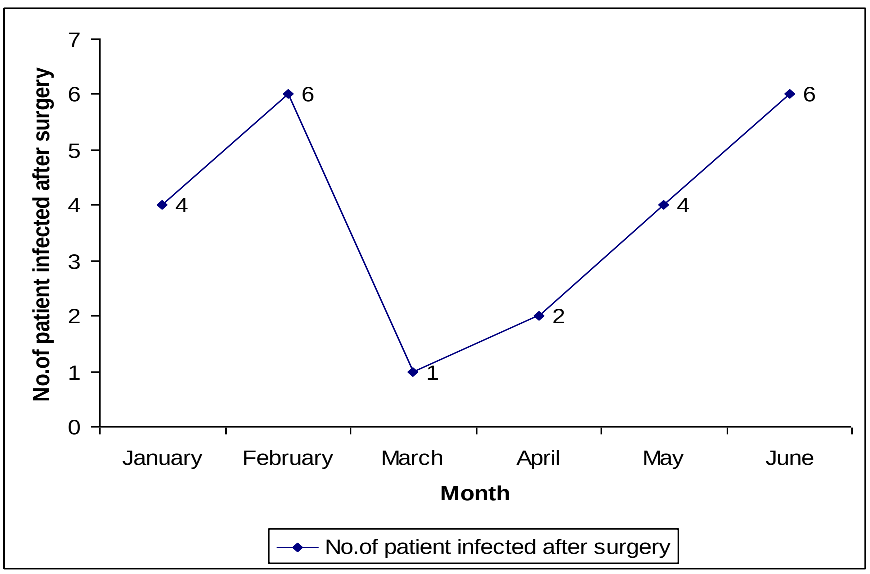
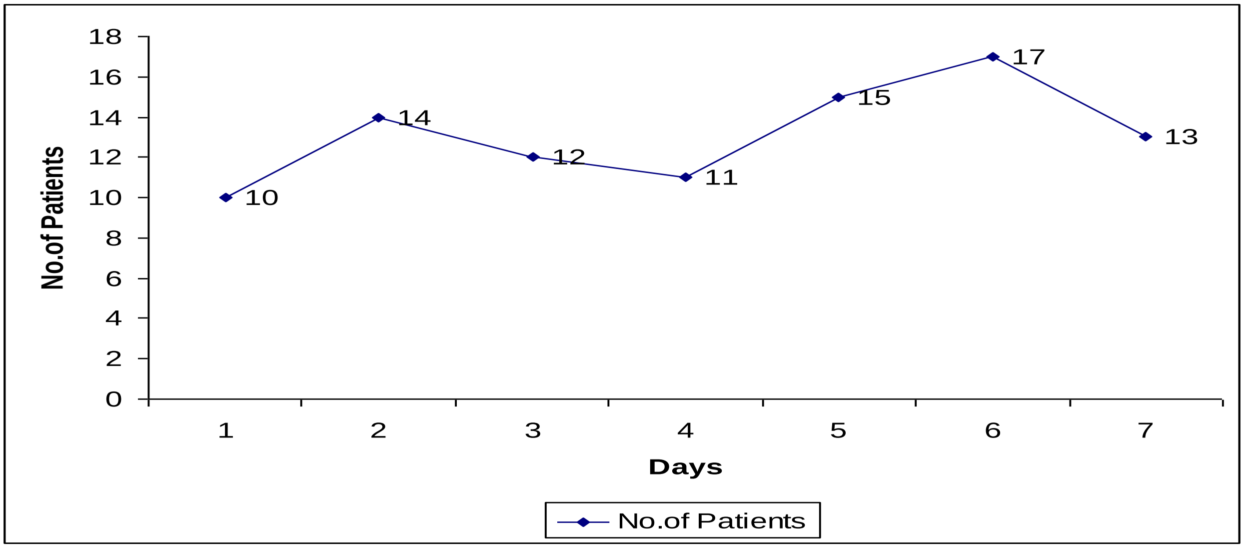


Illustration-2

The following data show the number of patients undergone x-ray test in a week.

Day	1	2	3	4	5	6	7
No.of Patients	10	14	12	11	15	17	13



9.7.3.1 Indication of Run chart

a) Non-linear variability.

b) Highest value and lowest value

c) Indicates time period at which highest and lowest value occurred.

d) Reasons for highest / lowest on time period can be analyzed.

9.7.3.2 Misinterpretation of Run chart

➔ Sometimes reader could conclude some trends which could be normal process variation as every process has its own variability.

➔ Reader some time may not recognize the trend.

9.7.3.3 Following tips can be used for better reading of chart

➔ Look at data for long period of time so that usual range of variation is evident.

➔ Look at recent data to ensure within usual range of variation.

9.8 MEASURES OF DISPERSION

The extent to which the data is scattered about the zone of central tendency is known as dispersion. Dispersion is the measure of the variation of the items.

Purpose of measuring variation

- To test the reliability of an average.
- To serve as a basis for control of variability.
- To compare two or more series.
- To facilitate as a basis for further statistical analysis.

These are several measures of dispersion out of which following are important.

- Range
- Quartile deviation
- Mean deviation
- Standard deviation
- Variance

9.8.1 Range

Range is the simplest measure of dispersion. Range is defined as the difference between the value of the smallest observation and the value of the largest observation present in the distribution.

Range = L – S

Where L = Largest value

S = Smallest value

From a grouped frequency distribution range is the difference between the upper limit of the highest class and the lower limit of the smallest class.

Range = Upper limit of the highest class – lower limit of the lowest class.

9.8.2 Coefficient of Range

The range calculated is not useful for comparison if the observations are in different units. For example, it is no possible compare the range of the weights of students with range of their heights, as the range of weights would be in kilograms and that of heights in centimeter. Therefore, for the purpose of comparison, a relative measure of range is required, which is called coefficient of range.

$$\text{Coefficient of Range} = \frac{\text{Largest value} - \text{smallest value}}{\text{Largest value} + \text{smallest value}}$$

$$= \frac{L-S}{L+S}$$

Example

A) Find the range for weight of 5 students 30, 32, 30, 38, 34

Solution

Range (R) = L – S = 38 – 30 = 8

Coefficient of range = $\frac{L-S}{L+S} = \frac{8}{38+30} = \frac{8}{68} = 0.117$

Example 1: The range of marks obtained by the students of a class having a total strength of 50 may be significantly less than the range of marks obtained by the students in an All India Level Exam like JEE. Since no of students who have taken exams also play an important role here.

Example 2: If the range of share price in a day of a share in the price of range of 1000/- is 50 we may say it is stable. But if the same range is seen for a share of value 200 we cannot say it is stable.

Hence we have to use coefficient of range as a measure of dispersion.

9.8.2.1 Merits and Limitations of Range

Merits

- Range is simple to understand and easy to compute.
- Range is the quickest way to get a measure of dispersion.

Limitations

- It is not based on all the observations in the data. It is computed based on highest and lowest values.
- It is influenced by extreme values and hence fluctuates from sample to sample of a population.
- Range cannot be computed from frequency distributions with open-end class.
- Range fails to explain about the character of the distribution.
- Range is unreliable because it is affected by extreme values.

Uses of Range

- Range is used in industry for the quality control of products without 100% inspection. Range plays an important role in construction of control charts used for quality control.
- Range is useful in studying variations in process quality.

9.8.3 Control Charts

A control chart is a graphical representation of collected information. The information may pertain to measured quality characteristics of judged quality characteristics of samples. It detects the variation in processing and warn if there is any departure from the specified limits. It also specifies the state of statistical control. It indicates natural capability of the process which helps for comparison and decision making between alternate process.

Commonly used control charts

a) Control charts for variables - $\bar{x}$ & R chart

b) Control chart for attributes – P –chart & C-chart.

9.8.3.1.1 Control charts for variables ($\bar{x}$ & R chart)

Control chart based on measurement of quality characteristics are called control chart for variables. $\bar{x}$ chart & R chart are used in combination for the control process. $\bar{x}$ chart shows the centering of the process & R chart shows uniformity or consistency of the process. Upper and lower control limits of the control chart show the maximum and minimum values that the process can be attainable, exceeding which process needs to be corrected.

9.8.3.1.2 Chart Construction Procedure

STEP – 1 : Calculate average ($\bar{x}$) Range of each sub group.

STEP – 2 : Calculate grand average ($\bar{\bar{x}}$) and Average Range ($\bar{R}$)

STEP – 3 : Calculate 3σ limits on control chart for $\bar{x}$ - chart

STEP – 4 : Calculate 3σ limits on control chart for $\bar{R}$ - chart

STEP – 5: Interpret the process using control chart.

Illustration 1

An operator has to keep the length of parts in the order of 19 ± 2 mm. Following are measured values. Construct $\bar{x}$ and R chart and give the conclusion.

	8:00 am	**9:00 am**	**10:00 am**	**11:00 am**	**12:00 noon**
1st day	17	16	16	18	20
2nd day	22	20	18	17	16
3rd day	20	20	18	18	18
4th day	18	18	18	22	10

Tabulating the same in a form as given below.

Above table indicates that four samples containing five sub-samples are chosen for plotting control chart.

Sample	**1**	**2**	**3**	**4**	**5**	$\bar{x}$ **- average**	**R Range**
1	17	16	16	18	20	17.4	4
2	22	20	18	17	16	18.6	6
3	20	20	18	18	18	18.8	2
4	18	18	18	22	20	19.2	4

Average for each sample is computed as follows –

a) $\bar{x}_1 = \dfrac{17+16+16+18+20}{5} = 17.4$

$$\bar{x}_2 = \frac{22+20+18+17+16}{5} = 18.6$$

Similarly - $\bar{x}_3, \bar{x}_4 \,\&\, \bar{x}_5$ are computed.

b) Range for each sample is calculated. Range = Large value – small value.

$R_1 = L_1 - S_1 = 20 - 16 = 4$

$R_2 = L_2 - S_2 = 22 - 16 = 6$

Similarly, $R_3 = 2$ and $R_4 = 4$

c) Grand average and average range is computed.

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4}{4} = \frac{17.4+18.6+18.8+19.2}{4} = 18.5$$

$$\bar{R} = \frac{R_1 + R_2 + R_3 + R_4}{4} = \frac{4+6+2+4}{4} = \frac{16}{4} = 4$$

d) Computation of 3 σ limits for $\bar{x}$ and R chart

Upper Control Limit UCL $\bar{x} = \bar{\bar{x}} + A_2\bar{R}$ and

Lower Control Limit LCL $\bar{x} = \bar{\bar{x}} - A_2\bar{R}$

Control limit for $\bar{x}$ chart

Upper Control Limit UCL $R = D_4\bar{R}$

Lower Control Limit LCL $R = D_3\bar{R}$

A_2, D_3 and D_4 are the constants for the given sub-samples size which can be obtained from statistical table.

Now, for given sub-sample size (n) = 5, from statistical table A_2, D_3 and D_4 are found as 0.58, 0 and 2.11 respectively. Similarly to find SD of process σ, $\sigma = \frac{\bar{R}}{d_2}$ equation can be used, where d_2 is a constant, which can be obtained from statistical table for the given sub-sample size for n=5, d_2=2.326

$$\sigma = \frac{\bar{R}}{d_2} = \frac{4}{2.326} = 1.719$$

Control limits for $\bar{x}$ chart

VCL $\bar{x} = \bar{\bar{x}} + A_2\bar{R} = 18.5 + 0.58x4 = 20.82$

LCL $\bar{x} = \bar{\bar{x}} - A_2\bar{R} = 18.5 - 0.58x4 = 16.18$

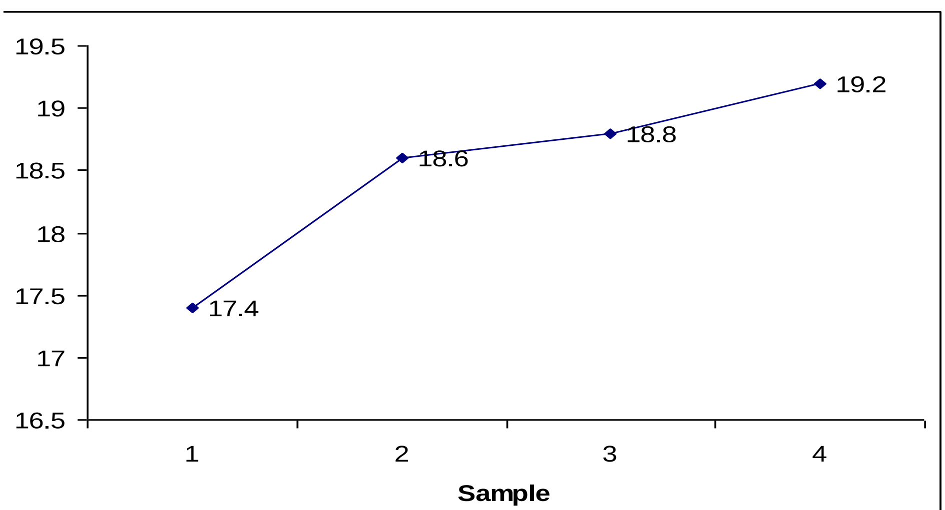
Control limit for R-chart

UCL R = $D_4\bar{R}$ = 2.11 x 4 = 8.44

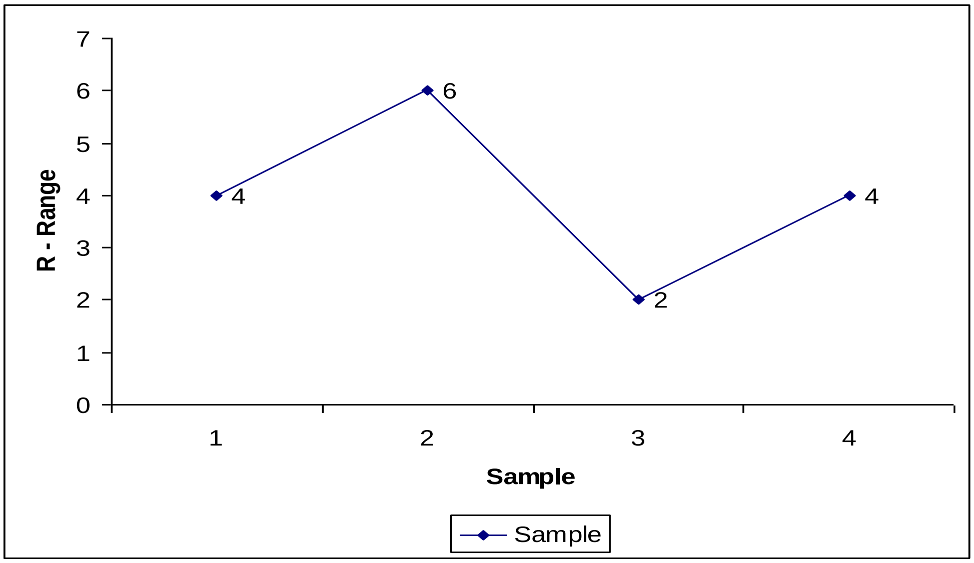
LCL R = $D_3\bar{R}$ = 0 x $\bar{R}$ = 0

Plot the Control Chart

$\bar{x} - chart$



R-chart



Conclusion from the chart: Since no points cross the control limits, we can say that the process is within the state of control.

9.8.3.1.2 Control chart for attribution

When data is collected on the basis of conforming or non-conforming (i.e segregation basis) to the specification then attribute charts are most appropriate to use.

9.8.3.1.3 Commonly Used Attribute Charts

P – Chart

It is most appropriate whenever sub-group size is variable and suitable to depict fractional defectives.

The control limits of P – chart is given by –

UCL $P = \overline{P} + 3\sqrt{\dfrac{\overline{P}\left(1-\overline{P}\right)}{n}}$

LCL $P = \overline{P} - 3\sqrt{\dfrac{\overline{P}\left(1-\overline{P}\right)}{n}}$

Where $\overline{P}$ = average fraction defective

= Total No. of defective during a period ÷ Total No. of inspected during a period.

Illustration

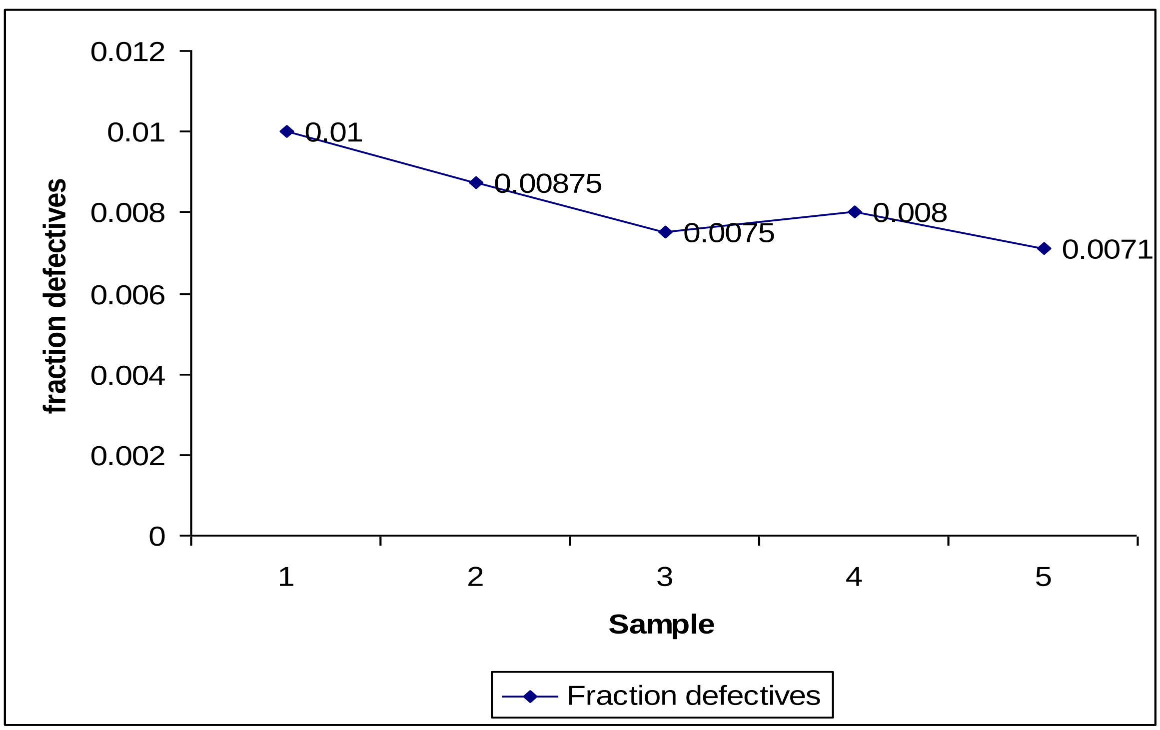
Month	No. of Parts Produced	Defective	Fraction defectives
1	500	5	0.0100
2	800	7	0.00875
3	400	3	0.0075
4	500	4	0.008
5	700	5	0.0071
	2900	24	

The average sample size (n) = 2900 ÷ 5 = 580

The average fraction defectives ($\overline{P}$) = 24 ÷ 2900 = 0.0082

$$\text{UCL } P = \overline{P} + 3\sqrt{\frac{\overline{P}(1-\overline{P})}{n}} = 0.082 + 3\sqrt{\frac{0.082x0.9918}{580}} = 0.019$$

$$\text{UCL } P = \overline{P} - 3\sqrt{\frac{\overline{P}(1-\overline{P})}{n}} = 0.082 - 3\sqrt{\frac{0.082x0.9918}{580}} = -3.0x10^{-3} = 0.00$$



Indication: No point exceeds process control limits and hence process is under control.

U – Chart

When subgroup size varies from sample to sample, it is necessary to use U-chart. Control limits for U-chart.

VCL $U = \overline{U} + 3\sqrt{\overline{U}/n}$

LCL $U = \overline{U} - 3\sqrt{\overline{U}/n}$

Illustration 2

Following table refers to the average number of wrong supply of parts to production for 10 lots of 100 parts each.

Lot No.	**No. of wrong supply parts**	**Wrong supply per lot of 100 parts (U)**
1	15	0.15
2	17	0.17
3	12	0.12
4	16	0.16
5	14	0.14
6	5	0.05
7	14	0.14
8	11	0.11
9	9	0.09
10	10	0.10
		1.23

Control limit calculations

$$\overline{U} = \frac{\sum U}{n} = \frac{1.23}{10} = 0.123$$

UCL $U = \overline{U} + 3\sqrt{\overline{U}/N} = 0.123 + 3\sqrt{0.123/100} = 0.228$

LCL $U = \overline{U} - 3\sqrt{\overline{U}/N} = 0.123 - 3\sqrt{0.123/100} = 0.018$

Conclusion: Since all 'U' values are within control limits; process is in control.

9.9 PRACTICAL LIMITATION OF CONTROL CHARTS FOR VARIABLES

$\overline{X}$ and R charts can be used for quality characteristics that can be measured and expressed in number. Many quality characteristics can be observed only viz, cracks in casting, blowholes, finish surface etc.

Further $\overline{X}$ and R charts can be used only for one measurable characteristics at a time –

If inspection following gauging techniques like inspection by GO-NOGO gauges for the economic purpose, then X and R charts may be plotted for most important and troublesome quality characteristics.

9.9.1 Fraction Defective

Fraction defective 'P' may be defined as the ratio of number of defectives articles found in any inspection to the total number of articles actually inspected. Fraction defective is always expressed in decimal fraction.

9.9.2 Comparison of X, R and P chart

- P chart is an attribute control chart : $\overline{X}$ and R chart is variable chart i.e. P chart for gauging inspection and $\overline{X}$, R chart for measurable quality characteristics are suitable respectively.
- The cost of collecting P chart is less than $\overline{X}$ and R
- The cost of computing and charting may also be less for P chart than cost of $\overline{X}$, R chart as P chart can be applied to any number of quality characteristics.
- P – chart is best suited for cases where the classification (sorting) the articles as good or bad.
- P chart though discloses the presence of assignable causes of variations, it is not sensitive as $\overline{X}$ and R chart
- The sample size generally larger for P chart than $\overline{X}$, R chart

9.9.3 Purpose of P chart – following purposes

a) To discover average proportion of defective articles.
b) To bring attention of Management about any changes of inferior quality.
c) In sampling inspection of large lots of purchased articles.

9.9.4 Comparison Between Attribute and Variable Control Chart

VARIABLE CHART	ATTRIBUTE CHART
Example $\overline{X}$, R	P, np, c, u charts
Type of data required – variable i.e, measurable value of characteristics	Attribute data (using Go and No Go gauge)
Field of application – control of individual characteristics	Control of proportion of defective or number of defects or number of defects or number of defects / unit area
Advantage a) Provides maximum utilization of information / data b) Provides detail information on process average and variation of industrial dimensions	Data required are often already available from records. Easily understandable by all persons It provides overall quality history
Disadvantage a) Not easily understood unless trained b) Can cause confusion between control charts c) Cannot be used for gauge inspection	a) They do not provide detailed information b) They do not recognize different degree of defectiveness

9.10 SOME MORE EXAMPLES FOR PRACTICE

1. Develop the trial control limits and interpret the results. The sample data is presented in below Table and each sample contains five sub-samples. You are given $A_2 = 0.73$, $D_3 = 0$ and $D_4=2.11$. You are required to revise the trial control limits if required to a maximum of two iterations. Clearly mention the inference.

Samples	1	2	3	4	5	6	7	8	9
Mean Value of the quality characteristics	6.96	6.40	6.35	6.39	6.50	6.85	6.30	6.34	6.01
Differences value among 5 sub-samples	0.10	0.18	0.20	0.35	0.20	0.18	0.13	0.1	0.16

Sample size	32	32	50	50	32	80	50	50	32	32

Answer: Means of Means = 6.455556. Average Range = 0.177778

Average Chart:

UCL = 6.4155 + 0.73 X 0.177778 = 6.545278.

LCL = 6.4155 – 0.73 X 0.177778 = 6.285722

Range Chart:

UCL = 2.111 X 0.177778 = 0.375289

LCL = 0 X 0.1555 =0.000

Result: Sample 1, 6 and 9 is out of control.

2. Develop the trial control limits for the sample data given below and each sample contains 100 sub-samples. Also interpret the results. Revise the control limit if required to a maximum of two iterations.

Samples	1	2	3	4	5	6	7	8	9
No. of Defectives	5	7	9	11	15	8	2	1	0

Average P = 54/900 = 0.06.

CL = 0.06 ± 3 x √ (0.06 x 0.96)/ 100

UCL = 0.131 and LCL = -0.11 ≈0.00

Note: Sample Number 5 is out of statistical control limit. Revision required.

3. Develop the trial control limits for the sample data given below and each sample contains 100 sub-samples. Also interpret the results.

Sample size	32	32	50	50	32	80	50	50	32	32
No. of Defectives	2	3	3	2	1	4	2	0	2	1

Soln.

No. of Defectives	2	3	3	2	1	4	2	0	2	1
Fraction Defectives	0.0625	0.094	0.06	0.04	0.03	0.05	0.04	0	0.062	0.031

Average P = 0.469/10 =0.0469

UCL = 0.0469+ ***3 x*** √ ***(***0.0469 ***x 0.9531)/ 32 =0.159***

UCL = 0.0469- ***3 x*** √ ***(***0.0469 ***x 0.9531)/ 32 =0.065***

4. Develop the trial control limits and *revise* the control limits if required for the sample data given below and each sample contains 100 sub-samples. Also interpret the results.

Samples	1	2	3	4	5	6	7	8	9
No. of Defectives	5	7	9	11	15	8	2	1	0

Average P = 54/900 = 0.06.

CL = 0.06 ± 3 x √ (0.06 x 0.96)/ 100

UCL = 0.131 and LCL = -0.11 ≈0.00

Note: Sample Number 5 is out of statistical control limit. Revision is required.

5. Develop the trial control limits and plot the suitable control chart and interpret the results. The sample data is presented in below Table and each sample contains five sub-samples. You are given A_2 = 0.73, D_3 = 0 and D_4=2.11.

Samples	1	2	3	4	5	6	7	8	9
Mean Value of the quality characteristics	6.36	6.40	6.35	6.39	6.50	6.55	6.30	6.34	6.55
Differences value among 5 sub-samples	0.10	0.18	0.20	0.15	0.20	0.18	0.13	0.1	0.16

- Means of Means = 6.4155.   Average Range = 0.1555
- Average Chart: UCL = 6.4155 + 0.73 X 0.1555 = 6.529:: LCL = 6.4155 - 0.73 X 0.1555 = 6.3019
- Range Chart: UCL = 2.111 X 0.1555 = 0.3281 :: LCL = 0 X 0.1555 =0.000

Result: Sample 5 and 9 is out of control.

Check Your Progress

1. Variation can cause due to

   A. Assignable Causes   B. Unassignable Causes   C. Both   D. None

2. Statistical data can be classified in to

   A. Variable and attribute   B. Range and standard deviation C. Primary and Secondary data

   D. P chart and R Chart

3. The objective of diagrammatic representation is to

   A. More accuracy   B. Visual impact   C. make data complicate   d. None of the above

4. Line chart is also called as run chart (True/False)

5. Control limits decides acceptance or rejection of samples (True/ False)

9.11 PROBABILITY DISTRIBUTION

The distribution which are based on actual data or experimentation are called observed frequency distribution and the distribution based on expectations on the basis of past experience in known as theoretical / expected frequency distribution.

9.11.1 Classification of important distribution

a) **Continuous Probability Distribution**

A probability distribution of the variable data is known as continuous probability distribution. The important continuous probability distributions are –

- Normal distribution
- Exponential distribution
- Weibull distribution

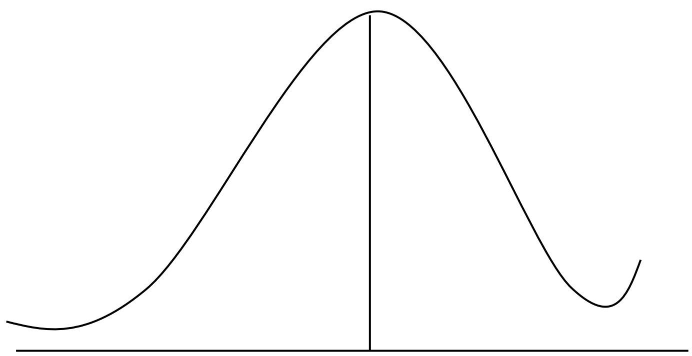
b) **Discontinuous Probability Distribution**

A distribution is discontinuous or discrete if specific values such as integer are present in the data. So, sin discrete data intermediate values cannot occur. Important discrete distribution is

- Hyper geometric distribution
- Binomial distribution
- Poisson distribution

c) **The Normal Probability Distribution**

Many naturally occurring distribution such as heights of adult makes, weight, dimension of part etc. conform the approximately the same shape. This characteristic distribution is belt shaped and symmetrical as shown in figure below.



$\bar{x} = \text{mod}\,e$

= median

A normal distribution is determined by the parameters, mean and standard deviation. For different values of mean and standard deviation, we get different normal distribution. The area under normal curve is always taken as unity so as to represent the total probability. The area has its own importance as it represents frequency of variable.

9.11.2 Importance of normal curve

- Normal distribution finds considerable application in theory of statistical quality control.
- Helps in reliability engineering applications.
- Most distribution occurring in practice like bionomical, Poisson can be approximated by a normal curve.

Important characteristics of normal curve

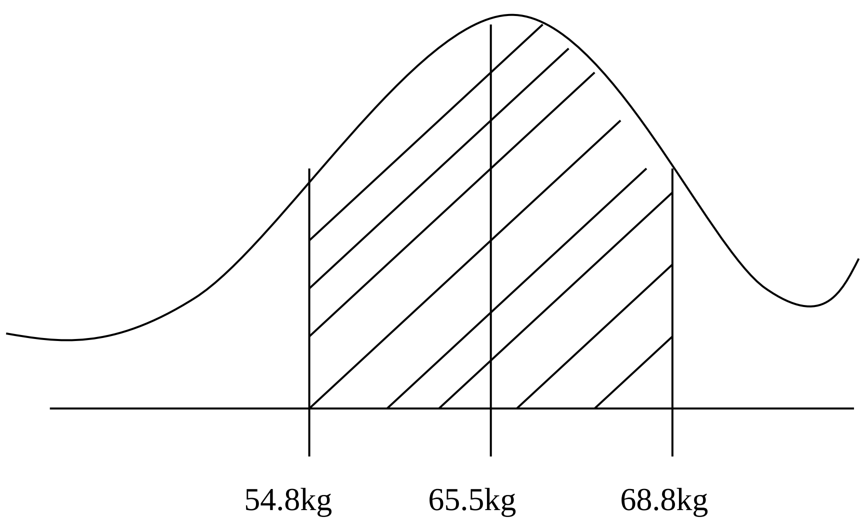
- The curve is symmetrical and bell shaped curve.
- The value of mean, median and mode will coincide because the distribution is symmetrical and single peaked.
- It has only one mode occurring at μ; It is unit-modal.

- It is governed by mean and standard deviation of the variable.
- It has two points of inflexion.
- Area under normal curve is always taken as unity to represent total probability. The area under normal curve between the mean and any other value is a distribution can be found using a normal curve table. The area refers to a proportion of the total number of items in the distribution.
- Standard normal distribution (Z) = $\dfrac{\text{var}\,iable - Mean}{s\tan darddeviation}$, $Z = \dfrac{xi - \overline{X}}{\sigma}$

9.11.3 Illustration and application of normal distribution

The male adults weight measurements which are large and follows normal distribution is distributed with a mean weight of 65.5kg and standard deviation of 6.2kg. Find the percentage of observations that fall between 54.8 kg and 68.8 kg. Given that (Z=-1.73, area = 0.4582 and for Z=0.53 is 0.2019)

X_i = variable, $\bar{x} = meanvalue, \sigma =$ standard deviation



It should be noted that area left of $Z = 0$ is 0.5 And area right of $Z = 0$ is 0.5

a) When x = 54.8kg, $Z = \dfrac{xi - \bar{x}}{\sigma} = \dfrac{54.8 - 65.5}{6.2} = -1.73$

from data area between $Z = 0 \,\&\, -1.73 = 0.4582$

b) when x = 68.8kg, $Z = \dfrac{68.8 - 65.5}{6.2} = 0.53$

From data area between $Z = 0$ and 0.53 is = 0.2019

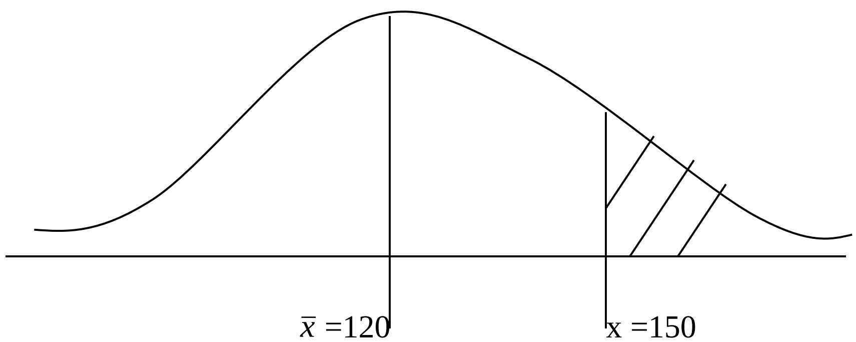
Total area between x = 54.8 to 68.8kg = (a) + (b)

$$= 0.2019 + 0.4582 = 0.6601$$

Percentage of Measurement = 0.6601 x 100 x = 66.01%

<u>Illustration (2)</u>

The customer accounts of a store have an average balance of ₹.120 and standard deviation is ₹.40. If the account is normally distributed. What proportion of account is over ₹.150? Given that area between 0 & 0.75 is 0.2734



$$Z = \frac{xi - \bar{x}}{\sigma} = \frac{150 - 120}{40} = 0.75$$

To find out proportion of accounts greater than ₹.150, it requires to refer to right of $Z = 0.75$

Area to right of $Z = 0$ is 0.5

Area between $Z = 0$ and 0.75 is 0.2734

The area to the right of $Z = 0.75$ is given by

$$Z = 0.5 - 0.2734 = 0.2266$$

Therefore, 22.66% of accounts will have balance in excess of ₹.150 (shaded area).

9.11.3 SOME MORE PRACTICE PROBLEMS

1. The University Admission Test (UAT) scores are normally distributed with a mean of 300 and a standard deviation of 25.

Determine the percentage of students scored more than 350.

Find the proportion of the students scored 300 marks.

Estimate percentage of the students scored between 220 and 260 marks.

i) Z=(x- μ)/σ=(350-300)/25=2.Z(2)=0.9772.i.e.Prob. (More than 350) =1 -0.9772 = 0.0228.i.e.,2.28%

ii) Z=(x- μ)/σ=(300-300)/25 =0.0 Therefore, 0.5 x 100 = 50 percent of Workers.

iii) $Z=(x-\mu)/\sigma=(220-300)/25=-3.2$:: $Z=(x-\mu)/\sigma=(260-300)/25=-1.6$. $Z(-3.2)=0.4903$ = 49%
$Z(-1.6)=0.4452$. Therefore, $Z(-3.2)- Z(-1.6)=0.4452-0.4903= 0.0541$. i.e., 5.41 percent have score between 220 and 260.

2. The life of the Indicator Lamp Bulb of a specific brand is normally distributed with mean 42 hours and a standard deviation of 24 hours.

I. Determine the percentage of bulbs survive more than 58 hours.
   Z = 0.667 Area = 0.5-0.2476 = 0.2524 i.e., 25.24%

II. Estimate percentage of bulbs survives between 30 and 66 hours. = 0.1915+0.3413
   = 0.5328 = 53.28%

3. The life of the ball bearing is normally distributed with mean life of 75 hours and standard deviation of 20 hours. Using the properties of normal distribution,

I. Determine the percentage of ball bearing have life more than 90 hours.

II. Determine the percentage of ball bearing have life less than 50 hours.

III. Estimate percentages of ball bearing have life between 55 hours and 85 hours.

- P (More than 90 hour) = (90-75)/20=0.75 Z (0.75) =0.2734.
  P (More than 90 hour) = 0.5 -0.2734 = 0.227=22.7%
- P (Less than 50 hours) = (50-75)/20=1.25 Z (1.25) =0.3944.
  P (Less than 50 hours) = 0.5 -0.3944 =0.1056 = 10.56 %.
- P (55 hours) = 55 - 75 / 20 = -1: Z (-1) = 0.3413
  P (55 hours) = 85 - 75 / 20 = 0.5: Z (0.5) = 0.1915
  P (Between 55 and 85 hours) = .3413+.1915 = 0.5408 =54%

4. The mean wages of daily employed workers in a large construction works is follows normal distribution and is ₹ 300 with standard deviation of ₹ 25. Find the probability of following that

i) The number of workers earning more than ₹ 350.

ii) The proportion of workers earning between ₹ 300 or less.

iii) The proportion of workers earning between ₹ 220 and ₹ 260. 10 marks

Given data: μ =300, σ = 25 and x = 350.

i) $Z= (x-\mu)/\sigma = (350-300)/25=2$.Z(2) =0.9772. i.e. Prob. (More than 350 hour) =1 -0.9772 = 0.0228.

ii) $Z= (x-\mu)/\sigma= (300-300)/25 =0.0$ Therefore, 0.5 x 100 = 50 percent of workers.

iii) $Z = (x - \mu)/\sigma = (220-300)/25 = -3.\ 2$: $Z = (x - \mu)/\sigma = (260-300)/25 = -1.6$. Z (-3.2) =0.4903. Z (-1.6) =0.4452

Therefore, Z (-3.2)- Z (-1.6) =0.4452-0.4903= 0.0541. i.e., 5.41 percent of workers earning between ₹ 220 and ₹ 260.

9.12 NOTES

9.13 SUMMARY

Growing emphasis on quantitative techniques of analysis and presentations, statistics has found place in many professional course and its applications. The concept of variation states that no two items will be perfectly identical and variation is a fact of nature. The variations may be small or large, variation exist in all processes. Diagrams helps us to see the pattern and shape of the complex situation. Diagrams helps us t visualize the whole meaning of numerical complex at a single glance.

9.14 KEYWORDS

Normal Distribution, Control Charts, Dispersion, Variability

9.15 ANSWER TO CHECK YOUR PROGRESS

1. C 2. A 3. B 4. True 5. True

9.16 SELF-ASSESSMENT QUESTIONS

1. Explain briefly the basic statistical quality concepts
2. Explain the classification of Data?
3. Explain the classification of quality classification?
4. Compare and contrast different types of charts used in statistical analysis?

9.17 REFERENCES

- Douglas C. Montgomery- Introduction to Statistical Quality Control-6th edition
- Dr. M. Mahajan- Statistical Quality Control- Dhanpath Rai Publishing Co Pvt Ltd

UNIT – 10: MARKOV CHAIN AND ANALYSIS

Structure:

10.0 Objectives

10.1 Introduction

10.2 Markov Chain and Analysis

10.3 The model

10.4 Assumptions in Markov Analysis

10.5 Stepped Transition Probabilities

10.6 Hidden Markov Model

10.7 Notes

10.8 Summary

10.9 Keywords

10.10 Answer to Check Your Progress

10.11 Self-Assessment Question

10.12 Reference

10.0 OBJECTIVES

After studying this unit, you would be able to:

- Define the meaning of Markov analysis and its application
- Analyse the advantages and disadvantages of Markov analysis
- Explain different Morkov models

10.1 INTRODUCTION

Russian mathematician Andrei Andreyevich Markov, pioneered the study of stochastic processes (involves operation of chance). Markov first applied this method to predict the movements of gas particles trapped in a container. Markov analysis is a method used to forecast the value of a variable whose predicted value is influenced only by its current state, and not by any prior activity. Markov chains are simple and common statistical model random processes and are popularly used in various domains, ranging from text generation to financial modelling. Markov Chains are conceptually intuitive, manageable and can be implemented without using advanced statistical or mathematical concepts.

10.2 MARKOV CHAIN AND ANALYSIS

Consider the happening event, if the next event is based on the preceding event, then we can apply Markov analysis there. Markov chain, which is also called as Markov process, describes the possibility of happening of the succeeding sequence of events based on the previous unit. The process of Markov analysis involves defining the likelihood of future action, which is dependent upon the current state of a variable. Decision tree can be drawn after determining the probabilities of future actions.

Application of Markov Analysis

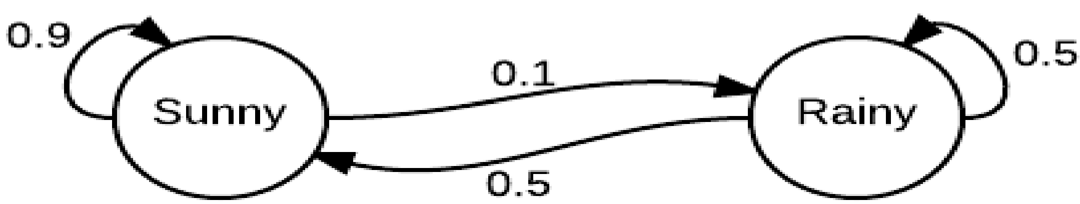
Markov analysis is adopted by experts across segments like, eeducators, medical practitioners, financial analysts, Quality Inspectors etc.

1) It is often employed to predict the number of defective pieces from assembly line
2) Proportion of a company's accounts receivable (AR) that will become bad debts in day to day business activities.

3) To forecast future brand loyalty of current customers and the outcome of these consumer decisions on a company's market share.
4) To predict stock price and option price too.
5) Predicting traffic flows, communications networks, genetic issues, and queues etc.
6) The PageRank of a webpage as used by Google is defined by a Markov chain.
7) To examine hypothesized differences between students', use of counselling skills in an introductory course.

A visualization of the weather example

Imagine that there are two possible states for weather: sunny or rainy. You can always directly observe the current weather state, and it will be one of the two states. You can predict the weather of next 5 minutes, 10 minutes etc. For you to predict what the weather will be like tomorrow, intuitively, you assume that there is an inherent transition in this process, in that the current weather has some bearing on what the next day's weather will be. If you are an expert in that area, you collect weather data over several years, and calculate that the chance of a sunny day occurring after a rainy day is 0.25. You also note that, by extension, the chance of a cloudy day occurring after a cloudy day must be 0.75, since there are only two possible states. You can now use this distribution to predict weather for days to come, based on what the current weather state is at the time.



The probability distribution is obtained solely by observing transitions from the current day to the next. This illustrates the Markov property,

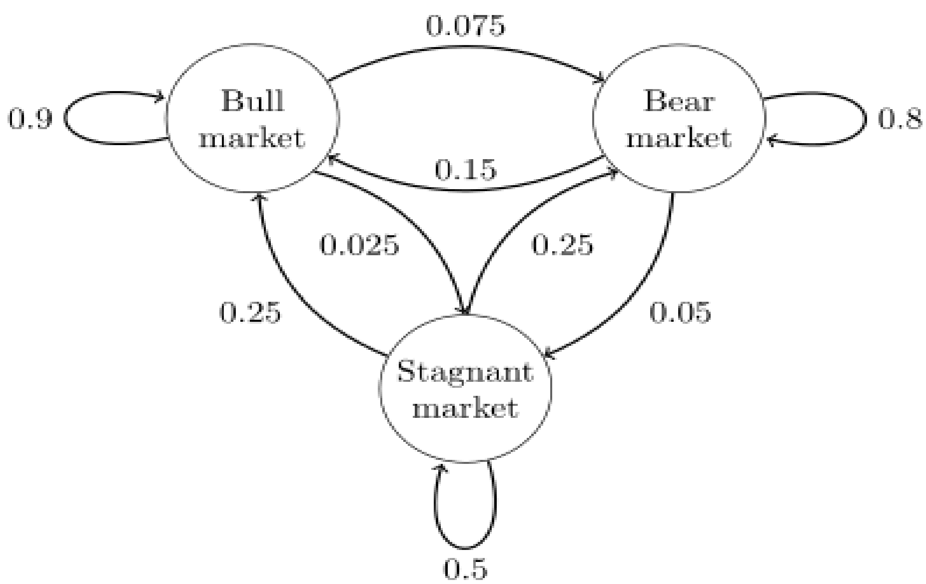
Advantages & Disadvantages of Markov analysis

Advantages	Disadvantages
1) Simplicity and out-of-sample forecasting accuracy. 2) Result is well-known in econometrics. 3) Relatively standard modelling approach can capture many of the features present in a clinical process	1) Not very useful for explaining events, 2) It cannot be the true model of the underlying situation. 3) Results explain little about why something happened.

4) Relatively easy to estimate conditional probabilities based on the current state.	4) Very large model is sometimes impossible to manage.

10.3 THE MODEL

Formally, a Markov chain is a probabilistic mechanism. The probability distribution of state transitions is typically represented as the Markov chain's transition matrix. If the Markov chain has N possible states, the matrix will be an N x N matrix, such that entry (I, J) is the probability of transitioning from state I to state J. Additionally, the transition matrix must be a stochastic matrix, a matrix whose entries in each row must add up to exactly 1. This makes complete sense, since each row represents its own probability distribution.



General view of a sample Markov chain, with states as circles, and edges as transitions

Sample transition matrix with 3 possible states

$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}.$$

We now know how to obtain the chance of transitioning from one state to another, but how about finding the chance of that transition occurring over multiple steps? To formalize this, **we now**

want to determine the probability of moving from state I to state J over M steps. As it turns out, this is actually very simple to find out. Given a transition matrix **P**, this can be determined by calculating the value of entry **(I, J)** of the matrix obtained by raising **P** to the power of **M**. For small values of **M**, this can easily be done by hand with repeated multiplication. However, for large values of **M**, if you are familiar with simple Linear Algebra, a more efficient way to raise a matrix to a power is to first diagonalizable the matrix.

CHECK YOUR PROGRESS

1. What is Markov Chain?
2. What is Markov Analysis?
3. How Google ranks webpages?
4. Mention any one limitation of Markov analysis
5. Mention any one application of Markov chain

10.4 ASSUMPTIONS IN MARKOV ANALYSIS

While discussing Markov chains, following assumptions are made

1. The system has a finite set of possible outcome.
2. The condition of the system in any given period depends on preceding period condition.
3. The transition probabilities are constant over a given length of time.
4. Changes in the system may occur only once during each period.
5. The transition periods occur with regularity.

In first order Markov chains, all the above assumptions are valid, whereas in second order Markov chains, the probability of next outcome may be based on the two previous outcomes, hence the concept of higher order chains.

A Second order Markov chain assumes that the probability of the next outcome may be based on the two-previous outcomes. In fact, this topic can be extended to describe higher order chains. But in this chapter, we will restrict ourselves to the first order Markov chains.

10.5 STEPPED TRANSITION PROBABILITIES

In the case of application of Markov Chains, we will consider the evaluation of the probabilistic movement of the system over a period of time. When we write $n = 0$, it means the system is in the same state or zero-step forward. When we want to establish the probability after some define steps, say n, we write the transition probability as $(p)^{(n)}$.

The corresponding transition matrix can be written as

$$p_2^{(n)} = \begin{array}{c} a_1 \\ a_2 \\ 1 \\ 1 \\ a_m \end{array} \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots & p_{1m}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \cdots & p_{2m}^{(n)} \\ & & & \\ & & & \\ p_{m1}^{(n)} & p_{m}^{(n)} & \cdots & p_{mn}^{(n)} \end{bmatrix}$$

The elements of this matrix can be interpreted as follows:

Here $p_{21}^{(n)}$ depicts the probability that the system which is in state a_2 will move to a_1 after *n*-steps.

Hence,

$$\sum_{i=1}^{m} p_i^{(n)} = 1$$

And

$$\sum_{j=1}^{m} p_{ij} = 1 \text{ for all } i's$$

When we want to write the probability at time t= n+1,

Then

$$p_j^{(n+1)} = \sum_{i=1}^{m} p_i^{(n)} = p_{ij}$$

Where $n = 0,1,2\ldots\ldots$

Generally, these relationships can be reduced to

$P_1^{(n+1)}\ P_2^{(n+1)} + \ldots\ldots\ldots\ldots$

$$\ldots\ldots p_n^{(n+1)} = p_1^{(n)}.\ P_2^{(n)} \ldots\ldots\ldots\ P_n^{(n)} \qquad \begin{bmatrix} p_{11} & p_{12} & \cdots p_{1m} \\ p_{21} & p_{22} & \cdots p_{2m} \\ p_{31} & p_{32} & \cdots p_{3m} \end{bmatrix}$$

$$p_{m1} \quad p_{m} \quad \cdots \; p_{mn}$$

or else R (n+1) = R(n)P

Where R is row vector and hence R(n+1) is row vector at t = n + 1 and R(n) at t = n.

Similarly R(n) = R (n-1)p = R(0) p_n.

Problem 1

The transition of state can also be depicted in the form of the Probability tree which is shown thus:

Fig.1

Work out the state probabilities.

Solution:

p_{11} = (0.70 x 0.70) + 0.30 x 0.60

= 0.67

p_{12} = (0.70 x 0.30) + 0.30 x 0.40

= 0.33

Probability calculation can be explained like this:

If we start with the present state at n= 0.

Than p_1 (0) denotes outcome at n=1.

Hence R(1) = R(0) p

$$= (1\ 0) \begin{pmatrix} 0.70 & 0.30 \\ 0.60 & 0.40 \end{pmatrix}$$

Similarly R(1) = R(1) p

$$= (0.7\ \ 0.30) \begin{pmatrix} 0.70 & 0.30 \\ 0.60 & 0.40 \end{pmatrix}$$

= (0.67 0.33)

This is what indicated in the form of probability tree above. It can be illustrated by taking another example.

Problem 2

Two traders are in for a tough competition. One trader has been lucky to retain his customers for 70% of the time and also in attracting the customers earlier using the product of his competitor trader, to switch over to his product 50% of the time.

Construct and interpret the transition matrix in terms of (a) retention and loss, (b) retention and gain.

Solution:

The transition matrix can be arranged in the following manner.

$$p = \begin{pmatrix} 0.70 & 0.30 \\ 0.50 & 0.50 \end{pmatrix} \quad \text{retention and gain}$$

retention and loss

This matrix indicates that probability of customer now using first trader's product initially and purchasing his product again next time is 0.70. Thus he retains his customers. Similarly the customers earlier using his competitor's product will change to his products with a probability of 0.50. Thus this is loss to the second competitor product. When we see the probability of customers earlier using first trader's product initially now purchase his competitors product is 0.30. This implies loss to first trader's business.

In a similar way, customers retaining purchase of second traders products with probability of 0.50 implies the retention of second trader's products.

Fig. 2.

10.6 HIDDEN MARKOV MODELS

An HMM models a system that satisfies the Markov property with unobserved, or hidden, states. Thus, it can be used to model systems where the state is not directly visible, but the output, which is dependent on the true underlying state, is visible. For example, when modelling the interactions between a technology and the hydrologic cycle where the underlying hydrologic process is inferred based on observed output variables. This inference is typically done by performing two calculation passes over the data. First, a forward pass, called filtering, is performed that computes the probability of the state of a hidden variable given the observations. Then, a backward pass, called smoothing, is performed that computes the probability of the observations given values of the true underlying state. Together, these result in a calculation of the most likely state of the system at any point in time.

While these algorithms are useful in solving the HMM, their computational complexity increases exponentially with the length of the sequence considered and the number of states of the system. In addition, if there are multiple processes within an overall network to be modelled, the computation further increases exponentially with the number of processes of interest. Therefore, approximate rather than exact analysis techniques are often used (Ghahramani and Jordan, 1997). Another alternative is to decompose an HMM into smaller, more computationally tractable sub models. This is done using HHMMs.

10.7 NOTES

10.8 SUMMARY

Markov analysis is a method used to forecast predicted value which is influenced only by its current state. The primary advantages of Markov analysis are simplicity and out-of-sample forecasting accuracy. Markov analysis is not very useful for explaining events. It cannot be the true model of the underlying situation in most cases. Markov analysis is useful for financial speculators, especially momentum investors. Markov analysis is a valuable tool for making predictions, but it does not provide explanations. Dear Learner, you can study further solving problem using Markov analysis using the books given in the reference.

10.9 KEYWORDS

- Markov Chain
- Markov Analysis
- Markov Model

10.10 ANSWER TO CHECK YOUR PROGRESS

1. Markov chain describes the possibility of happening of the succeeding sequence of events based on the previous unit.
2. The process of Markov analysis involves defining the likelihood of future action, which is dependent upon the current state of a variable.
3. Using Markov Analysis
4. Markov chain does not explain reason for happening of event
5. To predict stock price

10.11 SELF-ASSESSMENT QUESTIONS

1. What is Markov chain analysis?
2. Explain the application of Markov analysis with suitable example.
3. Describe the models of Markov analysis.
4. What is HMM?

10.12 REFERENCE

1) Alexander S. Poznyak, in Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Techniques, Volume 2, 2009
2) Kuntz, K., Sainfort, F., Butler, M., Taylor, B., Kulasingam , S., Gregory, S., ... & Kane, R. L. (2013). Decision and simulation modelling alongside systematic reviews. In Decision and Simulation Modelling in Systematic Reviews [Internet]. Agency for Healthcare Research and Quality (US).
3) Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Feb.

UNIT 11: SIMULATION

Structure:

11.0 Objectives
11.1 Introduction
11.2 Meaning and Definitions
11.3 Process of Simulation
11.4 Phases of Simulation
11.5 Applications of Simulation Models
11.6 Types of Simulation Models
11.7 Monte-Carlo techniques / simulation
11.8 Problem
11.9 Notes
11.10 Summary
11.11 Keywords
11.12 Answers to Check Your progress
11.13 Self -Assessment Questions
11.14 References

11.0 OBJECTIVES

After completing this unit, you should be able to:

- Explain the concept of simulation and applications
- Examine the process of simulation

11.1 INTRODUCTION

Simulation is to imitate reality to represent reality and it is a technique for conducting experiments. The simulation is descriptive and not optimizing technique. Simulation is a process often consists of repetition of an experiment in many, many times to obtain an estimate of the overall effect of certain actions. Simulation is usually called for only when the problem under investigation is too complex to be treated by analytical models or by numerical optimization techniques. In a simulation, a given system is copied and the variables and constants associated with it are manipulated in that artificial environment to examine the behavior of the system.

The availability of the computers makes it possible for us to deal with an extraordinary large quantity of details which can be incorporated into a model and the ability to manipulate the model over many experiments (i.e. replicating all the possibilities that may be imbedded in the external world and events would seem to recur).

For example,

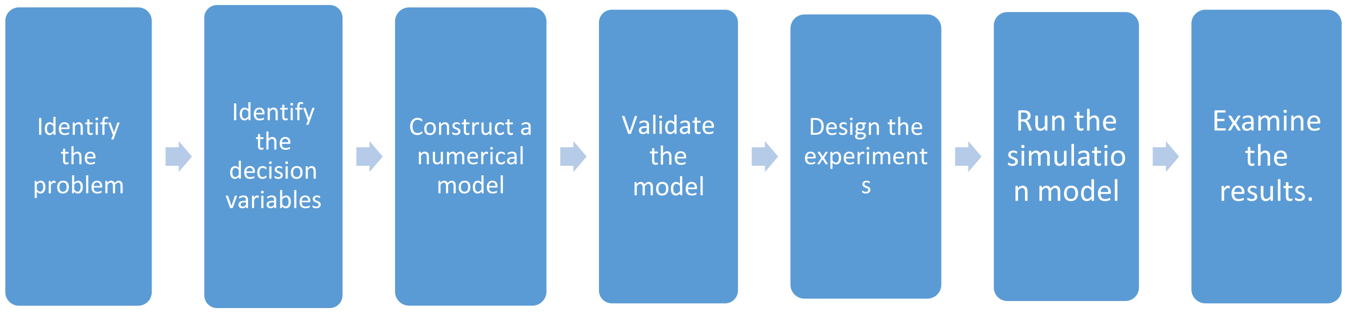
- Testing of an aircraft model in a wind tunnel to test the aerodynamic properties of an the model
- A model of a traffic signal system
- Military war games
- Business games for training
- Planetarium etc

11.2 MEANING AND DEFINITION

A simulation of a system or an organism is the operation of a model or simulator which is a representation of the system or organism. The model is amenable to manipulation which would be impossible, too expensive or unpractical to perform on the entity it portrays. The operation of the model can be studied and for it, properties concerning the behavior of the actual system can be inferred. - Shubik

Simulation is the process of designing a model of a real system and conducting the experiments with this model for the purpose of understanding the behavior (within the limits imposed by a criterion or set of criterion) for the operation of the system. - Shannon

11.3 PROCESS OF SIMULATION



Step 1: Identify the problem.

If an inventory system is being simulated, then the problem may concern the determination of the size of the order (number of units to be ordered) when the inventory falls up to the reorder level (point).

Step 2: Identify the decision variables, performance criterion (objective) and the decision rules.

In the context of the above defined inventory problem, the demand (consumption rate), lead time and safety stock are identified as the decision variable. These variables shall be responsible to measure the performance of the system in terms of total inventory cost under the decision rule- when to order.

Step 3: Construct a numerical model.

Numerical model is constructed to be analyzed on the computer. Some times the model is written in a particular simulation language which is suited for the problem under the analysis.

Step 4: Validate the model

Validation is necessary to ensure whether it is truly representing the system being analyzed and the results will be reliable.

Step 5: Design the experiments

Conduct experiments with the simulation model by listing specific values of variables to be tested (i.e. list courses of action for testing) at each trail (run).

Step 6: Run the simulation model.

Run the model on the computer to get the results in form of operating characteristics.

Step 7: Examine the results.

Examine the results of the problem as well as their reliability and correctness. If the simulation is complete, then select the best course of action (or alternative) otherwise make the desired changes in the model decision variables, parameters or design and return to step 3.

11.3.1 Advantages of simulation

1. Simulation is a straight forward and simple technique
2. The technique is very useful to analyze large and complex problems which are not amenable to mathematical or quantitative methods.
3. It is an interactive method, which enables the decision maker to study the changes and their effects on the performance of the system.
4. The experiments in a simulation are run on the model and not on the system itself.
5. Simulation can be used to design the service system before its actual installation.
6. Due to very characteristic of the model, a great amount of time saving can be achieved.
7. It is a very tempting technique that optimal solutions by analytical method can be overlooked.

11.3.2 Limitations of Simulation

1. At times simulation models can be very costly and expensive
2. It is trail and error technique to produce different solutions in repeated runs.
3. The solution obtained from the simulation may not be optimal.
4. The simulation model needs to be examined and analyzed for decision making. It only creates an alternative and not an optimal solution by itself.

11.4 PHASES OF SIMULATION

1. Definition of the problem and statement of objectives.
2. Identification of decision variables and construction of an appropriate model and its validation.
3. Experimentation with the model constructed.
4. Evaluation of the result of simulation.

For this purpose, collection of useful data and its intelligent use in formulation of simulation model is important. Most common method used is Probabilistic simulation or Monte Carlo Method. It is also called the ‘Computer Simulation’.

In general, the advantages of using simulation for complex problems are in its ability of experimentation with the system by interacting and observing the effects of modifications or deviations in variables in a very easy manner. We can make pre-trials and verify the solutions before their actual applications. Thus, simulation establishes its usefulness for solving complex problem in a very easy and inexpensive way.

11.5 APPLICATIONS OF SIMULATION MODELS

Some of the important areas of application of simulation are listed below:

1. Business games for training
2. Maintenance/service station
3. Simulation models for urban systems
4. Plant and warehousing locations
5. Queuing models
6. Inventory models
7. Financial models
8. Weather models
9. Traffic interaction/lights system models
10. Flight models
11. Pollution sources, concentration and related models
12. Organizational policy models.

Simulation can be usefully applied in the formulation of inventory policy or for capital budgeting. Due to uncertainty of supplies, inventory levels gain importance, when demands are also uncertain. Hence policy of Reorder and order quantity can be designed by simulation method. In case of capital budgeting, evaluation of market demand, Selling price, market price, investment requirement, growth rates or operating cost levels can be worked out by simulation.

There are a large number of computer based simulation models. Major use has been in the field of solving queuing problems. The GPSS programme was developed as a queuing simulating tool.

Similarly, a large number of production simulation programmes have been developed for individual operations of a given work order, where waiting time of a particular work centre is very uncertain, simulation helps in determining such waiting time quite accurately and thus is useful in line balancing problems.

11.6 TYPES OF SIMULATION MODELS

Broadly there can be four types of simulation models.

1. Deterministic models: Where input and output variables have a definite functional relationship.
2. Stochastic models: In this case, variables have the functional relationships given by the probabilistic assumption.
3. Static model: Variability in time is not considered.
4. Dynamic models: Where models deal with variable time functions.

11.7 MONTE-CARLO TECHNIQUES / SIMULATION.

The Monte-Carlo method is a simulation technique in which statistical distribution functions are created by using series of random numbers

11.7.1 Random numbers.

The underlying theory in random number is that, each number has an equal opportunity of being selected.

There are various ways in which random numbers may be generated. These could be: result of some device like coin or die; published table of random numbers, mid-square method, or some other sophisticated method.

It may be mentioned here that random numbers generated by some method may not be really random in nature. In fact such numbers are called pseudo-random-numbers.

Rand corporation (of USA): A million random digits, is a random number table used in simulation situations. The numbers in these tables are in random arrangement.

11.7.2 The Monte-Carlo simulation technique consists of the following steps.

1. Setting up a probability distribution for variables to be analyzed.
2. Building a cumulative probability distribution for each random variable.
3. Generating random numbers. Assign an appropriate set of random numbers to represent value or range (interval) of values for each random variable.
4. Conduct the simulation experiment by means of random sampling
5. Repeat Step 4 until the required number of simulation runs has been generated.
6. Design and implement a course of action and maintain control.

CHECK YOUR PROGRESS

1. In a simulation, ______ environment to examine the behavior of the system
2. There are ______ steps in Process of simulation

3. Most of the methods generate __________ type of random number

11.8 PROBLEM

Consider a construction project, with three parts. The parts have to be done one after the other, so the total time for the project will be the sum of the three parts. All the times are in months.

Task	Time Estimate			
Job 1	5 Months			
Job 2	4 Months			
Job 3	5 Months			
Total	14 Months			

Table 1: Basic Forecasting Model

In the simplest case, we create a single estimate for each of the three parts of the project. This model gives us a result for the total time: 14 months. But this value is based on three estimates, each of which is an unknown value. It might be a good estimate, but this model can't tell us anything about risk. How likely is it that the project will be completed on time?

To create a model we can use in a Monte Carlo simulation, we create three estimates for each part of the project. For each task, we estimate the minimum and maximum expected time (based on our experience, or expertise, or historical information). We use these with the "most likely" estimate, the one that we used above:

Task	Minimum	Most Likely	Maximum	
Job 1	4 Months	5 Months	7 Months	
Job 2	3 Months	4 Months	6 Months	
Job 3	4 Months	5 Months	6 Months	
Total	11 Months	14 Months	19 Months	

Table 2: Forecasting Model Using Range Estimates

This model contains a bit more information. Now there is a range of possible outcomes. The project might be completed in as little as 11 months, or as long as 19 months.

In the Monte Carlo simulation, we will randomly generate values for each of the tasks, then calculate the total time to completion1. The simulation will be run 500 times. Based on the results of the simulation, we will be able to describe some of the characteristics of the risk in the model. To test the likelihood of a particular result, we count how many times the model returned that result in the simulation. In this case, we want to know how many times the result was less than or equal to a particular number of months.

Time	Number of Times (Out of 500)	Percent of Total (Rounded)	
12 Months	1	0%	
13 Months	31	6%	
14 Months	171	34%	
15 Months	394	79%	
16 Months	482	96%	
17 Months	499	100%	
18 Months	500	100%	

Table 3: Results of a Monte Carlo Simulation

The original estimate for the "most likely", or expected case, was 14 months. From the Monte Carlo simulation, however, we can see that out of 500 trials using random values, the total time was 14 months or less in only 34% of the cases.

Put another way, in the simulation there is only a 34% chance – about 1 out of 3 – that any individual trial will result in a total time of 14 months or less. On the other hand, there is a 79% chance that the project will be completed within 15 months. Further, the model demonstrates that it is extremely unlikely, in the simulation, that we will ever fall at the absolute minimum or maximum total values.

This demonstrates the risk in the model. Based on this information, we might make different choices when planning the project. In construction, for example, this information might have an impact on our financing, insurance, permits, and hiring needs. Having more information about risk at the beginning means we can make a better plan for going forward.

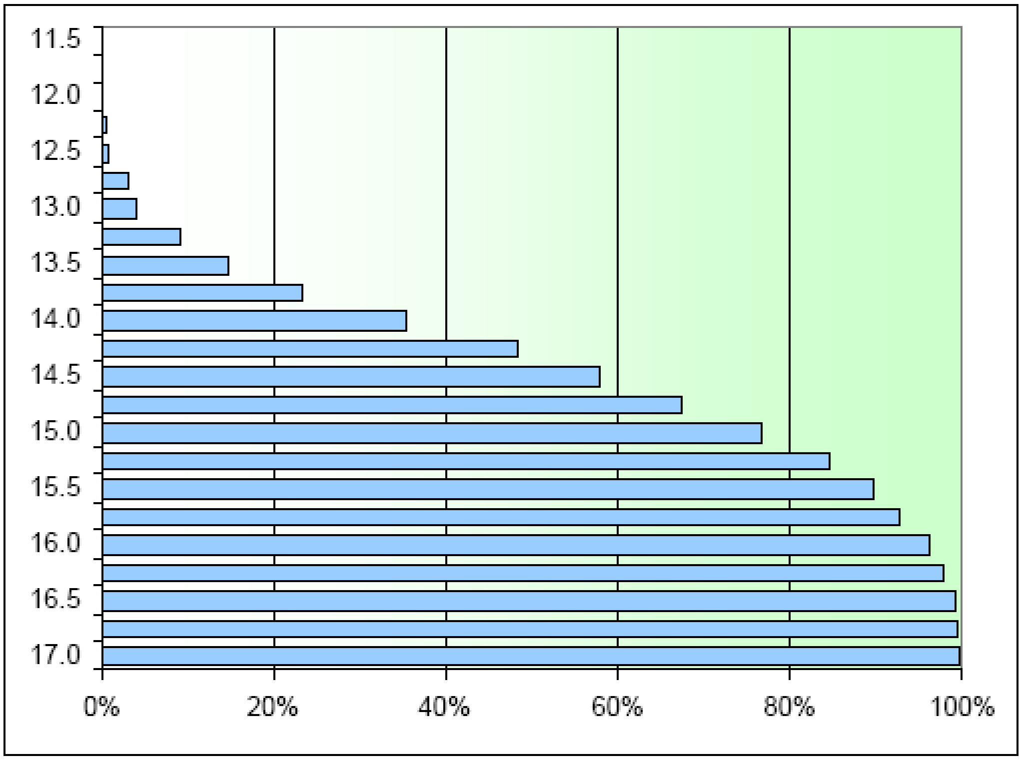


Figure 1: Probability of Completion Within Specified Time (Months)

How Reliable Is It?

Like any forecasting model, the simulation will only be as good as the estimates you make. It's important to remember that the simulation only represents probabilities and not certainty. Nevertheless, Monte Carlo simulation can be a valuable tool when forecasting an unknown future.

Applying to project management

In project management, the technique can be used to model the project cost, or it can be applied to certain project risks that you've identified. The more common use is in the creation of the project schedule and the determination of the project end date.

When you put together your project schedule, you typically create a series of tasks and an estimated duration for each task. When you're finished, you look at the resulting timeline to see the estimated end date. As we all know, the chances of hitting that exact end date aren't 100 percent; you're never 100 percent sure of the duration of the underlying activities. Since uncertainty is associated with each step, a Monte Carlo analysis can be performed.

11.9 NOTES

11.10 SUMMARY

Dear Learner, Mante Carlo simulation is a simulation technique used to create an artificial environment. You can try mante carlo simulation by installing this as an add on feature in Excel. It has saveral applications as discussed in this unit

11.11 KEYWORDS

Simulation

Mante Carlo Simulation

Random numbers

11.12 ANSWER TO CHECK YOUR PROGRESS

1. Artificial
2. 7
3. Psudo

11.13 SELF-ASSESSMENT QUESTIONS

1. What is the application of Mante Carlo Simulation ?
2. What are the advantages and disadvantages of simulation
3. Discuss the process of simulation

11.14 REFERENCE

1) Alexander S. Poznyak, in Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Techniques, Volume 2, 2009

2) Kuntz, K., Sainfort, F., Butler, M., Taylor, B., Kulasingam , S., Gregory, S., ... & Kane, R. L. (2013). Decision and simulation modeling alongside systematic reviews. In Decision and Simulation Modeling in Systematic Reviews [Internet]. Agency for Healthcare Research and Quality (US).

3) Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Feb.

12 : SIMULATION –MONTE CARLO METHOD

Structure:

12.0 Objectives

12.1 Introduction

12.2 Solved Problems

12.3 Notes

12.3 Summary

12.4 Keywords

12.5 Answers to Check Your progress

12.6 Self-Assessment Questions

12.7 References

12.0 OBJECTIVES

After completing this unit, you should be able to:

- Solve problems using Mante Carlo simulation

12.1 INTRODUCTION

Dear Learner, As discussed in the previous unit, in general terms, Simulation involves developing a model of some real life phenomenon and then performing experiments on the model evolved. Often we do not find a mathematical technique that; a model once constructed may permit us to predict what will be the consequences of taking a certain action. In particular we could 'experiment' on the model by 'trying' alternative actions or parameters and compare their consequences. This 'experimentation' allow us to answer 'what if' questions relating the effects of your assumption on the model response. Further to this let us see in this unit how to solve problems using Mante Carlo Simulation.

12.2 SOLVED PROBLEMS

1. M/s Thermo Heaters are the manufacturers of water heaters for both domestic and commercial applications. The company wants to keep stock of the product to meet the customer demand and also not to overstock as the inventory cost becomes higher. The company's past demand record is indicated below. Using the random numbers given below, simulate the demands for next 15 weeks and find the average demand per week for the simulation period.

Random Numbers									
22	19	16	78	03	93	78	23	15	58

Sales record is indicated in the below table.

Sl. No.	Demand	Number of Days
	5	4
	6	10
	7	16
	8	50
	9	62
	10	38
	11	12
	12	8

Solution:

Random Table					
Sl. No.	Demand	Number of Days	Probability	Cumulative Probability	Random Number Interval / Ran Tag
1	5	4	0.02	0.02	00--03
2	6	10	0.05	0.07	04--06
3	7	16	0.08	0.15	07--14
4	8	50	0.25	0.4	15--39
5	9	62	0.31	0.71	40--70
6	10	38	0.19	0.9	71--89
7	11	12	0.06	0.96	90--95
8	12	8	0.04	1	96--99
	Total = N=	200			

Demand Simulation

Day	Random Number	Demand
1	22	8
2	19	8
3	16	8
4	78	10
5	3	5
6	93	11
7	78	10
8	23	8
8	15	8
8	58	9

Average Demand = 85/10 = 8.5 Units

Total Demand = 85

2. Dr. Monte Carlo is a practicing general surgeon in city. He schedules all his patients for 30 minutes' appointments. The kind of the treatments and associated probability is presented in the Table below.

Category	Time required (Minutes)	Probability of the category.
Consultancy	30	0.40
Minor surgery	50	0.15
Follow-up treatment	15	0.15
Small extractions	45	0.10
Medical examination	15	0.20

Simulate the data and hence obtain the average waiting time of the patients. Also find the idle time of the doctor. It may be noted that doctor arrives exactly at 8.00 AM to the clinic. You may use the following random numbers for simulation.

40	82	11	34	25	66	17	79

Random Table			
Time required (Minutes)	Probability	Cumulative Probability	Random Number interval
30	0.4	0.4	00--03
50	0.15	0.55	04--54
15	0.15	0.7	55-69
45	0.1	0.8	70--79
15	0.2	1	80--99

No	Random Number	Patient Arrival	Service Time (min)	Patient In Time	Patient out Time	Waiting Time of Patients	Idle Time of Doctor
1	40	8.00 AM	50	8.00 AM	8.50 AM	0	0.00
2	82	8.30 AM	15	8.50 AM	9.05 AM	20	0.00
3	11	9.00 AM	50	9.05 AM	9.55 AM	5	0.00
4	34	9.30 AM	50	9.55 AM	10.45 AM	25	0.00
5	25	10.00 AM	50	10.45 AM	11.35 AM	45	0.00
6	66	10.30 AM	15	11.35 AM	11.50 AM	65	0.00
7	17	11.00 AM	50	11.50 AM	12.40 PM	50	0.00
8	79	11.30 AM	45	12.40 PM	1.25 PM	70	0.00
					Total =	280	0

Average waiting Time of Patients = 280/8= 35 Minutes

Idle Time of Doctor = 0 Minutes

3. M/s XYZ company is a manufactures of scooters. The company manufacturing around 200 scooters per day The daily production is slightly varies according to availability of resources. The production per day and its associated probabilities are given below table.

Production per day(units)	190	194	198	200	203	206	208	210
Probability	0.05	0.10	0.10	0.20	0.20	0.10	0.15	0.10

The finished vehicles are transported to retails through a truck. The truck can accommodate 200 vehicles exactly per load. Any extra vehicles will be kept in stock. If any shortages, truck will be under loaded.

Simulate the data and hence obtain the average vehicles manufactured and also waiting in the stock. Also find average empty space of truck. You may be use the following random numbers for simulation.

23	73	34	57	83	94	56	67

Random Table				
Sl No.	Production per day(units)	Probability	Cumulative Probability	Random Number Interval / Ran Tag
1	190	0.05	0.05	00--04
2	194	0.10	0.15	05--14
3	198	0.10	0.25	15--24
4	200	0.20	0.45	25--44
5	203	0.20	0.65	45--64
6	206	0.10	0.75	65--74
7	208	0.15	0.90	75--89
8	210	0.10	1.00	89--99

No	Random Number	Production per day(units)	The vehicles left in shop floor (Inventory) (units)	Empty Space in Truck (units)
1	23	198		2
2	73	206	6	
3	34	200		0
4	57	203	3	
5	83	208	8	
6	94	210	10	
7	56	203	3	
8	67	206	6	
	Total =	1634	36	2
	Average Production per day(units)=			204.75
	The average vehicles left in shop floor (Units) =			4.50

4. M/s Excel Electro Mechanical Industries manufactures an automatic water level controller for domestic sump-tank management system. The company outsources the required components and the production process involves both Mechanical and Electrical assembly respectively. The production time(Minutes) varies according to the motor skills of the semi-skilled employees.

The below Table shows the processing time taken along with associated probability.

Production Unit No.	Probability	Mechanical Assembly Time(Minutes)	Electrical Assembly Time (Minutes)
	0.10	15	30
	0.05	18	33
	0.15	21	35
	0.10	24	37
	0.15	26	40
	0.15	28	43
	0.20	31	45
	0.10	35	48

Using the following random number Table below, determine the average time anticipated for the production of next batch.

3441	4384	0238	0576	2875
7674	8312	9595	0090	2435
7349	1519	5415	8080	0995

Solution:

Note: Given random number Table has four digit random numbers. The first two digits may be used for mechanical assembly time estimation and remaining two digits for electrical assembly time estimation.

Random Table - Mechanical Assembly				
Product No	Mechanical Assembly Time (Minutes)	Proba bility	Cumulative Probability	Random Number interval
1	15	0.10	0.10	00--09
2	18	0.05	0.15	10--14
3	21	0.15	0.30	15--29
4	24	0.10	0.40	30--39
5	26	0.15	0.55	40--54
6	28	0.15	0.70	55--69
7	31	0.20	0.90	70--89
8	35	0.10	1.00	90--99

Random Table - Electrical Assembly				
Product No	Electrical Assembly Time (Minutes)	Probability	Cumulative Probability	Random Number interval
1	30	0.10	0.10	00--09
2	33	0.05	0.15	10--14
3	35	0.15	0.30	15--29
4	37	0.10	0.40	30--39
5	40	0.15	0.55	40--54
6	43	0.15	0.70	55--69
7	45	0.20	0.90	70--89
8	48	0.10	1.00	90--99

Product No	Random Number	Mechanical Assembly Time (Minutes)	Random Number	Electrical Assembly Time (Minutes)	Total Production Time (Min)
1	34	24	41	40	64
2	43	26	84	45	71
3	2	15	38	37	52
4	5	15	76	45	60
5	28	21	75	45	66
6	76	31	74	45	76
7	83	31	12	33	64
8	95	35	95	48	83
9	0	15	90	48	63
10	24	21	35	37	58
11	73	31	49	40	71
12	15	21	19	35	56
13	54	26	15	35	61
14	80	31	80	45	76
15	9	15	95	48	63
			Total Production Time(min)=		984
			Average Production Time(min)=		65.6

5. ABC Company manufactures 30 units/day. The sale of these items depends on demand as presented in the below Table.

No	Sales(Units)	Probability
1	25	0.15
2	27	0.20
3	29	0.25
4	31	0.25
5	33	0.05
6	35	0.10

The production cost and sales price/unit is Rs 50 and Rs 75 respectively. The unsold products have sold to scrap dealer at the rate of Rs 15 per unit at the end of the day. There is penalty of Rs 10 per unit if the demand is not met. Using the random number given in the below Table, determine the total profit/loss. If the company decides to manufacture 32 units per day, what is your advice?

03	23	34	57	83	44	56	67	94	74

Solution:

Random Table				
No	Sales(Units)	Probability	Cumulative Probability	Random Number interval
1	25	0.15	0.15	00--14
2	27	0.20	0.35	15--34
3	29	0.25	0.60	35--59
4	31	0.25	0.85	60--84
5	33	0.05	0.90	85--89
6	35	0.10	1.00	90--99

Simulation of profit/loss

			Profit/Loss per day with given production units			
Day	Random Number	Estimated Sales(Units)	30 Units		32 Units	
1	3	25	25 X 25-5 X15=	550	25X25-7X15=	520
2	23	27	27 X 25-3 X15=	630	27X25-5X15=	600
3	34	27	27 X 25-3 X15=	630	27X25-5X15=	600
4	57	29	29 X 25-1 X15=	710	29X25-3X15=	680
5	83	31	30 X 25-1 X10=	740	31X25-1X15=	760
6	44	29	29 X 25-1 X15=	710	29X25-3X15=	680
7	56	29	29 X 25-1 X15=	710	29X25-3X15=	680
8	67	31	30 X 25-1 X10=	740	31X25-1X15=	760
9	94	35	30 X 25-5 X10=	700	32X25-3X10=	770
10	74	31	30 X 25-1 X10=	740	31X25-1X15=	760
			Total=	6860	Total=	6810

6. The rain fall on the given day depends upon the previous day rain fall. Referring to the two Tables given below, you are required to assess the rain fall of the Mysore city using Monte-Carlo simulation method. State the (initial) assumption clearly.

No	Event (Rain on the previous day)	Probability
1	No Rain	0.40
2	1 cms rain	0.05
3	2 cms rain	0.20
4	3 cms rain	0.05
5	4 cms rain	0.10
6	5 cms rain	0.05
7	6 cms rain	0.15

No	Event (No-Rain on the previous day)	Probability
1	No Rain	0.45
2	1 cms rain	0.15
3	2 cms rain	0.15
4	3 cms rain	0.05
5	4 cms rain	0.10
6	5 cms rain	0.05
7	6 cms rain	0.05

Use the following random numbers: 87,77,24,99,05,55,66,10,26,02

Solution

Initial Assumption: The simulation is done based on the assumption that there was a rain on the previous day.

Note: If there is no rainfall on the given day of estimation, then it is required to referrer the random Table pertains to No rainfall.

Random Table - Rain fall on previous day				
No	Rain fall (cms)	Probability	Cumulative Probability	Random Number interval
1	No Rain	0.40	0.40	00--39
2	1 cms rain	0.05	0.45	40--44
3	2 cms rain	0.20	0.65	45--64
4	3 cms rain	0.05	0.70	65--69
5	4 cms rain	0.10	0.80	70--79
6	5 cms rain	0.05	0.85	80--84
7	6 cms rain	0.15	1.00	85--99

Random Table - No Rain fall on previous day				
No	Rain fall (cms)	Probability	Cumulative Probability	Random Number interval
1	No Rain	0.45	0.45	00--44
2	1 cms rain	0.15	0.60	45--59
3	2 cms rain	0.15	0.75	60--74
4	3 cms rain	0.05	0.80	75--79
5	4 cms rain	0.10	0.90	80--89
6	5 cms rain	0.05	0.95	90--94
7	6 cms rain	0.05	1.00	95--99

Simulation of Rain fall (cms)		
Day	Random Number	Rain fall (cms)
1	87	6
2	77	4
3	24	No rain
4	99	6
5	5	No rain
6	55	1
7	66	3
8	10	No rain
9	26	No rain
10	2	No rain
	Total Rain Fall (cms)=	20

7. The Super Motors Company, is an authorized sales and service station for two wheelers. The company sales the two wheelers also provide post services. The service station has a single server centralized stores having one store keeper. The general stores open at 9.00 AM. The data presented in below Table shows that the service station waiting time (min) along with associated probability.

Arrival Time & Probability	
Time(Min)	Probability
0	0.20
3	0.15
6	0.15
9	0.20
12	0.15
15	0.15

Service Time & Probability	
Time(Min)	Probability
4	0.20
8	0.15
12	0.15
15	0.20
18	0.15
20	0.15

Using the following random number Table below, determine the average waiting time of the mechanics who arrives to get the materials from general stores. Also determine the idle time of the store keeper.

Random Numbers for arrival time:	34,43,02,76,28,83,12,95,00, and 63
Random Numbers for Service time:	87,77,24,99,05,55,66,10,26,02

Solution

Random Table - Arrival Time			
Time required (Minutes)	Probability	Cumulative Probability	Random Number interval
2	0.20	0.20	00--19
3	0.15	0.35	20--34
6	0.15	0.50	35--49
9	0.20	0.70	50--69
12	0.15	0.85	70--85
15	0.15	1.00	86--99

Random Table - Service Time			
Time required (Minutes)	Probability	Cumulative Probability	Random Number interval
4	0.20	0.20	00--19
6	0.15	0.35	20--34
8	0.15	0.50	35--49
9	0.20	0.70	50--69
10	0.15	0.85	70--85
12	0.15	1.00	86--99

No	Random Number (Arrival)	Mechanic's Arrival	Arrival between Time (min)	Random Number (Service)	Service Time (min)	Mechanic In Time	Mechanic out Time	Waiting Time	Idle Time of store-keeper
1	34	9.03AM	3	87	12	9.03AM	9.15 AM	0	0.00
2	43	9.09 AM	6	77	10	9.15AM	9.25 AM	6	0.00
3	2	9.11 AM	2	24	6	9.25AM	9.31 AM	14	0.00
4	76	9.23 AM	12	99	12	9.31AM	9.43 AM	9	0.00
5	28	9.26 AM	3	5	4	9.43AM	9.47 AM	17	0.00
6	83	9.38 AM	12	55	9	9.47AM	9.56 AM	9	0.00
7	12	9.40AM	2	66	9	9.56AM	10.05 AM	16	0.00
8	95	9.55 AM	15	10	4	10.05AM	10.09AM	10	0.00
9	0	9.57AM	2	26	6	10.09AM	10.15AM	12	0.00
10	63	10.06AM	9	2	4	10.15AM	10.19AM	9	0.00
							Total =	102	0.00

Average Waiting Time of Mechanics at stores= 10.2 min

12.3 NOTES

12.4 SUMMARY

Dear Learner, As discussed Mante Carlo simulation is a simulation techinque used to create an artificial enivroment. You can try mante carlo simulation by installing this as an add on feature in Excel. In this unit different types of problems are solved to give your practical orientation.

12.5 KEYWORDS

- Simulation
- Mante Carlo Simulation
- Random numbers

12.6 SELF-ASSESSMENT QUESTIONS

1. What is the application of Mante Carlo Simulation ?
2. What are the advantages and disadvantages of simulation
3. Discuss the process of simulation

12.7 REFERENCE

1) Alexander S. Poznyak, in Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Techniques, Volume 2, 2009
2) Kuntz, K., Sainfort, F., Butler, M., Taylor, B., Kulasingam , S., Gregory, S., ... & Kane, R. L. (2013). Decision and simulation modelling alongside systematic reviews. In Decision and Simulation Modelling in Systematic Reviews [Internet]. Agency for Healthcare Research and Quality (US).
3) Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Feb.
4) Goodwin, P., & Wright, G. (1998). Decision Analysis for Management Judgment (2nd edition). Chichester: Wiley.
5) Clemen, R. T. (1996). Making hard decisions (2nd ed). Belmont, Calif: Duxbury Press.
6) Palisade Corporation (2002). Guide to using @risk. Newfield, NY: Palisade Corporation.

UNIT 13: INTRODUCTION TO BUSINESS ANALYTICS

Structure

13.0 Objectives

13.1 Introduction

13.2 Meaning and Definitions

13.3 Types of Business Analytics

13.4 Descriptive Analytics

13.5 Predictive Analytics

13.6 Prescriptive Analytics

13.7 Analytics using Excel

13.8 Case Study

13.9 Summary

13.10 Keywords

13.11 Answers to Check Your progress

13.12 Self-Assessment Questions

13.13 References

13.0 OBJECTIVES

After completing this unit, you should be able to:

- Explain the relevance of Analytics in business scenario, and how & where it is useful
- Apply Microsoft Excel for analysing business data
- Define Descriptive and prescriptive Analytics
- Utilise Microsoft Excel to derive descriptive and prescriptive analytics for a given data set

13.1 INTRODUCTION

Business analytics is the process of collecting and applying data on business activities so as to draw inferences and make calculated decisions with higher certainty. Business analytics enables data driven (fact based) decision making while extending accountability. Business analytics can be defined as 'the process of exploring, experimenting, simulating, and summarizing data to extract information'. Multiple levels and roles in an organisational hierarchy can benefit from Business analytics based on a gamut of analysis around business data to draw inferences.

The rapid strides in digital technology lead products in the Data systems domain, like Data aggregation, warehousing, Data Lakes, Data mining and Real time data visualisation has led to an exponential growth in adoption of Business analytics. Some of the typical examples of Data Analytics usage are:

- Pricing - Setting prices for consumer and industrial goods, Contracts etc.
- Customer segmentation - Identifying and targeting key customer groups across Industries - Retail, Insurance, Cards and Payments, Automotive, Healthcare etc.
- Merchandising - Determining brands to buy, quantities, and allocations
- Location - Finding the best location for businesses like Bank branches and ATMs
- Social Media - Understand trends and customer perceptions; assist marketing managers and product designers

Today it is imperative for every business to have a business analytics team, that specializes in unearthing business data, and present it with inferences to the stakeholders for them to make informed decisions. For a business to stay competitive in its domain there is a need for continuous improvement, increased efficiencies in business operations, and being agile in re-inventing often in response to market dynamics and opportunities. Business analytics helps in achieving this.

13.2 MEANINGS AND DEFINITIONS

Analytics

Analytics is the process of discovering, interpreting, and communicating significant patterns in data. Quite simply, analytics helps us see insights and meaningful data that we might not otherwise detect. Business analytics focuses on using insights derived from data to make more informed decisions that will help organizations improve on multiple business parameters, like sales, Return on Investment, operational efficiency, and make other business improvements.

Descriptive analytics

Descriptive analytics is a field of statistics that focuses on gathering, summarizing and representing raw data for easy interpretation. Descriptive analytics focuses on historical data, providing the context that is vital for understanding information and numbers. The representations are typically in the form of visual displays like line, bar and pie charts.

Predictive analytics

When probabilities are applied to historical data to make assessments on what could happen in the future, it is referred to as Predictive analytics. Just like descriptive analytics, predictive analytics also uses data mining.

Prescriptive analytics

Prescriptive analytics is the application of statistical methods and algorithms to derive actionable insights based on data gathered from a range of descriptive and predictive sources and applying them to the decision-making process. Prescriptive Analytics shows which option is the best for a given scenario.

13.3 TYPES OF BUSINESS ANALYTICS

Analytics can be classified based on the way they are derived and their functionality.

There are three types of business analytics based on the way they are derived:

1. Descriptive analytics: What has already happened
2. Predictive analytics: What could happen based on past data
3. Prescriptive analytics: What should happen in the future

There are multiple types of business analytics based on their function like:

1. Operations Analytics

2. HR Analytics
3. Financial Analytics
4. Customer Analytics
5. Supply Chain Analytics
6. Healthcare Analytics

13.4 DESCRIPTIVE ANALYTICS

Descriptive analytics uses historical data. Data is collected, organised and then presented in a way that is easily understood. Descriptive analytics focuses on what has already happened in a business and helps us understand changes that have occurred over time. It uses simple maths and statistical tools, such as arithmetic, averages and percent changes, rather than the complex calculations necessary for predictive and prescriptive analytics. It is used as a starting point to prepare data for further analysis.

The first step in the descriptive analytics process is to establish the metrics that can effectively evaluate performance against goals. Once the metrics are listed, the next step is to obtain the necessary data, which must then be collected, organised, and prepared for the next step - data analysis. Data aggregation and data mining are the two key techniques that are usually employed by descriptive analytics in its discovery of historical data. Data aggregation involves the collection and organization of data into manageable data sets, while data mining uses the created data sets to determine trends, patterns, and meaning. The information gained from these processes is then illustrated in a way that is easily understood by the target audience.

Below are some examples that illustrate the use of descriptive analytics:

- A variety of organisational reports, such as reports on Revenues, sales, Costs, Inventory, are all products of descriptive analytics.
- Insights into social media usage and engagement (e.g., Facebook or Instagram likes, posts, comments).
- Helps collate and summarise survey results.
- Reports general trends.

Advantages:

- Quickly and easily report on multiple parameters and identify gaps
- Easy to represent and well understood by a wider audience
- Describes relationships and exhibits the world as-is
- Helps in quick identification of variables of interest

13.5 PREDICTIVE ANALYTICS

In predictive Analytics, the likelihood of future outcomes are identified based on statistical modelling. These predictions on outcomes are then used to solve problems and identify opportunities for growth. For example, organisations are using predictive analytics to identify likelihood of credit card frauds based on patterns in criminal behaviour. Predictive Analytics utilizes statistical modelling, algorithms and machine learning to identify trends or patterns in historical or transactional data and predict future behaviours.

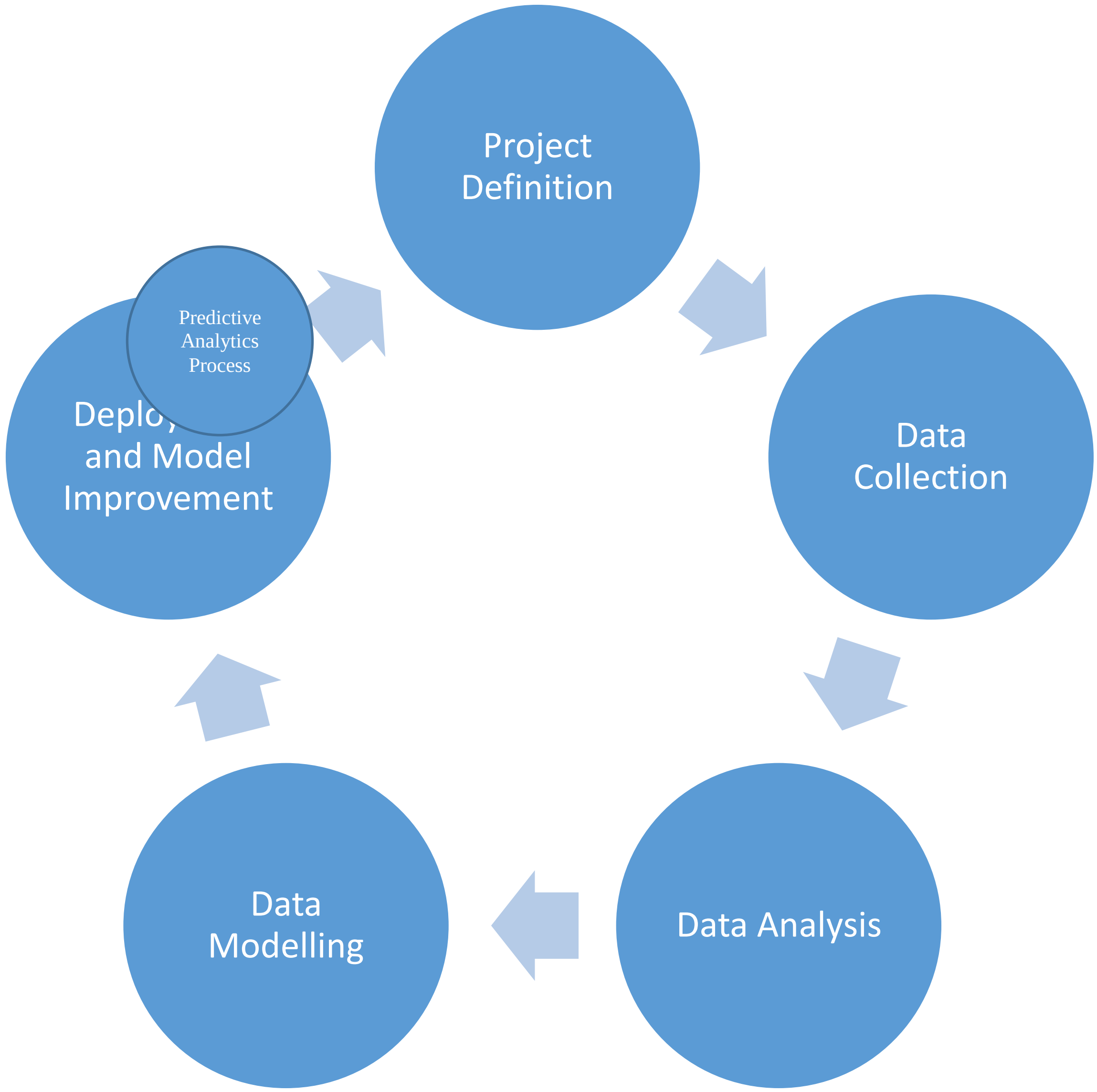
Advantages:

- Helps in Identifying patterns which then can be used for reducing risk in business.
- Helps in understanding the business constituents better. Example., Understanding customers using credit scores
- Can be used to optimise Operations and improving decision making
- Models can be reused across diverse organisational context, Example., Fraud detection model created for Telecom consumers can be applied to retail consumers with modifications as needed

Examples of Predictive Analytics use across domains:

- Banking and Financial Services: Detect and reduce fraud, measure credit risk, Debt Collection, estimate future cash flows and projections of expected receivables.
- Retail: Price optimisation, Measure the effectiveness of promotional events and determine which offers are most appropriate to customers.
- Pharmaceuticals: Optimising clinical trials, segmentation of patients based on their likelihood to respond well to a drug.
- Healthcare: Predicting patient health status over time, preventing decline in health.
- Insurance: Understand customer churn and improve customer retention
- Manufacturing: Predicting equipment failures and conducting preventive maintenance, Part replacements, managing spares inventory, improve safety and overall performance
- Government & Public Sector: Predict demographic trends, become proactive in governance, detect fraudulent claims and usage of government or public resources, prevent public unrest/ conflicts

Predictive Analytics Process:



The predictive analytics process involves multiple steps as follows: Project Definition: Define the business objectives and corresponding outcomes and deliverables desired, estimate the efforts, identify the metrics and the data sources

1. **Data Collection:** Aggregate the data sets required from all possible sources
2. **Data Analysis:** Inspect and Clean the data, transform the data, and finally model the data using statistical techniques with an objective of discovering useful insights

3. **Data Modelling:** Create accurate models to predict outcomes and evaluate multiple models using algorithms
4. **Model Deployment:** Once the model is finalised based on prediction accuracy and fit, deploy the model for decision making
5. **Model Monitoring:** Evaluate performance of the model periodically to ascertain the accuracy/deviations from desired outcomes and apply necessary remediation/correction.

13.6 PRESCRIPTIVE ANALYTICS

Prescriptive analytics generates recommendations and suggests decisions based on the computational findings of algorithmic models. Generating automated decisions or recommendations requires specific and unique algorithmic models.

The first steps is to know – what problem is to be solved? A recommendation cannot be generated without knowing what to look for. Thus prescriptive analytics begins with a problem.

Example., In a factory Predictive analytics suggests that a CNC Machine fails after 90 days of continuous operation. Obviously, maintenance has to be done prior to 90 days of continuous operation so as to prevent disruption. A further analysis of data using algorithms suggested possible alternative timelines for maintenance, prioritizing components for maintenance and possible production disruptions so as to conduct the preventive maintenance optimally.

Models are generally recommended to be tailored for each unique situation and need.

Descriptive Vs Predictive Vs Prescriptive Analytics

- Descriptive Analytics is focused solely on historical data, and describes facts or situations As-Is
- Predictive Analytics uses this historical data to develop statistical models that will then predict future possibilities
- Prescriptive Analytics takes the possible forecasted outcomes and predicts consequences for these outcomes

13.7 ANALYTICS USING EXCEL

Spread sheets are widely used for collecting, arranging, ordering, and analysing data. Microsoft Excel is one such spread sheet tool which has impressive capabilities to manage this. Before you begin with Excel, the following basic terminology should be well understood:

- **Ribbon**: The top part with Menu Options is referred to as the Ribbon. Excel selects the ribbon's Home tab when you open it. Learn how to use the ribbon.

- **Workbook**: A workbook is another word for your Excel file. When you start Excel, click on Blank workbook to create a new Excel workbook or Click in Open to work on an existing workbook
- **Worksheets**: A worksheet can be seen as a Page in the work book and is a collection of cells where you keep and manipulate the data. Each Excel workbook can contain multiple worksheets.
- **Format Cells**: When we format cells in Excel, we change the appearance and properties of data in that cell
- **Find & Select:** Learn how to use Excel's Find, Replace.
- **Templates:** Instead of creating an Excel workbook from scratch, you can create a workbook based on a template.
- **Range:** A range in Excel is a collection of two or more cells. This chapter gives an overview of some very important range operations.
- **Formulas and Functions:** A formula is an expression which calculates the value of a cell. Functions are predefined formulas and are already available in Excel.
- **Data Validation:** You can Use data validation to make sure that users enter correct values or type of data into a cell.
- **Keyboard Shortcuts:** Keyboard shortcuts allow you to do things with your keyboard instead of your mouse to increase your speed.
- **Print:** This about printing a worksheet/Workbook or parts of it
- **Share:** Learn how to share Excel data with Word documents and other files.
- **Protect:** Encrypt an Excel file with a password so that it requires a password to open it.

Excel Features that are useful in analysing data:

a) **Sort:** You can sort your Excel data on one column or multiple columns. You can sort in ascending or descending order.

b) **Filter:** Filter your Excel data if you only want to display records based on certain criteria.

c) Conditional Formatting: This enables you to highlight cells with a certain colour, depending on the cell's value.

d) **Charts:** A simple Excel chart can say more than a sheet full of numbers. As you'll see, creating charts is very easy. There are multiple types of charts like Bar, Line, Pie etc for you to use.

e) **Pivot Tables:** Pivot tables are one of Excel's most powerful features. A pivot table allows you to summarize from a large, detailed data set.

f) **Tables:** Helps you to analyse your data quickly and easily.

g) **Data Validation:** Helps you to standardise data to meet certain criteria and eliminate errors

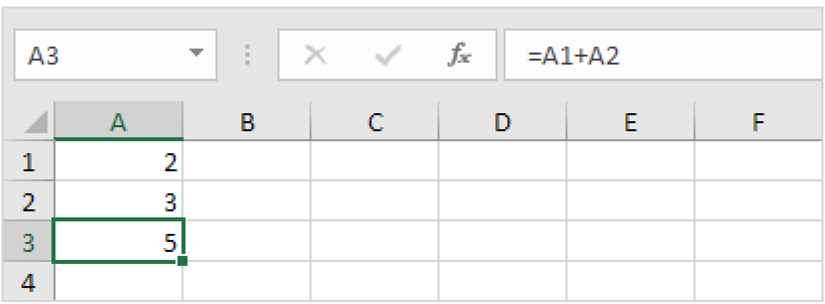
h) **What-If Analysis:** What-If Analysis in Excel allows you to try out different values (scenarios) for formulas.

i) **Solver:** Excel includes a tool called solver that uses techniques to find optimal solutions for decision problems.

j) **Analysis Tool Pak:** The Analysis ToolPak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis.

Functions in Excel

In its simplest form the following illustrates a function:



In the example above, the value of Cell A1 and Cell A2 as defined by an addition function in cell A3 becomes 2+3=5. Cell A3 has a function for addition, A3=A1+A2

- **Count and Sum:** You can count and sum based on one criteria or multiple criteria.
- **Logical:** logical functions, such as IF, AND, OR and NOT.
- **Cell References:** Cell references in Excel are very important. Understand the difference between relative, absolute and mixed reference.
- **Date & Time:** To enter a date, use the "/" or "-" characters. To enter a time, use the ":" (colon).
- **Text:** Excel has multiple functions to manipulating text strings.
- **Lookup & Reference:** Powerful functions like VLOOKUP, HLOOKUP, MATCH, INDEX and CHOOSE.
- **Financial:** Excel has multiple built in predefined financial functions, like average, Roi, Future value, Interest calculations, Amortisation etc.
- **Statistical:** There are multiple Statistical functions built in, average, Median, Mode, Regression, and so on.

- **Round/Ceiling/Floor:** There are functions to round numbers / restrict in Excel. ROUND, ROUNDUP and ROUNDDOWN, FLOOR, CEILING
- **Array Formulas:** Array formulas performing multiple calculations in one cell.

CHECK YOUR PROGRESS

1. Descriptive Analytics is based on _______________Data
   a. Current
   b. Historical
   c. Future
   d. Engineering
2. Predictive Analytics can predict ____________
   a. Outcomes
   b. Implications
   c. Variables
   d. Impact
3. Prescriptive Analytics is used to understand _____________
   a. Outcomes
   b. Implications
   c. Variables
   d. Engineering
4. Detection of Fraud is a good example of _____________Analytics
   a. Descriptive
   b. Predictive
   c. Prescriptive
   d. Engineering
5. State True of False. Microsoft Excel Add in " Solver" can be used for calculating Predictive Analytics
   a. True
   b. False

13.8 CASE STUDY

A business group, BriteLights, which produces Light fittings for Cars has four factories in different parts of the country. They are the authorised Original Equipment Manufacturer (OEM) for three major Car Manufacturers. The Business Pressures have forced the customers to demand the most optimal pricing from Brite Lights. The General Manager of Brite Lights has called in their Senior Business Analyst to recommend the right mix of supply from different Factory Locations to each customer so that they can continue to supply at the most competitive costs feasible and retain the customers for long term.

Unit Cost (INR)	Customer 1	Customer 2	Customer 3
Factory 1	80	94	160
Factory 2	144	72	116
Factory 3	48	132	142
Factory 4	24	142	65

Shipments in Dec 2021	**Customer 1**	**Customer 2**	**Customer 3**
Factory 1	200	300	50
Factory 2	100	200	100
Factory 3	200	50	100
Factory 4	500	50	400

Question to be asked:

What are the sequential steps the Senior Business Analyst must take in order to recommend the most optimal supply from the four factories to each customer?

Hints:

1. Understand the Descriptive Analytics like Average per Customer, Per factory, Understand the range with max and min cost
2. Understand what Metric should be optimised to minimise the cost

Recommended Solution:

Dear Learner, you may recall such transportation problems we have discussed in the first block.

Workout the Descriptive Analytics:

Unit Cost	**Customer 1**	**Customer 2**	**Customer 3**	**Average Unit Cost**
Factory 1	80	94	160	111.33
Factory 2	144	72	116	110.67
Factory 3	48	132	142	107.33
Factory 4	24	142	65	77.00
Average Customer Unit Cost	74	110	120.75	

From the above calculation it is clear that Customer 1 is receiving at least cost while Customer 3 is getting at the highest cost.

Shipments	**Customer 1**	**Customer 2**	**Customer 3**	**Average Shipments from Factory**
Factory 1	200	300	50	183.33
Factory 2	100	200	100	133.33

Factory 3	200	50	100	116.67
Factory 4	500	50	400	316.67
Average Shipment to Customer	250	150	162	

In terms of shipments, Factory 4 has the highest average while Customer 1 receives the maximum shipments. However the overall cost is a function of the shipments and unit cost corresponding. Hence the following calculation is made:

Shipments	**Customer 1**	**Customer 2**	**Customer 3**	**Average Shipments from Factory**	**Total Out**		**Supply**
Factory 1	200	300	50	183.33	550	=	550
Factory 2	100	200	100	133.33	400	=	400
Factory 3	200	50	100	116.67	350	=	350
Factory 4	500	50	400	316.67	950		950
Average Shipment to Customer	250	150	162.5				
	=	=	=				Total Cost (INR)
Total Demand	**1000**	**600**	**650**				**168100**

Here Total Cost is the summation of all shipments and corresponding Unit costs. The objective is to minimise this Metric. There are multiple methods to do this,

i) Using Trial and Error

ii) Using Solver Add-In of Excel.

In Trial and Error Method, you keep changing the mix of shipments from each factory to each customer so that Total cost reaches its lowest value.

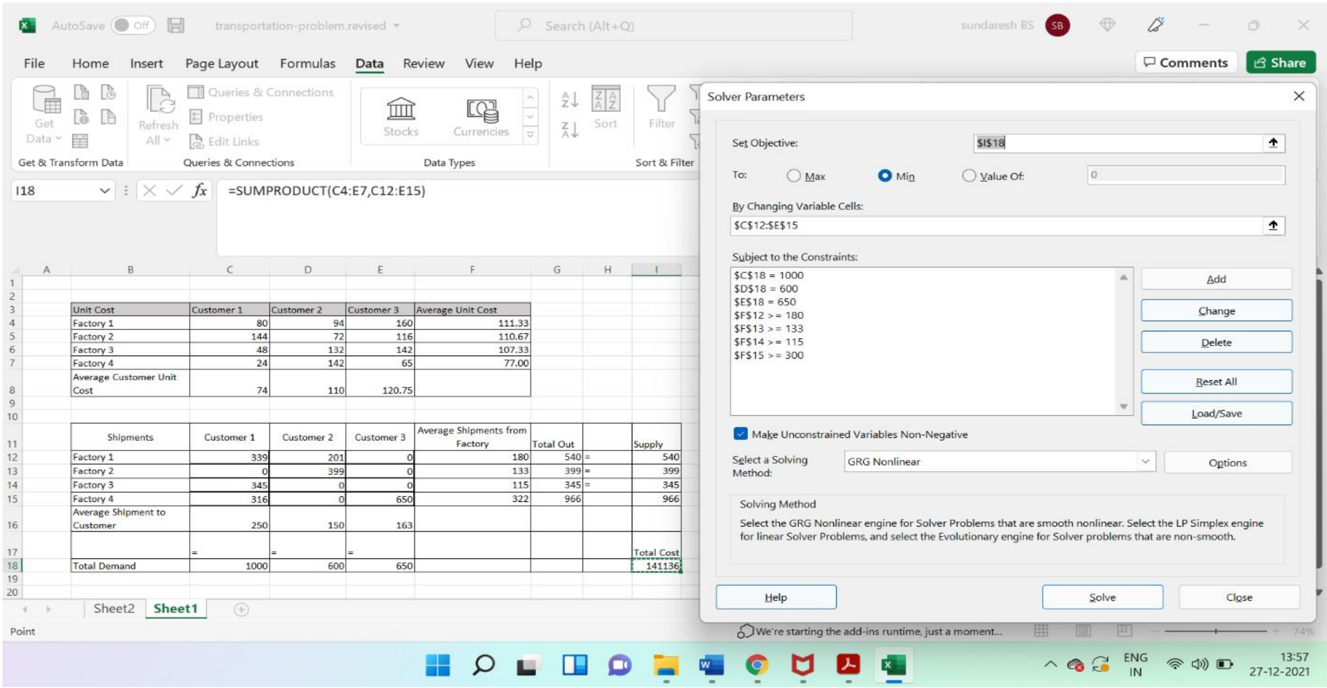
Let us try the solver Add-In which automates the above process. You can find the Solver under the Data Menu of Excel, In case you do not find it, Enable it using - File > More > Options > Add-In

The Screenshot shows how and where to provide the inputs for the solver. The following constraints have been given:

i) Total demand from Each customer has to be met

ii) No factory shall have shipments less than their current average

iii) There is no minimum shipment for any customer from each factory



The solver gives the following solution with the constraints given:

Shipments	Customer 1	Customer 2	Customer 3	Average Shipments from Factory	Total Out		Supply
Factory 1	339	201	0	180	540	=	540
Factory 2	0	399	0	133	399	=	399
Factory 3	345	0	0	115	345	=	345
Factory 4	316	0	650	322	966		966
Average Shipment to Customer	250	150	163				
	=	=	=				Total Cost INR
Total Demand	1000	600	650				141136

The Solver has redistributed the Supply to Customers while keeping the average production the same and all the demand of customers being met. Now note that the Total Cost reduces from INR 168100 to 141136, with Shipments optimised. A saving of 16% to Brite Lights with which they can keep the three main customers happy and engaged.

13.10 SUMMARY

Business Analytics is a vast yet important domain for Organisations as it provides data driven insights/inferences/recommendations to the stakeholders for making decisions. There are mainly three types of Analytics – Descriptive, Predictive and Prescriptive Analytics. Descriptive Analytics uses historical data and unearths facts and figures with in it and represents it in easy to understand graphs and tables. Most businesses still use Descriptive Analytics as their main decision making metrics and are now rapidly adopting Predictive and Prescriptive analytics to address the digital disruption.

Predictive Analytics builds on Descriptive analytics, adds statistical and machine learning rigour to unearth trends and make predictions of what could happen. It is highly useful across domains in determining future outcomes and take proactive steps to address the same.

Prescriptive Analytics goes to the next level of providing insights on the consequences of such outcomes.

Microsoft Excel has a host of capabilities in Data Aggregation and Analysis, which are easy to use and powerful.

13.11 KEYWORDS

Analytics, Descriptive Analytics, Predictive Analytics, Prescriptive Analytics, Data, Data Aggregation, Data Mining, Data Analysis, Microsoft Excel

13.12 ANSWERS TO CHECK YOUR PROGRESS

1. b
2. a
3. b
4. b
5. a

13.13 SELF ASSESSMENT QUESTIONS

1. Define Business Analytics and Explain the different types
2. Define Prescriptive analytics
3. Explain the Predictive Analytics Process with the steps involved

13.4 REFERENCES

- Business Analytics: Data Analysis and Decision Making, 5th Edition, Albright and Winston
- The Art of Modelling with Spreadsheet, Stephen Powel and Ken Baker URLs:
- https://www.excel-easy.com/
- https://support.microsoft.com/en-us/training

UNIT 14: BIG DATA MANAGEMENT

STRUCTURE

14.0 Objectives

14.1 Introduction

14.2 Data and Its Processing

14.3 Big Data Management

14.4 Data Visualisation

14.5 Case Study

14.6 Notes

14.7 Summary

14.8 Keywords

14.9 Self Assessment Questions

14.10 Answers to Check Your progress

14.11 References

14.0 OBJECTIVES

After completing this unit, you should be able to:

- Describe the fundamental concepts of Data, its types and characteristics
- Analyse the basic methods of data processing to a given problem
- Appreciate what constitutes Big data and its use in different sectors
- Examine the characteristics of effective Data visualisation

14.1 INTRODUCTION

"Data is the new Oil", such is the importance of data. With the ever increasing computing power with decreasing costs of computing, storage and communication, technology is evolving rapidly to consume huge volumes of data in real time and provide decision pointers to businesses.

What used to be numeric and text as data in the legacy world, has now expanded to include, texts, numeric, images, videos, social media messages, emojis, blogs, and so on. And at a much faster pace, much higher volumes. This has paved the way for huge developments in the data science, Engineering and data management domain. The business analytics as explained in the earlier section are a result of the myriad data crunching processes and techniques.

14.2 DATA AND ITS PROCESSING

Data is a collection of facts, figures, objects, symbols, and events gathered from different sources. Without data, it would be difficult for organizations to make appropriate decisions and hence, Organizations collect data to make better decisions.

Data Collection

Data is collected at various points in time from different sources. For example, before launching a new service for mobiles, the company needs to collect data on demand, customer preferences, their profile, and so on. Data is a valuable asset for every organization, but only when it is properly analysed. Data collection is a process of collecting information from all the relevant sources. In a Business, data is generated through multiple sources like, Operations, Transactions, Machinery, External interactions, and so on. Hence it is imperative for a business to create systems which collect such data in all its completeness and accuracy.

Data Management refers to the handling of data in its various forms, either collected digitally or converted from traditional forms into a digital form. The entire process of aggregating data, storing it, recovering it as and when required is referred to as Data Management.

Sources of Data:

A business may collect data from internal sources or External sources. Internally Data can come from Operations, Research, Customer Feedback, Machinery, Employees, Shop floor and so on. Externally data may come from vendors, suppliers, partners, research houses, publications, and so on. When a Business collects data directly from where it is created, it is referred to as Primary data. When the data is collected from other sources, it is referred to as Secondary data.

Data Processing

Data processing occurs when data is collected and translated into usable information. Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.), giving it the form and context necessary to be interpreted by computers and utilized by employees throughout an organization.

Six stages of data processing

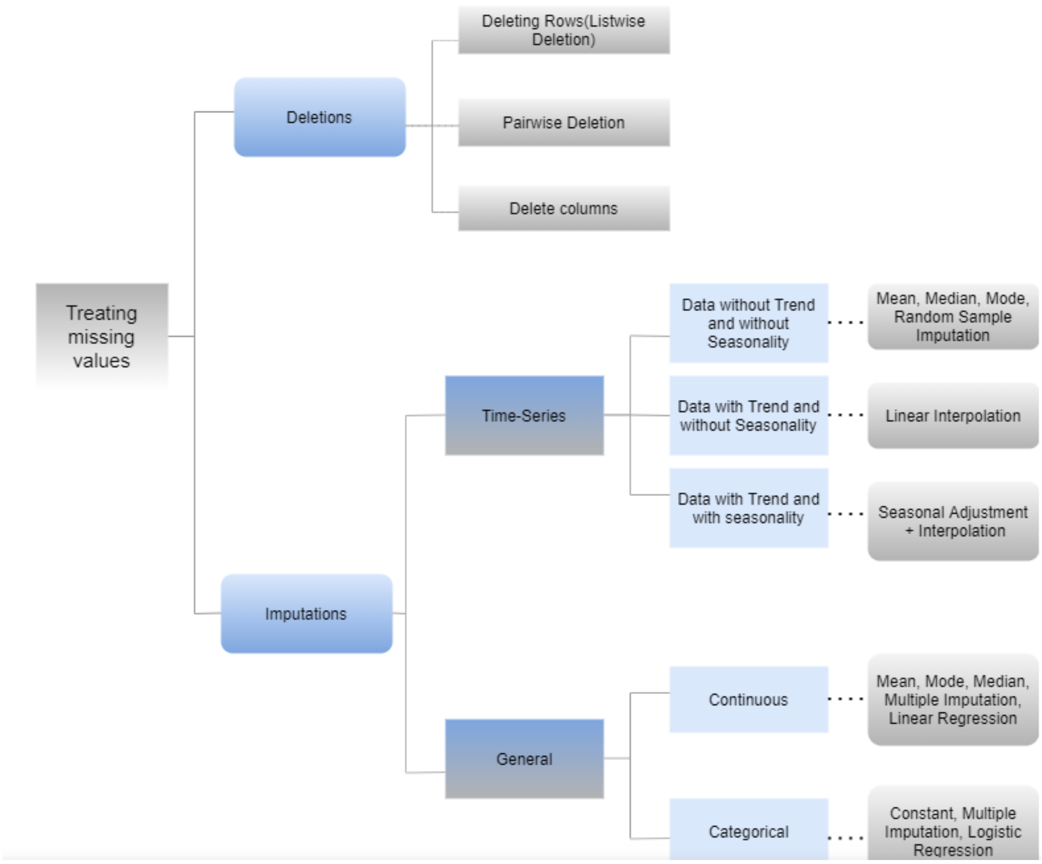
1. Data collection < As explained above>
2. Data preparation:

Once the data is collected, it then enters the data preparation stage. Data preparation, often referred to as “pre-processing” is the stage at which raw data is cleaned up and organized for the next stage of data processing. During preparation, raw data is diligently checked for any errors and omissions. The purpose of this step is to eliminate bad data (redundant, incomplete, or incorrect data) and create high-quality data for the right business intelligence.

Handling Missing Data: This is a major step in cleaning of data. There are two ways to handle this. Deletions or Imputation.

- The deletion methods only work for certain datasets where participants have missing fields. Deletion refers to eliminating rows which do not have complete data that’s required.
- Use regression analysis to systematically eliminate data.
- Data scientists can use data imputation techniques. In Imputation data is filled into missing cells using statistical techniques.

The following diagram explains the process of treating missing values through both the methods:



3. **Data input**: The clean data is then entered into its destination and translated into a language that it can understand. Data input is the first stage in which raw data begins to take the form of usable information.
4. **Processing:** During this stage, the data input to the computer is processed for interpretation. Processing is done using statistical modelling and machine learning algorithms, though the process itself may vary slightly depending on the source of data being processed and its intended use.
5. **Data output/interpretation**: The output/interpretation stage is the stage at which data is finally usable to non-data scientists. It is translated, readable, and often in the form of graphs, videos, images, plain text, etc.
6. **Data storage and Report Writing**: The final stage of data processing is storage. After all of the data is processed, it is then stored for future use. While some information may be put to use immediately, some of it will serve a purpose later on. Also, properly stored data is a necessity for compliance with data protection legislations. When data is properly stored, it can be quickly and easily accessed by members of the organization when needed.

Classification of Data

Classification or categorization is the process of grouping the statistical data under various understandable homogeneous groups for the purpose of convenient interpretation. A uniformity of attributes is the basic criterion for classification; and the grouping of data is made according to similarity. Classification becomes necessary when there is a diversity in the data collected, for meaningful presentation and analysis. However, it is not required for homogeneous data. A good classification should have the characteristics of clarity, homogeneity, equality of scale, purposefulness and accuracy.

Objectives of Classification are below:

1. The complex scattered and haphazard data is organized into concise, logical and intelligible form.
2. It is possible to make the characteristics of similarities and dis – similarities clear.
3. Comparative studies is possible.
4. Understanding of the significance is made easier
5. Underlying unity amongst different items is made clear and expressed.
6. Data is so arranged that analysis and generalization is possible.

14.3 BIG DATA MANAGEMENT

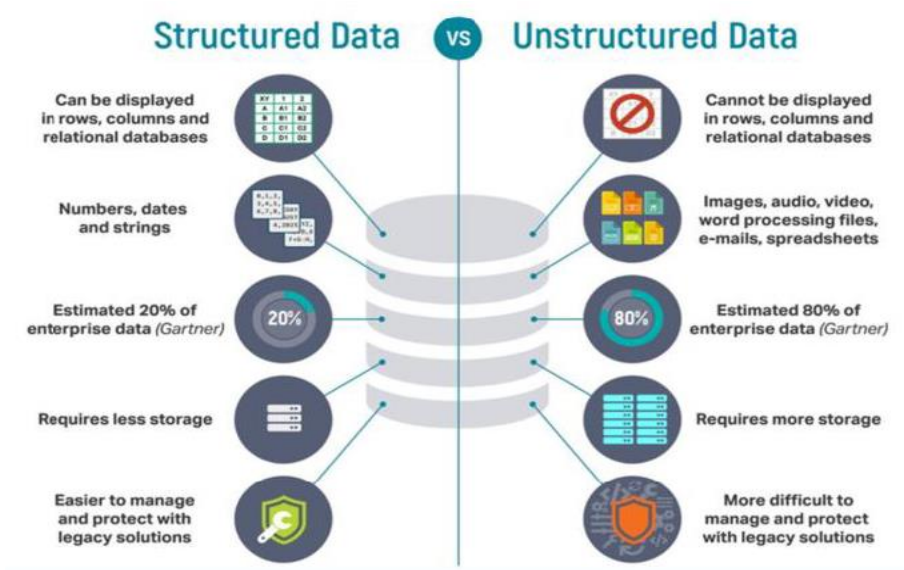
Big data is data (structured, semi structured and unstructured) whose scale, diversity, speed and complexity are huge.

Data is created continuously and at an ever-increasing rate. There are multiple sources of data creation such as Mobile phone apps, social media, imaging technologies, Communications, Videos, Sound files, sensors, and so on and that must be stored and processes in real time somewhere for some purpose. Keeping up with this huge influx of data and analysing the same is quite challenging, especially when it does not conform to traditional notions of data structure, to identify meaningful patterns and extract useful information.

Big Data is creating significant opportunities for organizations to utilise their most valuable asset – Data – to create compelling value propositions and competitive advantage. For businesses, Big Data helps drive efficiency, quality, and customer engagements through better personalisation of products and services. In many cases, Big Data analytics integrate structured, semi-structured and unstructured data with real time feeds, opening new paths to insights.

Several industries have led the way in developing their ability to gather and exploit big data: Example., Identifying credit card fraud situations (in near real time) to alert customers based on billions of transactions, Sending buying recommendations based on customer's buying behaviour in near real time

Difference between Structured and Unstructured data:



Big Data characteristics:

- **Volume of data:** Big Data can be billions of rows and millions of columns, in contrast to millions in traditional databases
- **Variety:** Complexity of data types and structures like Numbers, Text, Images, Videos, Messages, and so on.
- **Velocity:** Speed of generation of Data and growth: Big Data can describe high velocity data, with rapid data ingestion and near real time analysis.

Businesses can use advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics and natural language processing to gain new insights from previously untapped data sources independently or together with existing enterprise data.

Benefits & Advantages of Big Data Analytics

1. **Risk Management:** Credit card companies use Big Data analytics to identify fraudulent activities and discrepancies. The organization leverages it to narrow down a list of suspects or root causes of problems.
2. **Product Development and Innovations:** Rolls-Royce, one of the largest manufacturers of jet engines uses Big Data analytics to analyse how efficient the engine designs are and if there is any need for improvements.
3. **Quicker and Better Decision Making within Organizations:** Many organisations use Big Data analytics to make strategic decisions. For example, the companies leverage it to decide if a particular location would be suitable for a new outlet or not.
4. **Improve Customer Experience:** Air Lines uses Big Data analysis to improve customer experiences. They monitor tweets to find out their customers' experience regarding their journeys, delays, and so on

Big Data Analytics Process

- **Step 1 -** Business case: The Big Data analytics lifecycle begins with a business case, which defines the reason and goal behind the analysis.
- **Step 2 -**Data Sources and Collection: Key data sources are identified and data is collected
- **Step 3 -** Data Cleaning: All of the identified data from the previous step is filtered here to remove incomplete data, or impute incomplete data, and remove unclean data
- **Step 4 -** Data extraction - Data is transformed into a compatible form for the specific tool of analysis
- **Step 5 -** Data aggregation - In this stage, data with the same fields across different datasets are integrated.
- **Step 6 -** Data analysis - Data is analysed, and evaluated using analytical and statistical tools to discover useful information.
- **Step 7 -** Visualization of data - With tools like Tableau, Power BI, and QlikView, Big Data analysts produce graphic visualizations of the analysis.
- **Step 8 -** Final analysis result - This is the last step of the Big Data analytics lifecycle, where the final results of the analysis are made available to business stakeholders who will take action.

14.4 DATA VISUALISATION

Humans are constantly bombarded by information from multiple sources and through multiple channels. Our five sensory perceptions help our brain process this information. However, the brain is highly selective about what it processes and humans are also limited by the amount of information they can assimilate simultaneously. Data visualizations, when designed and executed well, help in reducing this cognitive load, and assist viewers in the evaluations. A good visualization that is focused on a task, highlights key information is of significantly more value.

Based on multiple researches six thumb rules have been defined for Data visualisation:

1) The simplest chart is usually the one that communicates most clearly.
2) Always directly represent the relationship you are trying to communicate. Do not leave it to the viewer to derive the relationship from other information
3) In general, do not ask viewers to compare in two dimensions. Comparing differences in length is easier than comparing differences in area
4) Never use colour on top of colour—colour is not absolute, but relative
5) Do not violate the primal perceptions of your viewers. Remember, up means more
6) Chart with graphical and ethical integrity. Do not lie, either by mistake or intentionally

Rule 1: Simplicity Vs Complexity

First, Is a Visualization Needed? An initial analysis of the data is required to determine the necessity of a visual representation. Example., if all kids in a class are 5' in height, a visual representation does not make sense.

In most cases, the choice of visualisation is between a Table and a graph. How to decide between a Table or a Graph?

Tables are best used when:

- The display will be used as a lookup for particular values.
- It will be used to compare individual values not groups or series of values to one another.
- Precision is required.
- Quantitative information to be provided involves more than one unit of measure.
- Both summary and detail values are included.

Which Graph to Use ?

The process of choosing the correct graph is fundamentally linked to the relationship being communicated. For each kind of relationship, there are multiple kinds of graphs that might be used

How Many Dimensions to Represent?

When representing data visually, we must decide how many dimensions to represent in a single graph. The maximum number of data dimensions that can be represented in a static graph is five and in an interactive graph is six. The table here provides a list of the most likely graphs to be used for various relationships and numbers of dimensions.

Dimensions:

1. X-axis placement
2. Y-axis placement
3. Size
4. Shape
5. Colour
6. Animation (interactive only, often used to display time)

Relationship	Most likely graph(s)	Keywords	Max. # of dimensions
Time series	Trend line Column chart Heat map Sparklines	Change Rise Increase Fluctuation Growth Decline/decrease Trend	4
Part to whole	Stacked column chart Stacked area chart Pareto chart (for two simultaneous parts to whole)	Rate or rate of total Percent or percentage of total Share "Accounts for X percent"	4
Ranking	Sorted bar/column chart	Larger than Smaller than Equal to Greater than Less than	4
Deviation	Line chart Column/bar chart Bullet graph	Plus or minus Variance Difference Relative to	4
Distribution	Box/whisker plot Histogram	Frequency Distribution Range Concentration Normal curve, bell curve	4
Correlation	Scatterplot Table pane	Increases with Decreases with Changes with Varies with	6
Geospatial	Choropleth (filled gradient) map	N/A	2

(Courtesy: Pochiraju and Sheshadri, Essentials of Business Analytics, Springer 2019)

Rule 2: Direct Representation

It is necessary that the data visualizer doesn't let the viewer interpret relationships, rather provides the story directly. For instance, if we wish to tell a story of differences, such as deviations from plan, and budgets vs. actual, do not rely on the viewer to calculate the differences themselves.

Rule 3: Single Dimensionality

Always stick to representing information in one dimension. Example comparing lengths is easier than calculating areas for the viewer. Creating 3D images could distort the story with its dependency on viewing angles.

Rule 4: Colours

Colour is not perceived absolutely by the human eye. Its always relative. Hence care needs to be exercised in choice of colours. When representing levels of a single variable, use a single-color gradient, when representing categories of a variable, use rainbow colours. Using colours that are near to nature are soothing to eyes and primary colours can be used for emphasis.

Rule 5: Go with established general conventions

Generally UP would mean increasing and DOWN would mean decreasing. Violating this could make the viewer read the graph wrong.

Rule 6: Tell the story with integrity

This deals with intentional or unintentional misrepresentation of a visual with respect to its data. The effect of the data must be accurately reflected in the visuals.

CHECK YOUR PROGRESS

1. The three Vs. of big data are _______,___________,__________
2. In Data visualisation a good representation constitutes ___ Dimensions (Options: 1,2,3,5)
3. In Data visualisation Colours are _________ (Options: Absolute, Relative)
4. The two methods of addressing missing values are __________, _______________
5. What is big data management?

14.5 CASE STUDY

Go to the census of India website and download the state demographics table for Karnataka State, which has the Education, Age, Gender, and migration reasons. Create visualisation of the same using the Six Rules.

Hint:

ii) The table will be available in xls. Download the same

iii) Eliminate all columns other than those required

iv) Look for the ideal representation choices – Table or Graph?

14.6 SUMMARY

Data processing and analysis is a key ingredient of Business analytics. Collecting Data, Cleaning the data, pre processing, Transforming and analysing and storing are all part of this. With the advent of digital technologies and rapid evolution and growth of data engineering and data science products, Big data analytics has entered into most businesses, giving them the benefit of Predictive and prescriptive Analytics almost in real time.

Data visualisation is very important to convey the right inferences to the viewers, and to do that there are rules based on scientific research.

14.8 KEYWORDS

Data, Data Processing, Data Analysis, Big Data, Data visualisation, Rules for Data visualisation

14.9 SELF ASSESMENT QUESTIONS

1. Explain the stages of data processing
2. Discuss the characteristics of big data management
3. Examine the rules of data visualization

14.10 ANSWERS TO CHECK YOUR PROGRESS

1. Volume, Velocity and Variety
2. 1
3. Relative
4. Deletion, Imputation
5. *Big data management* is the organization, administration and governance of large volumes of both structured and unstructured
6. data.

14.11 REFERENCES

Books

- Business Analytics: Data Analysis and Decision Making, 5th Edition, Albright and Winston
- Essentials of Business Analytics, Springer 2019, Pochiraju and Sheshadri

UNIT 15: DATA MINING AND ANALYSIS

STRUCTURE

15.0 Objectives

15.1 Introduction

15.2 Data Mining

15.3 Data Mining Methods

15.4 OLAP and multi-dimensional data analysis

15.5 Case Study

15.6 Notes

15.7 Summary

15.8 Keywords

15.9 Self-Assessment Questions

15.10 Answers to Check Your progress

15.11 References

15.0 OBJECTIVES

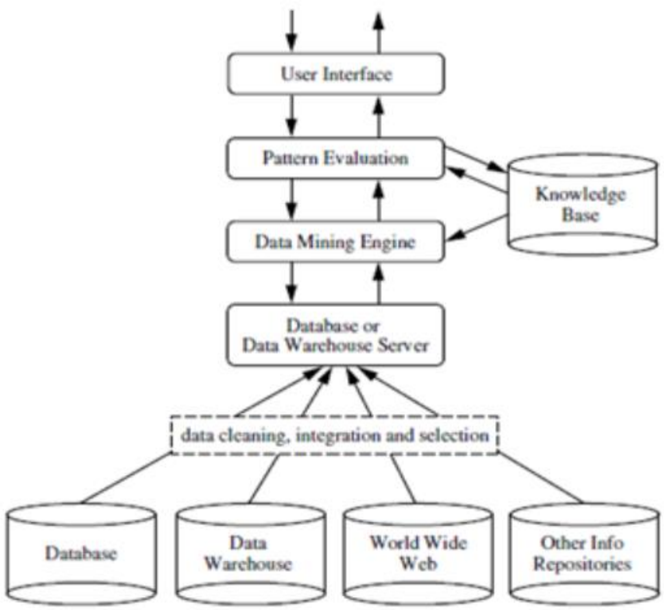
After completing this unit, you should be able to:

- Explain Data mining, its process and types
- Appreciate OLAP as a tool for decision making
- Practice key data mining methods of Cluster Analysis and Association rules

15.1 INTRODUCTION

The process of extracting information to identify patterns, trends, and useful data that would allow the business to take data-driven decisions from large sets of data is referred to as Data Mining. Data mining is also called as *Knowledge Discovery in Database (KDD).* KDD includes Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation, and Knowledge presentation.

15.2 DATA MINING



Data mining is one of the most useful techniques that help entrepreneurs, researchers, and individuals to extract valuable information from huge sets of data.

Data mining utilizes complex mathematical algorithms and evaluates the probability of future events. The process could include various types of services such as text mining, web mining, audio and video mining, pictorial data mining, and social media mining. There are many powerful techniques and software's available to mine data and find better insight from it.

History of Data Mining

In the 1990s, the term "Data Mining" was introduced. Early techniques of identifying patterns in data include Bayes theorem (1700s), and the evolution of regression(1800s). The rapid strides in both computing and storage technologies as well as Data Engineering and Data science software's have boosted data collection, storage, and manipulation. Explicit hands-on data investigation has progressively being replaced automatic data processing, and applying neural networks, clustering, genetic algorithms (1950s), decision trees(1960s), and supporting vector machines (1990s). Data mining origins are traced back to three family lines: Classical statistics, Artificial intelligence, and Machine learning.

Data Mining on different types of Databases

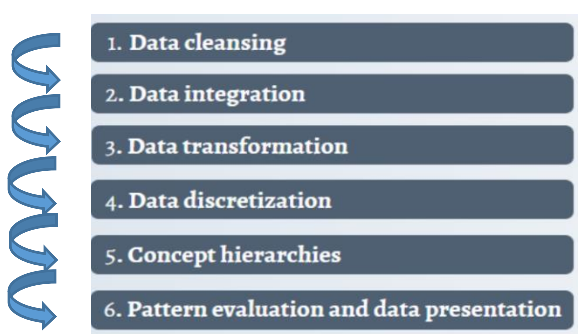
Data mining can be performed on the following types of data:

- Relational Database: A relational database is a collection of multiple data sets formally organized by tables, records, and columns. Tables convey and share information, which facilitates data search ability, reporting, and organization.
- Data warehouses: A Data Warehouse is the technology that collects the data from various sources within the organization into a cohesive system. The huge amount of data comes from multiple places such as Operations, Shop Floor, Marketing and Finance. The data warehouse is designed for the analysis of data rather than transaction processing.
- Object-Relational Database: A combination of an object-oriented database model and relational database model is called an object-relational model. It supports Classes, Objects, Inheritance, etc.
- Transactional Database: A transactional database refers to a database management system (DBMS) that has the potential to undo a database transaction if it is not performed appropriately.

Data Mining Process:

The diagram here depicts the Data Mining Process:

Data Mining Process



Data cleansing: It involves identifying and removing inaccurate and ambiguous data from a set of tables, databases, and record sets. This is achieved through standard Missing Data techniques like deletion or Imputation.

Data integration: This is about merging a new set of Data with an existing set. There could be multiple sources.

Data transformation: This requires transforming data within formats, generally from the source system to the required destination system. Some strategies include Smoothing, Aggregation, Normalization, Generalization, and attribute construction.

Data discretization: This is stage where the continuous attribute domain along intervals is split into manageable chunks. Two strategies involve Top-down discretization and bottom-up discretization.

Concept hierarchies: This minimizes the data by replacing and collecting low-level concepts from high-level concepts. Concept hierarchies define the multi-dimensional data with multiple levels of abstraction. The methods are Binning, histogram analysis, cluster analysis, etc.

Pattern evaluation and data presentation: Data is presented efficiently using Data visualisation best practices. Then the client and the customers can make use of it in the best possible way.

Advantages of Data Mining

- The Data Mining technique enables organizations to obtain knowledge-based data.
- Data mining enables organizations to transform operations and production.
- Data mining is cost-efficient.
- Data Mining helps the decision-making process
- Facilitates automated discovery of hidden patterns

Data Mining Applications

- Data Mining is primarily adopted by organizations needing a high level of customer engagement. Example., Data mining enables a retailer to use point-of-sale records of customer purchases to develop products and promotions that help the organization to attract the customer. Examples.,
- **Healthcare:** Data Mining can be used to forecast patients in each speciality of the health system. Example., The process ensures that the patients get intensive care at the right place and at the right time. Data mining also enables healthcare insurers to recognize fraud and misuse.
- **Market Basket Analysis:** Enables retailers to understand the purchase behaviour of a buyer and altering the store's layout accordingly.

Challenges of Implementation in Data mining

- Incomplete and noisy data: The data in the real-world is heterogeneous, incomplete, and noisy. These problems may occur due to data measuring instruments or because of human errors.
- Data Distribution: Quite a tough task to bring all the data to a centralized data repository mainly due to organizational and technical concerns.
- Complex Data: Real-world data is heterogeneous, and it could be multimedia data, including audio and video, images, complex data, spatial data, time series, and so on. Managing these various types of data and extracting useful information is a tough task.
- Performance: The data mining system's performance relies primarily on the efficiency of algorithms and techniques used.
- Data Privacy and Security: Data mining could lead to issues in terms of data security, governance, and privacy. For example, predicting consumer behaviour without their consent.

Data mining includes the utilization of refined data analysis tools to find previously unknown, patterns and relationships in huge data sets. These tools can incorporate statistical models, machine learning techniques, and mathematical algorithms, such as neural networks or decision trees. Thus, data mining incorporates analysis and prediction.

15.3 DATA MINING METHODS

There are multiple methods of Data Mining, but the important step is to select the most appropriate from them according to the business or the problem statement. These methods help in predicting the future outcomes and then making decisions accordingly.

Some of the Methods are listed here:

- Association
- Classification
- Clustering Analysis
- Prediction
- Sequential Patterns or Pattern Tracking
- Decision Trees
- Outlier Analysis or Anomaly Analysis
- Neural Network

Clustering and Association Analysis

Clustering refers to a technique of grouping objects so that objects with the same functionalities come together and objects with different functionalities go apart. In other words, we can say that clustering is a process of partitioning a data set into a set of meaningful subclasses, known as clusters. The grouping is achieved by determining similarities between data according to characteristics found in the real data. The groups are called Clusters.

Methods of clustering

- Partitioning methods
- Hierarchical clustering
- Fuzzy Clustering
- Density-based clustering
- Model-based clustering

Association Analysis:

This data mining technique helps to discover a link or relationship between two or more items. It finds a hidden pattern in the data set.

Association rules are if-then statements that support to show the probability of interactions between data items within large data sets in different types of databases. Association rule mining has several applications and is commonly used to help sales correlations in data or medical data sets. Example, Buying spices together with Oil or Buying socks with shoes or buying Mobile cover with a Mobile.

There are two types of Association Rules:

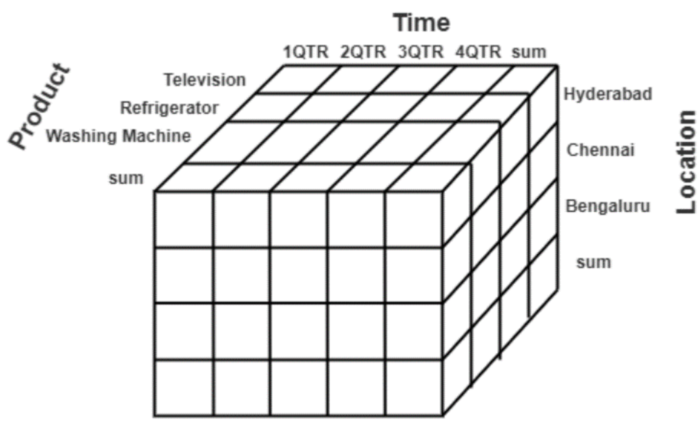
- Single dimensional association rule: These rules contain a single attribute that is repeated.
- Multidimensional association rule: These rules contain multiple attributes that are repeated.

15.4 OLAP AND MULTI-DIMENSIONAL DATA ANALYSIS

OLAP stands for Online Analytical Processing. It is a computing method that allows users to extract useful information and query the data in order to analyse it from multiple perspectives. For example, OLAP business intelligence queries usually aid in setting Production targets, Pricing, financial reporting, budgeting, predict future sales, trends analysis and other purposes. It enables the user to analyse information from different database systems simultaneously. OLAP data is stored in multidimensional databases.

A Multi-dimensional Data Model allows customers to interrogate analytical questions associated with market or business trends. They allow users to rapidly receive answers to the requests which they made by creating and examining the data comparatively faster. OLAP (online analytical processing) and data warehousing use multi-dimensional databases. It is used to show multiple dimensions of the data to users.

It represents data in the form of data cubes. Data cubes allow users to model and view the data from many dimensions and perspectives. It is defined by dimensions and facts and is represented by a fact table. Facts are numerical measures and fact tables contain measures of the related dimensional tables or names of the facts.



Multidimensional Data Representation

OLAP and data mining look similar since they operate on data to gain knowledge, but the major difference is how they operate on data. OLAP tools provide multidimensional data analysis and a summary of the data.

Key features of OLAP

- Supports complex calculations
- Time intelligence
- Multidimensional and flexible view of data
- Business-focused calculations
- Self-service reporting

Differences between Data Mining and OLAP

Data Mining	OLAP
Focused on data summary.	Focused on transaction-level data.

Discovery-driven.	Query driven.
Future dG39ata prediction.	Analysing past data.
Larger number of dimensions.	Limited number of dimensions.
Bottom-up approach.	Top-down approach.

CHECK YOUR PROGRESS

1. Which of the following is an essential process in which the intelligent methods are applied to extract data patterns?
   a. Warehousing
   b. Data Mining
   c. Text Mining
   d. Data Selection
2. Which of the following can be considered as the correct first step in the process of Data Mining?

a. Data Cleaning

b. Data Transformation

c. Data Analysis

d. Interpretation

3. What is KDD in data mining?
   a. Knowledge Discovery in Database
   b. Knowledge Discovery Data
   c. Knowledge Data definition
   d. Knowledge data house
4. What is the full form of OLAP?
5. What are neural networks?

15.5 CASE STUDY

< This is a case study from Springer website as a reference, to help you practice the concept of Datamining. >

The online retailer under consideration is a UK-based and registered non-store business with some 80 members of staff. The company was established in 1981 mainly selling unique all-occasion gifts. For years in the past, the merchant relied heavily on direct mailing catalogues, and orders were taken over phone calls. It was only 2 years ago that the company launched its own web site and shifted completely to the Web. Since then the company has maintained a steady and healthy number of customers from all parts of the United Kingdom and Europe, and has accumulated a huge amount of data about many customers. The company also uses Amazon.co.uk to market and sell its products.

The customer transaction dataset held by the merchant has 11 variables as shown in Table 1, and it contains all the transactions occurring in years 2010 and 2011. It should be noted that the variable *Post Code* is essential for the business as it provides vital information that makes each individual consumer recognizable and tractable, and therefore it makes some in-depth analyses possible in the present study.

As the first ever pilot study for the business to generate sensible customer intelligence, only the transactions created from 1 January 2011 to 31 December 2011 are explored. Over that particular period, there were 22 190 valid transactions in total, associated with 4381 valid distinct postcodes. Corresponding to these transactions, there are 406 830 instances (record rows) in the dataset, each for a particular item contained in a transaction. On average, each postcode is associated with five transactions, that is, each customer has purchased a product from the online retailer about once every 2 months. In addition, only consumers from the United Kingdom are analysed.

It is interesting to notice that the average number of distinct products (items) contained in each transaction occurring in 2011 was 18.3 (=406 830/22 190). This seems to suggest that many of the consumers of the business were organizational customers rather than individual customers.

The data is available in the article associated with the following URL as open source data.

https://link.springer.com/article/10.1057/dbm.2012.17

The retailer is concerned with the following common business issues:

1. Who are the most/least valuable customers to the business? What are the distinct characteristics of them?
2. Who are the most/least loyal customers, and how are they characterized?
3. What are customers’ purchase behaviour patterns? Which products/items have customers purchased together often? In which sequence the products have been purchased?

A sample of the data available is in the screenshot below:

Filter and Sort | Query Builder | Data • Describe • Graph • Analyze • | Export • Send To •

	Buyer	First_Purchase	Recency	Frequency	Monetary	Min	Max	Mean
2392	XX18 4ES	12	12	1	306.84	1.25	30	7.14
2393	XX2 4HG	4	4	1	1487.6	237.6	1250	743.80
2394	XX2 5XX	5	0	5	1952.45	0.39	76.32	22.19
2395	XX20 6AB	11	1	2	483.26	0.84	17.34	4.56
2396	XX20 6AQ	7	7	1	157.9	10.5	75	26.32
2397	XX20 6HJ	3	1	3	541.46	1.65	59.4	15.93
2398	XX21 8YN	1	1	1	167.62	0.19	15.12	3.64
2399	XX22 3RB	3	1	3	454.01	2.9	25.5	9.27
2400	XX24 8SA	11	11	1	344.14	1.45	40.56	11.87
2401	XX25 5HU	10	2	4	1498.51	0.12	45	5.01
2402	XX26 1AG	12	3	4	226.75	13.05	26.1	18.90
2403	XX26 3QN	8	8	1	185.65	15	35.7	23.21
2404	XX30 6AP	12	8	3	215.72	0.58	47.7	4.79
2405	XX30 6BJ	8	8	1	175.27	0.65	60	6.04
2406	XX31 7AD	11	4	4	1133.07	0.29	41.5	5.29
2407	XX31 7JF	8	4	3	1240.2	9.12	264	53.92
2408	XX31 7JN	8	2	4	2404.17	0.39	74.25	9.50
2409	XX33 0EA	9	3	2	993.18	8.5	39.6	19.86
2410	XX33 0EN	1	1	2	388.79	0.39	34.68	4.68
2411	XX33 0QL	12	1	7	2827.93	0.29	142.8	10.47
2412	XX34 1HA	7	7	1	2044.37	5.04	207.5	37.17
2413	XX34 2DS	1	1	1	161.67	2.5	19.8	12.44
2414	XX34 2HA	8	8	1	116.01	2.1	30	12.89
2415	XX34 3OA	11	7	2	309.36	1.1	17.4	4.91
2416	XX37 6QU	1	1	2	193.42	0.19	11.8	3.28
2417	XX39 4HL	2	2	1	390.07	0.42	18.72	3.90

Hint: Basic data warehouse (DW) implementation phases are:

- Current situation analysis
- Selecting data interesting for analysis, out of existing database
- Filtering and reducing data
- Extracting data into staging database
- Selecting fact table, dimensional tables and appropriate schemes
- Selecting measurements, percentages of aggregations and warehouse methods
- Creating and using the cube

15.6 SUMMARY

Data Mining and OLAP are two essential techniques for users to extract valuable information from their huge datasets. While they look similar, they have different connotations, Data mining focuses on discovery while OLAP focuses on Query.

There are multiple methods of Data mining which exploit inter relationships or similarities in data and affinity of different sets of information to each other.

15.7 KEYWORDS

Data Mining, OLAP, Clustering, Association Rule, Knowledge Discovery in Database (KDD)

15.8 ANSWERS TO CHECK YOUR PROGRESS

1. b
2. a
3. a
4. Online Analytical Processing
5. A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature.

15.9 SELF ASSESSMENT QUESTIONS

1. What is data mining? Explain its advantages and applications
2. Differentiate between cluster analysis and supply chain analytics
3. Explain multi-dimensional data analysis

15.11 REFERENCES

Books

- Business Analytics: Data Analysis and Decision Making, 5th Edition, Albright and Winston
- Essentials of Business Analytics, Springer 2019, Pochiraju and Sheshadri

URLS https://link.springer.com/article/10.1057/dbm.2012.17

UNIT 16: APPLICATIONS OF BUSINESS ANALYTICS

STRUCTURE

16.0 Objectives

16.1 Introduction

16.2 Meaning and Definitions

16.3 Application of business analysis

16.4 Retail analytics

16.5 Marketing Analytics

16.6 Financial Analytics

16.7 Healthcare Analytics

16.8 Supply chain analytics

16.9 Summary

16.10 Keywords

16.11 Self-Assessment Questions

16.12 Answers to Check Your progress

16.13 References

16.0 OBJECTIVES

After completing this unit, you should be able to:

- Appreciate the use of Business analytics for making decisions
- Examine the role of Analytics in Finance, HR, Healthcare, Supply chain, Retail and marketing domains

16.1 INTRODUCTION

Business analytics is defined as the process of understanding the data of a business to draw inferences and make calculated decisions with higher certainty of outcomes. Business analytics encompasses a gamut of analysis around business data to draw information that could be used by the leaders at various levels in an organization. Business analytics enables fact-based decision making while extending accountability in decision making.

The developments in real-time warehousing, Business intelligence and reporting capabilities in business systems, has evolved business analytics as a practical choice for strategic decision making.

16.2 APPLICATION OF BUSINESS ANALYSIS

The core of analytics is the application of logic and processes to find meaning or insights in data. Hence, business analytics measures the outcomes/results that are produced and then, provides a feedback loop that facilitates organizational learning. The table attached presents how analytics can engage different business functions and the associated benefits to the End users (internal or external). Thus the business gets sufficient and valid reasons to build a newer analytics driven value proposition for its stakeholders, such as customers, suppliers, employees, regulators, or others.

Analytics type	Purpose
Human resource analytics	Managing human capital
Employee performance	Managing employee performance
Staffing and scheduling	Managing employee productivity
Supply chain analytics	Managing supply networks better
Sourcing analytics	Vendor management
Supplier rating	Vendor risk management
Inventory analytics	Inventory management
Customer analytics	Unified customer value management
New business and acquisition	New portfolio management
Business process analytics	Resilience, response and agility of delivery
Collections and recovery analytics	Recovery management
Risk analytics	Risk management
Marketing analytics	Marketing process
Portfolio analytics	Business management
Sales analytics	Sales planning and sales force management
Spatial analytics	Accuracy and timeliness in delivery
Channel analytics	Improve channel efficiency and return on investments and brand equity

16.3 RETAIL ANALYTICS

Typically Retail is the last stop for a supply chain by delivering products directly to customers. Hence the retailers have to focus on collecting data on customers, their behaviour and preferences and incorporating these into business decisions. And so, retail has been an early adopter of analytics methodologies.

To understand how analytics could influence retail, it is useful to break down the business components and associated decisions taken in retail as follows:

- **Consumer:** Personalization is a key consumer-level implementation that retail firms make. Personalized pricing, discounts via coupons, loyalty cards, personalised product recommendations all could lead to better customer retention through better engagement and experience.
- **Product:** Retail product decisions like inventory decisions, pricing and assortment planning can be made more efficient with analytics.
- **Human resources:** The key decisions here are staffing related – Numbers, shifts, skills, and so on.
- **Advertising:** Here the key decision are around the media of advertisement and corresponding Product mix to advertise.

Examples of Retail Analytics

- An online e-commerce platform was seeing that customers were putting specific items into shopping cart but were not confirming the purchase with their debit or credit cards. Analytics revealed that such purchases were for a specific value range of the product. The company introduced Cash On Delivery (COD) for such items and the sales improved.

- Machine learning algorithms sniff out social media posts and Web browsing habits of consumers to elicit what products are creating a social buzz and correlate the same with Advertisement buying data. Such sentiment analysis can predict the top selling products.

Challenges in Retail Analytics

The typical challenges that arise in retail scenarios that need to be overcome for the successful use of retail analytics are classified into two buckets:

(a) Those that affect predictive modelling: Demand censoring (due to poor access to demand data) and inventory inaccuracies (software glitches, human errors)

(b) Those that affect decision-making: changing prices, physical constraints on assortments, supplier lead times, supplier contracts, and constraints on workforce.

16.4 MARKETING ANALYTICS

Robust use of analytics tools has helped firms increase performance in terms of sales, revenues, profits, customer satisfaction, and winning against the competition. Marketing analytics is the creation, aggregation and use of data to measure and optimize marketing decisions.

Marketing analytics revolve around the basic aspects of marketing such as target marketing and segmentation, price and promotion, customer valuation, resource allocation, response analysis, demand assessment, and new product development. However, these analytics can be applied at three specific levels:

- Firm: Tools are applied to the firm as a whole and helps in resource allocation.
- Brand/product: At the brand/product level, tools are applied to decide and evaluate strategies for a particular brand/product. For example, find how a particular brand advertisement will be received by the market.
- Customer: Tools applied at customer level provide insights that help in segmenting and targeting customers (Customer Lifetime Value, Customer Referral value, Customer Influence Value).

Key Techniques Used:

- **Conjoint Analysis:** Conjoint analysis is a marketing research technique to determine consumer preferences and potential customers. Conjoint analysis can be productively applied in the following areas:
  - Designing products that maximize the measured utilities for customers
  - Modifying existing products and developing new products
  - Selecting market segments for which a given product delivers high utility
  - Planning competitive strategy
  - Analysing pricing policies
- Stochastic frontier analysis this is a parametric approach, largely used to estimate the production or costs. SFA relies on the assumption that decision-making units behave sub-optimally and they can maximize or minimize their respective objective functions (costs, profits, operational efficiency)
- **Data Envelopment Analysis (DEA):** A very prominent analytic tool that can be used in decision making, it is designed to help managers measure and improve the performance of their organizations. DEA can help capture the efficiency of each unit and suggest the potential factors that may be responsible for making units efficient.

16.5 FINANCIAL ANALYTICS

Data analytics in finance is a part of quantitative finance. Quantitative finance primarily consists of three sectors in finance-asset management, banking, and insurance. Across these three sectors, there are four tightly connected functions in which quantitative finance is used-valuation, risk management, portfolio management, and performance analysis.

Financial Analytics helps a business to:

- Understand the performance
- Measure and manage the value of tangible and intangible assets
- Manage the investments
- Forecast the variations in the market
- Increase the functionalities of information systems
- Improve the business processes and profits

Basics of Finance

Some of the key terms in finance are explained here.

Liquidity: Liquidity in a business means the availability of cash and other assets to pay its debts, bills and other expenses. The liquidity level of the company differs from period to period because of certain factors like sales, economy and seasons.

Leverage: Leverage refers to the amount of finance which a company has borrowed from outside to run its operations as against its investment. Leverage is an important factor which is considered mainly by bankers and investors. A company will have a high leverage ratio when the debt of the company is high when compared to its equity.

Profitability: Profitability refers to the return that the business earns from the amount invested in the business.

Financial Ratios

There are also few ratios that will help in the overall financial analysis. Financial ratios are easy to calculate and simple to use. These ratios will tell where there needs to be an improvement in the business.

- **Current Ratio –** Exhibits the ability of a company to pay its near term obligations
- **Quick Ratio –** Explains the company's ability to pay its current liabilities
- **Liquidity Ratio –** This calculates the liquidity of the company by taking everything into consideration except cash
- **Debt or Equity Ratio –** This indicates the ratio of the company's investor vs supplied capital
- **Return on Equity Ratio –** This measures the company's level of profitability Quantitative finance primarily consists of three sectors in finance

    i) Asset management

    ii) Banking

    iii) Insurance

There are four tightly coupled building blocks across these three, in which Financial Analytics us used:

i) Valuation

ii) Risk management

iii) Portfolio management

iv) Performance analysis.

Quantitative finance can be distributed into two branches with a small overlap:

- Q-Quant: Risk neutral, and

- P-Quant: Risk-averse

In general, The Q-quants primarily work in the sell-side and are price-makers as opposed to P-quants who work in the buy-side and are typically price-takers.

Measuring Risk

Two key metrics used are:

- Value at Risk (VaR): Value at Risk describes the loss in a portfolio that can occur over a given period, at a given confidence level, due to exposure to market risk
- Threshold Persistence (TP): Given a threshold level of return for a given portfolio, traders and risk managers want to estimate how frequently the cumulative return on the portfolio goes below this threshold and stays below this threshold for a certain number of days.

16.6 HEALTHCARE ANALYTICS

The development of technologies related to Data capture, storage and analysis over the past 20 years has begun to revolutionize the use of data in all sectors within Healthcare. For example, data analysis of millions of records allows earlier detection of epidemics, detect insurance frauds, personalise premiums, identification of molecules, telemedicine, and new methods to evaluate the efficacy of vaccination programs.

For hospitals and healthcare managers, healthcare data analytics provide a combination of financial and administrative data alongside information that can aid patient care efforts, better services, and improve existing procedures. Healthcare Business Intelligence Software tend to emphasize broad categories of data for collection and parsing: costs and claims, research and development, clinical data alongside patient behaviour and sentiment.

i) Costs and Claims: Analytics here helps administrators in identifying areas to streamline operations and increase savings

ii) Research and development: This is crucial for providing new innovative solutions and treatment options

iii) Clinical data: This is vital for administrators to determine what areas of their service need to improve and offer more granular information regarding treatment effectiveness, success rates, and more.

iv) Patients sentiment: This is important to understand how patients are feeling and are reacting to service

Examples:

Telemedicine: Historically, diagnosing illness has required medical professionals to assess the condition of their patients face-to-face. Understanding various aspects about the body parameters that help doctors diagnose and prescribe a treatment often requires the transmission of information that is subtle and variable. Hearing the rhythm and rate of a heart, assessing the degradation in a patient's sense of balance, or seeing nuances in a change in the appearance of a wound are thought to require direct human contact. Big data techniques are expected to revolutionize this due to the practice of telemedicine.

Wearables: New wearable technologies can assist caregivers and nutritionists by collecting data of a person, over longer durations and consistently. Algorithms can use this data to suggest alternate courses of action while ensuring that new or unreported symptoms are not missed. Wearable technologies such as a Fit bit or Apple Watch are able to continuously track various health-related factors like heart rate, body temperature, and blood pressure with ease. Any changes that are life threatening can be addressed more proactively.

Drug Discovery: The ability to conduct in-silico simulations of drug interactions has reduced the drug discovery cycle significantly leading to faster releases to Markets.

16.7 SUPPLY CHAIN ANALYTICS

A supply chain consists of a series of activities that create or add value in the form of goods and services by transforming inputs into outputs. From a business perspective, such activities include buying raw materials from suppliers (Purchase), converting raw materials into finished goods (Produce), and moving and delivering goods and services to customers (Deliver). The twin goals of supply chain management are to optimise cost efficiency and customer satisfaction. Improved cost efficiency can lead to a lower price (increases market share) and/or a better margin (improves profitability). Better customer satisfaction, through improved service levels such as quicker delivery and/or higher stock availability, improves relationships with customers, which in turn may also lead to an increase in market share.

Supply chains have an opportunity to collect a lot of data, such as point-of-sale (POS) data from sales outlets, inventory and shipping data from logistics and distribution systems, and production and quality data from factories and suppliers. These real-time, high speed, large-volume data sets, if used effectively through supply chain analytics, can provide good opportunities for companies to track material flows, diagnose supply disruptions and bottle necks, predict market trends, and optimize business processes for cost reduction and service improvement. For instance, descriptive analytics can discover problems in current operations and provide insights on the root causes; predictive analytics can provide foresights on potential problems and opportunities not yet realized; and finally, prescriptive analytics can optimize the supply chains to balance the trade-offs between cost efficiency and customer service requirement.

Supply chain management involves planning, scheduling, and control of the flow of material, information, and funds in an organization. Typical gains include more accurate forecasting, improved inventory management, and better sourcing and transportation management.

Key areas of Supply Chain Analytics:

Demand Forecasting: In simple terms, demand forecasting is the science of predicting the future demand of products and services at every level of an organization. Demand forecasting is essential in planning for sourcing, manufacturing, logistics, distribution, and sales.

Inventory Optimisation: Inventory planning and control in its simplest form involves deciding when and how much to order to balance the trade-off between inventory investment and service levels. Fill rate measures the percentage of demand satisfied within the promised time window. Inventory investment is often measured by inventory turnover, which is the ratio between annual cost of goods sold (COGS) and average inventory investment.

CHECK YOUR PROGRESS

1. The analytics on Supplier Ratings helps in managing **__________**(Options: Vendor Risk, Health risk)
2. Inventory investment is often measured by **__________**(Options: Inventory turnover, Work in Progress) State if this is True or False.
3. Big data analytics can help in reducing drug discovery cycle.
4. In retail Analytics, predictive modelling can be affected by Demand censoring.

16.8 SUMMARY

Business Analytics play a major role in the survival and growth of businesses in this digital era where Data is being generated at a high speed, and Consumers connecting with the businesses using multiple channels. Multiple domains have benefitted from this Digital Disruption in entering new area of business and enhancing their existing services or products.

Analytics plays a huge role in the success of businesses in domains like, Healthcare, Finance, Supply chain, and marketing.

16.9 KEYWORDS

Business Analytics, Healthcare Analytics, Financial Analytics, Supply chain analytics, marketing analytics, and retail analytics.

16.10 SELF ASSESSMENT QUESTIONS

1. Prepare a list of Business Analytics that you would collect for a Retail store (Like DMart) in order to improve,

    (i) Overall Customer Experience

    (ii) Ease of store movement for the customer

    (iii) Reducing customers time at the billing desk

And show what methods you would use for analysing the same.

16.11 ANSWERS TO CHECK YOUR PROGRESS

1. Vendor Risk
2. Inventory turnover
3. True.
4. True

16. 12 REFERENCES

Books

- Business Analytics: Data Analysis and Decision Making, 5th Edition, Albright and Winston
- Essentials of Business Analytics, Springer 2019, Pochiraju and Sheshadri