[redacted]

CONTENTS

SYLLABUS

COMPUTER ORIENTED STATISTICAL AND OPTIMIZATION METHODS

C-121

Unit-I:

Collection of Data. Sampling & sampling designs. Classification and tabulation of Data, Graphical representation of Data.

Unit-II:

Measure of Central values. measure of dispersal, Skew. moments and kurtosis correlation and regression.

Unit-III:

Probability & Probability and distributions (Normal, Poisson's, Binomial)

Unit-IV:

Linear Programming. Graphical Methods. Simplex methods (Simple Applications)

Unit-V :

Transportation problems. Assignments problems. Game theory

UNIT 1 INTRODUCTION TO STATISTICS

NOTES

★ STRUCTURE ★

1.0 LEARNING OBJECTIVES

After studying this unit you will be able to:

- illustrate classification of data
- describe sampling and sampling designs
- explain frequency distribution
- describe graphical representation.

1.1 INTRODUCTION

Statistics is a branch of scientific method comprising of collection, presentation, analysis and interpretation of data which are obtained by measuring some characteristics. However, the word **statistics** is used in both singular and plural forms. For example, statistics is now taught in various disciplines—this is singular sense, whereas the statistics of industrial production of India for the last five years—this is plural sense.

Numerical figures which are the effect of a large number of causes only comprise statistical data. A single train accident is not a statistical data, but the total number of train accidents during a year constitutes the statistical data. A table of values of a mathematical function *viz.*, cos x, log x etc. will never be called statistical data. Statistics deals with quantitative data only. However, methods have been devised to transfer qualitative data to quantitative.

Statistics is a wide subject and find a very suitable place in various aspects of life. Statistical tools are used in agriculture, biology, behavioural science, geology, physics, psychology, medicines, engineering etc. In business and commerce the statistical tools *viz.*, demand analysis, forecasting, inventory control, network scheduling etc. are needed for proper organisation. For manufacturing industry statistical quality control and sampling theory are two important statistical tools.

Success in Operations Research in military operation and in other phases is because of statistics. The following steps are carried out for any statistical experiment.

(a) Collection of data. The problem which has been formulated requires data for investigation which are collected by any physical methods and techniques.

(b) Tabulation. The data which we have collected can be considered as raw data and we do not get any insight of the problem unless we go for tabulation, *i.e.,* represent the data in simple tabular form by diagrams, bar charts, pie charts etc. Construct the frequency distribution.

(c) Statistical inference. Apply the statistical methods on.the tabulated data and draw conclusions about the unknown properties of the population from which the data have been drawn.

1.2 CLASSIFICATION OF DATA

Definition

Classification is defined as the process of arranging data in groups (or classes), according to some common characteristic which separate them into different, but related parts. For example, we may classify the students of a college according to their age, by using the classes 16–18, 18–20, 20–22, 22–24 years etc. Here the students having common characteristics like age between 18 and 20 would be counted in the class 18–20. Here the classes are all different but still these are related in the sense that age is the common base of classification.

Objects of Classification

Collected data is classified in order to achieve the following objectives:

1. Data is classified to condense it into some classes formed according to the magnitude of the data.
2. Data is classified to bring out the points of similarities and dissimilarities in the data. For example, the data of population census can be classified according to the attributes male, female, literatre, illiterate, rich, poor etc.
3. Data is classified to facilitate comparison. For example, performance of students of two colleges can be easily compared if they are classified according to classes of percentage of marks like 0–10, 10–20, ..., 90–100.
4. Data is classified so that statistical methods may be applied easily on the data.

Requisites of a Good Classification

1. The raw data must be classified, keeping in view the object of the investigation.

NOTES

2. The classes must be exhaustive. In other words, there must exist *some class for each and every item.*
3. The classes must be mutually exclusive. It means that there must exist exactly one class for each item. For example, the classes 10–20 and 15–25 are not mutually exclusive, because the item 18 can be entered in any of the classes.
4. The classes must be homogeneous in the sense that the units of classes must be the same.
5. The classes must be flexible. It means that the classes may be *decreased or increased as per the need of the situation.*

Variable

The value of each item in the collected data is based on certain characteristics. The characteristics like height, weight, income, expenditure, population, marks etc. are measurable in nature. Such characteristics which are measurable in nature are called **quantitative variables.** A variable which can theoretically assume any value between two given values is called a **continuous variable,** otherwise it is called a **discrete variable.** The characteristics like beauty, honesty, intelligence, colour etc. are non-measurable in nature. Such characterstics which are non-measurable are called **qualitative variables** or **attributes.**

Statistical Series

The statistical data arranged according to some logical order is called a **statistical series.** Prof. L.R. Connor has defined statistical series as, "If two variable quantities can be arranged side by side so that measurable differences in the one corresponds with the measurable differences in other, the results is said to form a statistical series".

Types of Classification

The collected data can be classified as follows:

1. Geographical Classification
2. Chronological Classification
3. Qualitative Classification
4. Quantitative Classification.

1.3 SAMPLING AND SAMPLING DESIGNS

Definition

Sampling means the selection of a part of the aggregate with a view to draw some statistical informations about the whole. This aggregate of the investigation is called population and the selected part is called sample.

Methods of Sampling

There are many methods of sampling. The choice of method will be determined by the purpose of sampling. The various methods can be grouped under two groups:

1. Random Sampling (or Probability Sampling)
    (*a*) Simple or Unrestricted Random Sampling
    (*b*) *Restricted Random Sampling*
        (*i*) Stratified Random Sampling

  (*ii*) Systematic Sampling
  (*iii*) Cluster Sampling
  (*iv*) Multi-Phase Sampling (or Double Sampling)

2. Non-Random Sampling (or Non-Probability Sampling)
  (*a*) Judgement or Purposive Sampling
  (*b*) Quota Sampling
  (*c*) Convenience Sampling

1.4 FREQUENCY DISTRIBUTION

The frequency distribution is a tabulation of data which are obtained from measurement or observation or experiment, arranged in ascending or descending order.

Let us consider the resistance of 50 units of certain electrical product:

3.0	3.4	4.1	4.1	4.3	2.7	3.5	3.7	3.4	3.4
3.8	4.2	3.1	3.9	3.1	4.1	2.8	3.7	4.4	3.5
3.5	3.4	3.7	3.7	2.8	4.3	3.8	3.4	4.1	3.0
4.4	4.1	4.1	3.6	3.4	2.7	3.6	3.0	3.4	4.3
3.8	3.2	4.2	3.9	4:2	3.4	2.9	4.4	3.5	3.9

The following table shows the simple frequency distribution of these data with all data and their frequencies of occurrences.

Resistance	*Tabulation*	*Frequency*
2.7	ǀǀ	2
2.8	ǀǀ	2
2.9	ǀ	1
3.0	ǀǀǀ	3
3.1	ǀǀ	2
3.2	ǀ	1
3.4	𝍸 ǀǀǀ	8
3.5	ǀǀǀǀ	4
3.6	ǀǀ	2
3.7	ǀǀǀǀ	4
3.8	ǀǀǀ	3
3.9	ǀǀǀ	3
4.1	𝍸 ǀ	6
4.2	ǀǀǀ	3
4.3	ǀǀǀ	3
4.4	ǀǀǀ	3

When there is a large amount of highly variable data, the above frequency distribution can become large. The data may be grouped into classes to provide a better presentation. But there is no rule about the number of classes to be taken for the given data. In the above, the lowest data is 2.7 and the highest data is 4.4. Let us consider six classes of equal width and the following table is called grouped frequency distribution.

Class	*Frequency*
2.7 — 2.9	5
3.0 — 3.2	6
3.3 — 3.5	12
3.6 — 3.8	9
3.9 — 4.1	9
4.2 — 4.4	9

In the above table, the left side value of each class, *i.e.*, 2.7, 3.0, ..., 4.2 is called lower class limit and the right side of each class, *i.e.*, 2.9, 3.2, ..., 4.4 is called upper class limit. The width of each class is 0.2.

In this example the classes are not continuous. To make it continuous we add 0.05 to the upper limits and subtract 0.05 from the lower limits where 0.05 + 0.05 = 0.1 is the difference between the previous upper limit and the next lower limit between any two consecutive classes. In this case, the class limits are called class boundaries. The middle value of any class is called class mark. So we obtain the following table:

Class boundaries	*Class mark*	*Frequency*
2.65 — 2.95	2.8	5
2.95 — 3.25	3.1	6
3.25 — 3.55	3.4	12
3.55 — 3.85	3.7	9
3.85 — 4.15	4.0	9
4.15 — 4.45	4.3	9

1.5 GRAPHICAL REPRESENTATIONS

There are mainly four graphical representation of frequency distribution, *viz,* (*a*) Histogram, (*b*) Frequency polygon, (*c*) Bar chart, (*d*) Ogive.

(*a*) Histogram. In this graph the sides of the column represent the upper and lower class boundaries and their heights are proportional to the respective frequencies. Consider the following grouped frequency distribution.

NOTES

Weight (lbs.)	*No. of persons*
100 — 110	5
110 — 120	8
120 — 130	15
130 — 140	7
140 — 150	3
150 — 160	2

The histogram is drawn as follows (Fig. 1.1)

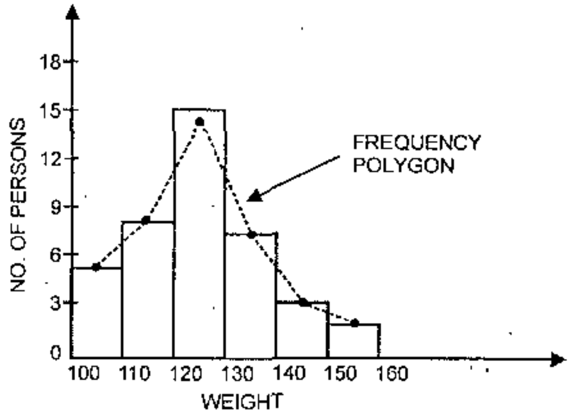


Fig. 1.1

Note. If the width of the class intervals are not same then calculate 'Relative frequency density' (*rfd*) for all classes as follows:

$$rfd = \frac{\text{Frequency}}{\text{Total frequency} \times \text{class width}}$$

Then take *rfd* on *y*-axis and class intervals on *x*-axis to draw histogram.

(b) Frequency polygon. Consider the mid-points with a height proportional to class frequency in the histogram. If these points are joined by straight lines then the resultant graph is called frequency polygon.

(c) Bar chart. A bar chart is a graphical representation of the frequency distribution in which the bars are centered at the mid-points of the cells. The heights of the bars are proportional to the respective class frequencies.

If a single attribute is presented then it is called simple bar chart (Fig. 1.2). When more than one attribute is presented then it is called multiple bar chart.

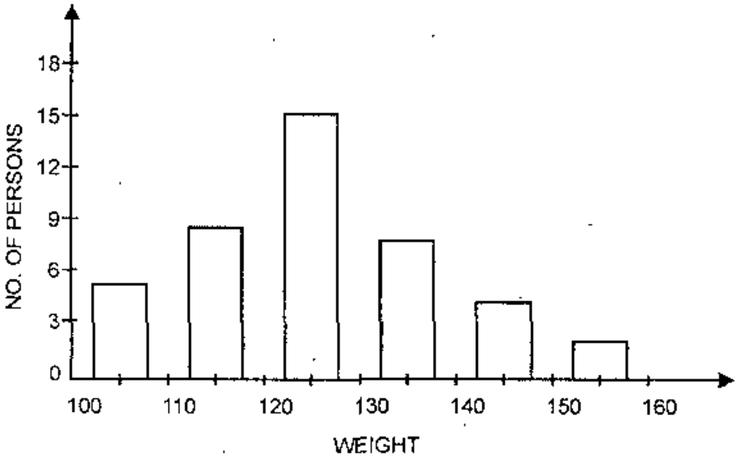


Fig. 1.2

(d) Ogive. There are two types of cumulative frequency distribution—less than type and more than type which are illustrated in the following table.

Daily Wages (in Rs.)	*No. of Workers (Frequency)*	*Cumulative frequency*	
		Less than	*More than*
22	6	6	120
27	12	18	114
32	14	32	102
37	16	48	88
42	19	67	72
47	22	89	53
52	31	120	31

Such a cumulative frequency distributions may be represented graphically and the graph is known as ogive because of its similarity to the ogee curve of the architect and the dam designer. The intersection point of the two curves give the median of the distribution.

For grouped frequency distribution, the 'less than' ogive must be plotted against the upper class boundary and not against the class mark, whereas for 'more than' ogive the cumulative frequency must be plotted against lower class boundary.

In this book if the type of the cumulative distribution is not mentioned it is to be understood that it is less than type.

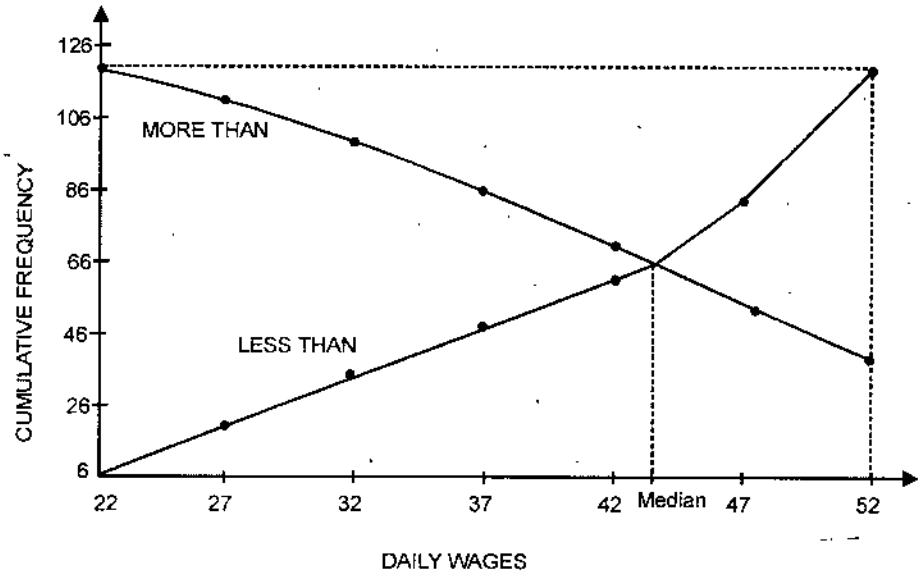


Fig. 1.3

SUMMARY

- Statistics is a branch of scientific method comprising of collection, presentation, analysis and interpretation of data which are obtained by measuring some characteristics.

- Statistics is a wide subject and find a very suitable place in various aspects of life. Statistical tools are used in agriculture, biology, behavioural science, geology, physics, psychology, medicines, engineering etc.
- The frequency distribution is a tabulation of data which are obtained from measurement or observation or experiment, arranged in ascending or descending order.

PROBLEMS

1. From the following prepare a frequency distribution table having class intervals of 5:

78	78	93	82	84	92	97	85
84	82	97	78	75	87	84	89
90	91	94	95	93	99	88	82
82	78	96	75	91	93	93	92
88	90	91	78	88	78	91	91

Also draw the histogram.

2. A machine shop produces steel pins. The width of 30 pins (in mm) was checked after machining and data was recorded as follows:

9.61	9.54	9.51	9.58	9.54	9.52
9.51	9.55	9.57	9.60	9.61	9.58
9.51	9.54	9.54	9.52	9.57	9.58
9.57	9.53	9.55	9.52	9.61	9.50
9.61	9.56	9.61	9.54	9.51	9.55

Construct the grouped frequency distribution by taking six classes.

3. For a machine making resistors the successive 40 items were checked and the following resistances in ohms were noted:

152	151	150	154	151	156	153	152
150	157	157	154	152	155	155	151
156	157	151	155	155	151	155	156
155	155	150	153	154	157	152	155
157	151	153	151	155	156	154	156

Prepare the simple frequency distribution table. Draw the ogives.

4. Exhibit the absolute and cumulative frequency in respect of the formations given below:

Height (in cm)	*No. of boys*
Less than 150	2
Less than 155	5
Less than 160	11
Less than 165	16
Less than 170	19
Less than 175	8

NOTES

5. Exhibit the absolute and cumulative frequency in respect of the formations given below:

Weight (in kg)	*No. of girls*
More than 44	2
More than 46	3
More than 48	6
More than 50	18
More than 52	12
More than 54	5
More than 56	6

6. Consider the following distribution:

Class	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	8	20	40	22	6	4

Draw the (*a*) Histogram, (*b*) frequency polygon, (*c*) ogives.

7. The hourly wage of employees (Rs. '00) in an organisation is given below:

Hourly wage	*No. of employees*
4.00 — 5.75	31
5.75 — 6.85	20
6.85 — 8.11	15
8.11 — 10.00	14
10.00 — 13.25	9

Draw the histogram.

NOTES

ANSWERS

4.

Class	*Frequency*	*Cumulative frequency (less than)*
0 — 150	2	2
150 — 155	5	7
155 — 160	11	18
160 — 165	16	34
165 — 170	19	53
170 — 175	8	61

5.

Class	*Frequency*	*Cumulative frequency (less than)*
44 — 46	2	2
46 — 48	3	5
48 — 50	6	11
50 — 52	18	29
52 — 54	12	41
54 — 56	5	46
56 —	6	52

FURTHER READINGS

1. Golden Statistics: N.P. Bali
2. Comprehensive Business Statistics: Paramanand Gupta

UNIT 2 MEASURES OF LOCATION, DISPERSION, CORRELATION AND REGRESSION

★ STRUCTURE ★

2.0 LEARNING OBJECTIVES

After studying this unit you will be able to:

- explain mean, median, mode and their types
- explain quartiles, deciles and percentiles
- define dispersion
- describe standard, mean and quartile deviation
- illustrate bivariate distribution and correlation.

NOTES

2.1 INTRODUCTION

For quantitative data it is observed that there is a tendency of the data to be distributed about a central value which is a typical value and is called a measure of central tendency. It is also called a measure of location because it gives the position of the distribution on the axis of the variable.

There are three commonly used measures of central tendency, *viz.* Mean, Median and Mode. The mean again may be of three types, *viz.* Arithmetic Mean (A.M.), Geometric Mean (G.M.) and Harmonic Mean (H.M.). Below we shall discuss these different measures.

2.2 ARITHMETIC MEAN (A.M.)

The arithmetic mean is simply called 'Average'. For the observations x_1, x_2, ..., x_n the A.M. is defined as

$$\bar{x} = \text{A.M.} = \frac{\sum_{i=1}^{n} x_i}{n}$$

For simple frequency distribution,

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\text{N}}, \text{ where N} = \Sigma f_i$$

For grouped frequency distribution, x_i is taken as class mark.

The A.M. is sometimes called as 'Average' or 'Sample Mean'.

Example 1. *Find the mean of the following data.*

No. of Matches	*2*	*3*	*4*	*6*	*7*	*8*
No. of goals	*0*	*5*	*2*	*4*	*6*	*10*

Solution. Here $\text{N} = \text{Total no. of matches} = 30$

Also $\Sigma xf = 0 + 15 + 8 + 24 + 42 + 80 = 169$

$$\text{Hence mean} = \frac{169}{30} = 5.633.$$

Note. Let $\bar{x}_1$ be the mean of n_1 observations and $\bar{x}_2$ be the mean of n_2 observations then the combined mean $\bar{x}$ is computed as follows:

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

2.3 GEOMETRIC MEAN (G.M.)

The geometric mean of the observations x_1, x_2, ...,x_n is defined as

$$\text{G.M.} = (x_1 . x_2 \ldots\ldots\ldots\ldots x_n)^{1/n}$$

For simple frequency distribution,

$$\text{G.M.} = \left(x_1^{f_1} . x_2^{f_2} \ldots\ldots\ldots\ldots\ldots x_n^{f_n}\right)^{1/\text{N}}, \quad \text{N} = \sum_{i=1}^{n} f_i$$

For grouped frequency distribution, x_i is taken as class mark.

Note. 1. The logarithm of the G.M. of a variate is the A.M. of its logarithm.

2. G.M. = 0 iff a single variate value is zero.

3. G.M. is not used if any variate value is negative.

Example 2. *Find the G.M. of the following distribution :*

Humidity reading	*No. of days*
60	3
62	2
64	4
68	2
70	4

Solution. Here N = No. of days = **15**.

x	f	log x	f log x
60	3	1.77815	5.33445
62	2	1.79239	3.58478
64	4	1.80618	7.22472
68	2	1.83251	3.66502
70	4	1.84510	7.38040
Σ	15	–	27.18937

Then, $$\log \text{G.M.} = \frac{1}{N}\Sigma f_i \log x_i = \frac{1}{15}(27.18937) = 1.81262$$

$$\Rightarrow \quad \text{G.M.} = 64.9561 \approx 64.96$$

2.4 HARMONIC MEAN (H.M.)

The reciprocal of the H.M. of a variate is the A.M. of its reciprocal.

For the observations $x_1, x_2, \ldots\ldots\ldots, x_n$

$$\text{H.M.} = \frac{n}{\Sigma(1/x_i)}$$

For simple frequency distribution,

$$\text{H.M.} = \frac{N}{\Sigma(f_i/x_i)}, \quad N = \Sigma f_i$$

For grouped frequency distribution x_i is taken as class mark.

Note. A.M. ≥ G.M. ≥ H.M.

Example 3. *Suppose a train moves 100 km with a speed of 40 km/hr, then 150 km with a speed of 50 km/hr and next 135 km with a speed of 45 km/hr. Calculate the average speed.*

Solution. To get average speed we require harmonic mean of 40,50 and 45 with 100, 150 and 135 as the respective frequency or weights.

$$\text{H.M.} = \frac{100 + 150 + 135}{100 \times \frac{1}{40} + 150 \times \frac{1}{50} + 135 \times \frac{1}{45}}$$

$$= \frac{385}{8.5} = 45.29$$

Hence the average speed per hour is 45.29 km.

Note. In the case of grouped frequency distributions with open end class at one extremity or at both the extremities, the A.M., G.M. and H.M. cannot be computed unless we make some plausible assumptions.

2.5 MEDIAN

(*a*) For the observations $x_1, x_2, \ldots\ldots\ldots x_n$ the median is the middle value if the number of observations is odd and have been arranged in ascending or descending order of magnitude. For even number of observations the median is taken as the average of two middle values after they are arranged in ascending or descending order of magnitude.

e.g., For the data, 10, 17, 15, 25, 18, let us arrange them in ascending order as 10, 15. 17, 18 and 25. Here the middle value is 17. Hence the median is 17.

Consider another sets of data, 21, 40, 19, 28, 33 30. Let us arrange them in ascending order as 19, 21, 28, 30, 33 and 40. There are two middle values 28 and 30. So the median is (28 + 30)/2 *i.e.*, 29.

(*b*) For simple frequency distribution the median is detained by using less than cumulative frequency distribution. Here the median is that value of the variable for which the cumulative frequency is just greater than $\frac{1}{2}\cdot\text{N}$ where N = total frequency.

e.g., consider

x	f	*Cumulative frequency*
10	2	2
15	5	7
20	11	18
25	7	25
30	5	30

Here N = 30, $\frac{\text{N}}{2} = 15$. So the cumulative frequency just greater than 15 is 18 and the corresponding variable value is 20. Then the median is 20.

(*c*) For grouped frequency distribution, the median is obtained by the following:

$$\text{Median} = \text{L} + \frac{h\left(\frac{\text{N}}{2} - \text{C}\right)}{f}$$

where,
L = Lower limit or boundary of the median class.
h = Width of the class interval
f = Frequency of the median class.
N = Total frequency
C = Cumulative frequency of the class preceding the median class.

Example 4. *Find the median of the following data :*

Marks	Less than 40	41–50	51–60	61–70	71–80	81 and above
No. of students	10	20	15	25	10	20

Solution.

Marks	*No. of students* (*f*)	*Cumulative frequency*
Less than 40	10	10
41 — 50	20	30
51 — 60	15	45
61 — 70	25	70
71 — 80	10	80
81 and above	20	100

Here $N = 100$, $\frac{N}{2} = 50$, $C = 45$, the median class is 61 – 70.

$L = 61$, $h = 9$, $f = 25$

$$\therefore \quad \text{Median} = 61 + \frac{9(50-45)}{25} = 62.8.$$

Note. In case of unequal class-intervals median is sometimes preferred to A.M.

2.6 MODE

Mode is the value of a variable which occurs most frequently in a set of observations.

For simple frequency distribution the mode is the value of a variable corresponding to the maximum frequency. For grouped frequency distribution the mode is obtained as follows.

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} . h$$

where,
L = Lower limit or boundary of the modal class
h = Width of the modal class
f_1 = Frequency of the modal class
f_0 = Frequency of the class preceeding the modal class
f_2 = Frequency of the class succeeding the modal class.

Note. 1. If $2f_1 - f_0 - f_2 = 0$, then the mode is obtained as follows :

$$\text{Mode} = L + \frac{f_1 - f_0}{|f_1 - f_0| + |f_1 - f_2|} \times h.$$

2. If the maximum frequency is repeated then the above technique is not practicable.
3. For symmetrical distribution,

$$\text{Mean} = \text{Median} = \text{Mode}. \qquad (\text{Mean} = \bar{x}).$$

However, for moderately skewed distribution there is an empirical relationship due to Karl Pearson *i.e.*,

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median}).$$

Example 5 : *Find the mode of the problem as given in example 4.*

Solution. Here maximum frequency is 25, so the modal class is 61–70.

$$L = 61, h = 9, f_1 = 25, f_0 = 15, f_2 = 10$$

$$\therefore \qquad \text{Mode} = 61 + \frac{25 - 15}{2 \times 25 - 15 - 10} \times 9 = 64.6.$$

2.7 QUARTILES, DECILES AND PERCENTILES

As the median divides an array into two parts, the quartiles divide the array into four parts, the deciles divide it into ten parts and the percentiles divide it into one hundred parts.

The first quartile/the lower quartile denoted by Q_1 is computed as follows.

$$Q_1 = L + \frac{\frac{N}{4} - C}{f} \times h$$

where, L = Lower limit of the class containing Q_1

f = Frequency of the class containing the Q_1

h = Width of the class containing the Q_1

C = Cumulative frequency of the class preceding the calss containing Q_1.

Here the cumulative frequency just greater than $\frac{N}{4}$ is the class containing Q_1.

The second quartile is the median.

The third quartile/the upper quartile denoted by Q_3 is computed as follows:

$$Q_3 = L + \frac{\frac{3N}{4} - C}{f} \times h$$

where, L = Lower limit of the class containing Q_3

f = Frequency of the class containing Q_3

h = Width of the class containing Q_3

C = Cumulative frequency of the class preceding the class containing Q_3.

Here the cumulative frequency *just greater than* $\frac{3N}{4}$ *is the class containing* Q_3.

The k-th decile denoted by D_k is computed as follows:

C = Cumulative frequency

$$D_k = L + \frac{\frac{k \times N}{10} - C}{f} \times h, \qquad (k = 1, 2, \ldots\ldots\ldots, 9)$$

where,
L = Lower limit of the class containing D_k
f = Frequency of the class containing D_k
h = Width of the class containing D_k
C = Cumulative frequency of the class preceding the class containing D_k,

Here the cumulative frequency just greater than $\frac{k \times N}{10}$ is the class containing D_k (k = 1, 2, ...,9).

The k-th percentile denoted by P_k is computed as follows:

$$P_k = L + \frac{\frac{k \times N}{100} - C}{f} \times h, \qquad (k = 1, 2, \ldots\ldots\ldots\ldots, 99)$$

where,
L = Lower limit of the class containing P_k
f = Frequency of the class containing P_k
h = Width of the class containing P_k
C = Cumulative frequency of the class preceding the class containing P_k.

Here the cumulative frequency just greater than $\frac{k \times N}{100}$ is the class containing P_k (k = 1,2, ...,99).

Example 6. *Determine (a)* Q_1, *(b)* Q_3, *(c)* D_5 *(d)* P_{80} *from the following distribution:*

Class	10–15	15–20	20–25	25–30	30–35	35–40
Frequency	20	15	31	22	10	2

Solution.

Class	*Frequency*	*Cumulative frequency*
10–15	20	20
15–20	15	35
20–25	31	66
25–30	22	88
30–35	10	98
35–40	2	100

(*a*) $N = 100, \frac{N}{4} = 25$. The cumulative frequency just greater than 25 is 35. So the class 15–20 contains Q_1. $L = 15, f = 15, h = 5, C = 20$. Therefore,

$$Q_1 = 15 + \frac{(25-20)}{15} \times 5 = 16.67$$

(*b*) Here $\frac{3N}{4} = 75$. The cumulative frequency just greater than 75 is 88. So the class 25 – 30 contains Q_3.

$L = 25, f = 22, h = 5, C = 66$. Therefore,

$$Q_3 = 25 + \frac{(75-66)}{22} \times 5 = 27.05$$

(*c*) Here $\frac{5N}{10} = 50$. The cumulative frequency just greater than 50 is 66. So the class 20 – 25 contains D_5.

$L = 20, f = 31, h = 5, C = 35$. Therefore,

$$D_5 = 20 + \frac{(50-35)}{31} \times 5 = 22.42$$

(*d*) Here $\frac{80N}{100} = 80$. The cumulative frequency just greater than 80 is 88. So the class 25 – 30 contains P_{80}.

$L = 25, f = 22, h = 5, C = 66$. Therefore,

$$P_{80} = 25 + \frac{(80-66)}{22} \times 5 = 28.18$$

Example 7. *A given machine is assumed to depreciate 30% in value in the first year, 35% in the second year and 80% per annum for the next three years, each percentage being calculated on the diminishing value. Calculate the average annual rate of depreciation.*

Solution. The proportional rates of depreciation for the 5 years are 0.30, 0.35, 0.80, 0.80 and 0.80 respectively.

Let r be the average proportional rate of depreciation.

Then $1 - r$ is the G.M. of (1 – 0.30), (1 – 0.35), (1 – 0.80), (1 – 0.80) and (1 – 0.80) *i.e.*, 0.70, 0.65, 0.20, 0.20 and 0.20.

$$\therefore \quad 1 - r = (0.70 \times 0.65 \times 0.20 \times 0.20 \times 0.20)^{1/5}$$

$$\Rightarrow \quad \log(l - r) = \frac{1}{5}\log(0.00364)$$

$$\Rightarrow \quad 1 - r = 0.3253$$

$$\Rightarrow \quad r = 0.6747$$

Hence the average annual rate of depreciation is 67.47%.

2.8 DEFINITION OF DISPERSION

After getting the idea of central value of the quantitative data as discussed in the previous chapter, it is observed that in some cases the values are very close around the central value and in other cases the values are scattered a little wide around the central value. The measure which gives the idea of the amount of scattering of the data around the central value is called the measure of *dispersion*.

There are four commonly used measures of dispersion *viz.* Standard Deviation (S.D.), Mean Deviation (M.D.), Quartile Deviation (Q.D.) and Range (R). Below we shall discuss these different measures.

2.9 STANDARD DEVIATION (S.D.)

Standard deviation (and variance) is a relative measure of the dispersion of a set of data–the larger the standard deviation, the more spread out the data.

If x_1, x_2, ..., x_n be a set of n observations forming a population, then its standard deviation is given by

$$\text{S.D.} = \sigma = \left[\frac{1}{n}.\ \Sigma(x_i - \bar{x})^2\right]^{1/2}, \quad \text{where } \bar{x} = \frac{1}{n}\Sigma x_i$$

$$= \left[\frac{1}{n}\Sigma x^2 - (\bar{x})^2\right]^{1/2}$$

For simple frequency distribution,

$$\text{S.D.} = \sigma = \left[\frac{1}{N}\Sigma f_i\,(x_i - \bar{x})^2\right]^{1/2}, \text{ where } N = \Sigma f_i$$

For grouped frequency distribution, x_i is taken as class mark.

Note. 1. The square of S.D. is known as variance.

2. When $\bar{x}$ is a real number, then the following alternative formula can be used to calculate S.D.

$$\text{S.D.} = \left[\frac{1}{N}\Sigma f x_i^2 - (\bar{x})^2\right]^{1/2}, \quad N = \Sigma f_i$$

3. For comparing the variability of two distributions the coefficient of variation (C.V.) is calculated as follows :

$$\text{C.V.} = \frac{\text{S.D.}}{\bar{x}} \times 100$$

The distribution with less C.V. is said to be more uniform or consistent or less variable or more homogeneous.

4. If the values of x and f are large then for simplicity step-deviation method can be employed in which the deviations of the given values of x from any arbitrary point A is taken.

Then $$\sigma^2 = \frac{1}{N}\Sigma fd^2 - \left(\frac{1}{N}\Sigma fd\right)^2 \text{ where, } d = x - A$$

which shows that the variance and S.D. of a distribution is independent of change of origin.

5. In case of grouped frequency distribution, if h be the width of the class-interval then we can use change of scale also *i.e.*, by taking $u = \frac{x - A}{h}$, we obtain
$$\sigma^2 = \left[\frac{1}{N}\Sigma fu^2 - \left(\frac{1}{N}\Sigma fu\right)^2\right] \times h^2$$
6. **(Combined Variance).** If $\bar{x}_1$ and σ_1 be the mean and S.D. of a group of observations of size n_1 whereas $\bar{x}_2$ and σ_2 be the mean and S.D. of another group of observations of size n_2, then the variance of the combined groups is given by :
$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1d_1^2 + n_2d_2^2}{n_1 + n_2}$$
where $d_1 = \bar{x}_1 - \bar{x}$, $d_2 = \bar{x}_2 - \bar{x}$ and $\bar{x}$ = Mean of the combined groups.
7. S.D. is independent of origin but not of scale.

Example 8. *A group of 100 items have a mean of 60 and a S.D. of 7. If the mean and S.D. of 60 of these items be 51 and 5.2 respectively, find the S.D. of the other 40 items.*

Solution. Consider the following:

		$\bar{x}$	S.D.
n	100	60	12
n_1	60	51	5.2
n_2	40	m	s

From the combined mean, we obtain

$$60 = \frac{60 \times 51 + 40 \times m}{100}$$

$$\Rightarrow \quad 6000 - 3060 = 40m$$

$$\Rightarrow \quad m = 294/4 = 73.5$$

From the combined variance we obtain,

$$144 = \frac{40 \times s^2 + 60 \times (5.2)^2 + 40 \times (73.5 - 60)^2 + 60 \times (60 - 51)^2}{100}$$

$$\Rightarrow \quad 14400 = 40s^2 + 13772.4$$

$$\Rightarrow \quad s^2 = 627.6/40 = 15.69$$

$$\Rightarrow \quad s = 3.961.$$

2.10 MEAN DEVIATION (M.D.)

If $x_1, x_2, \ldots\ldots, x_n$ be the n observations then M.D. is defined as

$$\text{M.D.} = \frac{1}{n}\Sigma|x - A|$$

where A may be mean $(\bar{x})$, or median or mode.

For simple frequency distribution, M.D. is defined as

$$\text{M.D.} = \frac{1}{N}\Sigma f\,|x - A|, \quad N = \Sigma f$$

where A may be mean $(\bar{x})$ or median or mode.

For grouped frequency distribution, x will be taken as class-mark.

Sometimes M.D. about mean $(\bar{x})$ is called simply M.D.

Note. 1. M.D. taken about median is the least, compared to M.D. taken about mean $(\bar{x})$ or mode.

2. It is not a very accurate measure of dispersion particularly when it is calculated from mode.

3. $\text{S.D.} \geq \text{M.D.}$

2.11 QUARTILE DEVIATION (Q.D.)

It is a location based measure of dispersion and is defined as follows:

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

where Q_1 is lower quartile and Q_3 is upper quartile. This Q.D. is also known as semi inter-quartile range, whereas the quantity $Q_3 - Q_1$ is called inter-quartile range.

Note. 1. For comparative studies of variability of two distributions a relative measure is given as follows :

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

2. If there is two much variability between two distributions then the following measure is calculated :

$$\text{Coefficient of variation} = \frac{\text{Q.D.}}{\text{Median}} \times 100.$$

3. In the case of unequal class-intervals Q.D. is sometimes preferred to S.D.

2.12 RANGE (R)

It is the simplest measure for dispersion. For the observations $x_1, x_2, \ldots\ldots, x_n$ it is defined as

$$R = (\text{largest value}) - (\text{smallest value})$$

For frequency distribution also, the same definition applies and frequencies do not taken into account.

Note. Of all the measures of dispersion, the S.D. is generally the one with the least sampling fluctuation. However, when the data contain a few extreme values widely different from the majority of the values, S.D. should not be used – Q.D. is the appropriate measure.

Example 9. *Find the mean deviation about the mean and variance for the following :*

NOTES

x	*Frequency*
47.5	7
48.1	17
45.9	46
44.0	44
40.7	54

Solution. The calculations are shown in the following table :

x	f	xf	$f\lvert x-\bar{x}\rvert$	$f(x-\bar{x})^2$
47.5	7	332.5	24.36	84.77
48.1	17	817.7	69.36	282.99
45.9	46	2111.4	86.48	162.58
44.0	44	1936.0	0.88	0.02
40.7	54	2197.8	179.28	595.21
Σ	168	7395.4	360.36	1125.57

$$\text{Mean} = \frac{\Sigma xf}{\Sigma f} = \frac{7395.4}{168} = 44.02$$

$$\text{Mean deviation about mean} = \frac{\Sigma f|x-\bar{x}|}{\Sigma f}$$

$$= \frac{360.36}{168}$$

$$= 2.145$$

$$\text{Variance} = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f}$$

$$= \frac{1125.57}{168} = 6.6998$$

Example 10. *Calculate S.D., Q.D., and M.D. of the following data.*

Data class	0 – 4	5 – 9	10 – 14	15 – 19	20 – 24	25 – 29	30 – 34
Frequency	4	5	10	8	7	9	7

Solution.

Class	*Class boundaries*	*Class mark (x)*	*Frequency (f)*	*fx*	fx^2	$f\lvert x-\bar{x}\rvert$	*Cumulative frequency*
0 – 4	(–0.5) – 4.5	2	4	8	16	65.6	4
5 – 9	4.5 – 9.5	7	5	35	245	57.0	9
10 – 14	9.5 – 14.5	12	10	120	1440	64.0	19

15 – 19	14.5 – 19.5	17	8	136	2312	11.2	27
20 – 24	19.5 – 24.5	22	7	154	3388	25.2	34
25 – 29	24.5 – 29.5	27	9	243	6561	77.4	43
30 – 34	29.5 – 34.5	32	7	224	7168	95.2	50
		Σ	50	920	21130	395.6	

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{920}{50} = 18.4$$

$$\text{S.D} = \left[\frac{1}{N}\Sigma fx^2 - (\bar{x})^2\right]^{1/2}$$

$$= \left[\frac{1}{50}(21130) - (18.4)^2\right]^{1/2}$$

$$= [84.04]^{1/2} = 9.167$$

Here $\frac{N}{4} = 12.5$, then the cumulative frequency just greater than 12.5 is 19, so 9.5 – 14.5 is the class containing Q_1. L = 9.5, h = 4, f = 10, C = 9

$$\therefore \qquad Q_1 = 9.5 + \frac{(12.5 - 9)}{10} \times 4 = 10.9$$

Again $\frac{3N}{4} = 37.5$, then the cumulative frequency just greater than 37.5 is 43, so 24.5 – 29.5 is the class containing Q_3. L = 24.5, h = 4, f = 9, C = 34

$$\therefore \qquad Q_3 = 24.5 + \frac{(37.5 - 34)}{9} \times 4 = 26.06$$

Then

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{26.06 - 10.9}{2} = 7.58$$

$$\text{M.D.} = \frac{1}{N}\Sigma f|x - \bar{x}| = \frac{395.6}{50} = 7.912.$$

Example 11. *A factory produces two types of electric bulbs A and B. In an experiment relating to their life, the following results were obtained.*

Length of life (in hours)	*No. of bulbs*	
	A	*B*
500 – 700	5	4
700 – 900	11	30
900 – 1100	26	12
1100 – 1300	10	8
1300 – 1500	8	6

Find which type of bulb is less variable in length of life.

NOTES

Solution.

Class	Class marks(m)	$u = \frac{m-1000}{100}$	Bulb A			Bulb B		
			f_1	f_1u	f_1u^2	f_2	f_2u	f_2u^2
500–700	600	–4	5	–20	80	4	–16	64
700–900	800	–2	11	–22	44	30	–60	120
900–1100	1000	0	26	0	0	12	0	0
1100–1300	1200	2	10	20	40	8	16	32
1300–1500	1400	4	8	32	128	6	24	96
		Total	60	10	292	60	–36	312

Bulb A

$$\bar{x}_A = A + \frac{h\,\Sigma f_1 u}{N_1} = 1000 + \frac{(100)\,(10)}{60} = 1016.67$$

$$\sigma_A = h\left[\frac{\Sigma f_1 u^2}{N_1} - \left(\frac{\Sigma f_1 u}{N_1}\right)^2\right]^{1/2}$$

$$= 100\left[\frac{292}{60} - \left(\frac{10}{60}\right)^2\right]^{1/2} = 219.97$$

$$\text{C.V.(A)} = \frac{\sigma_A}{\bar{x}_A} \times 100 = \frac{219.97}{1016.67} \times 100 = 21.64$$

Bulb B

$$\bar{x}_B = A + \frac{h\,\Sigma f_2 u}{N_2} = 1000 + \frac{(100)\,(-36)}{60} = 940$$

$$\sigma_B = h\left[\frac{\Sigma f_2 u^2}{N_2} - \left(\frac{\Sigma f_2 u}{N_2}\right)^2\right]^{1/2}$$

$$= 100\left[\frac{312}{60} - \left(-\frac{36}{60}\right)^2\right]^{1/2} = 220$$

$$\text{C.V.(B)} = \frac{\sigma_B}{\bar{x}_B} \times 100 = \frac{220}{940} \times 100 = 23.40$$

Since C.V. (A) < C.V. (B)

⇒ Bulb A is less variable in length of life.

2.13 MOMENTS

For n observations $x_1, x_2, \ldots\ldots\ldots, x_n$ and an arbitrary constant A, the rth moment about A is defined as,

$$\mu_r' = \frac{1}{n}\sum_i (x_i - A)^r, \quad r = 0,\ 1,\ 2......$$

For simple frequency distribution, the r th moment about A is defined as

$$\mu_r' = \frac{1}{\mathrm{N}}\sum_i f_i(x_i - \mathrm{A})^r,\ r = 0, 1, 2, ..., \mathrm{N} = \Sigma f_i$$

For grouped frequency distribution, x_i will be taken as class mark.

These moments, μ_r', are also known as 'raw moments'.

When $\mathrm{A} = \bar{x}$, then these moments are called 'central moments' and we denote as follows:

$$\mu_r = \frac{1}{\mathrm{N}}\sum_i f_i\left(x_i - \bar{x}\right)^r, \quad r = 0, 1, 2, ..., \mathrm{N} = \Sigma f_i$$

For $r = 0$, $\quad \mu_0 = \frac{1}{\mathrm{N}}\Sigma f = 1$

For $r = 1$, $\quad \mu_1 = \frac{1}{\mathrm{N}}\Sigma f(x - \bar{x})$

$$= \frac{1}{\mathrm{N}}\Sigma fx - \frac{1}{\mathrm{N}}\Sigma f\bar{x}$$

$$= \bar{x} - \bar{x}\frac{1}{\mathrm{N}}\Sigma f$$

$$= \bar{x} - \bar{x} = 0$$

For $r = 2$, $\mu_2 = \frac{1}{N}\Sigma f\,(x - \bar{x})^2$ which is called variance.

The third and fourth central moments *i.e.*, μ_3 and μ_4 are used to measure skewness and kurtosis which have been given in the following sections.

The important relations between the central and raw moments are as follows:

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$\mu_3 = \mu'_3 - 3\mu'_2\,\mu'_1 + 2(\mu'_1)^3$$

$$\mu_4 = \mu'_4 - 4\mu'_3\,\mu'_1 + 6\,\mu'_2(\mu'_1)^2 - (\mu'_1)^4$$

$$\bar{x} = \mu'_1 + \mathrm{A}.$$

Note. If the above moments are called as sample moments, then we have to use the notation m'_r instead of μ'_r. See the method of moments in Chapter 10A.

2.14 SKEWNESS

Skewness is a measure of symmetry of the shape of frequency distribution, *i.e.*, it reveals the dispersal of value on either side of an average is symmetrical or not.

There are four mathematical measures of relative skewness.

(*a*) **Karl Pearson's Coefficient of Skewness**

$$\mathrm{SK} = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

If mode is ill-defined, then we take

$$SK = \frac{3(\text{Mean} - \text{Median})}{\text{S.D.}}$$

(*b*) Bowley's Coefficient of Skewness

This is based on quartiles and median and is defined as

$$SK = \frac{Q_3 + Q_1 - 2\ \text{Median}}{Q_3 - Q_1}$$

This formula is useful when the mode is ill-defined or the distribution has open end classes or unequal class-intervals.

(*c*) Coefficient of skewness based on central moments

Using second and third central moments, the coefficient of skewness is defined as (due to Karl Pearson)

$$\beta_1 = SK = \frac{\mu_3^2}{\mu_2^3}$$

Also $\gamma_1 = \sqrt{\beta_1}$ which is due to R.A. Fishar.

Note. 1. If SK is positive then the frequency distribution is called positively skewed. If SK is negative then the frequency distribution is called negatively skewed. If SK is zero, then the frequency distribution is symmetric.

2. There is no theoretical limit to this measure.

2.15 KURTOSIS

This is a measure of peakness of a distribution and is defined in terms of second and fourth central moments as follows (due to Karl Pearson) :

$$\beta_2 = \frac{\mu_4}{\mu_2^2},$$

When $\beta_2 = 3$, the distribution is called **Mesokurtic**

$\beta_2 < 3$, the distribution is called **Platykurtic**

and $\beta_2 > 3$, the distribution is called **Leptokurtic.**

Another notation due to R.A. Fishar is $\gamma_2 = \beta_2 - 3$ which is also called excess of kurtosis.

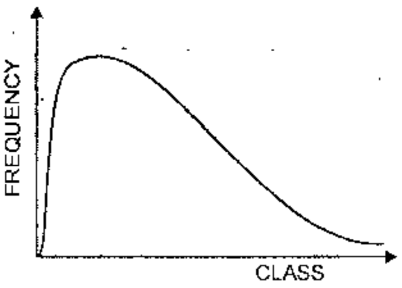


Fig. 2.1 Positively skewed

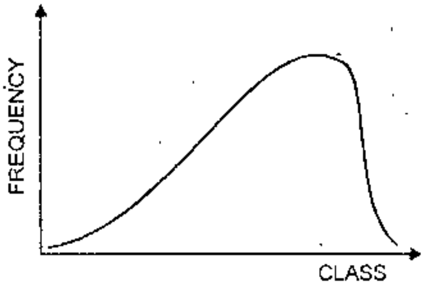


Fig. 2.2 Negatively skewed

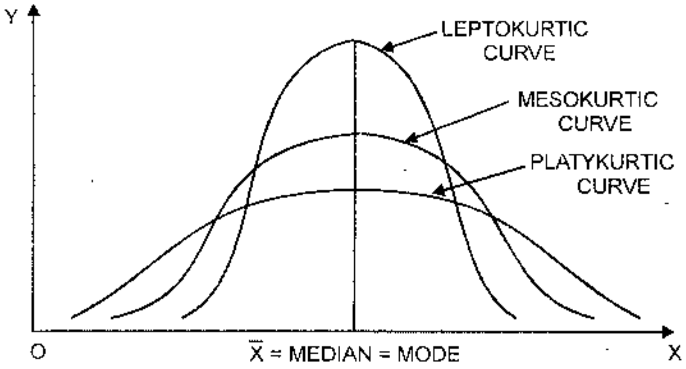


Fig. 2.3 Nature of kurtosis

Example 12. *Calculate the Karl Pearson's coefficient of skewness of the following frequency distribution.*

Class	383–387	388–392	393–397	398–402	403–407
Frequency	8	10	15	17	8

Solution.

Class mark (x)	*Frequency* (f)	$d = x - 395$	fd	fd^2
385	8	– 10	– 80	800
390	10	– 5	– 50	250
395	15	0	0	0
400	17	5	85	425
405	8	10	80	800
Σ	58	0	35	2275

$$\therefore \quad \bar{x} = 395 + \frac{\Sigma fd}{\Sigma f} = 395 + \frac{35}{58} = 395.603$$

$$\text{S.D.} = \left[\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2\right]^{1/2}$$

$$= \left[\frac{2275}{58} - \left(\frac{35}{58}\right)^2\right]^{1/2}$$

$$= [38.864]^{1/2} = 6.234$$

To calculate mode we require class boundaries. Since the maximum frequency is 17, this implies that 397.5 – 402.5 is the modal class.

$$\therefore \quad \text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

$$= 397.5 + \frac{(17-15)}{34-15-8} \times 4$$
$$= 398.227$$

NOTES

$$\therefore \quad \text{Coefficient of skewness} = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$
$$= \frac{395.603 - 398.227}{6.234}$$
$$= -0.421$$

which indicate the given distribution is negatively skewed.

Example 13. *The first four moments of a distribution about the value 3 of a variable are 1, 5, 14 and 46.*

Find the mean, variance and comment on the nature of the distribution.

Solution. Given A = 3, $\mu'_1 = 1$, $\mu'_2 = 5$, $\mu'_3 = 14$, $\mu'_4 = 46$.

$$\bar{x} = \mu'_1 + A = 1 + 3 = 4$$
$$\mu_2 = \text{Variance} = \mu'_2 - (\mu'_1)^2 = 5 - 1 = 4$$
$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3$$
$$= 14 - 3(5)(1) + 2(1)^3 = 1$$
$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu'_1)^4$$
$$= 46 - 4(14) + 6(5) - 3 = 17$$

Now $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{1}{64}, \quad \gamma_1 = \sqrt{\beta_1} = \frac{1}{8}$$

Since $\beta_1 > 0 \Rightarrow$ The distribution is positively skewed.

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{17}{16} = 1.0625, \qquad \gamma_2 = \beta_2 - 3 = -1.9375$$

Since $\beta_2 < 3 \Rightarrow$ The distribution is platykurtic

Example 14. *For a distribution, Bowley's coefficient of skewness is – 0.65, Q_1 = 15.28 and median = 25.2. What is its coefficient of quartile deviation?*

Solution. Bowley's coefficient of skewness is

$$SK = \frac{Q_3 + Q_1 - 2\ \text{Median}}{Q_3 - Q_1}$$

$$\Rightarrow \quad -0.65 = \frac{Q_3 + 15.28 - 2\,(25.2)}{Q_3 - 15.28}$$

$$\Rightarrow \quad Q_3 = 27.30$$

Then, coefficient of quartile deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{27.30 - 15.28}{27.30 + 15.28} = 0.28.$$

NOTES

2.16 BIVARIATE DISTRIBUTION

When a distribution has two variables then it is called bivariate. For example, if we measure the income and expenditure of a certain group of persons-one variable will measure income and the other variable will measure expenditure and the values will form the bivariate distribution.

There may be any correlation between the variables, *i.e.*, the change in one variable gives a specific change in the other variable. For example, if the increase (or decrease) of one variable results the increase (or decrease) of the other variable, then the correlation is said to be positive. If the increase (or decrease) leads to decrease (or increase) then the correlation is said to be negative.

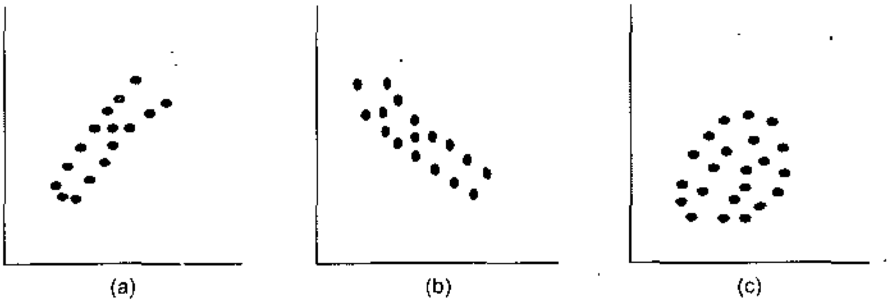


Fig. 2.4 Scatter diagram

The simplest way to represent the bivariate data in a diagram known as scatter diagram. For the bivariate distribution (x, y), the values (x_i, y_i), $i = 1, 2, ..., n$ of the variables are plotted in the xy-plane which is known as scatter diagram. This gives an idea about the correlation of the two variables.

2.17 COEFFICIENT OF CORRELATION

Karl Pearson has given a coefficient to measure the degree of linear relationship between two variables which is known as coefficient of correlation (or, correlation coefficient).

For a bivariate distribution (x, y) the coefficient of correlation denoted by r_{xy} and is defined as

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x . \sigma_y}$$

where $\quad cov\ (x, y)$ = Covariance between x and y

σ_x = S.D. of x

σ_y = S.D. of y

For the values (x_i, y_i), $i = 1, 2,......,n$ of a bivariate distribution,

$$r_{xy} = \frac{\frac{1}{n}\Sigma(x_i - \bar{x})\ (y_i - \bar{y})}{\sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2} . \sqrt{\frac{1}{n}\Sigma\ (y_i - \bar{y})^2}}$$

$$= \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2}\ \sqrt{\Sigma(y_i-\bar{y})^2}}$$

Limitations of r_{xy}

1. The coefficient of correlation can be used as a measure of linear relationship between two variables. In case of non-linear or any other relationship the coefficient of correlation does not provide any measure at all. So the inspection of scatter diagram is essential.
2. Correlation must be used to the data drawn from the same source. If distinct sources are used then the two variables may show correlation but in each source they may be uncorrelated.
3. For two variables with a positive or negative correlation it does not necessarily mean that there exists causal relationship. There may be the effect of some other variables in both of them. On elimination of this effect it may be found that the net correlation is nil.

Properties

1. The coefficient of correlation is independent of the origin and scale of reference.
2. $-1 \le r_{xy} \le 1$

Proof.

Let $u_i = \dfrac{x_i-\bar{x}}{\sigma_x}$ and $v_i = \dfrac{y_i-\bar{y}}{\sigma_y}$

Then $\dfrac{1}{n}\Sigma u_i^2 = 1,\ \dfrac{1}{n}\Sigma v_i^2 = 1,\ \dfrac{1}{n}\Sigma u_i\, v_i = r_{xy}$

Now $\dfrac{1}{n}\Sigma(u_i-v_i)^2 \ge 0$

$\Rightarrow \dfrac{1}{n}\Sigma u_i^2 + \dfrac{1}{n}\Sigma v_i^2 - \dfrac{2}{n}\Sigma u_i\, v_i \ge 0$

$\Rightarrow 2(1-r_{xy}) \ge 0$

$\Rightarrow r_{xy} \le 1$

Also, $\dfrac{1}{n}\Sigma(u_i+v_i)^2 \ge 0$

$\Rightarrow \dfrac{1}{n}\Sigma u_i^2 + \dfrac{1}{n}\Sigma v_i^2 + \dfrac{2}{n}.\Sigma u_i\, v_i \ge 0$

$\Rightarrow 2(1+r_{xy}) \ge 0$

$\Rightarrow r_{xy} \ge -1$

Combining we obtain $-1 \le r_{xy} \le 1$.

3. Two independent variables are uncorrelated (*i.e.*, $r_{xy} = 0$) but the converse is not always true.

4. If r_{xy} is the correlation coefficient in a sample of n pairs of observations, then the standard error of r_{xy} is defined by

$$\text{S.E.}(r_{xy}) = \frac{1-r_{xy}^2}{\sqrt{n}}$$

5. **Probable error** of the correlation coefficient is defined by

$$\text{P.E.}(r_{xy}) = 0.6745\,\frac{(1-r_{xy}^2)}{\sqrt{n}}$$

By Step Deviation Method

Let $d_x = x - A$, $d_y = y - B$ which are the deviations and A, B are assumed values, then

$$r_{xy} = \frac{\Sigma d_x\, d_y - \frac{\Sigma d_x . \Sigma d_y}{n}}{\sqrt{\left[\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{n}\right]} . \sqrt{\left[\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{n}\right]}}$$

where n = number of observations.

For grouped data,

$$r_{xy} = \frac{\Sigma f . d_x d_y - \frac{\Sigma f d_x . \Sigma f d_y}{N}}{\sqrt{\Sigma f d_x^2 - \frac{(\Sigma f d_x)^2}{N}} . \sqrt{\Sigma f d_y^2 - \frac{(\Sigma f d_y)^2}{N}}}$$

where $N = \Sigma f$

If $X = x - \bar{x}$ and $Y = y - \bar{y}$, then a short-cut formula is

$$r_{xy} = \frac{\Sigma XY}{\sqrt{\Sigma X^2} . \sqrt{\Sigma Y^2}}$$

Example 15. *Calculate the Karl Pearson's coefficient of correlation of the following data :*

x	25	27	30	35	33	28	36
y	19	22	27	28	30	23	28

Solution. Here $\bar{x} = 31$, $\bar{y} = 25$

x	y	$X = x-31$	$Y = y-25$	X^2	Y^2	XY
25	19	–6	–6	36	36	36
29	20	–2	–5	4	25	10
30	27	–1	2	1	4	–2
35	28	4	3	16	9	12
33	30	2	5	4	25	10
29	23	–2	–2	4	4	4
36	28	5	3	25	9	15
Σ: 217	175	0	0	90	112	85

The Karl Pearson's coefficient of correlation is given by

$$r_{XY} = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}} = \frac{85}{\sqrt{90 \times 112}} = 0.85.$$

NOTES

Example 16. *Compute Karl Pearson's coefficient of correlation in the following series relating to price and supply of commodity ;*

Price (Rs.)	60	65	70	75	80	85	90	95	100
Demand (Qts)	35	30	25	25	23	21	20	20	18

Solution. Computation table :

Price (x)	$d_x = x - 80$ $<d_x = x - A>$	d_x^2	*Demand (y)*	$d_y = y - 25$ $<d_y = y - B>$	d_y^2	$d_x . d_y$
60	–20	400	35	10	100	–200
65	–15	225	30	5	25	–75
70	–10	100	25	0	0	0
75	–5	25	25	0	0	0
80	0	0	23	–2	4	0
85	5	25	21	–4	16	–20
90	10	100	20	–5	25	–50
95	15	225	20	–5	25	–75
100	20	400	18	–7	49	–140
Σ	0	1500	–	–8	244	–560

$$r_{xy} = \frac{\Sigma d_x d_y - \frac{\Sigma d_x . \Sigma d_y}{9}}{\sqrt{\left[\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{9}\right]} \sqrt{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{9}}}$$

$$= \frac{-560}{\sqrt{1500} . \sqrt{244 - \frac{64}{9}}} = -0.94.$$

2.18 REGRESSION EQUATIONS

In regression analysis we can predict or estimate the value of one variable with the help of the value of other variable of the distribution after fitting to an equation. Hence there are two regression equations. The regression equation of Y on X is used to predict the value of Y with the value of X, whereas the regression equation of X on Y is used to predict the value of X with the value of Y. Here the independent variable is called predictor or explanator or regressor and the dependent variable is called explained or regressed variable.

When these regression equations are straight lines then they are called *regression lines.*

Regression Line of Y on X

$$Y = a + bX$$

Let $(x_1, y_1), (x_2, y_2), \ldots\ldots\ldots, (x_n, y_n)$ be the given observations using the method of least square, we can estimate the values of a and b and the resultant equation takes the form

$$Y - \overline{Y} = r.\frac{\sigma_Y}{\sigma_X}(X - \overline{X})$$

or,
$$y - \bar{y} = b_{yx}(x - \bar{x}) \text{ where } b_{yx} = r\frac{\sigma_y}{\sigma_x}$$
$$= \text{regression coefficient of } y \text{ on } x.$$

Regression Line of X on Y

$$X = c + dY$$

Similarly we obtain,

$$X - \overline{X} = r.\frac{\sigma_x}{\sigma_y}(Y - \overline{Y})$$

or,
$$x - \bar{x} = b_{xy}(y - \bar{y}),$$

where
$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = \text{regression coefficient of } x \text{ on } y.$$

Properties

1. Both the regression lines pass through the mean values $(\bar{x}, \bar{y})$.
2. $b_{xy}.b_{yx} = r^2$

   $\Rightarrow\ r = \pm\sqrt{b_{xy}.\,b_{yx}}$ and the sign of r is the same as of regression coefficients.
3. The two regression equations are different, unless $r = \pm 1$, in which case the two equations are identical.
4. The angle between the regression lines is given by, $\tan\theta = \frac{1-r^2}{r}.\frac{\sigma_x\,\sigma_y}{\sigma_x^2 + \sigma_y^2}$.

Formulas for Regression Coefficients

1. For, $X = x - \bar{x}$, $\quad Y = y - \bar{y}$,

$$b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}, \quad b_{yx} = \frac{\Sigma XY}{\Sigma X^2}$$

2. Generally,

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x.\Sigma y}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}, \quad b_{yx} = \frac{\Sigma xy - \frac{\Sigma x.\Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

where n = no. of observations.

NOTES

3. For $d_x = x - A$, $d_y = y - B$ where A and B are assumed values,

$$b_{xy} = \frac{\Sigma d_x . d_y - \frac{\Sigma d_x . \Sigma d_y}{n}}{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{n}}$$

$$b_{yx} = \frac{\Sigma d_x . d_y - \frac{\Sigma d_x . \Sigma d_y}{n}}{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{n}}$$

4. For grouped data

$$b_{xy} = \frac{\Sigma f d_x d_y - \frac{\Sigma f d_x . \Sigma f d_y}{N}}{\Sigma f d_y^2 - \frac{(\Sigma f d_y)^2}{N}}$$

$$b_{yx} = \frac{\Sigma f . d_x d_y - \frac{\Sigma f d_x . \Sigma f d_y}{N}}{\Sigma f d_x^2 - \frac{(\Sigma f d_x)^2}{N}}$$

where $N = \Sigma f$.

Standard Error of Estimates

Consider the regression equation of X on Y. Then the root mean square deviation of the points from the regression line of X on Y is called the standard error of estimate of X which is given by

$$S_x = \sigma_x \sqrt{1 - r^2}$$

Similarly, the standard error of estimate of Y from the regression equation Y on X is

$$S_y = \sigma_y \sqrt{1 - r^2}$$

The standard error of estimate serves a standard deviation of the size of the error of the predicted values of Y (from the equation Y on X) and of X (from the equation X on Y). The size of the standard error also helps up to assess the quality of our regression model.

Coefficient of Determination

This gives the percentage variation in the dependent variable that is accounted for or explained by the independent variable is given by

$$\text{Coefficient of determination, } R^2 = \frac{\text{Explained variance}}{\text{Total variance}}$$

Let Y be the dependent variable and X be the independent variable. If $R^2 = 0.85$ then we shall be able to reduce or explain 85% of the variation in Y with a knowledge of X.

If $\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i$ $(i = 1, 2, \ldots\ldots n,)$ be the fitted values to the observation (x_i, y_i) $i = 1, 2 \ldots\ldots n,$ then

$$R^2 = \frac{n\,\Sigma y_i^2 - (\Sigma y_i)^2}{n\,\Sigma(\hat{y}_i)^2 - (\Sigma y_i)^2}\,.\quad 0 \le R^2 \le 1$$

$R^2 = 1 \Rightarrow$ all n observations lie on the fitted regression line.

Example 17. *From the following data obtain the two regression lines and the correlation coefficient :*

Sales (*x*)	100	98	78	85	110	93	80
Purchase (*y*)	85	90	70	72	95	81	74

Find the value of y when x = 82.

Solution. Here $\Sigma x = 644$, $\Sigma y = 567$, $n = 7$

$$\therefore \quad \bar{x} = \frac{\Sigma x}{n} = 92, \qquad \bar{y} = \frac{\Sigma y}{n} = 81$$

x	$X = x - \bar{x}$	X^2	y	$Y = y - \bar{y}$	Y^2	XY
100	8	64	85	4	16	32
98	6	36	90	9	81	54
78	–14	196	70	–11	121	154
85	–7	49	72	–9	81	63
110	18	324	95	14	196	252
95	3	9	81	0	0	0
80	–12	144	70	–11	121	132
	Σ	822			616	687

The regression coefficients are

$$b_{yx} = \frac{\Sigma XY}{\Sigma X^2} = 0.84$$

$$b_{xy} = \frac{\Sigma XY}{\Sigma Y^2} = 1.12$$

Regression equation of y on x :

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

$$\Rightarrow \quad y - 81 = 0.84\,(x - 92)$$

$$\Rightarrow \quad y = 0.84x + 3.72$$

Regression equation of x on y

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

$$\Rightarrow \quad x - 92 = 1.12\,(y - 81)$$

$$\Rightarrow \quad x = 1.12y + 1.28$$

The correlation coefficient $= \sqrt{b_{xy}.b_{yx}}$ (since both coefficients are positive)

$$= \sqrt{(0.84).(1.12)} = 0.97$$

For $x = 82$, the value of y to be obtained from the regression equation of y on x. Hence $y = (0.84)\ (82) + 3.72 = 72.6$.

Example 18. *Consider the two regression lines : 3X + 2Y = 26 and 6X + Y = 31, (a) Find the mean value and correlation coefficient between X and Y. (b) If the variance of Y is 4, find the S.D. of X.*

Solution. (*a*) Intersection of two regression lines gives the mean value *i.e.*, $(\overline{X}, \overline{Y})$.

Solving the two equations, we obtain $\overline{X} = 4$ and $\overline{Y} = 7$.

Let 3X + 2Y = 26 be the regression line of X on Y and the other line as Y on X.

Then
$$X = -\frac{2}{3}Y + \frac{26}{3} \text{ (X on Y)} \Rightarrow b_{XY} = -\frac{2}{3}$$
$$Y = -6X + 31 \text{ (Y on X)} \Rightarrow b_{YX} = -6$$

but $r^2 = b_{xy}.\ b_{yx} = 4$ which cannot be true.

So we change our assumptions *i.e.*, the line 3X + 2Y = 26 represents Y on X and the other line as X on Y.

Then
$$Y = -\frac{3X}{2} + 13 \text{ (Y on X)} \Rightarrow b_{YX} = -\frac{3}{2}$$
$$X = -\frac{1}{6}Y + \frac{31}{6} \text{ (X on Y)} \Rightarrow b_{XY} = -\frac{1}{6}$$

$\therefore$
$$r = -\sqrt{b_{YX}.b_{XY}} \text{ (Since both the coefficients are negative)}$$
$$= -\sqrt{\frac{3}{2} \times \frac{1}{6}} = -\frac{1}{2}$$

(*b*) Given $\sigma_Y^2 = 4 \Rightarrow \sigma_Y = 2$

We have
$$b_{XY} = r.\frac{\sigma_X}{\sigma_Y}$$

$\Rightarrow$
$$\sigma_X = \frac{\sigma_Y.\ b_{XY}}{r} = \frac{2 \times (-1/6)}{(-1/2)} = \frac{2}{3}.$$

2.19 RANK CORRELATION

Let us suppose that a group of n individuals is given grades or ranks with respect to two characteristics. Then the correlation obtained between these ranks assigned on two characteristics is called rank correlation.

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be the ranks of the i-th individual in two characteristics. Then Spearman's Rank correlation coefficient is given as

$$r = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)}$$

where $d_i = x_i - y_i$.

It is also to be noted that $-1 \le r \le 1$ and the above formula is used when ranks are not repeated.

For repeated ranks, a correction factor is required in the formula. If m is the number of times an item is repeated then the factor $\frac{m(m^2-1)}{12}$ is to be added to Σd^2. For each repeated value, this correction factor is to be added.

Example 19. *The ranks of some 10 students in two subjects A and B are given below :*

Ranks in A	5	2	9	8	1	10	3	4	6	7
Ranks in B	10	5	1	3	8	6	2	7	9	4

Calculate the rank correlation coefficient.

Solution. Here $n = 10$.

Ranks in A	5	2	9	8	1	10	3	4	6	7	
Ranks in B	10	5	1	3	8	6	2	7	9	4	
d = A – B	–5	–3	8	5	–7	4	1	–3	–3	3	Σ
d^2	25	9	64	25	49	16	1	9	9	9	216

Rank correlation coefficient is given by

$$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 216}{10 \times 99} = -0.31.$$

Example 20. *Obtain the rank correlation coefficient for the following data :*

X	85	74	85	50	65	78	74	60	74	90
Y	78	91	78	58	60	72	80	55	68	70

Solution. Here $n = 10$

X	*Y*	*Rank X* (*x*)	*Rank Y* (*y*)	$d = x - y$	d^2
85	78	2.5	3.5	–1	1
74	91	6	1	5	25
85	78	2.5	3.5	–1	1
50	58	10	9	1	1
65	60	8	8	0	0
78	72	4	5	–1	1
74	80	6	2	4	16
60	55	9	10	–1	1
74	68	6	7	–1	1
90	70	1	6	–5	25
			Σ	0	72

In the X series 85 has repeated twice and given ranks 2.5 instead of 2 and 3. For this the correction factor is $\frac{2(4-1)}{12} = \frac{1}{2}$.

Also, 74 has repeated thrice in X series and given ranks 6 instead of 5, 6, 7. For this the correction factor is $\frac{3(9-1)}{12} = 2$.

In the Y series 78 has repeated twice and given ranks 3.5 instead of 3 and 4. For this the correction factor is $\frac{2(4-1)}{12} = \frac{1}{2}$.

So the total correction factors = $\frac{1}{2} + 2 + \frac{1}{2} = 3$

Then the rank correlation coefficient is given by

$$r = 1 - \frac{6(72+3)}{10(100-1)} = 0.545.$$

2.20 CORRELATION OF BIVARIATE FREQUENCY DISTRIBUTION

Consider the two way frequency table with marginal and joint distributions of the two variables X and Y.

X Y			Classes Mid Points x_1 x_2 ...... x_n	
Classes	Mid-Points	y_1 y_2 ⋮ y_m	$f(x, y)$	$g(y_1) = \sum_x f(x, y_1)$ $g(y_2) = \sum_x f(x, y_2)$ $g(y_m) = \sum_x f(x, y_m)$
			$h(x_1)$ ...... $h(x_n)$ $= \sum_y f(x_1, y)$ $= \sum_y f(x_n, y)$	$\mathbf{N} = \sum_x \sum_y f(x, y)$

Here
$$\bar{x} = \frac{1}{N} \Sigma x_i\, h(x_i), \quad \bar{y} = \frac{1}{N} \Sigma y_j\, g(y_j)$$

or simply
$$= \frac{1}{N}. \Sigma x h(x) \qquad = \frac{1}{N} \Sigma y\, g(y)$$

$$\sigma^2_x = \frac{1}{N} \Sigma\, x^2 h(x) - (\bar{x})^2, \quad \sigma^2_y = \frac{1}{N} \Sigma y^2\, g\,(y) - (\bar{y})^2$$

$$Cov\,(x,y) = \frac{1}{N} \sum_x \sum_y xy\ f(x,y) - \bar{x}.\bar{y}$$

Hence,
$$r_{xy} = \frac{Cov.\ (x,y)}{\sigma_x \,.\, \sigma_y}$$

Note. For large data, the two way frequency table is advantageous.

Example 21. *Calculate the correlation coefficient from the following table.*

x / y	0 – 8	8 – 16	16 – 24
1 – 5	2	0	4
5 – 9	3	2	2
9 – 13	2	5	1

Solution. Consider the following table :

	x	Mid-values			$g(y)$
	y	4	12	20	
Mid-Values	3	2	0	4	6
	7	3	2	2	7
	11	2	5	1	8
	$h(x)$	7	7	7	21

$$\bar{x} = \frac{1}{21}[(4)(7) + (12)(7) + (20)(7)] = 12$$

$$\bar{y} = \frac{1}{21}[(3)(6) + (7)(7) + (11)(8)] = 7.38$$

$$\sigma_x^2 = \frac{1}{21}[(16)(7) + (144)(7) + (400)(7)] - (12)^2 = 42.67$$

$$\sigma_y^2 = \frac{1}{21}[(9)(6) + (49)(7) + (121)(8)] - (7.38)^2 = 10.54$$

$$Cov\,(x,y) = \frac{1}{21}\sum_x \sum_y xy\, f(x,y) - \bar{x}\,.\,\bar{y}$$

$$= \frac{1764}{21} - (12)(7.38) = -4.56$$

The correlation coefficient is

$$r_{xy} = \frac{-4.56}{\sqrt{42.67}\,.\sqrt{10.54}} = -0.22.$$

2.21 MULTIPLE REGRESSION

If more than one predictor variable is present in the regression equation then it is called multiple regression. Let us consider two predictor variables and one regressed variable and the multiple linear regression model can be stated as follows:

$$y = a_0 + a_1 x_1 + a_2 x_2$$

To estimate a_0, a_1 and a_2 we take (x_{1i}, x_{2i}, y_i), $i = 1,2,\ldots\ldots, n$ as observed data where the x's are assumed to be known without error while the y values are random variables.

Let $S = \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i})]^2$ be the sum of squares of errors. Then to minimize S, we take $\frac{\partial S}{\partial a_0} = 0$, $\frac{\partial S}{\partial a_1} = 0$ and $\frac{\partial S}{\partial a_2} = 0$. From which we obtain three normal equations

$$\begin{aligned} \Sigma y_i &= n\, a_0 + a_1\, \Sigma x_{1i} + a_2\, \Sigma\, x_{2i} \\ \Sigma x_{1i}\, y_i &= a_0 \Sigma\, x_{1i} + a_1\, \Sigma x_{1i}^2 + a_2\, \Sigma\, x_{1i}\, x_{2i} \\ \Sigma\, x_{2i}\, y_i &= a_0\, \Sigma\, x_{2i} + a_1\, \Sigma\, x_{1i}\, x_{2i} + a_2\, \Sigma x_{2i}^2 \end{aligned}$$

By solving these equations we obtain the least square estimates of a_0, a_1 and a_2.

Note. 1. In the above model if a_2 vanishes then it is the regression line of y on x.

2. The above model represents a regression plane.

Example 22. *Consider the following data:*

x_1	2	4	5	6	3	1
x_2	1	2	1	3	5	2
y	14	16	17	20	18	12

Fit a least squares regression plane.

Solution. Here $n = 6$. Let the regression plane be $y = a_0 + a_1 x_1 + a_2 x_2$.

	x_1	x_2	x_1^2	x_2^2	$x_1\, x_2$	$x_1\, y$	$x_2\, y$	y
	2	1	4	1	2	28	14	14
	4	2	16	4	8	64	32	16
	5	1	25	1	5	85	17	17
	6	3	36	9	18	120	60	20
	3	5	9	25	15	54	90	18
	1	2	1	4	2	12	24	12
Σ	21	14	91	44	50	363	237	97

The three normal equations can be written as follows :

$$\begin{aligned} 97 &= 6a_0 + 21\, a_1 + 14a_2 \\ 363 &= 21\, a_0 + 91\, a_1 + 50a_2 \\ 237 &= 14a_0 + 50a_1 + 44\, a_2. \end{aligned}$$

By solving we obtain,

$$a_0 = 9.7, \quad a_1 = 1.3, \quad a_2 = 0.83.$$

2.22 CURVILINEAR REGRESSION

Consider one predictor variable and one regressed variable. Then there may be curvilinear (or non-linear) relationship between these variables. We consider some important relationships. To estimate the unknowns, the method of least square is to be applied. Briefly, we illustrate.

(*a*) $y = a + bx + cx^2$ (Parabolic equation)

Consider S = Sum of squares of errors

$$= \Sigma[y-(a+bx+cx^2)]^2$$

To minimize S, we take $\frac{\partial S}{\partial a}=0, \frac{\partial S}{\partial b}=0$ and $\frac{\partial S}{\partial c}=0$ and obtain the normal equations as follows:

$$\Sigma y_i = na + b\,\Sigma x_i + c\,\Sigma x_i^2$$

$$\Sigma x_{1i}\ y_i = a\,\Sigma x_i + b\,\Sigma x_i^2 + c\,\Sigma x_i^3$$

$$\Sigma\, x_i^2\ y_i = a\,\Sigma x_i^2 + b\,\Sigma x_i^3 + c\,\Sigma x_i^4$$

Solving these equations we obtain the estimates of a, b and c. For $y = a + bx$ (linear fit), there will be the first-two normal equations with $c = 0$.

(*b*) $y = ae^{bx}$ (Exponential equation)

To estimate a and b, take first log (base 10) on both sides

$$\log y = \log a + bx \log e$$

$\Rightarrow$ $\quad Y = A + Bx$ which is linear regression.

where $\quad Y = \log y,\ A = \log a,\ \ B = b \log e$

The normal equations are

$$\Sigma Y_i = nA + B\Sigma x_i$$

$$\Sigma x_i Y_i = A\ \Sigma x_i + B\ \Sigma x_i^2$$

The estimation of A and B will give the estimation of a and b.

(*c*) $y = ax^b$ (Geometric curve / Power function)

To estimate a and b, take first log (base 10) on both sides

$$\log y = \log a + b \log x$$

$\Rightarrow$ $\quad Y = A + bX$

where $Y = \log y$, $A = \log a$, $X = \log x$.

The normal equations are

$$\Sigma Y_i = nA + b\ \Sigma X_i$$

$$\Sigma X_i\ Y_i = A\,\Sigma X_i + b\Sigma X_i^2$$

Here the estimation of A will give the estimation of a, while the estimation of b is obtained directly from normal equations.

(*d*) $xy^a = b$ (Gas equation)

To estimate a and b, take first log (base 10) on both sides

$$\log x + a \log y = \log b.$$

$\Rightarrow$ $\quad \log y = \frac{1}{a}\log b - \frac{1}{a}\log x$

$\Rightarrow$ $\quad Y = A + BX$

where $\quad Y = \log y,\ A = \frac{1}{a}\log b,\ B = -1/a,\ X = \log x$

The normal equations are

$$\Sigma Y_i = nA + B\Sigma X_i$$

$$\Sigma X_i\ Y_i = A\Sigma X_i + B\Sigma X_i^2$$

The estimation of A and B will give the estimation of a and b.

(e) $y = ab^x$ **(Growth of bacteria)**

To estimate a and b, take first log (base 10) on both sides

$$\log y = \log a + x \log b$$

$$\Rightarrow \quad Y = A + xB.$$

The normal equations are

$$\Sigma Y_i = nA + B\Sigma x_i$$

$$\Sigma x_i Y_i = A\Sigma x_i + B\Sigma x_i^2$$

The estimation of A and B will give the estimation of a and b.

Example 23. *Find the best fitting regression equation of type $y = a + bx + cx^2$ to the following data :*

x	3	2	1	0	–1	–2	–3
y	10	8	3	1	2	6	8

Solution. Here $n = 7$.

	x	y	x^2	x^3	x^4	xy	x^2y
	3	10	9	27	81	30	90
	2	8	4	8	16	16	32
	1	3	1	1	1	3	3
	0	1	0	0	0	0	0
	–1	2	1	–1	1	–2	2
	–2	6	4	–8	16	–12	24
	–3	8	9	–27	81	–24	72
Σ	0	38	28	0	196	11	223

The normal equations are

$$7a + 28c = 38$$

$$28b = 11$$

$$28a + 196c = 223.$$

By solving these equations we obtain,

$a = 2.048 \qquad b = 0.393 \qquad c = 0.845$

Example 24. *Find the best fitting regression equation of type $y = ax^b$ to the following data :*

x	1	2	3	4	5	6
y	2	16	54	128	250	432

Solution. $y = ax^b$

Taking log on both sides we obtain

$$\log y = \log a + b \log x$$

$$\Rightarrow \quad Y = A + bX.$$

where, $Y = \log y$, $A = \log a$, $X = \log x$.

Here $n = 6$

x	X	y	Y	XY	X^2
1	0	2	0.3010	0	0
2	0.3010	16	1.2041	0.3624	0.0906
3	0.4771	54	1.7324	0.8265	0.2276
4	0.6021	128	2.1072	1.2687	0.3625
5	0.6990	250	2.3979	1.6761	0.4886
6	0.7782	432	2.6355	2.0509	0.6056
Σ	2.8574	–	10.3781	6.1846	1.7749

The normal equations are

$$6A + 2.8574 = 10.3781$$

$$2.8574A + 1.7749\,b = 6.1846$$

By solving these equations we obtain,

$$A = 0.3011 \Rightarrow a = 2.0004$$

and

$$b = 2.9996$$

Hence the best fitting regression equation is

$$y = 2.0004\,x^{2.9996}.$$

SUMMARY

- For quantitative data it is observed that there is a tendency of the data to be distributed about a central value which is a typical value and is called a measure of central tendency. It is also called a measure of location because it gives the position of the distribution on the axis of the variable.
- The arithmetic mean is simply called 'Average'. For the observations $x_1, x_2, ..., x_n$ the A.M. is defined as

$$\bar{x} = \text{A.M.} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- The geometric mean of the observations $x_1, x_2, ...,x_n$ is defined as

$$\text{G.M.} = (x_1.x_2............x_n)^{1/n}$$

- *The reciprocal of the H.M. of a variate is the A.M. of its reciprocal.* For the observations $x_1, x_2, ...,x_n$

$$\text{H.M.} = \frac{n}{\Sigma(1/x_i)}$$

- For the observations $x_1, x_2, ...,x_n$ the median is the middle value if the number of observations is odd and have been arranged in ascending or descending order of magnitude. For even number of observations the median is taken as the average of two middle values after they are arranged in ascending or descending order of magnitude.

- Mode is the value of a variable which occurs most frequently in a set of observations.
- As the median divides an array into two parts, the quartiles divide the array into four parts, the deciles divide it into ten parts and the percentiles divide it into one hundred parts.
- The measure which gives the idea of the amount of scattering of the data around the central value is called the measure of dispersion.
- Standard deviation (and variance) is a relative measure of the dispersion of a set of data—the larger the standard deviation, the more spread out the data.
- If x_1, x_2, ..., x_n be the n observations then M.D. is defined as

$$\text{M.D.} = \frac{1}{n}\Sigma|x-A|$$

- It is a location based measure of dispersion and is defined as follows:

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

- It is the simplest measure for dispersion. For the observations x_1, x_2, ..., x_n it is defined as

$$R = (\text{largest value}) - (\text{smallest value})$$

- For n observations x_1, x_2,..........,x_n and an arbitrary constant A, the rth moment about A is defined as,

$$\mu_r' = \frac{1}{n}\sum_i (x_i - A)^r, \quad r = 0,\ 1,\ 2, \ldots\ldots$$

- Skewness is a measure of symmetry of the shape of frequency distribution, *i.e.*, it reveals the dispersal of value on either side of an average is symmetrical or not.
- Kurtosis is a measure of peakness of a distribution and is defined in terms of second and fourth central moments.

PROBLEMS

1. Suppose a train moves 5 hrs at a speed of 40 km/hr, then 3 hrs at a speed of 45 km/hr and next 5 hrs at a speed of 60 km/hr. Calculate the average speed.
2. A factory has 4 sections, the no. of workers in the different sections being 50, 100, 60 and 150. The average wages per worker are Rs. 100, Rs. 120, Rs. 130 and Rs. 110 respectively, calculate the average wages for all the workers together.
3. Compute the A.M., G.M. and H.M. from the following:

Class	15–19	20–24	25–29	30–34	34–39	40–44
Frequency	13	32	4	42	58	51

4. The mean of 15 items is 34. It was found out, later on, that the two items 48 and 32 were wrongly copied as 84 and 23 respectively. Find out the correct mean.

5. The number of telephone calls received daily in a marketing department of a company for 200 days are given below :

No. of Calls	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50
Frequency	7	15	24	31	42	30	26	15	10

Calculate the mean, median and mode of the telephone calls.

6. Consider the following distribution of humidity readings in a certain place for 60 days.

Humidity	10–19	20–29	30–39	40–49	50–59	60–69
Frequency	3	15	15	20	5	2

Calculate (*i*) Q_1, (*ii*) Q_3, (*iii*) D_5, (*iv*) P_{90}.

7. From the following distribution of marks, calculate

(*i*) mean, (*ii*) median, (*iii*) mode, (*iv*) D_5 (*v*) P_{85}

Marks (more than)	0	10	20	30	40	50	60	70	80
No. of students	150	140	100	80	80	70	30	14	0

8. Find the missing frequencies in the following distribution when it is known that A.M. = 59.5

Daily wages (Rs.)	10–20	20–35	35–60	60–90	90–105	105–120	120–150
Percentage of wage earners	10	f_1	30	15	f_2	5	5

Where total of percentage of wage earners is 100.

9. An analysis of production rejects resulted in the following figures:

No. of rejects per operator	*No. of operators*	*No. of rejects per operator*	*No. of operators*
21–25	5	41–45	15
26–30	15	46–50	12
31–35	28	51–55	3
36–40	42		

Compute mean, median and mode of the production rejects.

10. An aeroplane travels distances of d_1, d_2 and d_3 km at speeds V_1, V_2 and V_3 km per hour respectively. Show that the average speed is given by v, where

$$\frac{d_1 + d_2 + d_3}{v} = \frac{d_1}{v_1} + \frac{d_2}{v_2} + \frac{d_3}{v_3}.$$

11. Find the M.D. and variance of the following distribution.

x	2	3	4	5	6	7	8	9	10
f	1	1	2	4	4	3	7	5	3

12. Find the M.D. and S.D. for the following data:

Score	4–5	6–7	8–9	10–11	12–13	14–15
Frequency	4	10	20	15	8	3

NOTES

13. With reference to the following data verify that the M.D. is the least when deviations are measured about the median instead of mean.

Class	5–9	10–14	15–19	20–24	25–29	30–34
Frequency	21	19	13	8	16	23

14. Calculate the mean age and the variance of the ages of the members of a society from the following table:

Age at nearest birthday	30–34	35–39	40–44	45–49	50–54	55–59	60–64
No. of members	3	9	18	25	26	15	4

15. Calculate the S.D. and M.D. about both mean and median for the first n integers.

16. In a series of 5 observations the values of mean and variance are 4.4 and 8.24. If three observations are 1, 2 and 6. find the other two.

17. Two batsman A and B made the following scores in a series of cricket matches:

A	14	13	26	53	17	29	79	36	84	49
B	37	22	56	52	14	10	37	48	20	4

Who is more consistent player?

18. From the data given below, find which series is more uniform.

Class	*Sereis A*	*Series B*
0 – 10	7	5
10 – 20	6	8
20 – 30	15	12
30 – 40	12	15
40 – 50	10	10

19. Calculate (*i*) Q.D., and (*ii*) S.D. of wages from the following data:

Weekly wages (Rs.)	35–36	36–37	37–38	38–39	39–40	40–41	41–42
No. of persons	14	20	42	54	45	18	7

20. Measurements of the lengths in metres of 50 iron rods are distributed as follows:

Class	*Frequency*	*Class*	*Frequency*
2.35 – 2.45	1	2.75 – 2.85	11
2.45 – 2.55	4	2.85 – 2.95	10
2.55 – 2.65	7	2.95 – 3.05	2
2.65 – 2.75	15		

Calculate (*i*) Q.D. and (*ii*) M.D. of the above data (*iii*) Coefficient of Q.D.

21. The mean, median and the coefficient of variation of 100 observations are found to be 60, 56 and 30 respectively. Find the coefficient of skewness of the above system of 100 observations.

22. Compute the Karl Pearson's coefficient of skewness of the following distribution:

Class	*Frequency*	*Class*	*Frequency*
2.7 – 2.9	2	4.2 – 4.4	113
3.0 – 3.2	16	4.5 – 4.7	71
3.3 – 3.5	46	4.8 – 5.0	22
3.6 – 3.8	88	5.1 – 5.3	4
3.9 – 4.1	138		

23. Compute mean, variance, β_1 and β_2 if the first four moments about a value 5 of a variable are given as 2, 20, 38 and 52.

24. The first three moments of a distribution about the value 3 are –1, 10, –28. Find the values of mean, standard deviation and the moment measure of skewness.

25. For a distribution the mean is 9, standard deviation is 5, $\sqrt{\beta_1} = 1$ and $\beta_2 = 3$. Obtain the first four moments about the origin *i.e.*, zero.

26. Find the appropriate measure of skewness from the following data :

Value	Less than 10	10–20	20–30	30–40	40–50	50 and above
Frequency	5	9	16	7	6	7

27. Find the skewness and kurtosis of the following distribution by central moments and comment on the type.

Class	0–10	10–20	20–30	30–40	40–50
Frequency	10	20	40	20	10

28. Find the C.V. of a frequency distribution given that its mean is 100, mode = 120 and Karl Pearson's coefficient of skewness = – 0.2.

29. Compare the skewness of two frequency distribution whose moments about the origin are as follows :

Distribution	μ'_1	μ'_2	μ'_3
A	2	5	14
B	2	5	1014

30. Calculate the coefficient of skewness and kurtosis of the following data :

Class	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20
Frequency	4	10	6	12	8

31. Calculate the (*i*) two regression coefficients, (*ii*) coefficient of correlation and (*iii*) the two regression equations from the following information :

$n = 10$, $\Sigma X = 350$, $\Sigma Y = 310$, $\Sigma(X - 35)^2 = 162$, $\Sigma\,(Y - 31)^2 = 222$; and

$\Sigma\,(X - 35)\,(Y - 31) = 92$.

32. Calculate the correlation coefficient of the following data :

x	45	46	46	47	48	49	50
y	44	48	45	48	52	51	49

33. In order to find the correlation coefficient between two variables X and Y from 20 pairs of observations, the following calculations were made :

$\Sigma x = 120, \quad \Sigma y = 70, \quad \Sigma x^2 = 780, \quad \Sigma y^2 = 450$ and $\Sigma xy = 500.$

On subsequent verification it was discovered that the two pairs ($x = 9, \ y = 11$) and ($x = 7$, y = 6) were copied wrongly, the correct values being ($x = 9, \ y = 12$) and ($x = 6, y = 7$). Obtain (*i*) the correct value of correlation coefficient, (*ii*) the two lines of regression, and (*iii*) angle between them.

34. (*a*) In a correlation analysis, the value of the Karl Pearson's coefficient of correlation and its probable error were found to be 0.90 and 0.40 respectively. Find the value of n.

(*b*) Given that $x = 4y + 5$ and $y = Kx + 4$ are the regression lines of x on y and y on x, respectively, show that $0 \le K \le 25$. If K = 0.10 actually, find the means of the variables x and y and also their coefficient of correlation.

35. Three judges give the following ranks of ten different models of a car having different attributes.

Models		1	2	3	4	5	6	7	8	9	10
Judge	A	3	9	6	8	5	4	2	1	10	7
	B	10	5	2	1	7	4	9	3	6	8
	C	7	8	1	5	3	10	6	2	4	9

Discuss which pair of judges have the nearest approach to common tastes of the car models.

36. The coefficient of rank correlation of the marks obtained by 10 students in two subjects was found to be 0.62. It was later discovered that the difference in ranks for one student was taken wrongly as 5 instead of 7. Find the correct coefficient of rank correlation.

37. *Fit a curve* $xy^a = b$ *to the following data:*

x	2	4	6	8	10
y	3	7	11	12	9

38. If the growth of a certain kind of bacteria follows the law $y = ab^x$, then find the best fitting values of a and b using the following data :

x	1	2	3	4	5
y	233.2	253.4	282.3	302.4	332.8

39. For the data given below, find the equation to the best fitting exponential curve of the form $y = ae^{bx}$

x	2	4	6	8	10	12
y	120	102	85	74	62	50

40. An employment bureau asked applicants their weekly wages on jobs last held. The actual wages were obtained for 54 of them and are recorded in the table below: x represents reported wage, y actual wage and the entry in the table represents frequency. *Find the correlation coefficient.*

y / x	15	20	25	30	35	40
40						2
35				3	5	
30			4	15		
25			20			
20		3	1			
15	1					

41. Calculate the Karl Pearson's coefficient of correlation between X and Y from the bivariate sample of 140 pairs of X and Y as distributed below :

X Y	10 – 20	20 – 30	30 – 40	40 – 50
10 – 20	20	26	–	–
20 – 30	8	14	37	–
30 – 40	–	4	18	3
40 – 50	–	–	4	6

ANSWERS

1. 48.85 km/hr.

2. Rs. 102.22

3. A.M. = 33.33, GM = 32.22, H.M. = 30.95.

4. Corrected mean = 32.2

5. Mean = 27.875, Median = 27.74, Mode = 27.39

6. $Q_1 = 27.5$, $Q_3 = 37.5$, $D_{10} = 45.5$, $P_{90} = 51.5$

7. Mean = 39.27, Median = 45, Mode = 56.46 (using empirical relation)

$D_5 = 45$, $P_{85} = 64.69$

8. $f_1 = 20$, $f_2 = 15$.

9. Mean = 36.96, Median = 36.64, Mode = 37.207.

11. M.D. = 1.773, Variance = 4.432

12. M.D. = 2.03, S.D. = 2.4756

14. Mean = 48.15, Variance = 49.43

15. S.D. = $\sqrt{(n^2-1)/12}$, M.D. about mean = M.D. about median

$$= \frac{n^2-1}{4n}, \text{ when } n = \text{odd integer}$$

$$= \frac{n}{4}, \text{ when } n = \text{even integer}$$

16. 4 and 9

17. C.V. (A) = 61.1% C.V. (B) = 58.47%, Batsman B is more consistent.

18. Series B. C.V. (A) = 47.05, C.V. (B) = 36.06

19. (*i*) Q.D. = 1.03, (*ii*) S.D. = 1.46

20. (*i*) Q.D. = 0.096, (*ii*) M.D. = 0.113, (*iii*) 0.035.

21. 0.67 **22.** 0.176

23. $\bar{x} = 7$, $\mu_2 = 16$, $\beta_1 = 1.06$, $\beta_2 = 0.70$ **24.** Mean = 2, S.D. = 3, $\beta_1 = 0$

25. $\mu'_1 = 9$, $\mu'_2 = 106$, $\mu'_3 = 1529$, $\mu'_4 = 25086$

26. Bowley's coefficient of skewness is suitable and = 0.24

27. $\beta_1 = 0$, $\beta_2 = 2.5$, symmetrical and platykurtic **28.** 100

29. A is symmetric, B is positively skew **30.** $\beta_1 = 0.038$, $\beta_2 = 1.806$

31. (*i*) $b_{xy} = 0.41$, $b_{yx} = 0.57$ (*ii*) 0.48, (*iii*) $y = 0.57x + 11.05$, $x = 0.41y + 22.29$

32. 0.754

33. (*i*) 0.754, (*ii*) $y = 3.918x - 20.3$, $x = 0.435y + 4.57$, (*iii*) 13°54'

34. (*a*) $n = 10$, (*b*) $\bar{x} = 26.67, \bar{y} = 7.5, r = 0.63$**35.** B and C

36. 0.47 **37.** $a = -1.27$, $b = 0.4$

38. $a = 213.45$, $b = 1.09$ **39.** $a = 143.64$, $b = -0.09$

40. 0.93 **41.** -0.705.

FURTHER READINGS

1. Statistics and Operations Research: A unified Approach; Dr Debashis Dutta.
2. Topics in Business Mathematics and statistics: Dr. Qazi shoeb Ahmad Dr. Mohd. Vaseen Jamail, Shadab Ahmad Khan.

UNIT 3 PROBABILITY AND PROBABILITY DISTRIBUTION

★ STRUCTURE ★

3.0 LEARNING OBJECTIONS

After studying this unit you will be able to:

- explain axioms of mathematical probability
- illustrate various theorems on probability
- define binomial, poisson and normal distribution.

3.1 INTRODUCTION

Random experiments are those experiments whose results depend on chance. For example, tossing a coin, where head or tail can turn up in a single toss. When a space shuttle takes off, then returning to the ground depends on several chance factors. To satisfy the demand of an item we deal with random experiment.

The single performance of a random experiment will be called an **outcome.** For example, head or tail is the outcome of tossing a coin.

A set of all possible outcomes of an experiment is called a **sample space.** For example, the sample space of tossing two coins is {(H, H), (H, T), (T, H), (T, T)} where H = head and T = tail. Consider two cities A and B connected by two roads R_1 and R_2. Another city C connected to B by a single road R_3.

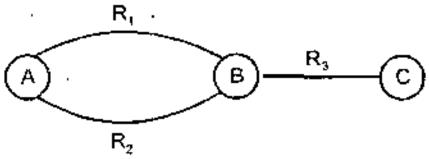


Fig. 3.1

NOTES

The possible travel times in the roads are as follows:

$$R_1 = 3\text{hr},\ \ 3.5\text{ hr}.$$

$$R_2 = 3\text{hr},\ \ 3.75\text{ hr}.$$

$$R_3 = 2\text{hr},\ \ 1.75\text{ hr}.$$

Then the sample space of possible travel times for this network from A to C is

{3 + 2, 3 + 1.75, 3.5 + 2, 3.5 + 1.75, 3 + 2, 3 + 1.75, 3.75 + 2, 3.75 + 1.75}

Any subset of a sample space is called an **event.** Events may be elementary or composite. In case of elementary event it cannot be decomposed into simpler events whereas the composite event is an aggregate of several elementary events.

A sample space is said to be **discrete** if it contains finite or countably infinite elements. For continuum elements, the sample space is said to be **continuous.**

If all the possible events in a random experiment are considered then this set is called **exhaustive.**

In tossing a coin, either head or tail will turn up. Both cannot turn up. These type of events are called **mutually exclusive.** In the above travel example, reaching to B from A can be made either by road R_1 or by road R_2 which is mutually exclusive. If none of the events in a sample space can be expected in preference to another then these events are said to be **equally likely.**

Consider a random experiment with possible results as cases. If there are n exhaustive, mutually exclusive and equally likely cases and of them m are favourable to an event A, then the **probability** of A is defined by

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

The above definition is known as **classical definition** of probability.

Example 1. *What is the probability of getting 3 tails in tossing 3 coins ?*

Solution. The sample space = {TTT, TTH, THT, THH, HTT, HTH, HHT, HHH}

Since three tails have occurred only once.

$$\therefore \qquad P(3T) = \frac{1}{8}$$

The probability of a composite event is the sum of the probabilities of the simple events of which it is composed.

Example 2. *Find the probability of exactly two tails in tossing 3 coins ?*

Solution. From the sample space (given in example 1) exactly two tails occur as TTH, THT and HTT.

Hence P(exactly two tails) $= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}.$

3.2 AXIOMS OF MATHEMATICAL PROBABILITY

Let A be an event in a sample space S, then

(*i*) $0 \leq P(A) \leq 1$

[Here P(A) = 0 means that the event will not occur and P(A) = 1 means that the event is certain]

(*ii*) P(S) = 1

(*iii*) If A and B are two mutually exclusive events then

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

3.3 COMPLEMENTATION RULE

Let A be an event and $\overline{A}$ be its complement, then

$$P(A) + P(\overline{A}) = 1$$

For example, if the probability of hitting a target by a missile is $\frac{1}{6}$ then its complement of not hitting the target will be given by $1 - \frac{1}{6} = \frac{5}{6}$.

Therefore if P(A) denotes the probability of occurrence of event A then $P(\overline{A})$ denotes the probability that event A fails to occur.

Example 3. *Two dice are thrown. What is the probability that the sum of two faces is multiple of 3 ?*

Solution. Each dice has the numbers 1, 2, 3, 4, 5, 6. Then the total number of cases *i.e.,* the sample space will consist of $6^2 = 36$ elements.

The favourable cases *i.e.,* the sum of two faces is multiple of 3 are given by (1,2), (2,1), (5,1), (1,5), (4,2), (2,4), (3,3), (3,6), (6,3), (5,4), (4,5), (6,6) *i.e.,* 12 cases.

Hence P (sum of two faces is multiple of 3) $= \frac{12}{36} = \frac{1}{3}$.

Example 4. *A batch contains 10 articles of which 3 are defective. If 4 articles are chosen at random, what is the probability that none of them is defective ?*

Solution. Total number of ways of selecting 4 articles out of 10 is $\binom{10}{4} = 210$.

If none of the selected articles is defective, which must come from 7 non-defective articles. So the number of favourable cases is $\binom{7}{4} = 35$.

Hence the required probability $= \frac{35}{210} = \frac{1}{6}$.

Example 5. *A room has three lamp sockets. From a collection of 10 light bulbs of which only 6 are good, a person selects 3 at random and puts them in the sockets. What is the probability that the room will have light?*

Solution. From the given 10 bulbs, 6 are good and 4 are damaged or bad bulbs. If the person selects at least one good bulb, then the room will have light.

P(room will not have light) = P(the person selects 3 bad bulbs)

$$= \frac{\binom{4}{3}\times\binom{6}{0}}{\binom{10}{3}} = \frac{1}{30}.$$

Using complementation rule; we have

P(the room will have light) =1 – P(Room will not have light) $= 1 - \frac{1}{30} = \frac{29}{30}$.

3.4 THEOREM OF TOTAL PROBABILITY/ADDITION THEOREM

I. If two events A and B are mutually exclusive, then the occurrence of either A or B is given by

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

[Also we can write $P(A+B) = P(A \cup B)$].

Proof. Let n possible outcomes from a random experiment which are mutually exclusive, exhaustive and equally likely. If n_1 of these outcomes are favourable to the event A, and n_2 outcomes are favourable to the event B, then

$$P(A) = \frac{n_1}{n}, \qquad P(B) = \frac{n_2}{n}$$

Since the events A and B are mutually exclusive *i.e.,* the n_1 outcomes are completely distinct from n_2 outcomes, then the number of outcomes favourable to either A or B is $n_1 + n_2$.

$$\therefore \qquad P(A+B) = \frac{n_1+n_2}{n} = \frac{n_1}{n} + \frac{n_2}{n} = P(A) + P(B).$$

II. When the two events A and B are not mutually exclusive, then the probability of occurrence of at least one of the 2 events is given by

$$P(A+B) = P(A) + P(B) - P(AB)$$

Proof. Here the event A + B means the occurrence of one of the following mutually exclusive events : AB, $A\overline{B}$ and $\overline{A}B$. Therefore

$$P(A+B) = P(AB + A\overline{B} + \overline{A}B) = P(AB) + P(A\overline{B}) + P(\overline{A}B)$$

Again, we have $\qquad P(A) = P(AB) + P(A\overline{B})$

$\Rightarrow \qquad P(A\overline{B}) = P(A) - P(AB)$

and $\qquad P(B) = P(AB) + P(\overline{A}B)$

$\Rightarrow \qquad P(\overline{A}B) = P(B) - P(AB)$

Hence,

$$P(A+B) = P(AB) + [P(A) - P(AB)] + [P(B) - P(AB)]$$
$$= P(A) + P(B) - P(AB).$$

Note. 1. For the events A, B and C which may not be mutually exclusive,

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

2. **Boole's inequality.**

$P(A+B) \leq P(A)+P(B)$.

Here equality sign holds when P(AB) = 0, *i.e.*, A and B are mutually exclusive.

3. **Bonferroni's inequality.**

$P(AB) \geq P(A)+P(B)-1$.

3.5 THEOREM OF COMPOUND PROBABILITY/ MULTIPLICATION THEOREM

Let A and B be two events in a sample space S and $P(A) \neq 0$, $P(B) \neq 0$, then the probability of happening of both the events are given by

(*i*) P(AB) = P(A).P(B/A)

(*ii*) P(AB) = P(B).P(A/B).

[Here P(B/A) denote the conditional probability which means the probability of event B such that the event A has already occurred otherwise event B will not occur. Similarly P(A/B).]

Proof. Let n possible outcomes from a random experiment which are mutually exclusive, exhaustive and equally likely. If n_1 of these outcomes are favourable to the event A, the unconditional probability of A is

$$P(A)=\frac{n_1}{n}$$

Out of these n_1 outcomes, let n_2 outcomes be favourable to another event B *i.e.*, the number of outcomes favourable to A as well as B is n_2. Hence,

$$P(AB)=\frac{n_2}{n}$$

Then the conditional probability of B assuming that A has already occurred is

$$P(B/A)=\frac{n_2}{n_1}$$

Therefore, $$\frac{n_2}{n}=\frac{n_1}{n}.\frac{n_2}{n_1}$$

⇒ P(AB) = P(A). P(B/A) which is (*i*).

Similarly, we can prove (*ii*).

Note. If the occurrence of the event A as well as B as well as C is given by

P(ABC) = P(A). P(B/A) P(C/AB).

3.6 INDEPENDENT EVENTS

Two events A and B in a sample space S are said to be independent if P(AB) = P(A).P(B).

For n independent events $A_1, A_2, ..., A_n$

$$P(A_1.A_2...A_n) = P(A_1)\,P(A_2)...P(A_n).$$

If A and B are two independent events, then

$$P(A) = P(A / B) = P(A / \overline{B})$$
$$P(B) = P(B / A) = P(B / \overline{A}).$$

NOTES

3.7 SUBJECTIVE PROBABILITY

This is another way to interpret probabilities using personal evaluation of an event.

If the odds in favour of A are $a : b$ then the subjective probability is taken as $P(A) = \frac{a}{a+b}$.

If the odds against A are $a : b$ then the subjective probability is taken as $P(A) = \frac{b}{a+b}$.

These subjective probabilities may or may not satisfy the third axiom of probability.

Example 6. *Given* $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{4}$ *and* $P(AB) = \frac{1}{6}$.

Find the values of $P(A / B)$, $P(\overline{A}B)$, $P(\overline{A}\,\overline{B})$ *and* $P(\overline{A} + B)$.

Solution. $$P(A / B) = \frac{P(AB)}{P(B)} = \frac{1/6}{1/4} = \frac{2}{3}$$

$$P(\overline{A}B) = P(B) - P(AB) = \frac{1}{4} - \frac{1}{6} = \frac{1}{12}$$

$$P(\overline{A}\,\overline{B}) = 1 - P(A + B)$$
$$= 1 - [P(A) + P(B) - P(AB)]$$
$$= 1 - \left[\frac{1}{3} + \frac{1}{4} - \frac{1}{6}\right] = \frac{7}{12}.$$

$$P(\overline{A} + B) = P(\overline{A}) + P(B) - P(\overline{A}B)$$
$$= \left(1 - \frac{1}{3}\right) + \frac{1}{4} - \frac{1}{6} = \frac{3}{4}.$$

Example 7. *An investment consultant predicts that the odds against the price of a certain stock will go up during the next week are 2 : 1 and the odds in favour of the price remaining the same are 1 : 3. What is the probability that the price of the stock will go down during the next week ?*

Solution. Let E = Event that stock price will go up

F = Event that stock price will remain same.

Then $P(E) = \frac{1}{3}$ and $P(F) = \frac{1}{4}$

$$P(E \cup F) = \text{P (stock price will either go up or remains same)}$$

$$= P(E) + P(F) \quad = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

So, P(stock price will go down)

$$= P(\overline{E}.\overline{F})$$

$$= 1 - P(E+F) = 1 - \frac{7}{12} = \frac{5}{12}.$$

Example 8. *A and B throw alternately with a pair of dice. One who first throws a total of 9 wins. Show that the chances of their winning are 9 : 8.*

Solution. Let E_1 = Event of A throwing a total of 9 with a pair of dice,

E_2 = Event of B throwing a total of 9 with a pair of dice,

and these are independent events.

$$\therefore \quad P(E_1) = P(E_2) = \frac{4}{36} = \frac{1}{9} \text{ and } P(\overline{E}_1) \quad = P(\overline{E}_2) = \frac{8}{9}.$$

Assume that A starts the game. Winning of A can be given by the following mutually exclusive cases :

(*i*) E_1, (*ii*) $\overline{E}_1\overline{E}_2E_1$, (*iii*) $\overline{E}_1\overline{E}_2\overline{E}_1\overline{E}_2E_1$ and so on.

Using the theorem of addition,

$$\text{P [winning of A]} = P(E_1) + P(\overline{E}_1\overline{E}_2E_1) + P(\overline{E}_1\overline{E}_2\overline{E}_1\overline{E}_2E_1) + \ldots.$$

$$= P(E_1) + P(\overline{E}_1)P(\overline{E}_2)P(E_1) + P(\overline{E}_1)P(\overline{E}_2)P(\overline{E}_1)P(\overline{E}_2)P(E_1) + \ldots.$$

$$= \frac{1}{9} + \frac{8}{9}.\frac{8}{9}.\frac{1}{9} + \frac{8}{9}.\frac{8}{9}.\frac{8}{9}.\frac{8}{9}.\frac{1}{9} + \ldots.$$

$$= \frac{1}{9}\left[1 + \left(\frac{8}{9}\right)^2 + \left(\frac{8}{9}\right)^4 + \ldots.\right]$$

$$= \frac{\frac{1}{9}}{1 - \left(\frac{8}{9}\right)^2} = \frac{9}{17}.$$

$$\therefore \quad \text{P [winning of B]} = 1 - \frac{9}{17} = \frac{8}{17}$$

Hence the chances of their winning are 9 : 8.

3.8 RANDOM VARIABLE

A random variable X is a function whose domain is the sample space S and taking a value in the range set which is the real line with chance.

If the sample space consists of discrete elements then the r.v. is called **discrete r.v.**

NOTES

If the sample space consists of continuous elements then the r.v. is called **continuous r.v.**

The distribution given by the random variable is called probability distribution. Again on the type of the r.v., the probability distribution is called discrete distribution or continuous distribution.

Any discrete distribution is represented by probability mass function (*pmf*). For example,

x	-1	0	1
$p(x)$	0.2	0.4	0.4

is a discrete distribution. Here the random variable X only takes the values −1, 0 and 1 with probability 0.2, 0.4 and 0.4 respectively.

The characteristic of *pmf* is

(*i*) $p(x) \geq 0$ for all x

(*ii*) $\sum_{x} p(x) = 1$

Any continuous distribution is represented by probability density function (*pdf*). For example,

$$f(x) = 1, \quad 0 < x < 1$$
$$= 0, \quad \text{elsewhere}$$

is a continuous distribution. Here the random variable can take any value between 0 and 1 with probability = 1, for any other value the probability = 0.

The characteristic of *pdf* is

(*i*) $f(x) \geq 0$ for all x

(*ii*) $\int_{-\infty}^{\infty} f(x)dx = 1.$

3.9 CHARACTERISTICS OF PROBABILITY DISTRIBUTION

(*a*) **Distribution function.** For discrete case, the distribution function denoted by F(x) is defined as

$$F(x) = P[X \leq x]$$

For the above example, the distribution function is given by

x	-1	0	1
F(x)	0.2	0.6	1

For continuous case, the distribution function is defined as

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(x)\,dx$$

For the above example,

$$F(x) = \int_{-\infty}^{x} f(x)\, dx = \int_{-\infty}^{0} f(x)\, dx + \int_{0}^{x} f(x)\, dx$$

$$= 0 + \int_{0}^{x} 1.\, dx = x.$$

Properties:

(*i*) Discrete case

$$P\ (a < X < b) = F(b) - F(a)$$
$$P\ (a \leq X < b) = P[X = a] + [F(b) - F(a)]$$
$$P\ (a < X \leq b) = [F(b) - F(a)] - P[X = b]$$
$$P\ (a \leq X \leq b) = F(b) - F(a) + P\ [X = a] - P\ [X = b]$$

Continuous case

$$P\ (a < X < b) = P(a \leq X < b) = P\ (a < X \leq b) = P(a \leq X \leq b)$$
$$= F(b) - F(a).$$

(*ii*) $$F(-\infty) = \underset{x \to -\infty}{\text{Lt}}\ F(x) = 0$$

$$F(\infty) = \underset{x \to \infty}{\text{Lt}}\ F(x) = 1$$

(*iii*) $F(x) \leq F(y)$ whenever $x < y$.

(*iv*) $F(a) - F(a - 0) = P[x = a]$ and $F(a + 0) = F(a)$

(*v*) For continuous case, $F'(x) = f(x) \geq 0 \Rightarrow F(x)$ is nondecreasing function.

(*b*) **Mean/expectation.** Let X be the random variable. Then mean/expectation is defined as

$$\mu = E[X] = \sum_{x} x.p(x) \qquad \text{(Discrete case)}$$

$$= \int_{-\infty}^{\infty} x f(x) dx \qquad \text{(Continuous case)}$$

This expectation is sometimes called as 'Population Mean'.

Properties :

(*i*) $E[X + Y] = E[X] + E[Y]$

(*ii*) $E\ [cX] = c\ E[X]$

(*iii*) $E[c] = c$ and $E\ [X + c] = E[X] + c$

(*iv*) If X and Y are independent then $E(XY) = E(X).\ E(Y)$

(*v*) Physically, expectation represents the centre of mass of the probability distribution.

(*c*) **Variance.** Variance of a probability distribution is given by

$$\sigma^2 = V[X] = \sum_{x} (x - \mu)^2\ p(x) \qquad \text{(Discrete case)}$$

$$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\ dx \qquad \text{(Continuous case)}$$

Alt. $$\sigma^2 = E[(X - \mu)^2] = E[X^2] - \{E[X]\}^2.$$

$$= \sum_x x^2 . p(x) - \mu^2 \qquad \text{(Discrete case)}$$

$$= \int_{-\infty}^{\infty} x^2 . f(x) dx - \mu^2 \qquad \text{(Continuous case)}$$

Properties :

(*i*) $V[aX + b] = a^2 V[X]$

(*ii*) Physically variance represents the moment of inertia of the probability mass distribution about a line through the mean perpendicular to the line of the distribution.

Example 9. *For the following distribution*

x	1	2	3	4	5
$p(x)$	0.1	k	.2	$3k$	.3

(*i*) *Find the value of k.*

(*ii*) *Find the mean and variance.*

(*iii*) *Find the distribution function.*

Solution. (*i*) Since this is a *pmf*, we have

$$\Sigma p(x) = 1$$

$\Rightarrow \quad 0.1 + k + 0.2 + 3k + 0.3 = 1$

$\Rightarrow \quad 4k + 0.6 = 1$

$\Rightarrow \quad 4k = 0.4$

$\Rightarrow \quad k = 0.1.$

(*ii*) μ = Mean = $\Sigma x.p(x) = 1(0.1) + 2(0.1) + 3(0.2) + 4(0.3) + 5(0.3) = 3.6$

$$\sigma^2 = \text{Variance} = \Sigma(x-\mu)^2 p(x)$$

$$= (1 - 3.6)^2 (0.1) + (2 - 3.6)^2 (0.1) + (3 - 3.6)^2 (0.2) + (4 - 3.6)^2 (0.3) + (5 - 3.6)^2 (0.3)$$

$$= 1.64.$$

(*iii*) Distribution function is given as follows :

x	1	2	3	4	5
$p(x)$	0.1	0.2	0.4	0.7	1

Example 10. *Find the mean, variance and distribution function of the pdf*

$$f(x) = ax^2, 0 \leq x \leq 1$$
$$= 0, \quad \text{elsewhere.}$$

Solution. Since this is a *pdf*, then $\int_{-\infty}^{\infty} f(x)\, dx = 1$

$$\Rightarrow \qquad \int_0^1 ax^2\, dx = 1$$

$$\Rightarrow \qquad a\left[\frac{x^3}{3}\right]_0^1 = 1$$

$$\Rightarrow \qquad a = 3$$

So the *pdf* can be taken as

$$f(x) = 3x^2, 0 \le x \le 1$$
$$= 0, \quad \text{elsewhere}$$

$$\mu = \text{Mean} = \int_0^1 x.f(x)\,dx = 3\int_0^1 x^3\,dx = 3\left[\frac{x^4}{4}\right]_0^1 = \frac{3}{4}$$

$$\sigma^2 = \text{Variance} = \int_0^1 \left(x - \frac{3}{4}\right)^2 .3x^2\,dx$$

$$= \frac{3}{16}\int_0^1 [16x^4 - 24x^3 + 9x^2]\,dx$$

$$= \frac{3}{16}\left[\frac{16}{5} - \frac{24}{4} + \frac{9}{3}\right] = \frac{3}{80}$$

$$\text{Distribution function} = 3\int_0^x x^2 dx, \quad 0 \le x \le 1$$

$$= \begin{bmatrix} x^3 & , 0 \le x \le 1 \\ 0 & , \text{elsewhere} \end{bmatrix}.$$

***(d)* Moments and moment generating function.** *r*th moment about the mean is defined as

$$\mu_r = E[(X - \mu)^r] = \sum_x (x - \mu)^r . P[X = x] \qquad \text{(Discrete case)}$$

$$= \int_{-\infty}^{\infty} (x - \mu)^r . f(x)\,dx \qquad \text{(Continuous case)}$$

Here $\mu_0 = 1$ and $\mu_1 = 0$ for all random variables.

These are called central moments. If we take moments about any point '*a*', then the moments are called raw moments and is denoted by μ'_r and $\mu'_0 = 1$, $\mu'_1 = \mu$. The moment generating function (*m.g.f.*) about a point '*a*' is defined as

$$M_X(t) = \sum_x e^{t(x-a)}.p(x) \qquad \text{(Discrete case)}$$

$$= \int_{-\infty}^{\infty} e^{t(x-a)}.f(x)\,dx \qquad \text{(Continuous case)}$$

Now consider the discrete case,

NOTES

$$M_X(t) = \Sigma p_i \, e^{t(x_i - a)}, \text{ denoting } p(x) \text{ by } p_i$$

$$= \Sigma p_i \left[1 + t(x_i - a) + \frac{t^2}{2!}.(x_i - a)^2 + \ldots. \right]$$

$$= \Sigma p_i + t\,\Sigma p_i (x_i - a) + \frac{t^2}{2!} \Sigma p_i (x_i - a)^2 + \ldots.$$

$$= 1 + t\mu'_1 + \frac{t^2}{2!}\mu'_2 + \ldots.. + \frac{t^r}{r!}.\mu'_r + \ldots$$

Therefore, μ'_r = Coefficient of $\frac{t^r}{r!}$ in the expansion of $M_X(t)$.

Alternatively, we have

$$\mu'_r = \left[\frac{d^r}{dt^r}.M_X(t) \right]_{t=0}$$

The relations between the central moments and raw moments are as follows:

$$\mu_2 = \mu'_2 - \mu^2 = \text{Variance}$$
$$\mu_3 = \mu'_3 - 3\mu'_2.\mu + 2\mu^3$$
$$\mu_4 = \mu'_4 - 4\mu'_3.\mu + 6\mu'_2\,\mu^2 - 3\mu^4$$

(The moments obtained from a distribution (discrete/continuous) is called "Population Moments").

Note. 1. For *m.g.f.* if the point a is not given, it is taken as zero.

2. A r.v. X may have no moments although its *m.g.f.* exists.

e.g., $f(x) = \frac{1}{(x+1)(x+2)}$, $x = 0, 1, 2, \ldots.$ (the reader can verify).

3. A r.v. X may have moments although its *m.g.f.* fail to generate the moments.

Example 11. *Given the probability distribution, calculate the mean deviation.*

x	0	1	2	3	4
$p(x)$	0.1	0.3	0.4	0.1	0.1

Solution. Here $\mu = \Sigma x\,p(x)$

$$= 0 + 1(0.3) + 2(0.4) + 3(0.1) + 4(0.1)$$

$$= 0.3 + 0.8 + 0.3 + 0.4 = 1.8$$

Therefore, $\text{M.D.} = \sum_x |x - \mu|.p(x)$

$$= |0 - 1.8|.0 + |1 - 1.8|\,(0.3) + |2 - 1.8|\,(0.4) + |3 - 1.8|\,(0.1) + |4 - 1.8|\,(0.1)$$

$$= 0.66.$$

Example 12. *Find the quartiles of the following distribution :*

$$f(x) = 3x^2, 0 \le x \le 1.$$

Solution. For lower quartile, $\int_0^{Q_1} f(x)\,dx = \frac{1}{4}$

$\Rightarrow$ $3\int_0^{Q_1} x^2\,dx = \frac{1}{4}$

$\Rightarrow$ $\left[x^3\right]_0^{Q_1} = \frac{1}{4}$

$\Rightarrow$ $Q_1^3 = \frac{1}{4} \Rightarrow Q_1 = \left(\frac{1}{4}\right)^{1/3}$

For upper quartile, $3\int_{Q_3}^{1} x^2\,dx = \frac{1}{4}$

$\Rightarrow$ $\left[x^3\right]_{Q_3}^{1} = \frac{1}{4}$

$\Rightarrow$

$1 - Q_3^3 = \frac{1}{4} \Rightarrow Q_3^3 = \frac{3}{4} \Rightarrow Q_3 = \left(\frac{3}{4}\right)^{1/3}.$

3.10 BINOMIAL DISTRIBUTION

If a random variable X takes two values 1 and 0 with probability p and q respectively and $q = 1 - p$ then this is called Bernoulli distribution. Here p is called the probability of success and q is called the probability of failure.

For n trials, the probability of x successes ($x \le n$) is given by Binomial distribution. The probability mass function is defined as follows :

$$P[X = x] = \binom{n}{x} p^x . q^{n-x},\ x = 0,1,\ldots..,n$$

where, n = No. of independent trials

x = No. of successes

p = Probability of success on any given trial

$q = 1 - p$

$$\binom{n}{x} = {}^nC_x$$

Generally, it is denoted as B(n, p).

Properties

(*i*) $\sum_{x=0}^{n} P[X = x] = 1.$

(*ii*) Distribution function

$$F(x) = P[X \leq x] = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}.$$

NOTES

(*iii*) First two moments about origin.

$$\mu_1' = \sum_{x=0}^{n} x \binom{n}{x} p^x . q^{n-x}$$

$$= np . \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} . q^{n-x}$$

$$= np . (p + q)^{n-1} = np.$$

$$\mu_2' = \sum_{x=0}^{n} x^2 . \binom{n}{x} p^x . q^{n-x}$$

$$= \sum_{x=0}^{n} [x(x-1) + x] \binom{n}{x} p^x . q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1) . \binom{n}{x} p^x . q^{n-x} + np$$

$$= n(n-1) p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} . q^{n-x} + np$$

$$= n(n-1) p^2 . (q + p)^{n-2} + np$$

$$= n(n-1) p^2 + np.$$

Now, $\mu_1 = \mu'_1 = np$, which is the mean.

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$= n(n-1)p^2 + np - n^2p^2$$

$$= np - np^2$$

$$= np\ (1-p)$$

$= npq$, which is the variance.

Similarly we obtain μ_3 and μ_4.

(*iv*) Skewness : $\beta_1 = \dfrac{(1-2p)^2}{npq}$, $\gamma_1 = \dfrac{1-2p}{\sqrt{npq}}$

(*v*) Kurtosis : $\beta_2 = 3 + \dfrac{1-6pq}{npq}$, $\gamma_2 = \dfrac{1-6pq}{npq}$

(*vi*) Mode is the value of x for which $P[X = n]$ is maximum.

When $(n + 1)\ p$ is not an integer,

Mode = Integral part of $(n + 1)p$.

When $(n + 1)\ p$ is an integer, we obtain two modes

i.e., Mode = $(n + 1)\ p$ and $(n + 1)\ p - 1$.

Example 13. *Determine the binomial distribution for which the mean is 8 and variance 4 and find its mode.*

Solution. Given that $np = 8$ and $npq = 4$

On division, we get $q = \frac{1}{2} \Rightarrow p = 1 - q = \frac{1}{2}$

Also $n = \frac{8}{p} = 16$

Thus the given binomial distribution is B(16, ½).

Now. $(n + 1)\, p = (16 + 1)\, \frac{1}{2} = \frac{17}{2} = 8 + \frac{1}{2}$

which implies that mode = 8 (integral part only).

Example 14. *Fit a binomial distribution to the following distribution:*

x	0	1	2	3	4	5
f	27	14	6	3	0	0

Solution. Mean $= \frac{\Sigma xf}{\Sigma f} = \frac{35}{50}$, $n = 5$

Therefore, $np = \frac{35}{50} \Rightarrow p = \frac{35}{250} = 0.14$ and $q = 0.86$

The expected frequencies of the fitted binomial distribution can be calculated from

$$50\ (0.86 + 0.14)^5$$

Hence we obtain

x	0	1	2	3	4	5
Expected F	24	19	6	1	0	0

3.11 POISSON DISTRIBUTION

When (*i*) the number of trials is indefinitely large *i.e.*, $n \to \infty$,

(*ii*) constant probability of success for each trial is very small *i.e.*, $p \to 0$,

and (*iii*) $n\bar{p} = \lambda$ a finite value.

Poisson distribution is obtained as a limiting case of binomial distribution.

The probability mass function is

$$P[X = x] = \frac{e^{-\lambda}.\lambda^x}{x!},\ x = 0,1,2,.....$$

where λ is called the parameter of this distribution.

Proof. Let $np = \lambda \Rightarrow p = \frac{\lambda}{n},\ q = 1 - \frac{\lambda}{n}$.

$$\therefore \binom{n}{x} p^x . q^{n-x} \approx \frac{n!}{n - x!\,.\,x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}.$$

when $n \to \infty$

$$\frac{n!}{(n-x)!\,.\,x!} = \frac{(n-k+1)\,(n-k+2)\ldots n}{n^k} \to 1$$

NOTES

Also, $$ln\left(1 - \frac{\lambda}{n}\right)^{n-x} \simeq (n-x)\left(-\frac{\lambda}{n}\right) \to -\lambda$$

Therefore, $$B\ (n,\ p) \to \frac{e^{-\lambda}.\ \lambda^x}{x!}$$

Properties

(*i*) $$\sum_{x=0}^{\infty} P[X=x] = 1$$

(*ii*) Distribution function

$$F(x) = P[X \leq x] = e^{-\lambda}.\sum_{k=0}^{x} \frac{\lambda^k}{k!}, \qquad x = 0,1,2,....$$

(*iii*) First two moments about origin.

$$\mu'_1 = \sum_{x=0}^{\infty} x\,.\,P[X=x] = \sum_{x=0}^{\infty} x\,.\,\frac{e^{-\lambda}.\lambda^x}{x!} = \lambda e^{-\lambda}.\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{x-1!}$$

$$= \lambda\,.\,e^{-\lambda}.\,e^{\lambda} = \lambda.$$

$$\mu'_2 = \sum_{x=0}^{\infty} x^2.\ P[X = x]$$

$$= \sum_{x=0}^{\infty} [x(x-1) + x]\,.\,\frac{e^{-\lambda}.\ \lambda^x}{x!}$$

$$= \left[\sum_{x=0}^{\infty} x(x-1)\,.\,\frac{e^{-\lambda}.\ \lambda^x}{x!}\right] + \lambda$$

$$= e^{-\lambda}.\ \lambda^2\,.\sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{x-2!} + \lambda$$

$$= e^{-\lambda}.\ \lambda^2.\ e^{\lambda} + \lambda = \lambda^2 + \lambda$$

∴ $\mu_1 = \mu'_1 = \lambda$ which is mean.

$\mu_2 = \mu'_2 - (\mu'_1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda,$ which is variance.

Similarly we obtain μ_3 and μ_4.

(*iv*) Skewness : $\beta_1 = \frac{1}{\lambda}$, $\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}$.

(*v*) Kurtosis : $\beta_2 = 3 + \frac{1}{\lambda}$, $\gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}$.

(*vi*) Mode :

When λ is not an integer,

$$\text{Mode} = \text{integral part of } \lambda.$$

When λ is an integer,

$$\text{Mode} = \lambda - 1 \text{ and } \lambda.$$

(*vii*) In queueing theory (to be discussed in part B), it can be shown under certain assumptions that the probability of arriving x customers in time t is

$$P_x(t) = \frac{e^{-\lambda t}.\,(\lambda t)^x}{x!}, \quad x = 0,1,2,....$$

which is a Poisson distribution with parameter λt. This is also called Poisson process *i.e.*, the number of customers generated (arriving) until any specific time has a Poisson distribution.

Example 15. *There are 150 misprints in a book of 520 pages. What is the probability that a given page will contain at most 2 misprints ?*

Solution. Here $\lambda = \frac{150}{520}$ and let the misprints follow Poisson distribution.

$$\begin{aligned}\text{Required probability} &= P[X \leq 2] \\ &= P[X = 0] + P[X = 1] + P[X = 2] \\ &= e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2}{2}.\,e^{-\lambda}\end{aligned}$$

$$= e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right) = e^{-15/52}\left(1 + \frac{15}{52} + \frac{1}{2}\left(\frac{15}{52}\right)^2\right)$$

$$= 0.9968.$$

Example 16. *A controlled manufacturing process is 0.2% defective. What is the probability of taking 2 or more defectives from a lot of 100 pieces ? (a) By using binomial distribution. (b) By using Poisson approximation.*

Solution. (*a*)

$$\begin{aligned}p &= \text{Probability of defective} = 0.002, \\ q &= 1 - p = 0.998 \\ n &= 100\end{aligned}$$

$\therefore$ Probability of finding 2 or more defective

$$\begin{aligned}&= 1 - [\text{Probability of zero and one defective}] \\ &= 1 - [P[X = 0] + P[X = 1]] \\ &= 1 - \left[(0.998)^{100} + 100\,(0.002)\,(0.998)^{99}\right] \\ &= 1 - 0.983 = 0.017.\end{aligned}$$

(*b*) Here $\lambda = np = 100 \times 0.002 = 0.2$

$\therefore$ Probability of finding 2 or more defective

$$\begin{aligned}&= 1 - [P[X = 0] + P[X = 1]] \\ &= 1 - \left[e^{-0.2} + (0.2)\,e^{-0.2}\right] \\ &= 1 - 0.982 = 0.018.\end{aligned}$$

3.12 NORMAL DISTRIBUTION

In the binomial distribution if (*i*) the number of trials is indefinitely large, *i.e.*, $n \to \infty$ and (*ii*) neither p nor q is very small, then the limiting form of the binomial distribution is called 'normal distribution'. It is a continuous distribution and probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-(x-\mu)^2/2\sigma^2}$$

or,

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot \text{exp.}\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

$-\infty < x < \infty, -\infty < \mu < \infty.$

It is denoted by N (μ, σ^2) where μ = Mean and σ^2 = Variance.

Since this distribution was developed by Carl Friedrich Gauss, sometimes it is also referred as 'Gaussian Distribution'.

Properties

(*i*) The normal curve is bell shaped and symmetrical about the line $x = \mu$.

(*ii*) Median : Let M be the median, then

$$\int_{-\infty}^{M} f(x)\, dx = \frac{1}{2}$$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{M} \text{exp.}\left[-(x-\mu)^2/2\sigma^2\right] dx = \frac{1}{2}$$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\mu} \text{exp.}\left[-(x-\mu)^2/2\sigma^2\right] dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\mu}^{M} \text{exp.}\left[-(x-\mu)^2/2\sigma^2\right] dx = \frac{1}{2}$$

In the first integral let $z = \frac{x-\mu}{\sigma}$, then we get

$$\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \text{exp.}\left[-z^2/2\right] dz = \frac{1}{2}$$

Therefore, $\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} \text{exp.}\left[-(x-\mu)^2/2\sigma^2\right] dx = \frac{1}{2}$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} \text{exp.}\left[-(x-\mu)^2/2\sigma^2\right] dx = 0$$

$$\Rightarrow \mu = \text{M}$$

(*iii*) Mode: It is the value of x for which $f(x)$ is maximum *i.e.*, $f'(x) = 0$ and $f''(x) < 0$.

Here we obtain, mode = μ.

Note. Mean, Median and Mode coincides.

(*iv*) Mean deviation (M.D.)

$$\text{M.D.} = \int_{-\infty}^{\infty} |x - \mu| \, . \, f(x) \, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-\mu| \, . \, \exp. \left[-(x - \mu)^2 / 2\sigma^2\right] dx$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| \exp. \left[-z^2/2\right] dz, \quad \text{taking } z = \frac{x-\mu}{\sigma}$$

$$= \sqrt{\frac{2}{\pi}} \, . \, \sigma \, . \int_{0}^{\infty} z \, . \, \exp. \left[-z^2/2\right] dz, \quad (\because \text{ the integrand is even})$$

$$= \sqrt{\frac{2}{\pi}} \, . \, \sigma \, . \int_{0}^{\infty} e^{-t} \, dt, \quad \text{taking } \frac{z^2}{2} = t$$

$$= \sqrt{\frac{2}{\pi}} \, . \, \sigma \, . \, 1 = \sqrt{\frac{2}{\pi}} \, . \sigma \approx \frac{4}{5} \sigma .$$

(*v*) $\text{Q.D : M.D : S.D.} = \frac{2}{3} : \frac{4}{5} : 1$

i.e., 10 : 12 : 15.

(*vi*) Points of inflexion of the normal curve is given at $x = \mu \pm \sigma$.

(*vii*) Central moments : $\mu_{2n+1} = 0$ (odd-order)

and $\mu_{2n} = 1, 3, 5, ..., (2n-1) \, \sigma^{2n}$. (even order).

(*viii*) Skewness : $\beta_1 = 0 \; (\because \mu_3 = 0), \;\; \gamma_1 = 0.$

(*ix*) Kurtosis : $\beta_2 = 3, \qquad \gamma_2 = \beta_2 - 3 = 0.$

(*x*) Distribution functionF(x) $= \int_{-\infty}^{x} f(x) \, dx.$

(*xi*) It is a *pdf* $\int_{-\infty}^{\infty} f(x) \, dx = 1.$

(*xii*) Standard Normal Distribution

If X ~ N(μ, σ^2), then $z = \frac{X - \mu}{\sigma}$ is called standard normal variate with E[z] = 0 and var [z] = 1 and z ~ N(0,1).

$$\phi\,(z) = \text{the } pdf = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty.$$

(*xiii*) $$P[x_1 < X < x_2] = \int_{x_1}^{x_2} f(x) \, dx$$

$$= \int_{0}^{x_2} f(x) \, dx - \int_{0}^{x_1} f(x) \, dx$$

$$= \int_0^{z_2} \phi(z)\, dz - \int_0^{z_1} \phi(z)\, dz \quad \text{taking } z = \frac{x - \mu}{\sigma}$$

$$= \Phi(z_2) - \Phi(z_1)$$

NOTES

These values are obtained from the standard normal table.

(*xiv*) Area under the normal curve :

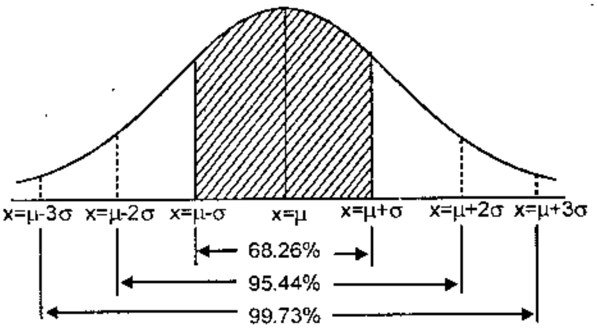


Fig. 3.2

Probable Error

Any manufactured items or measurement of any physical quantity shows slight error. All the errors in manufacturing or measurement are random in nature and follow a normal distribution. We define the probable error λ is such that the probability of an error falling within the limits $\mu - \lambda$ and $\mu + \lambda$ is exactly equal to the chance of an error falling outside these limits which implies that the chance of an error lying within $\mu - \lambda$ and $\mu + \lambda$ is ½.

$$\Rightarrow \qquad \int_{\mu-\lambda}^{\mu+\lambda} f(x)\, dx = \frac{1}{2}, \qquad f(x) = \text{normal } pdf.$$

$$\Rightarrow \qquad \int_0^{\lambda/\sigma} \phi(z)\, dz = \frac{1}{4} \qquad \left(\text{taking } z = \frac{x - \mu}{\sigma}\right)$$

$$\Rightarrow \qquad \frac{\lambda}{\sigma} = 0.6745 \qquad \text{(from normal table)}$$

$$\Rightarrow \qquad \lambda = 0.6745\ \sigma \sim \frac{2}{3}\sigma.$$

Normal Approximation To Binomial Distribution

If the number of trials is sufficiently large (*i.e.*, $n \geq 30$) the binomial distribution B(n, p) is approximated by the normal distribution N(μ, σ^2) with $\mu = np$ and $\sigma^2 = npq$. But a continuity correction is required. The discrete integer x in B(n, p) becomes the interval [x – 0.5, x + 0.5] in the N(μ, σ^2). Thus

$$P[X = x] \approx \frac{1}{\sigma\sqrt{2\pi}} \int_{x-0.5}^{x+0.5} e^{-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2} du$$

and $$P[x_1 < x < x_2] \approx \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1 - 0.5}^{x_2 + 0.5} e^{-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2} du$$

Example 17. *Given a random variable having the normal distribution with $\mu = 18.2$ and $\sigma = 1.25$, find the probabilities that it will take on a value*

(a) less than 16.5,

(b) greater than 18.8,

(c) between 16.5 and 18.8,

(d) between 19.2 and 20.1.

Solution. Let $z = \dfrac{x - 18.2}{1.25}$

(*a*) (Fig. 3.3) $z = \dfrac{16.5 - 18.2}{1.25} = -1.36$

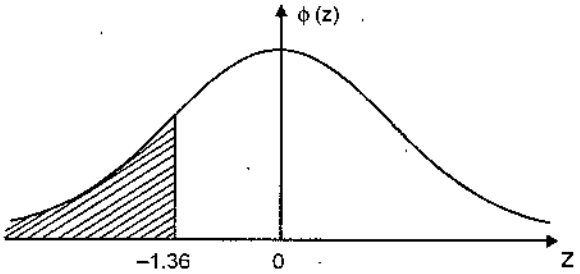


Fig. 3.3

$$\begin{aligned} P(z < -1.36) &= 0.5 - \Phi(1.36) \\ &= 0.5 - 0.4131 \\ &= 0.0869. \end{aligned}$$

(*b*) (Fig. 3.4) $z = \dfrac{18.8 - 18.2}{1.25} = 0.48$

$$\begin{aligned} P(z > 0.48) &= 0.5 - \Phi(0.48) \\ &= 0.5 - 0.1844 \\ &= 0.3156. \end{aligned}$$

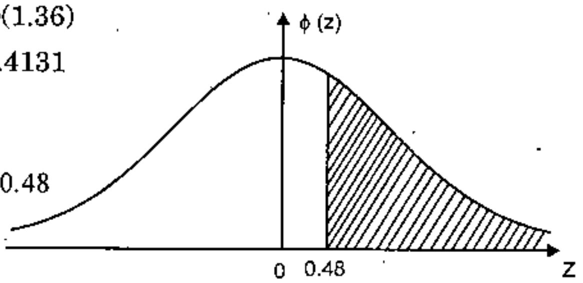


Fig. 3.4

(*c*) (Fig. 3.5) $$\begin{aligned} P(-1.36 < z < 0.48) &= \Phi(1.36) + \Phi(0.48) \\ &= 0.4131 + 0.1844 \\ &= 0.5975. \end{aligned}$$

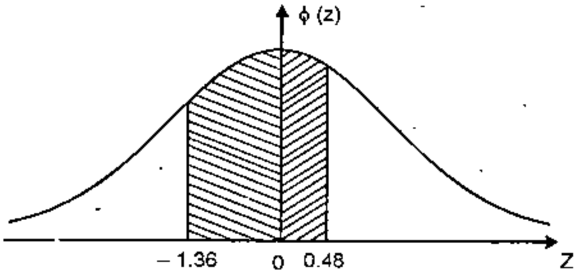


Fig. 3.5

(*d*) (Fig.3.6) When $x = 19.2$, $z_1 = 0.8$

When $x = 20.1$, $z_2 = 1.52$

Here $P(0.8 < z < 1.52) = \Phi(1.52) - \Phi(0.8)$

$= 0.4357 - 0.2881 = 0.1476.$

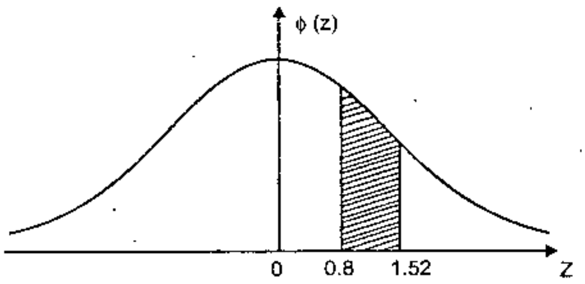


Fig. 3.6

Example 18. *Small stones are collected and weights are assumed to be normal. It is found that 5% of the stones are under 30 gm and 80% are under 50 gm. What are the mean and standard deviation of the distribution ?*

Solution. Let weight be represented by a random variable X such that $X \sim N(\mu, \sigma^2)$.

Given that, $P(X < 30) = 0.05$ and

$P(X < 50) = 0.8$ *i.e.,* $P(X > 50) = 0.2$

We obtain the Fig. 3.7.

While converting 30 and 50 into z_1 and z_2 using $z = \frac{x - \mu}{\sigma}$ the value z_1 will be in the negative side.

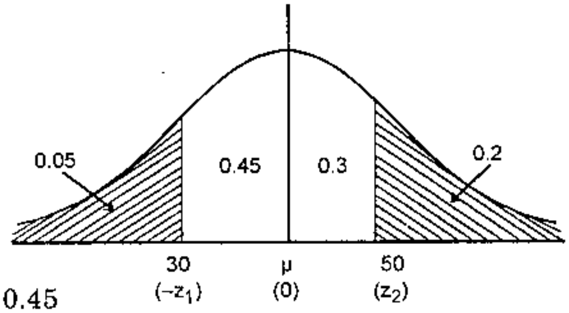


Fig. 3.7

Therefore $\Phi(-z_1) = 0.45$

$$\Rightarrow \quad \frac{30 - \mu}{\sigma} = -1.65$$

$$\Rightarrow \quad \mu - 1.65\sigma = 30 \quad \ldots(i)$$

and $\Phi(z_2) = 0.3$

$$\Rightarrow \quad \frac{50 - \mu}{\sigma} = 0.85$$

$$\Rightarrow \quad \mu + 0.85\sigma = 50 \quad \ldots(ii)$$

Solving (*i*) and (*ii*) we obtain $\mu = 43.2$ and $\sigma = 8$.

SUMMARY

- The single performance of a random experiment will be called an outcome.
- Axioms of Mathematical Probability

  Let A be an event in a sample space S, then:

  (*i*) $0 \leq P(A) \leq 1$

  [Here P(A) = 0 means that the event will not occur and P(A) = 1 means that the event is certain] (*ii*) P(S) = 1

(*iii*) If A and B are two mutually exclusive events then

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

- If two events A and B are mutually exclusive, then the occurrence of either A *or* B is given by

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

- When the two events A and B are not mutually exclusive, then the probability of occurrence of at least one of the 2 events is given by

$$P(A+B) = P(A) + P(B) - P(AB)$$

- Let A and B be two events in a sample space S and $P(A) \neq 0$, $P(B) \neq 0$, then the probability of happening of both the events are given by

  (*i*) P(AB) = P(A).P(B/A)

  (*ii*) P(AB) = P(B).P(A/B).

- Two events A and B in a sample space S are said to be independent if P(AB) = P(A).P(B).

- If A and B are two independent events, then

$$P(A) = P(A/B) = P(A/\overline{B})$$
$$P(B) = P(B/A) = P(B/\overline{A}).$$

- A random variable X is a function whose domain is the sample space S and taking a value in the range set which is the real line with chance.

- The probability that a specified magnitude (K) will be exceeded *i.e.*, $P[X > K] = p_0$ is called 'Exceedance Probability'.

- Median Exceedance Probability : In a sample of estimates of exceedance probability of a specified magnitude, this is the value that is exceeded by 50 percent of the estimates.

PROBLEMS

1. Let A, B and C be three mutually and exhaustive events. Find P(B), if $\frac{1}{3}.P(C) = \frac{1}{2}P(A) = P(B)$.

2. If P(A) = 0.50, P(B) = 0.40 and $P(A \cup B) = 0.70$, find P(A/B) and $P(\overline{A} \cup B)$. State whether A and B are independent.

3. Given $P(A) = \frac{1}{4}$, $P(A/B) = \frac{1}{4}$ and $P(B/A) = \frac{1}{2}$, find if

   (*i*) A and B are mutually exclusive, (*ii*) A and B are independent.

4. Out of the numbers 1 to 100, one is selected at random. What is the probability that it is divisible by 7 or 8 ?

5. A committee of 4 people is to be appointed from 3 offices of the production department, 4 officers of the purchase department, 2 officers of the sales department and 1 chartered accountant. Find the probability of forming the committee in the following manner.

NOTES

(*i*) There must be one from each category.

(*ii*) The committee should have at least one from the purchase department.

(*iii*) The chartered accountant must be in the committee.

6. What is the probability that a leap year selected at random will contain either 53 Thursdays or 53 Fridays ?

7. X can solve 80% of the problems while Y can solve 90% of the problems given in a Statistic book. A problem is selected at random. What is the probability that at least one of them will solve the same ?

8. A box P has 1000 items of which 100 are defective. Another box Q has 500 items of which 20 are defective. The items of both the boxes are mixed and one item is randomly taken out. It is found to be defective. What is the probability that the item belongs to box P ?

9. Villages A, B, C and D are connected by overhead telephone lines joining AB, AC, BC, BD and CD. As a result of severe gales, there is a probability p (the same for each link) that any particular link is broken. Find the probability of making a telephone call from A to B.

10. A man seeks advise regarding one of two possible courses of action from three advisers, who arrive at their recommendations independently. He follows the recommendation of the majority. The probabilities that the individual advisers are wrong are 0.1, 0.05 and 0.05 respectively. What is the probability that the man takes incorrect advise ?

11. An unbiased coin is tossed four times in succession and a man scores 2 or 1 according to the coin as shows head or tail in each throw. E_1, E_2 are the following events : E_1 : total score is even, E_2 : total score is divisible by 3. Determine whether E_1 and E_2 are independent or not.

12. There are two identical boxes containing respectively 4 white and 3 red balls and 3 white and 7 red balls. A box is chosen at random and a ball is drawn from it. Find the probability that the ball is white. If the ball is white, what is the probability that it is from the first box?

13. A speaks truth 4 out of 5 times. A die is tossed. He reports that there is a six. What is the chance that actually there was six ?

14. If a machine is set correctly it produces 10% defective items. If it is set incorrectly then it produces 10% good items. Chances for a setting to be correct and incorrect are in the ratio 7 : 3. After a setting is made, the first two items produced are found to be good items. What is the chance that the setting was correct ?

15. A random variable X has the following probability distribution :

x	-1	0	1	2	3
$p(x)$	0.2	0.1	k	$2k$	0.1

(*a*) Find the value of k.

(*b*) Calculate the mean and variance.

(*c*) Calculate the mean deviation.

(*d*) Find $P(1 \leq x \leq 3)$, $P(x > 1)$.

16. Given the probability distribution

x	1	2	3	4	5	6
$p(x)$	a	$2a$	a	a	$3a$	$2a$

(*i*) Find the value of a.

(*ii*) Find the distribution function.

(*iii*) Find the mean and variance.

(*iv*) Find $P(X < 4)$, $P(2 < X < 5)$.

17. A random variable X has the probability distribution :

x	-2	-1	0	1	2
$p(x)$	0.2	0.1	0.1	0.3	0.3

Calculate the first four central moments.

NOTES

18. Consider the following probability distribution :

$$f(x) = kx, \qquad 0 < x < 1$$

(*a*) Find the value of k.

(*b*) Find the mean and variance.

(*c*) Find the median, Q_1 and Q_3.

19. The Maxwell-Boltzmann distribution is given by

$$f(x) = 4a.\sqrt{\frac{a}{\pi}}.x^2 e^{-ax^2}, \qquad 0 \le x < \infty, a > 0$$

where, $a = \frac{m}{2kT}$, m = Mass, T = Temperature (K), k = Boltzmann constant and x = Speed of a gas molecule.

(*a*) Verify that this is a *pdf.*

(*b*) Calculate the mean and variance.

20. Find the *m.g.f.* of the following distributions :

(*i*) $f(x) = \frac{1}{b-a}$, $a \le x \le b$

$= 0,$ elsewhere.

(*ii*) $p(x) = \frac{e^{-\lambda}.\lambda^x}{x!}$, $x = 0,1,2,....$

Hence find the mean and variance in (*ii*).

21. In a shipment of 7 items there are three defective items. At random 4 items are selected. What is the expected number of defective items ?

22. The diameter of an electric cable is assumed to be a continuous random variable X with *pdf* $f(x) = 6x(1-x)$, $0 \le x \le 1$. Determine b such that $P[X < b] = P[X > b]$.

23. Five prizes are to be distributed among 20 students. Find the probability that a particular student will receive three prizes.

24. Consider an intersection approach in which studies have shown 25% right turns and no left turns. Find the probability of one out of the next four vehicles turning right.

25. The mean and variance of a binomial distribution are 6 and 2 respectively. Find $P[X > 1]$, $P[X = 2]$.

26. In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 and 0.2048. Find the parameter p of the distribution.

27. An experiment succeeds twice as often as it fails. What is the probability that in next five trials there will be (*i*) three successes, (*ii*) at least three successes ?

28. A quality control engineer inspects a random sample of 3 calculators from each lot of 20 calculators. If such a lot contains 4 slight defective calculators. What are the probabilities that the inspector's sample will contain

(*i*) no slight defective calculators,

(*ii*) one slight defective calculators,

(*iii*) at least two slight defective calculators.

NOTES

29. Fit a binomial distribution to the following distribution :

x	0	1	2	3	4	5
f	3	12	21	30	25	9

30. Fit a binomial distribution to the following data:

x	0	1	2	3	4
f	15	12	10	8	5

31. How many tosses of a coin are needed so that the probability of getting at least one head is 87.5% ?

32. A machine produces an average of 20% defective bolts. A batch is accepted if a sample of 5 bolts taken from that batch contains no defective and rejected if the sample contains 3 or more defectives. In other cases, a second sample is taken. What is the probability that the second sample is required ?

33. If the probability of a defective bolt is 0.1, find (*i*) mean, (*ii*) variance, (*iii*) moment coefficient of skewness and, (*iv*) Kurtosis for the distribution of defective bolts in a total of 400.

34. If on an average 1 vessel in every 10 is wrecked, find the probability that out of 5 vessels expected to arrive, at least 4 will arrive safely.

35. If X be a Poisson distributed random variable and P[X = 1] = 3P[X = 2], then find P[X > 2].

36. A medicine was supplied in 100 batches (each batch containing a fairly large number of items). A total of 50 items in all the batches were found to be defective. Find the probability that (*a*) a batch has no defective item, (*b*) a batch has at least three defective items.

37. If 2 per cent of electric bulbs manufactured by a certain company are defective, find the probability that in a sample of 200 bulbs (*a*) less than 2 bulbs are defective, (*b*) more than 3 bulbs are defective.

38. Let X follows Poisson distribution, find the value of the mean of the distribution of P[X = 1] = 3P[X = 2].

39. In a certain factory, blades are manufactured in packets of 10. There is a 0.1% probability for any blade to be defective. Using Poisson distribution calculate approximately the number of packets containing two defective blades in a consignment of 10000 packets.

40. Fit a Poisson distribution to the following:

x	0	1	2	3	4	5
f	20	16	11	7	4	2

41. Fit a Poisson distribution to the following:

x	0	1	2	3	4	5
f	120	82	52	22	4	0

42. Records show that the probability is 0.00002 that a car will have a flat tyre while driving over a certain bridge. Use the Poisson distribution to determine the probability that among 20000 cars driven over this bridge, not more than one will have a flat tyre.

43. A typist kept a record of mistakes made per day during 300 working days of a year :

Mistakes per day	0	1	2	3	4	5	6
No. of days	143	90	42	12	9	3	1

Fit an appropriate Poisson distribution to the data.

44. The probability that a Poisson variate X takes a positive value is $\left(1 - e^{-1.5}\right)$. Find the variance and also the probability that X lies between –1.5 and 1.5.

45. A manufacturer, who produces medicine bottles, finds that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys 100 boxes from the producer of bottles. Using Poisson distribution, find how many boxes will contain (*i*) no defectives, (*ii*) at least two defectives.

46. What is the probability that in a company of 1000 people only one person will have birth day on new year's day? (Assume that a year has 365 days).

47. The diameters of shafts manufactured by a machine are normal random variables with mean 10 cm and standard deviation 1 cm. Show that 81.85 percent of the shafts are expected to these diameters between 9 cm and 12 cm.

48. Tests have indicated that the tensile strengths of certain alloys averages 1885 kg/cm^2 with a standard deviation of 225 kg/cm^2. If the distribution is normal, what percentage of the casting will have tensile strength (*i*) less than 1500 kg/cm^2, (*ii*) more than 1600 kg/cm^2, (*iii*) between 2000 and 2100 kg/cm^2.

49. Assuming that the life in hours of an electric bulb is a random variable following normal distribution with mean of 2000 hours and standard deviation of 500 hours. Find the expected number of bulbs from a sample of 2000 bulbs having life (*i*) more than 2500 hours, (*ii*) between 2600 and 3000 hours.

50. The mean value of the modulus of rupture of a large number of test specimen has been found to be 400 kg/cm^2. If the standard deviation is 70 kg/cm^2 and the distribution is approximately normal, for what percentage of the specimens the modulus of rupture will fall (*i*) between 350 and 450 ? (*ii*) above 300 ?

51. A manufacturer makes chips for use in a mobile phone of which 10% are defective. For a random sample of 200 chips, find the approximate probability that more than 15 are defective.

52. E-mail messages are received by a general manager of a company at an average rate of 1 per hour. Find the probability that in a day the manager receives 24 messages or more.

53. A normal curve has an average of 150.2 and a standard deviation of 3.81. What percentage of the area under the curve will fall between limits of 137.5 and 153.4 ?

54. Suppose that life of a gas cylinder is normally distributed with a mean of 40 days and a S.D. of 5 days. If, at a time, 10,000 cylinders are issued to customers, how many will need replacement after 35 days ?

55. Fit a normal curve to the following :

x	0	1	2	3	4
f	9	21	42	20	8

56. Fit a normal curve to the following :

x	–2	–1	0	1	2	3
f	6	9	24	15	11	5

57. In a distribution exactly normal, 10.03% of the items are under 25 kilogram weight and 97% of the items are under 70 kilogram weight. What are the mean and standard deviation of the distribution ?

58. If the probability of committing an error follows normal distribution, compute the probable error from the following data :

    4.8, 4.2, 5.1, 3.8, 4.4, 4.7, 4.1 and 4.5.

59. The width of a slot on a forging is normally distributed with mean 0.9" and standard deviation 0.900" ± 0.005". What percentage of forgings will be defective ?

60. A sales tax officer has reported that the average sales of the 500 business that he has to deal with during a year is Rs. 36,000 with a standard deviation of 10000. Assuming that the sales in these business are normally distributed, find

    (*i*) the number of business as the sales of while are Rs. 40,000

    (*ii*) the percentage of business the sales of while are likely to range between Rs. 30,000 and Rs. 40,000.

ANSWERS

NOTES

1. $P(B) = \frac{2}{7}$ **2.** $P(A/B) = \frac{1}{2}$, $P(\overline{A} \cup B) = 0.7$, Independent

3. (*i*) Not mutually exclusive, (*ii*) Independent **4.** $\frac{1}{4}$

5. (*i*) $\frac{8}{70}$, (*ii*) $\frac{13}{14}$, (*iii*) $\frac{2}{5}$ **6.** $\frac{3}{7}$

7. $\frac{49}{50}$ **8.** $\frac{5}{6}$

9. $1 - 2p^2 + p^3$ **10.** 0.01175 **11.** Not independent

12. $\frac{61}{140}, \frac{40}{61}$ (Use Baye's theorem)

13. $\frac{4}{9}$ (Use Baye's theorem) **14.** 0.99 (Use Baye's theorem)

15. (*a*) 0.2 (*b*) 1.1, 1.69 (*c*) 1.1 (*d*) 0.7, 0.5.

16. (*i*) $a = 0.1$ (*ii*)

x	1	2	3	4	5	6
Cum. P(x)	.1	.3	.4	.5	.8	1

(*iii*) 3.9, 2.89 (*iv*) 0.4, 0.2

17. $\mu_1 = 0, \mu_2 = 2.24, \mu_3 = -1.75, \mu_4 = 9.03$

18. (*a*) 2, (*b*) $\frac{2}{3}, \frac{1}{18}$, (*c*) $\frac{1}{\sqrt{2}}, \frac{1}{2}, \frac{\sqrt{3}}{2}$

19. (*b*) Mean $= \frac{2}{\sqrt{a\pi}}$, Variance $= \frac{3}{2a} - \frac{4}{a\pi}$

20. (*i*) $\frac{e^{bt} - e^{at}}{t(b-a)}$, (*ii*) $e^{-\lambda}.e^{(\lambda e^t)}$ **21.** 12/7 **22.** $b = \frac{1}{2}$.

23. 0.088 **24.** 0.56 **25.** 0.999, 0.007

26. $p = \frac{1}{5}$ **27.** (*i*) $\frac{80}{243}$, (*ii*) $\frac{192}{243}$

28. (*i*) $\frac{64}{125}$, (*ii*) $\frac{48}{125}$, (*iii*) $\frac{13}{125}$

29.

x	0	1	2	3	4	5
EF	1	9	25	34	24	7

30.

x	0	1	2	3	4
EF	7	18	17	7	1

31. 3 tosses are required **32.** 0.6144

33. (*i*) 40, (*ii*) 36, (*iii*) $\frac{2}{15}$, (*iv*) $\frac{23}{1800}$ **34.** $\frac{45927}{50000}$

35. 0.0302 **36.** (*a*) 0.61 (*b*) 0.01

37. (*a*) 0.09 (*b*) 0.58 **38.** $\frac{2}{3}$ **39.** Two packets

40.

x	0	1	2	3	4	5
EF	15	20	15	7	2	1

(The EF corresponding to x = 1, has been adjusted to make Σ EF = 60)

41.

x	0	1	2	3	4	5
EF	108	102	49	16	4	1

42. 0.9380

43.

x	0	1	2	3	4	5	6
EF	123	110	49	14	3	1	0

44. Variance = 1.5, $P(-1.5 < X < 1.5) = 2.5\ e^{-1.5}$

45. (*i*) 61, (*ii*) 9 **46.** 0.18.

48. (*i*) 4.36%, (*ii*) 89.8%, (*iii*) 13.64%

49. (*i*) 317, (*ii*) 185 **50.** (*i*) 52.22%, (*ii*) 92.36%

51. 0.8554 **52.** 0.5398 **53.** 79.91% **54.** 8413

55.

x	0	1	2	3	4
EF	7	25	37	24	7

56.

x	−2	−1	0	1	2	3
EF	4	12	20	19	11	4

57. $\mu = 47.5$ and $\sigma = 15.578$ **58.** 0.26 **59.** 21.12%

60. (*i*) 78, (*ii*) 38.12%.

FURTHER READINGS

1. Statistics and Numerical Methods: Dr. Manish Goyal.
2. Golden Statistics: N.P. Bali.

UNIT 4 LINEAR PROGRAMMING

NOTES

★ STRUCTURE ★

4.0 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- describe about operations research
- explain linear programming problems
- illustrate graphical and simplex method.

4.1 INTRODUCTION TO OPERATIONS RESEARCH

The roots of Operations Research (O.R.) can be traced many decades ago. First this term was coined by Mc Closky and Trefthen of *United Kingdom in 1940 and* it came in existence during world war II when the allocations of scarce resources were done to the various military operations. Since then the field has developed very rapidly. Some chronological events are listed below :

1952 – Operations Research Society of America (ORSA).
1957 – Operations Research Society of India (ORSI)
– International Federation of O.R. Societies
1959 – *First Conference of ORSI*
1963 – Opsearch (the journal of O.R. by ORSI).

However the term 'Operations Research' has a number of different meaning. The Operational Research Society of Great Britain has adopted the following illàborate definition :

"Operational Research is the application of the methods of science to complex problems arising in the direction and management of large systems of men, machines, materials and money in industry, business, government and defence. The distinctive approach is to develop a scientific method of the system, incorporating measurements of factors such as chance and risk, with which to predict and compare the outcomes of alternative decisions, strategies and controls. The purpose is to help management to determine its policy and actions scientifically."

Whereas ORSA has offered the following shorter definition :

"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".

Many individuals have described O.R. according to their own view. Only three are quoted below :

"O.R. is the art of giving bad answers to problems which otherwise have worse answers" —*T.L. Saaty*

"O.R. is a scientific approach to problems solving for executive management" —*H.M. Wagner*

"O.R. is a scientific knowledge through interdisciplinary team effort for the purpose of determining the best utilization of limited resources". —*H.A. Taha*

An abbreviated *list of applications* of O.R. techniques are given below :

1.	Manufacturing	:	Production scheduling Inventory control Product mix Replacement policies
2.	Marketing	:	Advertising budget allocation Supply chain management
3.	Organizational behaviour	:	Personnel planning Scheduling of training programs Recruitment policies
4.	Facility planning	:	Factory location Hospital planning Telecommunication network planning Warehouse location
5.	Finance	:	Investment analysis Portfolio analysis
6.	Construction	:	Allocation of resources to projects Project scheduling
7.	Military		
8.	Different fields of engineering.		

4.2 INTRODUCTION TO LINEAR PROGRAMMING PROBLEMS (LPP)

1. When a problem is identified then the attempt is to make an mathematical model. In decision making all the decisions are taken through some variables which are known as decision variables. In engineering design, these variables are known as design vectors. So in the formation of mathematical model the following **three phases** are carried out :

(*i*) Identify the decision variables.

(*ii*) Identify the objective using the decision variables and

(*iii*) Identify the constraints or restrictions using the decision variables.

Let there be n decision variable $x_1, x_2, ..., x_n$ and the general form of the mathematical model which *is called as Mathematical programming problem under decision-making* can be stated as follows :

Maximize/Minimize $\quad z = f(x_1, x_2, ..., x_n)$

Subject to, $\quad g_i(x_1, x_2, ..., x_n) \{\leq, \geq \text{ or } =\} b_i$

$i = 1, 2, ..., m.$

and the type of the decisions *i.e.*, $x_j \geq 0$

or, $\quad x_j \leq 0$ or x_j's are unrestricted

or $\quad$ combination types decisions.

In the above, if the functions f and g_i ($i = 1, 2, ..., m$) are all linear, then the model is called "*Linear Programming Problem (LPP)*". If any one function is non-linear then the model is called "*Non-linear Programming Problem (NLPP)*".

II. We define some basic aspects of LPP in the following :

(*a*) Convex set : A set X is said to be convex if

$$x_1, x_2 \in X, \text{ then for } 0 \leq \lambda \leq 1,$$

$$x_3 = \lambda x_1 + (1 - \lambda)x_2 \in X$$

Some examples of convex sets are :

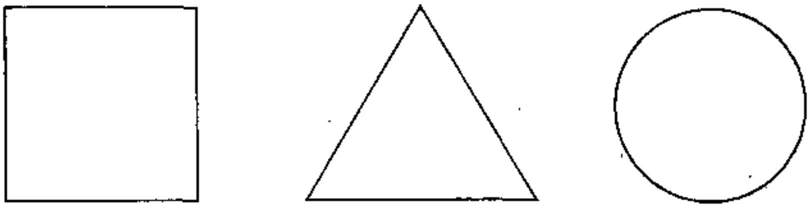


Fig. 4.1 Convex sets

Some examples of non-convex sets are :

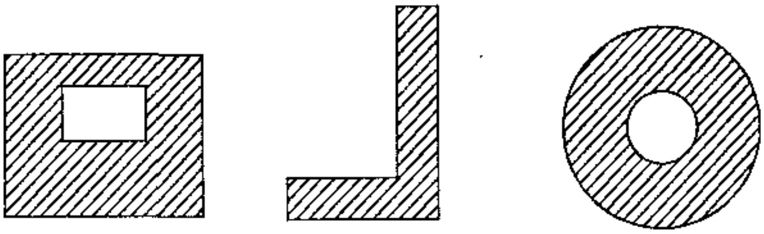


Fig. 4.2 Non-convex sets

Basically if all the points on a line segment forming by two points lies inside the set/geometric figure then it is called convex.

(*b*) Extreme point or vertex or corner point of a convex set : It is a point in the convex set which can not be expressed as $\lambda x_1 + (1 - \lambda)x_2$ where x_1 and x_2 are any two points in the convex set.

For a triangle, there are three vertices, for a rectangle there are four vertices and for a circle there are infinite number of vertices.

(*c*) Let $Ax = b$ be the constraints of an LPP. The set $X = \{x \mid Ax = b, x \geq 0\}$ is a convex set.

Feasible Solution : A solution which satisfies all the constraints in LPP is called feasible solution.

Basic Solution : Let m = no. of constraints and n = no. of variables and $m < n$. Then the solution from the system $Ax = b$ is called basic solution. In this system there are ${}^nc_m$ number of basic solutions. By setting $(n - m)$ variables to zero at a time, the basic solutions are obtained. The variables which is set to zero are known as 'non-basic' variables. Other variables are called basic variables.

Basic Feasible Solution (BFS) : A solution which is basic as well as feasible is called basic feasible solution.

Degenerate BFS : If a basic variable takes the value zero in a BFS, then the solution is said to be degenerate.

Optimal BFS : The BFS which optimizes the objective function is called optimal BFS.

4.3 GRAPHICAL METHOD

Let us consider the constraint $x_1 + x_2 = 1$. The feasible region of this constraint comprises the set of points on the straight line $x_1 + x_2 = 1$.

If the constraint is $x_1 + x_2 \geq 1$, then the feasible region comprises not only the set of points on the straight line $x_1 + x_2 = 1$ but also the points above the line. Here above means away from origin.

If the constraint is $x_1 + x_2 \leq 1$, then the feasible region comprises not only the set of points on the straight line $x_1 + x_2 = 1$ but also the points below the line. Here below means towards the origin.

The above three cases depicted below :

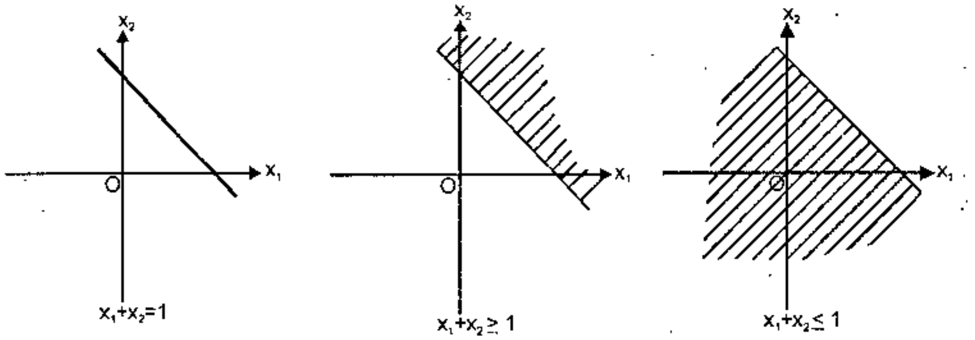


Fig. 4.3

For the constraints $x_1 \geq 1$, $x_1 \leq 1$, $x_2 \geq 1$, $x_2 \leq 1$ the feasible regions are depicted below :

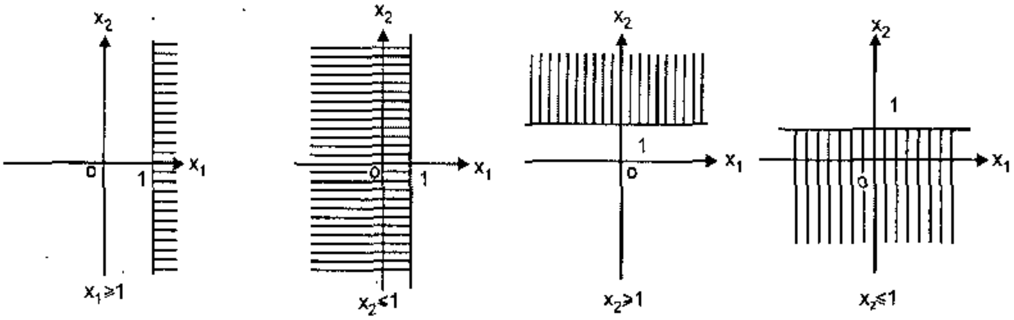


Fig. 4.4

For the constraints $x_1 - x_2 = 0$, $x_1 - x_2 \geq 0$ and $x_1 - x_2 \leq 0$ the feasible regions are depicted in Fig. 4.5.

The *steps of graphical method* can be stated as follows :

(*i*) Plot all the constraints and identify the individual feasible regions.

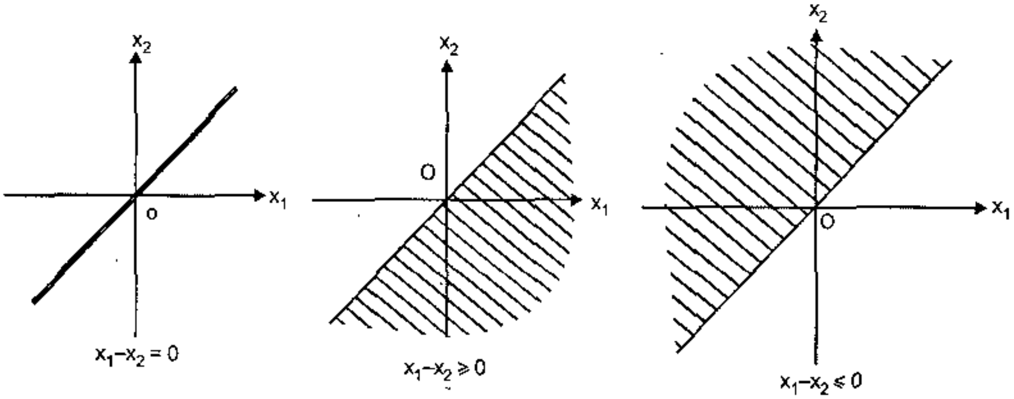


Fig. 4.5

(*ii*) Identify the common feasible region and identify the corner points *i.e.*, vertices of the common feasible region.

(*iii*) Identify the optimal solution at the corner points if exists.

Example 1. *Using graphical method solve the following LPP :*

$$\text{Maximize } z = 5x_1 + 3x_2$$

$$\text{Subject to, } 2x_1 + 5x_2 \leq 10,$$

$$5x_1 + 2x_2 \leq 10,$$

$$2x_1 + 3x_2 \geq 6,$$

$$x_1 \geq 0, x_2 \geq 0.$$

Solution. Let us present all the constraints in intercept form *i.e.*,

$$\frac{x_1}{5} + \frac{x_2}{2} \leq 1 \qquad \text{...(I)}$$

$$\frac{x_1}{2} + \frac{x_2}{5} \leq 1 \qquad \text{...(II)}$$

$$\frac{x_1}{3} + \frac{x_2}{2} \geq 1 \qquad \text{...(III)}$$

The common feasible region ABC is shown in Fig. 4.6 and the individual regions are indicated by arrows. (Due to non-negativity constraints *i.e.*, $x_1 \geq 0$, $x_2 \geq 0$, the common feasible region is obtained in the first quadrant).

The corner points are $A\left(\frac{18}{11}, \frac{10}{11}\right)$, $B\left(\frac{10}{7}, \frac{10}{7}\right)$ and C (0, 2). The value of the objective function at the corner points are $z_A = \frac{120}{11} = 10.91$, $z_B = \frac{80}{7} = 11.43$ and $z_C = 6$.

Here the common feasible region is bounded and the maximum has occurred at the corner point B. Hence the optimal solution is

$$x_1^* = \frac{10}{7},\ x_2^* = \frac{10}{7} \text{ and } z^* = \frac{80}{7} = 11.43.$$

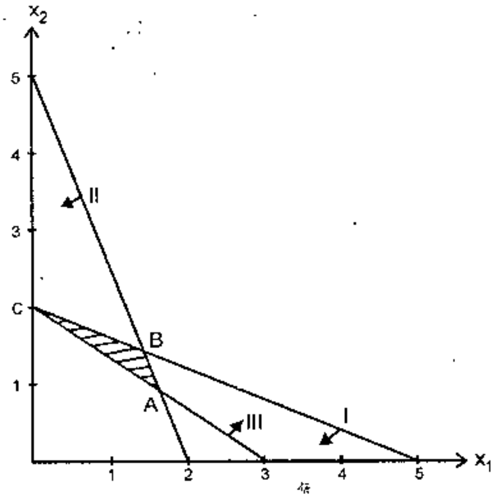


Fig. 4.6

Example 2. *Using graphical method solve the following LPP :*

$$\text{Minimize } z = 3x_1 + 10x_2$$

$$\text{Subject to, } 3x_1 + 2x_2 \geq 6,$$

$$4x_1 + x_2 \geq 4,$$

$$2x_1 + 3x_2 \geq 6,$$

$$x_1 \geq 0,\ x_2 \geq 0.$$

Solution. Let us present all the constraints in intercept form *i.e.,*

$$\frac{x_1}{2} + \frac{x_2}{3} \geq 1 \qquad ...(I)$$

$$\frac{x_1}{1} + \frac{x_2}{4} \geq 1 \qquad ...(II)$$

$$\frac{x_1}{3} + \frac{x_2}{2} \geq 1 \qquad ...(III)$$

Due to the non-negativity constraints *i.e.*, $x_1 \geq 0$ and $x_2 \geq 0$ the feasible region will be in the first quadrant.

The common feasible region is shown in Fig. 4.7 where the individual feasible regions are shown by arrows. Here the common feasible region is unbounded.

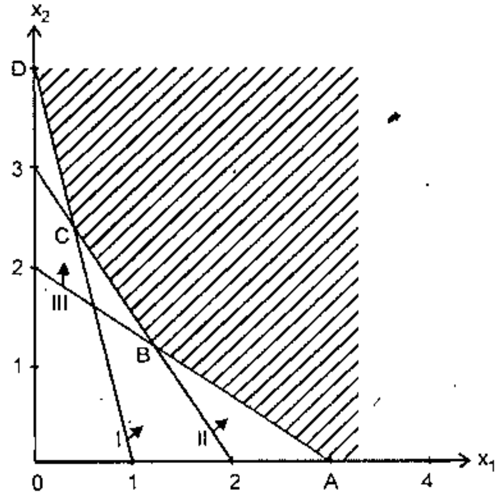


Fig. 4.7

i.e., open with the corner points A(3, 0), B$\left(\frac{3}{5}, \frac{8}{5}\right)$, C$\left(\frac{2}{5}, \frac{12}{5}\right)$ and D (0, 4). The value of the objective function at the corner points are $z_A = 9$, $z_B = \frac{89}{5} = 17.8$, $z_C = \frac{126}{5} = 25.2$, and $z_D = 40$.

Here the minimum has occurred at A and there is no other point in the feasible region at which the objective function value is lower than 9. Hence the optimal solution is

$$x_1^* = 3,\ x_2^* = 0 \text{ and } z^* = 9$$

Example 3. *Solve the following LPP by graphical method :*

$$\text{Maximize } z = 3x_1 - 15x_2$$

$$\text{Subject to, } x_1 + x_2 \le 8,$$

$$x_1 - 4x_2 \le 8,$$

$$x_1 \ge 0,\ x_2 \text{ unrestricted in sign.}$$

Solution. Since x_2 is unrestricted in sign this means x_2 may be ≥ 0 or ≤ 0. Also $x_1 \ge 0$. Then the common feasible region will be in the first and fourth quadrant. Let us present all the constraints in intercept forms *i.e.*,

$$\frac{x_1}{8} + \frac{x_2}{8} \le 1 \quad \text{...(I)}$$

$$\frac{x_1}{8} - \frac{x_2}{2} \le 1 \quad \text{...(II)}$$

The common feasible region is shown in Fig. 4.8 where the individual feasible regions are shown by arrows.

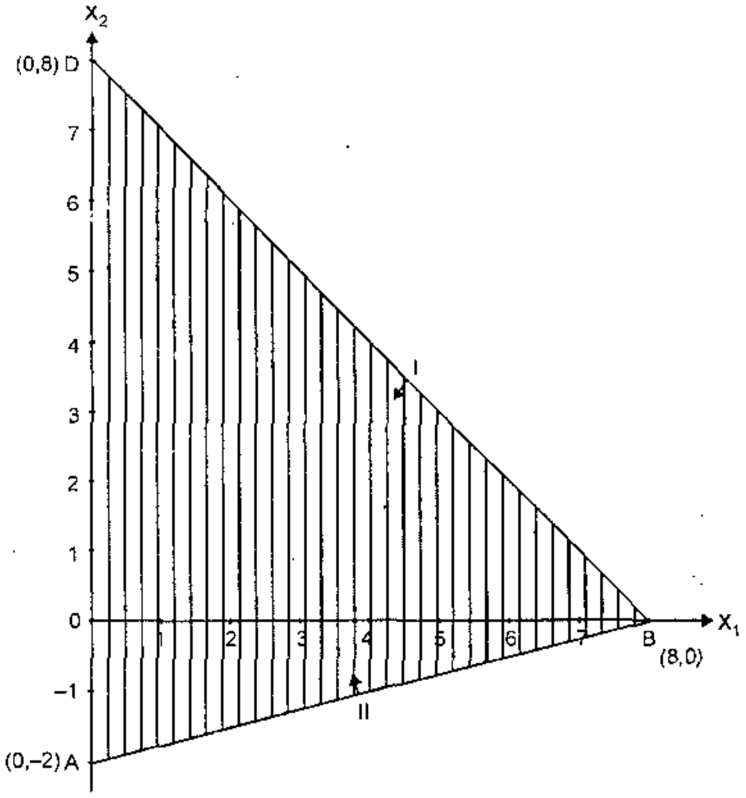


Fig. 4.8

The value of the objective function at the corner points are $z_A = 30$, $z_B = 24$ and $z_C = -120$. Since the common feasible region is bounded and the maximum has occurred at A, the optimal solution is

$$x_1^* = 0,\ x_2^* = -2 \text{ and } z^* = 30.$$

Exceptional Cases in Graphical Method

There are three cases may arise. When the value of the objective function is maximum/minimum at more than one corner points then '*multiple optima*' solutions are obtained.

Sometimes the optimum solution is obtained at infinity, then the solution is called '*unbounded solution*'. Generally, this type of solution is obtained when the common feasible region is unbounded and the type of the objective function leads to unbounded solution.

When there does not exist any common feasible region, then there does not exist any solution. Then the given LPP is called *infeasible i.e.,* having *no solution*. For example, consider the LPP which is infeasible

$$\text{Maximize } z = 5x_1 + 10x_2$$

$$\text{Subject to, } x_1 + x_2 \leq 2,$$

$$x_1 + x_2 \geq 3,$$

$$x_1, x_2 \geq 0.$$

Example 4. *Solve the following LPP using graphical method :*

Maximize $\quad z = x_1 + \frac{3}{5}x_2$

Subject to, $\quad 5x_1 + 3x_2 \leq 15,$

$$3x_1 + 4x_2 \leq 12,$$

$$x_1, x_2 \geq 0.$$

Solution. Let us present all the constraints in intercept forms *i.e.,*

$$\frac{x_1}{3} + \frac{x_2}{5} \leq 1 \qquad \text{...(I)}$$

$$\frac{x_1}{4} + \frac{x_2}{3} \leq 1 \qquad \text{...(II)}$$

Due to non-negativity constraints *i.e.,* $x_1 \geq 0, x_2 \geq 0$ the common feasible region is obtained in the first quadrant as shown in Fig. 4.9 and the individual feasible regions are shown by arrows.

The corner points are O(0, 0), A (3, 0), $B\left(\frac{24}{11}, \frac{15}{11}\right)$ and C(0, 3). The values of the objective function at the corner points are obtained as $z_O = 0$, $z_A = 3$, $z_B = 3$, $z_C = \frac{9}{4}$.

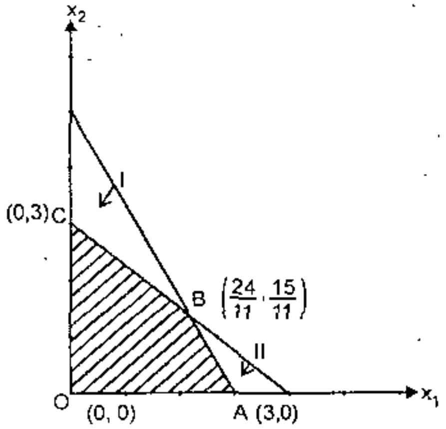


Fig. 4.9

Since the common feasible region is bounded and the maximum has occurred at two corner points *i.e.*, at A and B respectively, these solutions are called multiple optima. So the solutions are

$$x_1^* = 3,\ x_2^* = 0 \qquad \text{and} \qquad x_1^* = \frac{15}{11},\ x_2^* = \frac{24}{11} \text{ and } z^* = 3.$$

Example 5. *Using graphical method show that the following LPP is unbounded.*

$$\text{Maximize } z = 10x_1 + 3x_2$$

$$\text{Subject to, } -2x_1 + 3x_2 \leq 6,$$

$$x_1 + 2x_2 \geq 4,$$

$$x_1,\ x_2 \geq 0.$$

Solution. Due to the non-negativity constraints *i.e.*, $x_1 \geq 0$ and $x_2 \geq 0$ the common feasible region will be obtained in the first quadrant. Let us present the constraints in the intercept forms *i.e.*,

$$\frac{x_1}{-3} + \frac{x_2}{2} \leq 1 \qquad \text{...(I)}$$

$$\frac{x_1}{4} + \frac{x_2}{2} \geq 1 \qquad \text{...(II)}$$

The common feasible region is shown in Fig. 4.10 which is unbounded *i.e.*, open region.

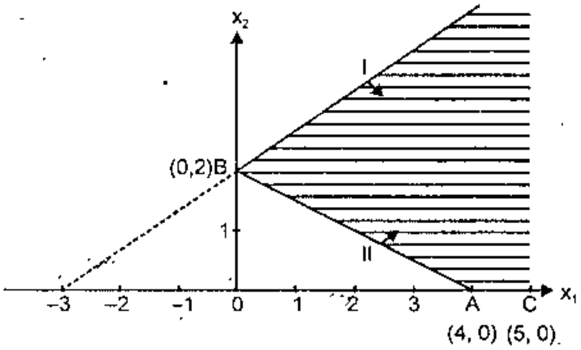


Fig. 4.10

There are two corner points A(4, 0) and B(0, 2). The objective function values are $z_A = 40$ and $z_B = 6$. Here the maximum is 40. Since the region is open, let us examine some other points.

Consider the point C(5, 0) and the value of the objective function is $z_C = 50$ which is greater than z_A. Therefore z_A is no longer optimal. If we move along x-axis, we observe that the next value is higher than the previous value and we reach to infinity for optimum value. Hence the problem is unbounded.

Note. For the same problem minimum exists which is the point B.

PROBLEMS

NOTES

Using graphical method solve the following LPP :

1. $$\text{Maximize } z = 13x_1 + 117x_2$$
$$\text{Subject to, } x_1 + x_2 \le 12,$$
$$x_1 - x_2 \ge 0,$$
$$4x_1 + 9x_2 \le 36,$$
$$0 \le x_1 \le 2 \text{ and } \quad 0 \le x_2 \le 10.$$

2. $$\text{Maximize } z = 3x_1 + 15x_2$$
$$\text{Subject to, } 4x_1 + 5x_2 \le 20,$$
$$x_2 - x_1 \le 1,$$
$$0 \le x_1 \le 4 \text{ and } \quad 0 \le x_2 \le 3.$$

3. $$\text{Maximize } z = 5x_1 + 7x_2$$
$$\text{Subject to, } 3x_1 + 8x_2 \le 12,$$
$$x_1 + x_2 \le 2,$$
$$2x_1 \le 3,$$
$$x_1, x_2 \ge 0.$$

4. $$\text{Minimize } z = 2x_1 + 3x_2$$
$$\text{Subject to, } x_2 - x_1 \ge 2,$$
$$5x_1 + 3x_2 \le 15,$$
$$2x_1 \ge 1,$$
$$x_2 \le 4,$$
$$x_1, x_2 \ge 0.$$

5. $$\text{Minimize } z = 10x_1 + 9x_2$$
$$\text{Subject to, } x_1 + 2x_2 \le 10,$$
$$x_1 - x_2 \le 0,$$
$$x_1 \le 0, x_2 \ge 0.$$

6. $$\text{Minimize } z = 4x_1 + 3x_2$$
$$\text{Subject to, } 2x_1 + 3x_2 \le 12,$$
$$3x_1 - 2x_2 \le 12,$$
$$x_1 \text{ unrestricted in sign, } x_2 \ge 0.$$

7. $$\text{Maximize } z = 10x_1 + 11x_2$$
$$\text{Subject to, } x_1 + x_2 \ge 4,$$
$$0 \le x_2 \le 3,$$
$$x_1 \ge 2,$$
$$x_1 \ge 0.$$

8. $$\text{Minimize } z = -x_1 + 2x_2$$
$$\text{Subject to, } x_1 - x_2 \ge 1,$$
$$x_1 + x_2 \ge 5,$$
$$x_1, x_2 \ge 0.$$

NOTES

9. $$\text{Maximize } z = 4x_1 + 5x_2$$
$$\text{Subject to, } 4x_1 - 5x_2 \leq 20,$$
$$x_2 - x_1 \leq 1,$$
$$0 \leq x_2 \leq 3$$
$$0 \leq x_1 \leq 4.$$

10. $$\text{Maximize } z = -3x_1 + 4x_2$$
$$\text{Subject to, } -3x_1 + 4x_2 \leq 12,$$
$$2x_1 - x_2 \geq -2,$$
$$x_1 \leq 4,$$
$$x_1 \geq 0, x_2 \geq 0.$$

ANSWERS

1. $x_1 = 2, x_2 = 2, z^* = 260$
2. $x_1 = \frac{5}{3},\ x_2 = \frac{8}{3},\ z^* = 45$
3. $x_1 = \frac{4}{5},\ x_2 = \frac{6}{5},\ z^* = \frac{62}{5}$
4. $x_1 = \frac{1}{2},\ x_2 = \frac{5}{2},\ z^* = \frac{17}{2}$
5. $x_1 = 0, x_2 = 0, z^* = 0$
6. $x_1 = 4, x_2 = 0, z^* = 16$
7. Unbounded solution
8. Unbounded solution
9. Multiple optima :

   $x_1 = \frac{5}{3},\ x_2 = \frac{8}{3}$, and $x_1 = 4,\ x_2 = \frac{4}{5}$ and $z^* = 20$

10. Multiple optima :

    $x_1 = \frac{4}{5},\ x_2 = \frac{18}{5}$ and $x_1 = 4, x_2 = 6$ and $z^* = 12$

4.4 SIMPLEX METHOD

The algorithm is discussed below with the help of a numerical example *i.e.*, consider

$$\text{Maximize } z = 4x_1 + 8x_2 + 5x_3$$
$$\text{Subject to, } x_1 + 2x_2 + 3x_3 \leq 18,$$
$$2x_1 + 6x_2 + 4x_3 \leq 15,$$
$$x_1 + 4x_2 + x_3 \leq 6,$$
$$x_1, x_2, x_3 \geq 0.$$

Step 1. If the problem is in minimization, then convert it to maximization as Min $z = -$ Max $(-z)$.

Step 2. All the right side constants must be positive. Multiply by -1 both sides for negative constants. All the variables must be non-negative.

Step 3. Make standard form by adding slack variables for '≤' type constraints, surplus variables for '≥' type constraints and incorporate these variables in the objective function with zero coefficients.

For example. $\text{Maximum } z = 4x_1 + 8x_2 + 5x_3 + 0.s_1 + 0.s_2 + 0.s_3$

$$\begin{aligned} \text{Subject to, } x_1 + 2x_2 + 3x_3 + s_1 &= 18, \\ 2x_1 + 6x_2 + 4x_3 + s_2 &= 15, \\ x_1 + 4x_2 + x_3 + s_3 &= 6, \\ x_1, x_2, x_3 &\geq 0, s_1, s_2, s_3 \geq 0 \end{aligned}$$

Note that an unit matrix due to s_1, s_2 and s_3 variables is present in the coefficient matrix which is the key requirement for simplex method.

Step 4. Simplex method is an iterative method. Calculations are done in a table which is called simplex table. For each constraint there will be a row and for each variable there will be a column. Objective function coefficients c_j are kept on the top of the table. x_B stands for basis column in which the variables are called 'basic variables'. Solution column gives the solution, but in iteration 1, the right side constants are kept. At the bottom $z_j - c_j$ row is called 'net evaluation' row.

In each iteration one variable departs from the basis and is called departing variable and in that place one variable enter which is called entering variable to improve the value of the objective function.

Minimum ratio column determines the departing variable.

Iteration 1.

		c_j	4	8	5	0	0	0	Min.
c_B	x_B	soln.	x_1	x_2	x_3	s_1	s_2	s_3	ratio
0	s_1	18	1	2	3	1	0	0	
0	s_2	15	2	6	4	0	1	0	
0	s_3	6	1	4	1	0	0	1	
	$z_j - c_j$								

Note. Variables which are forming the columns of the unit matrix enter into the basis column. In this table the solution is $s_1 = 18$, $s_2 = 15$, $s_3 = 6$, $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ and $z = 0$.

To test optimality we have to calculate $z_j - c_j$ for each column as follows :

$$z_j - c_j = c_B^T.[x_j] - c_j$$

For first column, $$(0, 0, 0)\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - 4 = -4$$

For second column, $$(0, 0, 0)\begin{pmatrix} 2 \\ 6 \\ 4 \end{pmatrix} - 8 = -8$$ and so on.

These are displayed in the following table :

		c_j	4	8	5	0	0	0	Min.
c_B	x_B	soln.	x_1	x_2	x_3	s_1	s_2	s_3	ratio
0	s_1	18	1	2	3	1	0	0	
0	s_2	15	2	6	4	0	1	0	
0	s_3	6	1	4	1	0	0	1	
	$z_j - c_j$		– 4	– 8	– 5	0	0	0	

↑

Decisions : If all $z_j - c_j \geq 0$. Then the current solution is optimal and stop. Else, Select the negative most value from $z_j - c_j$ and the variable corresponding to this value will be the entering variable and that column is called '*key column*'. Indicate this column with an upward arrow symbol.

In the given problem '– 8' is the most negative and variable x_2 is the entering variable. If there is a tie in the most negative, break it arbitrarily.

To determine the *departing variable*, we have to use minimum ratio. Each ratio is calculated as $\frac{[\text{soln.}]}{[\text{key column}]}$, componentwise division only for positive elements (*i.e.*, > 0) of the key column. In this example,

$$\min.\left\{\frac{18}{2}, \frac{15}{6}, \frac{6}{4}\right\} = \min.\{9, 2.5, 1.5\} = 1.5$$

The element corresponding to the min. ratio *i.e.*, here s_3 will be the departing variable and the corresponding row is called 'key row' and indicate this row by an outward arrow symbol. The intersection element of the key row and key column is called key element. In the present example, 4 is the key element which is highlighted. This is the end of this iteration, the final table is displayed below :

Iteration 1:

		c_j	4	8	5	0	0	0	Min.
c_B	x_B	soln.	x_1	x_2	x_3	s_1	s_2	s_3	ratio
0	s_1	18.	1	2	3	1	0	0	$\frac{18}{2} = 9$
0	s_2	15	2	6	4	0	1	0	$\frac{15}{6} = 2.5$ →
0	s_3	6	1	$\boxed{4}$	1	0	0	1	$\frac{6}{4} = 1.5$
	$z_j - c_j$		– 4	– 8	– 5	0	0	0	

↑

Step 5. For the construction of the next iteration (new) table the following calculations are to be made :

NOTES

(*a*) Update the x_B column and the c_B column.

(*b*) Divide the key row by the key element.

(*c*) Other elements are obtained by the following formula :

$$\begin{pmatrix}\text{new}\\ \text{element}\end{pmatrix} = \begin{pmatrix}\text{old}\\ \text{element}\end{pmatrix} - \frac{\begin{pmatrix}\text{element}\\ \text{corresponding to}\\ \text{key row}\end{pmatrix} \cdot \begin{pmatrix}\text{element}\\ \text{corresponding to}\\ \text{key column}\end{pmatrix}}{\text{key element}}$$

(*d*) Then go to step 4.

Iteration 2.

		c_j	4	8	5	0	0	0	Min.
c_B	x_B	soln.	x_1	x_2	x_3	s_1	s_2	s_3	ratio
0	s_1	15	$\frac{1}{2}$	0	$\frac{5}{2}$	1	0	$-\frac{1}{2}$	$15 \times \frac{3}{5} = 6$
0	s_2	6	$\frac{1}{2}$	0	$\boxed{\frac{5}{2}}$	0	1	$-\frac{3}{2}$	$6 \times \frac{2}{5} = 2.4$ →
8	x_2	$\frac{3}{2}$	$\frac{1}{4}$	1	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{3}{4} \times 4 = 3$
	$z_j - c_j$		-2	0	-3	0	0	2	

↑

Iteration 3.

		c_j	4	8	5	0	0	0	Min.
c_B	x_B	soln.	x_1	x_2	x_3	s_1	s_2	s_3	ratio
0	s_1	9	0	0	0	1	-1	1	-
5	x_3	$\frac{12}{5}$	$\frac{1}{5}$	0	1	0	$\frac{2}{5}$	$-\frac{3}{5}$	$\frac{12}{5} \times \frac{5}{1} = 12$
8	x_2	$\frac{9}{10}$	$\boxed{\frac{1}{5}}$	1	0	0	$-\frac{1}{10}$	$\frac{2}{5}$	$\frac{9}{10} \times \frac{5}{1} = 4.5$ →
	$z_j - c_j$		$-\frac{7}{5}$	0	0	0	$\frac{6}{5}$	$\frac{1}{5}$	

↑

Iteration 4.

		c_j	4	8	5	0	0	0	Min.
c_B	x_B	soln.	x_1	x_2	x_3	s_1	s_2	s_3	ratio
0	s_1	9	0	0	0	1	-1	1	

NOTES

5	x_3	$\frac{3}{2}$	0	-1	1	0	$\frac{1}{2}$	-1	
4	x_1	$\frac{9}{2}$	1	5	0	0	$-\frac{1}{2}$	2	
	$z_j - c_j$		0	7	0	0	$\frac{1}{2}$	3	

Since all $z_j - c_j \geq 0$, the current solution is optimal.

$$\therefore \qquad x_1^* = \frac{9}{2},\ x_2^* = 0,\ x_3^* = \frac{3}{2},\ z^* = \frac{51}{2}.$$

Note (exceptional cases).

(*a*) If in the key column, all the elements are non-positive *i.e.*, zero or negative, then min. ratio cannot be calculated and the problem is said to be unbounded.

(*b*) In the net evaluation of the optimal table all the basic variables will give the value zero. If any non-basic variable give zero net evaluation then it indicates that there is an alternative optimal solution. To obtain that solution, consider the corresponding column as key column and apply one simplex iteration.

(*c*) For negative variables, $x \leq 0$, set $x = -x'$, $x' \geq 0$.

For unrestricted variables set $x = x' - x''$ where $x', x'' \geq 0$.

Example 6. *Solve the following by simplex method :*

$$\text{Maximize } z = x_1 + 3x_2$$

$$\text{Subject to, } -x_1 + 2x_2 \leq 2,\ x_1 - 2x_2 \leq 2,\ x_1, x_2 \geq 0.$$

Solution. Standard form of the given LPP can be written as follows :

$$\text{Maximum } z = x_1 + 3x_2 + 0.s_1 + 0.s_2$$

$$\text{Subject to, } -x_1 + 2x_2 + s_1 = 2,\ x_1 - 2x_2 + s_2 = 2,$$

$$x_1, x_2 \geq 0,\ s_1, s_2 \text{ slacks} \geq 0.$$

Iteration 1.

		c_j	1	3	0	0	Min.	
c_B	x_B	soln.	x_1	x_2	s_1	s_2	ratio	
0	s_1	2	-1	$\boxed{2}$	1	0	$\frac{2}{2} = 1$	→
0	s_2	2	1	-2	0	1	–	
	$z_j - c_j$		-1	-3	0	0		
				↑				

Iteration 2.

		c_j	1	3	0	0	Min.
c_B	x_B	soln.	x_1	x_2	s_1	s_2	ratio
3	x_2	1	$-\frac{1}{2}$	1	$\frac{1}{2}$	0	
0	s_2	4	0	0	1	1	
	$z_j - c_j$		$-\frac{5}{2}$	0	$\frac{3}{2}$	0	
			↑				

Since all the elements in the key column are non-positive, we cannot calculate min. ratio. Hence the given LPP is said to be unbounded.

PROBLEMS

Solve the following LPP by simplex method:

11. Maximize $z = 3x_1 + 2x_2$

S/t, $5x_1 + x_2 \leq 10$, $4x_1 + 5x_2 \leq 60$; $x_1, x_2 \geq 0$

12. Maximize $z = 5x_1 + 4x_2 + x_3$

S/t, $6x_1 + x_2 + 2x_3 \leq 12$, $8x_1 + 2x_2 + x_3 \leq 30$,

$4x_1 + x_2 - 2x_3 \leq 16$, $x_1, x_2, x_3 \geq 0$

13. Maximize $z = 3x_1 + 2x_2$

S/t, $3x_1 + 4x_2 \leq 12$, $2x_1 + 5x_2 \leq 10$, $x_1, x_2 \geq 0$

14. Maximize $z = 3x_1 + 2x_2 + x_3$

S/t, $3x_1 + x_2 + 2x_3 \leq 20$, $x_1 + 3x_2 + 4x_3 \leq 16$, $x_1, x_2, x_3 \geq 0$

15. Maximize $z = 4x_1 - 2x_2 - x_3$

S/t, $x_1 + x_2 + x_3 \leq 3$, $2x_1 + 2x_2 + x_3 \leq 4$, $x_1 - x_2 \leq 0$,
$x_1, x_2, x_3 \geq 0$

16. Maximize $z = 5x_1 + 3x_2 + 3x_3$

S/t, $4x_1 + 4x_2 + 3x_3 \leq 12000$, $0.4x_1 + 0.5x_2 + 0.3x_3 \leq 1800$,

$0.2x_1 + 0.2x_2 + 0.1x_3 \leq 960$, $x_1, x_2, x_3 \geq 0$

17. Maximize $z = 3x_1 + 2x_2 + 2x_3$

S/t, $2x_1 - x_2 + 3x_3 \leq 18$, $x_1 + x_2 + 2x_3 \leq 12$, $x_1, x_2, x_3 \geq 0$

18. Maximize $z = 3x_1 + x_2 + x_3 + x_4$

S/t, $-2x_1 + 2x_2 + x_3 = 4$, $3x_1 + x_2 + x_4 = 6$, $x_i \geq 0$ for all i

19. Maximize $z = x_1 + x_2$

S/t, $x_1 - 2x_2 \leq 2$, $-x_1 + 2x_2 \leq 2$, $x_1, x_2 \geq 0$

20. Find all the optimal BFS to the following :

Maximize $z = x_1 + x_2 + x_3 + x_4$

S/t, $x_1 + x_2 \leq 2$, $x_3 + x_4 \leq 5$, $x_1, x_2, x_3, x_4 \geq 0$

ANSWERS

11. $x_1 = 0, x_2 = 10, z^* = 20$ (It 3)
12. $x_1 = 0, x_2 = 12, x_3 = 0, z^* = 48$ (It 3)
13. $x_1 = 4, x_2 = 0, z^* = 12$ (It 2)
14. $x_1 = \frac{11}{2}, x_2 = \frac{7}{2}, x_3 = 0, z^* = \frac{47}{2}$ (It 3)
15. $x_1 = 1, x_2 = 1, x_3 = 0, z^* = 2$ (It 3)
16. $x_1 = 3000, x_2 = 0, x_3 = 0, z^* = 15000$ (It 2)
17. $x_1 = 10, x_2 = 2, x_3 = 0, z^* = 34$ (It 3)
18. Solution 1 : $x_1 = 1, x_2 = 3, x_3 = x_4 = 0$ (It 2)
    Solution 2 : $x_1 = 0, x_2 = 2, x_3 = 0, x_4 = 4, z^* = 6$
19. Unbounded solution (It 2)
20. (2, 0, 5, 0), (0, 2. 5. 0), (0, 2, 0, 5), (2, 0, 0, 5).

FURTHER READINGS

1. Comprehensive Business Statistics: Premanand Gupta.
2. Statistics and Operations Research: Dr. Debashis Gupta.

UNIT 5 TRANSPORTATION, ASSIGNMENT PROBLEMS AND GAME THEORY

NOTES

★ STRUCTURE ★

5.0 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- explain mathematical formulation
- describe uv-method/modi method
- define degeneracy in T.P., Max-type and unbalanced T.P.
- illustrate game theory and basic defenitions
- explain dominance rules and graphical method for games
- define linear programming method for games.

Theorem 2. The number of basic variables in the basic feasible solution of an $m \times n$ T.P. is $m + n - 1$.

Proof. This is due to the fact that the one of the constraints is redundant in balanced T.P.

We have overall supply, $$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = \sum_{i=1}^{m} a_i$$

and overall demand $$\sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij} = \sum_{j=1}^{n} b_j$$

Since $\sum_i a_i = \sum_i b_i$, the above two equations are identical and we have only $m + n - 1$ independent constraints. Hence the theorem is proved.

Note. 1. If any basic variable takes the value zero then the basic feasible solution (BFS) is said to be degenerate. Like LPP, all non-basic variables take the value zero.

2. If a basic variable takes either positive value or zero, then the corresponding cell is called 'Basic cell' or 'Occupied cell'. For non-basic variable the corresponding cell is called 'Non-basic cell' or 'Non-occupied cell' or 'Non-allocated cell'.

Loop. This means a closed circuit in a transportation table connecting the occupied (or allocated) cells satisfying the following :

(*i*) It consists of vertical and horizontal lines connecting the occupied (or allocated) cells.

(*ii*) Each line connects only two occupied (or allocated) cells.

(*iii*) Number of connected cells is even.

(*iv*) Lines can skip the middle cell of three adjacent cells to satisfy the condition (*ii*).

The following are the examples of loops.

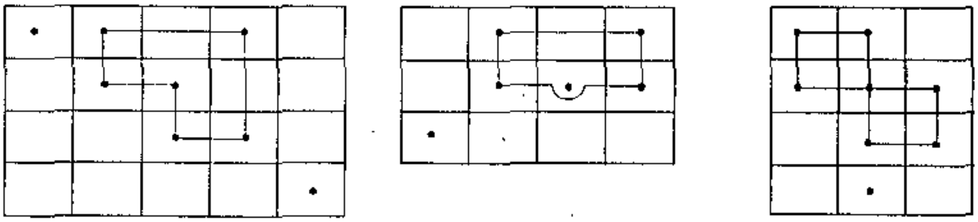


Fig. 5.1

Note. A solution of a T.P. is said to be basic if it does not consist of any loop.

5.2 FINDING INITIAL BASIC FEASIBLE SOLUTION

In this section three methods are to be discussed to find initial BFS of a T.P. In advance, it can be noted that the above three methods may give different initial BFS to the same T.P. Also allocation = minimum (supply, demand).

(*a*) North-West Corner Rule (NWC)

(*i*) Select the north west corner cell of the transportation table.

(*ii*) Allocate the min (supply, demand) in that cell as the value of the variable.

If supply happens to be minimum, cross-off the row for further consideration and adjust the demand.

If demand happens to be minimum, cross-off the column for further consideration and adjust the supply.

(*iii*) The table is reduced and go to step (*i*) and continue the allocation until all the supplies are exhausted and the demands are met.

Example 1. *Find the initial BFS of the following T.P. using NWC rule.*

To

From	M_1	M_2	M_3	M_4	Supply
F_1	3	2	4	1	20
F_2	2	4	5	3	15
F_3	3	5	2	6	25
F_4	4	3	1	4	40
	30	20	25	25	

Demand

Solution. Here, total supply = 100 = total demand. So the problem is balanced T.P. The north-west corner cell is (1, 1) cell. So allocate min. (20, 30) = 20 in that cell. Supply exhausted. So cross-off the first row and demand is reduced to 10. The reduced table is

	M_1	M_2	M_3	M_4	
F_2	2	4	5	3	15
F_3	3	5	2	6	25
F_4	4	3	1	4	40
	10	20	25	25	

Here the north-west corner cell is (2, 1) cell. So allocate min. (15, 10) = 10 in that cell. Demand met. So cross-off the first column and supply is reduced to 5. The reduced table is

	M_2	M_3	M_4	
F_2	4	5	3	5
F_3	5	2	6	25
F_4	3	1	4	40
	20	25	25	

Here the north-west corner cell is (2, 2) cell. So allocate min. (5, 20) = 5 in that cell. Supply exhausted. So cross-off the second row (due to F2) and demand is reduced to 15. The reduced table is

	M_2	M_3	M_4	
F_3	5	2	6	25
F_4	3	1	4	40
	15	25	25	

Here the north-west corner cell is (3, 2) cell. So allocate min. (25, 15) = 15 in that cell. Demand met. So cross-off the second column (due to M_2) and supply is reduced to 10. The reduced table is

	M_3	M_4	
F_3	2	6	10
F_4	1	4	40
	25	25	

Here the north-west corner cell is (3, 3) cell. So allocate min. (10, 25) = 10 in that cell. Supply exhausted. So cross-off the third row (due to F_3) and demand is reduced to 15. The reduced table is

	M_3	M_4	
F_4	1	4	40
	25	25	

continuing we obtain the allocation 15 to (4, 3) cell and 25 to (4, 4) cell so that supply exhausted and demand met. The **complete allocation** is shown below:

	M_1	M_2	M_3	M_4
F_1	$\boxed{20}$ 3	2	4	1
F_2	$\boxed{10}$ 2	$\boxed{5}$ 4	5	3
F_3	3	$\boxed{15}$ 5	$\boxed{10}$ 2	6
F_4	4	3	$\boxed{15}$ 1	$\boxed{25}$ 4

Thus, the initial BFS is

$$x_{11} = 20,\ x_{21} = 10,\ x_{22} = 5,\ x_{32} = 15,\ x_{33} = 10,\ x_{43} = 15,\ x_{44} = 25.$$

The transportation cost

$$= 20 \times 3 + 10 \times 2 + 5 \times 4 + 15 \times 5 + 10 \times 2 + 5 \times 1 + 25 \times 4$$

$$= \text{Rs. } 310.$$

(*b*) Least Cost Entry Method (LCM) (or Matrix Minimum Method)

(*i*) Find the least cost from transportation table. If the least value is unique, then go for allocation.

If the least value is not unique then select the cell for allocation for which the contributed cost is minimum.

(*ii*) If the supply is exhausted cross-off the row and adjust the demand.

If the demand is met cross-off the column and adjust the supply.

Thus the matrix is reduced.

(*iii*) Go to step (*i*) and continue until all the supplies are exhausted and all the demands are met.

Example 2. *Find the initial BFS of Example 1 using least cost entry method:*

	M_1	M_2	M_3	M_4	
F_1	3	2	4	1	20
F_2	2	4	5	3	15
F_3	3	5	2	6	25
F_4	4	3	1	4	40
	30	20	25	25	

Solution. Here the least value is 1 and occurs in two cells (1, 4) and (4, 3). But the contributed cost due to cell (1, 4) is 1 × min (20, 25) *i.e.,* 20 and due to cell (4, 3) is 1 × min. (40, 25) *i.e.,* 25. So we selected the cell (1, 4) and allocate 20. Cross-off the first row since supply exhausted and adjust the demand to 5. The reduced table is given below :

2	4	5	3	15
3	5	2	6	25
4	3	1	4	40
30	20	25	5	

The least value is 1 and unique. So allocate min. (40, 25) = 25 in that cell. Cross-off the third column (due to M_3) since the demand is met and adjust the supply to 15. The reduced table is given below :

2	4	3	15
3	5	6	25
4	3	4	15
30	20	5	

The least value is 2 and unique. So allocate min. (15, 30) = 15 in that cell. Cross-off the second row (due to F_2) since the supply exhausted and adjust the demand to 15. The reduced table is given below :

3	5	6	25
4	3	4	15
15	20	5	

The least value is 3 and occurs in two cells (3, 1) and (4, 2). The contributed cost due to cell (3, 1) is 3 × min. (25, 15) = 45 and due to cell (4, 2) is 3 × min. (15, 20) = 45. Let us select the (3, 1) cell for allocation and allocate 15. Cross-off the first column (due to M_1) since demand is met and adjust the supply to 10. The reduced table is given below :

5	6	10
3	4	15
20	5	

Continuing the above method and we obtain the allocations in the cell (4, 2) as 15, in the cell (3, 2) as 5 and in the cell (3, 4) as 5. The complete allocation is shown below :

	M_1	M_2	M_3	M_4
F_1	3	2	4	$\boxed{20}$ 1
F_2	$\boxed{15}$ 2	4	5	3
F_3	$\boxed{15}$ 3	$\boxed{5}$ 5	2	$\boxed{5}$ 6
F_4	4	$\boxed{15}$ 3	$\boxed{25}$ 1	4

The initial BFS is

$$x_{14} = 20,\ x_{21} = 15,\ x_{31} = 15,\ x_{32} = 5,\ x_{34} = 5,\ x_{42} = 15,\ x_{43} = 25.$$

The transportation cost

$$= 20 \times 1 + 15 \times 2 + 15 \times 3 + 5 \times 5 + 5 \times 6 + 15 \times 3 + 25 \times 1$$

$$= \text{Rs. } 220.$$

Note. If the least cost is only selected columnwise then it is called 'column minima' method. If the least cost is only selected row wise then it is called 'row minima' method.

(*c*) Vogel's Approximation Method (VAM)

(*i*) Calculate the row penalties and column penalties by taking the difference between the lowest and the next lowest costs of every row and of every column respectively.

(*ii*) Select the largest penalty by encircling it. For tie cases, it can be broken arbitrarily or by analyzing the contributed costs.

(*iii*) Allocate in the least cost cell of the row/column due to largest penalty.

(*iv*) If the demand is met, cross off the corresponding column and adjust the supply.

If the supply is exhausted, cross-off the corresponding row and adjust the demand.

Thus the transportation table is reduced.

(*v*) Go to Step (*i*) and continue until all the supplies exhausted and all the demands are met.

Example 3. *Find the initial BFS of example 1 using Vogel's approximation method.*

Solution.

	M_1	M_2	M_3	M_4	Row penalties
F_1	3	2	4	$\boxed{20}$ 1	20 (1)
F_2	2	4	5	3	15 (1)
F_3	3	5	2	6	25 (1)
F_4	4	3	1	4	40 (2)
Column penalties	30 (1)	20 (1)	25 (1)	25 ②	

Since there is a tie in penalties, let us break the tie by considering the contributed costs. Due to M_4, the contributed cost is 1 × min. (20, 25) = 20. While due to F_4, the contributed cost is 1 × min. (40, 25) = 25. So select the column due to M_4 for allocation and we allocate min. (20, 25) *i.e.*, 20 in (1, 4) cell. Then cross-off the first row as supply is exhausted and adjust the corresponding demand as 5. The reduced table is

	M_1	M_2	M_3	M_4	Row penalties
F_2	2	4	5	3	15 (1)
F_3	3	5	2	6	25 (1)
F_4	4	3	1	4	40 ②
Column penalties	30 (1)	20 (1)	25 (1)	5 (1)	

Here the largest penalty is 2 which is due to F_4. Allocate in (4, 3) cell as min. (40, 25) = 25. Cross-off the third column due to M_3, since demand is met and adjust the corresponding supply to 15. The reduced table is

	M_1	M_2	M_4	Row penalties
F_2	2	4	3	15 (1)
F_3	3	5	6	25 ②
F_4	4	3	4	15 (1)
Column penalties	30 (1)	20 (1)	5 (1)	

Here the largest penalty is 2 which is due to F_3. Allocate in (3, 1) cell as min. (25, 30) = 25. Cross-off the third row due to F_3 since supply is exhausted and adjust the corresponding demand to 5. The reduced table is

	M_1	M_2	M_4	Row penalties
F_2	2	4	3	15 (1)
F_4	4	3	4	15 (1)
Column penalties	5 ②	20 (1)	5 (1)	

Here the largest penalty is 2 which is due to M_1. Allocate in (2, 1) cell as min. (15, 5) = 5. Corss-off the first column due to M_1 since demand is met and adjust the supply to 10. The reduced table is

	M_2	M_4	Row penalties
F_2	4	3	10 (1)
F_4	3	4	15 (1)
	20	5	
Column penalties	(1)	①	

Here tie has occurred. The contributed cost is minimum due to (2, 4) cell which is 3 × min. (10, 5) = 15. So allocate min. (10, 5) = 5 in (2, 4) cell. Cross-off the fourth column which is due to M_4 since demand is met and adjust the corresponding supply to 5. On continuation we obtain the allocation of 5 in (2, 2) cell and 15 in (4, 2) cell. The complete allocation in shown below :

	M_1	M_2	M_3	M_4
F_1	3	2	4	20 1
F_2	5 2	5 4	5	5 3
F_3	25 3	5	2	6
F_4	4	15 3	25 1	4

The initial BFS is

$$x_{14} = 20,\ x_{21} = 5,\ x_{22} = 5,\ x_{24} = 5,\ x_{31} = 25,\ x_{42} = 15,\ x_{43} = 25.$$

The transportation cost

$$= 1 \times 20 + 2 \times 5 + 4 \times 5 + 3 \times 5 + 3 \times 25 + 3 \times 15 + 1 \times 25$$

$$= \text{Rs. } 210.$$

5.3 UV-METHOD/MODI METHOD

Taking the initial BFS by any method discussed above, this method find the optimal solution to the transportation problem. The steps are given below :

(*i*) For each row consider a variable u_i and for each column consider another variable v_j.

Find u_i and v_j such that

$u_i + v_j = c_{ij}$ for every basic cells.

(*ii*) For every non-basic cells, calculate the net evaluations as follows :

$\bar{c}_{ij} = u_i + v_j - c_{ij}$

If all $\bar{c}_{ij}$ are non-positive, the current solution is optimal.

If at least one $\bar{c}_{ij} > 0$, select the variable having the largest positive net evaluation to enter the basis.

(*iii*) Let the variable x_{rc} enter the basis. Allocate an unknown quantity θ to the cell (r, c).

Identify a loop that starts and ends in the cell (r, c).

Subtract and add θ to the corner points of the loop clockwise/anticlockwise.

(*iv*) Assign a minimum value of θ in such a way that one basic variable becomes zero and other basic variables remain non-negative. The basic cell which reduces to zero leaves the basis and the cell with θ enters into the basis.

If more than one basic variables become zero due to the minimum value of θ, then only one basic cell leaves the basis and the solution is called degenerate.

(*v*) Go to step (*i*) until an optimal BFS has been obtained.

NOTES

Note. In step (*ii*), if all $\bar{c}_{ij} < 0$, then the optimal solution is unique. If at least one $\bar{c}_{ij} < 0$, then we can obtain alternative solution. Assign θ in that cell and repeat one iteration (from step (*iii*)).

Example 4. *Consider the initial BFS by LCM of Example 2, find the optimal solution of the T.P.*

Solution. Iteration 1.

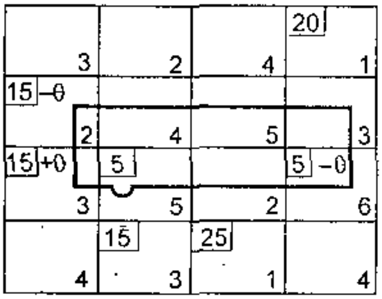
	M₁	M₂	M₃	M₄	
F₁	3	2	4	20 1	$u_1 = -5$
F₂	15 2	4	5	3	$u_2 = -1$
F₃	15 3	5 5	2	5 6	$u_3 = 0$ (Let)
F₄	4	15 3	25 1	4	$u_4 = -2$
	$V_1=3$	$V_2=5$	$V_3=3$	$V_4=6$	

For non-basic cells : $\bar{c}_{ij} = u_i + v_j - c_{ij}$

$$\bar{c}_{11} = -5,\ \bar{c}_{12} = -2,\ \bar{c}_{13} = -6,\ \bar{c}_{22} = 0,\ \bar{c}_{23} = -3,\ \bar{c}_{24} = 2,\ \bar{c}_{33} = 1,\ \bar{c}_{41} = -3,\ \bar{c}_{44} = 0.$$

Since all $\bar{c}_{ij}$ are not non-positive, the current solution is not optimal.

Select the cell (2, 4) due to largest positive value and assign an unknown quantity θ in that cell. Identify a loop and subtract and add θ to the corner points of the loop which is shown below :



3	2	4	20 1
15 −θ 2	4	5	3
15 +θ 3	5 5	2	5 −θ 6
4	15 3	25 1	4

Select θ = min. (5. 15) = 5. The cell (3, 4) leaves the basis and the cell (2, 4) enters into the basis. Thus the current solution is updated.

Iteration 2.

3	2	4	$\boxed{20}$ 1	$u_1 = -2$
$\boxed{10}$ 2	4	5	$\boxed{5}$ 3	$u_2 = 0$ (Let)
$\boxed{20}$ 3	$\boxed{5}$ 5	2	6	$u_3 = 1$
4	$\boxed{15}$ 3	$\boxed{25}$ 1	4	$u_4 = -1$
$V_1 = 2$	$V_2 = 4$	$V_3 = 2$	$V_4 = 3$	

For non-basic cells : $\bar{c}_{ij} = u_i + v_j - c_{ij}$

$\bar{c}_{11} = -3,\ \bar{c}_{12} = 0,\ \bar{c}_{13} = -4,\ \bar{c}_{22} = 0,\ \bar{c}_{23} = -3,\ \bar{c}_{33} = 1,\ \bar{c}_{34} = -2,\ \bar{c}_{41} = -3,\ \bar{c}_{44} = -2.$

Since all $\bar{c}_{ij}$ are not non-positive, the current solution is not optimal.

Select the cell (3, 3) due to largest positive value and assign an unknown quantity θ in that cell. Identify a loop and subtract and add θ to the corner points of the loop which is shown below :

3	2	4	$\boxed{20}$ 1
$\boxed{10}$ 2	4	5	$\boxed{5}$ 3
$\boxed{20}$ 3	$\boxed{5}-\theta$ 5	θ 2	6
4	$\boxed{15}+\theta$ 3	$\boxed{25}-\theta$ 1	4

Select θ = min. (5, 25) = 5. The cell (3, 2) leaves the basis and the cell (3, 3) enters into the basis. Thus the current solution is updated.

Iteration 3.

3	2	4	$\boxed{20}$ 1	$u_1 = -2$
$\boxed{10}$ 2	4	5	$\boxed{5}$ 3	$u_2 = 0$ (Let)
$\boxed{20}$ 3	5	$\boxed{5}$ 2	6	$u_3 = 1$
4	$\boxed{20}$ 3	$\boxed{20}$ 1	4	$u_4 = 0$
$V_1 = 2$	$V_2 = 3$	$V_3 = 1$	$V_4 = 3$	

For non-basic cells : $\bar{c}_{ij} = u_i + v_j - c_{ij}$

$\bar{c}_{11} = -3, \bar{c}_{12} = -1, \bar{c}_{13} = -5, \bar{c}_{22} = -1, \bar{c}_{23} = -5, \bar{c}_{32} = -1, \bar{c}_{34} = -2, \bar{c}_{41} = -2, \bar{c}_{44} = -1.$

Since all $\bar{c}_{ij}$ are non-positive, the current solution is optimal. Thus, the optimal solution is

$x_{14} = 20,\ x_{21} = 10,\ x_{24} = 5,\ x_{31} = 20,\ x_{33} = 5;\ x_{42} = 20,\ x_{43} = 20.$

The optimal transportation cost

$= 1 \times 20 + 2 \times 10 + 3 \times 5 + 3 \times 20 + 2 \times 5 + 3 \times 20 + 1 \times 20 =$ Rs. 205.

Example 5. *Consider the initial BFS by VAM of Example 3. find the optimal solution of the T.P.*

Solution. Iteration 1.

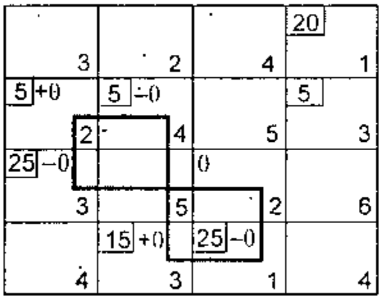
3	2	4	$\boxed{20}$ 1	$u_1 = -2$
$\boxed{5}$ 2	$\boxed{5}$ 4	5	$\boxed{5}$ 3	$u_2 = 0$ (Let)
$\boxed{25}$ 3	5	2	6	$u_3 = 1$
4	$\boxed{15}$ 3	$\boxed{25}$ 1	4	$u_4 = -1$
$V_1 = 2$	$V_2 = 4$	$V_3 = 2$	$V_4 = 3$	

For non-basic cells : $\bar{c}_{ij} = u_i + v_j - c_{ij}$

$\bar{c}_{11} = -3,\ \bar{c}_{12} = 0,\ \bar{c}_{13} = -4,\ \bar{c}_{23} = -3,\ \bar{c}_{32} = 0,\ \bar{c}_{33} = 1,\ \bar{c}_{34} = -2,\ \bar{c}_{41} = -3,\ \bar{c}_{44} = -2.$

Since all $\bar{c}_{ij}$ are not non-positive, the current solution is not optimal.

Select the cell (3, 3) due to largest positive value and assign an unknown quantity θ in that cell. Identify a loop and subtract and add θ to the corner points of the loop which is shown below :



3	2	4	$\boxed{20}$ 1
$\boxed{5}+\theta$ 2	$\boxed{5}-\theta$ 4	5	$\boxed{5}$ 3
$\boxed{25}-\theta$ 3	5	θ 2	6
4	$\boxed{15}+\theta$ 3	$\boxed{25}-\theta$ 1	4

Select θ = min. (5, 25, 25) = 5. The cell (2, 2) leaves the basis and the cell (3, 3) enters into the basis. Thus the current solution is updated.

Iteration 2.

3	2	4	$\boxed{20}$ 1	$u_1 = -2$
$\boxed{10}$ 2	4	5	$\boxed{5}$ 3	$u_2 = 0$ (Let)
$\boxed{20}$ 3	5	$\boxed{5}$ 2	6	$u_3 = 1$
4	$\boxed{20}$ 3	$\boxed{20}$ 1	4	$u_4 = 0$
$V_1 = 2$	$V_2 = 3$	$V_3 = 1$	$V_4 = 3$	

For non-basic cells : $\bar{c}_{ii} = u_i + v_j - c_{ij}$

$\bar{c}_{11} = -3, \bar{c}_{12} = -1, \bar{c}_{13} = -5, \bar{c}_{22} = -1, \bar{c}_{23} = -5, \bar{c}_{32} = -1, \bar{c}_{34} = -2, \bar{c}_{41} = -2, \bar{c}_{44} = -1.$

Since all $\bar{c}_{ij}$ are non-positive, the current solution is optimal. Thus the optimal solution is

$$x_{14} = 20,\ x_{21} = 10,\ x_{24} = 5,\ x_{31} = 20,\ x_{33} = 5,\ x_{42} = 20,\ x_{43} = 20.$$

The optimal transportation cost = Rs. 205.

NOTES

Note. To find optimal solution to a T.P., the number of iterations by uv-method is always more if we consider the initial BFS by NWC.

5.4 DEGENERACY IN T.P.

A BFS of a T.P. is said to be degenerate if one or more basic variables assume a zero value. This degeneracy may occur in initial BFS or in the subsequent iterations of *uv*-method. An initial BFS could become degenerate when the supply and demand in the intermediate stages of any one method (NWC/LCM/VAM) are equal corresponding to a selected cell for allocation. In *uv*-method it is identified only when more than one corner points in a loop vanishes due to minimum value of θ.

For the degeneracy in initial BFS, arbitrarily we can delete the row due to supply adjusting the demand to zero or delete the column due to demand adjusting the supply to zero whenever there is a tie in demand and supply.

For the degeneracy in *uv*-method, arbitrarily we can make one corner as non-basic cell and put zero in the other corner.

Example 6. *Find the optimal solution to the following T.P.* :

Source	Destination 1	 2	 3	Available
1	50	30	190	10
2	80	45	150	30
3	220	180	50	40
Requirement	40	20	20	80

Solution. Let us find the initial BFS using VAM :

	1	2	3	Row penalties
1	50	30	190	10 (20)
2	80	45	150	30 (35)
3	220	180	50	40 (130)
	40	20	20	
Column penalties	(30)	(15)	(100)	

Select (3, 3) cell for allocation and allocate min (40, 20) = 20 in that cell. Cross-off the third column as the requirement is met and adjust the availability to 20. The reduced table is given below :

	1	2	Row penalties
1	50	30	10 (20)
2	80	45	30 (35)
3	220	180	20 ㊵
Column penalties	40 (30)	20 (15)	

Select (3, 2) cell for allocation. Now there is a tie in allocation. Let us allocate 20 in (3, 2) cell and cross-off the second column and adjust the availability to zero. The reduced table is given below :

	1	
1	50	10
2	80	30
3	220	0
	40	

On continuation we obtain the remaining allocations as 0 in (3, 1) cell, 30 in (2, 1) cell and 10 in (1, 1) cell. The complete initial BFS is given below and let us apply the first iteration of *uv*-method :

Iteration 1.

$\boxed{10}$ 50	30	190	$u_1 = -170$
$\boxed{30}$ 80	45	150	$u_2 = -140$
$\boxed{0}$ 220	$\boxed{20}$ 180	$\boxed{20}$ 50	$u_3 = 0$ (Let)
$V_1 = 220$	$V_2 = 180$	$V_3 = 50$	

For non-basic cells : $\bar{c}_{ij} = u_i + v_j - c_{ij}$

$$\bar{c}_{12} = -20,\ \bar{c}_{13} = -310,\ \bar{c}_{22} = -5,\ \bar{c}_{23} = -240.$$

Since all $\bar{c}_{ij} < 0$, the current solution is optimal. Hence, the optimal solution is

$$x_{11} = 10,\ x_{21} = 30,\ x_{31} = 0,\ x_{32} = 20,\ x_{33} = 20.$$

The transportation cost

$$= 50 \times 10 + 80 \times 30 + 0 + 180 \times 20 + 50 \times 20$$

$$= \text{Rs. } 7500.$$

5.5 MAX-TYPE T.P.

Instead of unit cost in transportation table, unit profit is considered then the objective of the T.P. changes to maximize the total profits subject to supply and demand restrictions. Then this problem is called 'max-type' T.P.

To obtain optimal solution, we consider

$$\text{Loss} = -\text{ Profit}$$

and convert the max type transportation matrix to a loss matrix. Then all the methods described in the previous sections can be applied. Thus the optimal BFS obtained for the loss matrix will be the optimal BFS for the max-type T.P.

Example 7. *A company has three plants at locations A, B and C, which supply to four markets D, E, F and G. Monthly plant capacities are 500, 800 and 900 units respectively. Monthly demands of the markets are 600, 700, 400 and 500 units respectively. Unit profits (in rupees) due to transportation are given below :*

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

Determine an optimal distribution for the company in order to maximize the total transportation profits.

Solution. The given problem is balanced max type T.P. All profits are converted to losses by multiplying – 1.

	D	E	F	G	
A	–8	–5	–3	–6	500
B	–7	–4	–5	–2	800
C	–6	–8	–4	–2	900
	600	700	400	500	2200

The initial BFS by LCM is given below

500 –8	–5	–3	–6
100 –7	–4	400 –5	300 –2
–6	700 –8	–4	200 –2

To find optimal solution let us apply *uv*-method.

Iteration 1.

500 –θ –8	–5	–3	θ –6	$u_1 = -1$
100 +θ –7	–4	400 –5	300 –θ –2	$u_2 = 0$
–6	700 –8	–4	200 –2	$u_3 = 0$ (Let)
$V_1 = -7$	$V_2 = -8$	$V_3 = -5$	$V_4 = -2$	

For non-basic cells : $\bar{c}_{ij} = u_i + v_j - c_{ij}$

$$\bar{c}_{12} = -4,\ \bar{c}_{13} = -3,\ \bar{c}_{14} = 3,\ \bar{c}_{22} = -4,\ \bar{c}_{31} = -1,\ \bar{c}_{33} = -1.$$

Since all $\bar{c}_{ij}$ are not non-positive, the current solution is not optimal. Select the cell (1, 4) due to largest positive value and assign an unknown quantity θ in that cell. Identify a loop and subtract and add θ to the corner points of the loop which is shown above.

Select θ = min. (500, 300) = 300. The cell (2, 4) leaves the basis and the cell (1, 4) enters into the basis. Thus the current solution is updated.

Iteration 2.

200 −θ −8	−5	−3	300 +θ −6	$u_1 = -4$
400 −7	−4	400 −5	−2	$u_2 = -3$
0 −6	700 −8	−4	200 −θ −2	$u_3 = 0$ (Let)
$V_1 = -4$	$V_2 = -8$	$V_3 = -2$	$V_4 = -2$	

For non-basic cells,

$$\bar{c}_{12} = -7,\ \bar{c}_{13} = -3,\ \bar{c}_{22} = -7,\ \bar{c}_{24} = -3,\ \bar{c}_{31} = 2,\ \bar{c}_{33} = 2.$$

Since all the $\bar{c}_{ij}$ are not non-positive, the current solution is not optimal. There is a tie in largest positive values. Let us select the cell (3, 1) and assign an unknown quantity θ in that cell. Identify a loop and subtract and add θ to the corner points of the loop which is shown above.

Select θ = min. (200, 200) = 200. Since only one cell will leave the basis, let the cell (3, 3) leaves the basis and assign a zero in the cell (1, 1). The cell (3, 1) enters into the basis. Thus the current solution is updated.

Iteration 3.

0 −8	−5	−3	500 −6	$u_1 = -2$
400 −7	−4	400 −5	−2	$u_2 = -1$
200 −6	700 −8	−4	−2	$u_3 = 0$ (Let)
$V_1 = -6$	$V_2 = -8$	$V_3 = -4$	$V_4 = -4$	

For non-basic cells,

$$\bar{c}_{12} = -5,\ \bar{c}_{13} = -3,\ \bar{c}_{22} = -5,\ \bar{c}_{24} = -3,\ \bar{c}_{33} = 0,\ \bar{c}_{34} = -4.$$

Since all the $\bar{c}_{ij}$ are non-positive, the current solution is optimal.

Thus the optimal solution, which is degenerate, is

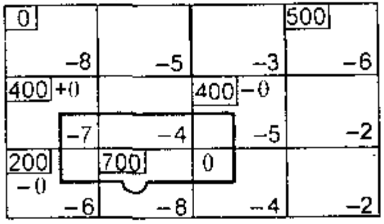
$$x_{11} = 0,\ x_{14} = 500,\ x_{21} = 400,\ x_{23} = 400,\ x_{31} = 200,\ x_{32} = 700.$$

The maximum transportation profit

$$= 0 + 3000 + 2800 + 2000 + 1200 + 5600 = \text{Rs. } 14600.$$

Since $\bar{c}_{33} = 0$, this indicates that there exists an alternative optimal solution. Assign an unknown quantity θ in the cell (3, 3). Identify a loop and subtract and add θ to the corner points of the loop which is shown below :

NOTES



0 −8	−5	−3	500 −6
400 +θ −7	−4	400 −θ −5	−2
200 −θ −6	700 −8	0 −4	−2

Select θ = min. (200, 400) = 200. The cell (3, 1) leaves the basis and the cell (3; 3) enters into the basis.

Iteration 4.

0 −8	−5	−3	500 −6	$u_1 = -2$
600 −7	−4	200 −5	−2	$u_2 = -1$
−6	700 −8	200 −4	−2	$u_3 = 0$ (Let)
$V_1 = -6$	$V_2 = -8$	$V_3 = -4$	$V_4 = -4$	

For non-basic cells,

$$\bar{c}_{12} = -5,\ \bar{c}_{13} = -3,\ \bar{c}_{22} = -5,\ \bar{c}_{24} = -3,\ \bar{c}_{31} = 0,\ \bar{c}_{34} = -2.$$

Since all the $\bar{c}_{ij}$ are non-positive, the current solution is optimal. Thus the alternative optimal solution is

$$x_{11} = 0,\ x_{14} = 500,\ x_{21} = 600,\ x_{23} = 200,\ x_{32} = 700,\ x_{33} = 200.$$

and the maximum transportation profit is Rs. 14,600.

5.6 UNBALANCED T.P.

If total supply ≠ total demand, the problem is called unbalanced T.P.. To obtain feasible solution, the unbalanced problem should be converted to balanced problem by introducing dummy source or dummy destination, whichever is required. Suppose, (supply =) $\sum a_i > \sum b_j$ (= demand). Then add one dummy destination with demand = $\left(\sum a_i - \sum b_j\right)$ with either zero transportation costs or some penalties, if they are given. Suppose (supply =) $\sum a_i < \sum b_j$ (= demand). Then add one dummy source with supply = $=\left(\sum b_j - \sum a_i\right)$ with either zero transportation costs or some penalties, if they are given.

After making it balanced the mathematical formulation is similar to the balanced T.P.

Example 8. *A company wants to supply materials from three plants to three new projects. Project I requires 50 truck loads, project II requires 40 truck loads and project III requires 60 truck loads. Supply capacities for the plants P_1, P_2 and P_3 are 30, 55 and 45 truck loads. The table of transportation costs are given below :*

	I	II	III
P_1	7	10	12
P_2	8	12	7
P_3	4	9	10

Determine the optimal distribution.

Solution. Here total supplies = 130 and total requirements = 150. The given problem is unbalanced T.P. To make it balanced consider a dummy plants with supply capacity of 20 truck loads and zero transportation costs to the three projects. Then the balanced T.P. is

		To I	II	III	
	P_1	7	10	12	30
From	P_2	8	12	7	55
	P_3	4	9	10	45
	P_4 (Dummy)	0	0	0	20
		50	40	60	

Using VAM, we obtain the initial BFS as given below :

5 7	20 10	5 12
8	12	55 7
45 4	9	10
0	20 0	0

To find optimal solution let us apply *uv*-method.

Iteration 1.

5 7	20 +0 10	5 –0 12	$u_1 = 0$ (Let)
8	12	55 7	$u_2 = -5$
45 4	9	10	$u_3 = -3$
0	20 –0 0	0 0	$u_4 = -10$
$V_1 = 7$	$V_2 = 10$	$V_3 = 12$	

For non-basic cells, $\bar{c}_{ij} = u_i + v_j - c_{ij}$

$\bar{c}_{21} = -6$, $\bar{c}_{22} = -7$, $\bar{c}_{32} = -2$, $\bar{c}_{33} = -1$, $\bar{c}_{41} = -3$, $\bar{c}_{43} = 2$,

NOTES

Since $\bar{c}_{43}$ is only positive value assign an unknown quantity θ in (4, 3) cell. Identify a loop and subtract and add θ to the corner points of the loop which is shown above.

Select $\theta = \min. (5, 20) = 5$ so that the cell (1, 3) leaves the basis and the cell (4, 3) enters into the basis.

Iteration 2.

$\boxed{5}$ 7	$\boxed{25}$ 10	12	$u_1 = 0$ (Let)
8	12	$\boxed{55}$ 7	$u_2 = -3$
$\boxed{45}$ 4	9	10	$u_3 = -3$
0	$\boxed{15}$ 0	$\boxed{5}$ 0	$u_4 = -10$
$V_1 = 7$	$V_2 = 10$	$V_3 = 10$	

For non-basic cells, we obtain

$\bar{c}_{13} = -2$, $\bar{c}_{21} = -4$, $\bar{c}_{22} = -5$, $\bar{c}_{32} = -2$, $\bar{c}_{33} = -3$, $\bar{c}_{41} = -3$

Since $\bar{c}_{ij} < 0$, the current solution is optimal. Thus the optimal solution is

Supply 15 truck loads from P_1 to I, 25 truck loads from P_1 to II, 55 truck loads from P_2 to III, 45 truck loads from P_3 to I. Demands of 15 truck loads for II and 5 truck loads for III will remain unsatisfied.

5.7 ASSIGNMENT PROBLEMS

Consider n machines $M_1, M_2, ..., M_n$ and n different jobs $J_1, J_2, ..., J_n$. These jobs to be processed by the machines one to one basis *i.e.*, each machine will process exactly one job and each job will be assigned to only one machine. For each job the processing cost depends on the machine to which it is assigned. Now we have to determine the assignment of the jobs to the machines one to one basis such that the total processing cost is minimum. This is called an *assignment problem.*

If the number of machines is equal to the number of jobs then the above problem is called *balanced* or *standard* assignment problem. Otherwise, the problem is called *unbalanced* or *non-standard* assignment problem. Let us consider a balanced assignment problem.

For linear programming problem formulation, let us define the decision variables as

$$x_{ij} = \begin{cases} 1, \text{ if job } j \text{ is assigned to machine } i \\ 0, \text{ otherwise} \end{cases}$$

and the cost of processing job j on machine i as c_{ij}. Then we can formulate the assignment problem as follows :

$$\text{Minimize } z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \qquad ...(1)$$

subject to, $$\sum_{j=1}^{n} x_{ij} = 1, \qquad i = 1, 2, ..., n$$

(Each machine is assigned exactly to one job)

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad j = 1, 2, ..., n.$$

(Each job is assigned exactly to one machine)

$$x_{ij} = 0 \text{ or } 1 \text{ for all } i \text{ and } j$$

In **matrix form**, $$\text{Minimize } z = Cx$$

$$\text{subject to, } Ax = 1,$$

$$x_{ij} = 0 \text{ or } 1,\ i, j = 1, 2, ..., n.$$

where A is a $2n \times n^2$ matrix and total unimodular *i.e.*, the determinant of every sub square matrix formed from it has value 0 or 1. This property permits us to replace the constraint $x_{ij} = 0$ or 1 by the constraint $x_{ij} \geq 0$. Thus we obtain

$$\text{Minimize } z = Cx$$

$$\text{subject to, } Ax = 1,\ x \geq 0$$

The **dual** of (1) with the non-negativity restrictions replacing the 0-1 constraints can be written as follows :

$$\text{Maximize } W = \sum_{i=1}^{n} u_i + \sum_{j=1}^{n} v_j$$

subject to, $u_i + v_j \leq c_{ij}$, $\qquad i, j = 1, 2, ..., n.$

u_i, v_j unrestricted in signs $\qquad i, j = 1, 2, ..., n.$

Example 9. *A company is facing the problem of assigning four operators to four machines. The assignment cost in rupees in given below :*

$$\begin{array}{cc|cccc|} & & & \text{Machine} & & \\ & & M_1 & M_2 & M_3 & M_4 \\ & I & 5 & 7 & - & 4 \\ \text{Operator} & II & 7 & 5 & 3 & 2 \\ & III & 9 & 4 & 6 & - \\ & IV & 7 & 2 & 7 & 6 \end{array}$$

In the above, operators I and III can not be assigned to the machines M_3 and M_4 respectively. Formulate the above problem as a LP model.

Solution. Let $x_{ij} = \begin{cases} 1, \text{if the } i\text{th operator is assigned to } j\text{th machine} \\ 0, \text{otherwise} \end{cases}$

$$i, j = 1, 2, 3, 4.$$

be the decision variables.

By the problem, $x_{13} = 0$ and $x_{34} = 0$.

The LP model is given below :

$$\text{Minimize } z = 5x_{11} + 7x_{12} + 4x_{14} + 7x_{21} + 5x_{22} + 3x_{23} + 2x_{24} + 9x_{31} + 4x_{32} + 6x_{33} + 7x_{41} + 2x_{42} + 7x_{43} + 6x_{44}.$$

subject to.

NOTES

(Operator assignment constraints)

$$x_{11} + x_{12} + x_{14} = 1$$
$$x_{21} + x_{22} + x_{23} + x_{24} = 1$$
$$x_{31} + x_{32} + x_{33} = 1$$
$$x_{41} + x_{42} + x_{43} + x_{44} = 1$$

(Machine assignment constraints)

$$x_{11} + x_{21} + x_{31} + x_{41} = 1$$
$$x_{12} + x_{22} + x_{32} + x_{42} = 1$$
$$x_{23} + x_{33} + x_{43} = 1$$
$$x_{14} + x_{24} + x_{44} = 1$$
$$x_{ij} \geq 0 \text{ for all } i \text{ and } j.$$

5.8 HUNGARIAN ALGORITHM

This is an efficient algorithm for solving the assignment problem developed by the Hungarian mathematician König. Here the optimal assignment is not affected if a constant is added or subtracted from any row or column of the balanced assignment cost matrix. The **algorithm** can be started as follows :

(*a*) Bring at least one zero to each row and column of the cost matrix by subtracting the minimum of the row and column respectively.

(*b*) Cover all the zeros in cost matrix by *minimum* number of horizontal and vertical lines.

(*c*) If number of lines = order of the matrix, then select the zeros as many as the order of the matrix in such a way that they cover all the rows and columns.

(Here $A_{n \times n}$ means *n*th order matrix)

(*d*) If number of lines ≠ order of the matrix, then perform the following and create a new matrix :

  (*i*) Select the minimum element from the uncovered elements of the cost matrix by the lines.

  (*ii*) Subtract the uncovered elements from the minimum element.

  (*iii*) Add the minimum element to the junction (*i.e.*, crossing of the lines) elements.

  (*iv*) Other elements on the lines remain unaltered.

  (*v*) Go to Step (*b*).

Example 10. *A construction company has four engineers for designing. The general manager is facing the problem of assigning four designing projects to these engineers. It is also found that Engineer 2 is not competent to design project 4. Given the time estimate required by each engineer to design a given project, find an assignment which minimizes the total time.*

		Projects			
		P1	P2	P3	P4
	E1	6	5	13	2
Engineers	E2	8	10	4	–
	E3	10	3	7	3
	E4	9	8	6	2

Solution. Let us first bring zeros rowwise by subtracting the respective minima from all the row elements respectively.

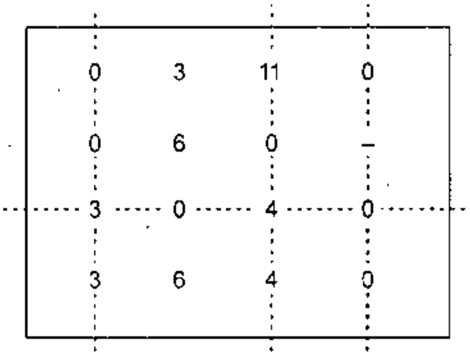
4	3	11	0
4	6	0	–
7	0	4	0
7	6	4	0

Let us bring zero columnwise by subtracting the respective minima from all the column elements respectively. Here the above operations is to be performed only on first column, since at least one zero has appeared in the remaining columns.

0	3	11	0
0	6	0	–
3	0	4	0
3	6	4	0

(This completes Step-*a*)

Now (Step-*b*) all the zeros are to be covered by minimum number of horizontal and vertical lines which is shown below. It is also to be noted that this covering is not unique.



0	3	11	0
0	6	0	–
3	0	4	0
3	6	4	0

It is seen that no. of lines = 4 = order of the matrix. Therefore by Step-*c*, we can go for assignment *i.e.*, we have to select 4 zeros such that they cover all the rows and columns which is shown below:

$\boxed{0}$	3	11	0
0	6	$\boxed{0}$	–
3	$\boxed{0}$	4	0
3	6	4	$\boxed{0}$

Therefore the optimal assignment is

E1→P1, E2→P3, E3→P2, E4→P4

and the minimum total time required = 6 + 4 + 3 + 2 = 15 units.

Example 11. *Solve the following job machine assignment problem. Cost data are given below :*

		Machines					
		1	2	3	4	5	6
	A	21	35	20	20	32	28
	B	30	31	22	25	28	30
	C	28	29	25	27	27	21
Jobs	D	30	30	26	26	31	28
	E	21	31	25	20	27	30
	F	25	29	22	25	30	21

Solution: Let us first bring zeros first rowwise and then columnwise by subtracting the respective minima elements from each row and each column respectively and the cost matrix, thus obtained, is as follows :

0	11	0	0	7	8
7	5	0	3	1	8
6	4	4	6	1	0
3	0	0	0	0	2
0	7	5	0	2	10
3	4	1	4	4	0

By Step-*b*, all the zeros are covered by minimum number of horizontal and vertical lines which is shown below :

NOTES

$$\begin{array}{|cccccc|}\hline 0 & 11 & 0 & 0 & 7 & 8 \\ 7 & 5 & 0 & 3 & 1 & 8 \\ 6 & 4 & 4 & 6 & 1 & 0 \\ 3 & 0 & 0 & 0 & 0 & 2 \\ 0 & 7 & 5 & 0 & 2 & 10 \\ 3 & 4 & 1 & 4 & 4 & 0 \\ \hline\end{array}$$

Here no. of lines ≠ order of the matrix. Hence, we have to apply Step-*d*. The minimum uncovered element is 1. By applying Step-*d* we obtain the following matrix :

$$\begin{array}{|cccccc|}\hline 0 & 11 & 1 & 0 & 7 & 9 \\ 6 & 4 & 0 & 2 & 0 & 8 \\ 5 & 3 & 4 & 5 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 3 \\ 0 & 7 & 6 & 0 & 2 & 11 \\ 2 & 3 & 1 & 3 & 3 & 0 \\ \hline\end{array}$$

Now, by Step-*b*, we cover all the zeros by minimum number of horizontal and vertical straight lines.

$$\begin{array}{|cccccc|}\hline 0 & 11 & 1 & 0 & 7 & 9 \\ 6 & 4 & 0 & 2 & 0 & 8 \\ 5 & 3 & 4 & 5 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 3 \\ 0 & 7 & 6 & 0 & 2 & 11 \\ 2 & 3 & 1 & 3 & 3 & 0 \\ \hline\end{array}$$

Now, the no. of lines = order of the matrix. So we can go for assignment by Step-*c*. The assignment is shown below :

$$\begin{array}{|cccccc|}\hline \boxed{0} & 11 & 1 & 0 & 7 & 9 \\ 6 & 4 & \boxed{0} & 2 & 0 & 8 \\ 5 & 3 & 4 & 5 & \boxed{0} & 0 \\ 3 & \boxed{0} & 1 & 0 & 0 & 3 \\ 0 & 7 & 6 & \boxed{0} & 2 & 11 \\ 2 & 3 & 1 & 3 & 3 & \boxed{0} \\ \hline\end{array}$$

The optimal assignment is A→1, B→3, C→5, D→2, E→4, F→6. An alternative assignment is also obtained as A→4, B→3, C→5, D→2, E→1, F→6. For both the assignments, the minimum cost is 21 + 22 + 27 + 30 + 20 + 21 *i.e.*, Rs. 141.

NOTES

5.9 UNBALANCED ASSIGNMENTS

For unbalanced or non-standard assignment problem no. of rows ≠ no. of columns in the assignment cost matrix *i.e.*, we deal with a rectangular cost matrix. To find an assignment for this type of problem, we have to first convert this unbalanced problem into a balanced problem by adding dummy rows or columns with zero costs so that the defective function will be unaltered. For machine-job problem, if no. of machines (say, *m*) > no. of jobs (say, *n*), then create *m-n* dummy jobs and the processing cost of dummy jobs as zero. When a dummy job gets assigned to a machine, that machine stays idle. Similarly the other case *i.e.*, *n* > *m*, is handled.

Example 12. *Find an optimal solution to an assignment problem with the following cost matrix :*

	M1	M2	M3	M4	M5
J1	13	5	20	5	6
J2	15	10	16	10	15
J3	6	12	14	10	13
J4	13	11	15	11	15
J5	15	6	16	10	6
J6	6	15	14	5	12

Solution. The above problem is unbalanced. We have to create a dummy machine M6 with zero processing time to make the problem as balanced assignment problem. Therefore we obtain the following :

	M1	M2	M3	M4	M5	M6 (dummy)
J1	13	5	20	5	6	0
J2	15	10	16	10	15	0
J3	6	12	14	10	13	0
J4	13	11	15	11	15	0
J5	15	6	16	10	6	0
J6	6	15	14	5	12	0

Let us bring zeros columnwise by subtracting the respective minima elements from each column respectively and the cost matrix, thus obtained, is as follows:

7	0	6	0	0	0
9	5	2	5	9	0
0	7	0	5	7	0
7	6	1	6	9	0
9	1	2	5	0	0
0	10	0	0	6	0

NOTES

Let us cover all the zeros by minimum number of horizontal and vertical lines and is given below :

7	0	6	0	0	0
9	5	2	5	9	0
0	7	0	5	7	0
7	6	1	6	9	0
9	1	2	5	0	0
0	10	0	0	6	0

Now, the number of lines ≠ order of the matrix. The minimum uncovered element by the lines is 1. Using Step-*d* of the Hungarian algorithm and covering all the zeros by minimum no. of lines we obtain as follows :

7	0	6	0	1	1
8	4	1	4	9	0
0	7	0	5	8	1
6	5	0	5	9	0
8	0	1	4	0	0
0	10	0	0	7	1

Now, the number of lines ≠ order of the matrix and we have to select 6 zeros such that they cover all the rows and columns. This is done in the following :

$$\begin{array}{cccccc}
7 & \boxed{0} & 6 & 0 & 1 & 1 \\
8 & 4 & 1 & 4 & 9 & \boxed{0} \\
\boxed{0} & 7 & 0 & 5 & 8 & 1 \\
6 & 5 & \boxed{0} & 5 & 9 & 0 \\
8 & 0 & 1 & 4 & \boxed{0} & 0 \\
0 & 10 & 0 & \boxed{0} & 7 & 1
\end{array}$$

Therefore, the optimal assignment is

J1→M2, J2→M6, J3→M1, J4→M3, J5→M5, J6→M4 and the minimum cost = Rs. (5 + 0 + 6 + 15 + 6 + 5) = Rs. 37.

In the above, the job J2 will not get processed since the machine M6 is dummy.

5.10 MAX-TYPE ASSIGNMENT PROBLEMS

When the objective of the assignment is to maximize, the problem is called 'Max-type assignment problem'. This is solved by converting the profit matrix to an opportunity loss matrix by subtracting each element from the highest element of the profit matrix. Then the minimization of the loss matrix is the same as the maximization of the profit matrix.

Example 13. *A company is faced with the problem of assigning 4 jobs to 5 persons. The expected profit in rupees for each person on each job are as follows :*

	Job			
Persons	J1	J2	J3	J4
I	86	78	62	81
II	55	79	65	60
III	72	65	63	80
IV	86	70	65	71
V	72	70	71	60

Find the assignment of persons to jobs that will result in a maximum profit.

Solution. The above problem is unbalanced max.-type assignment problem. The maximum element is 86. By subtracting all the elements from it obtain the following opportunity loss matrix.

$$\begin{bmatrix} 0 & 8 & 24 & 5 \\ 31 & 7 & 21 & 26 \\ 14 & 21 & 23 & 6 \\ 0 & 16 & 21 & 15 \\ 14 & 16 & 15 & 26 \end{bmatrix}$$

Now, a dummy job J5 is added with zero losses. Then bring zeros in each column by subtracting the respective minimum element from each column we obtain the following matrix.

$$\begin{bmatrix} 0 & 1 & 9 & 0 & 0 \\ 31 & 0 & 6 & 21 & 0 \\ 14 & 14 & 8 & 1 & 0 \\ 0 & 9 & 6 & 10 & 0 \\ 14 & 9 & 0 & 21 & 0 \end{bmatrix}$$

Let us cover all the zeros by minimum number of lines and is as follows:

$$\begin{bmatrix} 0 & 1 & 9 & 0 & 0 \\ 31 & 0 & 6 & 21 & 0 \\ 14 & 14 & 8 & 1 & 0 \\ 0 & 9 & 6 & 10 & 0 \\ 14 & 9 & 0 & 21 & 0 \end{bmatrix}$$

Since, the no. of lines ≠ order of the matrix, we have to select 5 zeros such that they cover all the rows and columns. This is done in the foilowing :

$$\begin{bmatrix} 0 & 1 & 9 & \boxed{0} & 0 \\ 31 & \boxed{0} & 6 & 21 & 0 \\ 14 & 14 & 8 & 1 & \boxed{0} \\ \boxed{0} & 9 & 6 & 10 & 0 \\ 14 & 9 & \boxed{0} & 21 & 0 \end{bmatrix}$$

The optimal assignment is

I→J4, II→J2, III→J5, IV→J1, V→J3 and maximum profit = Rs. (81 + 79 + 86 + 71) = Rs. 317. Here person III is idle.

Note. The max.-type assignment problem can also be converted to a minimization problem by multiplying all the elements of the profit matrix by – 1. Then the Hungarian method can be applied directly.

5.11 GAME THEORY

The mathematical theory of games was invented by John Von Neumann and Oskar Morgenstern (1944). Game theory is the study of the ways in which strategic interactions among rational players produce **outcomes** with respect to the **preferences** (or utilities) of those players, none of which might have been intended by any of them.

Game theory has found its applications in various fields such as Economics, Social Science, Political Science, Biology, Computer Science etc.

The famous example of a game is the **Prisoner's Dilemma** game. Suppose that *the police have arrested two people whom they know have committed an armed* robbery together. Unfortunately they lack enough admissible evidence to get a jury to convict. They do, however, have enough evidence to send each prisoner away for two years for theft. The chief inspector now makes the following offer to each prisoner. If you will confess the robbery implicating your partner and he does not also *confess, then you shall go free and he will get ten years. If you* both confess, you shall each get 5 years. If neither of you confess, then you shall each get two years for the theft.

5.12 BASIC DEFINITIONS

We assume that players are economically rational *i.e.*, a player can (*i*) assess outcomes, (*ii*) choose actions that yield their most preferred outcomes, given the actions of the other players.

(*i*) **Game :** All situations in which at least one player can only act to maximize his *utility through anticipating the responses to his actions by one or more other* players is called a game.

(*ii*) **Strategy :** A strategy is a possible course of action open to the player.

(*iii*) **Pure strategy :** A pure strategy is defined by a situation in which a course of action is played with probability one.

(*iv*) **Mixed strategy :** A mixed strategy is defined by a situation in which no course of action is taken with probability one.

(*v*) **Payoff matrix (or Reward matrix) :** A payoff matrix is an array in which any (*i, j*)th entry shows the outcome. Positive entry is the gain and negative entry is the loss for the row-player.

Matrix games are referred to as 'normal form' or 'strategic form' games, and games as trees are referred to as 'extensive form' games. The two sorts of games are not equivalent.

(vi) **Maximin criterion :** This is a criterion in which a player will choose the strategies with the largest possible payoff given an opponent's set of minimising countermoves.

(vii) **Minimax criterion :** This is a criterion in which a player will choose the strategies with the smallest possible payoff given an opponent's set of maximising countermoves.

(viii) **Saddle point :** If a payoff matrix has an entry that is simultaneously a maximum of row minima and a minimum of column maxima, then this entry is called a **saddle point** of the game and the game is said to be **strictly determined.**

(ix) **Value of the game :** If the game has a saddle point then the value at that entry is called the value of the game. If this value is zero then the game is said to be **fair.**

(x) **Zero-sum game :** A zero-sum game is a game in which the interests of the players are diametrically opposed *i.e.,* what one player wins the other loses. When two person play such game then it is called **two person zero-sum game.**

In this chapter we shall consider only matrix games.

Note. If in a game the total payoff to be divided among players is invariant *i.e.,* it does not depend upon the mix of strategies selected, then the game is called **constant-sum** game.

5.13 TWO-PERSON ZERO-SUM GAME WITH PURE STRATEGIES

To identify the saddle point and value of game the following procedure to be adopted on the payoff matrix :

(i) Identify the minimum from each row and place a symbol * in that cell/entry.

Take the maximum of these minima.

(ii) Identify the maximum from each column and place a symbol × in that cell/entry.

Take the minimum of these maxima.

(iii) If both the symbols * and × occurs in a/an cell/entry, then that cell/entry is called saddle point and the value in that cell/entry is called value of the game (v).

Also v = Maximum (row minima) = Minimum (column maxima). There may be more than one saddle point but the value of the game is unique.

Example 14. *Solve the following game :*

		Player B			
		B1	*B2*	*B3*	*B4*
	A1	1	5	4	2
Player A	*A2*	2	3	5	3
	A3	3	4	5	3

Solution. The calculations are displayed in the following table :

		Player B				
		B1	B2	B3	B4	Min.
	A1	1*	5×	4	2	1
Player A	A2	2*	3	5×	3×	2
	A3	3*×	4	5×	3*×	3
Max.		3	5	5	3	

Max. (Row Min.) = 3.

Min. (Column Max.) = 3

In the above game, there are two saddle points at (A3, B1) and (A3, B4).

The value of the game is 3. Here the optimal strategy for player A is A3 and the optimal strategy for player B is B1 and B4.

Example 15. *Determine the solution of the following game :*

		Player B				
		B1	*B2*	*B3*	*B4*	*B5*
	A1	*0*	*1*	*7*	*8*	*2*
Player A	*A2*	*6*	*4*	*5*	*5*	*4*
	A3	*7*	*3*	*2*	*1*	*2*
	A4	*1*	*4*	*1*	*4*	*5*

Solution. In the given game, player A has 4 strategies and player B has 5 strategies. The calculations are displayed in the following table :

	B1	*B2*	*B3*	*B4*	*B5*	Row Min.
A1	0*	1	7×	8×	2	0
A2	6	4*×	5	5	4*	4
A3	7×	3	2	1*	2	1
A3	1*	4×	1*	4	5×	1
Col. Max.	7	4	7	8	5	

Max. (Row Min.) = 4, Min. (Col Max.) = 4

In this game there is one saddle point at (A2, B2)

The value of the game is 4.

The optimal strategy for player A is A2

and the optimal strategy for player B is B2.

5.14 TWO-PERSON ZERO-SUM GAME WITH MIXED STRATEGIES

Consider the following game :

$$\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{\text{Player B}} \\
 & & \text{I} & \text{II} \\
\cline{3-4}
\text{Player A} & \text{I} & a_{11} & a_{12} \\
\cline{3-4}
 & \text{II} & a_{21} & a_{22} \\
\cline{3-4}
\end{array}$$

If this game does not have saddle point, then we assume that both players use mixed strategies.

Let player A select strategy I with probability p and strategy II with probability $1 - p$. Suppose player B select strategy I. then the expected gain to player A is given by $a_{11}\, p + a_{21}\, (1 - p)$.

If player B select strategy II, then the expected gain to player A is given by $a_{12}\, p + a_{22}\, (1 - p)$.

The optimal plan for player A requires that its expected gain to be equal for each strategies of player B. Thus we obtain

$$a_{11}\, p + a_{21}\, (1 - p) = a_{12}\, p + a_{22}\, (1 - p)$$

$$\Rightarrow \qquad p = \frac{a_{22} - a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}.$$

Similarly. let player B selects strategy I with probability q and strategy II with probability $1-q$. The expected loss to player B with respect to the strategies of player A are

$$a_{11}\, q + a_{12}\, (1 - q) \text{ and } a_{21}\, q + a_{22}\, (1 - q).$$

By equating the expected losses of player B we obtain

$$q = \frac{a_{22} - a_{12}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}.$$

The value of game v is found by substituting the value of p in one of the equations for the expected gain of A and on simplification, we obtain

$$v = \frac{a_{11}\, a_{22} - a_{12}\, a_{21}}{(a_{11} + a_{22}) - (a_{12} + a_{21})}.$$

Example 16. *Determine the solution of the following game :*

$$\begin{array}{ccc}
 & & \textit{Player B} \\
 & & \begin{array}{cc} B1 & B2 \end{array} \\
\textit{Player A} & \begin{array}{c} A1 \\ A2 \end{array} & \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}
\end{array}$$

Solution. Clearly the given game has no saddle point. So the players have to use mixed strategies.

NOTES

Let the mixed strategies for A as $S_A = \begin{pmatrix} A1 & A2 \\ p_1 & p_2 \end{pmatrix}$

where $p_2 = 1 - p_1$

and the mixed strategies for B as $S_B = \begin{pmatrix} B1 & B2 \\ q_1 & q_2 \end{pmatrix}$

where $q_2 = 1 - q_1$

$$\therefore \quad p_1 = \frac{4-2}{(3+4)-(1+2)} = \frac{2}{4} = \frac{1}{2},\ p_2 = 1 - p_1 = \frac{1}{2}$$

$$q_1 = \frac{4-1}{(3+4)-(1+2)} = \frac{3}{4},\ q_2 = 1 - q_1 = \frac{1}{4}$$

$$v = \frac{12-2}{(3+4)-(1+2)} = \frac{10}{4} = \frac{5}{2}.$$

Thus the optimal strategy for A is $S_A = \begin{pmatrix} A1 & A2 \\ 1/2 & 1/2 \end{pmatrix}$

and for B is $S_B = \begin{pmatrix} B1 & B2 \\ 3/4 & 1/4 \end{pmatrix}$

and the value of the game is 5/2.

5.15 DOMINANCE RULES

(a) For rows : (*i*) In the payoff matrix if all the entries in a row i_1 are **greater than or equal** to the corresponding entries of another row i_2, then row i_2 is said to be dominated by row i_1. In this situation row i_2 of the payoff matrix can be deleted.

e.g., $i_2 = (1, 2, -1)$ is dominated by $i_1 = (2, 2, 1)$, hence (1, 2, –1) can be deleted.

(*ii*) If sum of the entries of any two rows is greater than or equal to the corresponding entry of a third row, then that third row is said to be dominated by the above two rows and hence third row can be deleted.

(b) For columns : (*i*) In the payoff matrix if all the entries in a column j_1 are less than or equal to the corresponding entries of another column j_2, then column j_2 is said to be dominated by column j_1. In this situation column j_2 of the payoff matrix can be deleted.

e.g., $j_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ is dominated by $j_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Hence $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ can be deleted.

(*ii*) If sum of the entries of any two columns is less than or equal to the corresponding entry of a third column, then that third column is said to be dominated by the above two columns and hence third column can be deleted.

Example 17. *Using the rules for dominance solve the following game :*

$$\begin{array}{cc} & \textit{Player B} \\ & \begin{array}{ccc} I & II & III \end{array} \\ \textit{Player A} \begin{array}{c} I \\ II \\ III \end{array} & \begin{bmatrix} 5 & 2 & -2 \\ 2 & 3 & -1 \\ 3 & -2 & 3 \end{bmatrix} \end{array}$$

Solution. The given game has no saddle point. Let us apply the rules for dominance. It is observed that column 1 is dominated by column 3. Hence delete column 1 and the payoff matrix is reduced as follows :

$$\begin{array}{cc} & \begin{array}{cc} \text{II} & \text{III} \end{array} \\ \begin{array}{c} \text{I} \\ \text{II} \\ \text{III} \end{array} & \begin{bmatrix} 2 & -2 \\ 3 & -1 \\ -2 & 3 \end{bmatrix} \end{array}$$

Again row 1 is dominated by row 2. Hence delete row 1 and the payoff matrix is reduced to a 2 × 2 matrix.

$$\begin{array}{cc} & \begin{array}{cc} \text{II} & \text{III} \end{array} \\ \begin{array}{c} \text{II} \\ \text{III} \end{array} & \begin{bmatrix} 3 & -1 \\ -2 & 3 \end{bmatrix} \end{array}$$

Let the mixed strategy for player A be $S_A = \begin{pmatrix} \text{I} & \text{II} & \text{III} \\ 0 & p_1 & p_2 \end{pmatrix}$ with $p_2 = 1 - p_1$ and the mixed strategy for player B be

$$S_B = \begin{pmatrix} \text{I} & \text{II} & \text{III} \\ 0 & q_1 & q_2 \end{pmatrix} \text{ with } q_2 = 1 - q_1$$

$$\therefore \quad p_1 = \frac{3-(-2)}{(3+3)-(-1-2)} = \frac{5}{9}, p_2 = \frac{4}{9}$$

$$q_1 = \frac{3-(-1)}{(3+3)-(-1-2)} = \frac{4}{9}, q_2 = \frac{5}{9}$$

$$v = \frac{9-2}{(3+3)-(-1-2)} = \frac{7}{9}.$$

Hence the optimal mixed strategies are

$$S_A = \begin{pmatrix} \text{I} & \text{II} & \text{III} \\ 0 & 5/9 & 4/9 \end{pmatrix},$$

$$S_B = \begin{pmatrix} \text{I} & \text{II} & \text{III} \\ 0 & 4/9 & 5/9 \end{pmatrix},$$

$$v = 7/9.$$

5.16 GRAPHICAL METHOD FOR GAMES

(*a*) Let us consider a 2 × *n* game *i.e.*, the payoff matrix will consist of 2 rows and *n* columns. So player A (or, row-player) will have two strategies. Also assume that there is no saddle point. Then the problem can be solved by using the following procedure :

(*i*) Reduce the size of the payoff matrix using the rules of dominance, if it is applicable.

(*ii*) Let p be the probability of selection of strategy I and $1 - p$ be the probability of selection of strategy II by player A.

Write down the expected gain function of player A with respect to each of the strategies of player B.

(*iii*) Plot the gain functions on a graph. Keep the gain function on y-axis and p on x-axis. Here p will take the value 0 and 1.

(*iv*) Find the **highest intersection point** in the **lower boundary** (*i.e.*, lower envelope) of the graph. Since player A is a maximin player, then this point will be a maximin point.

(*v*) If the number of lines passing through the maximin point is only two, then obtain a 2 × 2 payoff matrix by retaining the columns corresponding to these two lines. Go to step (*vii*) else go to step (*vi*).

(*vi*) If more than two lines passing through the maximin point then identify two lines with opposite slopes and form the 2 × 2 payoff matrix as described in step (*v*).

(*vii*) Solve the 2 × 2 game.

Example 18. *Consider the following game and solve it using graphical method.*

$$\begin{array}{cc|ccccc|} & & \multicolumn{5}{c}{\textit{Player B}} \\ & & I & II & III & IV & V \\ \hline \textit{Player A} & I & 3 & 1 & 6 & -1 & 5 \\ & II & -2 & 4 & -1 & 2 & 1 \\ \hline \end{array}$$

Solution. It is observed that there is no saddle point. Column V is dominated by column I and column II is dominated by column IV. Therefore delete column V and column II and the payoff matrix is reduced as follows :

$$\begin{array}{cc|ccc|} & & \multicolumn{3}{c}{\text{Player B}} \\ & & \text{I} & \text{III} & \text{IV} \\ \hline \text{Player A} & \text{I} & 3 & 6 & -1 \\ & \text{II} & -2 & -1 & 2 \\ \hline \end{array}$$

Let p be the probability of selection of strategy I and $(1-p)$ be the probability of selection of strategy II by player A. Therefore, the expected gain (or payoff) function to player A with respect to different strategies of player B is given below :

B's strategy	A's expected gain function	A's expected gain $p = 0$	$p = 1$
I	$3p - 2(1 - p) = 5p - 2$	-2	3
III	$6p - (1 - p) = 7p - 1$	-1	6
IV	$-p + 2(1 - p) = -3p + 2$	2	-1

Now the A's expected gain function is plotted in Fig. It is observed that line I and IV passes through the highest point of the lower boundary. Hence we can form 2×2 payoff matrix by taking the columns due to I and IV for player A and it is displayed below :

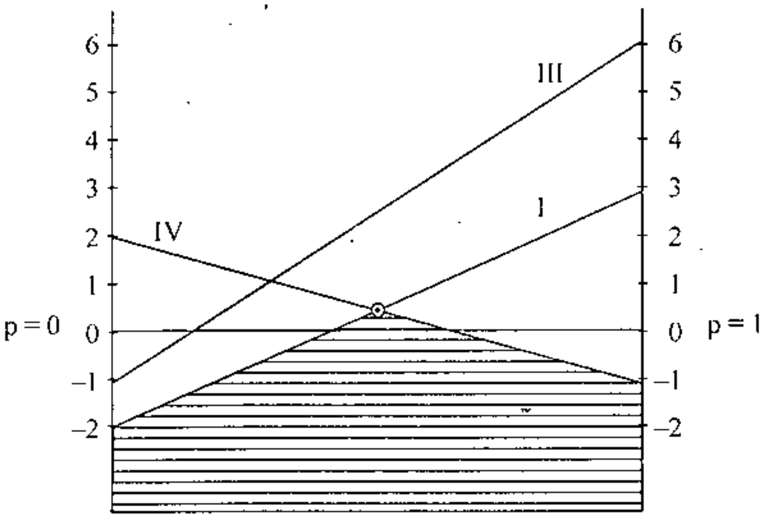


Fig. 5.2

$$\begin{array}{cc|cc|}
 & & \multicolumn{2}{c}{\text{Player B}} \\
 & & \text{I} & \text{IV} \\
\text{Player A} & \text{I} & 3 & -1 \\
 & \text{II} & -2 & 2
\end{array}$$

Let the mixed strategies for A be $S_A = \begin{pmatrix} \text{I} & \text{II} \\ p_1 & p_2 \end{pmatrix}$

with $p_2 = 1 - p_1$

and the mixed strategies for B be $S_B = \begin{pmatrix} \text{I} & \text{II} & \text{III} & \text{IV} & \text{V} \\ q_1 & 0 & 0 & q_2 & 0 \end{pmatrix}$

with $q_2 = 1 - q_1$

Therefore,

$$p_1 = \frac{2 - (-2)}{(3 + 2) - (-1 - 2)} = \frac{1}{2}, p_2 = 1 - p_1 = \frac{1}{2}$$

$$q_1 = \frac{2-(-1)}{(3+2)-(-1-2)} = \frac{3}{8}, q_2 = 1 - q_1 = \frac{5}{8}$$

$$v = \frac{6-2}{(3+2)-(-1-2)} = \frac{1}{2}$$

$\therefore$The optimal mixed strategies for A is

$$S_A = \begin{pmatrix} \text{I} & \text{II} \\ 1/2 & 1/2 \end{pmatrix},$$

the optimal mixed strategies for B is

$$S_B = \begin{pmatrix} \text{I} & \text{II} & \text{III} & \text{IV} & \text{V} \\ 3/8 & 0 & 0 & 5/8 & 0 \end{pmatrix}$$

and value of game $= \frac{1}{2}$.

(*b*) Let us consider a $m \times 2$ game *i.e.*, the payoff matrix will consist of m rows and 2 columns. Also assume that there is no saddle point. Then the problem can be solved by using the following procedure :

(*i*) Reduce the size of the payoff matrix using the rules of dominance, if it is applicable.

(*ii*) Let q be the probability of selection of strategy I and $1-q$ be the probability of selection of strategy II by the player B.

Write down the expected gain function of player B with respect to each of the strategies of player A.

(*iii*) Plot the gain functions on a graph. Keep the gain function on y-axis and q on x-axis. Here q will take the value 0 and 1.

(*iv*) Find the **lowest intersection point** in the **upper boundary** (*i.e.*, upper envelope) of the graph. Since player B is a minimax player, then this point will be a minimax point.

(*v*) If the number of lines passing through the minimax point is only two, then obtain a 2×2 payoff matrix by retaining the rows corresponding to these two lines. Go to step (*vii*) else goto step (*vi*).

(*vi*) If more than two lines passing through the minimax point then identify two lines with opposite slopes and form a 2×2 payoff matrix as described in step (*v*).

(*vii*) Solve the 2×2 game.

5.17 LINEAR PROGRAMMING METHOD FOR GAMES

The linear programming method is used in solving mixed strategies games of dimensions greater than (2×2) size. Consider an $m \times n$ payoff matrix in which player A (*i.e.*, the row player) has m strategies and player B (*i.e.*, the column player) has n strategies. The elements of payoff matrix be $\{(a_{ij})\ ;\ i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, n\}$.

Let p_i be the probability of selection of strategy i by player A and q_j be the probability of selection of strategy j by player B.

NOTES

LPP FOR PLAYER A

B's strategy	Expected gain function for A
1	$\sum_{i=1}^{m} a_{i1}\, p_i$
2	$\sum_{i=1}^{m} a_{i2}\, p_i$
- - -	- - -
n	$\sum_{i=1}^{m} a_{in}\, p_i$

Let
$$v = \text{Min.}\left\{\sum_i a_{i1}\, p_i,\ \sum_i a_{i2}\, p_i, \ldots, \sum_i a_{in}\, p_i\right\}$$

Since the player A is maximin type, the LPP can be written as follows :

Maximize v

Subject to,
$$\sum_i a_{i1}p_i \geq v$$
$$\sum_i a_{i2}p_i \geq v$$
— — — — — —
$$\sum_i a_{in}p_i \geq v$$
$$p_1 + p_2 + \ldots\ldots + p_m = 1$$
$$\text{all } p_i \geq 0$$

$\Rightarrow$ Maximize v

Subject to,
$$\sum_i a_{i1}(p_i / v) \geq 1$$
$$\sum_i a_{i2}(p_i / v) \geq 1$$
— — — — — —
$$\sum_i a_{in}(p_i / v) \geq 1$$
$$\frac{p_1}{v} + \frac{p_2}{v} + \ldots\ldots + \frac{p_m}{v} = 1$$
$$\text{all } p_i \geq 0$$

Set $$p_i/v = x_i,\ i = 1, 2, ..., m. \text{ Therefore}$$

$$\text{Maximize } v = \text{Minimize}\left(\frac{1}{v}\right)$$

$$= \text{Minimize}\left(\frac{p_1}{v} + \frac{p_2}{v} + ... + \frac{p_m}{v}\right)$$

$$= \text{Minimize } (x_1 + x_2 + ... + x_m)$$

Subject to, $$\sum_i a_{i1}\, x_i \geq 1$$

$$\sum_i a_{i2}\, x_i \geq 1$$

— — — — —

$$\sum_i a_{im}\, x_i \geq 1$$

and $$x_i \geq 0,\ i = 1, 2, ..., m.$$

LPP FOR PLAYER B

A's strategy	Expected loss/gain function to B
1	$\sum_j a_{1j}\, q_j$
2	$\sum_j a_{2j}\, q_j$
m	$\sum_j a_{mj}\, q_j$

Let $$u = \text{Max.}\left\{\sum_j a_{1j}\, q_j, \sum_j a_{2j}\, q_j, ..., \sum_j a_{mj}\, q_j\right\}$$

Since the player B is minimax type, the LPP can be written as follows : Minimize u

Subject to, $$\sum a_{1j}\, q_j \leq u$$

$$\sum_j a_{2j}\, q_j \leq u$$

— — — — —

$$\sum_j a_{mj}\, q_j \leq u$$

$$q_1 + q_2 + ... + q_n = 1$$

all $$q_j \geq 0$$

$\Rightarrow$ Minimize u

Subject to, $$\sum_j a_{1j}\left(q_j / u\right) \leq 1$$

$$\sum_j a_{2j}\left(q_j / u\right) \le 1$$

— — — — — — —

$$\sum_j a_{mj}\left(q_j / u\right) \le 1$$

$$\frac{q_1}{u}+\frac{q_2}{u}+\dots+\frac{q_n}{u} = 1$$

all $q_j \ge 0$

Set $q_j/u = y_j, j = 1, 2, \dots n$. Therefore

$$\text{Minimize } u = \text{Maximize}\left(\frac{1}{u}\right)$$

$$= \text{Maximize}\left(\frac{q_1}{u}+\frac{q_2}{u}+\dots+\frac{q_n}{u}\right)$$

$$= \text{Maximize } (y_1 + y_2 + \dots + y_n)$$

Subject to,

$$\sum_j a_{1j} y_j \le 1$$

$$\sum_j a_{2j} y_j \le 1$$

— — — — —

$$\sum_j a_{mj} y_j \le 1$$

and $y_j \ge 0, \quad j = 1, 2, \dots, n.$

Note. 1. In the above approach we may face two problems. Value of the game may be zero or less than zero. First case constraints will become infinite and in the second case, the type of each constraint will get changed. Therefore to obtain a non-negative value of the game, a constant c = Max. {abs. (negative values)} + 1 is to be added to each elements in the payoff matrix. The optimal strategy will not change. However the value of the original game will be the value of the new game minus constant.

2. The above LPP formulations for player A and B are primal-dual pair. So solving one problem, we can read the solution of the other problem from the optimal table.

Example 19. *Solve the following game by linear programming technique :*

$$\begin{array}{cc} & \text{Player } B \\ \text{Player } A & \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & 2 \\ 1 & 1 & -1 \end{bmatrix} \end{array}$$

Solution. The game has no saddle point. Since the payoff matrix has negative values, let us add a constant $c = 2$ to each element. The revised payoff matrix is given below :

$$\begin{array}{cc} & \text{Player B} \\ \text{Player A} & \begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & 4 \\ 3 & 3 & 1 \end{bmatrix} \end{array}$$

NOTES

Let the strategies of the two players be

$$S_A = \begin{pmatrix} I & II & III \\ p_1 & p_2 & p_3 \end{pmatrix}, \quad S_B = \begin{pmatrix} I & II & III \\ q_1 & q_2 & q_3 \end{pmatrix}$$

where $p_1 + p_2 + p_3 = 1$ and $q_1 + q_2 + q_3 = 1$.

The LPP for player A:

$$\text{Maximize } v = \text{Minimize } \frac{1}{v} = x_1 + x_2 + x_3$$

Subject to,

$$\begin{aligned} x_1 + x_2 + 3x_3 &\geq 1 \\ x_1 + 3x_2 + 3x_3 &\geq 1 \\ 3x_1 + 4x_2 + x_3 &\geq 1 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

where $x_j = p_j/v,\ j = 1, 2, 3.$

The LPP for player B:

$$\text{Minimize } u = \text{Maximize } \frac{1}{u} = y_1 + y_2 + y_3$$

Subject to,

$$\begin{aligned} y_1 + y_2 + 3y_3 &\leq 1 \\ y_1 + 3y_2 + 4y_3 &\leq 1 \\ 3y_1 + 3y_2 + y_3 &\leq 1 \\ y_1, y_2, y_3 &\geq 0 \end{aligned}$$

where $y_j = q_j/u,\ j = 1, 2, 3.$

Let us now solve the problem for player B.

The standard form can be written as follows :

$$\text{Maximize } \frac{1}{u} = y_1 + y_2 + y_3 + 0.s_1 + 0.s_2 + 0.s_3$$

Subject to,

$$\begin{aligned} y_1 + y_2 + 3y_3 + s_1 &= 1 \\ y_1 + 3y_2 + 4y_3 + s_2 &= 1 \\ 3y_1 + 3y_2 + y_3 + s_3 &= 1 \\ y_1, y_2, y_3 &\geq 0,\ s_1, s_2, s_3 \text{ slacks} \geq 0. \end{aligned}$$

Iteration 1

		c_j	1	1	1	0	0	0	Min.
c_B	x_B	Soln.	y_1	y_2	y_3	s_1	s_2	s_3	Ratio
0	s_1	1	1	1	3	1	0	0	1
0	s_2	1	1	3	4	0	1	0	1
0	s_3	1	3	3	1	0	0	1	1/3 →
$z_j - c_j$			−1	−1	−1	0	0	0	
			↑						

NOTES

Iteration 2

		c_j	1	1	1	0	0	0	Min.
c_B	x_B	Soln.	y_1	y_2	y_3	s_1	s_2	s_3	Ratio
0	s_1	2/3	0	0	8/3	1	0	–1/3	$\frac{2}{8}=\frac{1}{4}$
0	s_2	2/3	0	2	11/3	0	1	–1/3	2/11
1	y_1	1/3	1	1	1/3	0	0	1/3	1
$z_j - c_j$			0	0	–2/3	0	0	1/3	
					↑				

Iteration 3

		c_j	1	1	1	0	0	0	Min.
c_B	x_B	Soln.	y_1	y_2	y_3	s_1	s_2	s_3	Ratio
0	s_1	2/11	0	–16/11	0	1	–8/11	–1/11	
1	y_3	2/11	0	6/11	1	0	3/11	–1/11	
1	y_1	3/11	1	9/11	0	0	–1/11	4/11	
$z_j - c_j$			0	4/11	0	0	2/11	3/11	

$\therefore\ y_1^* = \frac{3}{11}, y_2^* = 0, y_3^* = \frac{2}{11}$, Max. $\frac{1}{u} = \frac{5}{11} \Rightarrow u^* = \frac{11}{5} = v^*$,

original $u^* = \frac{11}{5} - 2 = \frac{1}{5}$ = original v^*

Using duality, $x_1^* = 0,\ x_2^* = \frac{2}{11}, x_3^* = \frac{3}{11}$.

Now,

$$q_1 = y_1.u = \frac{3}{11}.\frac{11}{5} = \frac{3}{5}, \quad p_1 = x_1 . v = 0$$

$$q_2 = y_2.u = 0, \quad p_2 = x_2.v = \frac{2}{11}.\frac{11}{5} = \frac{2}{5}$$

$$q_3 = y_3.u = \frac{2}{11}.\frac{11}{5} = \frac{2}{5}, \quad p_3 = x_3.v = \frac{3}{11}.\frac{11}{5} = \frac{3}{5}$$

$$\therefore \quad S_A = \begin{pmatrix} I & II & III \\ 0 & 2/5 & 3/5 \end{pmatrix},$$

$$S_B = \begin{pmatrix} I & II & III \\ 3/5 & 0 & 2/5 \end{pmatrix} \text{ and } v^* = \frac{1}{5}.$$

SUMMARY

- Transportation problem (T.P.) is generally concerned with the distribution of a certain commodity/product from several origins/sources to several destinations with minimum total cost through single mode of transportation.
- A BFS of a T.P. is said to be degenerate if one or more basic variables assume a zero value. This degeneracy may occur in initial BFS or in the subsequent iterations of *uv*-method.
- For the degeneracy in initial BFS, arbitrarily we can delete the row due to supply adjusting the demand to zero or delete the column due to demand adjusting the supply to zero whenever there is a tie in demand and supply.
- If total supply ≠ total demand, the problem is called unbalanced T.P.. To obtain feasible solution, the unbalanced problem should be converted to balanced problem by introducing dummy source or dummy destination, whichever is required.
- This is an efficient algorithm for solving the assignment problem developed by the Hungarian mathematician König. Here the optimal assignment is not affected if a constant is added or subtracted from any row or column of the balanced assignment cost matrix.
- When the objective of the assignment is to maximize, the problem is called 'Max-type assignment problem'. This is solved by converting the profit matrix to an opportunity loss matrix by subtracting each element from the highest element of the profit matrix.
- There are various types of routing problems which occurs in a network. The most widely discussed problem is the 'Travelling Salesman Problem (TSP)'.
- All situations in which at least one player can only act to maximize his utility through anticipating the responses to his actions by one or more other players is called a game.
- A mixed strategy is defined by a situation in which no course of action is taken with probability one.
- If a payoff matrix has an entry that is simultaneously a maximum of row minima and a minimum of column maxima, then this entry is called a **saddle point** of the game and the game is said to be **strictly determined.**
- If the game has a saddle point then the value at that entry is called the value of the game.
- A zero-sum game is a game in which the interests of the players are diametrically opposed *i.e.*, what one player wins the other loses. When two person play such game then it is called **two person zero-sum game.**
- The linear programming method is used in solving mixed strategies games of dimensions greater than (2 × 2) size.

PROBLEMS

1. There are three sources which store a given product. The sources supply these products to four dealers. The capacities of the sources and the demands of the dealers are given. Capacities $S_1 = 150$, $S_2 = 40$, $S_3 = 80$, Demands $D_1 = 90$, $D_2 = 70$, $D_3 = 50$, $D_4 = 60$. The cost matrix is given as follows :

To

		D_1	D_2	D_3	D_4
	S_1	27	23	31	69
From	S_2	10	45	40	32
	S_3	30	54	35	57

Find the minimum cost of T.P.

2. There are three factories F_1, F_2, F_3 situated in different areas with supply capacities as 200, 400 and 350 units respectively. The items are shipped to five markets M_1, M_2, M_3, M_4 and M_5 with demands as 150, 120, 230, 200, 250 units respectively. The cost matrix is given as follows :

	M_1	M_2	M_3	M_4	M_5
F_1	2	5	6	4	7
F_2	4	3	5	8	8
F_3	4	6	2	1	5

Determine the optimal shipping cost and shipping patterns.

3. Find the initial basic feasible solution to the following T.P. using (*a*) NWC, (*b*) LCM, and (*c*) VAM :

(*i*)

To

		D	E	F	G	H	
	A	11	7	5	8	9	50
From	B	10	11	8	4	5	90
	C	9	6	12	5	5	60
		20	40	20	40	80	

(*ii*)

To

		A	B	C	D	E	
	I	9	10	0	8	9	90
	II	11	12	5	8	3	20
From	III	4	9	1	2	0	50
	IV	8	0	3	5	6	50
		80	60	20	40	10	

4. Identical products are produced in four factories and sent to four warehouses for delivery to the customers. The costs of transportation, capacities and demands are given as below :

		Warehouses				
		W_1	W_2	W_3	W_4	Capacities
Factories	F_1	9	6	11	5	200
	F_2	4	5	8	5	150
	F_3	7	8	4	6	350
	F_4	3	3	10	10	250
Demands		260	100	340	200	

Find the optimal schedule of delivery for minimization of cost of transportation. Is there any alternative solution ? If yes, then find it.

5. Starting with LCM initial BFS, find the optimal solution to the following T.P. problem :

			To			
From	5	1	2	4	3	60
	1	4	2	3	6	55
	4	2	3	5	2	40
	3	5	6	3	7	50
Demands	42	33	41	52	27	Supply

6. A company manufacturing air coolers has two plants located at Mumbai and Kolkata with a weekly capacity of 200 units and 100 units respectively. The company supplies air coolers to its 4 show-rooms situated at Ranchi, Delhi, Lucknow and Kanpur which have a demand of 75, 100, 100 and 30 units respectively. The cost per unit (in Rs.) is shown in the following table :

	Ranchi	*Delhi*	*Lucknow*	*Kanpur*
Mumbai	90	90	100	100
Kolkata	50	70	130	85

Plan the production programmes so as to minimize the total cost of transportation.

7. Solve the following assignment problems :

(*a*)

	A	B	C	D	E
I	12	20	20	18	17
II	20	12	5	11	8
III	20	5	12	5	9
IV	18	11	5	12	10
V	17	8	9	10	12

(*h*)

		Jobs					
		J1	J2	J3	J4	J5	J6
	A	18	10	25	10	11	22
	B	20	15	21	–	20	18
	C	11	17	19	15	18	17
Persons	D	18	16	20	16	20	21
	E	20	–	21	15	11	17
	F	11	15	19	12	15	20

8. A machine tool decides to make six sub-assemblies through six contractors A, B, C, D, E and F. Each contractor is to receive only one sub-assembly from A1, A2, A3, A4, A5 and A6. But the contractors C and E are not competent for the A4 and A2 assembly respectively. The cost of each subassembly by the bids submitted by each contractor is shown below (in hundred rupees) :

	A1	A2	A3	A4	A5	A6
A	15	10	11	18	13	22
B	9	12	18	10	14	11
C	9	15	11	–	22	11
D	14	13	9	12	15	10
E	10	–	11	22	13	18
F	10	14	15	12	13	14

Find the optimal assignments of the assemblies to contractors so as to minimize the total cost.

9. Five programmers, in a computer centre, write five programmes which run successfully but with different times. Assign the programmers to the programmes in a such a way that the total time taken by them is minimum taking the following time matrix:

		Programmes				
		P1	P2	P3	P4	P5
	A	80	66	65	65	73
	B	76	75	70	70	75
Programmers	C	74	73	72	70	66
	D	75	75	71	71	73
	E	76	66	66	70	75

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

		Jobs						
		I	II	III	IV	V	VI	VII
	A	9	6	12	11	13	15	11
	B	14	13	14	14	10	20	15
	C	18	6	17	11	15	13.	11
Persons	D	10	11	12	15	15	14	13
	E	15	6	18	15	10	14	12
	F	9	18	15	20	14	13	11
	G	14	15	12	13	11	17	20

Determine the optimal assignment schedule.

11. A construction company has to move six large cranes from old construction sites to new construction sites. The distances (in miles) between the old and the new sites are given below :

		New sites				
		A	B	C	D	E
	I	12	9	8	11	7
	II	11	10	8	12	7
Old sites	III	9	12	7	6	9
	IV	9	8	11	10	10
	V	10	9	9	6	11
	VI	11	11	7	8	9

Determine a plan for moving the cranes such that the total distance involved in the move will be minimum.

12. A company wants to assign five salesperson to five different regions to promote a product. The expected sales (in thousand) are given below :

		Regions				
		I	II	III	IV	V
	S1	27	54	37	100	85
	S2	55	66	45	80	32
Salesperson	S3	72	58	74	80	85
	S4	39	88	74	59	72
	S5	72	66	45	69	85

Solve the above assignment problem to find the maximum total expected sale.

13. A company makes profit (Rs.) while processing different jobs on different machines (one machine to one job only). Now, the company is facing problem of assigning 4 machines to 5 jobs. The profits are estimated as given below :

		Job				
		J1	J2	J3	J4	J5
	A	21	16	35	42	16
	B	15	20	30	35	15
Machine	C	20	16	30	27	18
	D	15	18	32	27	15

Determine the optimal assignment for maximum total profits.

ANSWERS

1. $x_{11} = 30, x_{12} = 70, x_{13} = 50, x_{24} = 40, x_{31} = 60, x_{34} = 20.$

   Minimum T.P. cost = Rs. 8190.

2. Solution I:

   $x_{11} = 150, x_{15} = 50, x_{22} = 120, x_{23} = 80, x_{25} = 200, x_{33} = 150, x_{34} = 200.$

Solution 2 :

$x_{11} = 150, x_{15} = 50, x_{22} = 120, x_{23} = 230, x_{25} = 50, x_{34} = 200, x_{35} = 150.$

Minimum shipping cost = Rs. 3510.

3. (*i*) (*a*) $x_{11} = 20, x_{12} = 30, x_{22} = 10, x_{23} = 20, x_{24} = 40, x_{25} = 20, x_{35} = 60.$

T.P. cost = Rs. 1260.

(*b*) $x_{11} = 20, x_{12} = 10, x_{13} = 20, x_{24} = 40, x_{25} = 50, x_{32} = 30, x_{35} = 30.$

T.P. cost = Rs. 1130.

(*c*) $x_{12} = 40, x_{13} = 10, x_{21} = 20, x_{23} = 10, x_{24} = 40, x_{25} = 20, x_{35} = 60.$

T.P. cost = Rs. 1170.

(*ii*) (*a*) $x_{11} = 80, x_{12} = 10, x_{22} = 20, x_{32} = 30, x_{33} = 20, x_{34} = 0, x_{44} = 40, x_{45} = 10.$

T.P. cost = Rs. 1610.

(*b*) $x_{11} = 70, x_{13} = 20, x_{21} = 10, x_{22} = 10, x_{31} = 0, x_{34} = 40, x_{35} = 10, x_{42} = 50.$

T.P. cost = Rs. 940.

(*c*) $x_{11} = 60, x_{12} = 10, x_{13} = 20, x_{21} = 10, x_{25} = 10, x_{31} = 10, x_{34} = 40, x_{42} = 50.$

T.P. cost = Rs. 900.

4. Solution 1 :

$x_{14} = 160, x_{21} = 110, x_{24} = 40, x_{33} = 340, x_{41} = 150, x_{42} = 100.$

Solution 2 :

$x_{14} = 200, x_{21} = 110, x_{33} = 340, x_{41} = 150, x_{42} = 100.$

Minimum T.P. cost = Rs. 3550.

5. Solution 1 :

$x_{12} = 33, x_{13} = 27, x_{21} = 42, x_{23} = 11, x_{24} = 2, x_{33} = 3, x_{35} = 27; x_{44} = 50.$

Solution 2 :

$x_{12} = 30, x_{13} = 30, x_{21} = 42, x_{23} = 11, x_{24} = 2, x_{32} = 3, x_{35} = 27, x_{44} = 50.$

Minimum T.P. cost = Rs. 370.

6. $x_{12} = 75, x_{13} = 95, x_{14} = 30, x_{21} = 75, x_{22} = 25.$

Minimum T.P. cost = Rs. 24750.

7. (*a*) I→A, II→E, III→D, IV→C, V→B, Min. cost = 38.

(*b*) A→J6, B→J2, C→J1, D→J3, E→J5, F→J4, Min. cost = 91.

8. A→A2, B→A4, C→A1, D→A6, E→A3, F→A5

A→A2, B→A4, C→A6, D→A3, E→A1, F→A5

A→A2, B→A4, C→A6, D→A3, F→A5, E→A1

For each assignment, min. cost = Rs. 6300.

9. A→P3, B→P4, C→P5, D→P1, E→P2

A→P4, B→P3, C→P5, D→P1, E→P2

Min. total time = 342 units.

10. A→VII, B→V, C→IV, D→I, E→II, F→VI, G→III

A→VII, B→V, C→IV, D→VI, E→II, F→I, G→III

A→I, B→V, C→IV, D→VI, E→II, F→VII, G→III

Min. total cost = 73.

11. I→E, III→A, IV→B, V→D, VI→C

Min. total distance = 37 miles

Crane II is not moved.

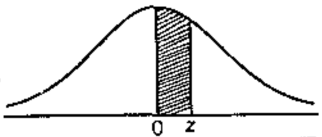
12. S1→IV, S2→I, S3→III, S4→II, S5→5, Max. total profit = Rs. 40200.
13. A→J4, B→J2, C→J1, D→J3, Job J5 is idle.

    Max. total profit = Rs. 114.

FURTHER READINGS

1. Statistics and Numerical Methods: Dr. Manish Goyal.
2. Golden statistics: N.P. Bali.

APPENDIX



STATISTICAL TABLES

Table I: Area under the Normal curve from 0 to z = Φ (z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993

Table II : Values of t_α

υ	α = 0.10	α = 0.05	α = 0.025	α = 0.01	α = 0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756

NOTES

Table III : Values of χ^2 with level of significance α and degrees of freedom ν

α ν	0.99	0.95	0.50	0.30	0.20	0.10	0.05	0.01
1	0.0002	0.004	0.46	1.07	1.64	2.71	3.84	6.64
2	0.020	0.103	1.39	2.41	3.22	4.60	5.99	9.21
3	0.115	0.35	2.37	3.66	4.64	6.25	7.82	11.34
4	0.30	0.71	3.36	4.88	5.99	7.78	9.49	13.28
5	0.55	1.14	4.35	6.06	7.29	9.24	11.07	15.09
6	0.87	1.64	5.35	7.23	8.56	10.64	12.59	16.81
7	1.24	2.17	6.35	8.38	9.80	12.02	14.07	18.48
8	1.65	2.73	7.34	9.52	11.03	13.36	15.51	20.09
9	2.09	3.32	8.34	10.66	12.24	14.68	16.92	21.67
10	2.56	3.94	9.34	11.78	13.44	15.99	18.31	23.21
11	3.05	4.58	10.34	12.90	14.63	17.28	19.68	24.72
12	3.57	5.23	11.34	14.01	15.81	18.55	21.03	26.22
13	4.11	5.89	12.34	15.12	16.98	19.81	22.36	27.69
14	4.66	6.57	13.34	16.22	18.15	21.06	23.68	29.14
15	5.23	7.26	14.34	17.32	19.31	22.31	25.00	30.58
16	5.81	7.96	15.34	18.42	20.46	23.54	26.30	32.00
17	6.41	8.67	16.34	19.51	21.62	24.77	27.59	33.41
18	7.02	9.39	17.34	20.60	22.76	25.99	28.87	34.80
19	7.63	10.12	18.34	21.69	23.90	27.20	30.14	36.19
20	8.26	10.85	19.34	22.78	25.04	28.41	31.41	37.57
21	8.90	11.59	20.34	23.86	26.17	29.62	32.67	38.93
22	9.54	12.34	21.34	24.94	27.30	30.81	33.92	40.29
23	10.20	13.09	22.34	26.02	28.43	32.01	35.01	41.64
24	10.86	13.85	23.34	27.10	29.55	33.20	36.42	42.98
25	11.52	14.61	24.34	28.17	30.68	34.68	37.65	44.31
26	12.20	15.38	25.34	29.25	31.80	35.56	38.88	45.64
27	12.88	16.15	26.34	30.32	32.91	36.74	40.11	46.96
28	13.56	16.93	27.34	31.39	34.03	37.92	41.34	48.28
29	14.26	17.71	28.34	32.46	35.14	39.09	42.56	49.59
30	14.95	18.49	29.34	33.53	36.25	40.26	43.77	50.89

Table IV : Values of $F_{0.05}$

v_2 of fre denom	ν_1 = Degrees of freedom for numerator																
	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252
2	18.51	19.00	19.16	19.25	19.30	19.33	19 35	19.37	19.38	19.40	19.41	19.43	19 45	19.46	19 46	19.47	19.48
3	10 13	9.55	9.28	9.12	9.01	8.94	8.89	8 85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5 96	5.91	5.86	5.80	5.77	5.75	5.72	5.69
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43
6	5.99	5 14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4 00	3.94	3.87	3.83	3.81	3.77	3 74
7	5.59	4.74	4.35	4.12	3.97	3 87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30
8	5.32	4.46	4.07	3 84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3 15	3.11	3.08	3.04	3.01
9	5.12	4.26	3.86	3 63	3.48	3.37	3.29	3.23	3.18	3 14	3.07	3.01	2 94	2.89	2.86	2 83	2.79
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62
11	4 84	3 98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2 79	2.72	2.65	2 60	2.57	2 53	2 49
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.38	2.38
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2 71	2.67	2.60	2 53	2.46	2.41	2.38	2.34	2.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2 64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06
18	4.41	3.55	3.16	2.93	2 77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98
20	4.35	3.49	3 10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2 02	1.98	1.94	1.89
23	4.28	3 42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.74
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.64
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.53

Table V : Values of $F_{0.01}$

ν_2 = Degrees of freedom for denominator	ν_1 = Degrees of freedom for numerator																
	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	4,052	5,000	5,403	5,625	5.764	5.859	5,928	5,982	6,023	6,056	6.106	6.157	6,209	6,240	6.261	6.287	6.313
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.57	99.47	99.48
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.32
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.65
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.20
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.06
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.82
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.03
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.48
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.08
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.78
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.54
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.34
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.18
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.05
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.93
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.83
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.75
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.67
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.61
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.55
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.50
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.45
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.40
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.36
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.45	2.39	2.30	2.21
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.27	2.20	2.11	2.02
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.84