CHAPTER 9

NUMBER SERIES, CODING DECODING AND ODD MAN OUT SERIES

LEARNING OBJECTIVES

- This Section deals with questions in which series or letters in some order, Coding and decoding
- This terms of the series or letters are follows certain pattern throughout

9.1 SERIES

Series Classified into Two Types, Namely

A. Number Series

B. Alphabet Series

A. NUMBER SERIES

Case 1: Missing terms of the series

In this type the questions we have to identify the missing term of the series real according to a specific pattern of the series rule to form its code. The students are required to detect the missing number of the series and answer the questions accordingly.

Example 1: Find the missing term of the series 2, 7, 16, _______ , 46, 67, 92

Explanation: Here the terms of the series are +5, +9, +13, +17, +21 , +25…

Thus, 2 + 5 = 6; and 7 + 9 = 16 …

So missing term = 16 + 13 = 29

Example 2: Find the wrong terms of the series 9, 29, 65, 126, 217, 344

Explanation: 2^3+1 1, 3^3 + 11, 4^3 – 1,…….

Here 29 is wrong term of series

Example 3: Find the missing term of the series 1,9, 25, 49, 81, 121, ...

Solution: The given terms of the series are consists square of consecutive odd number 1^2, 3^2, 5^2, 7^2, ...

So missing value = 13^2 = 169

B. ALPHABET SERIES

Alphabet series consists of letters of the alphabet placed in a specific pattern. For example, the series are in the following order of the numbers.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

Example 4: Find the next term of the series BKS, DJT, FIU, HHV?

Explanation: In each term, the first letter is moved two steps forward, the second letter one step backward and third letter one step forward to obtain the corresponding letter of the next term. So the missing term is JGW.

C. LETTER SERIES:

This type of question usually consist of a series of small letters which follow a certain pattern. However some letters are missing from the series. These missing letters are then given in a proper sequence as one of the alternatives.

Example 5: aab, _____, aaa, bba, _____

(a) baa (b) abb (c) bab (d) aab

1) The first blank space should be filled in by 'b' so that we have two a's by two b's.

2) The second blank place should be either `a', so that we have three a's followed by three b's.

3) The last space must be filled in by 'a'.

4) Thus we have two possible answers – 'baa'and 'bba'.

5) But only 'baa'appers in the alternatives.

So the answer (a) is correct.

9.2 CODING AND DECODING

Before transmitting, the data is encoded and at receiver side encode data is decoded in order to obtain original data by determining common key in encoded data.

The Coding and Decoding is classified into seven types according to the on what way it is doing. They are type

Type 1: Letter Coding

Type 2: Number Coding

Type 1: Letter Coding

In this type the real alphabets in a word are replaced by certain other alphabets according to a specific rule to form its code. The candidate is required to detect the common rule and answer the questions accordingly.

Case1: To form the code for another word

Example 6: If in a certain language MYSTIFY is coded as NZTUJGZ, how is MENESIS coded in that language?

Explanation: Clearly, each letter in the word MYSTIFY is moved one step forward to obtain the corresponding letter of the code.

M Y S T I F Y

+1 ↓

N Z T U J G Z

So, in MENESIS, N will be coded as O, E as F, M as N and so on. Thus, the code becomes NFOFTJT.

Example 7: If TAP is coded as SZO, then how is FRIEND coded?

Explanation: Clearly each letter in the word TAP is moved one step backward to obtain the corresponding letter of the code.

S Z O

-1 ↑

T A P

Thus, in FRIEND, F will be coded as E, R as Q , I as G, E as D, N as M and D as C. So, the code becomes EQGDMC.

Example 8: In a certain code, MENTION is written as LNEITNO. How is PRESENT written in that code?

Explanation: Clearly, to obtain the code, the first letter of the word MENTION is moved one step backward and the remaining letters are.

Reversed in order, taking two at a time. So, in PRESENT, P will be coded as O, and the sequence of the remaining letter in the code would be ERESTN. Thus the code becomes OERESTN. Hence, The answer is OERESTN.

Case 2: To find the word by analysing the given code (DECODING)

Example 9: If in a certain language CARROM is coded as BZQQNL, which word will be coded as HORSE?

Explanation: each letter of the word is one step ahead of the corresponding letter of the code

B Z Q Q N L	H O R S E
C A R R O M	I P S T F

So, H is coded as I, O as P, R as S, S as T and E as F. HORSE is coded a IPSTF.

Type 2: Number Coding

In these questions, either numerical code values are assigned to a word or alphabetical code letters are assigned to the numbers. The candidate is required to analyse the code as per the directions.

Case 1: When a numerical code values are assigned to words.

Example 10: If in a certain language A is coded as 1, B is coded as 2, and so on, how is AICCI is coded in that code?

Explanation: As given the letters are coded as

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

So in AICCI, A is coded as 1, I as 9,and C as 3. Thus, AICCI is coded as 19339.

Example 11: If PAINT is coded as 74128 and EXCEL is coded as 93596, then how would you encode ANCIENT ?

Explanation: Clearly, in the given code, the alphabets are coded as follows:

P	A	I	N	T	E	X	C	L
7	4	1	2	8	9	3	5	6

So, in ANCIENT, A is coded as 4, N is coded as 2, C as 5, I is coded as 3, E as 9, and T as 8. Hence, the correct code is 4251928.

Case 2: Number to letter coding.

Example 12: In a certain code, 2 is coded as P, 3 as N, 9 as Q, 5 as R, 4 as A and 6 as B. How is 423599 coded in that code?

Explanation: Clearly as given, 4 as A, 2 as P, 3 as N and 5 is coded as R, 9 as Q. So, 423599 is coded as APNRQQ.

9.3 ODD MAN OUT

Classification means 'to assort the items' of a given group on the basis of a certain common quality they possess and then spot the stranger or 'odd one out'.

These questions are based on words, letters and numerals. In these types of problems, we consider the defining quality of particular things. In these questions, four or five elements are given,out of which one does not belong to the group. You are required to find the 'odd one'.

Example 13: January, May, July, November

(a) January (b) May (c) July (d) November

Explanation: All the months above are 31 days , whereas , November 30 days

Answer: (d)

Example 14: 10, 14, 16, 18, 23, 24 and 26

(a) 26 (b) 17 (c) 23 (d) 9

Explanation: Each of the above series are even number, except 23.

Answer: (c)

Example 15: 6, 9, 15, 21, 24, 26, 30

(a) 9 (b) 26 (c) 24 (d) 30

Explanation: All are multiples of 3, except 26, answer (b)

Answer: (b)

Example 16: 1, 5, 14, 30, 51, 55, 91

(a) 5 (b) 55 (c) 51 (d) 91

Explanation: Each pattern is 1^2, $1^2 + 2^2$, $1^2 + 2^2 + 3^2$, $1^2 + 2^2 + 3^2 + 4^2$, $1^2 + 2^2 + 3^2 + 4^2 + 5^2$, $1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2$

But 51, is not of the form.

Answer: (c)

Example 17: 16, 25, 36, 62, 144, 196, 225

(a) 36 (b) 62 (c) 196 (d) 144

Explanation:

Each of the number except 62, is a perfect square.

Answer: (b)

EXERCISE 9(A)

(Note: Questions are taken from previous exam questions papers of Competitive exams like SSC, RRB, MAT, UPSC etc.)

Choose the most appropriate (a) or (b) or (c) or (d).

1) 6, 11, 21, 36, 56 ?
   (a) 42 (b) 51 (c) 81 (d) 91
2) 10 , 100, 200, 310 ?
   (a) 400 (b) 410 (c) 420 (d) 430
3) 11, 13, 17, 19, 23, 25, 27
   (a) 33 (b) 27 (c) 29 (d) 49
4) 6, 12, 21, 33 ?
   (a) 33 (b) 38 (c) 40 (d) 48
5) 2, 5, 9, 14, ? , 27
   (a) 20 (b) 16 (c) 18 (d) 24
6) 6, 11, 21, ? , 56, 81
   (a) 42 (b) 36 (c) 91 (d) 51
7) 10, 18, 28, 40, 54, ?, 88
   (a) 70 (b) 86 (c) 87 (d) 98
8) 120, 99, ?, 63, 48, 35
   (a) 80 (b) 36 (c) 45 (d) 40

9) 22, 24, 28, 36, ? , 84

(a) 44 (b) 52 (c) 38 (d) 54

10) 4832, 5840, 6848, 7856 ?

(a) 8864 (b) 8815 (c) 8846 (d) 8887

11) 10, 100, 200, 310, 430 ?

(a) 560 (b) 540 (c) 550 (d) 590

12) 28, 33, 31, 36, 34 ?

(a) 38 (b) 39 (c) 40 (d) 42

13) 120, 80, 40, 45, ?, 5

(a) 15 (b) 20 (c) 25 (d)

14) 2, 15, 41, 80, 132 ?

(a) 184 (b) 144 (c) 186 (d) 196

15) 6, 17, 39, ?, 116

(a) 72 (b) 75 (c) 85 (d) 80

16) 1, 4, 10, 22, ?, 94

(a) 46 (b) 48 (c) 49 (d) 47

17) 4, 9, 25, 48, ? , 169, 289, 361

(a) 120 (b) 121 (c) 122 (d) 164

18) 4, 12, 36, ? , 324

(a) 107 (b) 109 (c) 108 (d) 110

19) 1, 1, 4, 8 , 9, ? , 16, 64

(a) 27 (b) 28 (c) 32 (d) 40

20) 5760, 960, 192, ? 16, 8

(a) 47 (b) 48 (c) 52 (d) 50

21) 1, 2, 6, 7, 21, 22, 66, ? , 201

(a) 69 (b) 68 (c) 67 (d) 69

22) 48, 24, 96 , ? 192

(a) 48 (b) 47 (c) 44 (d) 54

23) 165, 195, 255, 285, ?, 435

(a) 345 (b) 390 (c) 335 (d) 395

24) 2, 3, 3, 5, 10, 13, 39, ?, 172, 177

(a) 42 (b) 44 (c) 43 (d) 40

25) 7, 26, 63, 214, 215, ?, 511

(a) 342 (b) 343 (c) 441 (d) 421

26) 3, 7, 15, 31, ? 127

(a) 62 (b) 63 (c) 64 (d) 65

27) 8, 28, 116, 584, ?

(a) 1752 (b) 3502 (c) 3504 (d) 3508

28) 6, 13, 28, 59, ?

(a) 122 (b) 114 (c) 113 (d) 112

29) 2, 7, 27, 107, 427, ?

(a) 1707 (b) 4027 (c) 4207 (d) 1207

30) 5, 2, 7, 9, 16, 25, 41, ?

(a) 65 (b) 66 (c) 67 (c) 68

31) In a certain language, MADRAS is coded NBESBT, how DELHI is coded in that code?

(a) EMMJI (b) EFMIJ (c) EMFIJ (d) JIFEM

32) If RAMAN is written as 12325 and DINESH as 675489 how HAMAM is written?

(a) 92323 (b) 92233 (c) 93233 (d) 93292

33) If RED is coded as 6720 then GREEN would be coded as

(a) 9207716 (b) 167129 (c) 1677209 (d) 1672091

34) If A = 1, FAT = 27, FAITH = ?

(a) 44 (b) 45 (c) 46 (d) 36

35) If BROTHER is coded 2456784, SISTER coded as 919684, what is coded for BORBERS?

(a) 2542889 (b) 2542898 (c) 2454889 (d) 2524889

36) If DELHI is coded 73541 and CALCUTTA as 82589662, How can CALICUT be coded?

(a) 5279431 (b) 5978213 (c) 8251896 (d) 8543962

37) If CLOCK is coded 34235 and TIME is 8679, what will be code of MOTEL?

(a) 72894 (b) 77684 (c) 72964 (d) 27894

38) If PALE is coded as 2134 and EARTH is coded as 41590, how is PEARL is code?

(a) 29530 (b) 24153 (c) 25430 (d) 254313

39) If LOSE is coded as 1357 and GAIN is coded as 2468, what do figure 82146 stands for?

(a) NGLAI (b) NGLIA (c) GNLIA (d) GNLIA

40) If MEKLF is coded as 91782 and LLLJK as 88867, how can IHJED is coded as?

(a) 97854 (b) 64512 (c) 54310 (d) 75632

41) If in a certain code language NAME is written as 4258 then what is coded as MEAN ?

(a) 2458 (b) 5842 (c) 8524 (d) 5824

42) If GOLD is written as IQNF, how WIND can be written as code?

(a) VHMC (b) VHCM (c) XJOE (d) DNIW

43) If ROSE is written as TQUG, how BISCUIT can be written in that code?

(a) DKUEWKV (b) CJTDVJU (c) DKVEWKV (d) DKUEWKY

LETTER: C Z N V R S W F D

CODE DIGIT: 8 6 4 7 2 9 3 5 1

(Q. No. 44-46) In each of the following questions find out the correctly coded alternative from amongst the given four alternatives (a), (b), (c), (d).

44) ZDRCVF

(a) 612875 (b) 619875 (c) 612845 (d) 612835

45) WNCSZV

(a) 348267 (b) 318267 (c) 348957 (d) 348967

46) RDNFVS

(a) 21679 (b) 216549 (c) 214579 (d) 218579

47) If DELHI is coded as CCIDD, how would you encode BOMBAY?

(a) AJMTVT (b) AMJXVS (c) MJXVSU (d) WXYZAX

48) In a certain code, RIPPLE is written as 613382 and LIFE is written as 8192. How is PILLER written in that code?

(a) 318826 (b) 318286 (c) 618826 (d) 338816

49) If PALAM could be given the code number 43, what code number can be given to SANTACRUZ?

(a) 123 (b) 85 (c) 120 (d) 125

49) **Directions: The number in each question below is to be codified in the following code:**

Digit	7	2	1	5	3	9	8	6	4
Letter	W	L	M	S	I	N	D	J	B

49) 184632

(a) MDJBSI (b) MDJBIL (c) MDJBWL (d) MDBJIL

50) In a certain code '256' means 'you are good', '637' means 'we are bad' and '358' means 'good and bad'. Which of the following represents 'and' in that code?

(a) 2 (b) 5 (c) 8 (d) 3

Directions: Find odd man out of the following (51-60):

51) 3, 5, 7, 15, 17, 19

(a) 15 (b) 17 (c) 19 (d) 7

52) 10, 14, 16, 18, 23, 24, 26

(a) 26 (b) 23 (c) 24 (d) 18

53) 1, 4, 9, 16, 24, 25, 36

(a) 9 (b) 24 (c) 25 (d) 36

54) 41, 43, 47, 53, 61, 71, 73, 75

(a) 75 (b) 73 (c) 71 (d) 53

55) 16, 25, 36, 73, 144, 196, 225

(a) 36 (b) 73 (c) 196 (d) 225

56) 1, 4, 9, 16, 19, 36, 49

(a) 19 (b) 9 (c) 49 (d) 16

57) 1, 5, 14, 30, 49, 55, 91

(a) 49 (b) 30 (c) 55 (d) 91

58) 835, 734, 642, 751, 853, 981, 532

(a) 751 (b) 853 (c) 981 (d) 532

59) 4, 5, 7, 10, 14, 18, 25, 32

(a) 7 (b) 14 (c) 18 (d) 33

60) 52, 51, 48, 43, 34, 27, 16

(a) 27 (b) 34 (c) 43 (d) 48

ANSWERS

EXERCISE-9 A

1. (c)	**2.** (d)	**3.** (b)	**4.** (d)	**5.** (a)	**6.** (b)	**7.** (a)	**8.** (a)	**9.** (b)	**10.** (a)
11. (a)	**12.** (b)	**13.** (b)	**14.** (d)	**15.** (a)	**16.** (a)	**17.** (b)	**18.** (c)	**19.** (a)	**20.** (b)
21. (c)	**22.** (a)	**23.** (a)	**24.** (c)	**25.** (b)	**26.** (b)	**27.** (d)	**28.** (a)	**29.** (a)	**30.** (c)
31. (b)	**32. (a)**	**33. (c)**	**34. (a)**	**35. (a)**	**36. (c)**	**37. (a)**	**38. (b)**	**39. (c)**	**40. (d)**
41. (a)	**42. (a)**	**43. (a)**	**44. (d)**	**45. (c)**	**46. (b)**	**47. (a)**	**48. (a)**	**49. (d)**	**50. (c)**
51. (a)	**52. (b)**	**53. (b)**	**54. (a)**	**55. (b)**	**56. (a)**	**57. (c)**	**58. (a)**	**59. (c)**	**60. (b)**

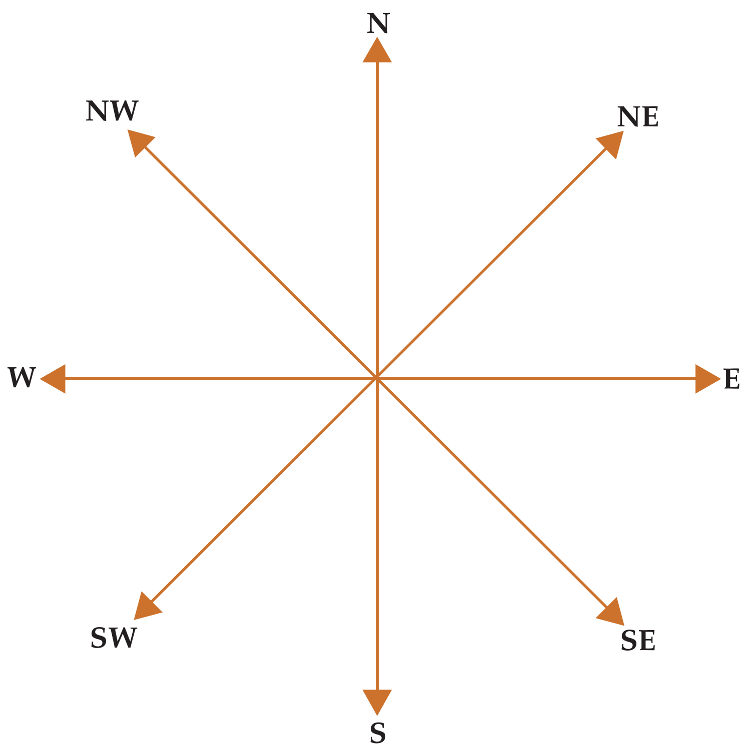
CHAPTER 10

DIRECTION SENSE TEST

INTRODUCTION

After reading this chapter, students will be able to understand:

- In this test, the questions consist of a sort of direction puzzle. A successive follow-up of direction is formulated and the students is required to ascertain the final direction. The test is meant to judge then ability to trace and follow correctly and sense the direction correctly.
- The adjoining figure shows the four main directions (North N, South S, East E, and West W) and four cardinal directions (North East (NE), North West (NW), South East (SE), South West (SW) to help the students know the directions.



Always Remember:

Left + left	Down
Left + right	Up
Right + left	Up
Right + right	Down
Up + left	Left
Up + right	Right
Down + left	Right
Down + right	Left

Examples:

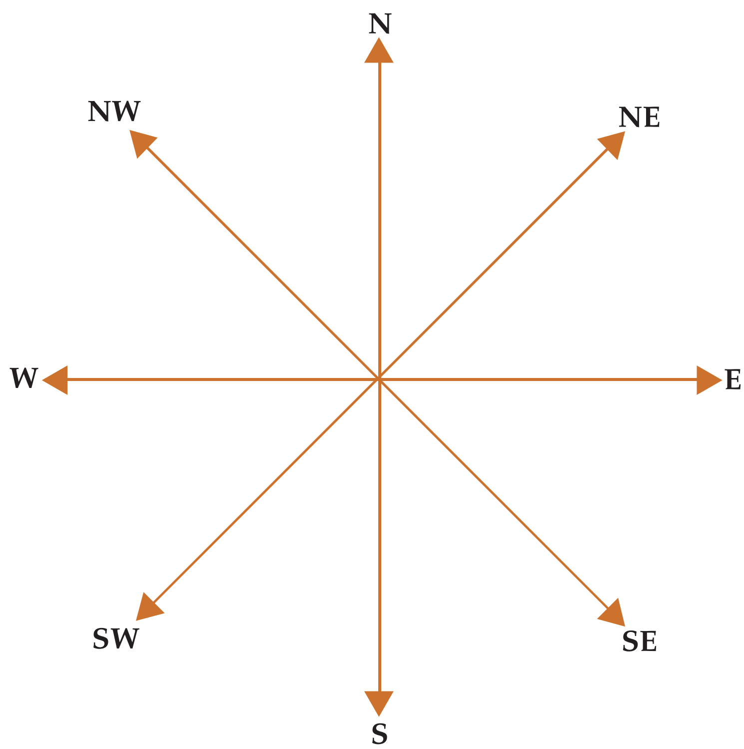
1. A man starts from a point and walks 2 km towards North, turns towards his right and walks 2 km, turns right again and walks. What is the direction now be is facing?

   (a) South (b) South-East

   (c) North (d) West

   Explanation: (a) The diagram given below helpful solving the questions and Direction Test.

   South.

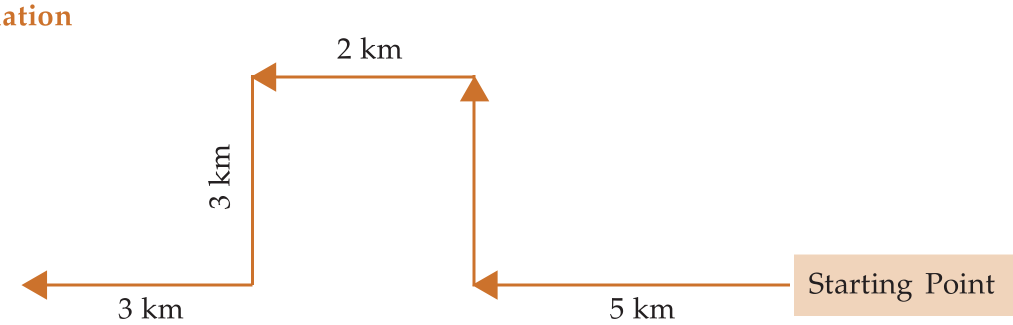


2. Ramu walks 5 kms starting from her house towards west then turns right and walks 3 km. Thereafter she takes left turn and walks 2 km. Further, she turn left and walks 3 km. Finally, she turns right and walks 3 kms. In what direction she is now from her house?

(a) West (b) North

(c) South (d) East

Explanation

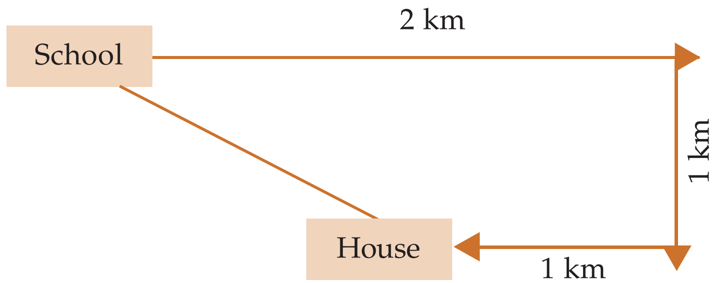


It's clear from the diagram Ramu is West of her house.

3. Gopal started walking 2 km straight from his school. Then he turned right and walked 1 km. Again he turned right and walked 1 km to reach his house. If his house is sourth-east from his school, then in which direction did Gopal start walking from the school?

(a) East (b) West

(c) South (d) North



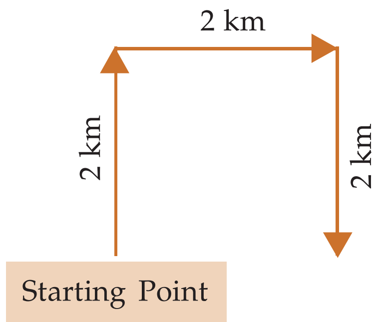
From the diagram that Gopal Started walking towards west from the school.

4. A man starts from a point, walks 2 km towards north, turns towards his right and walks 2 km, turns right again and walks. What is the direction now he is facing?

(a) South (b) East

(c) North (d) West

Explanation:



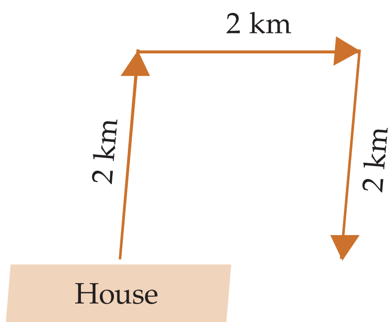
Based on the diagram the person facing towards south.

5. Janki started from her house and walked 2 km towards North. Then she took a right turn and covered one kilometre. Then she took again a right turn and walked for 2 kms. In what direction is she going?

(a) North (b) East

(c) South (d) West

Explanation:



Janaki is going on South.

EXERCISE – 10(A)

(Note: Questions are taken from previous exam questions papers of Competitive exams like SSC, RRB, MAT, UPSC etc.)

Choose the appropriate answer (a) or (b) or (c) or (d)

1. Mohan starts from point A and walks 1 km towards south, turns left and walks 1km. Then he turns left again and walks 1 km. Now he is facing.
   (a) East (b) West (c) North (d) South-west
2. Suresh starts from a point, walks 2 miles towards south, turns right and walks $1^1/_2$ miles, turns left and walks ½ miles and then he turns back. What is the direction he is facing now?
   (a) East (b) West (c) South (d) North
3. A man starts from a point, walks 4 miles towards north and turns left and walks 6 miles, turns right and walks for 3 miles and again turns right and walks 4 miles and takes rest for 30 minutes. He gets up and walks straight 2 miles in the same direction and turns right and walks one mile. What is the direction he is facing?
   (a) North (b) South (c) South-east (d) West
4. Arun started from point A and walked 10 km East to point B, then turned to North and walked 3 km to point C and then turned West and walked 12 kms to point D, then again turned South and walked 3 kms to point E. In which direction is he from his straight point?
   (a) East (b) South (c) West (d) North
5. A starts from a point and walks 5 kms north, then turns left and walks 3 kms. Then again turns left and walks 5 km. Point out the direction in which he is going now.
   (a) North (b) South (c) East (d) West
6. A rat run 20 towards East and turns to right runs 10 and turns to right runs 9 and again turns to left runs 5 and then turns to left runs 12 and finally turns to left and rusn 6. Now what direction is the rat facing?
   (a) East (b) North (c) West (d) South
7. A driver left his village and drove North for 20 km, after which he stopped for breakfast. Then he turned left and drove another 30 km, when he stopped for lunch. After some rest, he again turned left and drove 20 kms before stopping for evening tea. Once more he turned left and drove 30 kms to reach the town where he had supper. After evening tea in which direction did he drive?
   (a) West (b) East (c) North (d) South
8. A man is facing East, then he turns left and goes 10 m, then turns right and goes 5 m then goes 5 m to the South and from there 5 m to West. In which direction is be from his original place?
   (a) East (b) West (c) North (d) South

9. From her home Prerna wishes to go to school. From home she goes towards North and then turns left and then turns right, and finally she turns left and reaches school. In which direction her school is situated with respect to her home?

(a) North-East (b) North-West (c) South-East (d) South-West

10. A child walks 25 feet towards North, turns right and walks 40 feet, turns right again and walks 45 feet. He then turns left and walks 20 feet. He turns left again walks 20 feet. Finally, he turns to his left to walks another 20 feet. In which direction is the child from his starting point?

(a) North (b) South (c) West (d) East

11. Raju facing North and moves 20 km, then he turned to his right and moves 20 km and then he moves 10 km in North-East, then he turned to his right and moves 20 km and then he turned to his right and moves 20 km and again he turned to his left and moves 20 km. Now in which direction Rahu is facing?

(a) South-East (b) North-East (c) South-West (d) North-West

12. K is a place which is located 2 km away in the north-west direction from the capital P. R is another place that is located 2 km away in the south-west direction from K. M is another place and that is located 2 km away in the north-west direction from R. T is yet another place that is located 2 km away in the south-west direction from M. In which direction is T located in relation to P?

(a) South-west (b) North-west (c) West (d) North

13. Babu is Rahim's neighbour and his house is 200 meters away in the north-west direction. Joseph is Rahim's neighbour and his house is located 200 meter away in the south-west direction. Gopal is Joseph's neighbour and he stays 200 meters away in the south-east direction. Roy is Gopal's neighbour and his house is located 200 meters away in the north-east direction. Then where is the position of Roys' house in relation to Babu's ?

(a) South-east (b) south-west (c) North (d) North-east

14. A tourist drives 10 km towards west and turns to left and takes a drive of another 4 km. He then drives towards east another 4 km and then turns to his right and drives 5 km . Afterwards he turns to his left and travels 6 km. In which direction is je from the starting point?

(a) North (b) East (c) West (d) South

15. A man started walking West. He turned right, then right again and finally turned left. Towards which direction was he walking now?

(a) North (b) South (c) West (4) East

16. One evening, Raja started to walk toward the Sun. After walking a while, he turned to

his right and again to his right. After walking a while, he again turned right. In which direction is he facing?

(a) South (b) East (c) West (d) North

17. Five boys A, B, C, F, E, are sitting in a park in a circle. A is facing South-West, D is facing South-East, B and E are right opposite A and D respectively and C is equidistant between D and B. Which direction is C facing?

(a) West (b) South (c) North (4) East

18. If a man on a moped starts from a point and rides 4 km South then turns left and rides 2 km and turn again to the right to ride to go more towards which direction is he moving ?

(a) North (b) West (c) East (d) South

19. A man starts from a point, walk 8 km towards North, turns right and walks 12 km, turns left and walks 7 km turns and walks 20 km towards South, turns right and walks 12 km. In which direction is he from the starting point?

(a) North (b) South (c) West (d) East

20. Daily in the morning the shadow of Gol Gumbaz falls on Bara Kaman and in the evening the shadow of Bara Kaman falls on Gol Gumbaz exactly. So in which direction is Gol Gumbaz to Bara Kaman?

(a) Easter side (b) Western side (c) Northern side (d) Southern side

21. Ashok went 8 km South and turned West and walked 3 km again he turned North and walked 5 kms. He took a final turn to East and walked 3 kms . In which direction was Ashok from the starting point?

(a) East (b) North (c) West (d) South

22. If X stands on his head with his face towards south, to which direction will his left hand point ?

(a) East (b) West (c) North (d) South

23. I drove East for 5 miles then drove North 3 miles, then turned to my left and drove for 2 miles and again turned to my left. Which direction am I going now?

(a) South (b) North (c) West (d) North-west

24. If A stands on his head with his face towards north. In which direction will his left hand point ?

(a) North-East (b) North (c) East (d) North-West

25. A car travelling from south covers a distance of 8 km, then turns right and runs another 9 kms and again turns to the right and was stopped. Which direction does it face now?

(a) South (b) North (c) West (d) East

26. A taxi driver commenced his journey from a point and drove 10 km toward north and turned to his left and drove another 5 km. After waiting to meet a friend here, he turned to his right and continued to drive another 10 km. He has covered a distance of 25 km so far, but in which direction would he be now?

(a) South (b) North (c) East (d) South-east

27. A walks 3 kms northward and then he turns left and goes 2 km. He again turns left and goes 3 km. He turns right and walks straight. In which direction is he walking now?

(a) East (b) West (c) North (d) South

28. Áwalks southeards, then turns right, then left and then right. In which direction is he from the starting point?

(a) South (b) East (c) West (d) North

29. A man starts from a point, walks 15 metres towards East, turns left and walks 10 metres, turns right again and walks. Towards which direction is he now waking?

(a) North (b) East (c) West (d) South

30. A boy starts walking towards West, he turns right and again he turns right and then turns left at last. Towards which direction is he walking now?

(a) West (b) North (c) West (d) East

31. I stand with my right hand extended side-ways towards South. Towards which direction will my back be ?

(a) North (b) West (c) East (d) South

32. If a person moves 4 km towards west, then turns right and moves 3 km and then turns right and moves 6 km, which is the directions in which he is now moving ?

(a) East (b) West (c) North (d) South

33. If Mohan sees the rising sun behind the temple and the setting sun behind the railway station from his house, what is the direction of the temple from the railway station?

(a) South (b) North (c) East (d) West

34. Laxman went 15 km to North then he turned West and covered 10 kms. Then he turned south and covered 5 kms. Finally turning to East he covered 10 kms. In which direction he is from his house?

(a) East (b) West (c) North (d) South

35. A man starts from a point, walks 4 miles North, turns to his right and walks 2 miles, again turns to his right and walks 2 miles, again turns to his right and walks 2 miles. In

which direction would he be now?

(a) North (b) South (c) East (d) West

36. I started walking down a road in the morning facing the Sun. After walking for sometime I turned to my left. Then I turned to my right. In which direction was I going then ?

(a) East (b) West (c) North (d) South

37. Lakshmi walked 2 furlongs north from her house and took a turn to left and continued to walk another one kilometre and finally she turned left and reached the school. Which direction is she facing now?

(a) West (b) North (c) South (d) North

38. You are going straight, first eastwards, then turn to the right, then right again, then left. In which direction would you be going now?

(a) East (b) West (c) South (d) North

39. If Ahmed travels towards North from his house, then to left, then to South covering equal distances in each direction to reach Sohan's house, in which direction is Ahmed's house now?

(a) East (b) South (c) North (d) West

40. You go North, turn right, then right again and then go to the left. In which direction are you now?

(a) South (b) East (c) West (d) North

41. Roopa starts from a point and walks 15 metre towards west, turns left and walks 12 metre, turns right again and walks. What is the direction she is now facing?

(a) South (b) West (c) East (d) North

42. A man starts his journey facing the sun early morning. He then turns right and walks 2 km. He then walks 3 km after turning right again. Which is the direction he is facing now?

(a) North-East (b) North (c) West (d) South

43. Roy walks 2 km to East, then turns North-West and walks 3 km. Then he turns South and walks 5 km. Then again he turns West and walks 2 km. Finally he turns North and walks 6 km. In which direction, is he from the starting point?

(a) South-West (b) South-East (c) North-West (d) North-East

44. Seeta starts from a point, walks 2 km towards north, turns towards her right and walks 2 km, turns right again and walks. What is the direction she is facing now?

(a) East (b) West (c) South (d) North

45. Shyam was facing East. He walked 5 km forward and then after turning to his right walked 3 km. Again he turned to his right and walked 4 km. After this he turned back. Which direction was he facing at that time?

(a) East (b) West (c) North (d) South

46. Raju is standing facing north. He goes 30 metres ahead and turns left and goes for 15 metres. Now he turns right and goes for 50 metres and finally turns to his right and walks. In which direction is he heading?

(a) North (b) East (c) South (d) West

47. Sanmitra starts from his house and walks 3 km towards north. Then he turns right and walks 2 km and then turns right and walks 5 km, then turns right and walks 2 km and then again turns right and walks 2 km. Which direction is he facing now?

(a) North (b) South (c) West (d) East

48. Raju is Ramu's neighbour and he stays 100 metres away towards southeast. Venu is Raju's neighbour and he stays 100 metres away towards southwest. Khader is Venu's neighbour and he stays 100 metres away towards, north-west. Then where is the position of Khader's home in relation to Ramu's?

(a) South-East (b) South-West (c) North-West (d) East

49. Ramesh walked 3 km, towards West and turned to his left and walked 2 km. He, then turned to his right and walked 3 km. Finally, he turned to his right again and walked another 2 km. In which direction is Ramesh from his starting point now?

(a) East (b) West (c) North (d) South

50. Deepa starts walking north towards and after a while she turns to her right. After walking some distance, she turns to his left and walks a distance of 1 km. She then urns to her left again. In which direction she moving now?

(a) North (b) West (c) East (d) South

51. Raman starts walking in the morning facing the Sun. After sometime, he turned to the left later again he turned to his left. At what direction is Raman moving now?

(a) East (b) West (c) South (d) North

52. A starts walking towards North turns left, again turns left, turns right, again turns right once again turns left. In which direction is A walking now?

(a) East (b) South (c) West (d) South-East

53. X walks southwards and then turns right, then left and then right,. In which direction is he moving now?

(a) South (b) North (c) West (d) South-West

54. A man started to walk East. After moving a distance, he turned to his right. After moving a distance, he turned to his right again. After moving a little he turned in the end to his left. In which direction was he going now.?

(a) North (b) South (c) East (d) West

ANSWERS: EXERCISE 10(A)

1.	(c)	**2.**	(d)	**3.**	(b)	**4.**	(c)	**5.**	(b)
6.	(b)	**7.**	(b)	**8.**	(c)	**9.**	(b)	**10.**	(d)
11.	(a)	**12.**	(c)	**13.**	(a)	**14.**	(d)	**15.**	(a)
16.	(a)	**17.**	(d)	**18.**	(d)	**19.**	(b)	**20.**	(a)
21.	(d)	**22.**	(b)	**23.**	(a)	**24.**	(c)	**25.**	(a)
26.	(b)	**27.**	(b)	**28.**	(a)	**29.**	(b)	**30.**	(b)
31.	(b)	**32.**	(a)	**33.**	(c)	**34.**	(c)	**35.**	(a)
36.	(a)	**37.**	(d)	**38.**	(c)	**39.**	(a)	**40.**	(b)
41.	(b)	**42.**	(c)	**43.**	(c)	**44.**	(c)	**45.**	(a)
46.	(b)	**47.**	(a)	**48.**	(c)	**49.**	(b)	**50.**	(b)
51.	(b)	**52.**	(a)	**53.**	(c)	**54.**	(b)		

CHAPTER 11

SEATING ARRANGEMENTS

LEARNING OBJECTIVES

- To understand the Logical statements involved in the Seating Arrangements.
- To understand the types of Seating Arrangements.

The process of making a group of people to sit as per a prefixed manner is called Seating Arrangement these questions, some conditions are given on the basis of which students are required to arrange objects, either in a row or in a circular order.

INTRODUCTION

11.1 BASED ON VARIOUS PATTERN OF SITTING ARRANGEMENTS ARE CLASSIFIED INTO

1) Linear Arrangements
2) Circular Arrangements
3) Polygon Arrangements

Here we are limited to our topic linear and circular arrangements only. While making arrangements, it should be noted that all the conditions given are compiled with. These type of questions generally involve five to eight individuals arranged in a certain manner or pre-conditions. They may have to be arranged in a Circle or in a row accordingly.

Sometimes these questions are made more difficult by allowing an individual to a particular position with some conditions.

General instructions to Solve Seating Arrangement Questions are as follows.

1) First of all take a review on the given information. After performing this step, you would get an idea of the situation of people or objects.
2) Next, determine the usefulness of each information's and classify them accordingly into 'definite information', 'comparative information' and 'negative information'.
3) When the place of any objects or persons is definitely mentioned then we say that it is a definite information, X is sitting on the right end of the bench.
4) When the place of any object or person is not mentioned definitely but mentioned only in the comparison of another person or object, then we say that it is a comparative information.

   Example 1: A is sitting second to the right of E. This type of information can be helpful when we can get the definite information about E.

5) A part of definite information may consist of negative information. A negative information does not tell us anything definitely but it gives an idea to eliminate a possibility.

Example 2: C is not sitting on the immediate left of A.

11.2 TYPE-1 LINEAR ARRANGEMENT

In this type of arrangement, we arrange objects or persons in a line or row. The arrangement is done only on one 'axis' and hence, the position of persons or objects assumes importance in terms of order like positions. In this type of arrangement, we take directions according to our left and right.

Steps to Solve the Linear Arrangements:

(a) Identify the number of objects and their names.

(b) Use pictorial method to represent the people or objects and their positions.

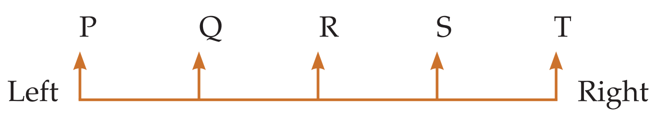
(c) Arrange the information with relevant facts and their positions and try to find out the solution.

(d) Answer the questions based on the arrangement having made.

There are few words which must be paid adequate attention, i.e., 'between' means sandwiched, 'immediate left' is different from 'to the left'. To understand it let us see some pictorial representation.

When direction of face is not clear, then we take **One Row Sequence**

(A) When direction of face is not clear, then we take based on diagram will be as follows:



From the above diagram, it is clear that

(i) Q, R, S, T are right of P but only Q is the immediate right of P.

(ii) S, R, Q, P are left of T but only S is the immediate left of T.

(iii) R, S, T are right of Q only R is the immediate right of Q.

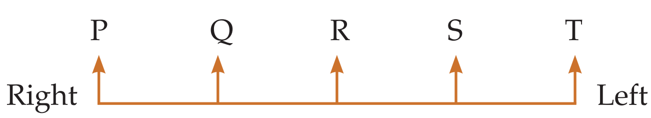
(iv) R, Q, P are left of S but only R is the immediate left of S.

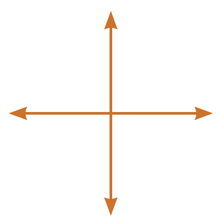
(v) S and T are right of R but only S is the immediate right of R.

(vi) Q and P are left of R but only Q is the immediate left of R.

(vii) A is the immediate left of Q while T is the immediate right of S.

(B) When direction of face is towards you, then the diagram will be as follows:





From the above diagram, it is clear that

(i) Left of P = P, R, S and T

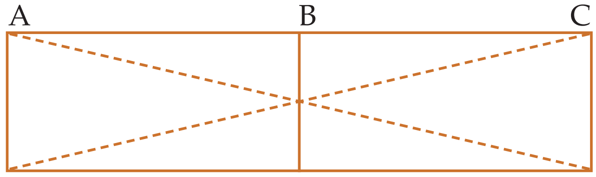
(ii) Right of T = S, R, Q and P

(iii) Q is immediate left of P; R is immediate left of Q; S is immediate left of R and T is immediate left of S.

(iv) S is immediate right of T; R is immediate right of S; Q is immediate right of R; and P is immediate right of Q.

Two Rows Sequence

Let us see 6 persons seating in two rows.



From the above diagram, it is clear that

(i) A is sitting opposite D

(ii) B is sitting opposite E

(iii) C is sitting opposite F

(iv) D and C are sitting at diagonally opposite positions

(iv) S and R are sitting at diagonally opposite positions.

Example 3: Four Children's are sitting in arrow. A is occupying seat next to B but not next to C. If C is not sitting next to D? Who is occupying seat next to adjacent to D.

(a) B (b) B and A (c) Impossible to tell (d) A

Solution: (d) The arrangements as per given information is possible only if C is sitting next to B and D is sitting next to A.

Therefore, two possible arrangements are C, B, A, D, or D, A, B, C

Clearly, only A is sitting adjacent to D:

Example 4: P, Q, R, S, T, U, V and W are sitting in a row facing North.

(i) P is fourth to the right of T

(ii) W is fourth to the left of S

(iii) R and U, which are not at the ends, are neighbours of Q and T respectively.

(iv) W is next to the left of P and P is the neighbour of Q, who are sitting at the extreme ends

Solution:

From information

(i) we get that there are three persons between P and TXXXP.

In the information (iv), it is given that W is next to the left of P and Q is the neighbour of P. Using the information with (i), we get TXXWPQ.

By the information (ii), TXXWPQXS

By the information (iii),

So, T and S are sitting at the extreme ends.

Example 5: There are Five houses P, Q , R, S, T . P is right of Q and T is left of R and right of P . Q is right of S. Which house in the middle.

(a) P (b) Q (c) R (d) T

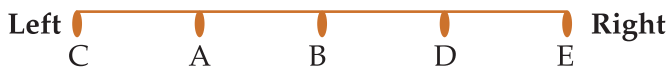
Solution: According to the question the houses can be arranged as follows.

Therefore, house P is middle.

Example 6: Friends are sitting on a bench. A is to the left of B but on the right of C, D is to the right of B but one the left of E. Who are at the extremes?

(a) A, B (b) A, D (c) C, E (d) B, D

Solution: Arrangements according to the question as follows.



Clearly C and E are the extremes.

Example 7: In a college party, 5 girls are sitting in a row. F is to the left of M and to the right of O. R is sitting to the right of N but to the left of O. Who is sitting in the middle?

(a) O (b) R (c) P (d) M

Solution: (a) arrangements of the question as follows.

Therefore, O is sitting in the middle.

Example 8: Five boys A, B, C, D and E are standing in a row. D is on the right of E, B is on the left of E but on the right of A. D is one the left of C, who is standing on the extreme right. Who is standing in the middle?

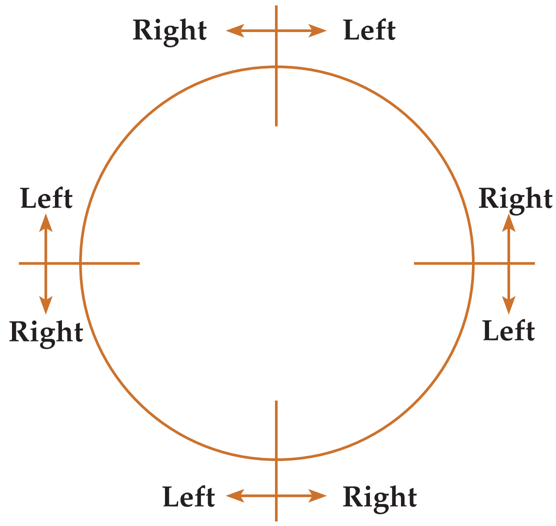
(a) B (b) C (c) D (d) E

Solution: The sequence of Boys as follows.

There E is standing in the middle.

Circular Arrangement:

In this arrangement, some persons are sitting around a circle and they are facing the centre.



1. Left movement is called clockwise rotation.
2. Right movement is called anti–clockwise rotation.
   (i) The above presentation is for 4 persons but for any number of persons, the direction is taken in the same manner.
   (ii) For rectangular and sequence arrangement, directions are taken as discussed in two rows sequence.

Example 9: (Q Nos. 1 to 3) **Study the following Question carefully and answer the given questions.**

Four ladies & A, B, C and D and Four Gentlemen E, F, G and H are sitting in a circle around a table facing each other .

I. No two ladies or gentlemen are sitting side by side.

II. C, who I sitting between G and E, is facing D.

III. F is between D and A and facing G.

IV. H is to the right of B.

(1) Who is sitting left of A?

(a) E (b) F (c) G (d) H

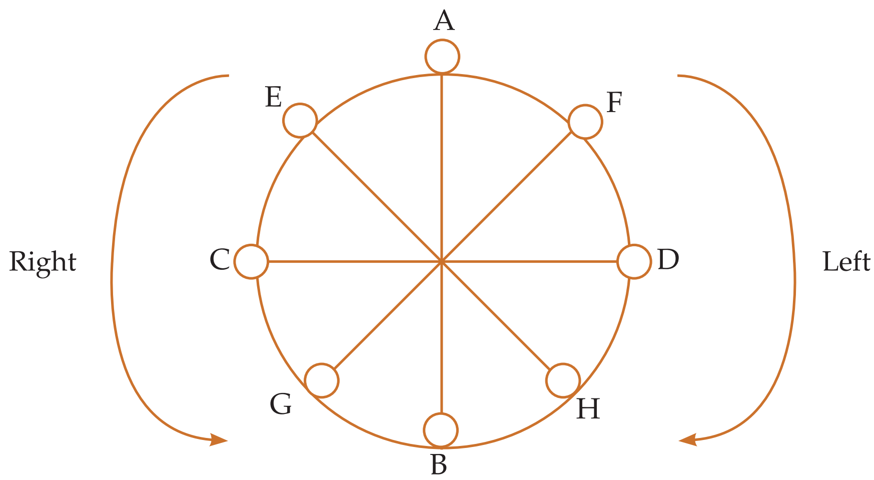
(2) E is facing whom?

(a) F (b) B (c) G (d) H

(3) Who is immediate neighbour of B?

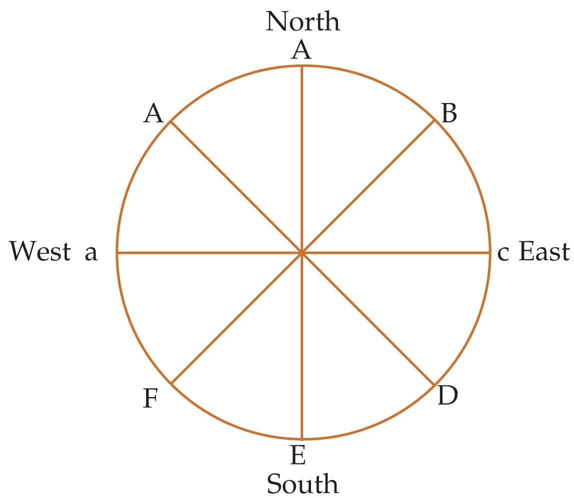
(a) G and H (b) E and F (c) E and H (d) F and H

Solution: On the basis of given information in the question, the seating arrangements of the persons are as follows.



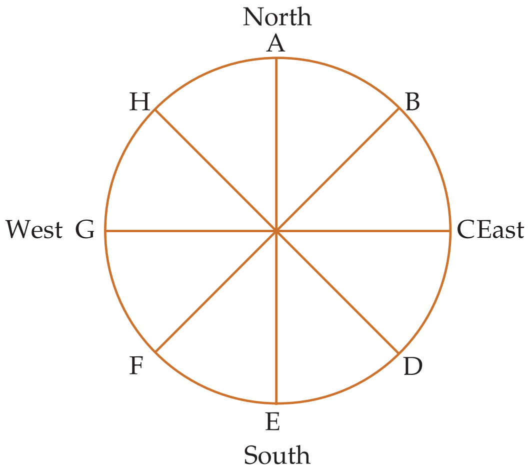
1) (b) Clearly , F is sitting left of A.
2) (d) Clearly E is facing H.
3) (a) G and H are neighbours of B.

Example 10: Eight persons A, B, C, D, E, F, G and H are sitting around the circle as given in the figure. They are facing the direction opposite to centre. If they move upto three places anti-clockwise, then.

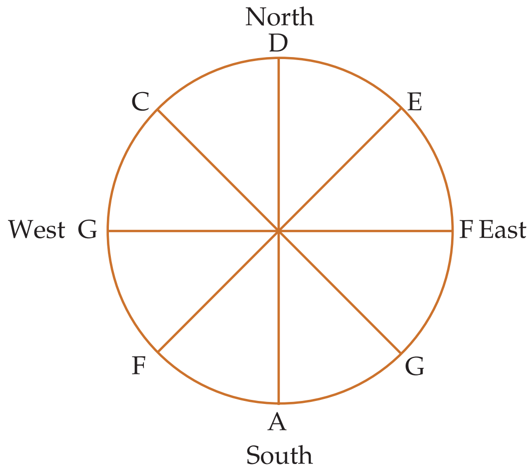


(a) B will face west
(b) E will face East
(c) H will face North-West
(d) A will face South

Solution: Following Seating arrangement is formed from the given in formation.



Original



New Portion

Clearly B will Face west

Example 11: Five People A, B, C, D and E are seated about a round table. Every chair is spaced equidistant from adjacent chairs. (UPSC (CSAT) 2013)

I. C is seated next to A .

II. A is seated two seats from D.

III. B is not seated next to A.

Which of the following must be true?

I. D is seated next to B.

II. E is seated next to A.

Select the correct from the options given below:

(a) Only I

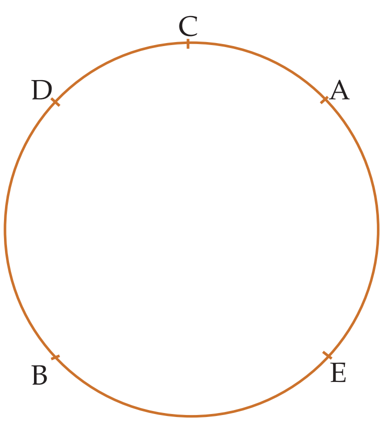
(b) Only II

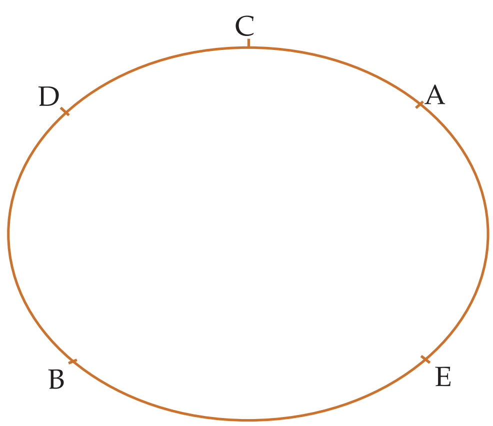
(c) Both I and II

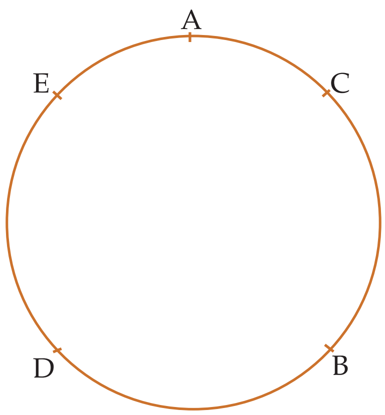
(d) Neither I nor II

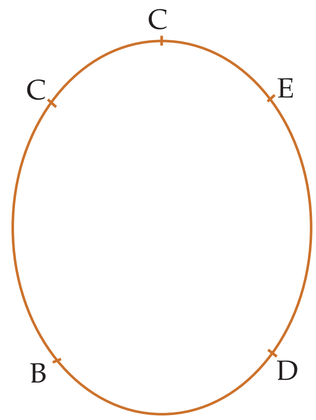
Solution:

According to the given information there are possible Seating arrangements.









From the above arrangements. It is clear that D is seated next to B . Also E is next to A.

Clearly both statements I and II are true.

Example 12: Study the following Question carefully and answer the given questions.

Eight friends A, B, C, D,E, G and H are sitting in a circle facing the centre, not necessarily in the same order. D sits third to the left of A. E sits to the immediate right of A. B is third to left of D. G is second to the right of B. C is neighbour of B. C is third to left of H. (GIC 2012)

1) Who amongst the following is sitting exactly between F and D?

   (a) C (b) E (c) H (d) A

2) Three of the following four are alike in a certain way based on the information given above and so form a group. Which is does not belong to that group.

   (a) DC (b) AH (c) EF (d) DF

3) Who amongst the following second to the left of H?

   (a) E (b) B (c) A (d) Noe of these

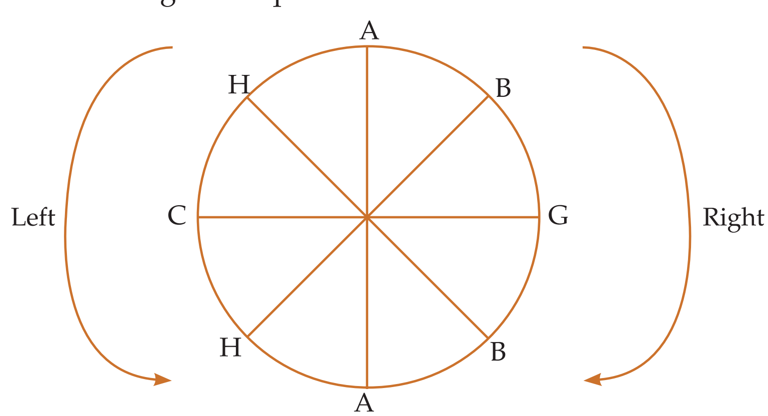
4) Who amongst the following are immediate neighbours of G?

   (a) CA (b) AF (c) DC (d) DF

5) Who amongst the following is sitting third to the right of A?

   (a) F (b) B (c) H (d) C

Solution: Arrangements according to the question is as follows.



1) (c), Clearly H is sitting exactly F and D

2) (d) DGC AFH EAF C none B

↑ ↑ ↑ ↑

Skipped Skipped Skipped No member is skipped in between

So, CB does not belong to the group.

3) (d) Clearly, H is sitting exactly between F and D

4) (c) Clearly D and C immediate neighbours of G

5) (d) Clearly, C is sitting third to the right of A

EXERCISE 11.A

(Note: Questions are taken from previous exam questions papers of Competitive exams like SSC, RRB, MAT, UPSC etc.)

Choose the appropriate answer (a) or (b) or (c) or (d)

1. Five boys A, B, C, D and E are sitting in a row A is to the right of B and E is to the left of B but to the right of C. A is to the left of D. Who is second from the left end? (U.P.B.Ed 2013)

   (a) D (b) A (c) E (d) B

2. There are five different houses, A to E, in a row. A is to the right of B and E is to the left of C and right of A, B is to the right of D. Which of the houses is in the middle? IB CA (IO) 2013)

   (a) A (b) B (c) C (d) D

3. Five friends P, Q, R, S and T are sitting in a row facing North. Here, S is between T and Q and Q is to the immediate left of R. P is to the immediate left of T. Who is in the middle? (SSC (Multi Task) 2014)

   (a) S (b) T (c) Q (d) R

4. Six children A, B, C, D, E and F are standing in a row. B is between F and D. E is between A and C. A does not stand next to eight F or D. C does not stand next to D. F is between which of the following pairs of children? (SSC (FCI) 2012)

   (a) B and E (b) B and C (c) B and D (d) B and A

5. There are eight books kept one over the other. Two books are on Organisation Behaviour, two books on TQM, three books on Industrial Relations and one book is on Economics. Counting from the top, the second, fifth and sixth books are on Industrial Relations. Two books on Industrial Relations are between two books on TQM. One book of Industrial Relations is between two books on Organizational Behaviour while the book above the book of Economics is a book of TQM. Which book is the last book from the top? (MAT 2011)

   (a) Economics (b) TQM

   (c) Industrial Relations (d) Organizational Behaviour

6. Five boys are standing in a row facing East. Pavan is left of Tavan, Vipin and Chavan to the left of Nakul. Chavan is between Tavan and Vipin. Vipin is fourth from the left, then how far is Tavan to the right? (CLAT 2014)

   (a) First (b) Second (c) Third (d) Fourth

7. Six persons M, N, O, P, Q and R are sitting in two row with three persons in each row. Both the row are in front of each other. Q is not at the end of any row. P is second the left of R. O is the neighbour of Q and diagonally opposite to P. N is the neighbour of R. Who is in front N? (UPSC (CSAT) 2011)

   (a) R (b) Q (c) P (d) M

8. Six persons A, B, C, D, E and F are sitting in two row, three in each row. (MAT 2011)

   (I) E is not at the end of any row

   (II) D is second to the left of F

   (III) C, the neighbor of E, is sitting diagonally opposite

   (IV) B is the neighbor of F.

   Which of the following are in one of the two rows?

   (a) D, B and F (b) C, E and B (c) A, E and F (d) F, B

Direction (Q.No.9): Read the following information carefully and answer that question that follows.

Five boys A_1, A_2, A_3, A_4 and A_5 are sitting in a stair in the following way. (RRB (TC/CC) 2010)

I. A_5 is above A_1

II. A_4 is under A_2

III. A_2 is under A_1

IV. A_4 is between A_2 and A_3.

9. Who is at the lowest position of the stair?

   (a) A_1 (b) A_3 (c) A_5 (d) A_2

10. Five children are sitting in a row. S is sitting next to P but not T. K is sitting next to R, who is sitting on the extreme left and T is not sitting next to K. Who is/are adjacent to S? (NIFT (UG) 2014)

   (a) K and P (b) R and P (c) Only P (d) P and T

11. Five senior citizens are living in a multi-storeyed building. Mr. Muan lives in a flat above Mr. Ashokan, Mr. Lokesh in a flat below Mr. Gaurav, Mr. Ashokan lives in a flat below Mr. Gaurav and Mr. Rakesh lives in a flat below Mr. Lokesh. Who lives in the topmost flat? (MAT 2013).

   (a) Mr. Lokesh (b) Mr. Gaurav (c) Mr. Muan (d) Mr. Rakesh

12. In a gathering seven members are sitting in a row. 'C' is sitting left to 'B' but on the right to 'D'. 'A' is sitting right to 'B', 'F; is sitting right to 'E' but left to 'D'. 'H' is sitting left to 'E'. Find the person sitting in the middle (SSC (10+2) 2013)

   (a) C (b) D (c) E (d) F

Directions (No: 13-17): Study the following information carefully to answer the given questions.

A to H are seated in straight line facing North. C sits fourth left of G. D sits second to right of G. Only two people sit between D and A. B and F are immediate neighbours of each other. B is not an immediate neighbour of A. H is not neighbour of D. (GIC 2012)

13. Who amongst the following sits exactly in the middle of the persons who sit fifth from the left and the person who sit sixth from the right?

    (a) C (b) H (c) E (d) F

14. Who amongst the following sits third to the right of C?

    (a) B (b) F (c) A (d) E

15. Which of the following represents persons seated at the two extreme ends of the line?

    (a) C, D (b) A, B (c) B, G (d) D, H

16. What is the position of H with respect to F?

    (a) Third to the left (b) Immediate right (c) Second to right (d) Fourth to left

17. How many persons are seated between A and E?

    (a) One (b) Two (c) Three (d) Four

Directions (Q. No. 18-22) (MAT 2012)

Study the following information carefully to answer the given questions.

Ten students are A to J are sitting in a row facing west.

I. B and F are not sitting on either of the edges.

II. G is sitting left of D and H is sitting to the right of J.

III. There are four persons between E and A.

IV. I is the north of B and F is the south of D.

V. J is between A and D and G is in E and F.

VI. There are two persons between H and C.

18. Who is sitting at the seventh place counting from left?

    (a) H (b) C (c) J (d) Either H or C

19. Who among the following is definitely sitting at one of the ends?

    (a) C (b) H (c) E (d) Cannot be determined

20. Who are immediate neighbours of I?

    (a) BC (b) BH (c) AH (d) Cannot determined

21. Who is sitting second left of D?

    (a) G (b) F (c) E (d) J

22. If G and A interchange their positions, then who become the immediate neighbours of E?

    (a) G and F (b) Only F (c) Only A (d) J and H

Directions (Q. Nos. 23-24) Read the following information carefully and then answer the questions that follow.

A group of singers, facing the audience, are standing in line on the stage as follows.

I. D is not right to C

II. F is not standing beside G.

III. B is not left of F

IV. E is not left of A

V. C and B have one person between E and F

VI. There are two persons H and C.

23. Who is on the Second extreme right?

(a) D (b) F (c) G (d) E

24. If we start counting from the left, on which number is B ?

(a) 1^{st} (b) 2^{nd} (c) 3^{rd} (d) 5th

Directions (Q. No. 25- 27): Study the following information carefully to answer the given questions.

Eight persons P to W are sitting in front of one another in two rows. Each row has four persons. P is between U and V and facing North. Q, who is to the immediate left of M is facing W. R is between T and M and W is to the immediate right of V.

(UCO Bank 2011)

25. Who is sitting in front of R?

(a) U (b) Q (c) V (d) P

26. Who is to the immediate right of R?

(a) M (b) U (c) M or (d) None of these

27. In which of the following pairs, persons are sitting in front of each other?

(a) MV (b) RV (c) TV (d) UR

28. Four girls A, B, C, D are sitting around a circle facing the centre. B and C infront of each other, which of the following is definitely true ? (MAT 2009)

(a) A and D infront of each other (b) A is not between B and C

(c) D is left of C (d) A is left of C

ANSWERS

1.	(c)	**2.**	(a)	**3.**	(a)	**4.**	(b)	**5.**	(a)	**6.**	(d)	**7.**	(b)
8.	(a)	**9.**	(b)	**10.**	(d)	**11.**	(c)	**12.**	(b)	**13.**	(d)	**14.**	(c)
15.	(d)	**16.**	(a)	**17.**	(a)	**18.**	(d)	**19.**	(c)	**20.**	(d)	**21.**	(a)
22.	(c)	**23.**	(b)	**24.**	(d)	**25.**	(d)	**26.**	(d)	**27.**	(a)	**28.**	(a)

CHAPTER 12

BLOOD RELATIONS

LEARNING OBJECTIVES

- Blood relations of a group of persons are given in jumbled form. In these tests, the questions which are asked in this section depend upon Relation.

12.1 DEFINITION

A person who is related to another by birth rather than by marriage.

Prerequisites:

To remember easily the relations may be divided into two sides as given below:

(i) Relations of Paternal side:

Father's father → Grandfather

Father's mother → Grandmother

Father's brother → Uncle

Father's sister → Aunt

Children of uncle → Cousin

Wife of uncle → Aunt

Children of aunt → Cousin

Husband of aunt → Uncle

(ii) Relations of Maternal side:

Mother's father → Maternal grandfather

Mother's mother → Maternal grandmother

Mother's brother → Maternal uncle

Mother's sister → Aunt

Children of maternal uncle → Cousin

Wife of maternal uncle → Maternal aunty

Relations:

1.	Grandfather's son	• Father or Uncle
2.	Grandmother's son	• Father or Uncle
3.	Grandfather's only son	• Father
4.	Grandmother's only son	• Father
5.	Mother's or father's mother	• Grandmother
6.	Son's wife	• Daughter-in-Law
7.	Daughter's husband	• Son-in-Law
8.	Husband's or wife's sister	• Sister-in-Law
9.	Brother's son	• Nephew
10.	Brother's daughter	• Niece
11.	Uncle or aunt's son or daughter	• Cousin
12.	Sister's husband	• Brother-in-Law
13.	Brother's wife	• Sister-in-Law
14.	Granson's or grand daughter's daughter	• Great grand Daughter

The efficiency in doing the problems of blood relations depends upon the knowledge of the blood relations. Some of the important relations are given below:

a) My mother's or father's son is my Brother.

b) My mother's or father's daughter is my Sister.

c) My mother's or father's father is my Grandfather.

d) My mother's or father's sister is my Aunt.

e) My mother's or father's brother is my Uncle.

f) My son's wife is my Daughter-in-law.

g) My daughter's husband is my Son-in-law.

h) My brother's son is my Nephew.

i) My brother's daughter is my Neice.

j) My sister's husband is my Brother-in-law.

k) My brother's wife is my Sister-in-law.

l) My husband's wifer's sister is my Sister-in-law.

m) My husband's or wife's brother is my Brother-in-law.

n) My uncle's or aunt's son or daughter is my Cousin.

o) My wife's father or husband's father is my Father-in-law.

p) My wife's mother or husband's mother is my Mother-in-law.

q) My father's wife is my Mother.

r) My mother's husband is my Father.

s) My son's or daughter's son is my Grandson.

t) My son's or daughter's daughter is my Grand-daughter.

Different types of questions with explanation:

(1) A is B's daughter, B is C's mother. D is C's brother. How is D related to A?

(a) Father (b) Grandfather (c) Brother (d) Son

Explanation: A is daughter B.

B is mother of C

Therefore, D is Son of B.

(2) P is Q's brother. R is Q's mother. S is R's father. T is S's mother. How is P related to T?

(a) Grand-daughter (b) Great grandson (c) Grandson (d) Grandmother

Explanation: P is brother of Q . Therefore, P is a male.

R is mother of P and Q and R is daughter of S. S is Son of T.

S is grandfather of P.

(3) A is B's brother. C is D's father. E is B's mother. A and D are brothers. How is E related to C?

(a) Sister (b) Sister-in-law (c) Niece (d) Wife

Explanation: A is brother of B . Therefore, A is male.

C is father of D. Therefore, C is male.

E is mother of B. Therefore, E is Female.

A and D are brothers.

Therefore, D is male.

Deductions:

(i) A and D are brothers of D

(ii) C is the father of A, B and D

(iii) C is the mother of A, B and D

(iv) E is wife of C

(4) A is the sister of B. B is the brother of C. C is the son of D. How is D related to A?

(a) Mother (b) Daughter (c) Son (d) Uncle

Explanation: (1) B is brother of C

C is son of D.

A is the sister of B and C.

According to the options given, we are left with no choice. But selection option (a) is correct.

5. B is the brother of A. whose only sister is mother of C. D is maternal grandmother of C. How is A related to D?

(a) Daughter-in-law (b) Daughter (c) Aunt (d) Nephew

Explanation: Although sex of A is not mentioned clearly in the question. On the basis of information given is A is daughter of B.

6. A and B are sisters. R and S are brothers. A's daughter is R's sister. What is B's relation to S?

(a) Mother (b) Grandmother (c) Sister (d) Aunt

Explanation: A's daughter R and S.

B is sister of A. B is aunt of S.

7. E is the sister of B. A is the father of C. B is the son of C. How is A related to E?

(a) Grandfather (b) Grand-daughter (c) Father (d) Great-grandfather

Explanation: is the Son of C and Grandson C and Grandson A.

E is sister of B.

Therefore, A is Grandfather of E.

8. Given that:

A is the mother of B.

C is the son of A.

D is the brother of E.

E is the daughter of B.

Who is grandmother of D?

(a) A (b) B (c) C (d) D

Explanation: E is the daughter of B and D is brother of E. Therefore B is son A and A is mother of B.

Thus A, is Grandmother of D.

9. A is D' brother. D is B's father. B and C are sisters. How is A related to C?

(a) Son (b) Grandson (c) Father (d) Uncle

Explanation: B and C daughters of D .

A is brother of D .

Therefore A is uncle of C.

10. A is B's sister. C is B's mother. D is C's father. E is D's mother, then how A is related to D?

(a) Grandfather (b) Daughter (c) Grandmother (d) Granddaughter

Explanation: D is Father of C and B is mother of C.

Thus, A is grandfather of D

11. (i) F is the brother of A.
    (ii) G is the daughter of A.
    (iii) K is the sister of F.
    (iv) G is the brother of C.

    Who is the uncle of G?

    (a) A (b) C (c) K (d) F

    Explanation: G is A and F is brother of A.

12. A is father of C and D is son of B. E is brother of A. If C is sister of D how is B related to E?

    (a) Sister-in-law (b) Sister (c) Brother (d) Brother-in-law

    Explanation: C and D Children of A and B.

    B is mother of C and D.

    Therefore, B is Sister-in-law of E.

13. C is wife of B. E is the son of C A is the brother of B and father of D. What is the relationship of E to D?

    (a) Mother (b) Sister (c) Brother (d) Cousin

    Explanation: E is B and C.

    A is uncle of E and Father of D.

    Therefore E is cousin of D.

14. M is the son of P. Q is the grand-daughter of O, who is the husband of P. How is M related to O?

    (a) Son (b) Daughter (c) Mother (d) Father

    Explanation: O is the Husband of P. M is the son of P.

    Therefore, M is son of O.

15. X and Y are brothers. R is the father of Y. S is the brother of T and maternal uncle of X. What is T to R?

    (a) Mother (b) Wife (c) Sister (d) Brother

    Explanation: R is the Father of X and Y.

    S is the maternal uncle of X and Y.

    Considering the option (b), T is wife of R.

EXERCISE 12 (A)

(Note: Questions are taken from previous exam questions papers of Competitive exams like SSC, RRB, MAT, UPSC etc.)

Choose the appropriate answer (a) or (b) or (c) or (d)

1 A is B's brother. C is A's mother. D is C's father, E is B's son. How is D related to A?

(a) Son (b) Grandson (c) Grandfather (d) Great Grandfather

2. As is B's brother. C is A's father. D is C's sister and E is D's mother. How is B related to E?

(a) Grand-daughter (b) Great grands daughter

(c) Grandaunt (d) Daughter

3. A is B's Sister. C is B's Mother. D is C's Father. E is D's Mother. Then how is A related to D?

(a) Grandmother (b) Grandfather (c) Daughter (d) Grands-daughter

4. A is the father of B. C is the daughter of B. D is the brother of B. E is the son of A. What is the relationship between C and E?

(a) Brother and sister (b) Cousins

(c) Niece and uncle (d) Uncle and aunt

5. If P is the husband of Q and R is the mother of S and Q. What is R to P?

(a) Mother (b) Sister (c) Aunt (d) Mother-in-law

6. P and Q are brothers. R and S are sister. P's son is S's brother. How is Q related to R?

(a) Uncle (b) Brother (c) Father (d) Grandfather

7. X is the husband of Y. W is the daughter of X. Z is husband of W. N is the daughter of Z. What is the relationship of N to Y?

(a) Cousin (b) Niece (c) Daughter (d) Grand-daughter

8. A reads a book and find the name of the author familiar. The author 'B' is the paternal uncle of C. C is the daughter of A. How is B related to A?

(a) Brother (b) Sister (c) Father (d) Uncle

9. A's mother is sister of B and she has a daughter C who is 21 years old. How is B related to D?

(a) Uncle (b) Maternal Uncle (c) Niece (d) Daughter

10. A is B's brother. C is A's mother. D is C's father. F is A's son. How is F related to D?

(a) Son (b) Grandson (c) Grand-grandson (d) Grand-daughter

11. A is B's brother. C is A's mother. D is C's father. E is B's son. How is B related to D?

(a) Son (b) Grand-daughter (c) Grandfather (d) Great grandfather

12. A is B's brother. C is A's mother. D is C's father. F a is A's son. How is B related to F's child?

(a) Aunt (b) Cousin (c) Nephew (d) Grandfather

13. A is B's daughter. B is C's mother. D is C's brother. How is D related to A?

(a) Father (b) Grandfather (c) Brother (d) Son

14. A is D's brother. D is B's father. B and C are sisters. How is C related to A?

(a) Cousin (b) Niece (c) Aunt (d) Nephew

15. A is B's brother. C is A's mother, D is C's father. E is B's son. How is D related to E ?

(a) Grandson (b) Great Grandson (c) Great Grandfather (d) Grandfather

16. X and Y are the children of A. A is the father of X but Y is not his son. How is Y related to A?

(a) Sister (b) Brother (c) Son (d) Daughter

17. A is B's brother. C is A's mother. D is C's father. E is B's son. How is E related to A?

(a) Cousin (b) Nephew (c) Uncle (d) Grandson

18. Based on the statements given below, find out who is the uncle of P?

(i) K is the bother of J

(ii) M is the sister of K

(iii) P is the brother of N

(iv) N is the daughter of J

(a) K (b) J (c) N (d) M

19. A and B are sisters. A is mother of D. D has a daughter C who is married to F. G is the husband of A. How is C related to D?

(a) Cousin (b) Niece (c) Aunt (d) Sister-in-law

20. R and S are brothers. X is the sister of Y and X is mother of R. What is Y to S?

(a) Uncle (b) brother (c) Father (d) Mother

21. A is B's brother. C is A's mother. D is C's father. B and D's grand-daughter. How is B related to D. Who is A's son?

(a) Aunt (b) Cousin (c) Niece (d) Grandaunt

22. A is the son of B while B and C are sisters to one another. E is the mother of C. If D is the son of E, which of the following statements is correct?

(a) D is the maternal uncle of A (b) E is the brother of B

(c) D is the cousin of A (d) B and D are brothers

23. P is the father of T. T is the daughter of M. M is the daughter of K. What is P to K?

(a) Father (b) Father-in-law (c) Brother (d) Son-in-law

24. A and B are brothers. E is the daughter of F. F is the wife of B. What is the relation of E to A?

(a) Sister (b) Daughter (c) Niece (d) Daughter

25. M and F are a married couple. A and B are sisters. A is the sister of F. Who is B to M?

(a) Sister (b) Sister-in-law (c) Niece (d) Daughter

26. If A is the mother of D. B is not the son of C. C is the father of D, D is the sister of B, then how is A related to B?

(a) Mother (b) Brother (c) Step son (d) Sister

27. A and B are brother and sister respectively. C is A's father. D is C's sister and E is D's mother. How is B related to E?

(a) Grand-daughter (b) Great grand-daughter

(c) Aunt (d) Daughter

28. Q is the son of P. X is the daughter of Q. R is the aunty (Bua) of X and L is the son of R, then what is L to P?

(a) Grandson (b) Grand-daughter (c) Daughter (d) Nephew

29. P and Q are brothers. R and S are sisters. P's son is S's brother. How is Q related to R?

(a) Uncle (b) Brother (c) Father (d) Grandfather

30. A and B are the young ones of C. If C is the mother of B but A is not the daughter of C, then what is the relationship between C and A?

(a) Nephew and Aunty (b) Brother and Sister

(c) Mother and son (d) Niece and Aunty

31. A is the mother of D and sister of B. B has a daughter C who is married to F. G is the husband of A. How is G related to D?

(a) Uncle (b) Husband (c) Son (d) Father

32. Pointing towards A, B said "your mother is the younger sister of my mother". How is A related to B?

(a) Uncle (b) Cousin (c) Nephew (d) Father

33. A is B's wife's husband's brother. C and D are sisters of B. How is A related to C?

(a) Brother (b) Sister-in-law (c) Wife (d) Sister

34. A and B are brothers. C and D are sisters. A's son is D's brother. How is B related to C?

(a) Father (b) Brother (c) Uncle (d) Son

35. A is B's sister. C is B's mother. D is C's father. E is D's mother. Then how is A related to D?

(a) Grandmother (b) Grandfather (c) Daughter (d) Grand-daughter

36. P, Q, R, S, T, U are 6 members of a family in which there are two married couples. T, a teacher is married to a doctor who is mother of R and U. Q the lawyer is married to P. P has one son and one grandson. Of the two married ladies one is a housewife. There is also one student and one male engineer in the family. Which of the following is true about the grand-daughter of the family?

(a) She is a lawyer (b) She is an engineer

(c) She is a student (d) She is a doctor

37. Six members of a family namely A, B, C, D, E and F are travelling together. 'B' is the son of C but C is not the mother of B. A and C are married couple. E is the brother of C. D is the daughter of A. F is the brother of B. How many male members are there in the family?

(a) 3 (b) 2 (c) 4 (d) 1

38. A's mother is sister of B and has a daughter C. How can A be related to B from among the following?

(a) Niece (b) Uncle (c) Daughter (d) Father

39. Rajiv is the brother of Atul. Sonia is the sister of Sunil. Atul is the son of Sonia. How is Rajiv related to Sonia?

(a) Nephew (b) Son (c) Brother (d) Father

40. Sita is the niece of Ashok. Ashok's mother is Lakshmi. Kalyani is Lakhshmi's mother. Kalyani's husband is Gopal. Parvathi is the mother-in-law of Gopal. How is Sita related to Gopal?

(a) Great grandson's daughter (b) Gopal's Sita's father

(c) Sita is Gopal's great grand-daughter (d) Grand niece

41. Seema is the daughter-in-law of Sudhir and sister-in-law of Ramesh. Mohan is the son of Sudhir and only brother of Ramesh. Find the relation between Seema and Mohan.

(a) Sister-in-law (b) Aunt

(c) Cousin (d) Wife

42. Suresh introduces a man as "He is the son of the woman who is the mother of the husband of my mother". How is Suresh related to the man?

(a) Uncle (b) Son (c) Cousin (d) Grandson

43. Pointing to a lady in a photograph. Meera said. "Her father's only son's wife is my mother-in-law "How is Meera's husband related to that lady in the photo?

(a) Nephew (b) Uncle (c) Son (d) Father

44. Pointing to a photograph Vikas said "She is the daughter of my grandfather's only son". How is the related to Vikas in the photograph?

(a) Father (b) Brother (c) Sister (d) Mother

45. Suresh's sister is the wife of Ram. Ram is Rani's brother. Ram's father is Madhur. Sheetal is Ram's grandmother. Rema is Sheetal is daughter-in-law. Rohit is Rani's brother's son. Who is Rohit to Suresh?

(a) Brother-in-law (b) Son

(c) Brother (d) Nephew

46. Vinod introduces Vishal as the son of the only brother of his father's wife. How is Vinod related to Vishal?

(a) Cousin (b) Brother (c) Son (d) Uncle

47. Among her children, Ganga's favourites are Ram and Rekha. Rekha is the mother of Sharat, who is loved most by his uncle Mithun. The head of the family is Ram Lal, who is succeeded by his sons Gopal and Mohan. Gopal and Ganga have been married for 35 years and have 3 children. What is the relation between Mithun and Mohan?

(a) Uncle (b) Son (c) Brother (d) No relation

48. Rahul and Robin are brothers. Promod is Rohin's father. Sheela is Pramod's sister. Prema is Promod's niece. Shubha is Sheela's grand-daughter. How is Rahul related to Shubha?

(a) Brother (b) Cousin (c) Uncle (d) Nephew

49. Preeti has a son, named Arun. Ram is Preeti's brother. Neeta too has a daughter named Reema. Neeta is Ram's sister. What is Arun's relationship to Reema?

(a) Brother (b) Nephew (c) Cousin (d) Uncle

50. There are 2 firm stars. One is the father of the other's son. What is the relationship of the two with each other?

(a) Grandfather and Grandson (b) Grandfather and son

(c) Husband and wife (d) Father and Son

51. Ramu's mother said to Ramu,"My mother has a son whose son is Achyut". How is Achyut relation to Ramu?

(a) Uncle (b) Cousin (c) Brother (d) Nephew

52. Ravi's father has a son Rohit who has an aunt Laxmi who has a husband Rao whose father-in-law is Mohan. What is the relation of Mohan to Ravi?

(a) Nephew (b) Grandfather (c) Son (d) Uncle

53. Vijay says, Ananda's mother is the only daughter of my mother". How is Ananda relation to Vijay?

(a) Brother (b) Father (c) Nephew (d) Grandfather

54. Introducing a man, a woman said, "His wife is the only daughter of my mother." How is the woman related with the man?

(a) Sister-in-law (b) Wife (c) Aunt (d) Mother-in-law

55. A prisoner introduced a boy who came to visit him to the jailor as "Brothers and sisters I have none, he is my father's son's son". Who is the boy?

(a) Nephew (b) Son (c) Cousin (d) Uncle

ANSWERS

EXERCISE 12(A)

1. (c)	**2.** (a)	**3.** (d)	**4.** (c)
5. (d)	**6.** (a)	**7.** (d)	**8.** (a)
9. (b)	**10.** (c)	**11.** (b)	**12.** (d)
13. (c)	**14.** (b)	**15.** (c)	**16.** (d)
17. (b)	**18.** (a)	**19.** (a)	**20.** (a)
21. (a)	**22.** (a)	**23.** (d)	**24.** (c)
25. (b)	**26.** (a)	**27.** (a)	**28.** (a)
29. (a)	**30.** (c)	**31.** (d)	**32.** (b)
33. (a)	**34.** (c)	**35.** (d)	**36.** (c)
37. (c)	**38.** (a)	**39.** (b)	**40.** (c)
41. (d)	**42.** (b)	**43.** (a)	**44.** (c)
45. (d)	**46.** (a)	**47.** (d)	**48.** (c)
49. (c)	**50.** (d)	**51.** (b)	**52.** (b)
53. (c)	**54.** (b)	**55.** (b)	

CHAPTER 13

SYLLOGISM

13.1 INTRODUCTION

Syllogism is a 'Greek' word that means inference or deduction. As such inferences are based on logic, then these inferences are called logical deduction. These deductions are based on propositions (premise).

Different types of questions covered in this chapter are as follows:

- Two Statements and Two Conclusions

'Syllogism' checks basic aptitude and ability of a candidate to derive inferences from given statements using step by step methods of solving problems.

Proposition

Let us consider the following sentences:

(i)

All	Rats	are	Rats
Quantifier	Subject	Copula	Predicate

(ii)

No	Rats	is	Rats
Quantifier	Subject	Copula	Predicate

(iii)

Some	Rats	are	Rats
Quantifier	Subject	Copula	Predicate

(iv)

Some	Rats	are not	Rats
Quantifier	Subject	Copula	Predicate

In all the sentences mentioned above, a **relation** is established between **subject** and **predicate** with the help of **quantifier** and **copula.**

Now, we can define proposition as under:

A **proposition** or **premise** is grammatical sentence comprising of four components.

- Quantifier
- Subject
- Copula
- Predicate

Components of Proposition

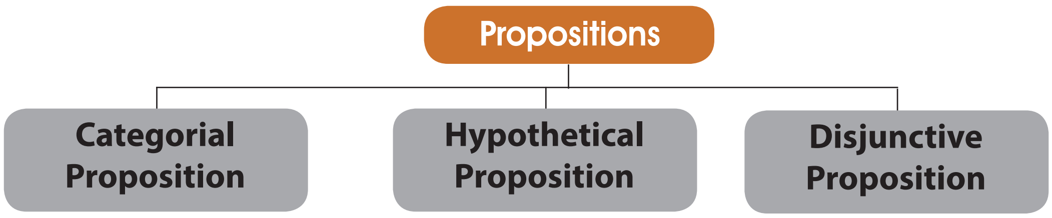
- **Quantifier –** The words 'All' 'No' and 'Some' are called quantifiers as they specify a quantity. Keep in mind that 'All' and 'No' are universal quantifiers because they refer to each and every object of a certain set.

  'Some' is a particular quantifier as it refers to atleast one existing object in a certain set.

- **Subject –** Subject is the part of the sentence something is said about. It is denoted by S.
- **Copula –** It is that part of a proposition that denotes the relation between subject and predicate.
- **Predicate –** It is that part of a proposition which is affirmed detail about that subject.

Classification of Proposition

A proposition can mainly be divided into three categories.



(i) **Categorical Proposition:** In categorical proposition, there exists a relationship between the subject and the predicate without any condition. It means predicate is either affirmation or denial of the subject unconditionally.

Example: I. All cups are pens.

II. No boy is girl.

(ii) **Hypothetical Proposition:** In a hypothetical proposition, relationship between subject and predicate is asserted conditionally.

Example: I. If it rains, he will not come.

II. If he comes, I will accompany him.

(iii) **Disjunctive Proposition:** In a disjunctive proposition, the assertion is of alteration.

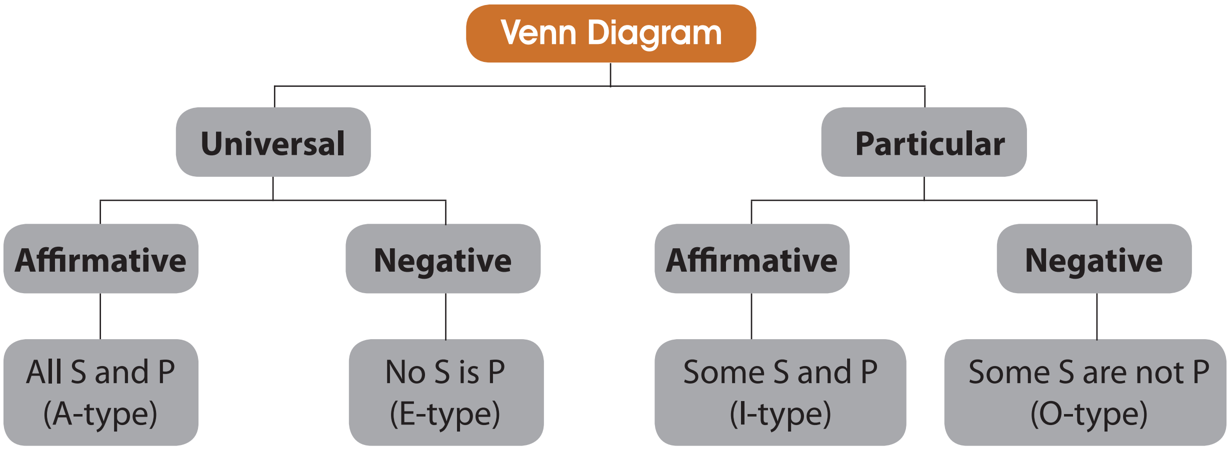
Example: I. Either he is sincere or he is loyal.

II. Either he is educated or he is scholar.

Keeping in view with the existing pattern of Syllogism in competitive examinations, we are concerned only with the categorical type of proposition.

Venn Diagram Representation of Two Propositions

Types of Venn diagram can be understood by the following diagram:



From the above diagram, following things are very much clear:

(i) Universal propositions, Either

  (a) completely include the subject (A-type)

  or

  (b) completely exclude the subject (E-type)

(ii) Particular propositions, Either

  (a) partly include the subject (I-type)

  or

  (b) partly exclude the subject (O-type)

Now, we can summarize the four standard types of propositions (premises) as below:

Format	**Type**
All S are P (Universal Affirmative)	A
No S is P (Universal Negative)	E
Some S are P (Particular Affirmative)	I
Some S are not P (Particular Negative)	O

Venn Diagram Representation

(i) A-Type (All S are P)

(ii) E-Type (No S is P)

(iii) E Type (No S is P)

(iv) O Type (Some S are not P)

Hidden Propositions

The type of propositions we have discussed earlier are of standard nature but there are propositions which do not appear in standard format and yet can be classified under any of the four types.

Let us now discuss the type of such propositions.

I. A-Type Propositions

(i) All positive propositions beginning with 'every' and 'any' are A type propositions.

Example:

(a) Every cat is dog ⇒ All dogs are cats

(b) Each of students of class has passed ⇒ All students of class X have passed

(c) Anyone can do this job ⇒ All (Women) can do this job

(ii) A positive sentence with a particular person as its subject is always an A-type proposition.

Example: (a) He should be awarded Noble Prize

He → Subject; Noble Prize → Subject

(b) Raman is a Great Scientist

Raman → Subject; Great → Predicate

(iii) A sentence with a definite exception is A type.

Example: (a) All girls except Sanjana is healthy

Sanjana → Definite exception

(b) All boys except Ram is intelligent

Ram → Definite exception

II. E-Type Propositions

(i) All negative sentences beginning with 'no one', 'not a single' etc., are E-type propositions.

Example:

(a) Not a single student could answer the question.

(b) None can cross the English channel.

(ii) A negative sentence with a very definite exception is also of E-type proposition.

Example: (a) No girl except Sanjana is healthy

Sanjana ↑ Definite exeption

(b) No boy except Ram is intelligent

Ram ↑ Definite exeption

(iii) When an Introgattive sentence is used to make an assertion, this could be reduced to an E-type proposition. example: Is there any person who can scale Mount Everest? ⇒ Non can climb Mount Everest.

(iv) A negative sentence with a particular person as its subject is E-type proposition.

Example: (a) He does not deserve Noble Prize

He ↑ Subject; Noble Prize ↑ Predicate

(b) Raman is not a Great Scientist

Raman ↑ Subject; Great ↑ Predicate

III. I-Type Propositions

(i) Positive propositions beginning with words such as 'most', 'a few' 'mostly', 'generally', 'almost', 'frequently', and 'often' are to be reduced to the I-type propositions.

Example:

(a) Almost all the Vegetables have been sold. ⇒ Some vegetables have been sold.

(b) Most of the students will qualify in the test. ⇒ Some of the students will qualify in the test.

(c) Boys are frequently physically weak ⇒ Some boys are physically weak.

(ii) Negative propositions beginning with words such as 'few' 'seldom', 'hardly', 'rarely', 'little' etc. are to be reduced to the I-type propositions.

Example:

(a) Seldom writers do not take rest. ⇒ Some writers take rest.

(b) Few Teachers do not tell a lie. ⇒ Some teachers tell a lie.

(c) Rarely Scientists do not get a good job ⇒ Some Scientists get a good job.

(iii) A positive sentence with an exception which is not definite, is reduced to an I-type proposition.

Example:

(a) All boys except (a few) are handsome

↑ Indefinite exception as names of girls are not given

(b) All girls except (four) have passed

↑ Indefinite exception as names of girls are not given

IV. O-Type Propositions

(i) All negative propositions beginning with words such as 'all', 'every', 'any', 'each' etc. are to be reduced to O-type propositions.

Example:

(a) All Psychos are not guilty. ⇒ Some Physchos are not guilty.

(b) All that glitters is not gold. ⇒ Some glittering objects are not gold.

(c) Everyone is not Scientist ⇒ Some are not Scientist.

(ii) Negative propositions with words as 'most', 'a few', 'mostly, 'generally', 'almost', 'frequently' are to be reduced to the O-type propositions.

Example:

(a) Boys are usually not physically weak. ⇒ Some boys are not physically weak.

(b) Priests are not frequently thiefs. ⇒ Some priests are not thiefs.

(c) Almost all the questions cannot be solved. ⇒ Some questions cannot be solved.

(iii) Positive propositions with starting words such as 'few', 'seldom', 'hardly', 'scarcely', 'rarely', 'little', etc., are to be reduced to the O-type propositions.

Example:

(a) Few boys are intelligent. ⇒ Some boys are not intelligent

(b) Seldom are innocents guilty. ⇒ Some innocent are not guilty.

(iv) A negative sentence with an exception, which is not definite is to be reduced to the O-type propositions.

e.g.

(a) No girls except two are beautiful

↑ Indefinite exception as names of girls are not given

(b) No cricketers except **a few**

↑ Indefinite exception as names of cricketer are not given

Exclusive Propositions

Such propositions start with 'only', 'alone', 'none but', 'none else but' etc., and they can be reduced to either A or E or l-type.

Example:

Only Post-graduates are officers.	(E-type)
None Post-graduate is officer.	(A-type)
All officers are Post-graduates.	(I-type)
Some Post-graduates are officers	

Types of Inferences

Inferences drawn from statements can be of two types:

1. **Immediate Inference:** When an inference is drawn from a single statement, then that inference is known as an immediate inference.

   Example: **Statement:** All books are pens.

   Conclusion: Some pens are books.

   In the above example, a conclusion is drawn from a single statement and does not require the second statement to be referred, hence the inference is called an immediate inference.

2. **Mediate Inference** In mediate inference, conclusion is drawn from two given statements.

   Example: **Statements:** All cats are dogs.

   All dogs are black.

   Conclusion: All cats are black.

   In the above example, conclusion is drawn from the two statements or in other words, both the statements are required to draw the conclusion. Hence, the above conclusion is known as mediate inference.

Method to Draw Immediate Inferences

There are various methods to draw immediate inferences like conversion, obversion, contraposition; etc. Keeping in view the nature of questions asked in various competitive examinations, we are required to study only two methods, implications and conversion.

(i) Implications (of a given proposition): Below we shall discuss the implications of all the four types of propositions. While drawing a conclusion through implication, subject remains the subject and predicate remains the predicate.

A-Type: All boys are blue.

From the above A-type proposition, it is very 'clear that if all boys are blue, then some boys will definitely be blue because some is a part of all. Hence, from A-type proposition, we can draw l-type conclusion (through implication).

E-Type: No cars are buses.

If no' cars are buses, it clearly means that some cars are not buses. Hence, from E-type proposition, O-type conclusion (through implication) can be drawn.

I-Type: Some chairs are tables.

'From the above I-type proposition, we cannot draw any valid conclusion (through implication).

O-Type: Some A are not B.

From the above O-type proposition, we can not draw any valid inference (through implication). On first look, it appears that if some A are not B, then conclusion that some A are B must be true but the possibility of this conclusion being true can be over ruled with the help of following example:

Case I A = {a, b, c} and B = {d, e, f}

Case II A = {a, b, c} and B ={b, c, d}

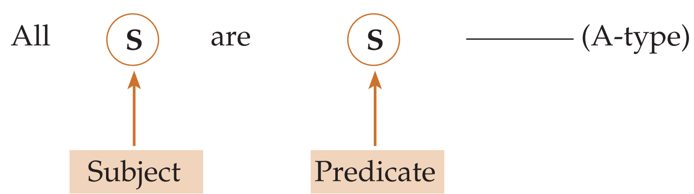
The above two cases show the relationship between A and B given by O-type proposition. "Some A are not B". Now, in case I, none of the element of set A is the element of set B. Hence, conclusion "Some A are B" cannot be valid. However, in case II, elements b and c are common to both sets A and B. Hence, here conclusion "Some A are B" is valid. But for any conclusion to be true, it should be true for all the cases. Hence, conclusion "Some A are B" is not a valid conclusion drawn from an O-type proposition.

All the results derived for immediate inference through implication can be presented in the table as below:

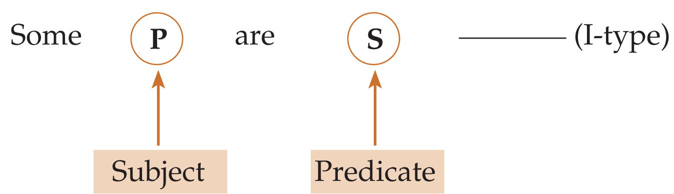
Types of propositions	Proposition	Types of Inferences	Inference
A	All S are P	I	Some S are P
E	No S is P	O	Some S are not P
I	Some S are P	–	–
O	Some S are not P	–	–

Conversion: Conversion is other way of getting immediate inferences. Unlike implication, in case of conversion, subject becomes predicate and predicate becomes subject. Let us see

- **Conversion of A-type**



After conversion it becomes



Clearly, A gets converted into I-type.

- **Conversion of E-type** No S is P

  S → Subject, P → Predicate

  After conversion it becomes

  No P is S (E-type)

  P → Subject, S → Predicate

  Clearly, E gets converted into E-type.

- **Conversion of I-type** Some S is P (E-type)

  S → Subject, P → Predicate

  Clearly I gets converted into I- type.

  Conversion of O – type.

  O type propositions cannot be converted.

Now we can make a conversion as follows:

Types of propositions	Gets Converted into
A	A
E	E
I	I
0	Never gets Converted

Immediate Inference Table

Types of propositions	Proposition	Valid immediate inference	Types of Immediate inference	Method
A	All S are P	Some S are P Some P are S	I I	Implication Conversion
E	No S is P	Some S are not P No P is S	E E	Implication Conversion
I	Some S are P	No Valid Inference Some P are S	- I	Implication Conversion
O	Some S are not P	No Valid Inference No Valid Inference	- -	Implication Conversion

Venn Diagram Representation of Immediate Inferences:

Immediate inferences are drawn from each type of Propositions (A, E, I , O)

1) A type All S are P

(i) S = { a, b, c } , P = { a, b, c, d, e}

(i) **S = { a, b, c} , P = { a, b, c, d, e, f}**

(ii) **S = { a, b, c} , P = { a, b, c}**

The above cases show all the possibilities of two sets S and P showing the relationship by the proposition.

All S are P in both cases.

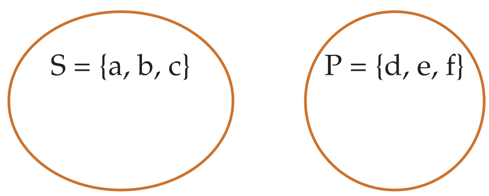
Some S are P.(Is true from relationship)

Some P are S (Its true)

Some P are not S is not valid because from it is case (i) but false from case (ii)

Inference All (P are S) is not valid because its true from case (ii) and False from case (i)

2) E type – No S Is P



We can draw the inferences as

(i) No P is S

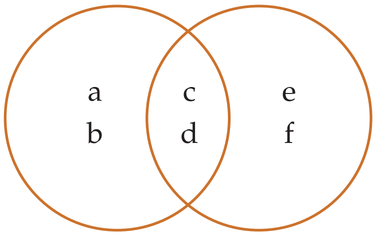
(ii) Some S are not P

(iii) Some P are not S

Any other inference drawn from E- type proposition is not valid.

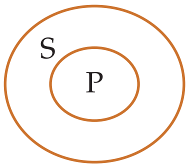
3) I-type: Some S are P

(i) S = { a, b, c, d} , P = {c, d, e, f}



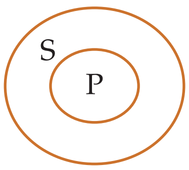
Some S are P

(ii) S = { a, b, c, d} and P = { a, b}



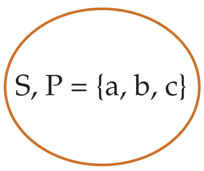
Set {a, b} is the part of S as well as Set P, hence some set S are P.

(iii) S = {a, b} , P = {a, b, c, d}



Set {b} is the part of the set S as well as Set P, hence some S are P.

(iv) S = {a, b, c} , P = {a, b, c}



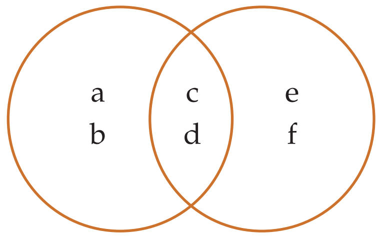
The above diagrams show the relationship between S and P from I-type relationship. From the possible combinations, it's clear that inference (Some P are S) is true. Inference (S are not P) is true from combinations (i) and (ii) but is not true from combinations (iii) and (iv).

Therefore inference (Some S are not P) is not a valid inference drawn from the above proposition.

Set {a, b} is part of set S and P, hence some s are P.

4) O-Type: Some S are not P

(i) S = {a, b, c, d} , P = {c, d, e, f}



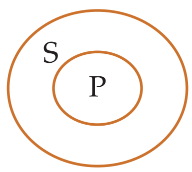
Set {a, b} is part of the set S but not Set P

Hence Some S are not P

(ii) S= {a, b, c} and P = {d, e, f}

Set {a, b} is the part of S but not set P

Hence the above relation represented by Some S are not P .



(iii) S = {a, b, c, d, e, f} , P = {e,f}

Set {a, b, c} is the part of set S abut not P. Hence Proposition Some S are not P.

On the basis of all possible combinations showing relationship between S and P, no valid inference can be drawn. Inference from Some P are not S) is true combination (i) and (iii) but not true for combination (iii). Hence it is invalid inference.

Inference (Some P are not S) is true from combination (i) and (ii) but not true for combination (iii), hence it is also invalid .

Following are the main rules for solving syllogism problems.

1) **All + All = All**

2) **All + No = No**

3) **All + Some = No conclusion**

4) **Some + No = Some Not**

5) **Some + Some = No conclusion**

6) **No + All = Some not (Reversed)**

7) **No + All = Some Not (Reversed)**

8) **No + Some = Some Not (Reserved)**

9) **No + No = No conclusion**

10) **Some Not/ Some not reserved + Anything = No conclusion**

11) **If all A are B then we can say – Some B are Not A is a possibility**

12) **If Some B are not A then we can say – All A are B is a possibility**

13) **If some A are B then we can say All A are B is a possibility. All B are A is a possibility.**

14) **All $\Leftrightarrow$ Some not reserved**

15) **Some $\Rightarrow$ All**

16) **No conclusion = Any possibility is true**

Implications (In case of Conclusions from Single Statement)

All $\Rightarrow$ Some That means If A are B then Some B is true.

Some $\Leftrightarrow$ **Some that means if Some A are B then Some B are A is true.**

No $\Leftrightarrow$ No that means if No A is B then No B is true.

Examples:

In this type of questions two statements and two conclusions are given. Its required to check.

Example 1:

Statement:

I. Some boys are student.

II. All students are Engineers.

Conclusions:

I. All Engineers are students.

II. Some boys are Engineers.

(a) Only I follows

(b) Only II follows

(c) Both I and II follow

(d) Neither I nor II follows.

Solution

(b) Statement I is an I-type proposition which distributes neither the subject nor the predicate. Statement II is an A-type propositions which distributes the subject 'Engineers' only.

Since, the Engineers is distributed in Conclusion I without being distributed in the premises. So, Conclusion I cannot follow. In second conclusion, where it is asked that some boys are Engineers but from Statement I nit is clear that some boys are not students. These boys may not be Engineers.

Example 2:

Statements:

I. All Lotus are flowers.

II. No Lily is a Lotus.

Conclusions:

I. No Lily is a flower

II. Some Lilies are flowers.

(a) Only I follows

(b) Only II follows

(c) Either I or II follows

(d) Neither I nor II follows

Solution: (c)

Here, the first premise is an A-type proposition and so the middle term 'Lotus' forming the subject is distributed. The second premise is an E proposition and so the middle term 'Lotus' forming the predicate is distributed. Since, the middle term is distributed twice, so the conclusion cannot be universal.

Example 3:

Statements

I All A's are C's

II All D's are C's

Conclusion

I All D's are C's

II. Some D's are not A's

(a) Only I follows

(b) Only II follows

(c) Both I and II follows

(d) None follows

Solution: (a) Now, taking conclusion I, it is clear that all D's are also C's but taking conclusion II, we cannot say that some D's are not A's because from Statement I it is clear that all D's are A's.

Hence, only Conclusion I follows.

Example 4:

Statements: All balls are bats.

All bats are stumps.

The sentences are already aligned. From the above given Table, A + A = A. Hence the conclusion is of type-A whose subject is the subject of the first proposition and the predicate is the predicate of the second proposition?

So the conclusion is *All balls are Stumps*.

Example 5:

Statements: All Professors are readers.

All Professors are writers.

This pair is not properly aligned because the subject of both the sentences is 'Professors'.

Since both the sentences are of type-A, we may convert any of them. So the aligned pair is Some readers are Professors.

All Professors are writers.

Here the conclusion will be of type - I

because I + A = I.

The conclusion is *Some readers are writers.*

Example 6:

Statements: Some Mangos are sweets.

All Mangos are Fruits.

The subject of both the sentences is the same. By the rule of IEA, we convert the

I - type statement. So the aligned pair is,

Some Sweets are Mangos.

All Mangos are Fruits

I+A=I. So the conclusion is

Some Sweets are Fruits.

Example 7:

Statements: All lights are bats.

No balls are lights.

By changing the order of the statements itself.

We can align the sentences. The aligned pair is

No balls are lights.

All lights are bats.

E + A = O*.

So the conclusion is,

Some bats are not balls.

Example 8: Statements: Some caps are blue.

No clip is blue.

Here the common term is 'blue' which is the predicate of both the sentences. By the rule of IEA, we convert the I-type statement. After conversion, the given pair becomes,

Some blue are caps.

No clip is blue.

Now by changing the order of the statements, we can align the sentences. So the aligned pair is, No clip is blue.

Some blue are caps.

The conclusion is of type O* since, E + I = O*. Hence the conclusion is *Some caps are not clips.*

Example 9:

Statements: Some powders are not soaps.

All soaps are detergents.

The given pair is properly aligned. But no definite conclusion can be drawn from this type because it is a O+A - type combination.

Complementary Pair

Consider the following.

Conclusions:

i) Some vans are trucks.

ii) Some vans are not trucks.

We know that either some vans will be trucks or some vans will not be trucks.

Hence either (i) or (ii) is true. Such pair of statements are called complementary pairs. So in a complementary pair, at least one of the two statements is always true. We can call a pairs a complementary pair if i) The subject and predicate of both the sentences are the same.

ii) They are an I + O - type pair or an A + O type pair or an I + E - type pair.

Some complementary pairs are given below.

i) All birds are Pigeons.

Some birds are not Pigeons.

ii) Some Chairs are watches.

Some Chairs are not watches.

iii) Some kids are cute.

No kids are cute.

Note: The steps to be followed to do a syllogism problem by analytical method are mentioned below.

i) Align the sentences properly.

ii) Draw conclusion using the table.

iii) Check for immediate inferences.

iv) Check for complementary pair if steps ii and iii fail.

EXERCISE 13 (A)

Directions (Qs. 1 - 25) : Each of the following questions contains two statements followed by two conclusions numbered I and II. You have to consider the two statements to be true, even if they seen to be at variance at the commonly known facts. You have to decide which of the given conclusions definitely follows from the given statements.

Give answer (a) if only I follows; (b) if only conclusion II follows; (c) if either I or II follows; (d) if neither I nor II follows and (e) if both I and II follow.

1. **Statement:** Some Chairs are glasses.

   All trees are Chairs.

   Conclusions: I. Some trees are glasses

   II. Some glasses are trees.

2. **Statement:** No man is a lion.

   Ram is a man.

   Conclusions: I. Ram is not a lion.

   II. All men are not Ram.

3. **Statement:** All boys are Fathers.

   All Fathers are Mothers.

   Conclusions: I. All Fathers are boys.

   II. All boys are Mothers.

4. **Statement:** All pens are cups.

   All cups are bowls.

   Conclusions: I. All pens are bowls.

   II. All cups are pots.

5. **Statement:** All students are boys.

   No boy is dull

Conclusions: I. There are no girls in the class

II. No student is dull.

6. **Statement:** Some cats are kittens.

All Rats are kittens.

Conclusions: I. Some cats are Rats.

II. Some Rats are cats.

7. **Statement:** All names are dogs.

No dogs are foxes.

Conclusions: I. All names are foxes.

II. No dogs are names.

8. **Statement:** All pens are dogs.

Some pens are lights.

Conclusions: I. Some dogs are lights.

II. Some lights are not dogs

9. **Statement:** Some animals are clouds.

Horse is a animal.

Conclusions: I. Some clouds are animal.

II. Hen is not a cloud.

10. **Statement:** All tables are rats.

Some Rats are chairs.

Conclusions: I. All rats are tables

II. Some chairs are not rats.

11. **Statement:** All tigers are birds.

Some birds are cows.

Conclusions: I. Some cows are birds.

II. Some tigers are cows.

12. **Statement:** All papers are pens.

All pens are erasers.

Conclusions: I. Some erases are papers.

II. Some pens are no papers.

13. **Statement:** Some trees are monkeys.

Some ships are trees.

Conclusions: I. Some Monkeys are ships.

II. Some trees are neither ships nor monkeys.

14. **Statement:** All glasses are mirrors.

Some mirrors are Black.

Conclusions: I. All mirrors are glasses.

II. Some glasses are black.

15. **Statement:** Some dogs are monkeys.

No monkey is black.

Conclusions: I. Some dogs are black.

II. Some monkeys are dogs.

16. **Statement:** All roads are poles.

No poles are Bungalows.

Conclusions: I. Some roads are Bungalows.

II. Some Bungalows are poles.

17. **Statement:** Many actors are directors.

All Directors are dancers.

Conclusions: I. Some actors are dancers.

II. No director is an actor.

18. **Statement:** Only dogs are animals.

No historian is an animal.

Conclusions: I. Some dogs are not historians.

II. Some historians are not dogs.

19. **Statement:** Some chairs are caps.

No cap is red.

Conclusions: I. Some caps are chairs.

II. No Chair is red.

20. **Statement:** Some cups are belts.

No belt is black.

Conclusions: I. Some cups are black.

II. Some cups are not black.

21. **Statement:** Some girls are flowers.

Some flowers are books.

Conclusions: I. Some girls are books.

II. No books are girls.

22. **Statement:** Some files are rats.

All animals are rats.

Conclusions: I. All files are rats.

II. Some rats are animals.

23. **Statement:** All cricketers are tall.

Rajesh is tall.

Conclusions: I. Rajesh is a cricketer.

II. Rajesh is not cricketer.

24. **Statement:** Some cats are cows.

All cows are horses.

Conclusions: I. Some horses are cats.

II. Some cats are horses.

25. **Statement:** All scientists are hard working.

No scientists are superstitious.

Conclusions: I. No scientists are superstitious.

II. All superstitious are not scientists.

ANSWERS:

1. (d)	**2.** (a)	**3.** (b)	**4.** (a)	**5.** (e)
6. (d)	**7.** (d)	**8.** (a)	**9.** (a)	**10.** (d)
11. (a)	**12.** (a)	**13.** (d)	**14.** (d)	**15.** (b)
16. (d)	**17.** (a)	**18.** (a)	19. (a)	**20.** (b)
21. (c)	**22.** (b)	**23.** (c)	**24.** (e)	**25.** (e)

CHAPTER 14

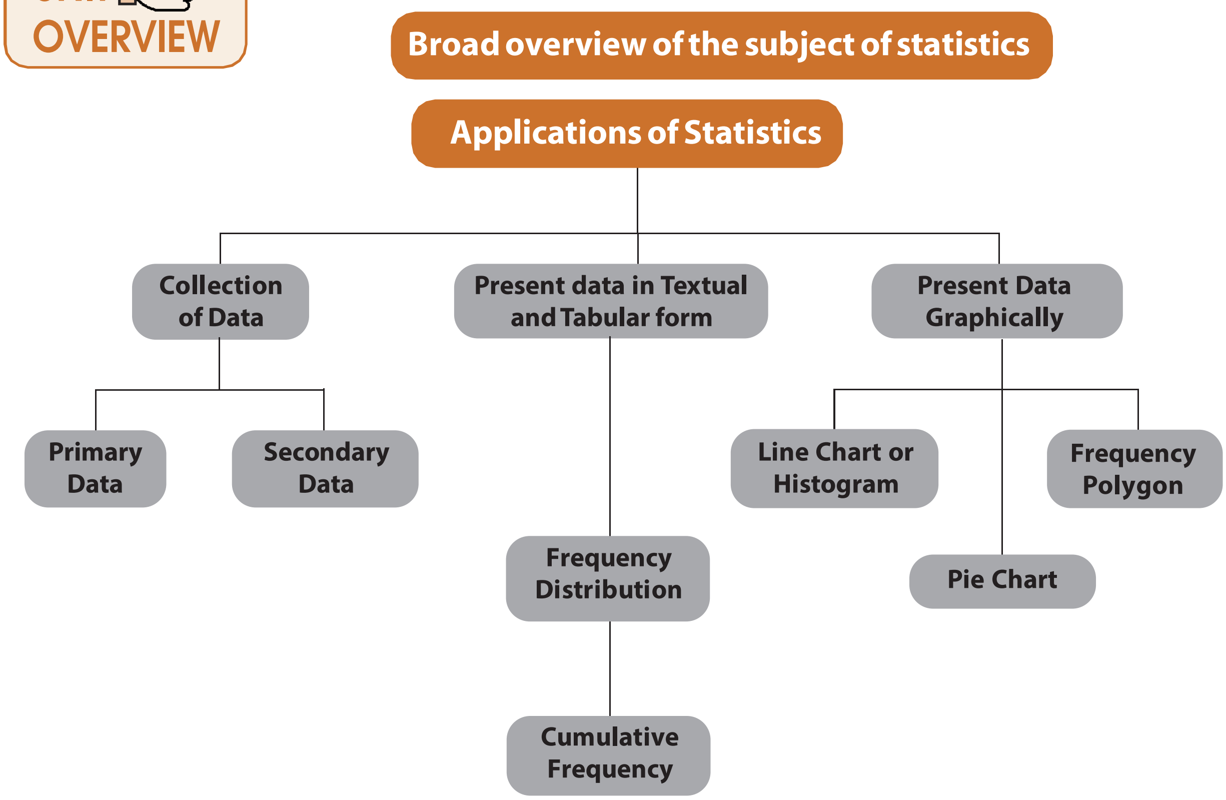
STATISTICAL DESCRITPION OF DATA

LEARNING OBJECTIVES

After reading this chapter, students will be able to understand:

- Have a broad overview of the subject of statistics and application thereof.
- Know about data collection technique including the distinction of primary and secondary data.
- Know how to present data in textual and tabular format including the technique of creating frequency distribution and working out cumulative frequency.
- Know how to present data graphically using histogram, frequency polygon and pie chart.

UNIT OVERVIEW



14.1 INTRODUCTION OF STATISTICS

The modern development in the field of not only Management, Commerce, Economics, Social Sciences, Mathematics and so on but also in our life like public services, defence, banking, insurance sector, tourism and hospitality, police and military etc. are dependent on a particular subject known as statistics. Statistics does play a vital role in enriching a specific domain by collecting data in that field, analysing the data by applying various statistical techniques and finally making statistical inferences about the domain. In the present world, statistics has almost a universal application. Our Government applies statistics to make the economic planning in an effective and a pragmatic way. The businessman plan and expand their horizons of business on the basis of the analysis of the feedback data. The political parties try to impress the general public by presenting the statistics of their performances and accomplishments. Most of the research scholars of today also apply statistics to present their research papers in an authoritative manner. Thus the list of people using statistics goes on and on and on. Due to these factors, it is necessary to study the subject of statistics in an objective manner.

History of Statistics

Going through the history of ancient period and also that of medieval period, we do find the mention of statistics in many countries. However, there remains a question mark about the origin of the word 'statistics'. One view is that statistics is originated from the Latin word 'status'. According to another school of thought, it had its origin in the Italian word 'statista'. Some scholars believe that the German word 'statistik' was later changed to statistics and another suggestion is that the French word 'statistique' was made as statistics with the passage of time.

In those days, statistics was analogous to state or, to be more precise, the data that are collected and maintained for the welfare of the people belonging to the state. We are thankful to Kautilya who had kept a record of births and deaths as well as some other precious records in his famous book 'Arthashastra' during Chandragupta's reign in the fourth century B.C. During the reign of Akbar in the sixteenth century A.D. We find statistical records on agriculture in Ain-i-Akbari written by Abu Fazl. Referring to Egypt, the first census was conducted by the Pharaoh during 300 B.C. to 2000 B.C.

Definition of Statistics

We may define statistics either in a singular sense or in a plural sense Statistics, when used as a plural noun, may be defined as data qualitative as well as quantitative, that are collected, usually with a view of having statistical analysis.

However, statistics, when used as a singular noun, may be defined, as the scientific method that is employed for collecting, analysing and presenting data, leading finally to drawing statistical inferences about some important characteristics it means it is 'science of counting' or 'science of averages'.

Application of statistics

Among various applications of statistics, let us confine our discussions to the fields of Economics, Business Management and Commerce and Industry.

Economics

Modern developments in Economics have the roots in statistics. In fact, Economics and Statistics are closely associated. Time Series Analysis, Index Numbers, Demand Analysis etc. are some overlapping areas of Economics and Statistics. In this connection, we may also mention Econometrics – a branch of Economics that interact with statistics in a very positive way. Conducting socio-economic surveys and analysing the data derived from it are made with the help of different statistical methods. Regression analysis, one of the numerous applications of statistics, plays a key role in Economics for making future projection of demand of goods, sales, prices, quantities etc. which are all ingredients of Economic planning.

Business Management

Gone are the days when the managers used to make decisions on the basis of hunches, intuition or trials and errors. Now a days, because of the never-ending complexity in the business and industry environment, most of the decision making processes rely upon different quantitative techniques which could be described as a combination of statistical methods and operations research techniques. So far as statistics is concerned, inferences about the universe from the knowledge of a part of it, known as sample, plays an important role in the development of certain criteria. Statistical decision theory is another component of statistics that focuses on the analysis of complicated business strategies with a list of alternatives – their merits as well as demerits.

Statistics in Commerce and Industry

In this age of cut-throat competition, like the modern managers, the industrialists and the businessmen are expanding their horizons of industries and businesses with the help of statistical procedures. Data on previous sales, raw materials, wages and salaries, products of identical nature of other factories etc are collected, analysed and experts are consulted in order to maximise profits. Measures of central tendency and dispersion, correlation and regression analysis, time series analysis, index numbers, sampling, statistical quality control are some of the statistical methods employed in commerce and industry.

Limitations of Statistics

Before applying statistical methods, we must be aware of the following limitations:

I Statistics deals with the aggregates. An individual, to a statistician has no significance except the fact that it is a part of the aggregate.

II Statistics is concerned with quantitative data. However, qualitative data also can be converted to quantitative data by providing a numerical description to the corresponding qualitative data.

III Future projections of sales, production, price and quantity etc. are possible under a specific set of conditions. If any of these conditions is violated, projections are likely to be inaccurate.

IV The theory of statistical inferences is built upon random sampling. If the rules for random sampling are not strictly adhered to, the conclusion drawn on the basis of these unrepresentative samples would be erroneous. In other words, the experts should be consulted before deciding the sampling scheme.

14.2 COLLECTION OF DATA

We may define 'data' as quantitative information about some particular characteristic(s) under consideration. Although a distinction can be made between a qualitative characteristic and a quantitative characteristic but so far as the statistical analysis of the characteristic is concerned, we need to convert qualitative information to quantitative information by providing a numeric descriptions to the given characteristic. In this connection, we may note that a quantitative characteristic is known as a variable or in other words, a variable is a measurable quantity. Again, a variable may be either discrete or continuous. When a variable assumes a finite or a countably infinite number of isolated values, it is known as a discrete variable. Examples of discrete variables may be found in the number of petals in a flower, the number of misprints a book contains, the number of road accidents in a particular locality and so on. A variable, on the other hand, is known to be continuous if it can assume any value from a given interval. Examples of continuous variables may be provided by height, weight, sale, profit and so on. Finally, a qualitative characteristic is known as an attribute. The gender of a baby, the nationality of a person, the colour of a flower etc. are examples of attributes.

We can broadly classify data as

(a) Primary;

(b) Secondary.

Collection of data plays the very important role for any statistical analysis. The data which are collected for the first time by an investigator or agency are known as primary data whereas the data are known to be secondary if the data, as being already collected, are used by a different person or agency. Thus, if Prof. Das collects the data on the height of every student in his class, then these would be primary data for him. If, however, another person, say, Professor Bhargava uses the data, as collected by Prof. Das, for finding the average height of the students belonging to that class, then the data would be secondary for Prof. Bhargava.

Collection of Primary Data

The following methods are employed for the collection of primary data:

(i) Interview method;

(ii) Mailed questionnaire method;

(iii) Observation method;

(iv) Questionnaries filled and sent by enumerators.

Interview method again could be divided into (a) Personal Interview method, (b) Indirect Interview method and (c) Telephone Interview method.

In personal interview method, the investigator meets the respondents directly and collects the required information then and there from them. In case of a natural calamity like a super cyclone or an earthquake or an epidemic like plague, we may collect the necessary data much more quickly and accurately by applying this method.

If there are some practical problems in reaching the respondents directly, as in the case of a rail accident, then we may take recourse for conducting Indirect Interview where the investigator collects the necessary information from the persons associated with the problems.

Telephone interview method is a quick and rather non-expensive way to collect the primary data where the relevant information can be gathered by the researcher himself by contacting the interviewee over the phone. The first two methods, though more accurate, are inapplicable for covering a large area whereas the telephone interview, though less consistent, has a wide coverage.

The amount of non-responses is maximum for this third method of data collection.

Mailed questionnaire method comprises of framing a well-drafted and soundly-sequenced questionnaire covering all the important aspects of the problem under consideration and sending them to the respondents with pre-paid stamp after providing all the necessary guidelines for filling up the questionnaire. Although a wide area can be covered using the mailed questionnaire method, the amount of non-responses is likely to be maximum in this method.

In observation method, data are collected, as in the case of obtaining the data on the height and weight of a group of students, by direct observation or using instrument. Although this is likely to be the best method for data collection, it is time consuming, laborious and covers only a small area. Questionnaire form of data collection is used for larger enquiries from the persons who are surveyed. Enumerators collects information directly by interviewing the persons having information : Question are explained and hence data is collected.

Sources of Secondary Data

There are many sources of getting secondary data. Some important sources are listed below:

(a) International sources like WHO, ILO, IMF, World Bank etc.

(b) Government sources like Statistical Abstract by CSO, Indian Agricultural Statistics by the Ministry of Food and Agriculture and so on.

(c) Private and quasi-government sources like ISI, ICAR, NCERT etc.

(d) Unpublished sources of various research institutes, researchers etc.

Scrutiny of Data

Since the statistical analyses are made only on the basis of data, it is necessary to check whether the data under consideration are accurate as well as consistence. No hard and fast rules can be recommended for the scrutiny of data. One must apply his intelligence, patience and experience while scrutinising the given information.

Errors in data may creep in while writing or copying the answer on the part of the enumerator. A keen observer can easily detect that type of error. Again, there may be two or more series of figures which are in some way or other related to each other. If the data for all the series are provided, they may be checked for internal consistency. As an example, if the data for population, area and density for some places are given, then we may verify whether they are internally consistent by examining whether the relation

$$\text{Density} = \frac{\text{Area}}{\text{Population}} \text{ holds.}$$

A good statistician can also detect whether the returns submitted by some enumerators are exactly of the same type thereby implying the lack of seriousness on the part of the enumerators. The bias of the enumerator also may be reflected by the returns submitted by him. This type of error can be rectified by asking the enumerator(s) to collect the data for the disputed cases once again.

14.3 PRESENTATION OF DATA

Once the data are collected and verified for their homogeneity and consistency, we need to present them in a neat and condensed form highlighting the essential features of the data. Any statistical analysis is dependent on a proper presentation of the data under consideration.

Classification or Organisation of Data

It may be defined as the process of arranging data on the basis of the characteristic under consideration into a number of groups or classes according to the similarities of the observations. Following are the objectives of classification of data:

(a) It puts the data in a neat, precise and condensed form so that it is easily understood and interpreted.

(b) It makes comparison possible between various characteristics, if necessary, and thereby finding the association or the lack of it between them.

(c) Statistical analysis is possible only for the classified data.

(d) It eliminates unnecessary details and makes data more readily understandable.

Data may be classified as -

(i) Chronological or Temporal or Time Series Data;

(ii) Geographical or Spatial Series Data;

(iii) Qualitative or Ordinal Data;

(iv) Quantitative or Cardinal Data.

When the data are classified in respect of successive time points or intervals, they are known as time series data. The number of students appeared for CA final for the last twenty years, the production of a factory per month from 2000 to 2015 etc. are examples of time series data.

Data arranged region wise are known as geographical data. If we arrange the students appeared for CA final in the year 2015 in accordance with different states, then we come across Geographical Data.

Data classified in respect of an attribute are referred to as qualitative data. Data on nationality, gender, smoking habit of a group of individuals are examples of qualitative data. Lastly, when the data are classified in respect of a variable, say height, weight, profits, salaries etc., they are known as quantitative data.

Data may be further classified as *frequency data* and *non-frequency data*. The qualitative as well as quantitative data belong to the frequency group whereas time series data and geographical data belong to the non-frequency group.

Mode of Presentation of Data

Next, we consider the following mode of presentation of data:

(a) Textual presentation;

(b) Tabular presentation or Tabulation;

(c) Diagrammatic representation.

(a) Textual presentation

This method comprises presenting data with the help of a paragraph or a number of paragraphs. The official report of an enquiry commission is usually made by textual presentation. Following is an example of textual presentation.

'In 2009, out of a total of five thousand workers of Roy Enamel Factory, four thousand and two hundred were members of a Trade Union. The number of female workers was twenty per cent of the total workers out of which thirty per cent were members of the Trade Union.

In 2010, the number of workers belonging to the trade union was increased by twenty per cent as compared to 2009 of which four thousand and two hundred were male. The number of workers not belonging to trade union was nine hundred and fifty of which four hundred and fifty were females.'

The merit of this mode of presentation lies in its simplicity and even a layman can present data by this method. The observations with exact magnitude can be presented with the help of textual presentation. Furthermore, this type of presentation can be taken as the first step towards the other methods of presentation.

Textual presentation, however, is not preferred by a statistician simply because, it is dull, monotonous and comparison between different observations is not possible in this method. For manifold classification, this method cannot be recommended.

(b) Tabular presentation or Tabulation

Tabulation may be defined as systematic presentation of data with the help of a statistical table having a number of rows and columns and complete with reference number, title, description of rows as well as columns and foot notes, if any.

We may consider the following guidelines for tabulation :

I A statistical table should be allotted a serial number along with a self-explanatory title.

II The table under consideration should be divided into caption, Box-head, Stub and Body. Caption is the upper part of the table, describing the columns and sub-columns, if any. The Box-head is the entire upper part of the table which includes columns and sub-column numbers, unit(s) of measurement along with caption. Stub is the left part of the table providing the description of the rows. The body is the main part of the table that contains the numerical figures.

III The table should be well-balanced in length and breadth.

IV The data must be arranged in a table in such a way that comparison(s) between different figures are made possible without much labour and time. Also the row totals, column totals, the units of measurement must be shown.

V The data should be arranged intelligently in a well-balanced sequence and the presentation of data in the table should be appealing to the eyes as far as practicable.

VI Notes describing the source of the data and bringing clarity and, if necessary, about any rows or columns known as footnotes, should be shown at the bottom part of the table.

The textual presentation of data, relating to the workers of Roy Enamel Factory is shown in the following table.

Table 14.1

Status of the workers of Roy Enamel factory on the basis of their trade union membership for 2009 and 2010.

Status	Member of TU			Non-member			Total		
Year	M (1)	F (2)	T (3)=(1)+ (2)	M (4)	F (5)	T (6)=(4)+ (5)	M (7)	F (8)	T (9)=(7)+ (8)
2009	3900	300	4200	300	500	800	4200	800	5000
2010	4200	840	5040	500	450	950	4700	1290	5990

Source:

Footnote: TU, M, F and T stand for trade union, male, female and total respectively.

The tabulation method is usually preferred to textual presentation as

(i) It facilitates comparison between rows and columns.

(ii) Complicated data can also be represented using tabulation.

(iii) It is a must for diagrammatic representation.

(iv) Without tabulation, statistical analysis of data is not possible.

(c) Diagrammatic representation of data

Another alternative and attractive representation of statistical data is provided by charts, diagrams and pictures. Unlike the first two methods of representation of data, diagrammatic representation can be used for both the educated section and uneducated section of the society. Furthermore, any hidden trend present in the given data can be noticed only in this mode of representation. However, compared to tabulation, this is less accurate. So if there is a priority for accuracy, we have to recommend tabulation.

We are going to consider the following types of diagrams :

I Line diagram or Historiagram;

II Bar diagram;

III Pie chart.

I **Line diagram or Historiagram**

When the data vary over time, we take recourse to line diagram. In a simple line diagram, we plot each pair of values of (t, y_t), y_t representing the time series at the time point t in the $t–y_t$ plane. The plotted points are then joined successively by line segments and the resulting chart is known as line-diagram.

When the time series exhibit a wide range of fluctuations, we may think of logarithmic or ratio chart where Log y_t and not y_t is plotted against t. We use Multiple line chart for representing two or more related time series data expressed in the same unit and multiple – axis chart in somewhat similar situations if the variables are expressed in different units.

II **Bar diagram**

There are two types of bar diagrams namely, Horizontal Bar diagram and Vertical Bar diagram. While horizontal bar diagram is used for qualitative data or data varying over space, the vertical bar diagram is associated with quantitative data or time series data. Bars i.e. rectangles of equal width and usually of varying lengths are drawn either horizontally or vertically. We consider Multiple or Grouped Bar diagrams to compare related series. Component or sub-divided Bar diagrams are applied for representing data divided into a number of components. Finally, we use Divided Bar charts or Percentage Bar diagrams for comparing different components of a variable and also the relating of the components to the whole. For this situation, we may also use Pie chart or Pie diagram or circle diagram.

ILLUSTRATIONS:

Example 14.1: The profits in lakhs of Rupees of an industrial house for 2009, 2010, 2011, 2012, 2013, 2014, and 2015 are 5, 8, 9, 6, 12, 15 and 24 respectively. Represent these data using a suitable diagram.

SOLUTION:

We can represent the profits for 7 consecutive years by drawing either a line chart or a vertical bar chart. Fig. 14.1 shows a line chart and figure 14.2 shows the corresponding vertical bar chart.

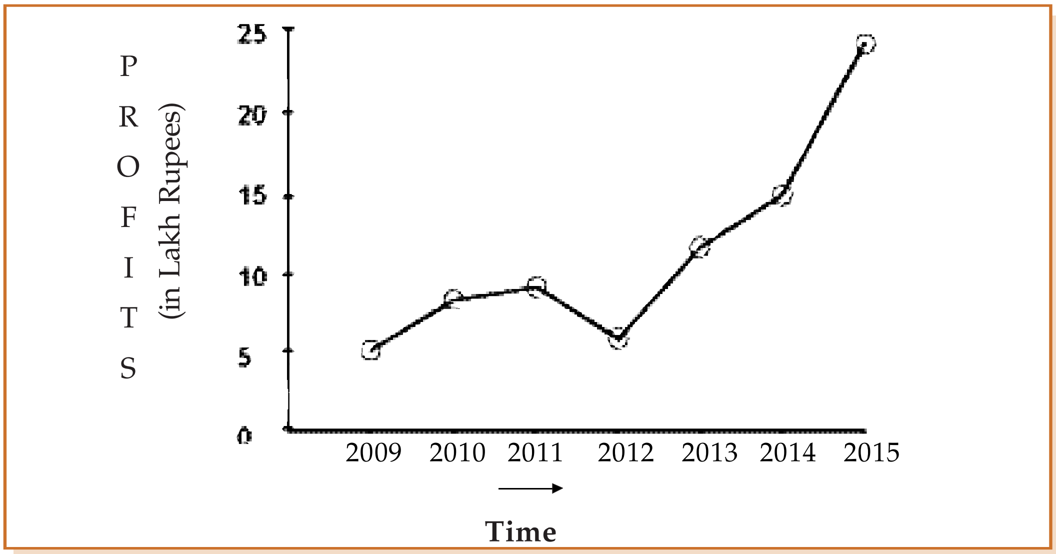


Figure 14.1

Showing line chart for the Profit of an Industrial House during 2002 to 2008.

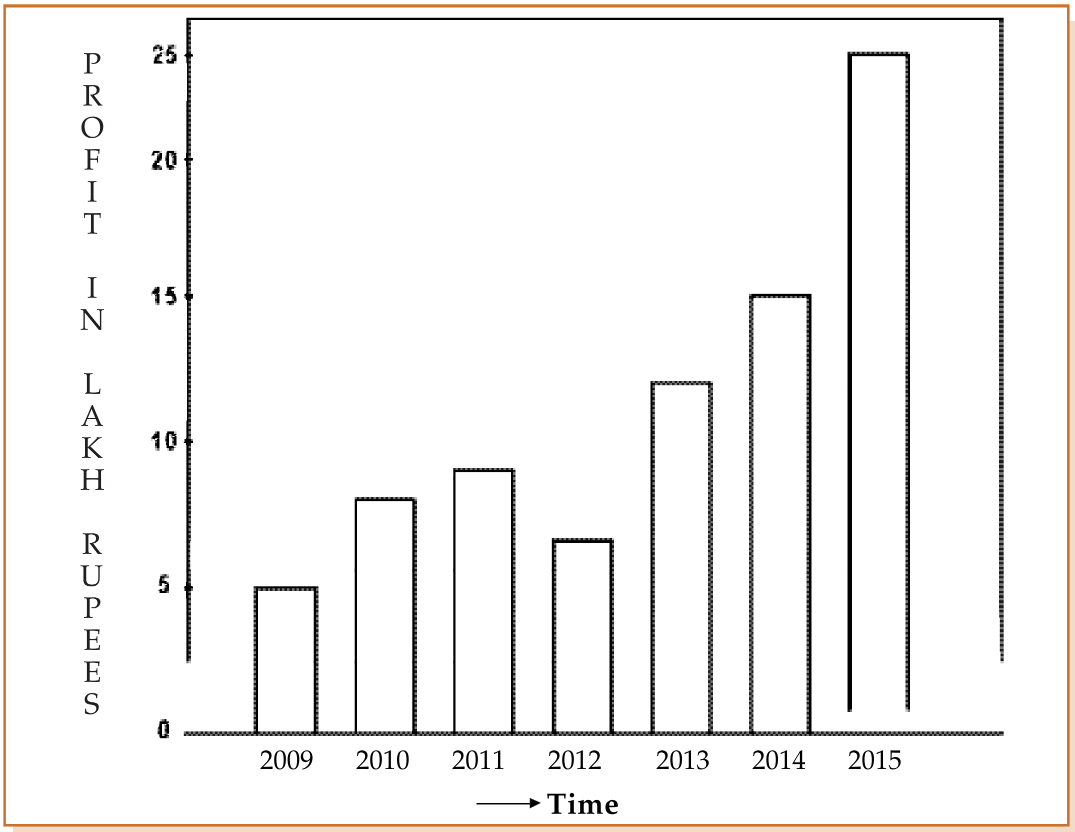


Figure 14.2

Showing vertical bar diagram for the Profit of an Industrial house from 2007 to 2015.

Example 14.2: The production of wheat and rice of a region are given below :

Year	Production in metric tones	
	Wheat	Rice
2012	12	25
2013	15	30
2014	18	32
2015	19	36

Represent this information using a suitable diagram.

Solution:

We can represent this information by drawing a multiple line chart. Alternately, a multiple bar diagram may be considered. These are depicted in figure 14.3 and 14.4 respectively.

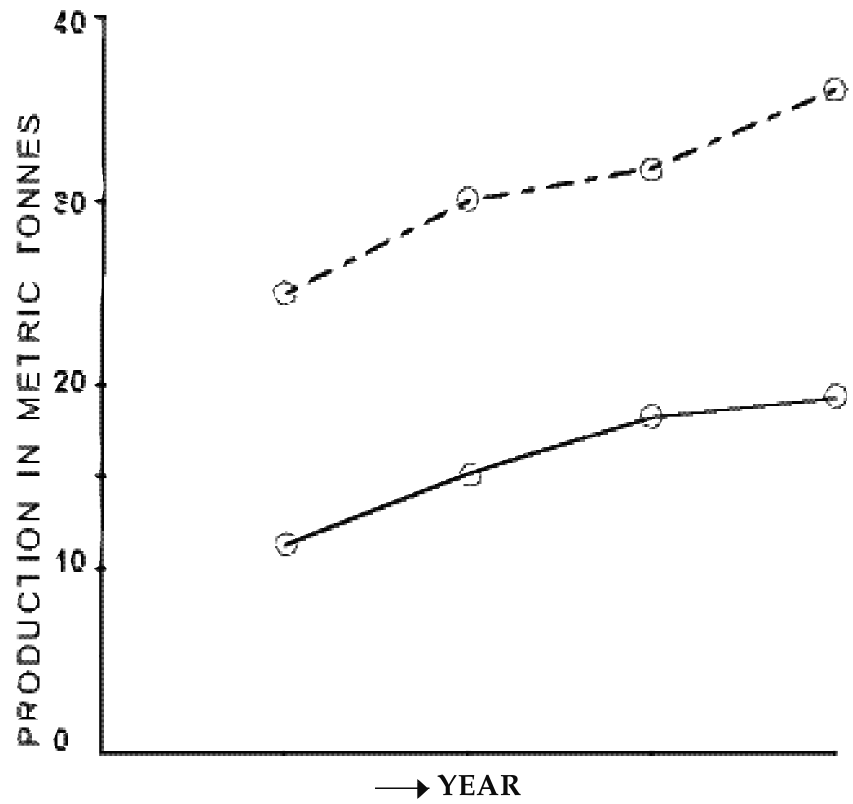


Figure 14.3

Multiple line chart showing production of wheat and rice of a region during 2012–2015. (Dotted line represent production of rice and continuous line that of wheat).

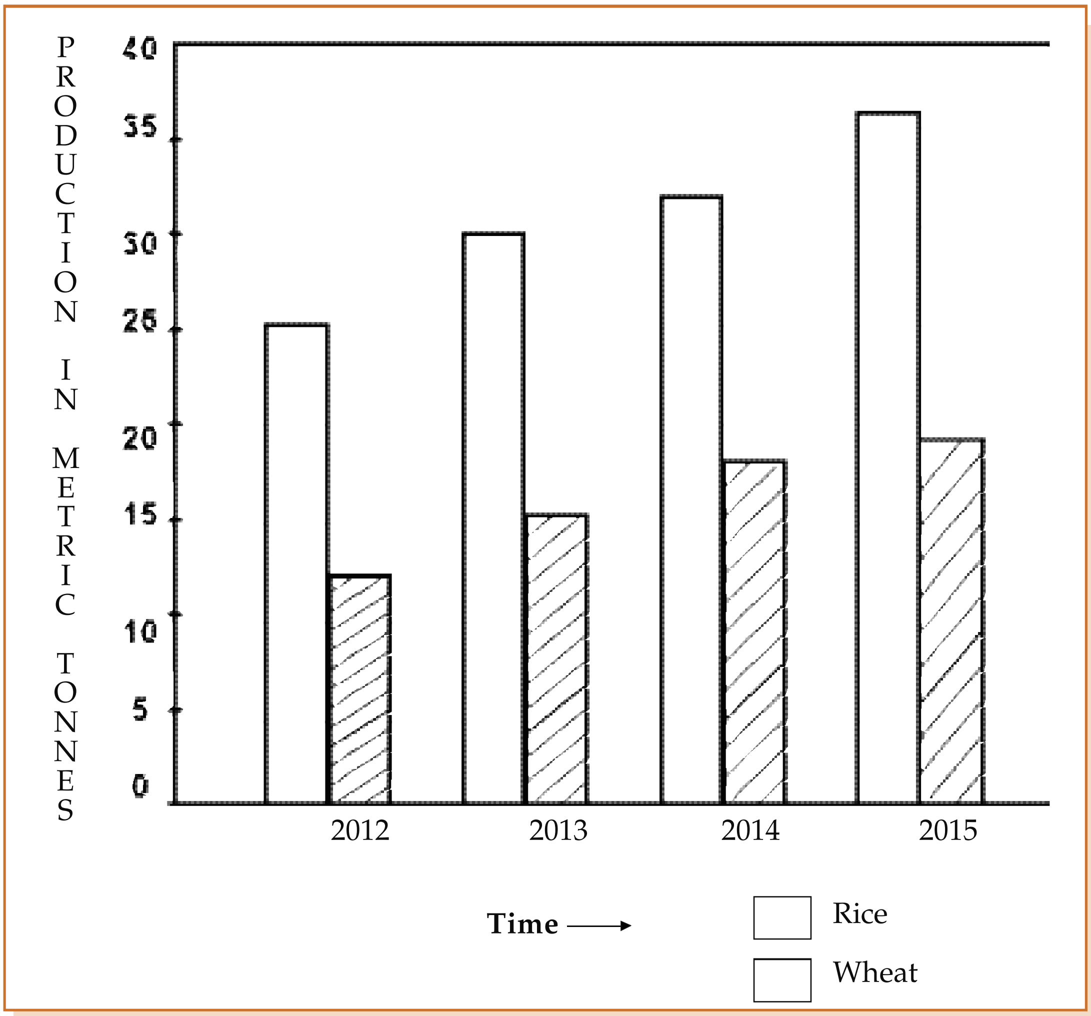


Figure 14.4

Multiple bar chart representing production of rice and wheat from 2012 to 2015.

Example 14.3: Draw an appropriate diagram with a view to represent the following data :

Source	Revenue in millions of (₹)
Customs	80
Excise	190
Income Tax	160
Corporate Tax	75
Miscellaneous	35

Solution:

Pie chart or divided bar chart would be the ideal diagram to represent this data. We consider Pie chart.

Table 14.2

Computation for drawing Pie chart

Source (1)	Revenue in Million rupees (2)	Central angle $= \frac{(2)}{\text{Total of (2)}} \times 360^o$
Customs	80	$\frac{80}{540} \times 360^o = 53^o$ (approx.)
Excise	190	$\frac{190}{540} \times 360^o = 127^o$
Income Tax	160	$\frac{160}{540} \times 360^o = 107^o$
Corporate Tax	75	$\frac{75}{540} \times 360^o = 50^o$
Miscellaneous	35	$\frac{35}{540} \times 360^o = 23^o$
Total	540	360^0

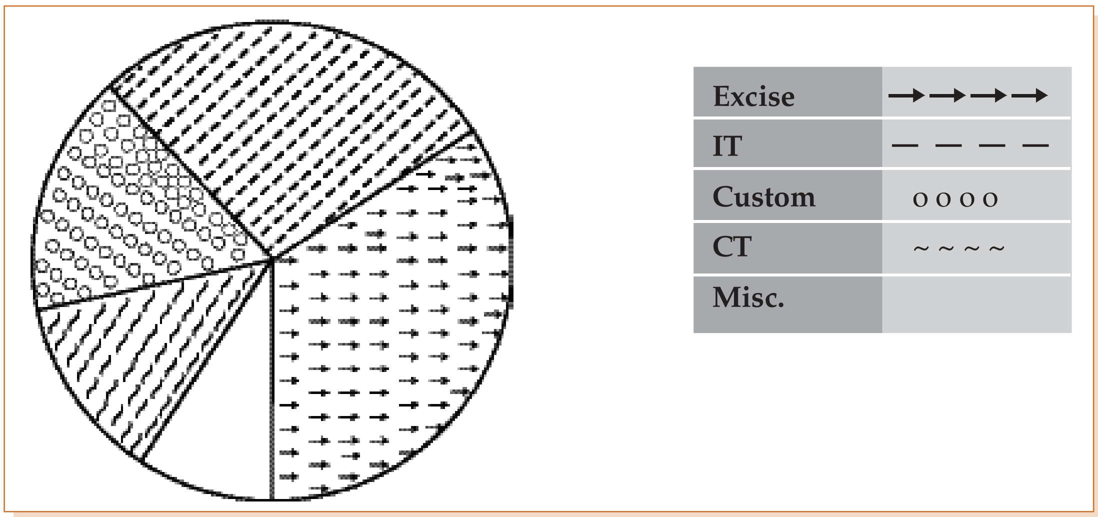


Figure 14.5

Pie chart showing the distribution of Revenue

14.4 FREQUENCY DISTRIBUTION

As discussed in the previous section, frequency data occur when we classify statistical data in respect of either a variable or an attribute. A frequency distribution may be defined as a tabular representation of statistical data, usually in an ascending order, relating to a measurable characteristic according to individual value or a group of values of the characteristic under study.

In case, the characteristic under consideration is an attribute, say nationality, then the tabulation is made by allotting numerical figures to the different classes the attribute may belong like, in this illustration, counting the number of Indian, British, French, German and so on. The qualitative characteristic is divided into a number of categories or classes which are mutually exclusive and exhaustive and the figures against all these classes are recorded. The figure corresponding to a particular class, signifying the number of times or how frequently a particular class occurs is known as the frequency of that class. Thus, the number of Indians, as found from the given data, signifies the frequency of the Indians. So frequency distribution is a statistical table that distributes the total frequency to a number of classes.

When tabulation is done in respect of a discrete random variable, it is known as Discrete or Ungrouped or simple Frequency Distribution and in case the characteristic under consideration is a continuous variable, such a classification is termed as Grouped Frequency Distribution. In case of a grouped frequency distribution, tabulation is done not against a single value as in the case of an attribute or a discrete random variable but against a group of values. The distribution of the number of car accidents in Delhi during 12 months of the year 2005 is an example of a ungrouped frequency distribution and the distribution of heights of the students of St. Xavier's College for the year 2004 is an example of a grouped frequency distribution.

Example 14.4: Following are the records of babies born in a nursing home in Bangalore during a week (B denoting Boy and G for Girl) :

B	G	G	B	G	G	B	B	G	G
G	G	B	B	B	G	B	B	G	B
B	B	G	B	B	B	G	G	B	G

Construct a frequency distribution according to gender.

Solution:

In order to construct a frequency distribution of babies in accordance with their gender, we count the number of male births and that of female births and present this information in the following table.

Table 14.3

Frequency distribution of babies according to Gender

Category	Number of births
Boy (B)	16
Girl (G)	14
Total	30

Frequency Distribution of a Variable

For the construction of a frequency distribution of a variable, we need to go through the following steps :

I Find the largest and smallest observations and obtain the difference between them, known as Range, in case of a continuous variable.

II Form a number of classes depending on the number of isolated values assumed by a discrete variable. In case of a continuous variable, find the number of class intervals using the relation, No. of class Interval X class length $\cong$ Range.

III Present the class or class interval in a table known as frequency distribution table.

IV Apply 'tally mark' i.e. a stroke against the occurrence of a particulars value in a class or class interval.

V Count the tally marks and present these numbers in the next column, known as frequency column, and finally check whether the total of all these class frequencies tally with the total number of observations.

Example 14.5: A review of the first 30 pages of a statistics book reveals the following printing mistakes:

0	1	3	3	2	5	6	0	1	0
4	1	1	0	2	3	2	5	0	4
2	3	2	2	3	3	4	6	1	4

Make a frequency distribution of printing mistakes.

Solution:

Since x, the printing mistakes, is a discrete variable, x can assume seven values 0, 1, 2, 3, 4, 5 and 6. Thus we have 7 classes, each class comprising a single value.

Table 14.4

Frequency Distribution of the number of printing mistakes of the first 30 pages of a book

Printing Mistake	Tally marks	Frequency (No. of Pages)
0	~~IIII~~	5
1	~~IIII~~	5
2	~~IIII~~ I	6
3	~~IIII~~ I	6
4	IIII	4
5	II	2
6	II	2
Total	–	30

Example 14.6: Following are the weights in kgs. of 36 BBA students of St. Xavier's College.

70	73	49	61	61	47	57	50	59
59	68	45	55	65	68	56	68	55
70	70	57	44	69	73	64	49	63
65	70	65	62	64	73	67	60	50

Construct a frequency distribution of weights, taking class length as 5.

Solution:

We have, Range = Maximum weight – Minimum weight

= 73 kgs. – 44 kgs.

= 29 kgs.

No. of class interval × class lengths ≅ Range

⇒ No. of class interval × 5 ≅ 29

⇒ No. of class interval $= \frac{29}{5} \cong 6.$

(We always take the next integer as the number of class intervals so as to include both the minimum and maximum values).

Table 14.5

Frequency Distribution of weights of 36 BBA Students

Weight in kg (Class Interval)	Tally marks	No. of Students (Frequency)
44-48	III	3
49-53	IIII	4
54-58	~~IIII~~	5
59-63	~~IIII~~ II	7
64-68	~~IIII~~ IIII	9
69-73	~~IIII~~ III	8
Total	–	36

Some important terms associated with a frequency distribution

Class Limit (CL)

Corresponding to a class interval, the class limits may be defined as the minimum value and the maximum value the class interval may contain. The minimum value is known as the lower class limit (LCL) and the maximum value is known as the upper class limit (UCL). For the frequency distribution of weights of BBA Students, the LCL and UCL of the first class interval are 44 kgs. and 48 kgs. respectively.

Class Boundary (CB)

Class boundaries may be defined as the actual class limit of a class interval. For overlapping classification or mutually exclusive classification that excludes the upper class limits like 10–20, 20–30, 30–40, ……… etc. the class boundaries coincide with the class limits. This is usually done for a continuous variable. However, for non-overlapping or mutually inclusive classification that includes both the class limits like 0–9, 10–19, 20–29,…… which is usually applicable for a discrete variable, we have

$$\text{LCB} = \text{LCL} - \frac{D}{2}$$

and $$\text{UCB} = \text{UCL} + \frac{D}{2}$$

where D is the difference between the LCL of the next class interval and the UCL of the given class interval. For the data presented in table 10.5, LCB of the first class interval

$$= 44 \text{ kgs.} - \frac{(49 - 48)}{2} \text{ kgs.}$$

$$= 43.50 \text{ kgs.}$$

and the corresponding UCB

$= 48\text{ kgs.} + \frac{49 - 48}{2}\text{ kgs.}$

= 48.50 kgs.

Mid-point or Mid-value or class mark

Corresponding to a class interval, this may be defined as the total of the two class limits or class boundaries to be divided by 2. Thus, we have

$$\text{mid-point} = \frac{LCL + UCL}{2}$$

$$= \frac{LCB + UCB}{2}$$

Referring to the distribution of weight of BBA students, the mid-points for the first two class intervals are

$$\frac{44\text{ kgs.} + 48\text{ kgs.}}{2} \text{ and } \frac{49\text{ kgs.} + 53\text{ kgs.}}{2}$$

i.e. 46 kgs. and 51 kgs. respectively.

Width or size of a class interval

The width of a class interval may be defined as the difference between the UCB and the LCB of that class interval. For the distribution of weights of BBA students, C, the class length or width is 48.50 kgs. – 43.50 kgs. = 5 kgs. for the first class interval. For the other class intervals also, C remains same.

Cumulative Frequency

The cumulative frequency corresponding to a value for a discrete variable and corresponding to a class boundary for a continuous variable may be defined as the number of observations less than the value or less than or equal to the class boundary. This definition refers to the less than cumulative frequency. We can define more than cumulative frequency in a similar manner. Both types of cumulative frequencies are shown in the following table.

Table 14.6

Cumulative Frequency Distribution of weights of 36 BBA students

Weight in kg (CB)	Cumulative Frequency	
	Less than	More than
43.50	0	33 + 3 or 36
48.50	0 + 3 or 3	29 + 4 or 33
53.50	3 + 4 or 7	24 + 5 or 29
58.50	7 + 5 or 12	17 + 7 or 24
63.50	12 + 7 or 19	8 + 9 or 17
68.50	19 + 9 or 28	0 + 8 or 8
73.50	28 + 8 or 36	0

Frequency density of a class interval

It may be defined as the ratio of the frequency of that class interval to the corresponding class length. The frequency densities for the first two class intervals of the frequency distribution of weights of BBA students are 3/5 and 4/5 i.e. 0.60 and 0.80 respectively.

Relative frequency and percentage frequency of a class interval

Relative frequency of a class interval may be defined as the ratio of the class frequency to the total frequency. Percentage frequency of a class interval may be defined as the ratio of class frequency to the total frequency, expressed as a percentage. For the last example, the relative frequencies for the first two class intervals are 3/36 and 4/36 respectively and the percentage frequencies are 300/36 and 400/36 respectively. It is quite obvious that whereas the relative frequencies add up to unity, the percentage frequencies add up to one hundred.

14.5 GRAPHICAL REPRESENTATION OF A FREQUENCY DISTRIBUTION

We consider the following types of graphical representation of frequency distribution :

(i) Histogram or Area diagram;

(ii) Frequency Polygon;

(iii) Ogives or cumulative Frequency graphs.

(i) **Histogram or Area diagram**

This is a very convenient way to represent a frequency distribution. Histogram helps us to get an idea of the frequency curve of the variable under study. Some statistical measure can be obtained using a histogram. A comparison among the frequencies for different class intervals is possible in this mode of diagrammatic representation.

In order to draw a histogram, the class limits are first converted to the corresponding class boundaries and a series of adjacent rectangles, one against each class interval, with the class interval as base or breadth and the frequency or frequency density usually when the class intervals are not uniform as length or altitude, is erected. The histogram for the distribution of weight of 36 BBA students is shown below. The mode of the weights has also been determined using the histogram.

i.e. Mode = 66.50 kgs.

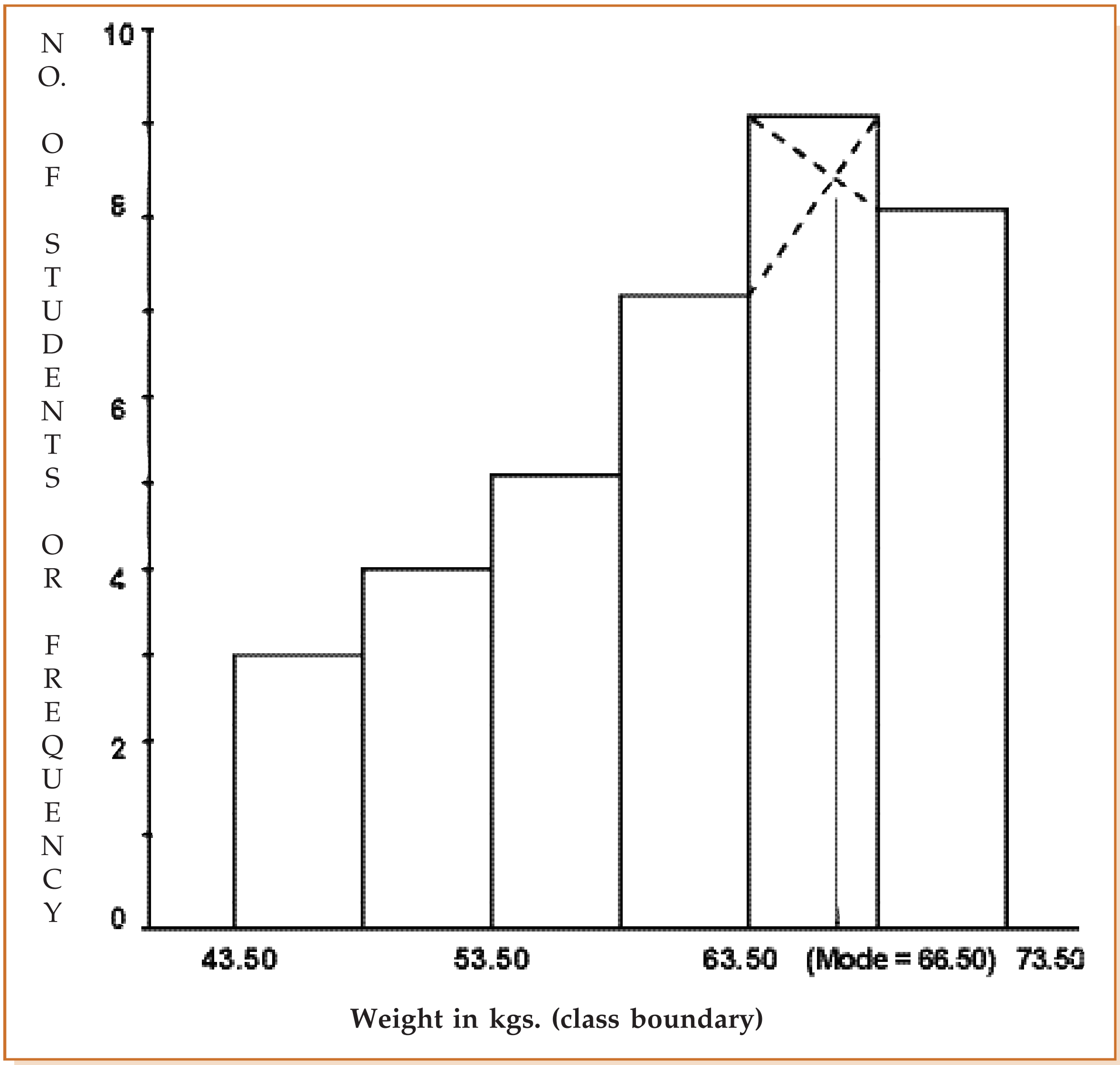


Figure 14.6

Showing histogram for the distribution of weight of 36 BBA students

(ii) **Frequency Polygon**

Usually frequency polygon is meant for single frequency distribution. However, we also apply it for grouped frequency distribution provided the width of the class intervals remains the same. A frequency curve can be regarded as a limiting form of frequency polygon. In order to draw a frequency polygon, we plot (x_i, f_i) for i = 1, 2, 3, ……….. n with x_i denoting the mid-point of the its class interval and f_i, the corresponding frequency, n being the number of class intervals. The plotted points are joined successively by line segments and the figure, so drawn, is given the shape of a polygon, a closed figure, by joining the two extreme ends of the drawn figure to two additional points $(x_0,0)$ and $(x_{n+1},0)$.

The frequency polygon for the distribution of weights of BBA students is shown in Figure 14.7. We can also obtain a frequency polygon starting with a histogram by adding the mid-points of the upper sides of the rectangles successively and then completing the figure by joining the two ends as before.

Mid-points	No. of Students (Frequency)
46	3
51	4
56	5
61	7
66	9
71	8

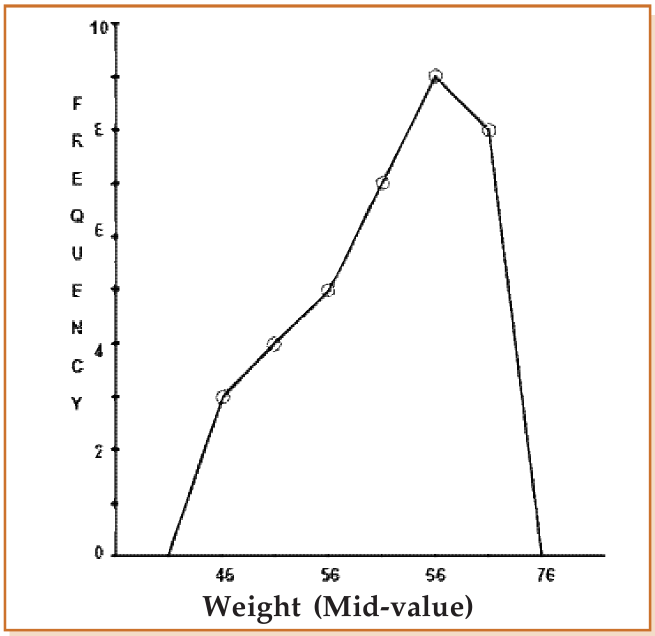


Figure 14.7

Showing frequency polygon for the distribution of height of 36 BBA students

(iii) Ogives or Cumulative Frequency Graph

By plotting cumulative frequency against the respective class boundary, we get ogives. As such there are two ogives – less than type ogives, obtained by taking less than cumulative frequency on the vertical axis and more than type ogives by plotting more than type cumulative frequency on the vertical axis and thereafter joining the plotted points successively by line segments. Ogives may be considered for obtaining quartiles graphically. If a perpendicular is drawn from the point of intersection of the two ogives on the horizontal axis, then the x-value of this point gives us the value of median, the second or middle quartile. Ogives further can be put into use for making short term projections.

Figure 14.8 depicts the ogives and the determination of the quartiles. This figure give us the following information.

1st quartile or lower quartile (Q_1) = 55 kgs.

2nd quartile or median (Q_2 or Me) = 62.50 kgs.

3rd quartile or upper quartile (Q_3) = 68 kgs.

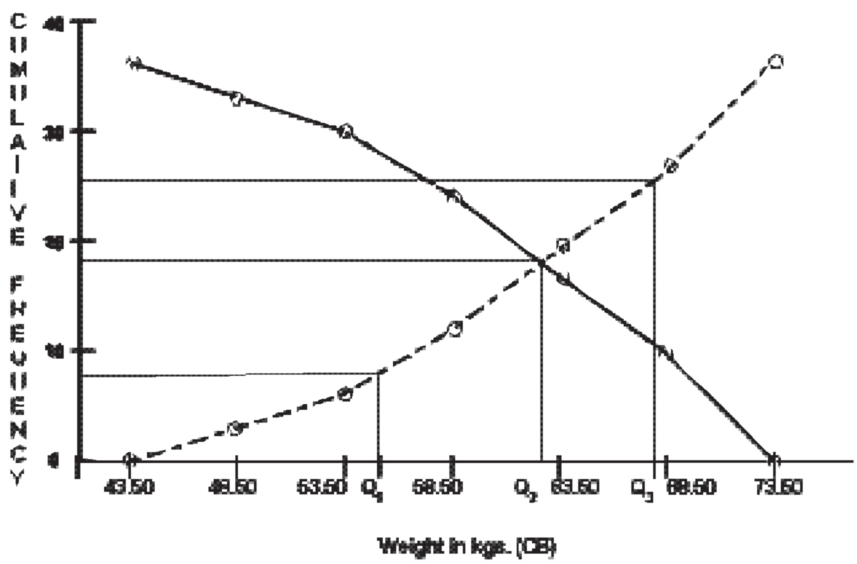


Figure 14.8

Showing the ogives for the distribution of weights of 36 BBA students

We find Q_1 = 55 kgs.

Q_2 = Me = 62.50 kgs.

Q_3 = 68 kgs.

Frequency Curve

A frequency curve is a smooth curve for which the total area is taken to be unity. It is a limiting form of a histogram or frequency polygon. The frequency curve for a distribution can be obtained by drawing a smooth and free hand curve through the mid-points of the upper sides of the rectangles forming the histogram.

There exist four types of frequency curves namely

(a) Bell-shaped curve;

(b) U-shaped curve;

(c) J-shaped curve;

(d) Mixed curve.

Most of the commonly used distributions provide bell-shaped curve, which, as suggested by the name, looks almost like a bell. The distribution of height, weight, mark, profit etc. usually belong to this category. On a bell-shaped curve, the frequency, starting from a rather low value, gradually reaches the maximum value, somewhere near the central part and then gradually decreases to reach its lowest value at the other extremity.

For a U-shaped curve, the frequency is minimum near the central part and the frequency slowly but steadily reaches its maximum at the two extremities. The distribution of Kolkata bound commuters belongs to this type of curve as there are maximum number of commuters during the peak hours in the morning and in the evening.

The J-shaped curve starts with a minimum frequency and then gradually reaches its maximum frequency at the other extremity. The distribution of commuters coming to Kolkata from the early morning hour to peak morning hour follows such a distribution. Sometimes, we may also come across an inverted J-shaped frequency curve.

Lastly, we may have a combination of these frequency curves, known as mixed curve. These are exhibited in the following figures.

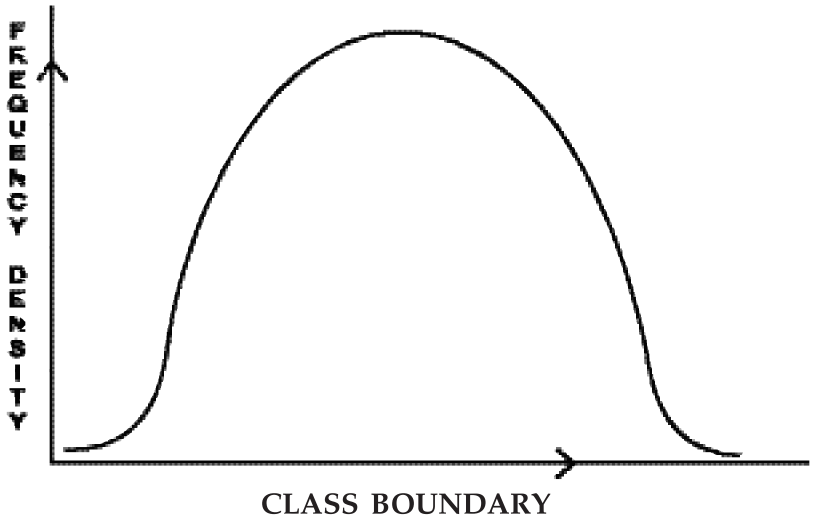


Figure 14.9

Bell-shaped curve

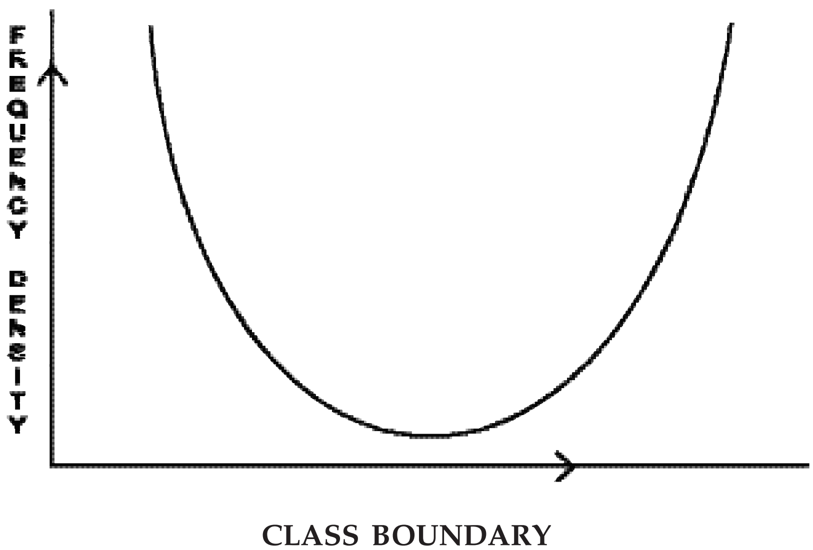


Figure 14.10

U-shaped curve

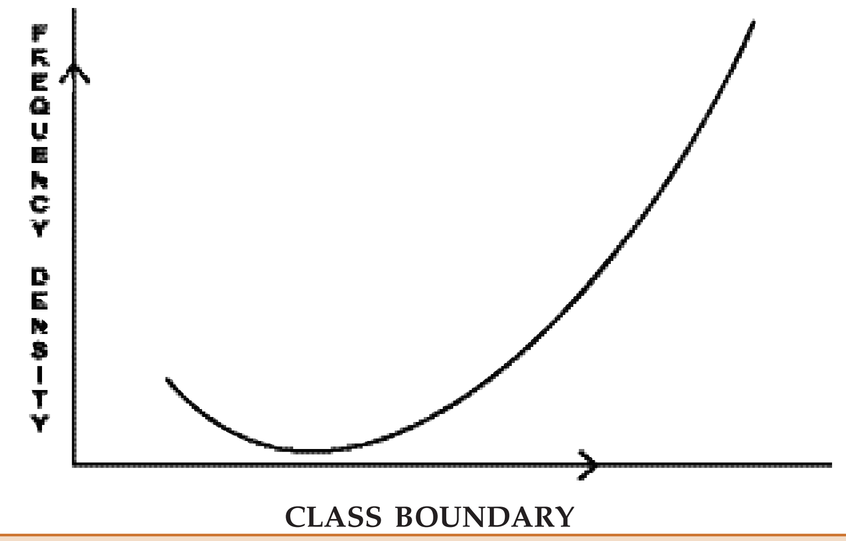


Figure 14.11

J-shaped curve

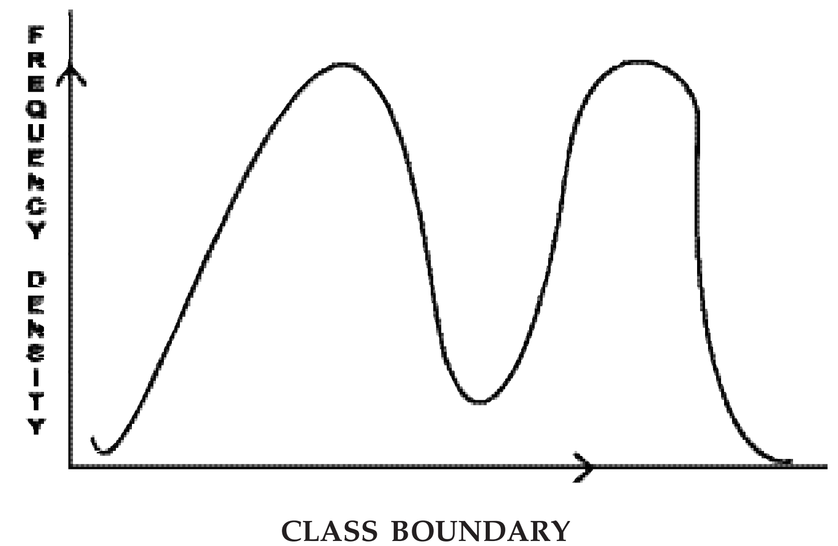


Figure 14.12

Mixed curve

SUMMARY

- Statistics deals with the aggregates. An individual, to a statistician has no significance except the fact that it is a part of the aggregate.
- Statistics is concerned with quantitative data. However, qualitative data also can be converted to quantitative data by providing a numerical description to the corresponding qualitative data.
- The theory of statistical inferences is built upon random sampling. If the rules for random sampling are not strictly adhered to, the conclusion drawn on the basis of these unrepresentative samples would be erroneous.
- We can broadly classify data as

  (a) Primary;

  (b) Secondary.
- Mode of Presentation of Data

  (a) Textual presentation;

  (b) Tabular presentation or Tabulation;

  (c) Diagrammatic representation.
- The types of diagrams:

  (a) Line diagram or Historiagram;

  (b) Bar diagram;

  (c) Pie chart.
- Frequency Distribution of a Variable

  (a) Find the largest and smallest observations and obtain the difference between them, known as Range, in case of a continuous variable.

  (b) Form a number of classes depending on the number of isolated values assumed by a discrete variable. In case of a continuous variable, find the number of class intervals using the relation, No. of class Interval X class length $\cong$ Range.

  (c) Present the class or class interval in a table known as frequency distribution table.

  (d) Apply 'tally mark' i.e. a stroke against the occurrence of a particulars value in a class or class interval.

  (e) Count the tally marks and present these numbers in the next column, known as frequency column, and finally check whether the total of all these class frequencies tally with thetotal number of observations.

EXERCISE

Set A

Answer the following questions. Each question carries 1 mark.

1. Which of the following statements is false?
   (a) Statistics is derived from the Latin word 'Status'
   (b) Statistics is derived from the Italian word 'Statista'
   (c) Statistics is derived from the French word 'Statistik'
   (d) None of these.
2. Statistics is defined in terms of numerical data in the
   (a) Singular sense (b) Plural sense
   (c) Either (a) or (b) (d) Both (a) and (b).
3. Statistics is applied in
   (a) Economics (b) Business management
   (c) Commerce and industry (d) All these.
4. Statistics is concerned with
   (a) Qualitative information (b) Quantitative information
   (c) (a) or (b) (d) Both (a) and (b).
5. An attribute is
   (a) A qualitative characteristic (b) A quantitative characteristic
   (c) A measurable characteristic (d) All these.
6. Annual income of a person is
   (a) An attribute (b) A discrete variable
   (c) A continuous variable (d) (b) or (c).
7. Marks of a student is an example of
   (a) An attribute (b) A discrete variable
   (c) A continuous variable (d) None of these.
8. Nationality of a student is
   (a) An attribute (b) A continuous variable
   (c) A discrete variable (d) (a) or (c).
9. Drinking habit of a person is
   (a) An attribute (b) A variable
   (c) A discrete variable (d) A continuous variable.

10. Age of a person is
    (a) An attribute (b) A discrete variable
    (c) A continuous variable (d) A variable.
11. Data collected on religion from the census reports are
    (a) Primary data (b) Secondary data
    (c) Sample data (d) (a) or (b).
12. The data collected on the height of a group of students after recording their heights with a measuring tape are
    (a) Primary data (b) Secondary data
    (c) Discrete data (d) Continuous data.
13. The primary data are collected by
    (a) Interview method (b) Observation method
    (c) Questionnaire method (d) All these.
14. The quickest method to collect primary data is
    (a) Personal interview (b) Indirect interview
    (c) Telephone interview (d) By observation.
15. The best method to collect data, in case of a natural calamity, is
    (a) Personal interview (b) Indirect interview
    (c) Questionnaire method (d) Direct observation method.
16. In case of a rail accident, the appropriate method of data collection is by
    (a) Personal interview (b) Direct interview
    (c) Indirect interview (d) All these.
17. Which method of data collection covers the widest area?
    (a) Telephone interview method (b) Mailed questionnaire method
    (c) Direct interview method (d) All these.
18. The amount of non-responses is maximum in
    (a) Mailed questionnaire method (b) Interview method
    (c) Observation method (d) All these.
19. Some important sources of secondary data are
    (a) International and Government sources
    (b) International and primary sources
    (c) Private and primary sources
    (d) Government sources.

20. Internal consistency of the collected data can be checked when

    (a) Internal data are given (b) External data are given

    (c) Two or more series are given (d) A number of related series are given.

21. The accuracy and consistency of data can be verified by

    (a) Internal checking (b) External checking

    (c) Scrutiny (d) Both (a) and (b).

22. The mode of presentation of data are

    (a) Textual, tabulation and diagrammatic (b) Tabular, internal and external

    (c) Textual, tabular and internal (d) Tabular, textual and external.

23. The best method of presentation of data is

    (a) Textual (b) Tabular

    (c) Diagrammatic (d) (b) and (c).

24. The most attractive method of data presentation is

    (a) Tabular (b) Textual

    (c) Diagrammatic (d) (a) or (b).

25. For tabulation, 'caption' is

    (a) The upper part of the table (b) The lower part of the table

    (c) The main part of the table (d) The upper part of a table that describes the column and sub-column.

26. 'Stub' of a table is the

    (a) Left part of the table describing the columns

    (b) Right part of the table describing the columns

    (c) Right part of the table describing the rows (d) All these.

    (d) Left part of the table describing the rows.

27. The entire upper part of a table is known as

    (a) Caption (b) Stub

    (c) Box head (d) Body.

28. The unit of measurement in tabulation is shown in

    (a) Box head (b) Body

    (c) Caption (d) Stub.

29. In tabulation source of the data, if any, is shown in the
    (a) Footnote (b) Body
    (c) Stub (d) Caption.
30. Which of the following statements is untrue for tabulation?
    (a) Statistical analysis of data requires tabulation
    (b) It facilitates comparison between rows and not columns
    (c) Complicated data can be presented
    (d) Diagrammatic representation of data requires tabulation.
31. Hidden trend, if any, in the data can be noticed in
    (a) Textual presentation (b) Tabulation
    (c) Diagrammatic representation (d) All these.
32. Diagrammatic representation of data is done by
    (a) Diagrams (b) Charts
    (c) Pictures (d) All these.
33. The most accurate mode of data presentation is
    (a) Diagrammatic method (b) Tabulation
    (c) Textual presentation (d) None of these.
34. The chart that uses logarithm of the variable is known as
    (a) Line chart (b) Ratio chart
    (c) Multiple line chart (d) Component line chart.
35. Multiple line chart is applied for
    (a) Showing multiple charts
    (b) Two or more related time series when the variables are expressed in the same unit
    (c) Two or more related time series when the variables are expressed in different unit
    (d) Multiple variations in the time series.
36. Multiple axis line chart is considered when
    (a) There is more than one time series (b) The units of the variables are different
    (c) (a) or (b) (d) (a) and (b).
37. Horizontal bar diagram is used for
    (a) Qualitative data (b) Data varying over time
    (c) Data varying over space (d) (a) or (c).

38. Vertical bar diagram is applicable when
    (a) The data are qualitative
    (b) The data are quantitative
    (c) When the data vary over time
    (d) (a) or (c).
39. Divided bar chart is considered for
    (a) Comparing different components of a variable
    (b) The relation of different components to the table
    (c) (a) or (b)
    (d) (a) and (b).
40. In order to compare two or more related series, we consider
    (a) Multiple bar chart
    (b) Grouped bar chart
    (c) (a) or (b)
    (d) (a) and (b).
41. Pie-diagram is used for
    (a) Comparing different components and their relation to the total
    (b) Representing qualitative data in a circle
    (c) Representing quantitative data in circle
    (d) (b) or (c).
42. A frequency distribution
    (a) Arranges observations in an increasing order
    (b) Arranges observation in terms of a number of groups
    (c) Relaters to a measurable characteristic
    (d) All these.
43. The frequency distribution of a continuous variable is known as
    (a) Grouped frequency distribution
    (b) Simple frequency distribution
    (c) (a) or (b)
    (d) (a) and (b).

44. The distribution of shares is an example of the frequency distribution of
    (a) A discrete variable
    (b) A continuous variable
    (c) An attribute
    (d) (a) or (c).
45. The distribution of profits of a blue-chip company relates to
    (a) Discrete variable
    (b) Continuous variable
    (c) Attributes
    (d) (a) or (b).
46. Mutually exclusive classification
    (a) Excludes both the class limits
    (b) Excludes the upper class limit but includes the lower class limit
    (c) Includes the upper class limit but excludes the upper class limit
    (d) Either (b) or (c).
47. Mutually inclusive classification is usually meant for
    (a) A discrete variable
    (b) A continuous variable
    (c) An attribute
    (d) All these.
48. Mutually exclusive classification is usually meant for
    (a) A discrete variable
    (b) A continuous variable
    (c) An attribute
    (d) Any of these.
49. The LCB is
    (a) An upper limit to LCL
    (b) A lower limit to LCL
    (c) (a) and (b)
    (d) (a) or (b).

50. The UCB is
    (a) An upper limit to UCL (b) A lower limit to LCL
    (c) Both (a) and (b) (d) (a) or (b).

51. length of a class is
    (a) The difference between the UCB and LCB of that class
    (b) The difference between the UCL and LCL of that class
    (c) (a) or (b)
    (d) Both (a) and (b).

52. For a particular class boundary, the less than cumulative frequency and more than cumulative frequency add up to
    (a) Total frequency (b) Fifty per cent of the total frequency
    (c) (a) or (b) (d) None of these.

53. Frequency density corresponding to a class interval is the ratio of
    (a) Class frequency to the total frequency (b) Class frequency to the class length
    (c) Class length to the class frequency (d) Class frequency to the cumulative frequency.

54. Relative frequency for a particular class
    (a) Lies between 0 and 1 (b) Lies between 0 and 1, both inclusive
    (c) Lies between –1 and 0 (d) Lies between –1 to 1.

55. Mode of a distribution can be obtained from
    (a) Histogram (b) Less than type ogives
    (c) More than type ogives (d) Frequency polygon.

56. Median of a distribution can be obtained from
    (a) Frequency polygon (b) Histogram
    (c) Less than type ogives (d) None of these.

57. A comparison among the class frequencies is possible only in
    (a) Frequency polygon (b) Histogram
    (c) Ogives (d) (a) or (b).

58. Frequency curve is a limiting form of
    (a) Frequency polygon (b) Histogram
    (c) (a) or (b) (d) (a) and (b).

59. Most of the commonly used frequency curves are

(a) Mixed (b) Inverted J-shaped

(c) U-shaped (d) Bell-shaped.

60. The distribution of profits of a company follows

(a) J-shaped frequency curve (b) U-shaped frequency curve

(c) Bell-shaped frequency curve (d) Any of these.

Set B

Answer the following questions. Each question carries 2 marks.

1. Out of 1000 persons, 25 per cent were industrial workers and the rest were agricultural workers. 300 persons enjoyed world cup matches on TV. 30 per cent of the people who had not watched world cup matches were industrial workers. What is the number of agricultural workers who had enjoyed world cup matches on TV?

(a) 260 (b) 240 (c) 230 (d) 250

2. A sample study of the people of an area revealed that total number of women were 40% and the percentage of coffee drinkers were 45 as a whole and the percentage of male coffee drinkers was 20. What was the percentage of female non-coffee drinkers?

(a) 10 (b) 15 (c) 18 (d) 20

3. Cost of sugar in a month under the heads raw materials, labour, direct production and others were 12, 20, 35 and 23 units respectively. What is the difference between the central angles for the largest and smallest components of the cost of sugar?

(a) 72^o (b) 48^o (c) 56^o (d) 92^o

4. The number of accidents for seven days in a locality are given below :

No. of accidents :	0	1	2	3	4	5	6
Frequency :	15	19	22	31	9	3	2

What is the number of cases when 3 or less accidents occurred?

(a) 56 (b) 6 (c) 68 (d) 87

5. The following data relate to the incomes of 86 persons :

Income in Rs. :	500–999	1000–1499	1500–1999	2000–2499
No. of persons :	15	28	36	7

What is the percentage of persons earning more than Rs. 1500?

(a) 50 (b) 45 (c) 40 (d) 60

6. The following data relate to the marks of a group of students:

Marks :	Below 10	Below 20	Below 30	Below 40	Below 50
No. of students :	15	38	65	84	100

How many students got marks more than 30?

(a) 65 (b) 50 (c) 35 (d) 43

7. Find the number of observations between 250 and 300 from the following data :

Value	More than 200	More than 250	More than 300	More than 350
No. of observations	56	38	15	0

(a) 56 (b) 23 (c) 15 (d) 8

Set C

Answer the following questions. Each question carries 5 marks.

1. In a study about the male and female students of commerce and science departments of a college in 5 years, the following datas were obtained :

1995	2000
70% male students	75% male students
65% read Commerce	40% read Science
20% of female students read Science	50% of male students read Commerce
3000 total No. of students	3600 total No. of students.

After combining 1995 and 2000 if x denotes the ratio of female commerce student to female Science student and y denotes the ratio of male commerce student to male Science student, then

(a) $x = y$ (b) $x > y$ (c) $x < y$ (d) $x \geq y$

2. In a study relating to the labourers of a jute mill in West Bengal, the following information was collected.

'Twenty per cent of the total employees were females and forty per cent of them were married. Thirty female workers were not members of Trade Union. Compared to this, out of 600 male workers 500 were members of Trade Union and fifty per cent of the male workers were married. The unmarried non-member male employees were 60 which formed ten per cent of the total male employees. The unmarried non-members of the employees were 80'. On the basis of this information, the ratio of married male non-members to the married female non-members is

(a) 1 : 3 (b) 3 : 1 (c) 4 : 1 (d) 5 : 1

3. The weight of 50 students in pounds are given below :

82,	95,	120,	174,	179,	176,	159,	91,	85,	175
88,	160,	97,	133,	159,	176,	151,	115,	105,	172
170,	128,	112,	101,	123,	117,	93,	117,	99,	90
113,	119,	129,	134,	178,	105,	147,	107,	155,	157
98,	117,	95,	135,	175,	97,	160,	168,	144,	175

If the data are arranged in the form of a frequency distribution with class intervals as 81-100, 101-120, 121-140, 141-160 and 161-180, then the frequencies for these 5 class intervals are

(a) 6, 9, 10, 11, 14 (b) 12, 8, 7, 11, 12 (c) 10, 12, 8, 11, 9 (d) 12, 11, 6, 9, 12

4. The following data relate to the marks of 48 students in statistics :

56,	10,	54,	38,	21,	43,	12,	22
48,	51,	39,	26,	12,	17,	36,	19
48,	36,	15,	33,	30,	62,	57,	17
5,	17,	45,	46,	43,	55,	57,	38
43,	28,	32,	35,	54,	27,	17,	16
11,	43,	45,	2,	16,	46,	28,	45

What are the frequency densities for the class intervals 30-39, 40-49 and 50-59

(a) 0.20, 0.50, 0.90

(b) 0.70, 0.90, 1.10

(c) 0.1875, 0.1667, 0.2083

(d) 0.90, 1.00, 0.80

5. The following information relates to the age of death of 50 persons in an area :

36,	48,	50,	45,	49,	31,	50,	48,	42,	57
43,	40,	32,	41,	39,	39,	43,	47,	45,	52
47,	48,	53,	37,	48,	50,	41,	49,	50,	53
38,	41,	49,	45,	36,	39,	31,	48,	59,	48
37,	49,	53,	51,	54,	59,	48,	38,	39,	45

If the class intervals are 31-33, 34-36, 37-39, …. Then the percentage frequencies for the last five class intervals are

(a) 18, 18, 10, 2 and 4. (b) 10, 15, 18, 4 and 2. (c) 14, 18, 20, 10 and 2.

(d) 10, 12, 16, 4 and 6.

ANSWERS

Set A

1.	(c)	**2.**	(b)	**3.**	(d)	**4.**	(d)	**5.**	(a)	**6.**	(b)
7.	(b)	**8.**	(a)	**9.**	(a)	**10.**	(c)	**11.**	(b)	**12.**	(a)
13.	(d)	**14.**	(c)	**15.**	(a)	**16.**	(c)	**17.**	(b)	**18.**	(a)
19.	(a)	**20.**	(d)	**21.**	(c)	**22.**	(a)	**23.**	(b)	**24.**	(c)

25.	(d)	**26.**	(d)	**27.**	(c)	**28.**	(a)	**29.**	(a)	**30.**	(b)
31.	(c)	**32.**	(d)	**33.**	(b)	**34.**	(b)	**35.**	(b)	**36.**	(d)
37.	(d)	**38.**	(b)	**39.**	(d)	**40.**	(c)	**41.**	(a)	**42.**	(d)
43.	(a)	**44.**	(a)	**45.**	(b)	**46.**	(b)	**47.**	(a)	**48.**	(b)
49.	(b)	**50.**	(a)	**51.**	(a)	**52.**	(a)	**53.**	(b)	**54.**	(a)
55.	(a)	**56.**	(c)	**57.**	(b)	**58.**	(d)	**59.**	(d)	**60.**	(c)

Set B

1.	(a)	**2.**	(b)	**3.**	(d)	**4.**	(d)	**5.**	(a)	**6.**	(c)
7.	(b)										

Set C

1.	(b)	**2.**	(c)	**3.**	(d)	**4.**	(d)	**5.**	(a)

ADDITIONAL QUESTION BANK

1. Graph is a

   (a) Line diagram (b) Bar diagram (c) Pie diagram (d) Pictogram

2. Details are shown by

   (a) Charts (b) Tabular presentation

   (c) both (d) none

3. The relationship between two variables are shown in

   (a) Pictogram (b) Histogram (c) Bar diagram (d) Line diagram

4. In general the number of types of tabulation are

   (a) two (b) three (c) one (d) four

5. A table has

   (a) four (b) two (c) five (d) none parts.

6. The number of errors in Statistics are

   (a) one (b) two (c) three (d) four

7. The number of "Frequency distribution" is

   (a) two (b) one (c) five (d) four

8. (Class frequency)/(Width of the class) is defined as

   (a) Frequency density (b) Frequency distribution

   (c) both (d) none

9. Tally marks determines

(a) class width (b) class boundary (c) class limit (d) class frequency

10. Cumulative Frequency Distribution is a

(a) graph (b) frequency (c) Statistical Table (d) distribution

11. To find the number of observations less than any given value

(a) Single frequency distribution (b) Grouped frequency distribution

(c) Cumulative frequency distribution (d) None is used.

12. An area diagram is

(a) Histogram (b) Frequency Polygon

(c) Ogive (d) none

13. When all classes have a common width

(a) Pie Chart (b) Frequency Polygon

(c) both (d) none is used.

14. An approximate idea of the shape of frequency curve is given by

(a) Ogive (b) Frequency Polygon

(c) both (d) none

15. Ogive is a

(a) Line diagram (b) Bar diagram (c) both (d) none

16. Unequal widths of classes in the frequency distribution do not cause any difficulty in the construction of

(a) Ogive (b) Frequency Polygon

(c) Histogram (d) none

17. The graphical representation of a cumulative frequency distribution is called

(a) Histogram (b) Ogive (c) both (d) none.

18. The most common form of diagrammatic representation of a grouped frequency distribution is

(a) Ogive (b) Histogram (c) Frequency Polygon (d) none

19. Vertical bar chart may appear somewhat alike

(a) Histogram (b) Frequency Polygon

(c) both (d) none

20. The number of types of cumulative frequency is

(a) one (b) two (c) three (d) four

21. A representative value of the class interval for the calculation of mean, standard deviation, mean deviation etc. is

    (a) class interval (b) class limit (c) class mark (d) none

22. The number of observations falling within a class is called

    (a) density (b) frequency (c) both (d) none

23. Classes with zero frequencies are called

    (a) nil class (b) empty class (c) class (d) none

24. For determining the class frequencies it is necessary that these classes are

    (a) mutually exclusive (b) not mutually exclusive

    (c) independent (d) none

25. Most extreme values which would ever be included in a class interval are called

    (a) class limits (b) class interval (c) class boundaries (d) none

26. The value exactly at the middle of a class interval is called

    (a) class mark (b) mid value (c) both (d) none

27. Difference between the lower and the upper class boundaries is

    (a) width (b) size (c) both (d) none

28. In the construction of a frequency distribution, it is generally preferable to have classes of

    (a) equal width (b) unequal width (c) maximum (d) none

29. Frequency density is used in the construction of

    (a) Histogram (b) Ogive

    (c) Frequency Polygon (d) none when the classes are of unequal width.

30. "Cumulative Frequency" only refers to the

    (a) less-than type (b) more-than type (c) both (d) none

31. For the construction of a grouped frequency distribution

    (a) class boundaries (b) class limits (c) both (d) none are used.

32. In all Statistical calculations and diagrams involving end points of classes

    (a) class boundaries (b) class value (c) both (d) none are used.

33. Upper limit of any class is ________ from the lower limit of the next class

    (a) same (b) different

    (c) both (d) none

34. Upper boundary of any class coincides with the Lower boundary of the next class.

    (a) true (b) false (c) both (d) none.

35. Excepting the first and the last, all other class boundaries lie midway between the upper limit of a class and the lower limit of the next higher class.

(a) true (b) false (c) both (d) none

36. The lower extreme point of a class is called

(a) lower class limit (b) lower class boundary

(c) both (d) none

37. For the construction of grouped frequency distribution from ungrouped data we use

(a) class limits (b) class boundaries (c) class width (d) none

38. When one end of a class is not specified, the class is called

(a) closed- end class (b) open- end class (c) both (d) none

39. Class boundaries should be considered to be the real limits for the class interval.

(a) true (b) false (c) both (d) none

40. Difference between the maximum & minimum value of a given data is called

(a) width (b) size (c) range (d) none

41. In Histogram if the classes are of unequal width then the heights of the rectangles must be proportional to the frequency densities.

(a) true (b) false (c) both (d) none

42. When all classes have equal width, the heights of the rectangles in Histogram will be numerically equal to the

(a) class frequencies (b) class boundaries (c) both (d) none

43. Consecutive rectangles in a Histogram have no space in between

(a) true (b) false (c) both (d) none

44. Histogram emphasizes the widths of rectangles between the class boundaries.

(a) false (b) true (c) both (d) none

45. To find the mode graphically

(a) Ogive (b) Frequency Polygon

(c) Histogram (d) none may be used.

46. When the width of all classes is same, frequency polygon has not the same area as the Histogram.

(a) True (b) false (c) both (d) none

47. For obtaining frequency polygon we join the successive points whose abscissa represent the corresponding class frequency_____

(a) true (b) false (c) both (d) none

48. In representing simple frequency distributions of a discrete variable

(a) Ogive (b) Histogram (c) Frequency Polygon (d) both is useful.

49. Diagrammatic representation of the cumulative frequency distribution is

(a) Frequency Polygon (b) Ogive (c) Histogram (d) none

50. For the overlapping classes 0–10 , 10–20 , 20–30 etc.the class mark of the class 0–10 is

(a) 5 (b) 0 (c) 10 (d) none

51. For the non-overlapping classes 0–19 , 20–39 , 40–59 the class mark of the class 0–19 is

(a) 0 (b) 19 (c) 9.5 (d) none

52.

Class :	0–10	10–20	20–30	30–40	40–50
Frequency :	5	8	15	6	4

For the class 20–30 , cumulative frequency is

(a) 20 (b) 13 (c) 15 (d) 28

53. An Ogive can be prepared in ____________ different ways.

(a) 2 (b) 3 (c) 4 (d) none

54. The curve obtained by joining the points, whose x- coordinates are the upper limits of the class-intervals and y coordinates are corresponding cumulative frequencies is called

(a) Ogive (b) Histogram (c) Frequency Polygon (d) Frequency Curve

55. The breadth of the rectangle is equal to the length of the class-interval in

(a) Ogive (b) Histogram (c) both (d) none

56. In Histogram, the classes are taken

(a) overlapping (b) non-overlapping (c) both (d) none

57. For overlapping class-intervals the class limit & class boundary are

(a) same (b) not same (c) zero (d) none

58. Classification is of

(a) four (b) Three (c) two (d) five kinds.

ANSWERS

1. (a)	**2.** (b)	**3.** (d)	**4.** (a)	**5.** (c)
6. (b)	**7.** (a)	**8.** (a)	**9.** (d)	**10.** (c)
11. (c)	**12.** (a)	**13.** (b)	**14.** (b)	**15.** (a)
16. (c)	**17.** (b)	**18.** (b)	**19.** (a)	**20.** (b)
21. (c)	**22.** (b)	**23.** (b)	**24.** (a)	**25.** (c)

26. (c)	**27.** (c)	**28.** (a)	**29.** (a)	**30.** (a)
31. (b)	**32.** (a)	**33.** (b)	**34.** (a)	**35.** (a)
36. (b)	**37.** (a)	**38.** (b)	**39.** (a)	**40.** (c)
41. (a)	**42.** (a)	**43.** (a)	**44.** (b)	**45.** (c)
46. (b)	**47.** (b)	**48.** (c)	**49.** (b)	**50.** (a)
51. (c)	**52.** (d)	**53.** (a)	**54.** (a)	**55.** (b)
56. (a)	**57.** (a)	**58.** (a)		

CHAPTER 15

MEASURES OF CENTRAL TENDENCY AND DISPERSION

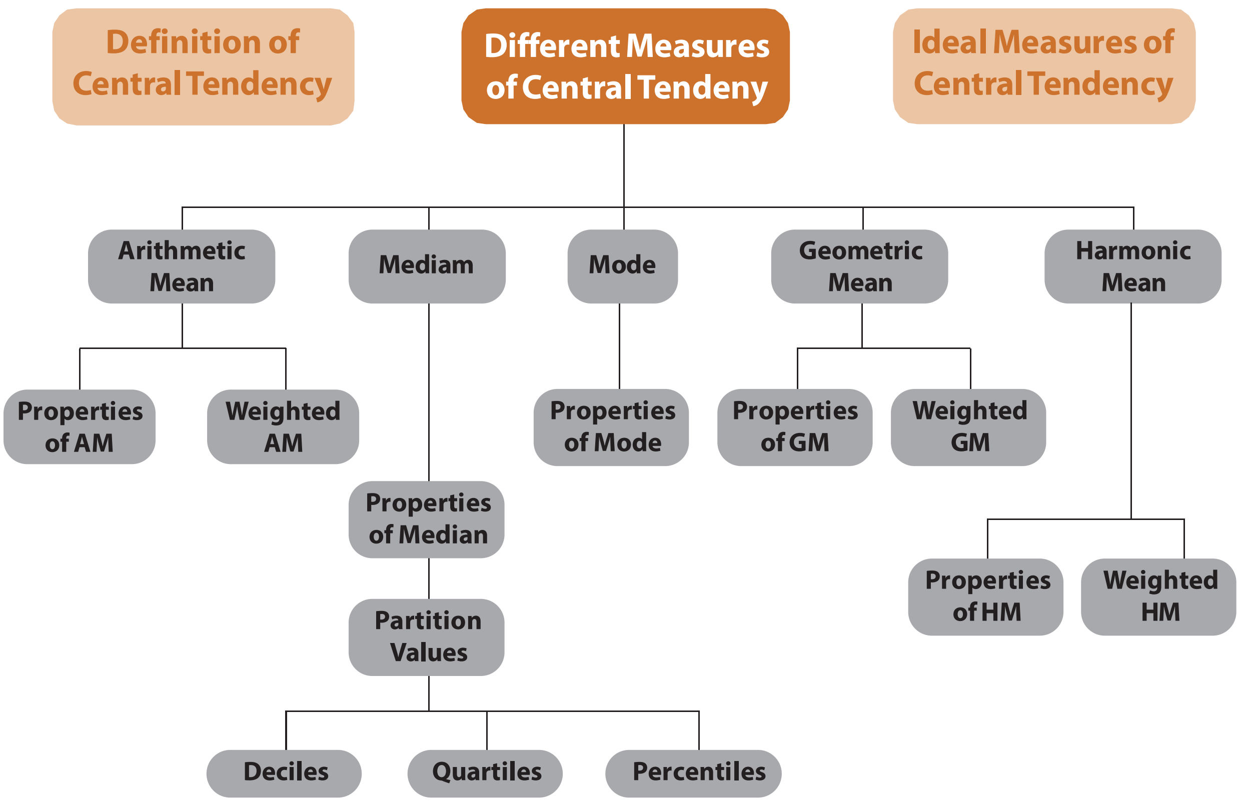
UNIT I: MEASURES OF CENTRAL TENDENCY

LEARNING OBJECTIVES

After reading this chapter, students will be able to understand:

- To understand different measures of central tendency, i.e. Arithmetic Mean, Median, Mode, Geometric Mean and Harmonic Mean, and computational techniques of these measures.
- To learn comparative advantages and disadvantages of these measures and therefore, which measures to use in which circumstance.

UNIT OVERVIEW



15.1.1 DEFINITION OF CENTRAL TENDENCY

In many a case, like the distributions of height, weight, marks, profit, wage and so on, it has been noted that starting with rather low frequency, the class frequency gradually increases till it reaches its maximum somewhere near the central part of the distribution and after which the class frequency steadily falls to its minimum value towards the end. Thus, central tendency may be defined as the tendency of a given set of observations to cluster around a single central or middle value and the single value that represents the given set of observations is described as a measure of central tendency or, location, or average. Hence, it is possible to condense a vast mass of data by a single representative value. The computation of a measure of central tendency plays a very important part in many a sphere. A company is recognized by its high average profit, an educational institution is judged on the basis of average marks obtained by its students and so on. Furthermore, the central tendency also facilitates us in providing a basis for comparison between different distribution. Following are the different measures of central tendency:

(i) Arithmetic Mean (AM)

(ii) Median (Me)

(iii) Mode (Mo)

(iv) Geometric Mean (GM)

(v) Harmonic Mean (HM)

15.1.2 CRITERIA FOR AN IDEAL MEASURE OF CENTRAL TENDENCY

Following are the criteria for an ideal measure of central tendency:

(i) It should be properly and unambiguously defined.

(ii) It should be easy to comprehend.

(iii) It should be simple to compute.

(iv) It should be based on all the observations.

(v) It should have certain desirable mathematical properties.

(vi) It should be least affected by the presence of extreme observations.

15.1.3 ARITHMETIC MEAN

For a given set of observations, the AM may be defined as the sum of all the observations divided by the number of observations. Thus, if a variable x assumes n values $x_1, x_2, x_3, \ldots\ldots\ldots x_n$, then the AM of x, to be denoted by $\overline{x}$, is given by,

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots\ldots\ldots\ldots + x_n}{n}$$

$$= \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{\sum x_i}{n} \qquad \text{..........................(15.1.1)}$$

In case of a simple frequency distribution relating to an attribute, we have

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots\ldots\ldots\ldots + f_nx_n}{f_1 + f_2 + f_3 + \ldots\ldots\ldots\ldots + f_n}$$

$$= \frac{\sum f_ix_i}{\sum f_i}$$

$$\overline{X} = \frac{\sum f_ix_i}{N} \qquad \text{..........................(15.1.2)}$$

assuming the observation x_i occurs f_i times, i=1,2,3,……..n and $N=\leq f_i$.
In case of grouped frequency distribution also we may use formula (15.1.2) with x_i as the mid value of the i-th class interval, on the assumption that all the values belonging to the i-th class interval are equal to x_i.
However, in most cases, if the classification is uniform, we consider the following formula for the computation of AM from grouped frequency distribution:

$$\overline{x} = A + \frac{\sum f_id_i}{N} \times C \qquad \text{................................(15.1.3)}$$

Where, $d_i = \dfrac{x_i - A}{C}$

A = Assumed Mean

C = Class Length

ILLUSTRATIONS:

Example 15.1.1: Following are the daily wages in Rupees of a sample of 9 workers: 58, 62, 48, 53, 70, 52, 60, 84, 75. Compute the mean wage.

Solution: Let x denote the daily wage in rupees.

Then as given, x_1=58, x_2=62, x_3= 48, x_4=53, x_5=70, x_6=52, x_7=60, x_8=84 and x_9=75.

Applying (15.1.1) the mean wage is given by,

$$\overline{x} = \frac{\sum_{i=1}^{9} x_i}{9}$$

$$= ₹\ \frac{(58 + 62 + 48 + 53 + 70 + 52 + 60 + 84 + 75)}{9}$$

$$= ₹\ \frac{562}{9}$$

$$= ₹\ 62.44.$$

Example 15.1.2: Compute the mean weight of a group of BBA students of St. Xavier's College from the following data:

Weight in kgs.	44 – 48	49 – 53	54 – 58	59 – 63	64 – 68	69 – 73
No. of Students	3	4	5	7	9	8

Solution: Computation of mean weight of 36 BBA students

Table 15.1.1

Weight in kgs. (1)	No. of Student (f_i) (2)	Mid-Value (x_i) (3)	$f_i x_i$ (4) = (2) x (3)
44 – 48	3	46	138
49 – 53	4	51	204
54 – 58	5	56	280
59 – 63	7	61	427
64 – 68	9	66	594
69 – 73	8	71	568
Total	36	–	2211

Applying (15.1.2), we get the average weight as

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$= \frac{2211}{36} \text{ kgs.}$$

$$= 61.42 \text{ kgs.}$$

Example 15.1.3: Find the AM for the following distribution:

Class Interval	350 – 369	370 – 389	390 – 409	410 – 429	430 – 449	450 – 469	470 – 489
Frequency	23	38	58	82	65	31	11

Solution: We apply formula (11.3) since the amount of computation involved in finding the AM is much more compared to **Example 15.1.2**. Any mid value can be taken as A. However, usually A is taken as the middle most mid-value for an odd number of class intervals and any one of the two middle most mid-values for an even number of class intervals. The class length is taken as C.

Table 15.1.2 Computation of AM

Class Interval (1)	Frequency(f_i) (2)	Mid-Value(x_i) (3)	$d_i = \frac{x_i - A}{c} = \frac{x_i - 419.50}{20}$ (4)	$f_i d_i$ (5) = (2)X(4)
350 – 369	23	359.50	– 3	– 69
370 – 389	38	379.50	– 2	– 76
390 – 409	58	399.50	– 1	– 58
410 – 429	82	419.50 (A)	0	0
430 – 449	65	439.50	1	65
450 – 469	31	459.50	2	62
470 – 489	11	479.50	3	33
Total	308	–	–	– 43

The required AM is given by

$$\bar{x} = A + \frac{\sum f_i d_i}{N} \times C$$

$$= 419.50 + \frac{(-43)}{308} \times 20$$

$$= 419.50 - 2.79$$

$$= 416.71$$

Example 15.1.4: Given that the mean height of a group of students is 67.45 inches. Find the missing frequencies for the following incomplete distribution of height of 100 students.

Height in inches	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74
No. of Students	5	18	–	–	8

Solution: Let x denote the height and f_3 and f_4 as the two missing frequencies.

Table 15.1.3

Estimation of missing frequencies

CI (1)	Frequency (f_i) (2)	Mid - Value (x_i) (3)	$d_i = \frac{x_i - A}{c}$ $\frac{x_i - 67}{3}$ (4)	$f_i d_i$ (5) = (2) x (4)
60-62	5	61	-2	-10
63 – 65	18	64	– 1	– 18
66 – 68	f_3	67 (A)	0	0
69 – 71	f_4	70	1	f_4
72 – 74	8	73	2	16
Total	$31 + f_3 + f_4$	–	–	$-12 + f_4$

As given, we have

$$31 + f_3 + f_4 = 100$$

$\Rightarrow \quad f_3 + f_4 = 69 \quad$ ....(1)

and $\quad \bar{x} = 67.45$

$$\Rightarrow \quad A + \frac{\sum f_i d_i}{N} \times C = 67.45$$

$$\Rightarrow \quad 67 + \frac{(-12 + f_4)}{100} \times 3 = 67.45$$

$$\Rightarrow \quad (-12 + f_4) \times 3 = (67.45 - 67) \times 100$$

$$\Rightarrow \quad -12 + f_4 = 15$$

$$\Rightarrow \quad f_4 = 27$$

On substituting 27 for f_4 in (1), we get

$f_3 + 27 = 69, \quad \Rightarrow f_3 = 42$

Thus, the missing frequencies would be 42 and 27.

Properties of AM

(i) **If all the observations assumed by a variable are constants, say k, then the AM is also k.** For example, if the height of every student in a group of 10 students is 170 cm, then the mean height is, of course, 170 cm.

(ii) the algebraic sum of deviations of a set of observations from their AM is zero

i.e. for unclassified data , $\sum(x_i - \bar{x}) = 0$

and for grouped frequency distribution, $\sum f_i(x_i - \bar{x}) = 0$ ....(15.1.4)

For example, if a variable x assumes five observations, say 58, 63, 37, 45, 29, then $\bar{x}$ =46.4. Hence, the deviations of the observations from the AM i.e. $(x_i - \bar{x})$ are 11.60, 16.60, –9.40, –1.40 and –17.40, then $\sum(x_i - \bar{x}) = 11.60 + 16.60 + (-9.40) + (-1.40) + (-17.40) = 0$.

(iii) AM is affected due to a change of origin and/or scale which implies that if the original variable x is changed to another variable y by effecting a change of origin, say a, and scale say b, of x i.e. y=a+bx, then the AM of y is given by $\bar{y} = a + b\bar{x}$.

For example, if it is known that two variables x and y are related by 2x+3y+7=0 and $\bar{x} = 15$, then the AM of y is given by $\bar{y} = \frac{-7 - 2\bar{x}}{3}$

$= \frac{-7 - 2 \times 15}{3} = \frac{-37}{3} = -12.33$.

(iv) If there are two groups containing n_1 and n_2 observations and $\bar{x}_1$ and $\bar{x}_2$ as the respective arithmetic means, then the combined AM is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \quad \text{..................................(15.1.5)}$$

This property could be extended to k>2 groups and we may write

$$\bar{x} = \frac{\sum n_i\bar{x}_i}{\sum n_i} \quad \text{..................................(15.1.6)} \quad i = 1, 2, \text{........}n.$$

Example 15.1.5: The mean salary for a group of 40 female workers is ₹ 5,200 per month and that for a group of 60 male workers is ₹ 6800 per month. What is the combined mean salary?

Solution: As given n_1 = 40, n_2 = 60, $\bar{x}_1$ = ₹ 5,200 and $\bar{x}_2$ = ₹ 6,800 hence, the combined mean salary per month is

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

$$= \frac{40 \times ₹\,5,200 + 60 \times ₹\,6,800}{40 + 60} = ₹\,6,160.$$

15.1.4 MEDIAN – PARTITION VALUES

As compared to AM, median is a positional average which means that the value of the median is dependent upon the position of the given set of observations for which the median is wanted. Median, for a given set of observations, may be defined as the middle-most value when the observations are arranged either in an ascending order or a descending order of magnitude.

As for example, if the marks of the 7 students are 72, 85, 56, 80, 65, 52 and 68, then in order to find the median mark, we arrange these observations in the following ascending order of magnitude: 52, 56, 65, 68, 72, 80, 85.

Since the 4^{th} term i.e. 68 in this new arrangement is the middle most value, the median mark is 68 i.e. Median (Me) = 68.

As a second example, if the wages of 8 workers, expressed in rupees are

56, 82, 96, 120, 110, 82, 106, 100 then arranging the wages as before, in an ascending order of magnitude, we get ₹56, ₹ 82, ₹ 82, ₹ 96, ₹ 100, ₹ 106, ₹ 110, ₹ 120. Since there are two middle-most values, namely, ₹ 96, and ₹ 100 any value between ₹ 96 and ₹ 100 may be, theoretically, regarded as median wage. However, to bring uniqueness, we take the arithmetic mean of the two middle-most values, whenever the number of the observations is an even number. Thus, the median wage in this example, would be

$$M = \frac{₹\,96 + ₹\,100}{2} = ₹\,98$$

In case of a grouped frequency distribution, we find median from the cumulative frequency distribution of the variable under consideration. We may consider the following formula, which can be derived from the basic definition of median.

$$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l} \right) \times C \quad \ldots\ldots (15.1.7)$$

Where,

l_1 = lower class boundary of the median class i.e. the class containing median.

N = total frequency.

N_l = less than cumulative frequency corresponding to l_1. (Pre median class)

N_u = less than cumulative frequency corresponding to l_2. (Post median class)

l_2 being the upper class boundary of the median class.

$C = l_2 - l_1$ = length of the median class.

Example 15.1.6: Compute the median for the distribution as given in **Example 15.1.3.**

Solution: First, we find the cumulative frequency distribution which is exhibited in **Table 15.1.4.**

Table 15.1.4

Computation of Median

Class boundary	Less than cumulative frequency
349.50	0
369.50	23
389.50	61
409.50 (l_1)	119 (N_l)
429.50 (l_2)	201(N_u)
449.50	266
469.50	297
489.50	308

We find, from the **Table 15.1.4**, $\frac{N}{2} = \frac{308}{2}$ = 154 lies between the two cumulative frequencies 119 and 201 i.e. 119 < 154 < 201 . Thus, we have N_l = 119, N_u = 201 l_1 = 409.50 and l_2 = 429.50. Hence C = 429.50 – 409.50 =20.
Substituting these values in (15.1.7), we get,

$$M = 409.50 + \frac{154-119}{201-119} \times 20$$
$$= 409.50+8.54$$
$$= 418.04.$$

Example 15.1.7: Find the missing frequency from the following data, given that the median mark is 23.

Mark :	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students :	5	8	?	6	3

Solution: Let us denote the missing frequency by f_3. Table 15.1.5 shows the relevant computation.

Table 15.1.5
(Estimation of missing frequency)

Mark	Less than cumulative frequency
0	0
10	5
20(l_1)	13(N_l)
30(l_2)	13+f_3(N_u)
40	19+f_3
50	22+f_3

Going through the mark column, we find that 20<23<30. Hence l_1=20, l_2=30 and accordingly N_l=13, N_u=13+f_3. Also the total frequency i.e. N is 22+f_3. Thus,

$$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l}\right) \times C$$

$$\Rightarrow \quad 23 = 20 + \frac{\left(\frac{22+f_3}{2}\right) - 13}{(13+f_3) - 13} \times 10$$

$$\Rightarrow \quad 3 = \frac{22+f_3-26}{f_3} \times 5$$

$$\Rightarrow \quad 3f_3 = 5f_3 - 20$$

$$\Rightarrow \quad 2f_3 = 20$$

$$\Rightarrow \quad f_3 = 10$$

So, the missing frequency is 10.

Properties of median

We cannot treat median mathematically, the way we can do with arithmetic mean. We consider below two important features of median.

(i) If x and y are two variables, to be related by y=a+bx for any two constants a and b, then the median of y is given by
$y_{me} = a + bx_{me}$

For example, if the relationship between x and y is given by 2x – 5y = 10 and if x_{me} i.e. the median of x is known to be 16.
Then 2x – 5y = 10

$\Rightarrow$ $y = -2 + 0.40x$

$\Rightarrow$ $y_{me} = -2 + 0.40\, x_{me}$

$\Rightarrow$ $y_{me} = -2 + 0.40\times 16$

$\Rightarrow$ $y_{me} = 4.40.$

(ii) For a set of observations, the sum of absolute deviations is minimum when the deviations are taken from the median. This property states that $\Sigma|x_i - A|$ is minimum if we choose A as the median.

PARTITION VALUES OR QUARTILES OR FRACTILES

These may be defined as values dividing a given set of observations into a number of equal parts. When we want to divide the given set of observations into two equal parts, we consider median. Similarly, quartiles are values dividing a given set of observations into four equal parts. So there are three quartiles – first quartile or lower quartile denoted by Q_1, second quartile or median to be denoted by Q_2 or Me and third quartile or upper quartile denoted by Q_3. First quartile is the value for which one fourth of the observations are less than or equal to Q_1 and the remaining three – fourths observations are more than or equal to Q_1. In a similar manner, we may define Q_2 and Q_3.

Deciles are the values dividing a given set of observation into ten equal parts. Thus, there are nine deciles to be denoted by D_1, D_2, D_3,.....D_9. D_1 is the value for which one-tenth of the given observations are less than or equal to D_1 and the remaining nine-tenth observations are greater than or equal to D_1 when the observations are arranged in an ascending order of magnitude.

Lastly, we talk about the percentiles or centiles that divide a given set of observations into 100 equal parts. The points of sub-divisions being P_1, P_2,..........P_{99}. P_1 is the value for which one hundredth of the observations are less than or equal to P_1 and the remaining ninety-nine hundredths observations are greater than or equal to P_1 once the observations are arranged in an ascending order of magnitude.

For unclassified data, the p^{th} quartile is given by the $(n+1)p^{th}$ value, where n denotes the total number of observations. p = 1/4, 2/4, 3/4 for Q_1, Q_2 and Q_3 respectively. p=1/10, 2/10,............9/10. For D_1, D_2,......,D_9 respectively and lastly p=1/100, 2/100,....,99/100 for P_1, P_2, P_3....P_{99} respectively.

In case of a grouped frequency distribution, we consider the following formula for the computation of quartiles.

$$Q = l_1 + \left(\frac{Np - N_l}{N_u - N_l}\right) \times C \quad \ldots\ldots\ldots\ldots \text{(15.1.8)}$$

The symbols, except p, have their usual interpretation which we have already discussed while computing median and just like the unclassified data, we assign different values to p depending on the quartile.

Another way to find quartiles for a grouped frequency distribution is to draw the ogive (less than type) for the given distribution. In order to find a particular quartile, we draw a line parallel to the horizontal axis through the point Np. We draw perpendicular from the point of intersection of this parallel line and the ogive. The x-value of this perpendicular line gives us the value of the quartile under discussion.

Example 15.1.8: Following are the wages of the labourers: ₹ 82, ₹ 56, ₹ 90, ₹ 50, ₹ 120, ₹ 75, ₹ 75, ₹ 80, ₹ 130, ₹ 65. Find Q_1, D_6 and P_{82}.

Solution: Arranging the wages in an ascending order, we get ₹ 50, ₹ 56, ₹ 65, ₹ 75, ₹ 75, ₹ 80, ₹ 82, ₹ 90, ₹ 120, ₹ 130.

Hence, we have

$$Q_1 = \frac{(n+1)}{4}\text{th value}$$

$$= \frac{(10+1)}{4}\text{th value}$$

$$= 2.75^{th} \text{ value}$$

$$= 2^{nd} \text{ value} + 0.75 \times \text{difference between the third and the } 2^{nd} \text{ values.}$$

$$= ₹\ [56 + 0.75 \times (65 - 56)]$$

$$= ₹\ 62.75$$

$$D_6 = (15 + 1) \times \frac{6}{10} \text{ th value}$$

$$= 6.60^{th} \text{ value}$$

$$= 6^{th} \text{ value} + 0.60 \times \text{difference between the } 7^{th} \text{ and the } 6^{th} \text{ values.}$$

$$= ₹\ (80 + 0.60 \times 2)$$

$$= ₹\ 81.20$$

$$P_{82} = (10+1)\times\frac{82}{100} \text{ th value}$$

$$= 9.02^{th} \text{ value}$$

$$= 9^{th} \text{ value} + 0.02 \times \text{difference between the } 10^{th} \text{ and the } 9^{th} \text{ values}$$

$$= ₹\ (120 + 0.02 \times 10)$$

$$= ₹\ 120.20$$

Next, let us consider one problem relating to the grouped frequency distribution.

Example 15.1.9: Following distribution relates to the distribution of monthly wages of 100 workers.

Wages in (₹)	: less than 500	500–699	700–899	900–1099	1100–1499	more than 1500
No. of workers :	5	23	29	27	10	6

Compute Q_3 , D_7 and P_{23} .

Solution: This is a typical example of an open end unequal classification as we find the lower class limit of the first class interval and the upper class limit of the last class interval are not stated, and theoretically, they can assume any value between 0 and 500 and 1500 to any number respectively. The ideal measure of the central tendency in such a situation is median as the median or second quartile is based on the fifty percent central values. Denoting the first LCB and the last UCB by the L and U respectively, we construct the following cumulative frequency distribution:

Table 15.1.7
Computation of quartiles

Wages in rupees (CB)	No. of workers (less than cumulative frequency)
L	0
499.50	5
699.50	28
899.50	57
1099.50	84
1499.50	94
U	100

For Q_3, $\frac{3N}{4} = \frac{3 \times 100}{4} = 75$

since, 57<75 <84, we take $N_l = 57$, $N_u=84$, $l_1=899.50$, $l_2=1099.50$, $c = l_2 - l_1 = 200$ in the formula (11.8) for computing Q_3.

Therefore, $Q_3 = ₹ \left[899.50 + \frac{75-57}{84-57} \times 200\right] = ₹\ 1032.83$

Similarly, for D_7, $\frac{7N}{10} = \frac{7 \times 100}{10} = 70$ which also lies between 57 and 84.

Thus, $D_7 = ₹ \left[899.50 + \frac{70-57}{84-57} \times 200\right] = ₹\, 995.80$

Lastly for P_{23}, $\frac{23N}{100} = \frac{23}{100} \times 100 = 23$ and as $5 < 23 < 28$, we have

$P_{23} = ₹\ [499.50 + \frac{23-5}{28-5} \times 200]$

$= ₹\ 656.02$

15.1.5 MODE

For a given set of observations, mode may be defined as the value that occurs the maximum number of times. Thus, mode is that value which has the maximum concentration of the observations around it. This can also be described as the most common value with which, even, a layman may be familiar with.

Thus, if the observations are 5, 3, 8, 9, 5 and 6, then Mode (Mo) = 5 as it occurs twice and all the other observations occur just once. The definition for mode also leaves scope for more than one mode. Thus sometimes we may come across a distribution having more than one mode. Such a distribution is known as a multi-modal distribution. Bi-modal distribution is one having two modes.

Furthermore, it also appears from the definition that mode is not always defined. As an example, if the marks of 5 students are 50, 60, 35, 40, 56, there is no modal mark as all the observations occur once i.e. the same number of times.

We may consider the following formula for computing mode from a grouped frequency distribution:

$$\text{Mode} = l_1 + \left(\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1}\right) \times c \qquad \text{..........................(15.1.9)}$$

where,

l_1 = LCB of the modal class.
i.e. the class containing mode.
f_0 = frequency of the modal class
f_{-1} = frequency of the pre-modal class
f_1 = frequency of the post modal class
C = class length of the modal class

Example 15.1.10: Compute mode for the distribution as described in Example. 15.1.3

Solution: The frequency distribution is shown below:

Table 15.1.8
Computation of mode

Class Interval	Frequency
350 - 369	23
370 - 389	38
390 - 409	58 (f_{-1})
410 - 429	82 (f_0)
430 - 449	65 (f_1)
450 - 469	31
470 - 489	11

Going through the frequency column, we note that the highest frequency i.e. f_0 is 82. Hence, f_{-1} = 58 and f_1 = 65. Also the modal class i.e. the class against the highest frequency is 410 – 429.

Thus l_1 = LCB=409.50 and c=429.50 – 409.50 = 20

Hence, applying formulas (11.9), we get

$$Mo = 409.5 + \frac{82 - 58}{2 \times 82 - 58 - 65} \times 20$$

= 421.21 which belongs to the modal class. (410 – 429)

When it is difficult to compute mode from a grouped frequency distribution, we may consider the following empirical relationship between mean, median and mode:

Mean – Mode = 3(Mean – Median) ……………………..(15.1.9A)

or Mode = 3 Median – 2 Mean

(11.9A) holds for a moderately skewed distribution. We also note that if y = a+bx, then $y_{mo} = a + bx_{mo}$ ……………………………………..(15.1.10)

Example 15.11: For a moderately skewed distribution of marks in statistics for a group of 200 students, the mean mark and median mark were found to be 55.60 and 52.40. What is the modal mark?

Solution: Since in this case, mean = 55.60 and median = 52.40, applying (11.9A), we get the modal mark as

$$\begin{aligned} \text{Mode} &= 3 \times \text{Median} - 2 \times \text{Mean} \\ &= 3 \times 52.40 - 2 \times 55.60 \\ &= 46. \end{aligned}$$

Example 15.1.12: If y = 2 + 1.50x and mode of x is 15, what is the mode of y?

Solution:

By virtue of (11.10), we have

$$\begin{aligned} y_{mo} &= 2 + 1.50 \times 15 \\ &= 24.50. \end{aligned}$$

15.1.6 GEOMETRIC MEAN AND HARMONIC MEAN

For a given set of n positive observations, the geometric mean is defined as the n-th root of the product of the observations. Thus if a variable x assumes n values $x_1, x_2, x_3, \ldots, x_n$, all the values being positive, then the GM of x is given by

$G = (x_1 \times x_2 \times x_3 \ldots \times x_n)^{1/n}$ ……………………………………. (15.1.11)

For a grouped frequency distribution, the GM is given by

$G = (x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \ldots \times x_n^{f_n})^{1/N}$ ……………………………………. (15.1.12)

Where $N = \sum f_i$

In connection with GM, we may note the following properties :

(i) Logarithm of G for a set of observations is the AM of the logarithm of the observations; i.e.

$$\log G = \frac{1}{r}\sum \log x \qquad \text{..................(15.1.13)}$$

(ii) if all the observations assumed by a variable are constants, say K > 0, then the GM of the observations is also K.

(iii) GM of the product of two variables is the product of their GM's i.e. if $z = xy$, then

$$\text{GM of } z = (\text{GM of } x) \times (\text{GM of } y) \qquad \text{..................(15.1.14)}$$

(iv) GM of the ratio of two variables is the ratio of the GM's of the two variables i.e. if $z = x/y$ then

$$\text{GM of } z = \frac{\text{GM of } x}{\text{GM of } y} \qquad \text{..................(15.1.15)}$$

Example 15.1.13: Find the GM of 3, 6 and 12.

Solution: As given $x_1=3$, $x_2=6$, $x_3=12$ and n=3.

Applying (15.1.11), we have $G= (3\times6\times12)^{1/3} = (6^3)^{1/3}=6$.

Example 15.1.14: Find the GM for the following distribution:

x :	2	4	8	16
f :	2	3	3	2

Solution: According to (15.1.12), the GM is given by

$$G = (x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \times x_4^{f_4})^{1/N}$$

$$= (2^2 \times 4^3 \times 8^3 \times 16^2)^{1/10}$$

$$= (2)^{2.50}$$

$$= 4\sqrt{2}$$

$$= 5.66$$

Harmonic Mean

For a given set of non-zero observations, harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observation. So, if a variable x assumes n non-zero values x_1, x_2, x_3,..............., x_n, then the HM of x is given by

$$H = \frac{n}{\sum(1/x_i)} \qquad \text{..................(15.1.16)}$$

For a grouped frequency distribution, we have

$$H = \frac{N}{\sum\left[\frac{f_i}{x_i}\right]} \quad \text{..................(15.1.17)}$$

Properties of HM

(i) If all the observations taken by a variable are constants, say k, then the HM of the observations is also k.

(ii) If there are two groups with n_1 and n_2 observations and H_1 and H_2 as respective HM's than the combined HM is given by

$$\frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}} \quad \text{..................(15.1.18)}$$

Example 15.15: Find the HM for 4, 6 and 10.

Solution: Applying (15.1.16), we have

$$H = \frac{3}{\frac{1}{4} + \frac{1}{6} + \frac{1}{10}}$$

$$= \frac{3}{0.25 + 0.17 + 0.10}$$

$$= 5.77$$

Example 15.1.16: Find the HM for the following data:

x:	2	4	8	16
f:	2	3	3	2

Solution: Using (15.1.17), we get

$$H = \frac{10}{\frac{2}{2} + \frac{3}{4} + \frac{3}{8} + \frac{2}{16}}$$

$$= 4.44$$

Relation between AM, GM, and HM

For any set of positive observations, we have the following inequality:

$$AM \geq GM \geq HM \quad \text{............. (15.1.19)}$$

The equality sign occurs, as we have already seen, when all the observations are equal.

Example 15.1.17: compute AM, GM, and HM for the numbers 6, 8, 12, 36.

Solution: In accordance with the definition, we have

$$AM = \frac{6+8+12+36}{4} = 15.5\ 0$$

$$GM = (6 \times 8 \times 12 \times 36)^{1/4}$$
$$= (2^8 \times 3^4)^{1/4} = 12$$

$$HM = \frac{4}{\frac{1}{6}+\frac{1}{8}+\frac{1}{12}+\frac{1}{36}} = 9.93$$

The computed values of AM, GM, and HM establish (15.1.19).

Weighted average

When the observations under consideration have a hierarchical order of importance, we take recourse to computing weighted average, which could be either weighted AM or weighted GM or weighted HM.

$$\text{Weighted AM} = \frac{\sum w_i x_i}{\sum w_i} \quad \text{............. (15.1.20)}$$

$$\text{Weighted GM} = \text{Ante log}\left(\frac{\sum w_i \log x_i}{\sum w_i}\right) \quad \text{............. (15.1.21)}$$

$$\text{Weighted HM} = \frac{\sum w_i}{\sum\left(\frac{w_i}{x_i}\right)} \quad \text{............. (15.1.22)}$$

Example 15.1.18: Find the weighted AM and weighted HM of first n natural numbers, the weights being equal to the squares of the corresponding numbers.

Solution: As given,

x	1	2	3	....	n
w	1^2	2^2	3^2	....	n^2

$$\text{Weighted AM} = \frac{\sum w_i x_i}{\sum w_i}$$

$$= \frac{1 \times 1^2 + 2 \times 2^2 + 3 \times 3^2 + \ldots\ldots\ldots\ldots\ldots\ldots n \times n^2}{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots\ldots\ldots\ldots + n^2}$$

$$= \frac{1^3 + 2^3 + 3^3 + \ldots\ldots\ldots + n^3}{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots + n^2}$$

$$= \frac{\left[\frac{n(n+1)}{2}\right]^2}{\frac{n(n+1)(2n+1)}{6}}$$

$$= \frac{3n(n+1)}{2(2n+1)}$$

$$\text{Weighted HM} = \frac{\sum w_i}{\sum \left(\frac{w_i}{x_i}\right)}$$

$$= \frac{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots n^2}{\frac{1^2}{1} + \frac{2^2}{2} + \frac{3^2}{3} + \ldots\ldots \frac{n^2}{n}}$$

$$= \frac{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots + n^2}{1 + 2 + 3 + \ldots\ldots\ldots + n}$$

$$= \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}}$$

$$= \frac{2n+1}{3}$$

A General review of the different measures of central tendency

After discussing the different measures of central tendency, now we are in a position to have a review of these measures of central tendency so far as the relative merits and demerits are concerned on the basis of the requisites of an ideal measure of central tendency which we have already mentioned in section 15.1.2. The best measure of central tendency, usually, is the AM. It is rigidly defined, based on all the observations, easy to comprehend, simple to calculate and amenable to mathematical properties. However, AM has one drawback in the sense that it is very much affected by sampling fluctuations. In case of frequency distribution, mean cannot be advocated for open-end classification.

Like AM, median is also rigidly defined and easy to comprehend and compute. But median is not based on all the observation and does not allow itself to mathematical treatment. However, median is not much affected by sampling fluctuation and it is the most appropriate measure of central tendency for an open-end classification.

Although mode is the most popular measure of central tendency, there are cases when mode remains undefined. Unlike mean, it has no mathematical property. Mode is also affected by sampling fluctuations.

GM and HM, like AM, possess some mathematical properties. They are rigidly defined and based on all the observations. But they are difficult to comprehend and compute and, as such, have limited applications for the computation of average rates and ratios and such like things.

Example 15.1.19: Given two positive numbers a and b, prove that **AH=G²**. Does the result hold for any set of observations?

Solution: For two positive numbers a and b, we have,

$$A = \frac{a+b}{2}$$

$$G = \sqrt{ab}$$

And $$H = \frac{2}{\frac{1}{a}+\frac{1}{b}}$$

$$= \frac{2ab}{a+b}$$

Thus $$AH = \frac{a+b}{2} \times \frac{2ab}{a+b}$$

$$= ab = G^2$$

This result holds for only two positive observations and not for any set of observations.

Example 15.1.20: The AM and GM for two observations are 5 and 4 respectively. Find the two observations.

Solution: If a and b are two positive observations then as given

$$\frac{a+b}{2} = 5$$

$\Rightarrow$ a+b = 10 ....(1)

and $\sqrt{ab} = 4$

$\Rightarrow$ ab = 16 ....(2)

$$\therefore (a-b)^2 = (a+b)^2 - 4ab$$

$$= 10^2 - 4 \times 16$$

$$= 36$$

$\Rightarrow \quad a - b = 6$ (ignoring the negative sign)..........................(3)

Adding (1) and (3) We get,

$$2a = 16$$

$$\Rightarrow \quad a = 8$$

From (1), we get $b = 10 - a = 2$

Thus, the two observations are 8 and 2.

Example 15.1.21: Find the mean and median from the following data:

Marks	:	less than 10	less than 20	less than 30
No. of Students	:	5	13	23
Marks	:	less than 40	less than 50	
No. of Students	:	27	30	

Also compute the mode using the approximate relationship between mean, median and mode.

Solution: What we are given in this problem is less than cumulative frequency distribution. We need to convert this cumulative frequency distribution to the corresponding frequency distribution and thereby compute the mean and median.

Table 15.1.19

Computation of Mean Marks for 30 students

Marks Class Interval (1)	No. of Students (f_i) (2)	Mid - Value (x_i) (3)	f_ix_i (4)= (2)×(3)
0 – 10	5	5	25
10 – 20	13 – 5 = 8	15	120
20 – 30	23 – 13 = 10	25	250
30 – 40	27 – 23 = 4	35	140
40 – 50	30 – 27 = 3	45	135
Total	30	–	670

Hence the mean mark is given by

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$= \frac{670}{30}$$

$$= 22.33$$

Table 15.1.10

Computation of Median Marks

Marks (Class Boundary)	No.of Students (Less than cumulative Frequency)
0	0
10	5
20	13
30	23
40	27
50	30

Since $\frac{N}{2} = \frac{30}{2} = 15$ lies between 13 and 23,

we have $l_1 = 20$, $N_l = 13$, $N_u = 23$

and $C = l_2 - l_1 = 30 - 20 = 10$

Thus,

$$\text{Median} = 20 + \frac{15 - 13}{23 - 13} \times 10$$

$$= 22$$

Since Mode = 3 Median – 2 Mean (approximately), we find that

Mode = 3x22 – 2x22.33

= 21.34

Example 15.1.22: Following are the salaries of 20 workers of a firm expressed in thousand rupees: 5, 17, 12, 23, 7, 15, 4, 18, 10, 6, 15, 9, 8, 13, 12, 2, 12, 3, 15, 14. The firm gave bonus amounting to ₹ 2,000, ₹ 3,000, ₹ 4,000, ₹ 5,000 and ₹ 6,000 to the workers belonging to the salary groups 1,000 – 5,000, 6,000 – 10,000 and so on and lastly 21,000 – 25,000. Find the average bonus paid per employee.

Solution: We first construct frequency distribution of salaries paid to the 20 employees. The average bonus paid per employee is given by $\frac{\sum f_i x_i}{N}$ Where x_i represents the amount of bonus paid to the i^{th} salary group and f_i, the number of employees belonging to that group which would be obtained on the basis of frequency distribution of salaries.

Table 15.1.11
Computation of Average bonus

Salary in thousand ₹ (Class Interval) (1)	Tally Mark (2)	No of workers (f_i) (3)	Bonus in Rupees x_i (4)	$f_i x_i$ (5) = (3) × (4)
1-5	IIII	4	2000	8000
6-10	卌	5	3000	15000
11-15	卌 III	8	4000	32000
16-20	II	2	5000	10000
21-25	I	1	6000	6000
TOTAL	–	20	–	71000

Hence, the average bonus paid per employee

$$= (₹)\frac{71000}{20}$$

(₹) = 3550

SUMMARY

- The best measure of central tendency, usually, is the AM. It is rigidly defined, based on all the observations, easy to comprehend, simple to calculate and amenable to mathematical properties. However, AM has one drawback in the sense that it is very much affected by sampling fluctuations. In case of frequency distribution, mean cannot be advocated for open-end classification.
- Median is also rigidly defined and easy to comprehend and compute. But median is not based on all the observation and does not allow itself to mathematical treatment. However, median is not much affected by sampling fluctuation and it is the most appropriate measure of central tendency for an open-end classification.
- Mode is the most popular measure of central tendency, there are cases when mode remains undefined. Unlike mean, it has no mathematical property. Mode is also affected by sampling fluctuations.
- Relationship between Mean, Median and Mode

  Mean – Mode = 3(Mean – Median)

  Mode = 3 Median – 2 Mean

- Relation between AM, GM, and HM

  AM ≥ GM ≥ HM
- GM and HM, like AM, possess some mathematical properties. They are rigidly defined and based on all the observations. But they are difficult to comprehend and compute and, as such, have limited applications for the computation of average rates and ratios and such like things.

EXERCISE — UNIT-I

Set A

Write down the correct answers. Each question carries 1 mark.

1. Measures of central tendency for a given set of observations measures
   (a) The scatterness of the observations (b) The central location of the observations
   (c) Both (a) and (b) (d) None of these.
2. While computing the AM from a grouped frequency distribution, we assume that
   (a) The classes are of equal length
   (b) The classes have equal frequency
   (c) All the values of a class are equal to the mid-value of that class
   (d) None of these.
3. Which of the following statements is wrong?
   (a) Mean is rigidly defined
   (b) Mean is not affected due to sampling fluctuations
   (c) Mean has some mathematical properties
   (d) All these
4. Which of the following statements is true?
   (a) Usually mean is the best measure of central tendency
   (b) Usually median is the best measure of central tendency
   (c) Usually mode is the best measure of central tendency
   (d) Normally, GM is the best measure of central tendency
5. For open-end classification, which of the following is the best measure of central tendency?
   (a) AM (b) GM (c) Median (d) Mode
6. The presence of extreme observations does not affect
   (a) AM (b) Median (c) Mode (d) Any of these.
7. In case of an even number of observations which of the following is median?
   (a) Any of the two middle-most value

(b) The simple average of these two middle values

(c) The weighted average of these two middle values

(d) Any of these

8. The most commonly used measure of central tendency is

(a) AM (b) Median (c) Mode (d) Both GM and HM.

9. Which one of the following is not uniquely defined?

(a) Mean (b) Median (c) Mode (d) All of these measures

10. Which of the following measure of the central tendency is difficult to compute?

(a) Mean (b) Median (c) Mode (d) GM

11. Which measure(s) of central tendency is(are) considered for finding the average rates?

(a) AM (b) GM (c) HM (d) Both (b) and (c)

12. For a moderately skewed distribution, which of he following relationship holds?

(a) Mean – Mode = 3 (Mean – Median) (b) Median – Mode = 3 (Mean – Median)

(c) Mean – Median = 3 (Mean – Mode) (d) Mean – Median = 3 (Median – Mode)

13. Weighted averages are considered when

(a) The data are not classified

(b) The data are put in the form of grouped frequency distribution

(c) All the observations are not of equal importance

(d) Both (a) and (c).

14. Which of the following results hold for a set of distinct positive observations?

(a) AM $\geq$ GM $\geq$ HM (b) HM $\geq$ GM $\geq$ AM

(c) AM > GM > HM (d) GM > AM > HM

15. When a firm registers both profits and losses, which of the following measure of central tendency cannot be considered?

(a) AM (b) GM (c) Median (d) Mode

16. Quartiles are the values dividing a given set of observations into

(a) Two equal parts (b) Four equal parts (c) Five equal parts (d) None of these

17. Quartiles can be determined graphically using

(a) Histogram (b) Frequency Polygon (c) Ogive (d) Pie chart.

18. Which of the following measure(s) possesses (possess) mathematical properties?

(a) AM (b) GM (c) HM (d) All of these

19. Which of the following measure(s) satisfies (satisfy) a linear relationship between two variables?

(a) Mean (b) Median (c) Mode (d) All of these

20. Which of he following measures of central tendency is based on only fifty percent of the central values?

(a) Mean (b) Median (c) Mode (d) Both (a) and (b)

Set B

Write down the correct answers. Each question carries 2 marks.

1. If there are 3 observations 15, 20, 25 then the sum of deviation of the observations from their AM is

(a) 0 (b) 5 (c) –5 (d) None of these.

2. What is the median for the following observations?

5, 8, 6, 9, 11, 4.

(a) 6 (b) 7 (c) 8 (d) None of these

3. What is the modal value for the numbers 5, 8, 6, 4, 10, 15, 18, 10?

(a) 18 (b) 10 (c) 14 (d) None of these

4. What is the GM for the numbers 8, 24 and 40?

(a) 24 (b) 12 (c) $8 \cdot \sqrt[3]{15}$ (d) 10

5. The harmonic mean for the numbers 2, 3, 5 is

(a) 2.00 (b) 3.33 (c) 2.90 (d) $-\sqrt[3]{30}$.

6. If the AM and GM for two numbers are 6.50 and 6 respectively then the two numbers are

(a) 6 and 7 (b) 9 and 4 (c) 10 and 3 (d) 8 and 5.

7. If the AM and HM for two numbers are 5 and 3.2 respectively then the GM will be

(a) 16.00 (b) 4.10 (c) 4.05 (d) 4.00.

8. What is the value of the first quartile for observations 15, 18, 10, 20, 23, 28, 12, 16?

(a) 17 (b) 16 (c) 12.75 (d) 12

9. The third decile for the numbers 15, 10, 20, 25, 18, 11, 9, 12 is

(a) 13 (b) 10.70 (c) 11 (d) 11.50

10. If there are two groups containing 30 and 20 observations and having 50 and 60 as arithmetic means, then the combined arithmetic mean is

(a) 55 (b) 56 (c) 54 (d) 52.

11. The average salary of a group of unskilled workers is ₹ 10,000 and that of a group of skilled workers is ₹ 15,000. If the combined salary is ₹ 12,000, then what is the percentage of skilled workers?

    (a) 40% (b) 50% (c) 60% (d) none of these

12. If there are two groups with 75 and 65 as harmonic means and containing 15 and 13 observation then the combined HM is given by

    (a) 65 (b) 70.36 (c) 70 (d) 71.

13. What is the HM of 1,1/2, 1/3,……………1/n?

    (a) n (b) 2n (c) $\frac{2}{(n+1)}$ (d) $\frac{n(n+1)}{2}$

14. An aeroplane flies from A to B at the rate of 500 km/hour and comes back from B to A at the rate of 700 km/hour. The average speed of the aeroplane is

    (a) 600 km. per hour (b) 583.33 km. per hour

    (c) $100\sqrt{35}$ km. per hour (d) 620 km. per hour.

15. If a variable assumes the values 1, 2, 3…5 with frequencies as 1, 2, 3…5, then what is the AM?

    (a) $\frac{11}{3}$ (b) 5 (c) 4 (d) 4.50

16. Two variables x and y are given by y= 2x – 3. If the median of x is 20, what is the median of y?

    (a) 20 (b) 40 (c) 37 (d) 35

17. If the relationship between two variables u and v are given by 2u + v + 7 = 0 and if the AM of u is 10, then the AM of v is

    (a) 17 (b) –17 (c) –27 (d) 27.

18. If x and y are related by x–y–10 = 0 and mode of x is known to be 23, then the mode of y is

    (a) 20 (b) 13 (c) 3 (d) 23.

19. If GM of x is 10 and GM of y is 15, then the GM of xy is

    (a) 150 (b) log 10 × log 15 (c) log 150 (d) None of these.

20. If the AM and GM for 10 observations are both 15, then the value of HM is

    (a) Less than 15 (b) More than 15

    (c) 15 (d) Can not be determined.

Set C

Write down the correct answers. Each question carries 5 marks.

1. What is the value of mean and median for the following data:

Marks:	5–14	15–24	25–34	35–44	45–54	55–64
No. of Students:	10	18	32	26	14	10

(a) 30 and 28 (b) 29 and 30 (c) 33.68 and 37.94 (d) 34.21 and 33.18

2. The mean and mode for the following frequency distribution

Class interval:	350–369	370–389	390–409	410–429	430–449	450–469
Frequency:	15	27	31	19	13	6

are

(a) 400 and 390 (b) 400.58 and 390 (c) 400.58 and 394.50 (d) 400 and 394.

3. The median and modal profits for the following data

Profit in '000 ₹:	below 5	below 10	below 15	below 20	below 25	below 30
No. of firms:	10	25	45	55	62	65

are

(a) 11.60 and 11.50 (b) ₹ 11556 and ₹ 11267

(c) ₹ 11875 and ₹ 11667 (d) 11.50 and 11.67.

4. Following is an incomplete distribution having modal mark as 44

Marks:	0–20	20–40	40–60	60–80	80–100
No. of Students:	5	18	?	12	5

What would be the mean marks?

(i) 45 (ii) 46 (iii) 47 (iv) 48

5. The data relating to the daily wage of 20 workers are shown below:

₹ 50, ₹ 55, ₹ 60, ₹ 58, ₹ 59, ₹ 72, ₹ 65, ₹ 68, ₹ 53, ₹ 50, ₹ 67, ₹ 58,
₹ 63, ₹ 69, ₹ 74, ₹ 63, ₹ 61, ₹ 57, ₹ 62, ₹ 64.

The employer pays bonus amounting to ₹ 100, ₹ 200, ₹ 300, ₹ 400 and ₹ 500 to the wage earners in the wage groups ₹ 50 and not more than ₹ 55 ₹ 55 and not more than ₹ 60 and so on and lastly ₹ 70 and not more than ₹ 75, during the festive month of October.

What is the average bonus paid per wage earner?

(a) ₹ 200 (b) ₹ 250 (c) ₹ 285 (d) ₹300

6. The third quartile and 65th percentile for the following data are

Profits in '000 ₹:	les than 10	10–19	20–29	30–39	40–49	50–59
No. of firms:	5	18	38	20	9	2

(a) ₹ 33,500 and ₹ 29,184 (b) ₹ 33,000 and ₹ 28,680

(c) ₹ 33,600 and ₹ 29,000 (d) ₹ 33,250 and ₹ 29,250.

7. For the following incomplete distribution of marks of 100 pupils, median mark is known to be 32.

Marks:	0–10	10–20	20–30	30–40	40–50	50–60
No. of Students:	10	–	25	30	–	10

What is the mean mark?

(a) 32 (b) 31 (c) 31.30 (d) 31.50

8. The mode of the following distribution is ₹ 66. What would be the median wage?

Daily wages (₹):	30–40	40–50	50–60	60–70	70–80	80–90
No of workers:	8	16	22	28	–	12

(a) ₹ 64.00 (b) ₹ 64.56 (c) ₹ 62.32 (d) ₹ 64.25

ANSWERS

Set A

1. (b)	**2.** (c)	**3.** (b)	**4.** (a)	**5.** (c)	**6.** (b)
7. (b)	**8.** (a)	**9.** (c)	**10.** (d)	**11.** (d)	**12.** (a)
13. (c)	**14.** (c)	**15.** (b)	**16.** (b)	**17.** (c)	**18.** (d)
19. (d)	**20.** (b)				

Set B

1. (a)	**2.** (b)	**3.** (b)	**4.** (c)	**5.** (c)	**6.** (b)
7. (d)	**8.** (c)	**9.** (b)	**10.** (c)	**11.** (a)	**12.** (c)
13. (c)	**14.** (b)	**15.** (a)	**16.** (c)	**17.** (c)	**18.** (b)
19. (a)	**20.** (c)				

Set C

1. (c)	**2.** (c)	**3.** (c)	**4.** (d)	**5.** (d)	**6.** (a)
7. (c)	**8.** (c)				

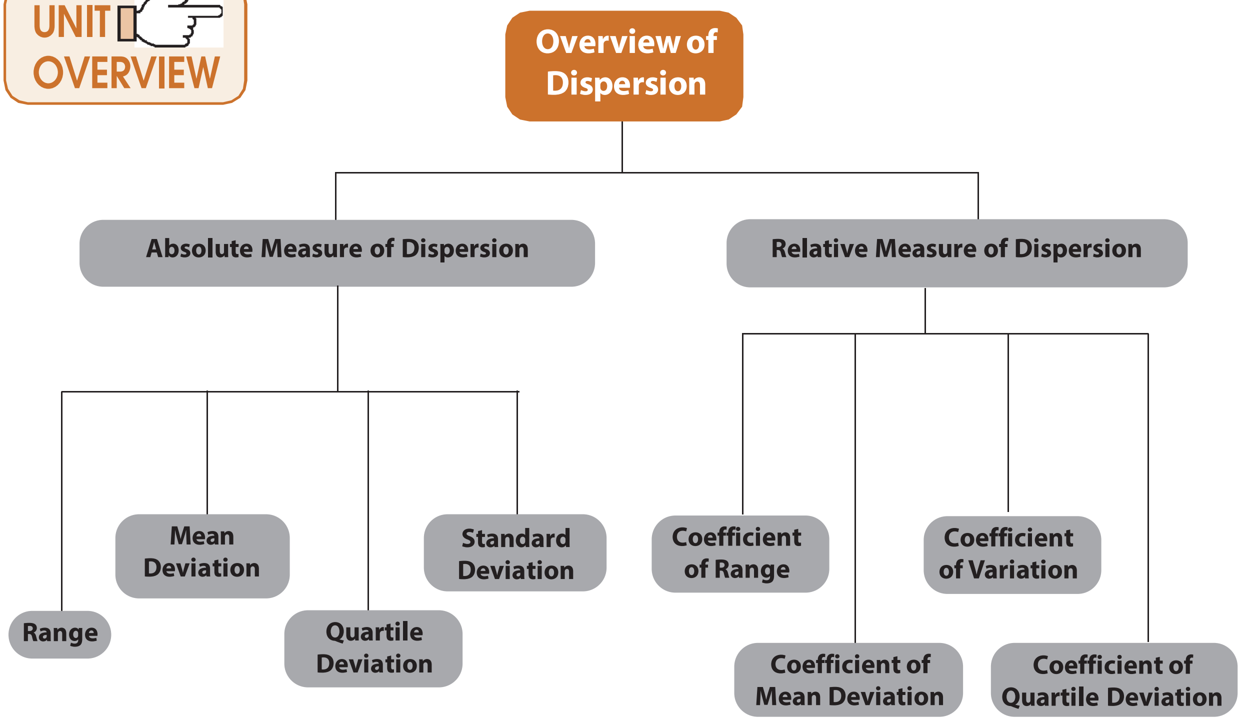
UNIT II: DISPERSION

LEARNING OBJECTIVES

After reading this chapter, students will be able to understand:

- To understand different measures of Dispersion i.e Range, Quartile Deviation, Mean Deviation and Standard Deviation and computational techniques of these measures.
- To learn comparative advantages and disadvantages of these measures and therefore, which measures to use in which circumstance.
- To understand a set of observation, it is equally important to have knowledge of dispersion which indicates the volatility. In advanced stage of chartered accountancy course, volatility measures will be useful in understanding risk involved in financial decision making.

UNIT OVERVIEW



15.2.1 DEFINITION OF DISPERSION

The second important characteristic of a distribution is given by dispersion. Two distributions may be identical in respect of its first important characteristic i.e. central tendency and yet they may differ on account of scatterness. The following figure shows a number of distributions having identical measure of central tendency and yet varying measure of scatterness. Obviously, distribution is having the maximum amount of dispersion.

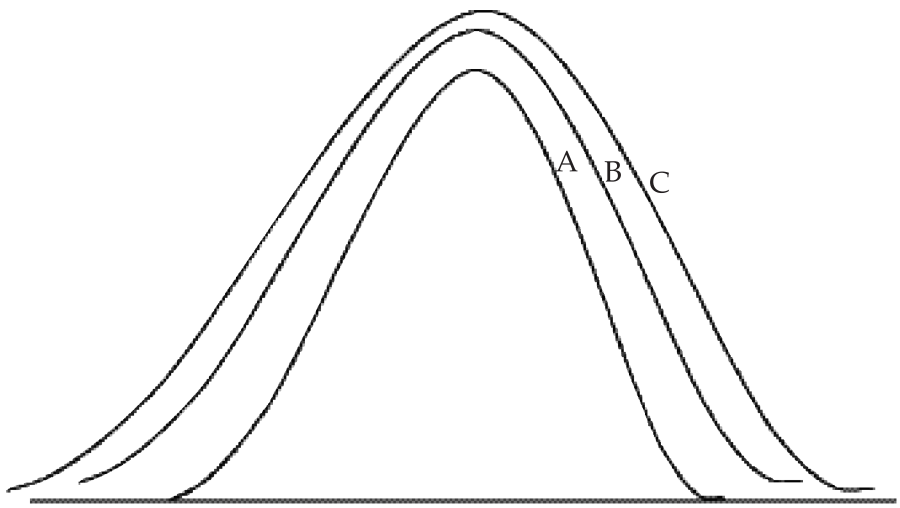


Figure 15.2.1

Showing distributions with identical measure of central tendency and varying amount of dispersion.

Dispersion for a given set of observations may be defined as the amount of deviation of the observations, usually, from an appropriate measure of central tendency. Measures of dispersion may be broadly classified into

1. Absolute measures of dispersion. **2. Relative measures of dispersion.**

Absolute measures of dispersion are classified into

(i) Range (ii) Mean Deviation

(iii) Standard Deviation (iv) Quartile Deviation

Likewise, we have the following relative measures of dispersion :

(i) Coefficient of Range. (ii) Coefficient of Mean Deviation

(iii) Coefficient of Variation (iv) Coefficient of Quartile Deviation.

We may note the following points of distinction between the absolute and relative measures of dispersion :

I Absolute measures are dependent on the unit of the variable under consideration whereas the relative measures of dispersion are unit free.

II For comparing two or more distributions, relative measures and not absolute measures of dispersion are considered.

III Compared to absolute measures of dispersion, relative measures of dispersion are difficult to compute and comprehend.

Characteristics for an ideal measure of dispersion

As discussed in section 15.2.1 an ideal measure of dispersion should be properly defined, easy to comprehend, simple to compute, based on all the observations, unaffected by sampling fluctuations and amenable to some desirable mathematical treatment.

15.2.2 RANGE

For a given set of observations, range may be defined as the difference between the largest and smallest of observations. Thus if L and S denote the largest and smallest observations respectively then we have

$$\text{Range} = L - S$$

The corresponding relative measure of dispersion, known as coefficient of range, is given by

$$\text{Coefficient of range} = \frac{L-S}{L+S} \times 100$$

For a grouped frequency distribution, range is defined as the difference between the two extreme class boundaries. The corresponding relative measure of dispersion is given by the ratio of the difference between the two extreme class boundaries to the total of these class boundaries, expressed as a percentage.

We may note the following important result in connection with range:

Result:

Range remains unaffected due to a change of origin but affected in the same ratio due to a change in scale i.e., if for any two constants a and b, two variables x and y are related by $y = a + bx$,

Then the range of y is given by

$$R_y = |b| \times R_x \quad \ldots\ldots\ldots\ldots (15.2.1)$$

Example 15.2.1: Following are the wages of 8 workers expressed in Rupees.
82, 96, 52, 75, 70, 65, 50, 70. Find the range and also its coefficient.

Solution: The largest and the smallest wages are L = ₹ 96 and S= ₹ 50
Thus range = ₹ 96 – ₹ 50 = ₹ 46

$$\text{Coefficient of range} = \frac{96-50}{96+50} \times 100$$
$$= 31.51$$

Example 15.2.2: What is the range and its coefficient for the following distribution of weights?

Weights in kgs. :	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74
No. of Students :	12	18	23	10	3

Solution: The lowest class boundary is 49.50 kgs. and the highest class boundary is 74.50 kgs. Thus we have

Range = 74.50 kgs. – 49.50 kgs.
= 25 kgs.

Also, coefficient of range $= \frac{74.50 - 49.50}{74.50 + 49.50} \times 100$

$= \frac{25}{124} \times 100$

$= 20.16$

Example 15.2.3 : If the relationship between x and y is given by 2x+3y=10 and the range of x is ₹ 15, what would be the range of y?

Solution: Since 2x+3y=10

Therefore, $y = \frac{10}{3} - \frac{2}{3}x$

Applying (11.23) , the range of y is given by

$$R_y = |b| \times R_x$$
$$= 2/3 \times ₹\ 15$$
$$= ₹\ 10.$$

15.2.3 MEAN DEVIATION

Since range is based on only two observations, it is not regarded as an ideal measure of dispersion. A better measure of dispersion is provided by mean deviation which, unlike range, is based on all the observations. For a given set of observation, mean deviation is defined as the arithmetic mean of the absolute deviations of the observations from an appropriate measure of central tendency. Hence if a variable x assumes n values $x_1, x_2, x_3 \ldots x_n$, then the mean deviation of x about an average A is given by

$$MD_A = \frac{1}{n}\sum|x_i - A| \qquad \ldots\ldots(15.2.2)$$

For a grouped frequency distribution, mean deviation about A is given by

$$MD_A = \frac{1}{n}\sum|x_i - A|f_i \qquad \ldots\ldots(15.2.2)$$

Where x_i and f_i denote the mid value and frequency of the i-th class interval and

$$N = \sum f_i$$

In most cases we take A as mean or median and accordingly, we get mean deviation about mean or mean deviation about median.

A relative measure of dispersion applying mean deviation is given by

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation about A}}{\text{A}} \times 100 \quad \text{...............(15.2.3)}$$

Mean deviation takes its minimum value when the deviations are taken from the median. Also mean deviation remains unchanged due to a change of origin but changes in the same ratio due to a change in scale i.e. if y = a + bx, a and b being constants,

then MD of y = |b| × MD of x ....(15.2.4)

Example 15.2.4: What is the mean deviation about mean for the following numbers?

5, 8, 10, 10, 12, 9.

Solution:

The mean is given by

$$\overline{X} = \frac{5+8+10+10+12+9}{6} = 9$$

Table 15.2.1

Computation of MD about AM

x_i	$\lvert x_i - \overline{x} \rvert$
5	4
8	1
10	1
10	1
12	3
9	0
Total	10

Thus mean deviation about mean is given by

$$\frac{\sum \lvert x_i - \overline{x} \rvert}{n} = \frac{10}{6} = 1.67$$

Example. 15.2.5: Find mean deviations about median and also the corresponding coefficient for the following profits ('000 ₹) of a firm during a week.

82, 56, 75, 70, 52, 80, 68.

Solution:

The profits in thousand rupees is denoted by x. Arranging the values of x in an ascending order, we get

52, 56, 68, 70, 75, 80, 82.

Therefore, Me = 70. Thus, Median profit = ₹ 70,000.

Table 15.2.2

Computation of Mean deviation about median

x_i	$\lvert x_i - Me \rvert$
52	18
56	14
68	2
70	0
75	5
80	10
82	12
Total	61

Thus mean deviation about median = $\frac{\sum \lvert x_i - \text{Median} \rvert}{n}$

$= (₹)\frac{61}{7}$

= ₹ 8714.28

Coefficient of mean deviation = $\frac{\text{MD about median}}{\text{Median}} \times 100$

$= \frac{8714.28}{70000} \times 100$

= 12.45

Example 15.2.6 : Compute the mean deviation about the arithmetic mean for the following data:

x :	1	3	5	7	9
f :	5	8	9	2	1

lso find the coefficient of the mean deviation about the AM.

Solution: We are to apply formula (15.1.2) as these data refer to a grouped frequency distribution the AM is given by

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$= \frac{5\times1+8\times3+9\times5+2\times7+1\times9}{5+8+9+2+1} = 3.88$$

Table 15.2.3

Computation of MD about the AM

x (1)	f (2)	$\lvert x-\bar{x}\rvert$ (3)	$f\lvert x-\bar{x}\rvert$ (4) = (2) × (3)
1	5	2.88	14.40
3	8	0.88	7.04
5	9	1.12	10.08
7	2	3.12	6.24
9	1	5.12	5.12
Total	25	–	42.88

Thus, MD about AM is given by

$$\frac{\sum f\lvert x-\bar{x}\rvert}{N}$$

$$= \frac{42.88}{25}$$

$$= 1.72$$

Coefficient of MD about its AM $= \frac{\text{MD about AM}}{\text{AM}} \times 100$

$$= \frac{1.72}{3.88}\times100$$

$$= 44.33$$

Example 15.2.7: Compute the coefficient of mean deviation about median for the following distribution:

Weight in kgs.	:	40-50	50-60	60-70	70-80
No. of persons	:	8	12	20	10

Solution: We need to compute the median weight in the first stage

Table 15.2.4

Computation of median weight

Weight in kg (CB)	No. of Persons (Cumulative Frequency)
40	0
50	8
60	20
70	40
80	50

Hence, $$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l}\right) \times C$$

$$= \left[60 + \frac{25 - 20}{40 - 20} \times 10\right] \text{kg.} = 62.50\text{kg.}$$

Table 15.2.5

Computation of mean deviation of weight about median

weight (kgs.) (1)	mid-value (x_i) kgs. (2)	No. of persons (f_i) (3)	$\|x_i - Me\|$ (kgs.) (4)	$f_i\|x_i - Me\|$ (kgs.) (5)=(3)×(4)
40–50	45	8	17.50	140
50–60	55	12	7.50	90
60–70	65	20	2.50	50
70–80	75	10	12.50	125
Total	–	50	–	405

$$\text{Mean deviation about median} = \frac{\sum f_i |x_i - \text{Median}|}{N}$$

$$= \frac{405}{50} \text{kg.}$$

$$= 8.10 \text{ kg.}$$

$$\text{Coefficient of mean deviation about median} = \frac{\text{Mean deviation about median}}{\text{Median}} \times 100$$

$$= \frac{8.10}{62.50} \times 100$$

$$= 12.96$$

Example 15.2.8: If x and y are related as 4x+3y+11 = 0 and mean deviation of x is 5.40, what is the mean deviation of y?

Solution: Since 4x + 3y + 11 = 0

Therefore, $y = \left(\frac{-11}{3}\right) + \left(\frac{-4}{3}\right)x$

Hence MD of y= |b| × MD of x

$$= \frac{4}{3} \times 5.40$$

$$= 7.20$$

15.2.4 STANDARD DEVIATION

Although mean deviation is an improvement over range so far as a measure of dispersion is concerned, mean deviation is difficult to compute and further more, it cannot be treated mathematically. The best measure of dispersion is, usually, standard deviation which does not possess the demerits of range and mean deviation.

Standard deviation for a given set of observations is defined as the root mean square deviation when the deviations are taken from the AM of the observations. If a variable x assumes n values x_1, x_2, x_3 ………..x_n then its standard deviation(s) is given by

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \qquad \text{............................(15.2.5)}$$

For a grouped frequency distribution, the standard deviation is given by

$$s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}} \qquad \text{............................ (15.2.6)}$$

(15.2.5) and (15.2.6) can be simplified to the following forms

$$s = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \quad \text{for unclassified data}$$

$$= \sqrt{\frac{\sum f_i x_i^2}{N} - \bar{x}^2} \quad \text{for a grouped frequency distribution.}$$

.... (15.2.7)

Sometimes the square of standard deviation, known as variance, is regarded as a measure of dispersion. We have, then,

$$\text{Variance} = s^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad \text{for unclassified data}$$

$$= \frac{\sum f_i (x_i - \bar{x})^2}{N} \quad \text{for a grouped frequency distribution} \ldots\ldots\ldots\ldots\ldots\ldots (15.2.8)$$

A relative measure of dispersion using standard deviation is given by coefficient of variation (cv) which is defined as the ratio of standard deviation to the corresponding arithmetic mean, expressed as a percentage.

$$\text{Coefficient of Variation (CV)} = \frac{SD}{AM} \times 100 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15..2.9)$$

ILLUSTRATIONS:

Example 15.2.9: Find the standard deviation and the coefficient of variation for the following numbers: 5, 8, 9, 2, 6

Solution: We present the computation in the following table.

Table 15.2.6
Computation of standard deviation

x_i	x_i^2
5	25
8	64
9	81
2	4
6	36
30	$\sum x_i^2 = 210$

Applying (15.2.7), we get the standard deviation as

$$s = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{210}{5} - \left(\frac{30}{5}\right)^2} \qquad \left(\text{since } \bar{x} = \frac{\Sigma x_i}{n}\right)$$

$$= \sqrt{42 - 36}$$

$$= \sqrt{6}$$

$$= 2.45$$

The coefficient of variation is

$$CV = 100 \times \frac{SD}{AM}$$

$$= 100 \times \frac{2.45}{6}$$

$$= 40.83$$

Example 15.2.10: Show that for any two numbers a and b, standard deviation is given by $\frac{|a-b|}{2}$.

Solution: For two numbers a and b, AM is given by $\bar{x} = \frac{a+b}{2}$

The variance is

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{2}$$

$$= \frac{\left(a - \frac{a+b}{2}\right)^2 + \left(b - \frac{a+b}{2}\right)^2}{2}$$

$$= \frac{\frac{(a-b)^2}{4} + \frac{(a-b)^2}{4}}{2}$$

$$= \frac{(a-b)^2}{4}$$

$$\Rightarrow \quad s = \frac{|a-b|}{2}$$

(The absolute sign is taken, as SD cannot be negative).

Example 15.2.11: Prove that for the first n natural numbers, SD is $\sqrt{\frac{n^2-1}{12}}$.

Solution: for the first n natural numbers AM is given by

$$\bar{x} = \frac{1+2+3+.................+n}{n}$$

$$= \frac{n(n+1)}{2n}$$

$$= \frac{n+1}{2}$$

$$\therefore \text{ SD } = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{1^2+2^2+3^2.................+n^2}{n} - \left(\frac{n+1}{2}\right)^2}$$

$$= \sqrt{\frac{n(n+1)(2n+1)}{6n} - \frac{(n+1)^2}{4}}$$

$$= \sqrt{\frac{(n+1)(4n+2-3n-3)}{12}} = \sqrt{\frac{n^2-1}{12}}$$

Thus, SD of first n natural numbers is SD = $\sqrt{\frac{n^2-1}{12}}$

We consider the following formula for computing standard deviation from grouped frequency distribution with a view to saving time and computational labour:

$$S = \sqrt{\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2} \qquad ....(15.2.10)$$

Where $d_i = \frac{x_i - A}{C}$

Example 15.2.12: Find the SD of the following distribution:

Weight (kgs.)	:	50-52	52-54	54-56	56-58	58-60
No. of Students	:	17	35	28	15	5

Solution:

Table 15.2.7
Computation of SD

Weight (kgs.) (1)	No. of Students (f_i) (2)	Mid-value (x_i) (3)	$d_i = \frac{x_i - 55}{2}$ (4)	$f_i d_i$ (5)=(2)×(4)	$f_i d_i^2$ (6)=(5)×(4)
50-52	17	51	–2	–34	68
52-54	35	53	–1	–35	35
54-56	28	55	0	0	0
56-58	15	57	1	15	15
58-60	5	59	2	10	20
Total	100	–	–	– 44	138

Applying (11.33), we get the SD of weight as

$$= \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2} \times C$$

$$= \sqrt{\frac{138}{100} - \frac{(-44)^2}{100}} \times 2\text{kgs.}$$

$$= \sqrt{1.38 - 0.1936} \times 2\text{ kgs.}$$

$$= 2.18\text{ kgs.}$$

Properties of standard deviation

I. If all the observations assumed by a variable are constant i.e. equal, then the SD is zero. This means that if all the values taken by a variable x is k, say, then s = 0. This result applies to range as well as mean deviation.

II. SD remains unaffected due to a change of origin but is affected in the same ratio due to a change of scale i.e., if there are two variables x and y related as y = a+bx for any two constants a and b, then SD of y is given by

$$s_y = |b|s_x \qquad \text{..........................(15.2.11)}$$

III. If there are two groups containing n_1 and n_2 observations, $\bar{x}_1$ and $\bar{x}_2$ as respective AM's, s_1 and s_2 as respective SD's , then the combined SD is given by

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}} \qquad \text{..........................(15.2.12)}$$

where, $d_1 = \bar{x}_1 - \bar{x}$

$d_2 = \bar{x}_2 - \bar{x}$

and $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$ = combined AM

This result can be extended to more than 2 groups. For x(7²) groups, we have

$$s = \sqrt{\frac{\sum n_i s_i^2 + \sum n_i d_i^2}{\sum n_i}} \quad \text{.............................} \quad (15.2.13)$$

With $d_i = x_i - \bar{x}$

and $\bar{x} = \dfrac{\sum n_i \bar{x}_i}{\sum n_i}$

Where $\bar{x}_1 = \bar{x}_2$ (11.35) is reduced to

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$$

Example 15.2.13: If AM and coefficient of variation of x are 10 and 40 respectively, what is the variance of (15–2x)?

Solution: let y = 15 – 2x

Then applying (11.34), we get,

$s_y = 2 \times s_x$ .... (1)

As given cv_x = coefficient of variation of x = 40 and $\bar{x}$ = 10

This $cv_x = \dfrac{s_x}{x} \times 100$

$\Rightarrow \quad 40 = \dfrac{S_x}{10} \times 100$

$\Rightarrow \quad S_x = 4$

From (1), $S_y = 2 \times 4 = 8$

Therefore, variance of $(15 - 2x) = S_y^2 = 64$

Example 15.2.14: Compute the SD of 9, 5, 8, 6, 2.

Without any more computation, obtain the SD of

Sample I	–1,	–5,	–2,	–4,	–8,
Sample II	90,	50,	80,	60,	20,
Sample III	23,	15,	21,	17,	9.

Solution:

Table 15.2.7
Computation of SD

x_i	x_i^2
9	81
5	25
8	64
6	36
2	4
30	210

The SD of the original set of observations is given by

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

$$= \sqrt{\frac{210}{5} - \left(\frac{30}{5}\right)^2}$$

$$= \sqrt{42 - 36}$$

$$= \sqrt{6}$$

$$= 2.45$$

If we denote the original observations by x and the observations of sample I by y, then we have

$$y = -10 + x$$

$$y = (\ \ 10) + (1)\,x$$

$$\therefore S_y = |1| \times S_x$$

$$= 1 \times 2.45$$

$$= 2.45$$

In case of sample II, x and y are related as

$$Y = 10x$$

$$= 0 + (15)x$$

$\therefore s_y = |10| \times s_x$

$= 10 \times 2.45$

$= 24.50$

And lastly, y= (5)+(2)x

$\Rightarrow s_y = 2 \times 2.45$

$= 4.90$

Example 15.2.15: For a group of 60 boy students, the mean and SD of stats. marks are 45 and 2 respectively. The same figures for a group of 40 girl students are 55 and 3 respectively. What is the mean and SD of marks if the two groups are pooled together?

Solution: As given $n_1 = 60$, $\bar{x}_1 = 45$, $s_1 = 2$ $n_2 = 40$, $\bar{x}_2 = 55$, $s_2 = 3$

Thus the combined mean is given by

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

$$= \frac{60 \times 45 + 40 \times 55}{60 + 40}$$

$= 49$

Thus $d_1 = \bar{x}_1 - \bar{x} = 45 - 49 = -4$

$d_2 = \bar{x}_2 - \bar{x} = 55 - 49 = 6$

Applying (11.35), we get the combined SD as

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

$$s = \sqrt{\frac{60 \times 2^2 + 40 \times 3^2 + 60 \times (-4)^2 + 40 \times 6^2}{60 + 40}}$$

$= \sqrt{30}$

$= 5.48$

Example 15.2.16: The mean and standard deviation of the salaries of the two factories are provided below :

Factory	No. of Employees	Mean Salary	SD of Salary
A	30	₹ 4800	₹ 10
B	20	₹ 5000	₹ 12

i) Find the combined mean salary and standard deviation of salary.
ii) Examine which factory has more consistent structure so far as satisfying its employees are concerned.

Solution: Here we are given

$n_1 = 30,\ \bar{x}_1 = ₹\ 4800,\ s_1 = ₹\ 10,$

$n_2 = 20,\ \bar{x}_2 = ₹\ 5000,\ s_2 = ₹\ 12$

i) $\dfrac{30 \times ₹\ 4800 + 20 \times ₹\ 5000}{30 + 20} = ₹\ 4800$

$d_1 = \bar{x}_1 - \bar{x} = ₹\ 4{,}800 - ₹\ 4880 = -\ ₹\ 80$

$d_2 = \bar{x}_2 - \bar{x} = ₹\ 5{,}000 - ₹\ 4880 = \ ₹\ 120$

hence, the combined SD in rupees is given by

$$s = \sqrt{\frac{30\times10^2 + 20\times12^2 + 30\times(-80)^2 + 20\times120^2}{30+20}}$$

$$= \sqrt{9717.60}$$

$$= 98.58$$

thus the combined mean salary and the combined standard deviation of salary are ₹ 4880 and ₹ 98.58 respectively.

ii) In order to find the more consistent structure, we compare the coefficients of variation of the two factories. Letting $CV_A = 100 \times \dfrac{S_A}{\bar{x}_A}$ and $CV_B = 100 \times \dfrac{S_B}{\bar{x}_B}$

We would say factory A is more consistent

if $CV_A < CV_B$. Otherwise factory B would be more consistent.

Now $CV_A = 100 \times \dfrac{s_A}{\bar{x}_A} = 100 \times \dfrac{s_1}{\bar{x}_1} = \dfrac{100 \times 10}{4800} = 0.21$

and $CV_B = 100 \times \dfrac{s_B}{\bar{x}_B} = 100 \times \dfrac{s_2}{\bar{x}_2} = \dfrac{100 \times 12}{5000} = 0.24$

Thus we conclude that factory A has more consistent structure.

Example 15.2.17: A student computes the AM and SD for a set of 100 observations as 50 and 5 respectively. Later on, she discovers that she has made a mistake in taking one observation as 60 instead of 50. What would be the correct mean and SD if

i) The wrong observation is left out?

ii) The wrong observation is replaced by the correct observation?

Solution: As given, n = 100, $\bar{x} = 50$, S = 5

Wrong observation = 60(x), correct observation = 50(V)

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

$$\Rightarrow \quad \sum x_i = n\bar{x} = 100 \times 50 = 5000$$

$$\text{and} \quad s^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

$$\Rightarrow \sum x_i^2 = n(\bar{x}^2 + s^2) = 100(50^2 + 5^2) = 252500$$

i) Sum of the 99 observations = 5000 – 60 = 4940

AM after leaving the wrong observation = 4940/99 = 49.90

Sum of squares of the observation after leaving the wrong observation

= 252500 – 60^2 = 248900

Variance of the 99 observations = 248900/99 – $(49.90)^2$

= 2514.14 – 2490.01

= 24.13

∴ SD of 99 observations = 4.91

ii) Sum of the 100 observations after replacing the wrong observation by the correct observation = 5000 – 60 + 50 = 4990

$$\text{AM} = \frac{4990}{100} = 49.90$$

Corrected sum of squares = $252500 + 50^2 - 60^2 = 251400$

Corrected SD $= \sqrt{\frac{251400}{100} - (49.90)^2}$

$= \sqrt{45.99}$

$= 6.78$

15.2.5 QUARTILE DEVIATION

Another measure of dispersion is provided by **quartile deviation** or **semi-inter–quartile** range which is given by

$$Q_d = \frac{Q_3 - Q_1}{2} \qquad \text{.................................(15.2.14)}$$

A relative measure of dispersion using quartiles is given by coefficient of quartile deviation which is

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100 \qquad \text{.................................(15.2.15)}$$

Quartile deviation provides the best measure of dispersion for open-end classification. It is also less affected due to sampling fluctuations. Like other measures of dispersion, quartile deviation remains unaffected due to a change of origin but is affected in the same ratio due to change in scale.

Example 15.2.18 : Following are the marks of the 10 students : 56, 48, 65, 35, 42, 75, 82, 60, 55, 50. Find quartile deviation and also its coefficient.

Solution:

After arranging the marks in an ascending order of magnitude, we get 35, 42, 48, 50, 55, 56, 60, 65, 75, 82

$$\text{First quartile } (Q_1) = \frac{(n+1)}{4} \text{th observation}$$

$$= \frac{(10+1)}{4} \text{th observation}$$

$$= 2.75^{th} \text{ observation}$$

$= 2^{nd}$ observation + 0.75 × difference between the third and the 2^{nd} observation.

$$= 42 + 0.75 \times (48 - 42)$$

$$= 46.50$$

$$\text{Third quartile } (Q_3) = \frac{3(n+1)}{4} \text{th observation}$$

$$= 8.25^{th} \text{ observation}$$

$$= 65 + 0.25 \times 10$$

$$= 67.50$$

Thus applying (11.37), we get the quartile deviation as

$$\frac{Q_3 - Q_1}{2} = \frac{67.50 - 46.50}{2} = 10.50$$

Also, using (11.38), the coefficient of quartile deviation

$$= \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

$$= \frac{67.50 - 46.50}{67.50 + 46.50} \times 100$$

$$= 18.42$$

Example 15.2.19 : If the quartile deviation of x is 6 and 3x + 6y = 20, what is the quartile deviation of y?

Solution: $3x + 6y = 20$

$$\Rightarrow \quad y = \left(\frac{20}{6}\right) + \left(\frac{-3}{6}\right)x$$

Therefore, quartile deviation of $y = \frac{|-3|}{6} \times \text{quartile deviation of } x$

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

$= 3.$

Example 15.2.20: Find an appropriate measures of dispersion from the following data:

Daily wages (₹)	:	upto 20	20-40	40-60	60-80	80-100
No. of workers (₹)	:	5	11	14	7	3

Solution: Since this is an open-end classification, the appropriate measure of dispersion would be quartile deviation as quartile deviation does not taken into account the first twenty five percent and the last twenty five per cent of the observations.

Table 15.2.8
Computation of Quartile

Daily wages in (₹) (Class boundary)	No. of workers (less than cumulative frequency)
a	0
20	5
40	16
60	30
80	37
100	40

Here a denotes the first Class Boundary

$Q_1 = ₹\left[20 + \frac{10-5}{16-5} \times 20\right] = ₹\ 29.09$

$Q_3 = ₹\ 60$

Thus quartile deviation of wages is given by

$$\frac{Q_3 - Q_1}{2}$$

$$= \frac{₹\,60 - ₹\,29.09}{2}$$

$= ₹\ 15.46$

Example 15.2.21: The mean and variance of 5 observations are 4.80 and 6.16 respectively. If three of the observations are 2, 3 and 6, what are the remaining observations?

Solution: Let the remaining two observations be a and b, then as given

$$\frac{2+3+6+a+b}{5} = 4.80$$

$\Rightarrow \quad 11+a+b = 24$

$\Rightarrow \quad a+b = 13 \quad$ ....(1)

and $\dfrac{2^2+a^2+b^2+3^2+6^2}{5} - (4.80)^2$

$\Rightarrow \quad \dfrac{49+a^2+b^2}{5} - 23.04 = 6.16$

$\Rightarrow \quad 49 + a^2 + b^2 = 146$

$\Rightarrow \quad a^2 + b^2 = 97 \quad$ ....(2)

From (1), we get $a = 13 - b$ ....(3)

Eliminating a from (2) and (3), we get

$(13 - b)^2 + b^2 = 97$

$\Rightarrow \quad 169 - 26b + 2b^2 = 97$

$\Rightarrow \quad b^2 - 13\,b + 36 = 0$

$\Rightarrow \quad (b-4)(b-9) = 0$

$\Rightarrow \quad b = 4$ or 9

From (3), $a = 9$ or 4

Thus the remaining observations are 4 and 9.

Example 15.2.22: After shift of origin and change of scale, a frequency distribution of a continuous variable with equal class length takes the following form of the changed variable (d):

d	:	–2	–1	0	1	2
Frequency	:	17	35	28	15	5

If the mean and standard deviation of the original frequency distribution are 54.12 and 2.1784 respectively, find the original frequency distribution.

Solution: We need find out the origin A and scale C from the given conditions.

Since $d_i = \dfrac{x_i - A}{C}$

$\Rightarrow \quad x_i = A + Cd_i$

Once A and C are known, the mid- values x_i's would be known. Finally, we convert the mid-values to the corresponding class boundaries by using the formula:

$$LCB = x_i - C/2$$

and $$UCB = x_i + C/2$$

On the basis of the given data, we find that

$$\Sigma f_i d_i = -44,\ \Sigma f_i d_i^2 = 138 \text{ and } N = 100$$

$$\text{Hence } s = \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2} \times C$$

$$\Rightarrow \quad 2.1784 = \sqrt{\frac{138}{100} - \left(\frac{-44}{100}\right)^2} \times C$$

$$\Rightarrow \quad 2.1784 = \sqrt{1.38 - 0.1936} \times C$$

$$\Rightarrow \quad 2.1784 = 1.0892 \times C$$

$$\Rightarrow \quad C = 2$$

$$\text{Further, } \bar{x} = A + \frac{\Sigma f_i d_i}{N} \times C$$

$$\Rightarrow \quad 54.12 = A + \frac{-44}{100} \times 2$$

$$\Rightarrow \quad 54.12 = A - 0.88$$

$$\Rightarrow \quad A = 55$$

$$\text{Thus } x_i = A + Cd_i$$

$$\Rightarrow \quad x_i = 55 + 2d_i$$

Table 15.2.9

Computation of the Original Frequency Distribution

d_i	f_i	$x_i = 55 + 2d_i$	Class interval $x_i \pm \frac{C}{2}$
–2	17	51	50-52
–1	35	53	52-54
0	28	55	54-56
1	15	57	56-58
2	5	59	58-60

Example 15.2.23: Compute coefficient of variation from the following data:

Age	:	under 10	under 20	under 30	under 40	under 50	under 60
No. of persons Dying	:	10	18	30	45	60	80

Solution: What is given in this problem is less than cumulative frequency distribution. We need first convert it to a frequency distribution and then compute the coefficient of variation.

Table 15.2.10

Computation of coefficient of variation

Age in years class Interval	No. of persons dying (f_i)	Mid-value (x_i)	$d_i = \frac{x_i - 25}{10}$	$f_i d_i$	$f_i d_i^2$
0-10	10	5	–2	–20	40
10-20	18–10= 8	15	–1	–8	8
20-30	30–18=12	25	0	0	0
30-40	45–30=15	35	1	15	15
40-50	60–45=15	45	2	30	60
50-60	80–60=20	55	3	60	180
Total	80	–	–	77	303

The AM is given by:

$$\bar{x} = A + \frac{\sum f_i d_i}{N} \times C$$

$$= \left(25 + \frac{77}{80} \times 10\right) \text{ years}$$

$$= 34.63 \text{ years}$$

The standard deviation is

$$s = \sqrt{\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2} \times C$$

$$= \sqrt{\frac{303}{80} - \left(\frac{77}{80}\right)^2} \times 10 \text{ years}$$

$= \sqrt{3.79 - 0.93} \times 10$ years

$= 16.91$ years

Thus the coefficient of variation is given by

$$CV = \frac{S}{X} \times 100$$

$$= \frac{16.91}{34.63} \times 100$$

$= 48.83$

Example 15.2.24: You are given the distribution of wages in two factors A and B

Wages in ₹	:	100-200	200-300	300-400	400-500	500-600	600-700
No. of workers in A	:	8	12	17	10	2	1
No. of workers in B	:	6	18	25	12	2	2

State in which factory, the wages are more variable.

Solution:

As explained in example 11.36, we need compare the coefficient of variation of A(i.e. v_A) and of B (i.e v_B).

If $v_A > v_B$, then the wages of factory A would be more variable. Otherwise, the wages of factory B would be more variable where

$$V_A = 100 \times \frac{s_A}{\overline{x}_A} \quad \text{and} \quad V_B = 100 \times \frac{s_B}{\overline{x}_B}$$

Table 15.2.11

Computation of coefficient of variation of wages of Two Factories A and B

Wages in rupees (1)	Mid-value x (2)	d= (3)	No. of workers of A f_A (4)	No. of workers of B f_B (5)	f_Ad (6)=(3)×(4)	f_Ad^2 (7)=(3)×(6)	f_Bd (8)=(3)×(5)	f_Bd^2 (9)=(3)×(8)
100-200	150	–2	8	6	–16	32	–12	24
200-300	250	–1	12	18	–12	12	–18	18
300-400	350	0	17	25	0	0	0	0
400-500	450	1	10	12	10	10	12	12
500-600	550	2	2	2	4	8	4	8
600-700	650	3	1	2	3	9	6	18
Total	–	–	50	65	–11	71	– 8	80

For Factory A

$$\bar{x}_A = ₹\left(350 + \frac{-11}{50} \times 100\right) = ₹\,328$$

$$S_A = ₹\sqrt{\frac{71}{50} - \left(\frac{-11}{50}\right)^2} \times 100 = \text{Nu.}117.12$$

$$\therefore V_A = \frac{S_A}{\bar{x}_A} \times 100 = 35.71$$

For Factory B

$$\bar{x}_B = ₹\left(350 + \frac{-8}{65} \times 100\right) = ₹\,337.69$$

$$S_B = ₹\sqrt{\frac{80}{65} - \left(\frac{-8}{65}\right)^2} \times 100$$

$$= ₹\,110.25$$

$$\therefore V_B = \frac{110.25}{337.69} \times 100 = 32.65$$

As $V_A > V_B$, the wages for factory A is more variable.

SUMMARY

- All these merits of standard deviation make SD as the most widely and commonly used measure of dispersion.
- Range is the quickest to compute and as such, has its application in statistical quality control. However, range is based on only two observations and affected too much by the presence of extreme observation(s).
- Mean deviation is rigidly defined, based on all the observations and not much affected by sampling fluctuations. However, mean deviation is difficult to comprehend and its computation is also time consuming and laborious. Furthermore, unlike SD, mean deviation does not possess mathematical properties.
- Quartile deviation is also rigidly defined, easy to compute and not much affected by sampling fluctuations. The presence of extreme observations has no impact on quartile deviation since quartile deviation is based on the central fifty-percent of the observations. However, quartile deviation is not based on all the observations and it has no desirable mathematical properties. Nevertheless, quartile deviation is the best measure of dispersion for open-end classifications.

EXERCISE — UNIT-II

Set A

Write down the correct answers. Each question carries one mark.

1. Which of the following statements is correct?
   (a) Two distributions may have identical measures of central tendency and dispersion.
   (b) Two distributions may have the identical measures of central tendency but different measures of dispersion.
   (c) Two distributions may have the different measures of central tendency but identical measures of dispersion.
   (d) All the statements (a), (b) and (c).
2. Dispersion measures
   (a) The scatterness of a set of observations
   (b) The concentration of a set of observations
   (c) Both (a) and (b)
   (d) Neither (a) and (b).
3. When it comes to comparing two or more distributions we consider
   (a) Absolute measures of dispersion (b) Relative measures of dispersion
   (c) Both (a) and (b) (d) Either (a) or (b).
4. Which one is easier to compute?
   (a) Relative measures of dispersion (b) Absolute measures of dispersion
   (c) Both (a) and (b) (d) Range
5. Which one is an absolute measure of dispersion?
   (a) Range (b) Mean Deviation
   (c) Standard Deviation (d) All these measures
6. Which measure of dispersion is most usefull?
   (a) Standard deviation (b) Quartile deviation
   (c) Mean deviation (d) Range
7. Which measures of dispersions is not affected by the presence of extreme observations?
   (a) Range (b) Mean deviation
   (c) Standard deviation (d) Quartile deviation
8. Which measure of dispersion is based on the absolute deviations only?
   (a) Standard deviation (b) Mean deviation
   (c) Quartile deviation (d) Range

9. Which measure is based on only the central fifty percent of the observations?
   (a) Standard deviation (b) Quartile deviation
   (c) Mean deviation (d) All these measures
10. Which measure of dispersion is based on all the observations?
   (a) Mean deviation (b) Standard deviation
   (c) Quartile deviation (d) (a) and (b) but not (c)
11. The appropriate measure of dispersion for open-end classification is
   (a) Standard deviation (b) Mean deviation
   (c) Quartile deviation (d) All these measures.
12. The most commonly used measure of dispersion is
   (a) Range (b) Standard deviation
   (c) Coefficient of variation (d) Quartile deviation.
13. Which measure of dispersion has some desirable mathematical properties?
   (a) Standard deviation (b) Mean deviation
   (c) Quartile deviation (d) All these measures
14. If the profits of a company remains the same for the last ten months, then the standard deviation of profits for these ten months would be ?
   (a) Positive (b) Negative (c) Zero (d) (a) or (c)
15. Which measure of dispersion is considered for finding a pooled measure of dispersion after combining several groups?
   (a) Mean deviation (b) Standard deviation
   (c) Quartile deviation (d) Any of these
16. A shift of origin has no impact on
   (a) Range (b) Mean deviation
   (c) Standard deviation (d) All these and quartile deviation.
17. The range of 15, 12, 10, 9, 17, 20 is
   (a) 5 (b) 12 (c) 13 (d) 11.
18. The standard deviation of 10, 16, 10, 16, 10, 10, 16, 16 is
   (a) 4 (b) 6 (c) 3 (d) 0.
19. For any two numbers SD is always
   (a) Twice the range (b) Half of the range
   (c) Square of the range (d) None of these.

20. If all the observations are increased by 10, then
    (a) SD would be increased by 10
    (b) Mean deviation would be increased by 10
    (c) Quartile deviation would be increased by 10
    (d) All these three remain unchanged.

21. If all the observations are multiplied by 2, then
    (a) New SD would be also multiplied by 2
    (b) New SD would be half of the previous SD
    (c) New SD would be increased by 2
    (d) New SD would be decreased by 2.

Set B

Write down the correct answers. Each question carries two marks.

1. What is the coefficient of range for the following wages of 8 workers?
   ₹ 80, ₹ 65, ₹ 90, ₹ 60, ₹ 75, ₹ 70, ₹ 72, ₹ 85.
   (a) ₹ 30 (b) ₹ 20 (c) 30 (d) 20

2. If R_x and R_y denote ranges of x and y respectively where x and y are related by 3x+2y+10=0, what would be the relation between x and y?
   (a) $R_x = R_y$ (b) $2 R_x = 3 R_y$ (c) $3 R_x = 2 R_y$ (d) $R_x = 2 R_y$

3. What is the coefficient of range for the following distribution?

Class Interval :	10-19	20-29	30-39	40-49	50-59
Frequency:	11	25	16	7	3

   (a) 22 (b) 50 (c) 72.46 (d) 75.82

4. If the range of x is 2, what would be the range of –3x +50 ?
   (a) 2 (b) 6 (c) –6 (d) 44

5. What is the value of mean deviation about mean for the following numbers?
   5, 8, 6, 3, 4.
   (a) 5.20 (b) 7.20 (c) 1.44 (d) 2.23

6. What is the value of mean deviation about mean for the following observations?
   50, 60, 50, 50, 60, 60, 60, 50, 50, 50, 60, 60, 60, 50.
   (a) 5 (b) 7 (c) 35 (d) 10

7. The coefficient of mean deviation about mean for the first 9 natural numbers is
   (a) 200/9 (b) 80 (c) 400/9 (d) 50.

8. If the relation between x and y is 5y–3x = 10 and the mean deviation about mean for x is 12, then the mean deviation of y about mean is

(a) 7.20 (b) 6.80 (c) 20 (d) 18.80.

9. If two variables x and y are related by 2x + 3y –7 =0 and the mean and mean deviation about mean of x are 1 and 0.3 respectively, then the coefficient of mean deviation of y about its mean is

(a) –5 (b) 12 (c) 50 (d) 4.

10. The mean deviation about mode for the numbers 4/11, 6/11, 8/11, 9/11, 12/11, 8/11 is
(a) 1/6 (b) 1/11 (c) 6/11 (d) 5/11.

11. What is the standard deviation of 5, 5, 9, 9, 9, 10, 5, 10, 10?

(a) $\sqrt{14}$ (b) $\frac{\sqrt{42}}{3}$ (c) 4.50 (d) 8

12. If the mean and SD of x are a and b respectively, then the SD of $\frac{x-a}{b}$ is

(a) –1 (b) 1 (c) ab (d) a/b.

13. What is the coefficient of variation of the following numbers?
53, 52, 61, 60, 64.
(a) 8.09 (b) 18.08 (c) 20.23 (d) 20.45

14. If the SD of x is 3, what us the variance of (5–2x)?
(a) 36 (b) 6 (c) 1 (d) 9

15. If x and y are related by 2x+3y+4 = 0 and SD of x is 6, then SD of y is
(a) 22 (b) 4 (c) $\sqrt{5}$ (d) 9.

16. The quartiles of a variable are 45, 52 and 65 respectively. Its quartile deviation is

(a) 10 (b) 20 (c) 25 (d) 8.30.

17. If x and y are related as 3x+4y = 20 and the quartile deviation of x is 12, then the quartile deviation of y is

(a) 16 (b) 14 (c) 10 (d) 9.

18. If the SD of the 1st n natural numbers is 2, then the value of n must be

(a) 2 (b) 7 (c) 6 (d) 5.

19. If x and y are related by y = 2x+ 5 and the SD and AM of x are known to be 5 and 10 respectively, then the coefficient of variation is

(a) 25 (b) 30 (c) 40 (d) 20.

20. The mean and SD for a, b and 2 are 3 and $\frac{2}{\sqrt{3}}$ respectively, The value of ab would be

(a) 5 (b) 6 (c) 12 (d) 3.

Set C

Write down the correct answer. Each question carries 5 marks.

1. What is the mean deviation about mean for the following distribution?

Variable:	5	10	15	20	25	30
Frequency:	3	4	6	5	3	2

(a) 6.00 (b) 5.93 (c) 6.07 (d) 7.20

2. What is the mean deviation about median for the following data?

X :	3	5	7	9	11	13	15
F :	2	8	9	16	14	7	4

(a) 2.50 (b) 2.46 (c) 2.43 (d) 2.37

3. What is the coefficient of mean deviation for the following distribution of heights? Take deviation from AM.

Height in inches:	60-62	63-65	66-68	69-71	72-74
No. of students:	5	22	28	17	3

(a) 2.30 inches (b) 3.45 inches (c) 3.82 inches (d) 2.48 inches

4. The mean deviation of weights about median for the following data:

Weight (lb) :	131-140	141-150	151-160	161-170	171-180	181-190
No. of persons :	3	8	13	15	6	5

Is given by

(a) 10.97 (b) 8.23 (c) 9.63 (d) 11.45.

5. What is the standard deviation from the following data relating to the age distribution of 200 persons?

Age (year) :	20	30	40	50	60	70	80
No. of people:	13	28	31	46	39	23	20

(a) 15.29 (b) 16.87 (c) 18.00 (d) 17.52

6. What is the coefficient of variation for the following distribution of wages?

Daily Wages (₹):	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
No. of workers	17	28	21	15	13	6

(a) ₹ 14.73 (b) 14.73 (c) 26.93 (d) 20.82

7. Which of the following companies A and B is more consistent so far as the payment of dividend is concerned ?

Dividend paid by A :	5	9	6	12	15	10	8	10
Dividend paid by B :	4	8	7	15	18	9	6	6

(a) A (b) B (c) Both (a) and (b) (d) Neither (a) nor (b)

8. The mean and SD for a group of 100 observations are 65 and 7.03 respectively. If 60 of these observations have mean and SD as 70 and 3 respectively, what is the SD for the group comprising 40 observations?

   (a) 16 (b) 25 (c) 4 (d) 2

9. If two samples of sizes 30 and 20 have means as 55 and 60 and variances as 16 and 25 respectively, then what would be the SD of the combined sample of size 50?

   (a) 5.00 (b) 5.06 (c) 5.23 (d) 5.35

10. The mean and SD of a sample of 100 observations were calculated as 40 and 5.1 respectively by a CA student who took one of the observations as 50 instead of 40 by mistake. The current value of SD would be

    (a) 4.90 (b) 5.00 (c) 4.88 (d) 4.85.

11. The value of appropriate measure of dispersion for the following distribution of daily wages

Wages (₹):	Below 30	30-39	40-49	50-59	60-79	Above 80
No. of workers	5	7	18	32	28	10

is given by

(a) ₹ 11.03 (b) ₹ 10.50 (c) 11.68 (d) ₹ 11.68.

UNIT-II: ANSWERS

Set A

1. (d)	**2.** (a)	**3.** (b)	**4.** (d)	**5.** (d)	**6.** (a)
7. (d)	**8.** (b)	**9.** (b)	**10.** (d)	**11.** (c)	**12.** (b)
13. (a)	**14.** (c)	**15.** (b)	**16.** (d)	**17.** (d)	**18.** (c)
19. (b)	**20.** (d)	**21.** (a)			

Set B

1. (d)	**2.** (c)	**3.** (c)	**4.** (b)	**5.** (c)	**6.** (a)
7. (c)	**8.** (a)	**9.** (b)	**10.** (b)	**11.** (b)	**12.** (b)
13. (a)	**14.** (a)	**15.** (b)	**16.** (a)	**17.** (d)	**18.** (b)
19. (c)	**20.** (a)				

Set C

1. (c)	**2.** (d)	**3.** (b)	**4.** (a)	**5.** (b)	**6.** (c)
7. (a)	**8.** (c)	**9.** (b)	**10.** (b)	**11.** (a)	

ADDITIONAL QUESTION BANK

1. The number of measures of central tendency is
   (a) two (b) three (c) four (d) five
2. The words "mean" or "average" only refer to
   (a) A.M (b) G.M (c) H.M (d) none
3. ———————— is the most stable of all the measures of central tendency.
   (a) G.M (b) H.M (c) A.M (d) none.
4. Mean is of ——— types.
   (a) 3 (b) 4 (c) 8 (d) 5
5. Weighted A.M is related to
   (a) G.M (b) frequency (c) H.M (d) none.
6. Frequencies are also called weights.
   (a) True (b) false (c) both (d) none
7. The algebraic sum of deviations of observations from their A.M is
   (a) 2 (b) -1 (c) 1 (d) 0
8. G.M of a set of n observations is the ——— root of their product.
   (a) n/2 th (b) (n+1)th (c) nth (d) (n -1)th
9. The algebraic sum of deviations of 8, 1, 6 from the A.M viz.5 is
   (a) -1 (b) 0 (c) 1 (d) none
10. G.M of 8, 4,2 is
    (a) 4 (b) 2 (c) 8 (d) none
11. ———————— is the reciprocal of the A.M of reciprocal of observations.
    (a) H.M (b) G.M (c) both (d) none
12. A.M is never less than G.M
    (a) True (b) false (c) both (d) none
13. G.M is less than H.M
    (a) true (b) false (c) both (d) none
14. The value of the middlemost item when they are arranged in order of magnitude is called
    (a) standard deviation (b) mean (c) mode (d) median
15. Median is unaffected by extreme values.
    (a) true (b) false (c) both (d) none

16. Median of 2, 5, 8, 4, 9, 6, 71 is

(a) 9 (b) 8 (c) 5 (d) 6

17. The value which occurs with the maximum frequency is called

(a) median (b) mode (c) mean (d) none

18. In the formula Mode = $L_1 + (d_1 \text{ X } c)/ (d_1 + d_2)$

d_1 is the difference of frequencies in the modal class & the ———— class.

(a) preceding (b) following (c) both (d) none

19. In the formula Mode = $L_1 + (d_1 \text{ X } c)/ (d_1 + d_2)$

d_2 is the difference of frequencies in the modal class & the ———— class.

(a) preceding (b) succeeding (c) both (d) none

20. In formula of median for grouped frequency distribution N is

(a) total frequency (b) frequency density
(c) frequency (d) cumulative frequency

21. When all observations occur with equal frequency ———— does not exit.

(a) median (b) mode (c) mean (d) none

22. Mode of the observations 2, 5, 8, 4, 3, 4, 4, 5, 2, 4, 4 is

(a) 3 (b) 2 (c) 5 (d) 4

23. For the observations 5, 3, 6, 3, 5, 10, 7, 2 there are ———— modes.

(a) 2 (b) 3 (c) 4 (d) 5

24. ———— of a set of observations is defined to be their sum, divided by the no. of observations.

(a) H.M (b) G.M (c) A.M (d) none

25. Simple average is sometimes called

(a) weighted average (b) unweighted average
(c) relative average (d) none

26. When a frequency distribution is given, the frequencies themselves treated as weights.

(a) True (b) false (c) both (d) none

27. Each value is considered only once for

(a) simple average (b) weighted average
(c) both (d) none

28. Each value is considered as many times as it occurs for

(a) simple average (b) weighted average
(c) both (d) none

29. Multiplying the values of the variable by the corresponding weights and then dividing the sum of products by the sum of weights is

(a) simple average (b) weighted average
(c) both (d) none

30. Simple & weighted average are equal only when all the weights are equal.

(a) True (b) false (c) both (d) none

31. The word "average " used in "simple average" and "weighted average" generally refers to

(a) median (b) mode (c) A.M , G.M or H.M (d) none

32. ———— average is obtained on dividing the total of a set of observations by their number

(a) simple (b) weighted (c) both (d) none

33. Frequencies are generally used as

(a) range (b) weights (c) mean (d) none

34. The total of a set of observations is equal to the product of their number of observations and the

(a) A.M (b) G.M (c) H.M (d) none

35. The total of the deviations of a set of observations from their A.M is always

(a) 0 (b) 1 (c) -1 (d) none

36. Deviation may be positive or negative or zero

(a) true (b) false (c) both (d) none

37. The sum of the squares of deviations of a set of observations has the smallest value, when the deviations are taken from their

(a) A.M (b) H.M (c) G.M (d) none

38. For a given set of positive observations H.M is less than G.M

(a) true (b) false (c) both (d) none

39. For a given set of positive observations A.M is greater than G.M

(a) true (b) false (c) both (d) none

40. Calculation of G.M is more difficult than

(a) A.M (b) H.M (c) median (d) none

41. ———— has a limited use

(a) A.M (b) G.M (c) H.M (d) none

42. A.M of 8, 1, 6 is

(a) 5 (b) 6 (c) 4 (d) none

43. ———— can be calculated from a frequency distribution with open end intervals

(a) Median (b) Mean (c) Mode (d) none

44. The values of all items are taken into consideration in the calculation of

(a) median (b) mean (c) mode (d) none

45. The values of extreme items do not influence the average in case of

(a) median (b) mean (c) mode (d) none

46. In a distribution with a single peak and moderate skewness to the right, it is closer to the concentration of the distribution in case of

(a) mean (b) median (c) both (d) none

47. If the variables x & z are so related that z = ax + b for each $x = x_i$ where a & b are constants, then $\bar{z} = a\bar{x} + b$

(a) true (b) false (c) both (d) none

48. G.M is defined only when

(a) all observations have the same sign and none is zero

(b) all observations have the different sign and none is zero

(c) all observations have the same sign and one is zero

(d) all observations have the different sign and one is zero

49. ———— is useful in averaging ratios, rates and percentages.

(a) A.M (b) G.M (c) H.M (d) none

50. G.M is useful in construction of index number.

(a) true (b) false (c) both (d) none

51. More laborious numerical calculations involves in G.M than A.M

(a) True (b) false (c) both (d) none

52. H.M is defined when no observation is

(a) 3 (b) 2 (c) 1 (d) 0

53. When all values occur with equal frequency, there is no

(a) mode (b) mean (c) median (d) none

54. ———— cannot be treated algebraically

(a) mode (b) mean (c) median (d) none

55. For the calculation of ———— , the data must be arranged in the form of a frequency distribution.

(a) median (b) mode (c) mean (d) none

56. ———— is equal to the value corresponding to cumulative frequency

(a) mode (b) mean (c) median (d) none

57. ————— is the value of the variable corresponding to the highest frequency

(a) mode (b) mean (c) median (d) none

58. The class in which mode belongs is known as

(a) median class (b) mean class (c) modal class (d) none

59. The formula of mode is applicable if classes are of ————— width.

(a) equal (b) unequal (c) both (d) none

60. For calculation of ——— we have to construct cumulative frequency distribution

(a) mode (b) median (c) mean (d) none

61. For calculation of ——— we have to construct a grouped frequency distribution

(a) median (b) mode (c) mean (d) none

62. Relation between mean, median & mode is

(a) mean - mode = 2 (mean - median)
(b) mean - median = 3 (mean - mode)
(c) mean - median = 2 (mean - mode)
(d) mean - mode = 3 (mean - median)

63. When the distribution is symmetrical, mean, median and mode

(a) coincide (b) do not coincide (c) both (d) none

64. Mean, median & mode are equal for the

(a) Binomial distribution
(b) Normal distribution
(c) both
(d) none

65. In most frequency distributions, it has been observed that the three measures of central tendency viz. mean, median & mode, obey the approximate relation, provided the distribution is

(a) very skew (b) not very skew (c) both (d) none

66. —————— divides the total number of observations into two equal parts.

(a) mode (b) mean (c) median (d) none

67. Measures which are used to divide or partition the observations into a fixed number of parts are collectively known as

(a) partition values (b) quartiles (c) both (d) none

68. The middle most value of a set of observations is

(a) median (b) mode (c) mean (d) none

69. The number of observations smaller than ——— is the same as the number larger than it.

(a) median (b) mode (c) mean (d) none

70. ———— is the value of the variable corresponding to cumulative frequency N /2

(a) mode (b) mean (c) median (d) none

71. ———————— divide the total no. observations into 4 equal parts.

(a) median (b) deciles (c) quartiles (d) percentiles

72. ———————— quartile is known as Upper quartile

(a) First (b) Second (c) Third (d) none

73. Lower quartile is

(a) first quartile (b) second quartile (c) upper quartile (d) none

74. The number of observations smaller than lower quartile is the same as the no. lying between lower and middle quartile.

(a) true (b) false (c) both (d) none

75. ———— are used for measuring central tendency, dispersion & skewness.

(a) Median (b) Deciles (c) Percentiles (d) Quartiles.

76. The second quartile is known as

(a) median (b) lower quartile (c) upper quartile (d) none

77. The lower & upper quartiles are used to define

(a) standard deviation (b) quartile deviation
(c) both (d) none

78. Three quartiles are used in

(a) Pearson's formula (b) Bowley's formula
(c) both (d) none

79. Less than First quartile, the frequency is equal to

(a) N /4 (b) 3N /4 (c) N /2 (d) none

80. Between first & second quartile, the frequency is equal to

(a) 3N/4 (b) N /2 (c) N /4 (d) none

81. Between second & upper quartile, the frequency is equal to

(a) 3N/4 (b) N /4 (c) N /2 (d) none

82. Above upper quartile, the frequency is equal to

(a) N /4 (b) N /2 (c) 3N /4 (d) none

83. Corresponding to first quartile, the cumulative frequency is

(a) N /2 (b) N / 4 (c) 3N /4 (d) none

84. Corresponding to second quartile, the cumulative frequency is

(a) N/4 (b) 2 N/4 (c) 3N/4 (d) none

85. Corresponding to upper quartile, the cumulative frequency is

(a) 3N/4 (b) N/4 (c) 2N/4 (d) none

86. The values which divide the total number of observations into 10 equal parts are

(a) quartiles (b) percentiles (c) deciles (d) none

87. There are ———— deciles.

(a) 7 (b) 8 (c) 9 (d) 10

88. Corresponding to first decile, the cumulative frequency is

(a) N/10 (b) 2N/10 (c) 9N/10 (d) none

89. Fifth decile is equal to

(a) mode (b) median (c) mean (d) none

90. The values which divide the total number of observations into 100 equal parts is

(a) percentiles (b) quartiles (c) deciles (d) none

91. Corresponding to second decile, the cumulative frequency is

(a) N/10 (b) 2N/10 (c) 5N/10 (d) none

92. There are ———— percentiles.

(a) 100 (b) 98 (c) 97 (d) 99

93. 10^{th} percentile is equal to

(a) 1^{st} decile (b) 10^{th} decile (c) 9^{th} decile (d) none

94. 50^{th} percentile is known as

(a) 50^{th} decile (b) 50^{th} quartile (c) mode (d) median

95. 20^{th} percentile is equal to

(a) 19^{th} decile (b) 20^{th} decile (c) 2^{nd} decile (d) none

96. (3^{rd} quartile —— 1^{st} quartile)/2 is

(a) skewness (b) median (c) quartile deviation (d) none

97. 1^{st} percentile is less than 2^{nd} percentile.

(a) true (b) false (c) both (d) none

98. 25^{th} percentile is equal to

(a) 1^{st} quartile (b) 25^{th}quartile (c) 24^{th} quartile (d) none

99. 90^{th} percentile is equal to

(a) 9^{th} quartile (b) 90^{th} decile (c) 9^{th} decile (d) none

100. 1st decile is greater than 2nd decile

(a) True (b) false (c) both (d) none

101. Quartile deviation is a measure of dispersion.

(a) true (b) false (c) both (d) none

102. To define quartile deviation we use

(a) lower & middle quartiles (b) lower & upper quartiles
(c) upper & middle quartiles (d) none

103. Calculation of quartiles, deciles ,percentiles may be obtained graphically from

(a) Frequency Polygon (b) Histogram (c) Ogive (d) none

104. 7th decile is the abscissa of that point on the Ogive whose ordinate is

(a) 7N/10 (b) 8N /10 (c) 6N /10 (d) none

105. Rank of median is

(a) (n+ 1)/2 (b) (n+ 1)/4 (c) 3(n + 1)/4 (d) none

106. Rank of 1st quartile is

(a) (n+ 1)/2 (b) (n+ 1)/4 (c) 3(n + 1)/4 (d) none

107. Rank of 3rd quartile is

(a) 3(n+ 1)/4 (b) (n+ 1)/4 (c) (n + 1)/2 (d) none

108. Rank of k th decile is

(a) (n+ 1)/2 (b) (n+ 1)/4 (c) (n + 1)/10 (d) k(n +1)/10

109. Rank of k th percentile is

(a) (n+ 1)/100 (b) k(n+ 1)/10 (c) k(n + 1)/100 (d) none

110. ———— is equal to value corresponding to cumulative frequency (N + 1)/2 from simple frequency distribution

(a) Median (b) 1st quartile (c) 3rd quartile (d) 4th quartile

111. ——— is equal to the value corresponding to cumulative frequency (N + 1)/4 from simple frequency distribution

(a) Median (b) 1st quartile (c) 3rd quartile (d) 1st decile

112. ——— is equal to the value corresponding to cumulative frequency 3 (N + 1)/4 from simple frequency distribution

(a) Median (b) 1st quartile (c) 3rd quartile (d) 1st decile

113. ——— is equal to the value corresponding to cumulative frequency k (N + 1)/10 from simple frequency distribution

(a) Median (b) kth decile (c) kth percentile (d) none

114. ———— is equal to the value corresponding to cumulative frequency k(N + 1)/100 from simple frequency distribution

(a) k^{th} decile (b) k^{th} percentile (c) both (d) none

115. For grouped frequency distribution ———— is equal to the value corresponding to cumulative frequency N /2

(a) median (b) 1^{st} quartile (c) 3^{rd} quartile (d) none

116. For grouped frequency distribution ———— is equal to the value corresponding to cumulative frequency N /4

(a) median (b) 1^{st} quartile (c) 3^{rd} quartile (d) none

117. For grouped frequency distribution ———— is equal to the value corresponding to cumulative frequency 3N /4

(a) median (b) 1^{st} quartile (c) 3^{rd} quartile (d) none

118. For grouped frequency distribution ———— is equal to the value corresponding to cumulative frequency kN/10

(a) median (b) kth percentile (c) kth decile (d) none

119. For grouped frequency distribution ———— is equal to the value corresponding to cumulative frequency kN /100

(a) k^{th} quartile (b) k^{th} percentile (c) k^{th} decile (d) none

120. In Ogive, abscissa corresponding to ordinate N/2 is

(a) median (b) 1^{st} quartile (c) 3^{rd} quartile (d) none

121. In Ogive, abscissa corresponding to ordinate N/4 is

(a) median (b) 1^{st} quartile (c) 3^{rd} quartile (d) none

122. In Ogive, abscissa corresponding to ordinate 3N/4 is

(a) median (b) 3^{rd} quartile (c) 1^{st} quartile (d) none

123. In Ogive, abscissa corresponding to ordinate ———— is kth decile.

(a) kN/10 (b) kN/100 (c) kN/50 (d) none

124. In Ogive , abscissa corresponding to ordinate ———— is kth percentile.

(a) kN/10 (b) kN/100 (c) kN/50 (d) none

125. For 899, 999, 391, 384, 590, 480, 485, 760, 111, 240
Rank of median is

(a) 2.75 (b) 5.5 (c) 8.25 (d) none

126. For 333, 999, 888, 777, 666, 555, 444
Rank of 1^{st} quartile is

(a) 3 (b) 1 (c) 2 (d) 7

127. For 333, 999, 888, 777, 1000, 321, 133
Rank of 3rd quartile is

(a) 7 (b) 4 (c) 5 (d) 6

128. Price per kg.(₹) : 45 50 35; Kgs.Purchased : 100 40 60 Total frequency is

(a) 300 (b) 100 (c) 150 (d) 200

129. The length of a rod is measured by a tape 10 times. You are to estimate the length of the rod by averaging these 10 determinations.

What is the suitable form of average in this case?

(a) A.M (b) G.M (c) H.M (d) none

130. A person purchases 5 rupees worth of eggs from 10 different markets.You are to find the average no. of eggs per rupee for all the markets taken together. What is the suitable form of average in this case?

(a) A.M (b) G.M (c) H.M (d) none

131. You are given the population of India for the courses of 1981 & 1991. You are to find the population of India at the middle of the period by averaging these population figures, assuming a constant rate of increase of population.

What is the suitable form of average in this case?

(a) A.M (b) G.M (c) H.M (d) none

132. ____________ is least affected by sampling fluctions.

(a) Standard deviation (b) Quartile deviation
(c) both (d) none

133. "Root –Mean Square Deviation from Mean" is

(a) Standard deviation (b) Quartile deviation

(c) both (d) none

134. Standard Deviation is

(a) absolute measure (b) relative measure (c) both (d) none

135. Coefficient of variation is

(a) absolute measure (b) relative measure (c) both (d) none

136. ____________ deviation is called semi-interquartile range.

(a) Percentile (b) Standard (c) Quartile (d) none

137. ________________ Deviation is defined as half the difference between the lower & upper quartiles.

(a) Quartile (b) Standard (c) both (d) none

138. Quartile Deviation for the data 1, 3, 4, 5, 6, 6, 10 is

(a) 3 (b) 1 (c) 6 (d) 1.5

139. Coefficient of Quartile Deviation is

(a) (Quartile Deviation x 100)/Median (b) (Quartile Deviation x 100)/Mean
(c) (Quartile Deviation x 100) /Mode (d) none

140. Mean for the data 6, 4, 1, 6, 5, 10, 3 is

(a) 7 (b) 5 (c) 6 (d) none

141. Coefficient of variation = (Standard Deviation x 100)/Mean

(a) true (b) false (c) both (d) none

142. If mean = 5, Standard deviation = 2.6 then the coefficient of variation is

(a) 49 (b) 51 (c) 50 (d) 52

143. If median = 5, Quartile deviation = 1. 5 then the coefficient of quartile deviation is

(a) 33 (b) 35 (c) 30 (d) 20

144. A.M of 2, 6, 4, 1, 8, 5, 2 is

(a) 4 (b) 3 (c) 4 (d) none

145. Most useful among all measures of dispersion is

(a) S.D (b) Q.D (c) Mean deviation (d) none

146. For the observations 6, 4, 1, 6, 5, 10, 4, 8 Range is

(a) 10 (b) 9 (c) 8 (d) none

147. A measure of central tendency tries to estimate the

(a) central value (b) lower value (c) upper value (d) none

148. Measures of central tendency are known as

(a) differences (b) averages (c) both (d) none

149. Mean is influenced by extreme values.

(a) true (b) false (c) both (d) none

150. Mean of 6, 7, 11, 8 is

(a) 11 (b) 6 (c) 7 (d) 8

151. The sum of differences between the actual values and the arithmetic mean is

(a) 2 (b) -1 (c) 0 (d) 1

152. When the algebraic sum of deviations from the arithmetic mean is not equal to zero, the figure of arithmetic mean ———————— correct.

(a) is (b) is not (c) both (d) none

153. In the problem

No. of shirts:	30–32	33–35	36–38	39–41	42–44
No. of persons:	15	14	42	27	18

The assumed mean is

(a) 34 (b) 37 (c) 40 (d) 43

154. In the problem

Size of items:	1–3	3–8	8–15	15–26
Frequency:	5	10	16	15

The assumed mean is

(a) 20.5 (b) 2 (c) 11.5 (d) 5.5

155. The average of a series of overlapping averages, each of which is based on a certain number of item within a series is known as

(a) moving average (b) weighted average
(c) simple average (d) none

156. ———— averages is used for smoothening a time series.

(a) moving average (b) weighted average
(c) simple average (d) none

157. Pooled Mean is also called

(a) Mean (b) Geometric Mean (c) Grouped Mean (d) none

158. Half of the numbers in an ordered set have values less than the ———— and half will have values greater than the ———— .

(a) mean, median (b)median, median (c) mode, mean (d) none.

159. The median of 27, 30, 26, 44, 42, 51, 37 is

(a) 30 (b) 42 (c) 44 (d) 37

160. For an even number of values the median is the

(a) average of two middle values (b) middle value
(c) both (d) none

161. In the case of a continuous frequency distribution, the size of the ———— item indicates class interval in which the median lies.

(a) $(n-1)/2^{th}$ (b) $(n+1)/2^{th}$ (c) $n/2^{th}$ (d) none

162. The deviations from median are ———— if negative signs are ignored as compared to other measures of central tendency.

(a) minimum (b) maximum (c) same (d) none

163. Ninth Decile lies in the class interval of the item

(a) n/9 (b) 9n/10 (c) 9n/20 (d) none item.

164. Ninety Ninth Percentile lies in the class interval of the item

(a) 99n/100 (b) 99n/10 (c) 99n/200 (d) none item.

165. ———— is the value of the variable at which the concentration of observation is the densest.

(a) mean (b) median (c) mode (d) none

166. Height in cms: 60–62 63–65 66–68 69–71 72–74

Height in cms:	60–62	63–65	66–68	69–71	72–74
No. of students:	15	118	142	127	18

Modal group is

(a) 66–68 (b) 69–71 (c) 63–65 (d) none

167. A distribution is said to be symmetrical when the frequency rises & falls from the highest value in the ———— proportion.

(a) unequal (b) equal (c) both (d) none

168. ———— always lies in between the arithmetic mean & mode.

(a) G.M (b) H.M (c) Median (d) none

169. Logarithm of G.M is the ———— of logarithms of the different values.

(a) weighted mean (b) simple mean (c) both (d) none

170. ———— is not much affected by fluctuations of sampling.

(a) A.M (b) G.M (c) H.M (d) none

171. The data 1, 2, 4, 8, 16 are in

(a) Arithmetic progression (b) Geometric progression

(c) Harmonic progression (d) none

172. ———— & ———— can not be calculated if any observation is zero.

(a) G.M & A.M (b) H.M & A.M (c) H.M & G. M (d) None.

173. ———— & ———— are called ratio averages.

(a) H.M & G.M (b) H. M & A.M (c) A.M & G.M (d) none

174. ———— is a good substitute to a weighted average.

(a) A.M (b) G.M (c) H.M (d) none

175. For ordering shoes of various sizes for resale, a ———— size will be more appropriate.

(a) median (b) modal (c) mean (d) none

176. ———— is called a positional measure.

(a) mean (b) mode (c) median (d) none

177. 50% of actual values will be below & 50% of will be above ————

(a) mode (b) median (c) mean (d) none

178. Extreme values have ———— effect on mode.

(a) high (b) low (c) no (d) none

179. Extreme values have ———— effect on median.

(a) high (b) low (c) no (d) none

180. Extreme values have ———— effect on A.M.

(a) greatest (b) least (c) some (d) none

181. Extreme values have ———— effect on H.M.

(a) least (b) greatest (c) medium (d) none

182. ———— is used when representation value is required & distribution is asymmetric.

(a) mode (b) mean (c) median (d) none

183. ———— is used when most frequently occurring value is required (discrete variables).

(a) mode (b) mean (c) median (d) none

184. ———— is used when rate of growth or decline required.

(a) mode (b) A.M (c) G.M (d) none

185. In finding ————, the distribution has open-end classes.

(a) median (b) mean (c) standard deviation (d) none

186. The cumulative frequency distribution is used for

(a) median (b) mode (c) mean (d) none

187. In ——— the quantities are in ratios.

(a) A.M (b) G.M (c) H.M (d) none

188. ———— is used when variability has also to be calculated.

(a) A.M (b) G.M (c) H.M (d) none

189. ———— is used when the sum of absolute deviations from the average should be least.

(a) Mean (b) Mode (c) Median (d) None

190. ———— is used when sampling variability should be least.

(a) Mode (b) Median (c) Mean (d) none

191. ———— is used when distribution pattern has to be studied at varying levels.

(a) A.M (b) Median (c) G.M (d) none

192. The average discovers

(a) uniformity in variability (b) variability in uniformity of distribution
(c) both (d) none

193. The average has relevance for

(a) homogeneous population (b) heterogeneous population
(c) both (d) none

194. The correction factor is applied in

(a) inclusive type of distribution (b) exclusive type of distribution
(c) both (d) none

195. "Mean has the least sampling variability" prove the mathematical property of mean

(a) True (b) false (c) both (d) none

196. "The sum of deviations from the mean is zero" —— is the mathematical property of mean

(a) True (b) false (c) both (d) none

197. "The mean of the two samples can be combined" — is the mathematical property of mean

(a) True (b) false (c) both (d) none

198. "Choices of assumed mean does not affect the actual mean"— prove the mathematical property of mean

(a) True (b) false (c) both (d) none

199. "In a moderately asymmetric distribution mean can be found out from the given values of median & mode"— is the mathematical property of mean

(a) True (b) false (c) both (d) none

200. The mean wages of two companies are equal. It signifies that the workers of both the companies are equally well-off.

(a) True (b) false (c) both (d) none

201. The mean wage in factory A is Nu.6,000 whereas in factory B it is Nu.5,500. It signifies that factory A pays more to all its workers than factory B.

(a) True (b) false (c) both (d) none

202. Mean of 0, 3, 5, 6, 7, 9, 12, 0, 2 is

(a) 4.9 (b) 5.7 (c) 5.6 (d) none

203. Median of 15, 12, 6, 13, 12, 15, 8, 9 is

(a) 13 (b) 8 (c) 12 (d) 9

204. Median of 0.3, 5, 6, 7, 9, 12, 0, 2 is

(a) 7 (b) 6 (c) 3 (d) 5

205. Mode of 0, 3, 5, 6, 7, 9, 12, 0, 2 is

(a) 6 (b) 0 (c) 3 (d) 5

206. Mode 0f 15, 12, 5, 13, 12, 15, 8, 8, 9, 9, 10, 15 is

(a)15 (b) 12 (c) 8 (d) 9

207. Median of 40, 50, 30, 20, 25, 35, 30, 30, 20, 30 is

(a) 25 (b) 30 (c) 35 (d) none

208. Mode of 40, 50, 30, 20, 25, 35, 30, 30, 20, 30 is

(a) 25 (b) 30 (c) 35 (d) none

209. ———————— in particular helps in finding out the variability of the data.

(a) Dispersion (b) Median (c) Mode (d) None

210. Measures of central tendency are called averages of the ———order.

(a) 1^{st} (b) 2^{nd} (c) 3^{rd} (d) none

211. Measures of dispersion are called averages of the ———order.

(a) 1^{st} (b) 2^{nd} (c) 3^{rd} (d) none

212. In measuring dispersion, it is necessary to know the amount of ———— & the degree of —————.

(a) variation, variation (b) variation, median
(c) median, variation (d) none

213. The amount of variation is designated as ———————— measure of dispersion.

(a) relative (b) absolute (c) both (d) none

214. The degree of variation is designated as ———————— measure of dispersion.

(a) relative (b) absolute (c) both (d) none

215. For purposes of comparison between two or more series with varying size or no. of items, varying central values or units of calculation, only ———————— measures can be used.

(a) absolute (b) relative (c) both (d) none

216. The relation Relative range = Absolute range/Sum of the two extremes. is

(a) True (b) false (c) both (d) none

217. The relation Absolute range = Relative range/Sum of the two extremes is

(a) True (b) false (c) both (d) none

218. In quality control ———— is used as a substitute for standard deviation.

(a) mean deviation (b) median (c) range (d) none

219. ———————— factor helps to know the value of standard deviation.

(a) Correction (b) Range (c) both (d) none

220. ———————— is extremely sensitive to the size of the sample

(a) Range (b) Mean (c) Median (d) Mode

221. As the sample size increases, ———————— also tends to increase.

(a) Range (b) Mean (c) Median (d) Mode

222. As the sample size increases, range also tends to increase though not proportionately.

(a) true (b) false (c) both (d) none.

223. As the sample size increases, range also tends to

(a) decrease (b) increase (c) same (d) none

224. The dependence of range on extreme items can be avoided by adopting

(a) standard deviation (b) mean deviation (c) quartile deviation (d) none

225. Quartile deviation is called

(a) semi inter quartile range (b) quartile range (c) both (d) none

226. When 1st quartile = 20, 3rd quartile = 30, the value of quartile deviation is

(a) 7 (b) 4 (c) -5 (d) 5

227. $(Q_3 - Q_1)/(Q_3 + Q_1)$ is

(a) coefficient of Quartile Deviation (b) coefficient of Mean Deviation
(c) coefficient of Standard deviation (d) none

228. Standard deviation is denoted by

(a) σ^2 (b) σ (c) $\sqrt{\sigma}$ (d) none

229. The square of standard deviation is known as

(a) variance (b) standard deviation
(c) mean deviation (d) none

230. Mean of 25, 32, 43, 53, 62, 59, 48, 31, 24, 33 is

(a) 44 (b) 43 (c) 42 (d) 41

231. For the following frequency distribution

Class interval:	10–20	20–30	30–40	40–50	50–60	60–70
Frequency:	20	9	31	18	10	9

assumed mean can be taken as

(a) 55 (b) 45 (c) 35 (d) none

232. The value of the standard deviation does not depend upon the choice of the origin.

(a) True (b) false (c) both (d) none

233. Coefficient of standard deviation is

(a) S.D/Median (b) S.D/Mean (c) S.D/Mode (d) none

234. The value of the standard deviation will change if any one of the observations is changed.

(a). True (b) false (c) both (d) none

235. When all the values are equal then variance & standard deviation would be

(a) 2 (b) -1 (c) 1 (d) 0

236. For values lie close to the mean, the standard deviations are

(a) big (b) small (c) moderate (d) none

237. If the same amount is added to or subtracted from all the values, variance & standard deviation shall

(a) changed (b) unchanged (c) both (d) none

238. If the same amount is added to or subtracted from all the values, the mean shall increase or decrease by the ———— amount

(a) big (b) small (c) same (d) none

239. If all the values are multiplied by the same quantity, the ———— & ———— also would be multiple of the same quantity.

(a) mean, standard deviation (b) mean , median
(c) mean, mode (d) median , deviations

240. For a moderately non-symmetrical distribution, Mean deviation = 4/5 of standard deviation

(a) true (b) false (c) both (d) none

241. For a moderately non-symmetrical distribution, Quartile deviation = Standard deviation/3

(a) true (b) false (c) both (d) none

242. For a moderately non-symmetrical distribution, probable error of standard deviation = Standard deviation/3

(a) true (b) false (c) both (d) none

243. Quartile deviation = Probable error of Standard deviation.

(a) true (b) false (c) both (d) none

244. Coefficient of Mean Deviation is

(a) Mean deviation x 100/Mean or mode (b) Standard deviation x 100/Mean or median

(c) Mean deviation x 100/Mean or median (d) none

245. Coefficient of Quartile Deviation = Quartile Deviation x 100/Median

(a) true (b) false (c) both (d) none

246. Karl Pearson's measure gives

(a) coefficient of Mean Variation (b) coefficient of Standard deviation
(c) coefficient of variation (d) none

247. In ——— range has the greatest use.

(a) Time series (b) quality control (c) both (d) none

248. Mean is an absolute measure & standard deviation is based upon it. Therefore standard deviation is a relative measure.

(a) true (b) false (c) both (d) none

249. Semi-quartile range is one-fourth of the range in a normal symmetrical distribution.

(a) Yes (b) No (c) both (d) none

250. Whole frequency table is needed for the calculation of

(a) range (b) variance (c) both (d) none

251. Relative measures of dispersion make deviations in similar units comparable.

(a) true (b) false (c) both (d) none

252. Quartile deviation is based on the

(a) highest 50% (b) lowest 25%
(c) highest 25% (d) middle 50% of the item.

253. S.D is less than Mean deviation

(a) true (b) false (c) both (d) none

254. Coefficient of variation is independent of the unit of measurement.

(a) true (b) false (c) both (d) none

255. Coefficient of variation is a relative measure of

(a) mean (b) deviation (c) range (d) dispersion.

256. Coefficient of variation is equal to

(a) Standard deviation x 100 / median (b) Standard deviation x 100 / mode
(c) Standard deviation x 100 / mean (d) none

257. Coefficient of Quartile Deviation is equal to

(a) Quartile deviation x 100 / median (b) Quartile deviation x 100 / mean
(c) Quartile deviation x 100 / mode (d) none

258. If each item is reduced by 15 A.M is

(a) reduced by 15 (b) increased by 15 (c) reduced by 10 (d) none

259. If each item is reduced by 10, the range is

(a) increased by 10 (b) decreased by 10 (c) unchanged (d) none

260. If each item is reduced by 20, the standard deviation

(a) increased (b) decreased (c) unchanged (d) none

261. If the variables are increased or decreased by the same amount the standard deviation is

(a) decreased (b) increased (c) unchanged (d) none

262. If the variables are increased or decreased by the same proportion, the standard deviation changes by

(a) same proportion (b) different proportion (c) both (d) none

263. The mean of the 1st n natural no. is

(a) n/2 (b) (n-1)/2 (c) (n+1)/2 (d) none

264. If the class interval is open-end then it is difficult to find

(a) frequency (b) A.M (c) both (d) none

265. Which one is true—

(a) A.M = assumed mean + arithmetic mean of deviations of terms

(b) G.M = assumed mean + arithmetic mean of deviations of terms

(c) Both (d) none

266. If the A.M of any distribution be 25 & one term is 18. Then the deviation of 18 from A.M is

(a) 7 (b) -7 (c) 43 (d) none

267. For finding A.M in Step-deviation method, the class intervals should be of

(a) equal lengths (b) unequal lengths (c) maximum lengths (d) none

268. The sum of the squares of the deviations of the variable is ———— when taken about A.M

(a) maximum (b) zero (c) minimum (d) none

269. The A.M of 1, 3, 5, 6, x, 10 is 6 . The value of x is

(a) 10 (b) 11 (c) 12 (d) none

270. The G.M of 2 & 8 is

(a) 2 (b) 4 (c) 8 (d) none

271. (n+1)/2 th term is median if n is

(a) odd (b) even (c) both (d) none

272. For the values of a variable 5, 2, 8, 3, 7, 4, the median is

(a) 4 (b) 4.5 (c) 5 (d) none

273. The abscissa of the maximum frequency in the frequency curve is the

(a) mean (b) median (c) mode (d) none

274.

Variable:	2	3	4	5	6	7
No. of men:	5	6	8	13	7	4

Mode is

(a) 6 (b) 4 (c) 5 (d) none

275. The class having maximum frequency is called

(a) modal class (b) median class (c) mean class (d) none

276. For determination of mode, the class intervals should be

(a) overlapping (b) maximum (c) minimum (d) none

277. First Quartile lies in the class interval of the

(a) $n/2^{th}$ item (b) $n/4^{th}$ item (c) $3n/4^{th}$ item (d) $n/10^{th}$ item

278. The value of a variate that occur most often is called

(a) median (b) mean (c) mode (d) none

279. For the values of a variable 3, 1, 5, 2, 6, 8, 4 the median is

(a) 3 (b) 5 (c) 4 (d) none

280. If $y = 5x - 20$ & $\overline{x} = 30$ then the value of $\overline{y}$ is

(a) 130 (b) 140 (c) 30 (d) none

281. If $y = 3x - 100$ and $\overline{x} = 50$ then the value of $\overline{y}$ is

(a) 60 (b) 30 (c) 100 (d) 50

282. The median of the numbers 11, 10, 12, 13, 9 is

(a) 12.5 (b) 12 (c) 10.5 (d) 11

283. The mode of the numbers 7, 7, 7, 9, 10, 11, 11, 11, 12 is

(a) 11 (b) 12 (c) 7 (d) 7 & 11

284. In a symmetrical distribution when the 3rd quartile plus 1st quartile is halved, the value would give

(a) mean (b) mode (c) median (d) none

285. In Zoology ———————— is used.

(a) median (b) mean (c) mode (d) none

286. For calculation of Speed & Velocity

(a) G.M (b) A.M (c) H.M (d) none is used.

287. The S.D is always taken from

(a) median (b) mode (c) mean (d) none

288. Coefficient of Standard deviation is equal to

(a) S.D/A.M (b) A.M/S.D (c) S.D/GM (d) none

289. The distribution, for which the coefficient of variation is less, is ——— consistent.

(a) less (b) more (c) moderate (d) none

ANSWERS

1.	(b)	2.	(a)	3.	(c)	4.	(a)	5.	(b)
6.	(a)	7.	(d)	8.	(c)	9.	(b)	10.	(a)
11.	(a)	12.	(a)	13.	(b)	14.	(d)	15.	(a)
16.	(d)	17.	(b)	18.	(a)	19.	(b)	20.	(a)
21.	(b)	22.	(d)	23.	(a)	24.	(c)	25.	(b)
26.	(a)	27.	(a)	28.	(b)	29.	(b)	30.	(a)
31.	(c)	32.	(a)	33.	(b)	34.	(a)	35.	(a)
36.	(a)	37.	(a)	38.	(a)	39.	(a)	40.	(a)
41.	(c)	42.	(a)	43.	(a)	44.	(b)	45.	(a)
46.	(b)	47.	(a)	48.	(a)	49.	(b)	50.	(a)
51.	(a)	52.	(d)	53.	(a)	54.	(a)	55.	(b)
56.	(c)	57.	(a)	58.	(c)	59.	(c)	60.	(b)
61.	(b)	62.	(d)	63.	(a)	64.	(b)	65.	(b)
66.	(c)	67.	(c)	68.	(a)	69.	(a)	70.	(c)
71.	(c)	72.	(c)	73.	(a)	74.	(a)	75.	(d)
76.	(a)	77.	(b)	78.	(b)	79.	(a)	80.	(c)
81.	(b)	82.	(a)	83.	(b)	84.	(b)	85.	(a)
86.	(c)	87.	(c)	88.	(a)	89.	(b)	90.	(a)
91.	(b)	92.	(d)	93.	(a)	94.	(d)	95.	(c)
96.	(c)	97.	(a)	98.	(a)	99.	(c)	100.	(b)
101.	(a)	102.	(b)	103.	(c)	104.	(a)	105.	(a)
106.	(b)	107.	(a)	108.	(d)	109.	(c)	110.	(a)
111.	(b)	112.	(c)	113.	(b)	114.	(b)	115.	(a)
116.	(b)	117.	(c)	118.	(c)	119.	(b)	120.	(a)
121.	(b)	122.	(b)	123.	(a)	124.	(b)	125.	(b)
126.	(c)	127.	(d)	128.	(d)	129.	(a)	130.	(c)
131.	(b)	132.	(a)	133.	(a)	134.	(a)	135.	(b)
136.	(c)	137.	(a)	138.	(d)	139.	(a)	140.	(b)
141.	(a)	142.	(d)	143.	(c)	144.	(c)	145.	(a)
146.	(b)	147.	(a)	148.	(b)	149.	(a)	150.	(d)
151.	(c)	152.	(b)	153.	(b)	154.	(c)	155.	(a)

156. (a)	157. (c)	158. (b)	159. (d)	160. (a)
161. (c)	162. (a)	163. (b)	164. (a)	165. (c)
166. (a)	167. (b)	168. (c)	169. (a)	170. (b)
171. (b)	172. (c)	173. (c)	174. (c)	175. (b)
176. (c)	177. (b)	178. (d)	179. (c)	180. (c)
181. (b)	182. (b)	183. (b)	184. (c)	185. (a)
186. (a)	187. (b)	188. (a)	189. (c)	190. (c)
191. (b)	192. (a)	193. (b)	194. (b)	195. (b)
196. (a)	197. (a)	198. (a)	199. (b)	200. (b)
201. (b)	202. (a)	203. (c)	204. (d)	205. (b)
206. (a)	207. (b)	208. (b)	209. (a)	210. (a)
211. (b)	212. (a)	213. (b)	214. (a)	215. (b)
216. (a)	217. (b)	218. (c)	219. (a)	220. (a)
221. (a)	222. (a)	223. (b)	224. (c)	225. (a)
226. (d)	227. (a)	228. (b)	229. (a)	230. (d)
231. (c)	232. (a)	233. (b)	234. (a)	235. (d)
236. (b)	237. (b)	238. (c)	239. (a)	240. (b)
241. (b)	242. (b)	243. (a)	244. (c)	245. (a)
246. (c)	247. (b)	248. (b)	249. (a)	250. (c)
251. (b)	252. (d)	253. (a)	254. (a)	255. (d)
256. (c)	257. (a)	258. (a)	259. (c)	260. (c)
261. (c)	262. (a)	263. (c)	264. (a)	265. (a)
266. (b)	267. (a)	268. (c)	269. (b)	270. (b)
271. (a)	272. (b)	273. (c)	274. (c)	275. (a)
276. (a)	277. (b)	278. (c)	279. (c)	280. (a)
281. (d)	282. (d)	283. (d)	284. (c)	285. (c)
286. (c)	287. (c)	288. (a)	289. (b)	

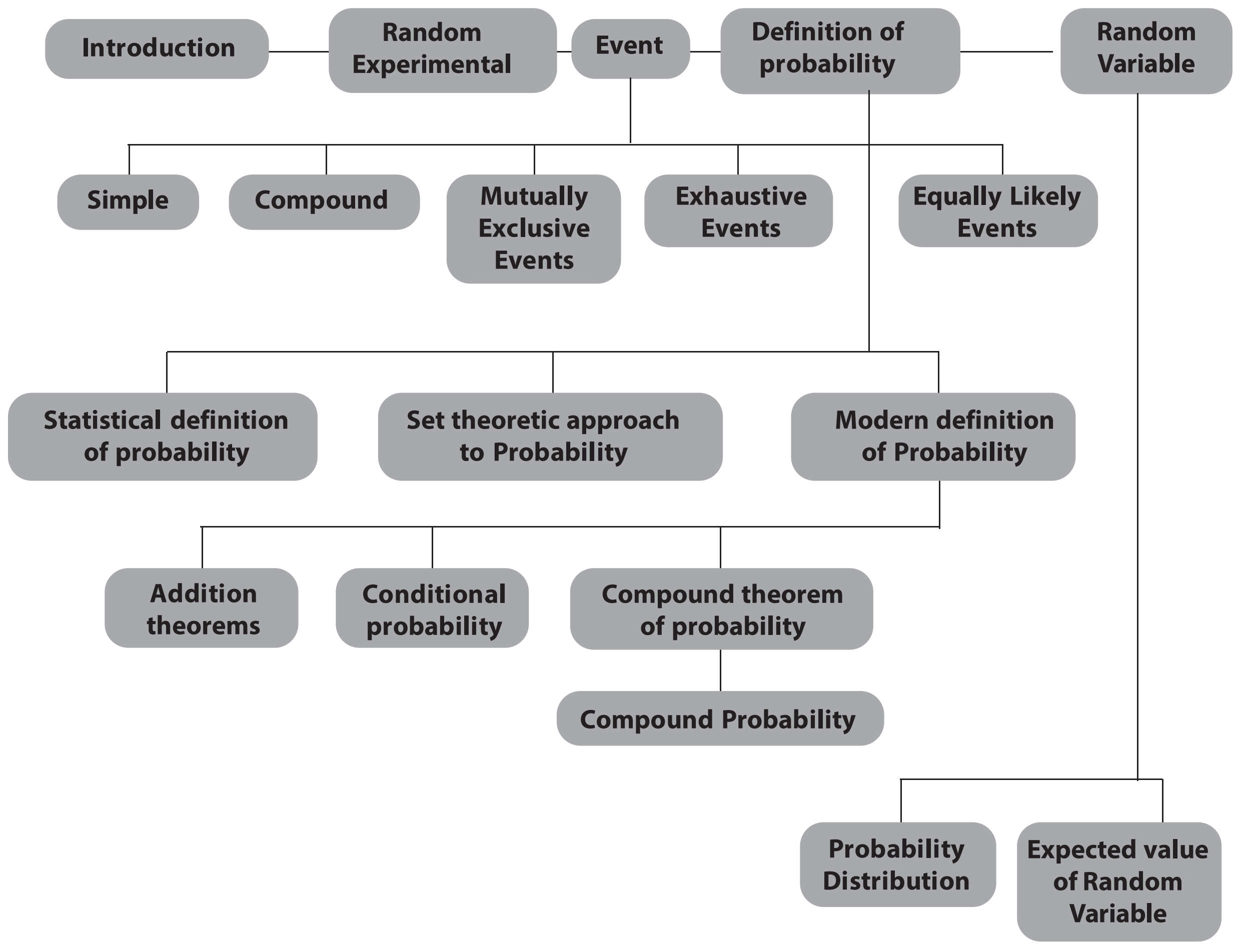
CHAPTER 16

PROBABILITY

LEARNING OBJECTIVES

Concept of probability is used in accounting and finance to understand the likelihood of occurrence or non-occurrence of a variable. It helps in developing financial forecasting in which you need to develop expertise at an advanced stage of chartered accountancy course.

UNIT OVERVIEW



16.1 INTRODUCTION

The terms 'Probably' 'in all likelihood', 'chance', 'odds in favour', 'odds against' are too familiar nowadays and they have their origin in a branch of Mathematics, known as Probability. In recent time, probability has developed itself into a full-fledged subject and become an integral part of statistics. The theories of Testing Hypothesis and Estimation are based on probability.

It is rather surprising to know that the first application of probability was made by a group of mathematicians in Europe about three hundreds years back to enhance their chances of winning in different games of gambling. Later on, the theory of probability was developed by Abraham De Moicere and Piere-Simon De Laplace of France, Reverend Thomas Bayes and R. A. Fisher of England, Chebyshev, Morkov, Khinchin, Kolmogorov of Russia and many other noted mathematicians as well as statisticians.

Two broad divisions of probability are Subjective Probability and Objective Probability. Subjective Probability is basically dependent on personal judgement and experience and, as such, it may be influenced by the personal belief, attitude and bias of the person applying it. However in the field of uncertainty, this would be quite helpful and it is being applied in the area of decision making management. This Subjective Probability is beyond the scope of our present discussion. We are going to discuss Objective Probability in the remaining sections.

16.2 RANDOM EXPERIMENT

In order to develop a sound knowledge about probability, it is necessary to get ourselves familiar with a few terms.

Experiment: An experiment may be described as a performance that produces certain results.

Random Experiment: An experiment is defined to be random if the results of the experiment depend on chance only. For example if a coin is tossed, then we get two outcomes—Head (H) and Tail (T). It is impossible to say in advance whether a Head or a Tail would turn up when we toss the coin once. Thus, tossing a coin is an example of a random experiment. Similarly, rolling a dice (or any number of dice), drawing items from a box containing both defective and non—defective items, drawing cards from a pack of well shuffled fifty—two cards etc. are all random experiments.

Events: The results or outcomes of a random experiment are known as events. Sometimes events may be combination of outcomes. The events are of two types:

(i) Simple or Elementary,

(ii) Composite or Compound.

An event is known to be simple if it cannot be decomposed into further events. Tossing a coin once provides us two simple events namely Head and Tail. On the other hand, a composite event is one that can be decomposed into two or more events. Getting a head when a coin is tossed twice is an example of composite event as it can be split into the events HT and TH which are both elementary events.

Mutually Exclusive Events or Incompatible Events: A set of events A1, A2, A3, …… is known to be mutually exclusive if not more than one of them can occur simultaneously. Thus occurrence of one such event implies the non-occurrence of the other events of the set. Once a coin is tossed, we get two mutually exclusive events Head and Tail.

Exhaustive Events: The events A1, A2, A3, ……….. are known to form an exhaustive set if one of these events must necessarily occur. As an example, the two events Head and Tail, when a coin is tossed once, are exhaustive as no other event except these two can occur.

Equally Likely Events or Mutually Symmetric Events or Equi-Probable Events: The events of a random experiment are known to be equally likely when all necessary evidence are taken into account, no event is expected to occur more frequently as compared to the other events of the set of events. The two events Head and Tail when a coin is tossed is an example of a pair of equally likely events because there is no reason to assume that Head (or Tail) would occur more frequently as compared to Tail (or Head).

16.3 CLASSICAL DEFINITION OF PROBABILITY OR A PRIOR DEFINITION

Let us consider a random experiment that result in n finite elementary events, which are assumed to be equally likely. We next assume that out of these n events, n_A $(\leq n)$ events are favourable to an event A. Then the probability of occurrence of the event A is defined as the ratio of the number of events favourable to A to the total number of events. Denoting this by P(A), we have

$$P(A) = \frac{n_A}{n} = \frac{\text{No. of equally likely events favourable to A}}{\text{Total no. of equally likely events}} \quad \ldots\ldots\ldots\ldots\ldots (16.1)$$

However if instead of considering all elementary events, we focus our attention to only those composite events, which are mutually exclusive, exhaustive and equally likely and if $m(\leq n)$ denotes such events and is furthermore $m_A(\leq n_A)$ denotes the no. of mutually exclusive, exhaustive and equally likely events favourable to A, then we have

$$P(A) = \frac{m_A}{m} = \frac{\text{No. of mutually exclusive, exhaustive and equally likely events favourable to A}}{\text{Total no. of mutually exclusive, exhaustive and equally likely events}}$$

……………… (16.2)

For this definition of probability, we are indebted to Bernoulli and Laplace. This definition is also termed as a priori definition because probability of the event A is defined on the basis of prior knowledge.

This classical definition of probability has the following demerits or limitations:

(i) It is applicable only when the total no. of events is finite.

(ii) It can be used only when the events are equally likely or equi-probable. This assumption is made well before the experiment is performed.

(iii) This definition has only a limited field of application like coin tossing, dice throwing, drawing cards etc. where the possible events are known well in advance. In the field of uncertainty or where no prior knowledge is provided, this definition is inapplicable.

In connection with classical definition of probability, we may note the following points:

(a) The probability of an event lies between 0 and 1, both inclusive.

i.e. $0 \leq P(A) \leq 1$ .... (16.3)

When P(A) = 0, A is known to be an impossible event and when P(A) = 1, A is known to be a sure event.

(b) Non-occurrence of event A is denoted by A′ or A^C or $\bar{A}$ and it is known as complimentary event of A. The event A along with its complimentary A′ forms a set of mutually exclusive and exhaustive events.

i.e. P(A) + P (A′) = 1

$$\Rightarrow \quad P(A') = 1 - P(A)$$

$$1 - \frac{m_A}{m}$$

$$= \frac{m - m_A}{m} \quad \text{............... (16.4)}$$

(c) The ratio of no. of favourable events to the no. of unfavourable events is known as odds in favour of the event A and its inverse ratio is known as odds against the event A.

i.e. odds in favour of A $= m_A : (m - m_A)$ .... (16.5)

and odds against A $= (m - m_A) : m_A$ .... (16.6)

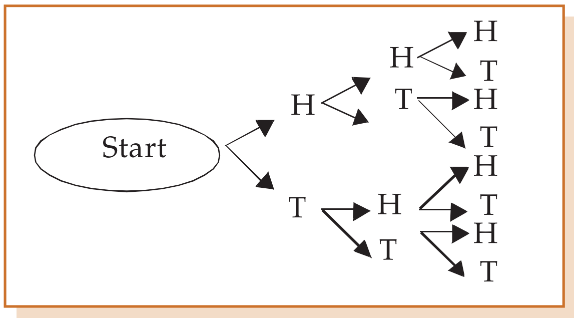
ILLUSTRATIONS:

Example 16.1: A coin is tossed three times. What is the probability of getting:

(i) 2 heads

(ii) at least 2 heads.

Solution: When a coin is tossed three times, first we need enumerate all the elementary events. This can be done using 'Tree diagram' as shown below:



Hence the elementary events are

HHH, HHT, HTH, HTT, THH, THT, TTH, TTT

Thus the number of elementary events (n) is 8.

(i) Out of these 8 outcomes, 2 heads occur in three cases namely HHT, HTH and THH. If we denote the occurrence of 2 heads by the event A and if assume that the coin as well as performer of the experiment is unbiased then this assumption ensures that all the eight elementary events are equally likely. Then by the classical definition of probability, we have

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

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

$$= 0.375$$

(ii) Let B denote occurrence of at least 2 heads i.e. 2 heads or 3 heads. Since 2 heads occur in 3 cases and 3 heads occur in only 1 case, B occurs in 3 + 1 or 4 cases. By the classical definition of probability,

$$P(B) = \frac{4}{8}$$

$$= 0.50$$

Example 16.2: A dice is rolled twice. What is the probability of getting a difference of 2 points?

Solution: If an experiment results in p outcomes and if the experiment is repeated q times, then the total number of outcomes is pq. In the present case, since a dice results in 6 outcomes and the dice is rolled twice, total no. of outcomes or elementary events is 6^2 or 36. We assume that the dice is unbiased which ensures that all these 36 elementary events are equally likely. Now a difference of 2 points in the uppermost faces of the dice thrown twice can occur in the following cases:

1st Throw	2nd Throw	Difference
6	4	2
5	3	2
4	2	2
3	1	2
1	3	2
2	4	2
3	5	2
4	6	2

Thus denoting the event of getting a difference of 2 points by A, we find that the no. of outcomes favourable to A, from the above table, is 8. By classical definition of probability, we get

$$P(A) = \frac{8}{36}$$

$$= \frac{2}{9}$$

Example 16.3: Two dice are thrown simultaneously. Find the probability that the sum of points on the two dice would be 7 or more.

Solution: If two dice are thrown then, as explained in the last problem, total no. of elementary events is 6^2 or 36. Now a total of 7 or more i.e. 7 or 8 or 9 or 10 or 11 or 12 can occur only in the following combinations:

SUM = 7: (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)

SUM = 8: (2, 6), (3, 5), (4, 4), (5, 3), (6, 2)

SUM = 9: (3, 6), (4, 5), (5, 4), (6, 3)

SUM = 10: (4, 6), (5, 5), (6, 4)

SUM = 11: (5, 6), (6, 5)

SUM = 12: (6, 6)

Thus the no. of favourable outcomes is 21. Letting A stand for getting a total of 7 points or more, we have

$$P(A) = \frac{21}{36}$$

$$= \frac{7}{12}$$

Example 16.4: What is the chance of picking a spade or an ace not of spade from a pack of 52 cards?

Solution: A pack of 52 cards contain 13 Spades, 13 Hearts, 13 Clubs and 13 Diamonds. Each of these groups of 13 cards has an ace. Hence the total number of elementary events is 52 out of which 13 + 3 or 16 are favourable to the event A representing picking a Spade or an ace not of Spade. Thus we have

$$P(A) = \frac{16}{52} = \frac{4}{13}$$

Example 16.5: Find the probability that a four digit number comprising the digits 2, 5, 6 and 7 would be divisible by 4.

Solution: Since there are four digits, all distinct, the total number of four digit numbers that can be formed without any restriction is 4! or $4 \times 3 \times 2 \times 1$ or 24. Now a four digit number would be divisible by 4 if the number formed by the last two digits is divisible by 4. This could happen when the four digit number ends with 52 or 56 or 72 or 76. If we fix the last two digits by 52, and then the 1st two places of the four digit number can be filled up using the remaining 2 digits in 2! or 2 ways. Thus there are 2 four digit numbers that end with 52. Proceeding in this manner, we find that the number of four digit numbers that are divisible by 4 is 4×2 or 8. If (A) denotes the event that any four digit number using the given digits would be divisible by 4, then we have

$$P(A) = \frac{8}{24}$$

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

Example 16.6: A committee of 7 members is to be formed from a group comprising 8 gentlemen and 5 ladies. What is the probability that the committee would comprise:

(a) 2 ladies,

(b) at least 2 ladies.

Solution: Since there are altogether 8 + 5 or 13 persons, a committee comprising 7 members can be formed in

$^{13}C_7$ or $\frac{13!}{7!6!}$ or $\frac{13\times12\times11\times10\times9\times8\times7!}{7!\times6\times5\times4\times3\times2\times1}$

or $11 \times 12 \times 13$ ways.

(a) When the committee is formed taking 2 ladies out of 5 ladies, the remaining (7–2) or 5 committee members are to be selected from 8 gentlemen. Now 2 out of 5 ladies can be selected in $^5C_2$ ways and 5 out of 8 gentlemen can be selected in $^8C_5$ ways. Thus if A denotes the event of having the committee with 2 ladies, then A can occur in $^5C_2\times {}^8C_5$ or

$\frac{5\times4}{2\times1}\times\frac{8\times7\times6}{3\times2}$ or 10×56 ways.

Thus $P(A) = \frac{10\times56}{11\times12\times13}$

$= \frac{140}{429}$

(b) Since the minimum number of ladies is 2, we can have the following combinations:

Population:	5L		8G
Sample:	2L	+	5G
or	3L	+	4G
or	4L	+	3G
or	5L	+	2G

Thus if B denotes the event of having at least two ladies in the committee, then B can occur in

$^5C_2 \times {}^8C_5 + {}^5C_3 \times {}^8C_4 + {}^5C_4 \times {}^8C_3 + {}^5C_5 \times {}^8C_2$

i.e. 1568 ways.

Hence $P(B) = \frac{1568}{11\times12\times13}$

$= \frac{392}{429}$

16.4 RELATIVE FREQUENCY DEFINITION OF PROBABILITY

Owing to the limitations of the classical definition of probability, there are cases when we consider the statistical definition of probability based on the concept of relative frequency. This definition of probability was first developed by the British mathematicians in connection with the survival probability of a group of people.

Let us consider a random experiment repeated a very good number of times, say n, under an identical set of conditions. We next assume that an event A occurs f_A times. Then the limiting value of the ratio of f_A to n as n tends to infinity is defined as the probability of A.

i.e. $$P(A) = \lim_{n \to \infty} \frac{F_A}{n} \qquad \text{.................. (16.7)}$$

This statistical definition is applicable if the above limit exists and tends to a finite value.

Example 16.7: The following data relate to the distribution of wages of a group of workers:

Wages in ₹:	50-60	60-70	70-80	80-90	90-100	100-110	110-120
No. of workers:	15	23	36	42	17	12	5

If a worker is selected at random from the entire group of workers, what is the probability that

(a) his wage would be less than ₹ 50?

(b) his wage would be less than ₹ 80?

(c) his wage would be more than ₹ 100?

(d) his wages would be between ₹ 70 and ₹ 100?

Solution: As there are altogether 150 workers, n = 150.

(a) Since there is no worker with wage less than ₹ 50, the probability that the wage of a randomly selected worker would be less than ₹ 50 is $P(A) = \frac{0}{150} = 0$

(b) Since there are (15+23+36) or 74 worker having wages less than ₹ 80 out of a group of 150 workers, the probability that the wage of a worker, selected at random from the group, would be less than ₹ 80 is

$$P(B) = \frac{74}{150} = \frac{37}{75}$$

(c) There are (12+5) or 17 workers with wages more than ₹ 100. Thus the probability of finding a worker, selected at random, with wage more than ₹ 100 is

$$P(C) = \frac{17}{150}$$

(d) There are (36+42+17) or 95 workers with wages in between ₹ 70 and ₹ 100. Thus

$$P(D) = \frac{95}{150} = \frac{19}{30}$$

16.5 OPERATIONS ON EVENTS-SET THEORETIC APPROACH TO PROBABILITY

Applying the concept of set theory, we can give a new dimension of the classical definition of probability. A sample space may be defined as a non-empty set containing all the elementary events of a random experiment as sample points. A sample space is denoted by S or Ω. An event A may be defined as a non-empty subset of S. This is shown in Figure 16.1

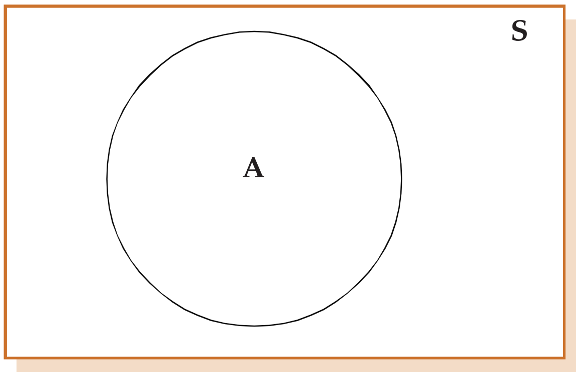


Figure 16.1

Showing an event A and the sample space S

As for example, if a dice is rolled once than the sample space is given by

S = {1, 2, 3, 4, 5, 6}.

Next, if we define the events A, B and C such that

A = {x: x is an even no. of points in S}

B = {x: x is an odd no. of points in S}

C = {x: x is a multiple of 3 points in S}

Then, it is quite obvious that

A = {2, 4, 6}, B = {1, 3, 5} and C = {3, 6}.

The classical definition of probability may be defined in the following way.

Let us consider a finite sample space S i.e. a sample space with a finite no. of sample points, n (S). We assume that all these sample points are equally likely. If an event A which is a subset of S, contains n (A) sample points, then the probability of A is defined as the ratio of the number of sample points in A to the total number of sample points in S. i.e.

$$P(A) = \frac{n(A)}{n(S)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (16.8)$$

Union of two events A and B is defined as a set of events containing all the sample points of event A or event B or both the events. This is shown in Figure 17.2 we have $A \cap B = \{x : x \in A$ on $x \in B\}$.

where x denotes the sample points.

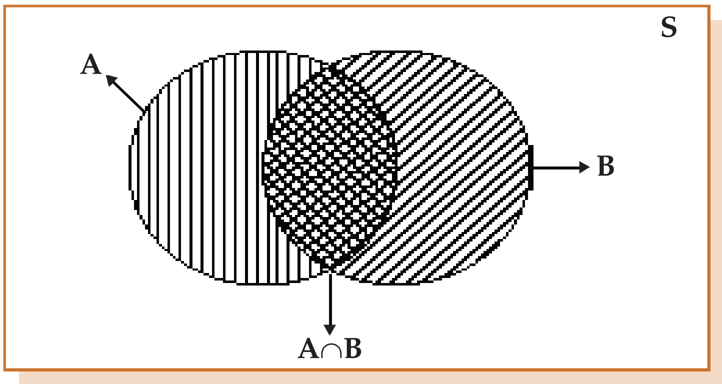


Figure 16.2

Showing the union of two events A and B and also their intersection

In the above example, we have $A \cup C = \{2, 3, 4, 6\}$

and $A \cup B = \{1, 2, 3, 4, 5, 6\}$.

The intersection of two events A and B may be defined as the set containing all the sample points that are common to both the events A and B. This is shown in Figure 16.2. we have

$A \cap B = \{x : x \in A \text{ and } x \in B\}$.

In the above example, $A \cap B = \phi$

$A \cap C = \{6\}$

Since the intersection of the events A and B is a null set ϕ, it is obvious that A and B are mutually exclusive events as they cannot occur simultaneously.

The difference of two events A and B, to be denoted by A – B, may be defined as the set of sample points present in set A but not in B. i.e.

$A - B = \{x : x \in A \text{ and } x \notin B\}$.

Similarly, B – A = {x:$x \in$ B and $x \notin$ A}.

In the above examples,

A – B = ϕ

And A – C = {2, 4}.

This is shown in Figure 16.3.

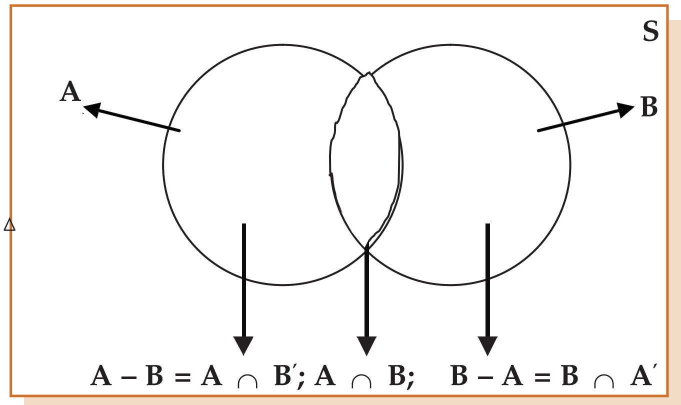


Figure 16.3

Showing (A – B) and (B – A)

The complement of an event A may be defined as the difference between the sample space S and the event A. i.e.

A′= {x: x $\in$ S and x $\notin$ A}.

In the above example A′= S – A

= {1, 3, 5}

Figure 16.4 depicts A′

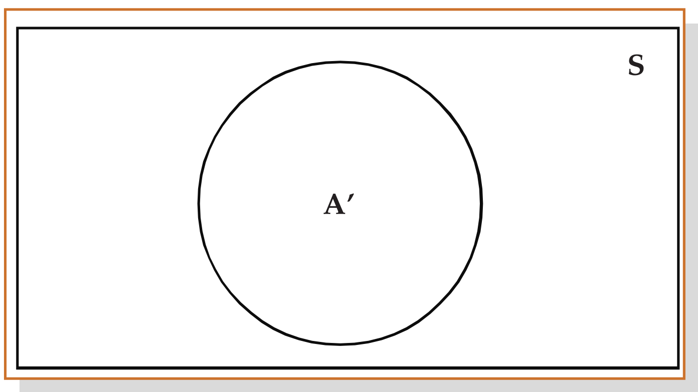


Figure 16.4

Showing A′

Now we are in a position to redefine some of the terms we have already discussed in section (17.2). Two events A and B are mutually exclusive if P (A $\cap$ B) = 0 or more precisely,....(16.9)

$P(A \cup B) = P(A) + P(B)$ ....(16.10)

Similarly three events A, B and C are mutually exclusive if

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$ ....(16.11)

Two events A and B are exhaustive if

$P(A \cup B) = 1$ ....(16.12)

Similarly three events A, B and C are exhaustive if

$P(A \cup B \cup C) = 1$ .... (16.13)

Three events A, B and C are equally likely if

$P(A) = P(B) = P(C)$ .... (16.14)

Example 16.8: Three events A, B and C are mutually exclusive, exhaustive and equally likely. What is the probability of the complementary event of A?

Solution: Since A, B and C are mutually exclusive, we have

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$ .... (1)

Since they are exhaustive, $P(A \cup B \cup C) = 1$ .... (2)

Since they are also equally likely, $P(A) = P(B) = P(C) = K$, Say ....(3)

Combining equations (1), (2) and (3), we have

$1 = K + K + K$

$\Rightarrow K = 1/3$

Thus $P(A) = P(B) = P(C) = 1/3$

Hence $P(A') = 1 - 1/3 = 2/3$

16.6 AXIOMATIC OR MODERN DEFINITION OF PROBABILITY

Let us consider a sample space S in connection with a random experiment and let A be an event defined on the sample space S i.e. $A \subseteq S$. Then a real valued function P defined on S is known as a probability measure and P(A) is defined as the probability of A if P satisfies the following axioms:

(i) $P(A) \geq 0$ for every $A \subseteq S$ (subset) .... (16.15)

(ii) $P(S) = 1$ .... (16.16)

(iii) For any sequence of mutually exclusive events $A_1, A_2, A_3, ..$

$P(A_1 \cup A_2 \cup A_3 \cup ....) = P(A_1) + P(A_2) + P(A_3) +$ .... (16.17)

16.7 ADDITION THEOREMS OR THEOREMS ON TOTAL PROBABILITY

Theorem 1 For any two mutually exclusive events A and B, the probability that either A or B occurs is given by the sum of individual probabilities of A and B.

i.e. $P(A \cup B)$

or $P(A + B) = P(A) + P(B)$ ………. (16.18)

or P(A or B) whenever A and B are mutually exclusive

This is illustrated in the following example.

Example 16.9: A number is selected from the first 25 natural numbers. What is the probability that it would be divisible by 4 or 7?

Solution: Let A be the event that the number selected would be divisible by 4 and B, the event that the selected number would be divisible by 7. Then AUB denotes the event that the number would be divisible by 4 or 7. Next we note that A = {4, 8, 12, 16, 20, 24} and B = {7, 14, 21} whereas S = {1, 2, 3, ……... 25}. Since $A \cap B = \phi$, the two events A and B are mutually exclusive and as such we have

$P(A \cup B) = P(A) + P(A)$ ……… (1)

Since $P(A) = \frac{n(A)}{n(S)} = \frac{6}{25}$

and $P(B) = \frac{n(B)}{n(S)} = \frac{3}{25}$

Thus from (1), we have

$$P(A \cup B) = \frac{6}{25} + \frac{3}{25}$$

$$= \frac{9}{25}$$

Hence the probability that the selected number would be divisible by 4 or 7 is 9/25 or 0.36

Example 16.10: A coin is tossed thrice. What is the probability of getting 2 or more heads?

Solution: If a coin is tossed three times, then we have the following sample space.

S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} 2 or more heads imply 2 or 3 heads.

If A and B denote the events of occurrence of 2 and 3 heads respectively, then we find that

A = {HHT, HTH, THH} and B = {HHH}

$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}$

and $P(B) = \frac{n(B)}{n(S)} = \frac{1}{8}$

As A and B are mutually exclusive, the probability of getting 2 or more heads is

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

$= \frac{3}{8} + \frac{1}{8}$

$= 0.50$

Theorem 2 For any $K(\geq 2)$ mutually exclusive events $A_1, A_2, A_3 \ldots, A_K$ the probability that at least one of them occurs is given by the sum of the individual probabilities of the K events.

i.e. $P(A_1 \cup A_2 \cup \ldots \cup A_K) = P(A_1) + P(A_2) + \ldots . P(A_K)$ .... (16.19)

Obviously, this is an extension of Theorem 1.

Theorem 3 For any two events A and B, the probability that either A or B occurs is given by the sum of individual probabilities of A and B less the probability of simultaneous occurrence of the events A and B.

i. e. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .... (16.20)

This theorem is stronger than Theorem 1 as we can derive Theorem 1 from Theorem 3 and not Theorem 3 from Theorem 1. For want of sufficient evidence, it is wiser to apply Theorem 3 for evaluating total probability of two events.

Example 16.11: A number is selected at random from the first 1000 natural numbers. What is the probability that it would be a multiple of 5 or 9?

Solution: Let A, B, $A \cup B$ and $A \cap B$ denote the events that the selected number would be a multiple of 5, 9, 5 or 9 and both 5 and 9 i.e. LCM of 5 and 9 i.e. 45 respectively.

Since $1000 = 5 \times 200$

$= 9 \times 111 + 1$

$= 45 \times 22 + 10,$

it is obvious that

$P(A) = \frac{200}{1000}, P(B) = \frac{111}{1000}, P(A \cap B) = \frac{22}{1000}$

Hence the probability that the selected number would be a multiple of 4 or 9 is given by

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

$= \frac{200}{1000} + \frac{111}{1000} - \frac{22}{1000}$

$= 0.29$

Example 16.12: The probability that an Accountant's job applicant has a B. Com. Degree is 0.85, that he is a CA is 0.30 and that he is both B. Com. and CA is 0.25 out of 500 applicants, how many would be B. Com. or CA?

Solution: Let the event that the applicant is a B. Com. be denoted by B and that he is a CA be denoted by C Then as given,

$P(B) = 0.85$, $P(C) = 0.30$ and $P(B \cap C) = 0.25$

The probability that an applicant is B. Com. or CA is given by

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

$= 0.85 + 0.30 - 0.25$

$= 0.90$

Expected frequency $= N \times P(B \cup C)$

Expected frequency $= 500 \times 9.90 = 450$

Example 16.13: If $P(A–B) = 1/5$, $P(A) = 1/3$ and $P(B) = 1/2$, what is the probability that out of the two events A and B, only B would occur?

Solution: A glance at Figure 17.3 suggests that

$P(A–B) = P(A \cap B') = P(A) - P(A \cap B)$ ....(16.21)

And $P(B–A) = P(B \cap A') = P(B) - P(A \cap B)$ ....(16.22)

Also (16.21) and (16.22) describe the probabilities of occurrence of the event only A and only B respectively.

As given $P(A–B) = \frac{1}{5}$

$\Rightarrow \quad P(A) - P(A \cap B) = \frac{1}{5}$

$\Rightarrow \quad \frac{1}{3} - P(A \cap B) = \frac{1}{5}$ [Since P(A) = 1/3]

$\Rightarrow \quad P(A \cap B) = \frac{2}{15}$

The probability that the event B only would occur

$= P(B–A)$

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

$= \frac{1}{2} - \frac{2}{15}$ [Since $P(B) = \frac{1}{2}$]

$= \frac{11}{30}$

Theorem 4 For any three events A, B and C, the probability that at least one of the events occurs is given by

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \quad \text{..........} \; (16.23)$$

Following is an application of this theorem.

Example 16.14: There are three persons A, B and C having different ages. The probability that A survives another 5 years is 0.80, B survives another 5 years is 0.60 and C survives another 5 years is 0.50. The probabilities that A and B survive another 5 years is 0.46, B and C survive another 5 years is 0.32 and A and C survive another 5 years 0.48. The probability that all these three persons survive another 5 years is 0.26. Find the probability that at least one of them survives another 5 years.

Solution: As given P(A) = 0.80, P(B) = 0.60, P(C) = 0.50,

$P(A \cap B) = 0.46$, $P(B \cap C) = 0.32$, $P(A \cap C) = 0.48$ and

$P(A \cap B \cap C) = 0.26$

The probability that at least one of them survives another 5 years in given by

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

$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$ .... (16.23)

$= 0.80 + 0.60 + 0.50 - 0.46 - 0.32 - 0.48 + 0.26$

$= 0.90$

Expected Frequency = N X P(BUC)
=500 X 0.90= 450

16.8 CONDITIONAL PROBABILITY AND COMPOUND THEOREM OF PROBABILITY

Compound Probability or Joint Probability

The probability of an event, discussed so far, is technically known as unconditional or marginal probability. However, there are situations that demand the probability of occurrence of more than one event. The probability of occurrence of two events A and B simultaneously is known as the Compound Probability or Joint Probability of the events A and B and is denoted by $P(A \cap B)$. In a similar manner, the probability of simultaneous occurrence of K events A_1, A_2, ... A_k, is denoted by $P(A_1 \cap A_2 \cap \dots \cap A_k)$.

In case of compound probability of 2 events A and B, we may face two different situations. In the first case, if the occurrence of one event, say B, is influenced by the occurrence of another event A, then the two events A and B are known as dependent events. We use the notation P(B/A), to be read as 'probability of the event B given that the event A has already occurred (or 'the conditional probability of B given A) to suggest that another event B will happen if and only if the first event A has already happened. This is given by

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} \qquad \text{........ (16.24)}$$

Provided P(A) > 0 i.e. A is not an impossible event.

Similarly, $P(A/B) = \frac{P(A \cap B)}{P(B)}$ .... (16.25)

if P(B) > 0.

As an example if a box co ntains 5 red and 8 white balls and two successive draws of 2 balls are made from it without replacement then the probability of the event 'the second draw would result in 2 white balls given that the first draw has resulted in 2 Red balls' is an example of conditional probability since the drawings are made without replacement, the composition of the balls in the box changes and the occurrence of 2 white balls in the second draw (B_2) is dependent on the outcome of the first draw (R_2). This event may b denoted by

$P(B_2/R_2)$.

In the second scenario, if the occurrence of the second event B is not influenced by the occurrence of the first event A, then B is known to be independent of A. It also follows that in this case, a is also independent of B and A and B are known as mutually independent or just independent. In this case, we have

$P(B/A) = P(B)$ .... (16.26)

and also $P(A/B) = P(A)$ .... (16.27)

There by implying, $P(A \cap B) = P(A) \times P(B)$ .... (16.28)

[From (16.24) or (16.25)]

In the above example, if the balls are drawn with replacement, then the two events B_2 and R_2 are independent and we have

$P(B_2 / R_2) = P(B_2)$

(16.28) is the necessary and sufficient condition for the independence of two events. In a similar manner, three events A, B and C are known as independent if the following conditions hold :

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

$P(A \cap C) = P(A) \times P(C)$

$P(B \cap C) = P(B) \times P(C)$

$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$ .... (16.29)

It may be further noted that if two events A and B are independent, then the following pairs of events are also independent:

(i) A and B′

(ii) A′ and B

(iii) A′ and B′ .... (16.30)

Theorems of Compound Probability

Theorem 5 For any two events A and B, the probability that A and B occur simultaneously is given by the product of the unconditional probability of A and the conditional probability of B given that A has already occurred

i.e. $P(A \cap B) = P(A) \times P(B/A)$ Provided $P(A) > 0$ ……………(16.31)

Theorem 6 For any three events A, B and C, the probability that they occur jointly is given by

$P(A \cap B \cap C) = P(A) \times P(B/A) \times P(C/(A \cap B))$ Provided $P(A \cap B) > 0$ … (16.32)

In the event of independence of the events

(16.31) and (16.32) are reduced to

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

and $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

which we have already discussed.

Example 16.15: Rupesh is known to hit a target in 5 out of 9 shots whereas David is known to hit the same target in 6 out of 11 shots. What is the probability that the target would be hit once they both try?

Solution: Let A denote the event that Rupesh hits the target and B, the event that David hits the target. Then as given,

$P(A) = \frac{5}{9},\ P(B) = \frac{6}{11}$

and $P(A \cap B) = P(A) \times P(B)$

$= \frac{5}{9} \times \frac{6}{11}$

$= \frac{10}{33}$ (as A and B are independent)

The probability that the target would be hit is given by

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

$= \frac{5}{9} + \frac{6}{11} - \frac{10}{33}$

$= \frac{79}{99}$

Alternately $P(A \cup B) = 1 - P(A \cup B)'$

$= 1 - P(A' \cap B')$ (by De-Morgan's Law)

$= 1 - P(A') \times P(B')$

$= 1 - [1 - P(A)] \times [1 - P(B)]$ (by 12.30)

$= 1 - (1 - \frac{5}{9}) \times (1 - \frac{6}{11})$

$= 1 - \frac{4}{9} \times \frac{5}{11}$

$= \frac{79}{99}$

Example 16.16: A pair of dice is thrown together and the sum of points of the two dice is noted to be 10. What is the probability that one of the two dice has shown the point 4?

Solution: Let A denote the event of getting 4 points on one of the two dice and B denote the event of getting a total of 10 points on the two dice. Then we have

$P(A) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$

and $P(A \cap B) = \frac{2}{36}$

[Since a total of 10 points may result in (4, 6) or (5, 5) or (6, 4) and two of these combinations contain 4]

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

$= \frac{2/36}{1/12}$

$= \frac{2}{3}$

Alternately The sample space for getting a total of 10 points when two dice are thrown simultaneously is given by

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

Out of these 3 cases, we get 4 in 2 cases. Thus by the definition of probability, we have

$P(B/A) = \frac{2}{3}$

Example 16.17: In a group of 20 males and 15 females, 12 males and 8 females are service holders. What is the probability that a person selected at random from the group is a service holder given that the selected person is a male?

Solution: Let S and M stand for service holder and male respectively. We are to evaluate P (S / M).

We note that $(S \cap M)$ represents the event of both service holder and male.

Thus $P(S/M) = \frac{P(S \cap M)}{P(M)}$

$= \frac{12/35}{20/35}$

$= 0.60$

Example 16.18: In connection with a random experiment, it is found that

$P(A) = \frac{2}{3}$, $P(B)\ \frac{3}{5}$ = and $P(A \cup B) = \frac{5}{6}$

Evaluate the following probabilities:

(i) P(A/B) (ii) P(B/A) (iii) P(A′/ B) (iv) P(A/ B′) (v) P(A′/ B′)

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

$=> \frac{5}{6} = \frac{2}{3} + \frac{3}{5} - P(A \cap B)$

$=> P(A \cap B) = \frac{2}{3} + \frac{3}{5} - \frac{5}{6}$

$= \frac{13}{30}$

Hence (i) $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{13/30}{3/5} = \frac{13}{18}$

(ii) $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{13/30}{2/3} = \frac{13}{20}$

(iii) $P(A'/B) = \frac{P(A' \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} = \frac{\frac{3}{5} - \frac{13}{30}}{\frac{3}{5}} = \frac{5}{18}$

(iv) $(A/B') = \frac{P(A \cap B')}{P(B')} = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{7}{12}$

(v) $P(A'/B') = \frac{P(A' \cap B')}{P(B')}$

$$= \frac{P(A \cup B)'}{P(B')} \quad \text{[by De-Morgan's Law } A' \cap B' = (AUB)' \text{]}$$

$$= \frac{1 - P(A \cup B)}{1 - P(B)}$$

$$= \frac{1 - 5/6}{1 - 3/5}$$

$$= \frac{5}{12}$$

Example 16.19: The odds in favour of an event is 2 : 3 and the odds against another event is 3 : 7. Find the probability that only one of the two events occurs.

Solution: We denote the two events by A and B respectively. Then by (16.5) and (16.6), we have

$$P(A) = \frac{2}{2+3} = \frac{2}{5}$$

$$\text{and } P(B) = \frac{7}{7+3} = \frac{7}{10}$$

As A and B are independent, $P(A \cap B) = P(A) \times P(B)$

$$= \frac{2}{5} \times \frac{7}{10} = \frac{7}{25}$$

Probability that either only A occurs or only B occurs

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

$= [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)]$

$= P(A) + P(B) - 2\, P(A \cap B)$

$$= \frac{2}{5} + \frac{7}{10} - 2 \times \frac{7}{25}$$

$$= \frac{20 + 35 - 28}{50}$$

$$= \frac{27}{50}$$

Example 16.20: There are three boxes with the following compositions :

Box \ Colour	Blue	Red	White	Total
I	5	8	10	23
II	4	9	8	21
III	3	6	7	16

One ball in drawn from each box. What is the probability that they would be of the same colour?

Solution: Either the balls would be Blue or Red or White. Denoting Blue, Red and White balls by B, R and W respectively and the box by lower suffix, the required probability is

$$= P(B_1 \cap B_2 \cap B_3) + P(R_1 \cap R_2 \cap R_3) + P(W_1 \cap W_2 \cap W_3)$$

$$= P(B_1) \times P(B_2) \times P(B_3) + P(R_1) \times P(R_2) \times P(R_3) + P(W_1) \times P(W_2) \times P(W_3)$$

$$= \frac{5}{23}\times\frac{4}{21}\times\frac{3}{16}+\frac{8}{23}\times\frac{9}{21}\times\frac{6}{16}+\frac{10}{23}\times\frac{8}{21}\times\frac{7}{16}$$

$$= \frac{60+432+560}{7728}$$

$$= \frac{1052}{7728}$$

Example 16.21: Mr. Roy is selected for three separate posts. For the first post, there are three candidates, for the second, there are five candidates and for the third, there are 10 candidates. What is the probability that Mr. Roy would be selected?

Solution: Denoting the three posts by A, B and C respectively, we have

$$P(A) = \frac{1}{3},\ P(B) = \frac{1}{5} \text{ and } P(C) = \frac{1}{10}$$

The probability that Mr. Roy would be selected (i.e. selected for at least one post).

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

$= 1 - P[(A \cup B \cup C)']$

$= 1 - P(A' \cap B' \cap C')$ (by De-Morgan's Law)

$= 1 - P(A') \times P(B') \times P(C')$ (As A , B and C are independent, so are their complements)

$$= 1-\left(1-\frac{1}{3}\right)\times\left(1-\frac{1}{5}\right)\times\left(1-\frac{1}{10}\right) = \frac{13}{25}$$

Example 16.22: The independent probabilities that the three sections of a costing department will encounter a computer error are 0.2, 0.3 and 0.1 per week respectively what is the probability that there would be

(i) at least one computer error per week?

(ii) one and only one computer error per week?

Solution: Denoting the three sections by A, B and C respectively, the probabilities of encountering a computer error by these three sections are given by P(A) = 0.20, P(B) = 0.30 and P(C) = 0.10

(i) Probability that there would be at least one computer error per week.

= 1 – Probability of having no computer error in any at the three sections.

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

$= 1 - P(A') \times P(B') \times P(C')$ [Since A, B and C are independent]

$= 1 - (1 - 0.20) \times (1 - 0.30) \times (1 - 0.10)$

= 0.50

(ii) Probability of having one and only one computer error per week

$= P(A \cap B' \cap C') + P(A' \cap B \cap C') + P(A' \cap B' \cap C)$

$= P(A) \times P(B') \times P(C') + P(A') \times P(B) \times P(C') + P(A') \times P(B') \times P(C)$

$= 0.20 \times 0.70 \times 0.90 + 0.80 \times 0.30 \times 0.90 + 0.80 \times 0.70 \times 0.10$

= 0.40

Example 16.23: A lot of 10 electronic components is known to include 3 defective parts. If a sample of 4 components is selected at random from the lot, what is the probability that this sample does not contain more than one detectives?

Solution: Denoting detective component and non-defective components by D and D′ respectively, we have the following situation :

	D	D′	T
Lot	3	7	10
Sample (1)	0	4	4
(2)	1	3	4

Thus the required probability is given by

$= ({}^3C_0 \times {}^7C_4 + {}^3C_1 \times {}^7C_3) / {}^{10}C_4$

$= \dfrac{1 \times 35 + 3 \times 35}{210}$

$= \dfrac{2}{3}$

Example 16.24: There are two urns containing 5 red and 6 white balls and 3 red and 7 white balls respectively. If two balls are drawn from the first urn without replacement and transferred to the second urn and then a draw of another two balls is made from it, what is the probability that both the balls drawn are red?

Solution: Since two balls are transferred from the first urn containing 5 red and 6 white balls to the second urn containing 3 red and 7 white balls, we are to consider the following cases :

Case A : Both the balls transferred are red. In this case, the second urn contains 5 red and 7 white balls.

Case B : The two balls transferred are of different colours. Then the second urn contains 4 red and 8 white balls.

Case C : Both the balls transferred are white. Now the second urn contains 3 red and 7 white balls.

The required probability is given by

$P(R \cap A) + P(R \cap B) + P(R \cap C)$

$= P(R/A) \times P(A) + P(R/B) \times P(B) + P(R/C) \times P(C)$

$$= \frac{^5C_2}{^{12}C_2} \times \frac{^5C_2}{^{11}C_2} + \frac{^4C_2}{^{12}C_2} \times \frac{^5C_1 \times {^6C_1}}{^{11}C_2} \times \frac{^3C_2}{^{12}C_2} \times \frac{^6C_2}{^{11}C_2}$$

$$= \frac{10}{66} \times \frac{10}{55} + \frac{6}{66} \times \frac{30}{55} + \frac{3}{66} \times \frac{15}{55}$$

$$= \frac{325}{66 \times 55} = \frac{65}{726}$$

Example 16.25: If 8 balls are distributed at random among three boxes, what is the probability that the first box would contain 3 balls?

Solution: The first ball can be distributed to the 1st box or 2nd box or 3rd box i.e. it can be distributed in 3 ways. Similarly, the second ball also can be distributed in 3 ways. Thus the first two balls can be distributed in 3^2 ways. Proceeding in this way, we find that 8 balls can be distributed to 3 boxes in 3^8 ways which is the total number of elementary events.

Let A be the event that the first box contains 3 balls which implies that the remaining 5 both must go to the remaining 2 boxes which, as we have already discussed, can be done in 2^5 ways. Since 3 balls out of 8 balls can be selected in $^8C_3$ ways, the event can occur in $^8C_3 \times 2^5$ ways, thus we have

$$P(A) = \frac{^8C_3 \times 2^5}{3^8}$$

$$= \frac{56 \times 32}{6561}$$

$$= \frac{1792}{6561}$$

Example 16.26: There are 3 boxes with the following composition :

Box I : 7 Red + 5 White + 4 Blue balls

Box II : 5 Red + 6 White + 3 Blue balls

Box III : 4 Red + 3 White + 2 Blue balls

One of the boxes is selected at random and a ball is drawn from it. What is the probability that the drawn ball is red?

Solution: Let A denote the event that the drawn ball is blue. Since any of the 3 boxes may be drawn, we have $P(B_I) = P(B_{II}) = P(B_{III}) = \frac{1}{3}$

Also $P(R_1/B_{II})$ = probability of drawing a red ball from the first box

$$= \frac{7}{16}$$

$P(R_2 \;/\; B_{II}) = \frac{5}{14}$ and $P(R_3 \;/\; B_{III}) = \frac{4}{9}$

Thus we have

$$P(A)= P(R_1 \cap B_I) + P(R_2 \cap B_{II}) + P(R_3 \cap B_{III})$$

$$= P(R_1 \;/\; B_I) \times P(B_I) + P(R_2 \;/\; B_{II}) \times P(B_{II}) + P(R_3 \;/\; B_{III}) \times P(B_{III})$$

$$= \frac{7}{16}\times\frac{1}{3}+\frac{5}{14}\times\frac{1}{3}+\frac{4}{9}\times\frac{1}{3}$$

$$= \frac{7}{48}+\frac{5}{42}+\frac{4}{27}$$

$$= \frac{1249}{3024}$$

16.9 RANDOM VARIABLE - PROBABILITY DISTRIBUTION

A random variable or stochastic variable is a function defined on a sample space associated with a random experiment assuming any value from R and assigning a real number to each and every sample point of the random experiment. A random variable is denoted by a capital letter. For example, if a coin is tossed three times and if X denotes the number of heads, then X is a random variable. In this case, the sample space is given by

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

and we find that X = 0 if the sample point is TTT

X = 1 if the sample point is HTT, THT or TTH

X = 2 if the sample point is HHT, HTH or THH

and X = 3 if the sample point is HHH.

We can make a distinction between a discrete random variable and a continuous variable. A random variable defined on a discrete sample space is known as a discrete random variable and it can assume either only a finite number or a countably infinite number of values. The number of car accident, the number of heads etc. are examples of discrete random variables.

A continuous random variable, like height, weight etc. is a random variable defined on a continuous sample space and assuming an uncountably infinite number of values.

The probability distribution of a random variable may be defined as a statement expressing the different values taken by a random variable and the corresponding probabilities. Then if a random variable X assumes n finite values X, X_2, X_3, ..., X_n with corresponding probabilities P_1, P_2, P_3, ..., P_n such that

(i) $p_i \geq 0$ for every i .... (16.33)

and (ii) $\sum p_i = 1$ (over all i) .... (16.34)

then the probability distribution of the random variable X is given by

Probability Distribution of X

X :	X_1	X_2	X_3	......X_n	Total
P :	P_1	P_2	P_3	.... P_n	1

For example, if an unbiased coin is tossed three times and if X denotes the number of heads then, as we have already discussed, X is a random variable and its probability distribution is given by

Probability Distribution of Head when a Coin is Tossed Thrice

X :	0	1	2	3	Total
P :	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

There are cases when it is possible to express the probability (P) as a function of X. In case X is a discrete variable and if such a function f(X) really exists, then f(X) is known as Probability Mass Function (PMF) of X, f(X), then, must satisfy the conditions:

(i) $f(X) \geq 0$ for every X .... (16.35)

and (ii) $\sum_X f(X) = 1$ .... (16.36)

Where f(X) is given by

$f(X) = P(X = X)$ .... (16.37)

When x is a continuous random variable defined over an interval [α, β], where $\beta > \alpha$, then x can assume an infinite number of values from its interval and instead of assigning individual probability to every mass point x, we assign probabilities to interval of values. Such a function of x, provided it exists, is known as probability density function (pdf) of x. f(x) satisfies the following conditions:

(i) $f(x) \geq 0$ for $x \in [\alpha, \beta]$ .... (16.38)

(ii) $\int_{\alpha}^{\beta} f(x)dx = 1$ .... (16.39)

and the probability that x lies between two specified values a and b, where $\alpha \leq a < b \leq \beta$, is given by

$\int_{a}^{b} f(x)dx$ .... (16.40)

16.10 EXPECTED VALUE OF A RANDOM VARIABLE

Expected value or Mathematical Expectation or Expectation of a random variable may be defined as the sum of products of the different values taken by the random variable and the corresponding probabilities. Hence, if a random variable x assumes n values x_1, x_2, x_3 …., x_n with corresponding probabilities p_1, p_2, p_3 …., p_n , where p_i's satisy (16.33) and (16.34), then the expected value of x is given by

$\mu = E(x) = \sum p_i x_i$ .... (16.41)

Expected value of x^2 in given by

$E(x^2) = \sum p_i x_i^2$ .... (16.42)

In particular expected value of a monotonic function g (x) is given by

$E[g(x)] = \sum p_i g(x_i)$ .... (16.43)

Variance of x, to be denoted by , σ^2 is given by

$V(x) = \sigma^2 = E(x - \mu)^2$

$= E(x^2) - \mu^2$ .... (16.44)

The positive square root of variance is known as standard deviation and is denoted by σ.

If y = a + b x, for two random variables x and y and for a pair of constants a and b, then the mean i.e. expected value of y is given by

$\mu_y = a + b\ \mu_x$ .... (16.45)

and the standard deviation of y is

$\sigma_y = |b| \times \sigma_X$ .... (16.46)

When x is a discrete random variable with probability mass function f(x), then its expected value is given by

$$\mu = \sum_x x f(x) \qquad \text{......... (16.47)}$$

and its variance is

$$\sigma^2 = E(x^2) - \mu^2$$

Where $E(x^2) = \sum_x x^2 f(x)$ ....(16.48)

For a continuous random variable x defined in [–∞, ∞], its expected value (i.e. mean) and variance are given by

$$E(x) = \int_{-\infty}^{\infty} x f(x)dx \qquad \text{......... (16.49)}$$

and $\sigma^2 = E(x^2) - \mu^2$

where $E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$ .... (16.50)

Properties of Expected Values

1. Expectation of a constant k is k

   i.e. E(k) = k for any constant k. ....(16.51)

2. Expectation of sum of two random variables is the sum of their expectations.

   i.e. E(x + y) = E(x) + E(y) for any two random variables x and y. .... (16.52)

3. Expectation of the product of a constant and a random variable is the product of the constant and the expectation of the random variable.

   i.e. E(k x) = k.E(x) for any constant k .... (16.53)

4. Expectation of the product of two random variables is the product of the expectation of the two random variables, provided the two variables are independent.

   i.e. $E(xy) = E(x) \times E(y)$ .... (16.54)

   Whenever x and y are independent.

Example 16.27: An unbiased coin is tossed three times. Find the expected value of the number of heads and also its standard deviation.

Solution: If x denotes the number of heads when an unbiased coin is tossed three times, then the probability distribution of x is given by

X :	0	1	2	3
P :	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

The expected value of x is given by

$\mu = E(x) = \sum p_i x_i$

$= \frac{1}{8}\times 0+\frac{3}{8}\times 1+\frac{3}{8}\times 2+\frac{1}{8}\times 3$

$= \frac{0+3+6+3}{8} = 1.50$

Also $\quad E(x^2) = \sum p_i x_i^2$

$= \frac{1}{8}\times 0^2+\frac{3}{8}\times 1^2+\frac{3}{8}\times 2^2+\frac{1}{8}\times 3^2$

$= \frac{0+3+12+9}{8} = 3$

$= \sigma^2 = E(x^2) - \mu^2$

$= 3 - (1.50)^2$

$= 0.75$

$\therefore SD = \sigma = 0.87$

Example 16.28: A random variable has the following probability distribution:

X :	4	5	7	8	10
P :	0.15	0.20	0.40	0.15	0.10

Find $E[x - E(x)]^2$. Also obtain v(3x –4)

Solution: The expected value of x is given by

$E(x) = \sum p_i x_i$

$= 0.15 \times 4 + 0.20 \times 5 + 0.40 \times 7 + 0.15 \times 8 + 0.10 \times 10$

$= 6.60$

Also, $\quad E[x - E(x)]^2 = \sum \mu_i^2 P_i$ where $= \mu_i = x_i - E(x)$

Let $y = 3x - 4 = (-4) + (3)x$. Then variance of y = var y = $b^2 \times \sigma_x^2 = 9 \times \mu_x^2$ (From 16.46)

Table 16.1

Computation of E [x – E(x)]²

x_i	p_i	$\mu_i = x_i - E(x)$	^2_i	$^2_i p_i$
4	0.15	–2.60	6.76	1.014
5	0.20	–1.60	2.56	0.512
7	0.40	0.40	0.16	0.064
8	0.15	1.40	1.96	0.294
10	0.10	3.40	11.56	1.156
Total	1.00	–	–	3.040

Thus $E[x - E(x)]^2 = 3.04$

As $\mu_x^2 = 3.04$, $v(y) = 9 \times 3.04 = 27.36$

Example 16.29: In a business venture, a man can make a profit of ₹ 50,000 or incur a loss of ₹ 20,000. The probabilities of making profit or incurring loss, from the past experience, are known to be 0.75 and 0.25 respectively. What is his expected profit?

Solution: If the profit is denoted by x, then we have the following probability distribution of x:

X :	₹ 50,000	₹ –20,000
P :	0.75	0.25

Thus his expected profit

$E(x) = p_1x_1 + p_2x_2$

$= 0.75 \times ₹\ 50{,}000 + 0.25 \times (₹ - 20{,}000)$

$= ₹\ 32{,}500$

Example 16.30: A box contains 12 electric lamps of which 5 are defectives. A man selects three lamps at random. What is the expected number of defective lamps in his selection?

Solution: Let x denote the number of defective lamps x can assume the values 0, 1, 2 and 3.

P(x = 0) = Prob. of having 0 defective out of 5 defectives and 3 non defective out of 7 non defectives

$$= \frac{^5C_0 \times {}^7C_3}{^{12}C_3} = \frac{35}{220}$$

Similarly $$P(x = 1) = \frac{^5C_1 \times {}^7C_2}{^{12}C_3} = \frac{105}{220}$$

$$P(x = 2) = \frac{{}^5C_2 \times {}^7C_1}{{}^{12}C_3} = \frac{70}{220}$$

and
$$P(x = 3) = \frac{{}^5C_3 \times {}^7C_0}{{}^{12}C_3} = \frac{10}{220}$$

Probability Distribution of No. of Defective Lamp

X :	0	1	2	3
P :	$\frac{35}{220}$	$\frac{105}{220}$	$\frac{70}{220}$	$\frac{10}{220}$

Thus the expected number of defectives is given by

$$\frac{35}{220} \times 0 + \frac{105}{220} \times 1 + \frac{70}{220} \times 2 + \frac{10}{220} \times 3$$

$= 1.25$

Example 16.31: Moidul draws 2 balls from a bag containing 3 white and 5 Red balls. He gets ₹ 500 if he draws a white ball and ₹ 200 if he draws a red ball. What is his expectation? If he is asked to pay ₹ 400 for participating in the game, would he consider it a fair game and participate?

Solution: We denote the amount by x. Then x assumes the value 2 x ₹ 500 i.e. ₹ 1000 if 2 white balls are drawn, the value ₹ 500 + ₹ 200 i.e. ₹ 700 if 1 white and 1 red balls are drawn and the value 2 x ₹ 200 i.e. ₹ 400 if 2 red balls are drawn. The respective probabilities are given by

$$P(WW) = \frac{{}^3C_2}{{}^8C_2} = \frac{3}{28}$$

$$P(WR) = \frac{{}^3C_1 \times {}^5C_1}{{}^8C_2} = \frac{15}{28}$$

$$\text{and } P(RR) = \frac{{}^5C_2}{{}^8C_2} = \frac{10}{28}$$

Probability Distribution of x

X :	₹ 1000	₹ 700	₹ 400
P :	$\frac{3}{28}$	$\frac{15}{28}$	$\frac{10}{28}$

Hence $E(x) = \frac{3}{28} \times$ ₹ 1000 + $\frac{15}{28} \times$ ₹ 700 + $\frac{10}{28} \times$ ₹ 400

$$= \frac{₹\,3000 + ₹\,10500 + ₹\,4000}{28}$$

= ₹ 625 > 400. Therefore the game is fair and he would participate.

Example 16.32: A number is selected at random from a set containing the first 100 natural numbers and another number is selected at random from another set containing the first 200 natural numbers. What is the expected value of the product?

Solution: We denote the number selected from the first set by x and the number selected from the second set by y. Since the selections are independent of each other, the expected value of the product is given by

$$E(xy) = E(x) \times E(y) \qquad \text{.......... (1)}$$

Now x can assume any value between 1 to 100 with the same probability 1/100 and as such the probability distribution of x is given by

X :	1	2	3	100
P :	$\frac{1}{100}$	$\frac{1}{100}$	$\frac{1}{100}$	.... $\frac{1}{100}$

$$\text{Thus } E(x) = \frac{1}{100} \times 1 + \frac{1}{100} \times 2 + \frac{1}{100} \times 3 + \text{..........................} \frac{1}{100} \times 100$$

$$= \frac{1+2+3+\text{..........}+100}{100}$$

$$= \frac{100\times101}{2\times100} \qquad \left[\text{Since } 1+2+\text{.....} + n = \frac{n(n+1)}{2}\right]$$

$$= \frac{101}{2}$$

Similarly, $E(y) = \frac{201}{2}$

$$\therefore E(xy) = \frac{101}{2} \times \frac{201}{2} \qquad \text{[From (1)]}$$

$$= \frac{20301}{4}$$

$$= 5075.25$$

Example 16.33: A dice is thrown repeatedly till a 'six' appears. Write down the sample space. Also find the expected number of throws.

Solution: Let p denote the probability of getting a six and q = 1 – p, the probability of not getting a six. If the dice is unbiased then

$p = \frac{1}{6}$ and $q = \frac{5}{6}$

If a six obtained with the very first throw then the experiment ends and the probability of getting a six, as we have already seen, is p. However, if the first throw does not produce a six, the dice is thrown again and if a six appears with the second throw, the experiment ends. The probability of getting a six preceded by a non–six is qp. If the second thrown does not yield a six, we go for a third throw and if the third throw produces a six, the experiment ends and the probability of getting a Six in the third attempt is q^2p. The experiment is carried on and we get the following countably infinite sample space.

$S = \{ p, qp, q^2p, q^3p, \ldots\}$

If x denotes the number of throws necessary to produce a six, then x is a random variable with the following probability distribution :

X :	1	2	3	4 ……….
P :	p	qp	q^2p	q^3p ……….

Thus $E(x) = p \times 1 + qp \times 2 + q^2p \times 3 + q^3p \times 4 + \ldots\ldots$

$= p(1+ 2q + 3q^2 + 4q^3 + \ldots\ldots)$

$= p\,(1 - q)^{-2}$

$= \frac{p}{p^2}$ (as $1-q = p$)

$= \frac{1}{p}$

In case of an unbiased dice, $p = 1/6$ and $E(x) = 6$

Example 16.34: A random variable x has the following probability distribution :

X	:	0	1	2	3	4	5	6	7
P(X)	:	0	2k	3k	k	2k	k^2	$7k^2$	$2k^2+k$

Find (i) the value of k

(ii) $P(x < 3)$

(iii) $P(x \geq 4)$

(iv) $P(2 < x \leq 5)$

Solution: By virtue of (17.36), we have

$\sum P(x) = 1$

$\Rightarrow \quad 0 + 2k + 3k + k + 2k + k^2 + 7k^2 + 2k^2 + k = 1$

$\Rightarrow$ $10k^2 + 9k - 1 = 0$

$\Rightarrow$ $(k + 1)(10k - 1) = 0$

$\Rightarrow$ $k = 1/10$ (as $k \neq -1$ by virtue of (17.36))

(i) Thus the value of k is 0.10

(ii) $P(x < 3) = P(x = 0) + P(x = 1) + P(x = 2)$

$= 0 + 2k + 3k$

$= 5k$

$= 0.50$ (as $k = 0.10$)

(iii) $P(x \geq 4) = P(x = 4) + P(x = 5) + P(x = 6) + P(x = 7)$

$= 2k + k^2 + 7k^2 + (2k^2 + k)$

$= 10k^2 + 3k$

$= 10 \times (0.10)^2 + 3 \times 0.10$

$= 0.40$

(iv) $P(2 < x \leq 5) = P(x = 3) + P(x = 4) + P(x = 5)$

$= k + 2k + k^2$

$= k^2 + 3k$

$= (0.10)^2 + 3 \times 0.10$

$= 0.31$

SUMMARY

- **Experiment:** An experiment may be described as a performance that produces certain results.
- **Random Experiment:** An experiment is defined to be random if the results of the experiment depend on chance only.
- **Events:** The results or outcomes of a random experiment are known as events. Sometimes events may be combination of outcomes. The events are of two types:
  (i) Simple or Elementary,
  (ii) Composite or Compound.
- **Mutually Exclusive Events or Incompatible Events:** A set of events $A_1, A_2, A_3, \ldots\ldots$ is known to be mutually exclusive if not more than one of them can occur simultaneously
- **Exhaustive Events:** The events $A_1, A_2, A_3, \ldots\ldots\ldots\ldots$ are known to form an exhaustive set if one of these events must necessarily occur.

- **Equally Likely Events or Mutually Symmetric Events or Equi-Probable Events:** The events of a random experiment are known to be equally likely when all necessary evidence are taken into account, no event is expected to occur more frequently as compared to the other events of the set of events.
- The probability of occurrence of the event A is defined as the ratio of the number of events favourable to A to the total number of events. Denoting this by P(A), we have

$$P(A) = \frac{n_A}{n} = \frac{\text{No. of equally likely events favourable to A}}{\text{Total no. of equally likely events}}$$

(a) The probability of an event lies between 0 and 1, both inclusive.

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

When P(A) = 0, A is known to be an impossible event and when P(A) = 1, A is known to be a sure event.

(b) Non-occurrence of event A is denoted by A′ or A^C or $\bar{A}$ and it is known as complimentary event of A. The event A along with its complimentary A′ forms a set of mutually exclusive and exhaustive events.

i.e. $P(A) + P(A') = 1$

$\Rightarrow \quad P(A') = 1 - P(A)$

$$1 - \frac{m_A}{m}$$

$$= \frac{m - m_A}{m}$$

(c) The ratio of no. of favourable events to the no. of unfavourable events is known as odds in favour of the event A and its inverse ratio is known as odds against the event A.

i.e. odds in favour of A $= m_A : (m - m_A)$

and odds against A $= (m - m_A) : m_A$

(d) For any two mutually exclusive events A and B, the probability that either A or B occurs is given by the sum of individual probabilities of A and B.

i.e. $P(A \cup B)$

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

(e) For any $K(\geq 2)$ mutually exclusive events $A_1, A_2, A_3 \ldots, A_K$ the probability that at least one of them occurs is given by the sum of the individual probabilities of the K events.

i.e. $P(A_1 \cup A_2 \cup \ldots \cup A_K) = P(A_1) + P(A_2) + \ldots. P(A_K)$

(f) For any two events A and B, the probability that either A or B occurs is given by the sum of individual probabilities of A and B less the probability of simultaneous occurrence of the events A and B.

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

(g) For any three events A, B and C, the probability that at least one of the events occurs is given by

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

(h) For any two events A and B, the probability that A and B occur simultaneously is given by the product of the unconditional probability of A and the conditional probability of B given that A has already occurred

i.e. $P(A \cap B) = P(A) \times P(B/A)$ Provided $P(A) > 0$

(i) Compound Probability or Joint Probability

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

(j) For any three events A, B and C, the probability that they occur jointly is given by $P(A \cap B \cap C) = P(A) \times P(B/A) \times P(C/(A \cap B))$ Provided $P(A \cap B) > 0$

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

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

$$P(A'/B) = \frac{P(A' \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)}$$

(l) $$P(A/B') = \frac{P(A \cap B')}{P(B')} = \frac{P(A) - P(A \cap B)}{1 - P(B)}$$

(m) $$P(A'/B') = \frac{P(A' \cap B')}{P(B')}$$

$$= \frac{P(A \cup B)'}{P(B')}$$ [by De-Morgan's Law $A' \cap B' = (AUB)'$]

$$= \frac{1 - P(A \cup B)}{1 - P(B)}$$

- A random variable or stochastic variable is a function defined on a sample space associated with a random experiment assuming any value from R and assigning a real number to each and every sample point of the random experiment.
- Expected value or Mathematical Expectation or Expectation of a random variable may be defined as the sum of products of the different values taken by the random variable and the corresponding probabilities.

  When x is a discrete random variable with probability mass function f(x), then its expected value is given by

  $$\mu = \sum_x x f(x)$$

  and its variance is

  $$\sigma^2 = E(x^2) - \mu^2$$

  Where $E(x^2) = \sum_x x^2 f(x)$

  For a continuous random variable x defined in [–∞, ∞], its expected value (i.e. mean) and variance are given by

  $$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

  $$\text{and } \sigma^2 = E(x^2) - \mu^2$$

  $$\text{where } E(x^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

  Properties of Expected Values

  (i) Expectation of a constant k is k

  i.e. E(k) = k for any constant k.

  (ii) Expectation of sum of two random variables is the sum of their expectations.

  i.e. E(x + y) = E(x) + E(y) for any two random variables x and y.

  (iii) Expectation of the product of a constant and a random variable is the product of the constant and the expectation of the random variable.

  i.e. E(k x) = k.E(x) for any constant k

  (iv) Expectation of the product of two random variables is the product of the expectation of the two random variables, provided the two variables are independent.

  i.e. $E(xy) = E(x) \times E(y)$

  Whenever x and y are independent.

EXERCISE

Set A

Write down the correct answers. Each question carRies 1 mark.

1. Initially, probability was a branch of
   (a) Physics (b) Statistics
   (c) Mathematics (d) Economics.
2. Two broad divisions of probability are
   (a) Subjective probability and objective probability
   (b) Deductive probability and non–deductive probability
   (c) Statistical probability and Mathematical probability
   (d) None of these.
3. Subjective probability may be used in
   (a) Mathematics (b) Statistics
   (c) Management (d) Accountancy.
4. An experiment is known to be random if the results of the experiment
   (a) Can not be predicted (b) Can be predicted
   (c) Can be split into further experiments (d) Can be selected at random.
5. An event that can be split into further events is known as
   (a) Complex event (b) Mixed event
   (c) Simple event (d) Composite event.
6. Which of the following pairs of events are mutually exclusive?
   (a) A : The student reads in a school. B : He studies Philosophy.
   (b) A : Raju was born in India. B : He is a fine Engineer.
   (c) A : Ruma is 16 years old. B : She is a good singer.
   (d) A : Peter is under 15 years of age. B : Peter is a voter of Kolkata.
7. If P(A) = P(B), then
   (a) A and B are the same events (b) A and B must be same events
   (c) A and B may be different events (d) A and B are mutually exclusive events.
8. If $P(A \cap B) = 0$, then the two events A and B are
   (a) Mutually exclusive (b) Exhaustive
   (c) Equally likely (d) Independent.

9. If for two events A and B, P(AUB) = 1, then A and B are
   (a) Mutually exclusive events (b) Equally likely events
   (c) Exhaustive events (d) Dependent events.
10. If an unbiased coin is tossed once, then the two events Head and Tail are
   (a) Mutually exclusive (b) Exhaustive
   (c) Equally likely (d) All these (a), (b) and (c).
11. If P(A) = P(B), then the two events A and B are
   (a) Independent (b) Dependent
   (c) Equally likely (d) Both (a) and (c).
12. If for two events A and B, $P(A \cap B) \neq P(A) \times P(B)$, then the two events A and B are
   (a) Independent (b) Dependent
   (c) Not equally likely (d) Not exhaustive.
12. If P(A/B) = P(A), then
   (a) A is independent of B (b) B is independent of A
   (c) B is dependent of A (d) Both (a) and (b).
14. If two events A and B are independent, then
   (a) A and the complement of B are independent
   (b) B and the complement of A are independent
   (c) Complements of A and B are independent
   (d) All of these (a), (b) and (c).
15. If two events A and B are independent, then
   (a) They can be mutually exclusive (b) They can not be mutually exclusive
   (c) They can not be exhaustive (d) Both (b) and (c).
16. If two events A and B are mutually exclusive, then
   (a) They are always independent (b) They may be independent
   (c) They can not be independent (d) They can not be equally likely.
17. If a coin is tossed twice, then the events 'occurrence of one head', 'occurrence of 2 heads' and 'occurrence of no head' are
   (a) Independent (b) Equally likely
   (c) Not equally likely (d) Both (a) and (b).
18. The probability of an event can assume any value between
   (a) – 1 and 1 (b) 0 and 1
   (c) – 1 and 0 (d) none of these.

19. If P(A) = 0, then the event A

(a) will never happen (b) will always happen

(c) may happen (d) may not happen.

20. If P(A) = 1, then the event A is known as

(a) symmetric event (b) dependent event

(c) improbable event (d) sure event.

21. If p : q are the odds in favour of an event, then the probability of that event is

(a) $\frac{p}{q}$ (b) $\frac{p}{p+q}$

(c) $\frac{q}{p+q}$ (d) none of these.

22. If P(A) = 5/9, then the odds against the event A is

(a) 5 : 9 (b) 5 : 4

(c) 4 : 5 (d) 5 : 14

23. If A, B and C are mutually exclusive and exhaustive events, then P(A) + P(B) + P(C) equals to

(a) $\frac{1}{3}$ (b) 1

(c) 0 (d) any value between 0 and 1.

24. If A denotes that a student reads in a school and B denotes that he plays cricket, then

(a) $P(A \cap B) = 1$ (b) $P(A \cup B) = 1$

(c) $P(A \cap B) = 0$ (d) P(A) = P(B).

25. P(B/A) is defined only when

(a) A is a sure event (b) B is a sure event

(c) A is not an impossible event (d) B is an impossible event.

26. P(A/B') is defined only when

(a) B is not a sure event (b) B is a sure event

(c) B is an impossible event (d) B is not an impossible event.

27. For two events A and B, $P(A \cup B) = P(A) + P(B)$ only when

(a) A and B are equally likely events (b) A and B are exhaustive events

(c) A and B are mutually independent (d) A and B are mutually exclusive.

28. Addition Theorem of Probability states that for any two events A and B,
    (a) $P(A \cup B) = P(A) + P(B)$
    (b) $P(A \cup B) = P(A) + P(B) + P(A \cap B)$
    (c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    (d) $P(A \cup B) = P(A) \times P(B)$
29. For any two events A and B,
    (a) $P(A) + P(B) > P(A \cap B)$
    (b) $P(A) + P(B) < P(A \cap B)$
    (c) $P(A) + P(B) \geq P(A \cap B)$
    (d) $P(A) \text{ x } P(B) \leq P(A \cap B)$
30. For any two events A and B,
    (a) $P(A–B) = P(A) – P(B)$
    (b) $P(A–B) = P(A) – P(A \cap B)$
    (c) $P(A–B) = P(B) – P(A \cap B)$
    (d) $P(B–A) = P(B) + P(A \cap B)$.
31. The limitations of the classical definition of probability
    (a) it is applicable when the total number of elementary events is finite
    (b) it is applicable if the elementary events are equally likely
    (c) it is applicable if the elementary events are mutually independent
    (d) (a) and (b).
32. According to the statistical definition of probability, the probability of an event A is the
    (a) limiting value of the ratio of the no. of times the event A occurs to the number of times the experiment is repeated
    (b) the ratio of the frequency of the occurrences of A to the total frequency
    (c) the ratio of the frequency of the occurrences of A to the non-occurrence of A
    (d) the ratio of the favourable elementary events to A to the total number of elementary events.
33. The Theorem of Compound Probability states that for any two events A and B.
    (a) $P(A \cap B) = P(A) \times P(B/A)$
    (b) $P(A \cup B) = P(A) \times P(B/A)$
    (c) $P(A \cap B) = P(A) \times P(B)$
    (d) $P(A \cup B) = P(B) + P(B) – P(A \cap B)$.
34. If A and B are mutually exclusive events, then
    (a) $P(A) = P(A–B)$.
    (b) $P(B) = P(A–B)$.
    (c) $P(A) = P(A \cap B)$.
    (d) $P(B) = P(A \cap B)$.
35. If $P(A–B) = P(B–A)$, then the two events A and B satisfy the condition
    (a) $P(A) = P(B)$.
    (b) $P(A) + P(B) = 1$
    (c) $P(A \cap B) = 0$
    (d) $P(A \cup B) = 1$
36. The number of conditions to be satisfied by three events A, B and C for complete independence is
    (a) M2
    (b) 3
    (c) 4
    (d) any number.

37. If two events A and B are independent, then $P(A \cap B)$

(a) equals to $P(A) + P(B)$
(b) equals to $P(A) \times P(B)$
(c) equals to $P(A) \times P(B/A)$
(d) equals to $P(B) \times P(A/B)$.

38. Values of a random variable are

(a) always positive numbers.
(b) always positive real numbers.
(c) real numbers.
(d) natural numbers.

39. Expected value of a random variable

(a) is always positive
(b) may be positive or negative
(c) may be positive or negative or zero
(d) can never be zero.

40. If all the values taken by a random variable are equal then

(a) its expected value is zero
(b) its standard deviation is zero
(c) its standard deviation is positive
(d) its standard deviation is a real number.

41. If x and y are independent, then

(a) $E(xy) = E(x) \times E(y)$
(b) $E(xy) = E(x) + E(y)$
(c) $E(x - y) = E(x) + E(y)$
(d) $E(x - y) = E(x) + x\,E(y)$

42. If a random variable x assumes the values x_1, x_2, x_3, x_4 with corresponding probabilities p_1, p_2, p_3, p_4 then the expected value of x is

(a) $p_1 + p_2 + p_3 + p_4$
(b) $x_1 p_1 + x_2 p_3 + x_3 p_2 + x_4 p_4$
(c) $p_1 x_1 + p_2 x_2 + p_3 x_3 + p_4 x_4$
(d) none of these.

43. f(x), the probability mass function of a random variable x satisfies

(a) $f(x) > 0$
(b) $\sum_x f(x) = 1$
(c) both (a) and (b)
(d) $f(x) \geq 0$ and $\sum_x f(x) = 1$

44. Variance of a random variable x is given by

(a) $E(x - \mu)^2$
(b) $E[x - E(x)]^2$
(c) $E(x^2 - \mu)$
(d) (a) or (b)

45. If two random variables x and y are related by $y = 2 - 3x$, then the SD of y is given by

(a) $-3 \times$ SD of x
(b) $3 \times$ SD of x.
(c) $9 \times$ SD of x
(d) $2 \times$ SD of x.

46. Probability of getting a head when two unbiased coins are tossed simultaneously is

(a) 0.25
(b) 0.50
(c) 0.20
(d) 0.75

47. If an unbiased coin is tossed twice, the probability of obtaining at least one tail is

(a) 0.25
(b) 0.50
(c) 0.75
(d) 1.00

48. If an unbiased die is rolled once, the odds in favour of getting a point which is a multiple of 3 is

(a) 1:2 (b) 2:1

(c) 1:3 (d) 3:1

49. A bag contains 15 one rupee coins, 25 two rupee coins and 10 five rupee coins. If a coin is selected at random from the bag, then the probability of not selecting a one rupee coin is

(a) 0.30 (b) 0.70

(c) 0.25 (d) 0.20

50. A, B, C are three mutually independent with probabilities 0.3, 0.2 and 0.4 respectively. What is $P(A \cap B \cap C)$?

(a) 0.400 (b) 0.240

(c) 0.024 (d) 0.500

51. If two letters are taken at random from the word HOME, what is the Probability that none of the letters would be vowels?

(a) 1/6 (b) 1/2

(c) 1/3 (d) 1/4

52. If a card is drawn at random from a pack of 52 cards, what is the chance of getting a Spade or an ace?

(a) 4/13 (b) 5/13

(c) 0.25 (d) 0.20

53. If x and y are random variables having expected values as 4.5 and 2.5 respectively, then the expected value of (x–y) is

(a) 2 (b) 7

(c) 6 (d) 0

54. If variance of a random variable x is 23, then what is the variance of 2x+10?

(a) 56 (b) 33

(c) 46 (d) 92

55. What is the probability of having at least one 'six' from 3 throws of a perfect die?

(a) 5/6 (b) $(5/6)^3$

(c) $1-(1/6)^3$ (d) $1-(5/6)^3$

Set B

Write down the correct answers. Each question carries 2 marks.

1. Two balls are drawn from a bag containing 5 white and 7 black balls at random. What is the probability that they would be of different colours?

(a) 35/66 (b) 30/66

(c) 12/66 (d) None of these

2. What is the chance of throwing at least 7 in a single cast with 2 dice?

(a) 5/12 (b) 7/12

(c) 1/4 (d) 17/36

3. What is the chance of getting at least one defective item if 3 items are drawn randomly from a lot containing 6 items of which 2 are defective item?

(a) 0.30 (b) 0.20

(c) 0.80 (d) 0.50

4. If two unbiased dice are rolled together, what is the probability of getting no difference of points?

(a) 1/2 (b) 1/3

(c) 1/5 (d) 1/6

5. If A, B and C are mutually exclusive independent and exhaustive events then what is the probability that they occur simultaneously?

(a) 1 (b) 0.50

(c) 0 (d) any value between 0 and 1

6. There are 10 balls numbered from 1 to 10 in a box. If one of them is selected at random, what is the probability that the number printed on the ball would be an odd number greater that 4?

(a) 0.50 (b) 0.40

(c) 0.60 (d) 0.30

7. Following are the wages of 8 workers in rupees:

50, 62, 40, 70, 45, 56, 32, 45

If one of the workers is selected at random, what is the probability that his wage would be lower than the average wage?

(a) 0.625 (b) 0.500

(c) 0.375 (d) 0.450

8. A, B and C are three mutually exclusive and exhaustive events such that P (A) = 2 P (B) = 3P(C). What is P (B)?

(a) 6/11 (b) 3/11

(c) 1/6 (d) 1/3

9. For two events A and B, P (B) = 0.3, P (A but not B) = 0.4 and P (not A) = 0.6. The events A and B are

(a) exhaustive (b) independent

(c) equally likely (d) mutually exclusive

10. A bag contains 12 balls which are numbered from 1 to 12. If a ball is selected at random, what is the probability that the number of the ball will be a multiple of 5 or 6 ?

(a) 0.30 (b) 0.25

(c) 0.20 (d) 1/3

11. Given that for two events A and B, P (A) = 3/5, P (B) = 2/3 and P ($A \cup B$) = 3/4, what is P (A/B)?

(a) 0.655 (b) 13/60

(c) 31/60 (d) 0.775

12. For two independent events A and B, what is P (A+B), given P(A) = 3/5 and P(B) = 2/3?

(a) 11/15 (b) 13/15

(c) 7/15 (d) 0.65

12. If P (A) = p and P (B) = q, then

(a) P(A/B) ≤ p/q (b) P(A/B)≤ p/q

(c) P(A/B) ≤ q/p (d) None of these

14. If P ($\bar{A} \cup \bar{B}$) = 5/6, P(A) = ½ and P ($\bar{B}$) = 2/3, , what is P ($A \cup B$) ?

(a) 1/3 (b) 5/6

(c) 2/3 (d) 4/9

15. If for two independent events A and B, P ($A \cup B$) = 2/3 and P (A) = 2/5, what is P (B)?

(a) 4/15 (b) 4/9

(c) 5/9 (d) 7/15

16. If P (A) = 2/3, P (B) =3/4, P (A/B) = 2/3, then what is P (B / A)?

(a) 1/3 (b) 2/3

(c) 3/4 (d) 1/2

17. If P (A) = a, P (B) = b and P (P ($A \cap B$) = c then the expression of P ($A' \cap B'$) in terms of a, b and c is

(a) 1 – a – b – c (b) a + b – c

(c) 1 + a – b – c (d) 1 – a – b + c

18. For three events A, B and C, the probability that only A occur is

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

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

19. It is given that a family of 2 children has a girl, what is the probability that the other child is also a girl ?

(a) 0.50 (b) 0.75

(c) 1/3 (d) 2/3

20. Two coins are tossed simultaneously. What is the probability that the second coin would show a tail given that the first coin has shown a head?

(a) 0.50 (b) 0.25

(c) 0.75 (d) 0.125

21. If a random variable x assumes the values 0, 1 and 2 with probabilities 0.30, 0.50 and 0.20, then its expected value is

(a) 1.50 (b) 3

(c) 0.90 (d) 1

22. If two random variables x and y are related as y = –3x + 4 and standard deviation of x is 2, then the standard deviation of y is

(a) – 6 (b) 6

(c) 18 (d) 3.50

23. If 2x + 3y + 4 = 0 and v(x) = 6 then v (y) is

(a) 8/3 (b) 9

(c) – 9 (d) 6

Set C

Write down the correct answers. Each question carries 5 marks.

1. What is the probability that a leap year selected at random would contain 53 Saturdays?

(a) 1/7 (b) 2/7

(c) 1/12 (d) 1/4

2. If an unbiased coin is tossed three times, what is the probability of getting more that one head?

(a) 1/8 (b) 3/8

(c) 1/2 (d) 1/3

3. If two unbiased dice are rolled, what is the probability of getting points neither 6 nor 9?

(a) 0.25 (b) 0.50

(c) 0.75 (d) 0.80

4. What is the probability that 4 children selected at random would have different birthdays?

(a) $\frac{364 \times 363 \times 362}{(365)^3}$ (b) $\frac{6 \times 5 \times 4}{7^3}$

(c) 1/365 (d) $(1/7)^3$

5. A box contains 5 white and 7 black balls. Two successive drawn of 3 balls are made (i) with replacement (ii) without replacement. The probability that the first draw would produce white balls and the second draw would produce black balls are respectively

(a) 6/321 and 3/926 (b) 1/20 and 1/30

(c) 35/144 and 35/108 (d) 7/968 and 5/264

6. There are three boxes with the following composition:

Box I: 5 Red + 7 White + 6 Blue balls Box II: 4 Red + 8 White + 6 Blue balls

Box III: 3 Red + 4 White + 2 Blue balls

If one ball is drawn at random, then what is the probability that they would be of same colour?

(a) 89/729 (b) 97/729

(c) 82/729 (d) 23/32

7. A number is selected at random from the first 1000 natural numbers. What is the probability that the number so selected would be a multiple of 7 or 11?

(a) 0.25 (b) 0.32

(c) 0.22 (d) 0.33

8. A bag contains 8 red and 5 white balls. Two successive draws of 3 balls are made without replacement. The probability that the first draw will produce 3 white balls and the second 3 red balls is

(a) 5/223 (b) 6/257

(c) 7/429 (d) 3/548

9. There are two boxes containing 5 white and 6 blue balls and 3 white and 7 blue balls respectively. If one of the the boxes is selected at random and a ball is drawn from it, then the probability that the ball is blue is

(a) 115/227 (b) 83/250

(c) 137/220 (d) 127/250

10. A problem in probability was given to three CA students A, B and C whose chances of solving it are 1/3, 1/5 and 1/2 respectively. What is the probability that the problem would be solved?

(a) 4/15 (b) 7/8

(c) 8/15 (d) 11/15

11. There are three persons aged 60, 65 and 70 years old. The survival probabilities for these three persons for another 5 years are 0.7, 0.4 and 0.2 respectively. What is the probability that at least two of them would survive another five years?

(a) 0.425 (b) 0.456

(c) 0.392 (d) 0.388

12. Tom speaks truth in 30 percent cases and Dick speaks truth in 25 percent cases. What is the probability that they would contradict each other?

(a) 0.325 (b) 0.400

(c) 0.925 (d) 0.075

13. There are two urns. The first urn contains 3 red and 5 white balls whereas the second urn contains 4 red and 6 white balls. A ball is taken at random from the first urn and is transferred to the second urn. Now another ball is selected at random from the second arm. The probability that the second ball would be red is

(a) 7/20 (b) 35/88

(c) 17/52 (d) 3/20

14. For a group of students, 30 %, 40% and 50% failed in Physics , Chemistry and at least one of the two subjects respectively. If an examinee is selected at random, what is the probability that he passed in Physics if it is known that he failed in Chemistry?

(a) 1/2 (b) 1/3

(c) 1/4 (d) 1/6

15. A packet of 10 electronic components is known to include 2 defectives. If a sample of 4 components is selected at random from the packet, what is the probability that the sample does not contain more than 1 defective?

(a) 1/3 (b) 2/3

(c) 13/15 (d) 3/15

16. 8 identical balls are placed at random in three bags. What is the probability that the first bag will contain 3 balls?

(a) 0.2731 (b) 0.3256

(c) 0.1924 (d) 0.3443

17. X and Y stand in a line with 6 other people. What is the probability that there are 3 persons between them?

(a) 1/5 (b) 1/6

(c) 1/7 (d) 1/3

18. Given that P (A) = 1/2, P (B) = 1/3, P $(A \cap B)$ = 1/4, what is P (A′/B′)

(a) 1/2 (b) 7/8

(c) 5/8 (d) 2/3

19. Four digits 1, 2, 4 and 6 are selected at random to form a four digit number. What is the probability that the number so formed, would be divisible by 4?

(a) 1/2 (b) 1/5

(c) 1/4 (d) 1/3

20. The probability distribution of a random variable x is given below:

x :	1	2	4	5	6
P :	0.15	0.25	0.20	0.30	0.10

What is the standard deviation of x?

(a) 1.49 (b) 1.56

(c) 1.69 (d) 1.72

21. A packet of 10 electronic components is known to include 3 defectives. If 4 components are selected from the packet at random, what is the expected value of the number of defective?

(a) 1.20 (b) 1.21

(c) 1.69 (d) 1.72

22. The probability that there is at least one error in an account statement prepared by 3 persons A, B and C are 0.2, 0.3 and 0.1 respectively. If A, B and C prepare 60, 70 and 90 such statements, then the expected number of correct statements

(a) 170 (b) 176

(c) 178 (d) 180

23. A bag contains 6 white and 4 red balls. If a person draws 2 balls and receives ₹ 10 and ₹ 20 for a white and red balls respectively, then his expected amount is

(a) ₹ 25 (b) ₹ 26

(c) ₹ 29 (d) ₹ 28

24. The probability distribution of a random variable is as follows:

x :	1	2	4	6	8
P :	k	2k	3k	3k	k

The variance of x is

(a) 2.1 (b) 4.41

(c) 2.32 (d) 2.47

ANSWERS

Set A

1.	(c)	2.	(a)	3.	(c)	4.	(d)	5.	(d)	6.	(d)
7.	(c)	8.	(a)	9.	(c)	10.	(d)	11.	(c)	12.	(b)
12.	(d)	14.	(d)	15.	(b)	16.	(c)	17.	(c)	18.	(b)
19.	(a)	20.	(d)	21.	(b)	22.	(c)	23.	(b)	24.	(c)
25.	(c)	26.	(a)	27.	(d)	28.	(c)	29.	(c)	30.	(b)
31.	(d)	32.	(a)	33.	(a)	34.	(a)	35.	(a)	36.	(c)
37.	(b)	38.	(c)	39.	(c)	40	(b)	41.	(a)	42.	(c)
43.	(d)	44.	(d)	45.	(b)	46.	(b)	47.	(c)	48.	(a)
49.	(b)	50.	(c)	51.	(a)	52.	(a)	53.	(a)	54.	(d)
55.	(d)										

Set B

1.	(a)	2.	(b)	3.	(c)	4.	(d)	5.	(c)	6.	(d)
7.	(b)	8.	(b)	9.	(d)	10.	(d)	11.	(d)	12.	(b)
12.	(a)	14.	(c)	15.	(b)	16.	(c)	17.	(d)	18.	(d)
19.	(c)	20.	(a)	21.	(c)	22.	(b)	23.	(a)		

Set C

1.	(b)	2.	(c)	3.	(c)	4.	(a)	5.	(d)	6.	(a)
7.	(c)	8.	(c)	9.	(c)	10.	(d)	11.	(d)	12.	(b)
12.	(b)	14.	(a)	15.	(c)	16.	(a)	17.	(c)	18.	(c)
19.	(d)	20.	(c)	21.	(a)	22.	(c)	23.	(d)	24.	(b)

ADDITIONAL QUESTION BANK

1. All possible outcomes of a random experiment forms the
   (a) events (b) sample space (c) both (d) none
2. If one of outcomes cannot be expected to occur in preference to the other in an experiment the events are
   (a) simple events (b) compound events
   (c) favourable events (d) equally likely events
3. If two events cannot occur simultaneously in the same trial then they are
   (a) mutually exclusive events (b) simple events
   (c) favourable events (d) none
4. When the number of cases favourable to the event A is none then P(A) is equal to
   (a) 1 (b) 0 (c) $\frac{1}{2}$ (d) none
5. A card is drawn from a well-shuffled pack of playing cards. The probability that it is a spade is
   (a) $\frac{1}{13}$ (b) $\frac{1}{4}$ (c) $\frac{3}{13}$ (d) none
6. A card is drawn from a well-shuffled pack of playing cards. The probability that it is a king is
   (a) $\frac{1}{13}$ (b) $\frac{1}{4}$ (c) $\frac{4}{13}$ (d) none
7. A card is drawn from a well-shuffled pack of playing cards. The probability that it is the ace of clubs is
   (a) $\frac{1}{13}$ (b) $\frac{1}{4}$ (c) $\frac{1}{52}$ (d) none
8. In a single throw with two dice the probability of getting a sum of five on the two dice is
   (a) $\frac{1}{9}$ (b) $\frac{5}{36}$ (c) $\frac{5}{9}$ (d) none
9. In a single throw with two dice, the probability of getting a sum of six on the two dice is
   (a) $\frac{1}{9}$ (b) $\frac{5}{36}$ (c) $\frac{5}{9}$ (d) none
10. The probability that exactly one head appears in a single throw of two fair coins is
    (a) $\frac{3}{4}$ (b) $\frac{1}{2}$ (c) $\frac{1}{4}$ (d) none
11. The probability that at least one head appears in a single throw of three fair coins is
    (a) $\frac{1}{8}$ (b) $\frac{7}{8}$ (c) $\frac{1}{3}$ (d) none
12. The definition of probability fails when the no of possible outcomes of the experiment is infinite
    (a) True (b) false (c) both (d) none

13. The following table gives distribution of wages of 100 workers –

Wages (in ₹)	120–140	140–160	160–180	180–200	200–220	220–240	240–260
No. of workers	9	20	0	10	8	35	18

The probability that his wages are under ₹ 140 is

(a) 20/100 (b) 9/100 (c) 29/100 (d) none

14. An individual is selected at random from the above group. The probability that his wages are under ₹160 is

(a) 9/100 (b) 20/100 (c) 29/100 (d) none

15. For the above table the probability that his wages are above ₹ 200 is

(a) 43/100 (b) 35/100 (c) 53/100 (d) 61/100

16. For the above table the probability that his wages between ₹ 160 and 220 is

(a) 30/100 (b) 10/100 (c) 38/100 (d) 18/100

17. The table below shows the history of 1000 men :

Life (in years) :	60	70	80	90
No. survived :	1000	500	100	60

The probability that a man will survived to age 90 is

(a) 60/1000 (b) 160/1000 (c) 660/1000 (d) none

18. The terms "chance" and probability are synonymous

(a) True (b) false (c) both (d) none

19. If probability of drawing a spade from a well-shuffled pack of playing cards is $\frac{1}{4}$ then the probability that of the card drawn from a well-shuffled pack of playing cards is 'not a spade' is

(a) 1 (b) $\frac{1}{2}$ (c) $\frac{1}{4}$ (d) $\frac{3}{4}$

20. Probability of the sample space is

(a) 0 (b) $\frac{1}{2}$ (c) 1 (d) none

21. Sum of all probabilities of mutually exclusive and exhaustive events is equal to

(a) 0 (b) $\frac{1}{2}$ (c) $\frac{3}{4}$ (d) 1

22. Let a sample space be S = $\{X_1, X_2, X_3\}$ which of the fallowing defines probability space on S ?

(a) $P(X_1)= \frac{1}{4}$, $P(X_2)= \frac{1}{3}$, $P(X_3)= \frac{1}{3}$ (b) $P(X_1)= 0$, $P(X_2)= \frac{1}{3}$, $P(X_3)= \frac{2}{3}$

(c) $P(X_1)= \frac{2}{3}$, $P(X_2)= \frac{1}{3}$, $P(X_3)= \frac{2}{3}$ (d) none

23. Let P be a probability function on S = $\{X_1, X_2, X_3\}$ if $P(X_1) = \frac{1}{4}$ and $P(X_3) = \frac{1}{3}$ then $P(X_2)$ is equal to

(a) 5/12 (b) 7/12 (c) 3/4 (d) none

24. The chance of getting a sum of 10 in a single throw with two dice is

(a) 10/36 (b) 1/12 (c) 5/36 (d) none

25. The chance of getting a sum of 6 in a single throw with two dice is

(a) 3/36 (b) 4/36 (c) 6/36 (d) 5/36

26. P (B/A) defines the probability that event B occurs on the assumption that A has happened

(a) Yes (b) no (c) both (d) none

27. The complete group of all possible outcomes of a random experiment given an ________ set of events.

(a) mutually exclusive (b) exhaustive (c) both (d) none

28. When the event is 'certain' the probability of it is

(a) 0 (b) 1/2 (c) 1 (d) none

29. The classical definition of probability is based on the feasibility at subdividing the possible outcomes of the experiments into

(a) mutually exclusive and exhaustive

(b) mutually exclusive and equally likely

(c) exhaustive and equally likely

(d) mutually exclusive,exhaustive and equally likely cases.

30. Two unbiased coins are tossed. The probability of obtaining 'both heads' is

(a) $\frac{1}{4}$ (b) $\frac{2}{4}$ (c) $\frac{3}{4}$ (d) none

31. Two unbiased coins are tossed. The probability of obtaining one head and one tail is

(a) $\frac{1}{4}$ (b) $\frac{2}{4}$ (c) $\frac{3}{4}$ (d) none

32. Two unbiased coins are tossed. The probability of obtaining both tail is

(a) $\frac{2}{4}$ (b) $\frac{3}{4}$ (c) $\frac{1}{4}$ (d) none

33. Two unbiased coins are tossed. The probability of obtaining at least one head is

(a) $\frac{1}{4}$ (b) $\frac{2}{4}$ (c) $\frac{3}{4}$ (d) none

34. When two unbiased coins are tossed, the probability of obtaining 3 heads is

(a) $\frac{2}{4}$ (b) $\frac{1}{4}$ (c) $\frac{3}{4}$ (d) 0

35. When two unbiased coins are tossed, the probability of obtaining not more than 3 heads is

(a) $\frac{3}{4}$ (b) $\frac{1}{2}$ (c) 1 (d) 0

36. When two unbiased coins are tossed, the probability of getting both heads or both tails is

(a) $\frac{1}{2}$ (b) $\frac{3}{4}$ (c) $\frac{1}{4}$ (d) none

37. Two dice with face marked 1, 2, 3, 4, 5, 6 are thrown simultaneously and the points on the dice are multiplied together. The probability that product is 12 is

(a) 4/36 (b) 5/36 (c) 12/36 (d) none

38. A bag contain 6 white and 5 black balls. One ball is drawn. The probability that it is white is

(a) 5/11 (b) 1 (c) 6/11 (d) 1/11

39. Probability of occurrence of at least one of the events A and B is denoted by

(a) P(AB) (b) P(A+B) (c) P(A/B) (d) none

40. Probability of occurrence of A as well as B is denoted by

(a) P(AB) (b) P(A+B) (c) P(A/B) (d) none

41. Which of the following relation is true ?

(a) $P(A)- P(A^C)= 1$ (b) $P(A)+ P(A^C)= 1$ (c) $P(A)\ P(A^C)= 1$ (d) none

42. If events A and B are mutually exclusive, the probability that either A or B occurs is given by

a) P (A+B)= P(A)– P(B) (b) P (A+B)=P(A)+ P(B)– P(AB)

c) P (A+B)= P(A)– P(B)+ P(AB) (d) P (A+B)= P(A)+ P(B)

43. The probability of occurrence of at least one of the 2 events A and B (which may not be mutually exclusive) is given by

a) P(A+B)= P(A)– P(B) (b) P(A+B)= P(A)+ P(B)– P(AB)

c) P(A+B)= P(A)– P(B)+ P(AB) (d) P(A+B)= P(A)+P (B)

44. If events A and B are independent, the probability of occurrence of A as well as B is given by

(a) P(AB)= P(A/B) (b) P(AB)= P(A)/ P(B)

(c) P(AB)= P(A)P(B) (d) None

45. For the condition P(AB)= P(A)P(B)two events A and B are said to be

(a) dependent (b) independent (c) equally like (d) none

46. The conditional probability of an event B on the assumption that another event A has actually occurred is given by

(a) P(B/A)= P(AB)/P(A) (b) P(A/B)= P(AB)/ P(B)

(c) P(B/A)= P(AB) (d) P(A/B)= P(AB)/ P(A)P(B)

47. Given P(A)= $\frac{1}{2}$, P(B)= $\frac{1}{3}$, P(AB)= $\frac{1}{4}$, the value of P(A+B) is

a) $\frac{3}{4}$ b) $\frac{7}{12}$ c) $\frac{5}{6}$ d) $\frac{1}{6}$

48. Given P(A)= $\frac{1}{2}$, P(B)= $\frac{1}{3}$, P (AB)= $\frac{1}{4}$, the value of P (A/B) is

(a) $\frac{1}{2}$ (b) $\frac{1}{6}$ (c) $\frac{2}{3}$ (d) $\frac{3}{4}$

49. If P (A)= $\frac{1}{3}$, P(B)= $\frac{1}{4}$, the events A & B are

a) not equally likely b) mutually exclusive
c) equally likely d) none

50. If events A and B are independent then

a) A^C and B^C are dependent b) A^C and B are dependent
c) A and B^C are dependent d) A^C and B^C are also independent

51. A card is drown from each of two well-shuffled packs of cards.The probability that at least one of them is an ace is

a) $\frac{1}{69}$ b) $\frac{25}{169}$ c) $\frac{2}{13}$ d) none

52. When a die is tossed, the sample space is

a) S =(1,2,3,4,5) b) S =(1,2,3,4) c) S =(1,2,3,4,5,6) d) none

53. If P (A)= $\frac{1}{4}$, P(B)= $\frac{2}{5}$, P(A+B)= $\frac{1}{2}$ then P(AB)is equal to

a) $\frac{3}{4}$ b) $\frac{2}{20}$ c) $\frac{13}{20}$ d) $\frac{3}{20}$

54. If events A and B are independent and P(A)= 2/3 , P(B)= 3/5 then P(A+B)is equal to

a) $\frac{13}{15}$ b) $\frac{6}{15}$ c) $\frac{1}{15}$ d) none

55. The expected number of head in 100 tosses of an unbiased coin is

a) 100 b) 50 c) 25 d) none

56. A and B are two events such that P(A)= $\frac{1}{3}$, P(B) = $\frac{1}{4}$, P(A+B)= $\frac{1}{2}$, than P(B/A) is equal to

a) $\frac{1}{4}$ b) $\frac{1}{3}$ c) $\frac{1}{2}$ d) none

57. Probability mass function is always

a) 0 b) greater than 0
c) greater than equal to 0 d) less than 0

58. The sum of probability mass function is equal to

a) –1 b) 0 c) 1 d) none

59. When X is a continues function f(x)is called

a) probability mass function b) probability density function
c) both d) none

60. Which of the following set of function define a probability space on S = {a_1, a_2, a_3}

a) $P(a_1)= \frac{1}{3}, P(a_2) = \frac{1}{2}, P(a_3)= \frac{1}{4}$ b) $P(a_1)= \frac{1}{3}, P(a_2)= \frac{1}{6}, P(a_3)= \frac{1}{2}$

c) $P(a_1)= P(a_2)= \frac{2}{3}, P(a_3)= \frac{1}{4}$ d) None

61. If $P(a_1)= 0, P(a_2)= \frac{1}{3}, P(a_3) = \frac{2}{3}$ then S = {a_1, a_2, a_3} is a probability space

a) true b) false c) both d) none

62. If two events are independent then

a) P(B/A)= P(AB) P(A) b) P(B/A)= P(AB) P(B)

c) P(B/A)= P(B) d) P(B/A)P(A)

63. When expected value is negative the result is

a) favourable b) unfavourable

c) both d) none to the above

64. The expected value of X, the sum of the scores, when two dice are rolled is

a) 9 b) 8 c) 6 d) 7

65. Let A and B be the events with P(A)= 1/3, P(B) = 1/4 and P(AB)= 1/12 then P(A/B) is equal to

a) $\frac{1}{3}$ b) $\frac{1}{4}$ c) $\frac{3}{4}$ d) $\frac{2}{3}$

66. Let A and B be the events with P(A)= 2/3, P(B)= 1/4 and P(AB)= 1/12 then P(B/A) is equal to

a) $\frac{7}{8}$ b) $\frac{1}{3}$ c) $\frac{1}{8}$ d) none

67. The odds in favour of one student passing a test are 3:7.The odds against another student passing at are 3:5.The probability that both pass is

a) $\frac{7}{16}$ b) $\frac{21}{80}$ c) $\frac{9}{80}$ d) $\frac{3}{16}$

68. The odds in favour of one student passing a test are 3:7.The odds against another student passing at are 3:5. The probability that both fail is

a) $\frac{7}{16}$ b) $\frac{21}{80}$ c) $\frac{9}{80}$ d) $\frac{3}{16}$

69. In formula P(B/A), P(A) is

a) greater than zero b) less than zero

c) equal to zero d) greater than equal to zero

70. Two events A and B are mutually exclusive means they are

a) not disjoint b) disjoint c) equally likely d) none

71. A bag contains 10 white and 10 black balls A ball is drawn from it. The probability that it will be white is

(a) $\frac{1}{10}$ (b) 1 (c) $\frac{1}{2}$ (d) none

72. Two dice are thrown at a time. The probability that the numbers shown are equal is

(a) $\frac{2}{6}$ (b) $\frac{5}{6}$ (c) $\frac{1}{6}$ (d) none

73. Two dice are thrown at a time. The probability that 'the difference of numbers shown is 1' is

(a) $\frac{11}{18}$ (b) $\frac{5}{18}$ (c) $\frac{7}{18}$ (d) none

74. Two dice are thrown together. The probability that 'the event the difference of numbers shown is 2' is

(a) 2/9 (b) 5/9 (c) 4/9 (d) 7/9

75. The probability space in tossing two coins is

(a) {(H,H),(H,T),(T,H)} (b) {(H,T),(T,H),(T,T)}
(c) {(H,H),(H,T),(T.H), (T,T)} (d) none

76. The probability of drawing a white ball from a bag containing 3 white and 8 balls is

(a) 3/5 (b) 3/11 (c) 8/11 (d) none

77. Two dice are thrown together. The probability of the event that the sum of numbers shown is greater than 5 is

(a) 13/18 (b) 15/18 (c) 1 (d) none

78. A traffic census show that out of 1000 vehicles passing a junction point on a highway 600 turned to the right. The probability of an automobile turning the right is

(a) 2/5 (b) 3/5 (c) 4/5 (d) none

79. Three coins are tossed together. The probability of getting three tails is

(a) 5/8 (b) 3/8 (c) 1/8 (d) none

80. Three coins are tossed together.The probability of getting exactly two heads is

(a) 5/8 (b) 3/8 (c) 1/8 (d) none

81. Three coins are tossed together. The probability of getting at least two heads is

(a) 1/2 (b) 3/8 (c) 1/8 (d) none

82. 4 coins are tossed. The probability that there are 2 heads is

(a) 1/2 (b) 3/8 (c) 1/8 (d) none

83. If 4 coins are tossed. The chance that there should be two tails is

(a) 1/2 (b) 3/8 (c) 1/8 (d) none

84. If A is an event and A^C its complementary event then

(a) $P(A)=P(A^C)-1$ (b) $P(A^C)=1-P(A)$ (c) $P(A)=1 + P(A^C)$ (d) none

85. If P(A)= 3/8, P(B)= 1/3 and P(AB)= $\frac{1}{4}$ then $P(A^C)$ is equal to

(a) 5/8 (b) 3/8 (c) 1/8 (d) none

86. If P(A)= 3/8, P(B)= 1/3 then $P(\overline{B})$ is equal to

(a) 1 (b) 1/3 (c) 2/3 (d) none

87. If P(A)= 3/8, P(B)= 1/3 and P(AB)= $\frac{1}{4}$ then P(A + B)is

(a) 13/24 (b) 11/24 (c) 17/24 (d) none

88. If P(A)= 1/5, P(B)= 1/2 and A and B are mutually exclusive then P(AB) is

(a) 7/10 (b) 3/10 (c) 1/5 (d) none

89. The probability of throwing more than 4 in a single throw from an ordinary die is

(a) 2/3 (b) 1/3 (c) 1 (d) none

90. The probability that a card drawn at random from the pack of playing cards may be either a queen or an ace is

(a) 2/13 (b) 11/13 (c) 9/13 (d) none

91. The chance of getting 7 or 11 in a throw of 2 dice is

(a) 7/9 (b) 5/9 (c) 2/9 (d) none

92. If the probability of a horse A winning a race is 1/6 and the probability of a horse B winning the same race is 1/4 , what is the probability that one of the horses will win

(a) 5/12 (b) 7/12 (c) 1/12 (d) none

93. If the probability of a horse A winning a race is 1/6 and the probability of a horse B winning the same race is 1/4 , What is the probability that none of them will win

(a) 5/12 (b) 7/12 (c) 1/12 (d) none

94. If P (A)= 7/8 then(P(A^C) is equal to

(a) 1 (b) 0 (c) 7/8 (d) 1/8

95. The value of P(S) were S is the sample space is

(a) –1 (b) 0 (c) 1 (d) none

96. A man can kill a bird once in three shots.The probabilities that a bird is not killed is

(a) 1/3 (b) 2/3 (c) 1 (d) 0

97. If on an average 9 shops out of 10 return safely to a port, the probability of one ship returns safely is

(a) 1/10 (b) 8/10 (c) 9/10 (d) none

98. If on an average 9 shops out of 10 return safely to a port, the probability of one ship does not reach safely is

(a) 1/10 (b) 8/10 (c) 9/10 (d) none

99. The probability of winning of a person is 6/11 and at a result he gets ₹ 77/-. The expectation of this person is

(a) ₹ 35/- (b) ₹ 42/- (c) ₹ 58/- (d) none

100. A family has 2 children. The probability that both of them are boys if it is known that one of them is a boy

(a) 1 (b) 1/2 (c) 3/4 (d) none

101. The Probability of the occurrence of a number greater then 2 in a throw of a die if it is known that only even numbers can occur is

(a) 1/3 (b) 1/2 (c) 2/3 (d) none

102. A player has 7 cards in hand of which 5 are red and of these five 2 are kings. A card is drawn at random. The probability that it is a king, it being known that it is red is

(a) 2/5 (b) 3/5 (c) 4/5 (d) none

103. In a class 40 % students read Mathematics, 25 % Biology and 15 % both Mathematics and Biology. One student is select at random. The probability that he reads Mathematics if it is known that he reads Biology is

(a) 2/5 (b) 3/5 (c) 4/5 (d) none

104. In a class 40 % students read Mathematics, 25 % Biology and 15 % both Mathematics and Biology. One student is select at random.The probability that he reads Biology if he reads Mathematics

(a) 7/8 (b) 1/8 (c) 3/8 (d) none

105. Probability of throwing an odd no with an ordinary six faced die is

(a) 1/2 (b) 1 (c) –1/2 (d) 0

106. For a event A which is certain, P (A) is equal to

(a) 1 (b) 0 (c) –1 (d) none

107. When none of the outcomes is favourable to the event then the event is said to be

(a) certain (b) sample (c) impossible (d) none

ANSWERS

1. (b)	2. (d)	3. (a)	4. (b)	5. (b)
6. (a)	7. (c)	8. (a)	9. (b)	10. (b)
11. (b)	12. (a)	13. (b)	14. (c)	15. (d)
16. (d)	17. (a)	18. (a)	19. (d)	20. (c)
21. (d)	22. (b)	23. (a)	24. (b)	25. (d)
26. (a)	27. (b)	28. (c)	29. (d)	30. (a)
31. (b)	32. (c)	33. (c)	34. (d)	35. (c)
36. (a)	37. (a)	38. (c)	39. (b)	40. (a)
41. (b)	42. (d)	43. (b)	44. (c)	45. (b)
46. (a)	47. (b)	48. (d)	49. (a)	50. (d)
51. (b)	52. (c)	53. (d)	54. (a)	55. (b)
56. (a)	57. (c)	58. (c)	59. (b)	60. (b)
61. (a)	62. (c)	63. (b)	64. (d)	65. (a)
66. (c)	67. (d)	68. (b)	69. (a)	70. (b)
71. (c)	72. (c)	73. (b)	74. (a)	75. (c)
76. (b)	77. (a)	78. (b)	79. (c)	80. (b)
81. (a)	82. (b)	83. (b)	84. (b)	85. (a)
86. (c)	87. (b)	88. (d)	89. (b)	90. (a)
91. (c)	92. (a)	93. (b)	94. (d)	95. (c)
96. (b)	97. (c)	98. (a)	99. (b)	100. (d)
101. (c)	102. (a)	103. (b)	104. (c)	105. (a)
106. (a)	107. (c)			

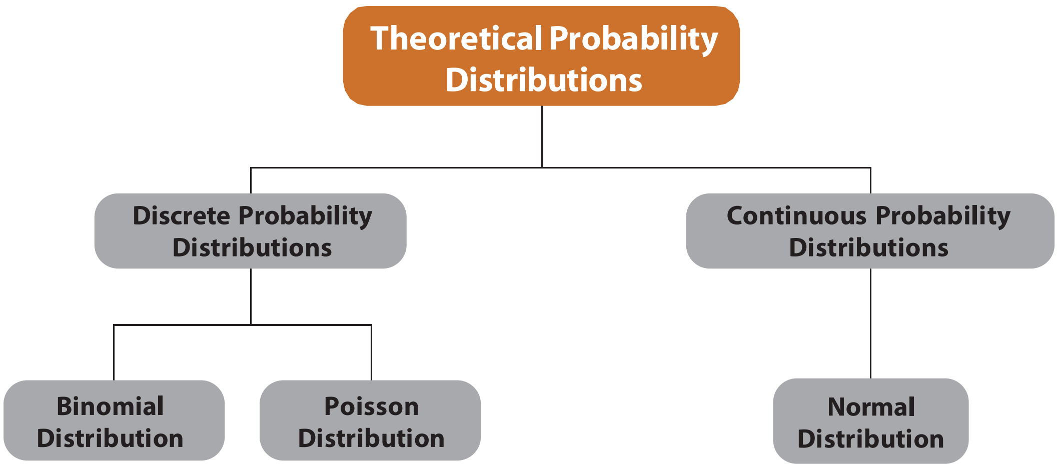
CHAPTER 17

THEORETICAL DISTRIBUTIONS

LEARNING OBJECTIVES

The Students will be introduced in this chapter to the techniques of developing discrete and continuous probability distributions and its applications.

UNIT OVERVIEW



17.1 INTRODUCTION

In chapter nine, it may be recalled, we discussed frequency distribution. In a similar manner, we may think of a probability distribution where just like distributing the total frequency to different class intervals, the total probability (i.e. one) is distributed to different mass points in case of a discrete random variable or to different class intervals in case of a continuous random variable. Such a probability distribution is known as Theoretical Probability Distribution, since such a distribution exists in theory. We need to study theoretical probability distribution for the following important factors:

(a) An observed frequency distribution, in many a case, may be regarded as a sample i.e. a representative part of a large, unknown, boundless universe or population and we may be interested to know the form of such a distribution. By fitting a theoretical probability distribution to an observed frequency distribution of, say, the lamps produced by a manufacturer, it may be possible for the manufacturer to specify the length of life of the lamps produced by him up to a reasonable degree of accuracy. By studying the effect of a particular type of missiles, it may be possible for our scientist to suggest the number of such missiles necessary to destroy an army position. By knowing the distribution of smokers, a social activist may warn the people of a locality about the nuisance of active and passive smoking and so on.

(b) Theoretical probability distribution may be profitably employed to make short term projections for the future.

(c) Statistical analysis is possible only on the basis of theoretical probability distribution. Setting confidence limits or testing statistical hypothesis about population parameter(s) is based on the probability distribution of the population under consideration.

A probability distribution also possesses all the characteristics of an observed distribution. We define mean (μ), median $(\tilde{\mu})$, mode (μ_0), standard deviation (σ) etc. exactly same way we have done earlier. Again a probability distribution may be either a discrete probability distribution or a Continuous probability distribution depending on the random variable under study. **Two important discrete probability distributions are (a) Binomial Distribution and (b) Poisson distribution.**

Some important continuous probability distributions are

Normal Distribution

17.2 BINOMIAL DISTRIBUTION

One of the most important and frequently used discrete probability distribution is Binomial Distribution. It is derived from a particular type of random experiment known as Bernoulli process named after the famous mathematician Bernoulli. Noting that a 'trial' is an attempt to produce a particular outcome which is neither certain nor impossible, the characteristics of Bernoulli trials are stated below:

(i) Each trial is associated with two mutually exclusive and exhaustive outcomes, the occurrence of one of which is known as a 'success' and as such its non occurrence as a 'failure'. As an example, when a coin is tossed, usually occurrence of a head is known as a success and its non–occurrence i.e. occurrence of a tail is known as a failure.

(ii) The trials are independent.

(iii) The probability of a success, usually denoted by p, and hence that of a failure, usually denoted by q = 1–p, remain unchanged throughout the process.

(iv) The number of trials is a finite positive integer.

A discrete random variable x is defined to follow binomial distribution with parameters n and p, to be denoted by x ~ B (n, p), if the probability mass function of x is given by

$$f(x) = p(X = x) = {}^{n}c_{x}\, p^{x} q^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$
$$= 0, \text{ otherwise} \qquad \ldots\ldots\ldots (17.1)$$

We may note the following important points in connection with binomial distribution:

(a) As n >0, p, q ≥ 0, it follows that f(x) ≥ 0 for every x

Also $\sum_{x} f(x) = f(0) + f(1) + f(2) + \ldots + f(n) = 1$ **………(17.2)**

(b) Binomial distribution is known as biparametric distribution as it is characterised by two parameters n and p. This means that if the values of n and p are known, then the distribution is known completely.

(c) The mean of the binomial distribution is given by $\mu = np$ **…. (17.3)**

(d) Depending on the values of the two parameters, binomial distribution may be unimodal or bi- modal. μ_0**, the mode of binomial distribution, is given by** μ_0 **= the largest integer contained in (n+1)p if (n+1)p is a non-integer (n+1)p and (n+1)p - 1 if (n+1)p is an integer ….(17.4)**

(e) The variance of the binomial distribution is given by

$$\sigma^2 = npq \qquad \ldots\ldots\ldots\ldots (17.5)$$

Since p and q are numerically less than or equal to 1, npq < np

⇒ variance of a binomial variable is always less than its mean.

Also variance of X attains its maximum value at p = q = 0.5 and this maximum value is n/4.

(f) Additive property of binomial distribution.

If X and Y are two independent variables such that

$X \sim B(n_1, P)$

and $Y \sim B(n_{21} P)$

Then $(X+Y) \sim B(n_1 + n_2, P)$ **……………………………. (17.6)**

Applications of Binomial Distribution

Binomial distribution is applicable when the trials are independent and each trial has just two outcomes success and failure. It is applied in coin tossing experiments, sampling inspection plan, genetic experiments and so on.

Example 17.1: A coin is tossed 10 times. Assuming the coin to be unbiased, what is the probability of getting

(i) 4 heads?

(ii) at least 4 heads?

(iii) at most 3 heads?

Solution: We apply binomial distribution as the tossing are independent of each other. With every tossing, there are just two outcomes either a head, which we call a success or a tail, which we call a failure and the probability of a success (or failure) remains constant throughout.

Let X denotes the no. of heads. Then X follows binomial distribution with parameter n = 8 and p = 1/2 (since the coin is unbiased). Hence q = 1 – p = 1/2

The probability mass function of X is given by

$f(x) = {}^{n}c_{x}\ p^{x}\ q^{n-x}$

$= {}^{10}c_{x} \cdot (1/2)^{x} \cdot (1/2)^{10-x}$

$$= \frac{{}^{10}c_{x}}{2^{10}}$$

$= {}^{10}c_{x} / 1024$ for x = 0, 1, 2, ...10

(i) probability of getting 4 heads

$= f(4)$

$= {}^{10}c_{4} / 1024$

$= 210 / 1024$

$= 105 / 512$

(ii) probability of getting at least 4 heads

$= P(X \geq 4)$

$= P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8)$

$= {}^{10}c_{4} / 1024 + {}^{10}c_{5} / 1024 + {}^{10}c_{6} / 1024 + {}^{10}c_{7} / 1024 + {}^{10}c_{8} /1024$

$= \frac{210 + 252 + 210 + 120 + 45}{1024}$

= 837 / 1024

(iii) probability of getting at most 3 heads

= P (X ≤ 3)

= P (X = 0) + P (X = 1) + P (X = 2) + P (X = 3)

= f (0) + f (1) + f (2) + f (3)

$= {}^{10}c_0 / 1024 + {}^{10}c_1 / 1024 + {}^{10}c_2 / 1024 + {}^{10}c_3 / 1024$

$= \frac{1 + 10 + 45 + 120}{1024}$

= 176 / 1024

= 11/64

Example 17.2: If 15 dates are selected at random, what is the probability of getting two Sundays?

Solution: If X denotes the number at Sundays, then it is obvious that X follows binomial distribution with parameter n = 15 and p = probability of a Sunday in a week = 1/7 and q = 1 – p = 6 / 7.

Then $f(x) = {}^{15}c_x (1/7)^x . (6/7)^{15-x}$.

for x = 0, 1, 2,……….. 15.

Hence the probability of getting two Sundays

= f(2)

$= {}^{15}c_2 (1/7)^2 . (6/7)^{15-2}$

$= \frac{105 \times 6^{13}}{7^{15}}$

≅ 0.29

Example 17.3: The incidence of occupational disease in an industry is such that the workmen have a 10% chance of suffering from it. What is the probability that out of 5 workmen, 3 or more will contract the disease?

Solution: Let X denote the number of workmen in the sample. X follows binomial with parameters n = 5 and p = probability that a workman suffers from the occupational disease = 0.1

Hence q = 1 – 0.1 = 0.9.

Thus $f(x) = {}^5c_x . (0.1)^x . (0.9)^{5-x}$

For $x = 0, 1, 2, \ldots\ldots.,5.$

The probability that 3 or more workmen will contract the disease

$= P(x \geq 3)$

$= f(3) + f(4) + f(5)$

$= {}^5c_3\ (0.1)^3\ (0.9)^{5-3} + {}^5c_4\ (0.1)^4 . (0.9)^{5-4} + {}^5c_5\ (0.1)^5$

$= 10 \times 0.001 \times 0.81 + 5 \times 0.0001 \times 0.9 + 1 \times 0.00001$

$= 0.0081 + 0.00045 + 0.00001$

$\cong 0.0086.$

Example 17.4: Find the probability of a success for the binomial distribution satisfying the following relation 4 P (x = 4) = P (x = 2) and having the parameter n as six.

Solution: We are given that n = 6. The probability mass function of x is given by

$f(x) = {}^nc_x\ p^x\ q^{\,n-x}$

$= {}^6c_x\ p^x\ q^{\,n-x}$

for $x = 0, 1, \ldots\ldots ,6.$

Thus $P(x = 4) = f(4)$:

$= {}^6c_4\ p^4\ q^{\,6-4}$

$= 15\ p^4\ q^2$

and $P(x = 2) = f(2)$

$= {}^6c_2\ p^2\ q^{\,6-2}$

$= 15p^2\ q^4$

Hence $4\ P(x = 4) = P(x = 2)$

$\Rightarrow \quad 60\ p^4\ q^2 = 15\ p^2\ q^4$

$\Rightarrow \quad 15\ p^2\ q^2\ (4p^2 - q^2) = 0$

$\Rightarrow \quad 4p^2 - q^2 = 0$ (as $p \neq 0, q \neq 0$)

$\Rightarrow \quad 4p^2 - (1 - p)^2 = 0$ (as $q = 1 - p$)

$\Rightarrow \quad (2p + 1 - p) = 0$ or $(2p - 1 + p) = 0$

$\Rightarrow \quad p = -1$ or $p = 1/3$

Thus $p = 1/3$ (as $p \neq -1$)

Example 17.5: Find the binomial distribution for which mean and standard deviation are 6 and 2 respectively.

Solution: Let x ~ B (n, p)

Given that mean of x = np = 6 … (1)

and SD of x = 2

$\Rightarrow$ variance of x = npq = 4 ….. (2)

Dividing (2) by (1), we get $q = \frac{2}{3}$

Hence $p = 1 - q = \frac{1}{3}$

Replacing p by $\frac{1}{3}$ in equation (1), we get $n \times \frac{1}{3} = 6$

$\Rightarrow n = 18$

Thus the probability mass function of x is given by

$f(x) = {}^{n}c_{x}\, p^{x}\, q^{n-x}$

$= {}^{18}c_{x}\,(1/3)^{x} \cdot (2/3)^{18-x}$

for x = 0, 1, 2,…… ,18

Example 17.6: Fit a binomial distribution to the following data:

x:	0	1	2	3	4	5
f:	3	6	10	8	3	2

Solution: In order to fit a theoretical probability distribution to an observed frequency distribution it is necessary to estimate the parameters of the probability distribution. There are several methods of estimating population parameters. One rather, convenient method is 'Method of Moments'. This comprises equating p moments of a probability distribution to p moments of the observed frequency distribution, where p is the number of parameters to be estimated. Since n = 5 is given, we need estimate only one parameter p. We equate the first moment about origin i.e. AM of the probability distribution to the AM of the given distribution and estimate p.

i.e. $n\hat{p} = \bar{x}$

$\Rightarrow \hat{p} = \frac{\bar{x}}{n}$ ($\hat{p}$ is read as p hat)

The fitted binomial distribution is then given by

$f(x) = {}^{n}c_{x}\, \hat{p}^{\,x}(1 - \hat{p})^{n-x}$

For x = 0, 1, 2, …… n

On the basis of the given data, we have

$$\bar{x} = \sum \frac{f_i x_i}{N}$$

$$= \frac{3\times0+6\times1+10\times2+8\times3+3\times4+2\times5}{3+6+10+8+3+2} = 2.25$$

Thus $\hat{p} = \bar{x}/n = \frac{2.25}{n} = 0.45$

and $\hat{q} = 1 - \hat{p} = 0.55$

The fitted binomial distribution is

$f(x) = {}^5c_x\,(0.45)^x\,(0.55)^{5-x}$

For x = 0, 1, 2, 3, 4, 5.

Table 17.1

Fitting Binomial Distribution to an Observed Distribution

X	f (x)	Expected frequency	Observed frequency
	$= {}^5c_x\,(0.4)^x\,(0.6)^{5-x}$	Nf (x) = 32 f (x)	
0	0.07776	2.49 ≅ 3	3
1	0.25920	8.29 ≅ 8	6
2	0.34560	11.06 ≅ 11	10
3	0.23040	7.37 ≅ 7	8
4	0.07680	2.46 ≅ 3	3
5	0.01024	0.33 ≅ 0	2
Total	1.000 00	32	32

A look at Table 17.1 suggests that the fitting of binomial distribution to the given frequency distribution is satisfactory.

Example 17.7: 6 coins are tossed 512 times. Find the expected frequencies of heads. Also, compute the mean and SD of the number of heads.

Solution: If x denotes the number of heads, then x follows binomial distribution with parameters n = 6 and p = prob. of a head = ½, assuming the coins to be unbiased. The probability mass function of x is given by

$f(x) = {}^6c_x\,(1/2)^x.\,(1/2)^{6-x}$

$= {}^6c_x/2^6$

for x = 0, 1, …..6.

The expected frequencies are given by Nf (x).

Table 17.2

Finding Expected Frequencies when 6 coins are tossed 512 times

x	f (x)	Nf (x) Expected frequency	x f (x)	x^2f (x)
0	1/64	8	0	0
1	6/64	48	6/64	6/64
2	15/64	120	30/64	60/64
3	20/64	160	60/64	180/64
4	15/64	120	60/64	240/64
5	6/64	48	30/64	150/64
6	1/64	8	6/64	36/64
Total	1	512	3	10.50

Thus mean = $\mu = \sum_x x.f(x) = 3$

$$E(x^2) = \sum_x x^2.f(x) = 10.50$$

$$\text{Thus } \sigma^2 = \sum_x x^2.f(x) - \mu^2$$

$$= 10.50 - 3^2 = 1.50$$

$$\therefore SD = \sigma = \sqrt{1.50} \cong 1.22$$

Applying formula for mean and SD, we get

$\mu = np = 6 \times 1/2 = 3$

and $\sigma = \sqrt{npq} = \sqrt{6 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{1.50} \cong 1.22$

Example 17.8: An experiment succeeds thrice as after it fails. If the experiment is repeated 5 times, what is the probability of having no success at all ?

Solution: Denoting the probability of a success and failure by p and q respectively, we have,

$p = 3q$
$\Rightarrow p = 3(1 - p)$
$\Rightarrow p = 3/4$
$\therefore q = 1 - p = 1/4$
when n = 5 and p = 3/4, we have

$f(x) = {}^5c_x \; (3/4)^x \; (1/4)^{5-x}$

for n = 0, 1, ... , 5.

So probability of having no success

= f (0)

$= {}^5c_0 \; (3/4)^0 \, (1/4\;)^{5-0}$

= 1/1024

Example 17.9: What is the mode of the distribution for which mean and SD are 10 and $\sqrt{5}$ respectively.

Solution: As given np = 10 .... (1)

and $\sqrt{npq} = \sqrt{5}$

$\Rightarrow$ npq = 5 .... (2)

on solving (1) and (2), we get n = 20 and p = 1/2

Hence mode = Largest integer contained in (n+1)p

= Largest integer contained in (20+1) × 1/2

= Largest integer contained in 10.50

= 10.

Example 17.10: If x and y are 2 independent binomial variables with parameters 6 and 1/2 and 4 and 1/2 respectively, what is P (x + y ≥ 1)?

Solution: Let z = x + y.

It follows that z also follows binomial distribution with parameters

(6 + 4) and 1/2

i.e. 10 and 1/2

Hence P (z ≥ 1)

= 1 – P (z < 1)

= 1 – P (z = 0)

$= 1 - {}^{10}c_0 \; (1/2\;)^0 . \; (1/2\;)^{10-0}$

$= 1 - 1 \;/\; 2^{10}$

= 1023 / 1024

17.3 POISSON DISTRIBUTION

Poisson distribution is a theoretical discrete probability distribution which can describe many processes. Simon Denis Poisson of France introduced this distribution way back in the year 1837.

Poisson Model

Let us think of a random experiment under the following conditions:

I. The probability of finding success in a very small time interval (t, t + dt) is kt, where k (>0) is a constant.

II. The probability of having more than one success in this time interval is very low.

III. The probability of having success in this time interval is independent of t as well as earlier successes.

The above model is known as Poisson Model. The probability of getting x successes in a relatively long time interval T containing m small time intervals t i.e. T = mt. is given by

$$\frac{e^{-kt}.(kt)^x}{x!}$$

for x = 0, 1, 2, ... ∞ …… (17.7)

Taking kT = m, the above form is reduced to

$$\frac{e^{-m}.m^x}{x!}$$

for x = 0, 1, 2, ... ∞ …..... (17.8)

Definition of Poisson Distribution

A random variable X is defined to follow Poisson distribution with parameter λ, to be denoted by X ~ P (m) if the probability mass function of x is given by

$$f(x) = P(X = x) = \frac{e^{-m}.m^x}{x!} \text{ for } x = 0, 1, 2, \ldots \infty$$

$$= 0 \quad \text{otherwise} \qquad \ldots\ldots (17.9)$$

Here e is a transcendental quantity with an approximate value as 2.71828.

It is wiser to remember the following important points in connection with Poisson distribution:

(i) Since $e^{-m} = 1/e^m > 0$, whatever may be the value of m, m > 0, it follows that f (x) $\geq$ 0 for every x.

Also it can be established that $\sum_x$ f(x) = 1 i.e. f(0) + f(1) + f(2) +....... = 1.... (17.10)

(ii) Poisson distribution is known as a uniparametric distribution as it is characterised by only one parameter m.

(iii) The mean of Poisson distribution is given by m i,e μ = m. (17.11)

(iv) The variance of Poisson distribution is given by σ^2 = m (17.12)

(v) Like binomial distribution, Poisson distribution could be also unimodal or bimodal depending upon the value of the parameter m.

We have μ_0 = The largest integer contained in m if m is a non-integer

= m and m–1 if m is an integer .... (17.13)

(vi) Poisson approximation to Binomial distribution

If n, the number of independent trials of a binomial distribution, tends to infinity and p, the probability of a success, tends to zero, so that m = np remains finite, then a binomial distribution with parameters n and p can be approximated by a Poisson distribution with parameter m (= np).

In other words when n is rather large and p is rather small so that m = np is moderate then

$\beta\ (n, p) \cong P\ (m)$. .... (17.14)

(vii) Additive property of Poisson distribution

If X and y are two independent variables following Poisson distribution with parameters m_1 and m_2 respectively, then Z = X + Y also follows Poisson distribution with parameter $(m_1 + m_2)$.

i.e. if $X \sim P\ (m_1)$

and $Y \sim P\ (m_2)$

and X and Y are independent, then

$Z = X + Y \sim P\ (m_1 + m_2)$ .... (17.15)

Application of Poisson distribution

Poisson distribution is applied when the total number of events is pretty large but the probability of occurrence is very small. Thus we can apply Poisson distribution, rather profitably, for the following cases:

a) The distribution of the no. of printing mistakes per page of a large book.
b) The distribution of the no. of road accidents on a busy road per minute.
c) The distribution of the no. of radio-active elements per minute in a fusion process.
d) The distribution of the no. of demands per minute for health centre and so on.

Example 17.11: Find the mean and standard deviation of x where x is a Poisson variate satisfying the condition P (x = 2) = P (x = 3).

Solution: Let x be a Poisson variate with parameter m. The probability max function of x is then given by

$$f\ (x) = \frac{e^{-m}.m^{x}}{x!} \qquad \text{for } x = 0, 1, 2, ...\ \infty$$

now, P (x = 2) = P (x = 3)

$\Rightarrow f(2) = f(3)$

$$\Rightarrow \frac{e^{-m}.m^2}{2!} = \frac{e^{-m}.m^3}{3!}$$

$$\Rightarrow \frac{e^{-m}.m^2}{2}(1 - m/3) = 0$$

$\Rightarrow 1 - m / 3 = 0$ (as $e^{-m} > 0, m > 0$)

$\Rightarrow m = 3$

Thus the mean of this distribution is m = 3 and standard deviation = $\sqrt{3} \cong 1.73$.

Example 17.12: The probability that a random variable x following Poisson distribution would assume a positive value is $(1 - e^{-2.7})$. What is the mode of the distribution?

Solution: If x ~ P (m), then its probability mass function is given by

$$f(x) = \frac{e^{-m}.m^2}{x!} \quad \text{for } x = 0, 1, 2, ... \infty$$

The probability that x assumes a positive value

$= P (x > 0)$

$= 1 - P (x \leq 0)$

$= 1 - P (x = 0)$

$= 1 - f(0)$

$= 1 - e^{-m}$

As given,

$1 - e^{-m} = 1 - e^{-2.7}$

$\Rightarrow e^{-m} = e^{-2.7}$

$\Rightarrow m = 2.7$

Thus μ_0 = largest integer contained in 2.7

= 2

Example 17.13: The standard deviation of a Poisson variate is 1.732. What is the probability that the variate lies between –2.3 to 3.68?

Solution: Let x be a Poisson variate with parameter m.

Then SD of x is $\sqrt{m}$.

As given $\sqrt{m} = 1.732$

$\Rightarrow m = (1.732)^2 \cong 3.$

The probability that x lies between –2.3 and 3.68

$= P(-2.3 < x < 3.68)$

$= f(0) + f(1) + f(2) + f(3)$ (As x can assume 0, 1, 2, 3, 4 .....)

$$= \frac{e^{-3}.3^0}{0!} + \frac{e^{-3}.3^1}{1!} + \frac{e^{-3}.3^2}{2!} + \frac{e^{-3}.3^3}{3!}$$

$= e^{-3} (1 + 3 + 9/2 + 27/6)$

$= 13e^{-3}$

$$= \frac{13}{e^3}$$

$$= \frac{13}{(2.71828)^3} \text{ (as e = 2.71828)}$$

$\cong 0.65$

Example 17.14: X is a Poisson variate satisfying the following relation:

P (X = 2) = 9P (X = 4) + 90P (X = 6).

What is the standard deviation of X?

Solution: Let X be a Poisson variate with parameter m. Then the probability mass function of X is

$$P(X = x) = f(x) = \frac{e^{-m}.m^x}{x!} \text{ for } x = 0, 1, 2, \ldots.. \infty$$

Thus P (X = 2) = 9P (X = 4) + 90P (X = 6)

$\Rightarrow f(2) = 9\, f(4) + 90\, f(6)$

$$\Rightarrow \frac{e^{-m} m^2}{2!} = \frac{9e^{-m}.m^4}{4!} + \frac{90.e^{-m} m^6}{6!}$$

$$\Rightarrow \frac{e^{-m} m^2}{2}\left(\frac{90m^4}{360} + \frac{9m^2}{12} - 1\right) = 0$$

$$\Rightarrow \frac{e^{-m} m^2}{8}(m^4 + 3m^2 - 4) = 0$$

$\Rightarrow e^{-m} .m^2 (m^2 + 4) (m^2 - 1) = 0$

$\Rightarrow m^2 - 1 = 0$ (as $e^{-m} > 0$ $m > 0$ and $m^2 + 4 \neq 0$)

$\Rightarrow m = 1$ (as $m > 0$, $m \neq -1$)

Thus the standard deviation of X is $\sqrt{1} = 1$

Example 17.15: Between 9 and 10 AM, the average number of phone calls per minute coming into the switchboard of a company is 4. Find the probability that during one particular minute, there will be,

1. no phone calls
2. at most 3 phone calls (given $e^{-4} = 0.018316$)

Solution: Let X be the number of phone calls per minute coming into the switchboard of the company. We assume that X follows Poisson distribution with parameters m = average number of phone calls per minute = 4.

1. The probability that there will be no phone call during a particular minute

$= P(X = 0)$

$= \dfrac{e^{-4}.4^0}{0!}$

$= e^{-4}$

$= 0.018316$

2. The probability that there will be at most 3 phone calls

$= P(X \leq 3)$

$= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$= \dfrac{e^{-4}.4^0}{0!} + \dfrac{e^{-4}.4^1}{1!} + \dfrac{e^{-4}.4^2}{2!} + \dfrac{e^{-4}.4^3}{3!}$

$= e^{-4}(1 + 4 + 16/2 + 64/6)$

$= e^{-4} \times 71/3$

$= 0.018316 \times 71/3$

$\cong 0.43$

Example 17.16: If 2 per cent of electric bulbs manufactured by a company are known to be defectives, what is the probability that a sample of 150 electric bulbs taken from the production process of that company would contain

1. exactly one defective bulb?
2. more than 2 defective bulbs?

Solution: Let x be the number of bulbs produced by the company. Since the bulbs could be either defective or non-defective and the probability of bulb being defective remains the same, it follows that x is a binomial variate with parameters n = 150 and p = probability of a bulb being defective = 0.02. However since n is large and p is very small, we can approximate this binomial distribution with Poisson distribution with parameter m = np = 150 x 0.02 = 3.

1. The probability that exactly one bulb would be defective

$= P(X = 1)$

$= \frac{e^{-3}.3^1}{1!}$

$= e^{-3} \times 3$

$= \frac{3}{e^3}$

$= 3/(2.71828)^3$

$\cong 0.15$

2. The probability that there would be more than 2 defective bulbs

$= P(X > 2)$

$= 1 - P(X \leq 2)$

$= 1 - [f(0) + f(1) + f(2)]$

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

$= 1 - 8.5 \times e^{-3}$

$= 1 - 0.4232$

$= 0.5768 \cong 0.58$

Example 17.17: The manufacturer of a certain electronic component is certain that two per cent of his product is defective. He sells the components in boxes of 120 and guarantees that not more than two per cent in any box will be defective. Find the probability that a box, selected at random, would fail to meet the guarantee? Given that $e^{-2.40} = 0.0907$.

Solution: Let x denote the number of electric components. Then x follows binomial distribution with n = 120 and p = probability of a component being defective = 0.02. As before since n is quite large and p is rather small, we approximate the binomial distribution with parameters n and p by a Poisson distribution with parameter m = n.p = 120 × 0.02 = 2.40. Probability that a box, selected at random, would fail to meet the specification = probability that a sample of 120 items would contain more than 2.40 defective items.

$= P(X > 2.40)$

$= 1 - P(X \leq 2.40)$

$= 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$

$= 1 - [e^{-2.40} + e^{-2.40} \times 2.4 + e^{-2.40} \times \left(\frac{2.40}{2}\right)^2]$

$= 1 - e^{-2.40}\left(1 + 2.40 + \frac{(2.40)^2}{2}\right)$

$= 1 - 0.0907 \times 6.28$

$\cong 0.43$

Example 17.18: A discrete random variable x follows Poisson distribution. Find the values of

(i) P (X = at least 1)

(ii) $P(X \leq 2/ X \geq 1)$

You are given E (x) = 2.20 and $e^{-2.20} = 0.1108$.

Solution: Since X follows Poisson distribution, its probability mass function is given by

$$f(x) = \frac{e^{-m}.m^x}{x!} \text{ for } x = 0, 1, 2, \ldots\ldots \infty$$

(i) P (X = at least 1)

$= P(X \geq 1)$

$= 1 - P(X < 1)$

$= 1 - P(X = 0)$

$= 1 - e^{-m}$

$= 1 - e^{-2.20}$ (as E (x) = m = 2.20, given)

$= 1 - 0.1108$ (as $e^{-2.20} = 0.1108$ as given)

$\cong 0.89$.

(ii) $P(x \leq 2 / x \geq 1)$

$$= P\frac{[(X \leq 2) \cap (X \geq 1)]}{P(X \geq 1)} \qquad \left(\text{as } P(A/B) = P\frac{(A \cap B)}{P(B)}\right)$$

$$= \frac{P(X=1)+P(X=2)}{1-P(X<1)}$$

$$= \frac{f(1)+f(2)}{1-f(0)}$$

$$= \frac{e^{-m}.m + e^{-m}.m^2/2}{1-e^{-m}}$$

$$= \frac{e^{-2.20} \times 2.2 + e^{-2.20} \times (2.20)^2/2}{1-e^{-2.20}} \qquad (\because m = 2.2)$$

$$= \frac{0.5119}{0.8892}$$

$$\cong 0.58$$

Fitting a Poisson distribution

As explained earlier, we can apply the method of moments to fit a Poisson distribution to an observed frequency distribution. Since Poisson distribution is uniparametric, we equate m, the parameter of Poisson distribution, to the arithmetic mean of the observed distribution and get the estimate of m.

i.e. $\hat{m} = \bar{x}$

The fitted Poisson distribution is then given by

$$\hat{f}(x) = \frac{e^{-\hat{m}}.(\hat{m})^x}{x!} \qquad \text{for } x = 0, 1, 2.................\infty$$

Example 17.19: Fit a Poisson distribution to the following data :

Number of death:	0	1	2	3	4
Frequency:	122	46	23	8	1

(Given that $e^{-0.6} = 0.5488$)

Solution: The mean of the observed frequency distribution is

$$\bar{x} = \frac{\sum f_i x_i}{N}$$

$$= -\frac{122 \times 0 + 46 \times 1 + 23 \times 2 + 8 \times 3 + 1 \times 4}{122 + 46 + 23 + 8 + 1}$$

$$= \frac{120}{200}$$

$$= 0.6$$

Thus $\hat{m} = 0.6$

Hence $\hat{f}(0) = e^{-\hat{m}} = e^{-0.6} = 0.5488$

$$\hat{f}(1) = \frac{e^{-\hat{m}} \times m}{1!} = 0.6 \times e^{-0.6} = 0.3293$$

$$\frac{(0.6)^2}{2!} \times 0.5488 = 0.0988$$

$$\frac{(0.6)^3}{3!} \times 0.5488 = 0.0198$$

Lastly $P(X \geq 4) = 1 - P(X < 4)$.

Table 17.3

Fitting Poisson Distribution to an Observed Frequency Distribution of Deaths

X	f (x)	Expected frequency $N \times f(x)$	Observed frequency
0	0.5488	109.76 = 110	122
1	0.6 x 0.5488 = 0.3293	65.86 = 65	46
2	$(0.6)^2/2$ x 0.5488 = 0.0988	19.76 = 20	23
3	$(0.6)^3/3$ x 0.5488 = 0.0198	3.96 = 4	8
4 or more	0.0033 (By subtraction)	0.66 = 1	1
Total	1	200	200

17.4 NORMAL OR GAUSSIAN DISTRIBUTION

The two distributions discussed so far, namely binomial and Poisson, are applicable when the random variable is discrete. In case of a continuous random variable like height or weight, it is impossible to distribute the total probability among different mass points because between any two unequal values, there remains an infinite number of values. Thus a continuous random variable is defined in term of its probability density function f (x), provided, of course, such a function really exists, f (x) satisfies the following condition:

$$f(x) \geq 0 \text{ for } x \in (-\infty, \infty)$$

and $$\int_{-\infty}^{+\infty} f(x) = 1.$$

The most important and universally accepted continuous probability distribution is known as normal distribution. Though many mathematicians like De-Moivre, Laplace etc. contributed towards the development of normal distribution, Karl Gauss was instrumental for deriving normal distribution and as such normal distribution is also referred to as Gaussian Distribution.

A continuous random variable x is defined to follow normal distribution with parameters μ and σ^2, to be denoted by

X ~ N (μ, σ ²) ………………. (17.16)

If the probability density function of the random variable x is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-(\bar{x}-u)^2/2\sigma^2}$$

for $-\infty < x < \infty$ ……… (17.17)

where μ and σ are constants, and $\sigma > 0$

Some important points relating to normal distribution are listed below:

(a) The name Normal Distribution has its origin some two hundred years back as the then mathematician were in search for a normal model that can describe the probability distribution of most of the continuous random variables.

(b) If we plot the probability function y = f (x), then the curve, known as probability curve, takes the following shape:

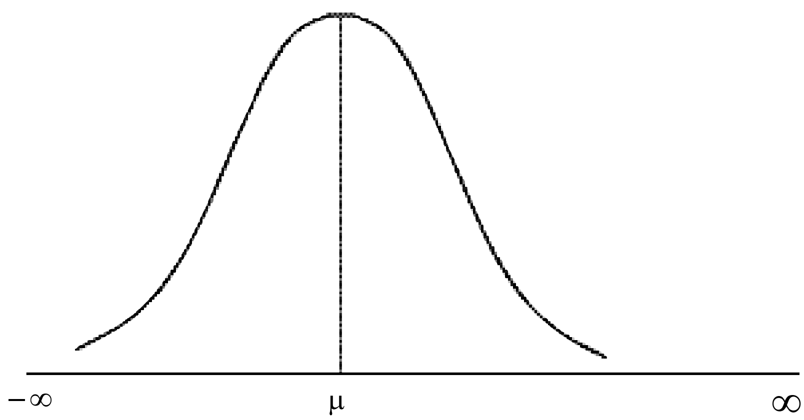


Fig. 17.1

Showing Normal Probability Curve

A quick look at figure 17.1 reveals that the normal curve is bell shaped and has one peak, which implies that the normal distribution has one unique mode. The line drawn through x = μ has divided the normal curve into two parts which are equal in all respect. Such a curve is known as symmetrical curve and the corresponding distribution is known as symmetrical distribution. Thus, we find that the normal distribution is symmetrical about x = μ. It may also be noted that the binomial distribution is also symmetrical about p = 0.5. We next note that the two tails of the normal curve extend indefinitely on both sides of the curve and both the left and right tails never touch the horizontal axis. The total area of the normal curve or for that any probability curve is taken to be unity i.e. one. Since the vertical line drawn through x = μ divides the curve into two equal halves, it automatically follows that, The area between $-\infty$ to μ = the area between μ to ∞ = 0.5

When the mean is zero, we have

the area between $-\infty$ to 0 = the area between 0 to ∞ = 0.5

(c) If we take $\mu = 0$ and $\sigma = 1$ in (18.17), we have

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{for } -\infty < z < \infty \quad \text{........ (17.18)}$$

The random variable z is known as standard normal variate (or variable) or standard normal deviate. The probability that a standard normal variate X would take a value less than or equal to a particular value say X = x is given by

$\phi(x) = p(X \le x)$ .... (17.19)

$\phi(x)$ is known as the cumulative distribution function.

We also have $\phi(0) = P(X \le 0)$ = Area of the standard normal curve between $-\infty$ and 0 = 0.5 .... (17.20)

(d) **The normal distribution is known as biparametric distribution as it is characterised by two parameters μ and σ^2. Once the two parameters are known, the normal distribution is completely specified.**

Properties of Normal Distribution

1. Since $\pi = 22/7$, $e^{-\theta} = 1 / e^{\theta} > 0$, whatever θ may be,

   it follows that $f(x) \ge 0$ for every x.

   It can be shown that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

2. The mean of the normal distribution is given by μ. Further, since the distribution is symmetrical about $x = \mu$, it follows that the mean, median and mode of a normal distribution coincide, all being equal to μ.

3. The standard deviation of the normal distribution is given by σ.

   Mean deviation of normal distribution is

   $\sigma\sqrt{2} \cong \mathbf{0.8}\sigma$ .... (17.21)

   The first and third quartiles are given by

   $Q_1 = \mu - 0.675\,\sigma$ .... (17.22)

   and $Q_3 = \mu + 0.675\,\sigma$ .... (17.23)

   so that, quartile deviation = $0.675\,\sigma$ .... (17.24)

4. The normal distribution is symmetrical about $x = \mu$. As such, its skewness is zero i.e. the normal curve is neither inclined move towards the right (negatively skewed) nor towards the left (positively skewed).

5. The normal curve y = f (x) has two points of inflexion to be given by $x = \mu - \sigma$ and $x = \mu + \sigma$ i.e. at these two points, the normal curve changes its curvature from concave to convex and from convex to concave.

6. If $x \sim N(\mu, \sigma^2)$ then $z = x - \mu/\sigma \sim N(0, 1)$, z is known as standardised normal variate or normal deviate.

   We also have $P(z \le k) = \phi(k)$ ……………. (17.25)

   The values of $\phi(k)$ for different k are given in a table known as "Biometrika."

   Because of symmetry, we have

   $\phi(-k) = 1 - \phi(k)$ …………………. (17.26)

   We can evaluate the different probabilities in the following manner:

   $$P(x < a) = P\left[\frac{x-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right]$$

   $= P(z < k), (k = a - \mu/\sigma)$

   $= \phi(k)$ ………………….. (17.27)

   Also $P(x \le a) = P(x < a)$ as x is continuous.

   $P(x > b) = 1 - P(x \le b)$

   $= 1 - \phi(b - \mu/\sigma)$ …………. (17.28)

   and $P(a < x < b) = \phi(b - \mu/\sigma) - \phi(a - \mu/\sigma)$ …. (17.29)

   ordinate at x = a is given by

   $(1/\sigma)\, \phi(a - \mu/\sigma)$ …………….. (17.30)

   Also, $\phi(-k) = \phi(k)$ ……….. (17.31)

   The values of $\phi(k)$ for different k are also provided in the Biometrika Table.

7. Area under the normal curve is shown in the following figure :

$\mu - 3\sigma$	$\mu - 2\sigma$	$\mu - \sigma$	$x = \mu$	$\mu + \sigma$	$\mu + 2\sigma$	$\mu + 3\sigma$
$(z = -3)$	$(z = -2)$	$(z = -1)$	$(z = 0)$	$(z = 1)$	$(z = 2)$	$(z = 3)$

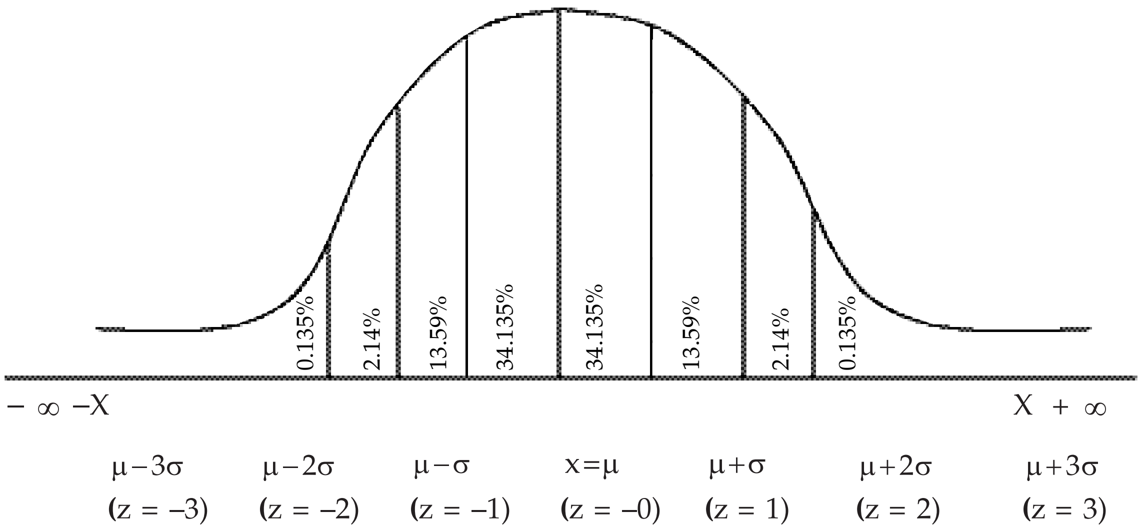


Fig. 17.2

Area Under Normal Curve

From this figure, we find that

$P (\mu - \sigma < x < \mu) = P (\mu < x < \mu + \sigma) = 0.34135$

or alternatively, $P (-1 < z < 0) = P (0 < z < 1) = 0.34135$

$P (\mu - 2 \sigma < x < \mu) = P (\mu < x < \mu + 2 \sigma) = 0.47725$

i.e. $P (- 2 < z < 1) = P (1 < z < 2) = 0.47725$

$P (\mu - 3 \sigma < x < \mu) = P (\mu < x < \mu + 3\sigma) = 0.49865$

i.e. $P(-3 < z < 0) = P (0 < z < 3) = 0.49865$

……. (17.32)

combining these results, we have

$P (\mu - \sigma < x < \mu + \sigma) = 0.6828$

=> $P (-1 < z < 1) = 0.6828$

$P (\mu - 2 \sigma < x < \mu + 2\sigma) = 0.9546$

=> $P (- 2 < z < 2) = 0.9546$

and $P (\mu - 3 \sigma < x < \mu + 3 \sigma) = 0.9973$

=> $P (- 3 < z < 3) = 0.9973.$

………… (17.33)

We note that 99.73 per cent of the values of a normal variable lies between $(\mu - 3 \sigma)$ and $(\mu + 3 \sigma)$. Thus the probability that a value of x lies outside that limit is as low as 0.0027.

8. If x and y are independent normal variables with means and standard deviations as μ_1 and μ_2 and σ_1, and σ_2 respectively, then z = x + y also follows normal distribution with mean $(\mu_1 + \mu_2)$ and SD = $\sqrt{\sigma_1^2+\sigma_2^2}$ respectively.

i.e. If $x \sim N(\mu_1, \sigma_1^2)$

and $y \sim N(\mu_2, \sigma_2^2)$ and x and y are independent,

then $z = x + y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

…… (17.34)

Applications of Normal Distribution

The applications of normal distribution is not restricted to statistics only. Many science subjects, social science subjects, management, commerce etc. find many applications of normal distributions. Most of the continuous variables like height, weight, wage, profit etc. follow normal distribution. If the variable under study does not follow normal distribution, a simple transformation of the variable, in many a case, would lead to the normal distribution of the changed variable. When n, the number of trials of a binomial distribution, is large and p, the probability of a success, is moderate i.e. neither too large nor too small then the binomial distribution, also, tends to normal distribution. Poisson distribution, also for large value of m approaches normal distribution. Such transformations become necessary as it is easier to compute probabilities under the assumption of a normal distribution. Not only the distribution of discrete random variable, the probability distributions of t, chi-square and F also tend to normal distribution under certain specific conditions. In order to infer about the unknown universe, we take recourse to sampling and inferences regarding the universe is made possible only on the basis of normality assumption. Also the distributions of many a sample statistic approach normal distribution for large sample size.

Example 17.20: For a random variable x, the probability density function is given by

$$f(x) = \frac{e^{-(x-4)^2}}{\sqrt{\pi}}$$

for $-\infty < x < \infty$.

Identify the distribution and find its mean and variance.

Solution: The given probability density function may be written as

$$f(x) = \frac{1}{1/\sqrt{2}\times\sqrt{2\pi}} e^{-(x-4)^2/2\times 1/2} \qquad \text{for } -\infty < x < \infty$$

$$= \frac{1}{\sigma\times\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad \text{for } -\infty < x < \infty$$

with $\mu = 4$ and $\sigma^2 = ½$

Thus the given probability density function is that of a normal distribution with $\mu = 4$ and variance = ½.

Example 17.21: If the two quartiles of a normal distribution are 47.30 and 52.70 respectively, what is the mode of the distribution? Also find the mean deviation about median of this distribution.

Solution: The 1st and 3rd quartiles of N (μ, σ^2) are given by $(\mu - 0.675\ \sigma)$ and $(\mu + 0.675\ \sigma)$ respectively. As given,

$\mu - 0.675\ \sigma = 47.30$ …. (1)

$\mu + 0.675\ \sigma = 52.70$ …. (2)

Adding these two equations, we get

$2\ \mu = 100$ or $\mu = 50$

Thus Mode = Median = Mean = 50. Also $\sigma = 4$.

Also Mean deviation about median

= mean deviation about mode

= mean deviation about mean

$\cong 0.80\ \sigma$

= 3.20

Example 17.22: Find the points of inflexion of the normal curve

$$f(x) = \frac{1}{4\sqrt{2\pi}} . e^{-(x-10)^2/32}$$

for $-\infty < x < \infty$

Solution: Comparing f (x) to the probability densities function of a normal variable x , we find that $\mu = 10$ and $\sigma = 4$.

The points of inflexion are given by

$\mu - \sigma$ and $\mu + \sigma$

i.e. 10 – 4 and 10 + 4

i.e. 6 and 13.

Example 17.23: If x is a standard normal variable such that

$P(0 \le x \le b) = a$, what is the value of $P(|x| \ge b)$?

Solution: $P((x) \ge b)$

$= 1 - P(|x| \le b)$

$= 1 - P(-b \le x \le b)$

$= 1 - [P(0 \le x \le b) - P(-b \le x \le 0)]$

$= 1 - [P (0 \le x \le b) + P (0 \le x \le b)]$

$= 1 - 2a$

Example 17.24: X follows normal distribution with mean as 50 and variance as 100. What is $P(x \ge 60)$? Given $\phi(1) = 0.8413$

Solution: We are given that $x \sim N(\mu, \sigma^2)$ where

$\mu = 50$ and $\sigma^2 = 100 => \sigma = 10$

Thus $P(x \ge 60)$

$= 1 - P(x \le 60)$

$$= 1 - P\left(\frac{x-50}{10} \le \frac{60-50}{10}\right) = 1 - P(z \le 1)$$

$= 1 - \phi(1)$ (From 17.26)

$= 1 - 0.8413$

$\cong 0.16$

Example 17.25: If a random variable x follows normal distribution with mean as 120 and standard deviation as 40, what is the probability that $P(x \le 150 / x > 120)$?

Given that the area of the normal curve between $z = 0$ to $z = 0.75$ is 0.2734.

Solution: $P(x \le 150 / x > 120)$

$$= \frac{P(120 < x \le 150)}{P(x > 120)}$$

$$= \frac{P(120 < x \le 150)}{1 - P(x \le 120)}$$

$$= \frac{P\left(\frac{120-120}{40} \le \frac{x-120}{40} \le \frac{150-120}{40}\right)}{1 - P\left(\frac{x-120}{40} \le \frac{120-120}{40}\right)}$$

$$= \frac{P(0 < z \le 0.75)}{1 - P(z \le 0)}$$

$$= \frac{\phi(0.75) - \phi(0)}{1 - \phi(0)}$$ (From 17.29)

$$= \frac{0.7734 - 0.50}{1 - 0.50}$$

$\cong 0.55$ (ϕ (0.75) = Area of the normal curve between $z = -\infty$ to $z = 0.75$ = area between $-\infty$ to 0 + Area between 0 to 0.75 = 0.50 + 0.2734 = 0.7734)

Example 17.26: X is a normal variable with mean = 25 and SD 10. Find the value of b such that the probability of the interval [2 5, b] is 0.4772 given ϕ (2) = 0.9772.

Solution: We are given that $x \sim N(\mu, \sigma^2)$ where $\mu = 25$ and $\sigma = 10$

and $P[25 < x < b] = 0.4772$

$$\Rightarrow \left[\frac{25-25}{10} < \frac{x-25}{10} < \frac{b-25}{10}\right] = 0.4772$$

$$\Rightarrow P[0 < z < \frac{b-25}{10}] = 0.4772$$

$$\Rightarrow \phi\left(\frac{b-25}{10}\right) - \phi(0) = 0.4772$$

$$\Rightarrow \phi\left(\frac{b-25}{10}\right) - 0.50 = 0.4772$$

$$\Rightarrow \phi\left(\frac{b-25}{10}\right) = 0.9772$$

$$\Rightarrow \phi \frac{b-25}{10} = \phi(2) \qquad \text{(as given)}$$

$$\Rightarrow \frac{b-25}{10} = 2$$

$$\Rightarrow b = 25 + 2 \times 10 = 45.$$

Example 17.27: In a sample of 500 workers of a factory, the mean wage and SD of wages are found to be ₹ 500 and ₹ 48 respectively. Find the number of workers having wages:

(i) more than ₹ 600

(ii) less than ₹ 450

(iii) between ₹ 548 and ₹ 600.

Solution: Let X denote the wage of the workers in the factory. We assume that X is normally distributed with mean wage as ₹ 500 and standard deviation of wages as ₹ 48 respectively.

(i) Probability that a worker selected at random would have wage more than ₹ 600

$= P (X > 600)$

$= 1 - P (X \le 600)$

$$= 1 - P \left(\frac{X - 500}{48} \le \frac{600 - 500}{48} \right)$$

$= 1 - P (z \le 2.08)$

$= 1 - \phi (2.08)$

$= 1 - 0.9812$ (From Biometrika Table)

$= 0.0188$

Thus the number of workers having wages less than ₹ 600

$= 500 \times 0.0188$

$= 9.4$

$\cong 9$

(ii) Probability of a worker having wage less than ₹ 450

$= P (X < 450)$

$$= P \left(\frac{X - 500}{48} < \frac{450 - 500}{48} \right)$$

$= P(z < - 1.04)$

$= \phi (- 1.04)$

$= 1 - \phi (1.04)$ (from 17.26)

$= 1 - 0.8508$ (from Biometrika Table)

$= 0.1492$

Hence the number of workers having wages less than ₹ 450

$= 500 \times 0.1492$

$\cong 75$

(iii) Probability of a worker having wage between ₹ 548 and ₹ 600.

$= P (548 < x < 600)$

$$= P \left(\frac{548 - 500}{48} < \frac{x - 500}{48} < \frac{600 - 500}{48} \right)$$

$= P (1 < z < 2.08)$

$= \phi (2.08) - \phi (1)$

$= 0.9812 - 0.8413$ (consulting Biometrika)

$= 0.1399$

So the number of workers with wages between ₹ 548 and ₹ 600

$= 500 \times 0.1399$

$\cong 70.$

Example 17.28: The distribution of wages of a group of workers is known to be normal with mean ₹ 500 and SD ₹ 100. If the wages of 100 workers in the group are less than ₹ 430, what is the total number of workers in the group?

Solution: Let X denote the wage. It is given that X is normally distributed with mean as ₹ 500 and SD as ₹ 100 and P (X < 430) = 100/N, N being the total no. of workers in the group

$$\Rightarrow P\left(\frac{X-500}{100} < \frac{430-500}{100}\right) = \frac{100}{N}$$

$$\Rightarrow P(z < -0.70) = \frac{100}{N}$$

$$\Rightarrow \phi(-0.70) = \frac{100}{N}$$

$$\Rightarrow 1-\phi(0.70) = \frac{100}{N}$$

$$\Rightarrow 1-0.758 = \frac{100}{N}$$

$$\Rightarrow 0.242 = \frac{100}{N}$$

$$\Rightarrow N \cong 413.$$

Example 17.29: The mean height of 2000 students at a certain college is 165 cms and SD 9 cms. What is the probability that in a group of 5 students of that college, 3 or more students would have height more than 174 cm?

Solution: Let X denote the height of the students of the college. We assume that X is normally distributed with mean (μ) 165 cms and SD (σ) as 9 cms. If p denotes the probability that a student selected at random would have height more than 174 cms., then

$$p = P(X > 174)$$
$$= 1 - P(X \le 174)$$
$$= 1 - P\left(\frac{X-165}{9} \le \frac{174-165}{9}\right)$$
$$= 1 - P(z \le 1)$$
$$= 1 - \phi(1)$$
$$= 1 - 0.8413$$
$$= 0.1587$$

If y denotes the number of students having height more than 174 cm. in a group of 5 students then $y \sim \beta(n, p)$ where n = 5 and p = 0.1587. Thus the probability that 3 or more students would be more than 174 cm.

$$= p(y \ge 3)$$
$$= p(y = 3) + p(y = 4) + p(y = 5)$$
$$= 5_{C_3}(0.1587)^3.(0.8413)^2 + 5_{C_4}(0.1587)^4 \times (0.8413) + 5_{C_5}(0.1587)^5$$
$$= 0.02829 + 0.002668 + 0.000100$$
$$= 0.03106.$$

Example 17.30: The mean of a normal distribution is 500 and 16 per cent of the values are greater than 600. What is the standard deviation of the distribution?

(Given that the area between $z = 0$ to $z = 1$ is 0.34)

Solution: Let σ denote the standard deviation of the distribution.

We are given that

$$P(X > 600) = 0.16$$
$$\Rightarrow 1 - P(X \le 600) = 0.16$$
$$\Rightarrow P(X \le 600) = 0.84$$
$$\Rightarrow P\left(\frac{X-500}{\sigma} \le \frac{600-500}{\sigma}\right) = 0.84$$
$$\Rightarrow P\left(z \le \frac{100}{\sigma}\right) = 0.84$$
$$\Rightarrow \phi\left(\frac{100}{\sigma}\right) = \phi(1)$$

$$\Rightarrow \frac{(100)}{} = 1$$

$$\Rightarrow \sigma = 100.$$

Example 17.31: In a business, it is assumed that the average daily sales expressed in Rupees follows normal distribution.

Find the coefficient of variation of sales given that the probability that the average daily sales is less than ₹ 124 is 0.0287 and the probability that the average daily sales is more than ₹ 270 is 0.4599.

Solution: Let us denote the average daily sales by x and the mean and SD of x by μ and σ respectively. As given,

$$P(x < 124) = 0.0287 \text{ ....(1)}$$

$$P(x > 270) = 0.4599 \text{ ....(2)}$$

From (1), we have

$$P\left(\frac{X-\mu}{\sigma} < \frac{124-\mu}{\sigma}\right) = 0.0287$$

$$\Rightarrow P\left(z < \frac{124-\mu}{\sigma}\right) = 0.0287$$

$$\Rightarrow \phi\left(\frac{124-\mu}{\sigma}\right) = 0.0287$$

$$\Rightarrow 1-\phi\left(\frac{\mu-124}{\sigma}\right) = 0.0287$$

$$\Rightarrow \phi\left(\frac{\mu-124}{\sigma}\right) = 0.9713$$

$$\Rightarrow \phi\left(\frac{\mu-124}{\sigma}\right) = \phi(2.085) \text{ (From Biometrika)}$$

$$\Rightarrow \left(\frac{\mu-124}{\sigma}\right) = 2.085 \text{ ....(3)}$$

From (2) we have,

$$1 - P(x \le 270) = 0.4599$$

$$\Rightarrow P\left(\frac{X-\mu}{\sigma} \le \frac{270-\mu}{\sigma}\right) = 0.5401$$

$$\Rightarrow \phi\left(\frac{270-\mu}{\sigma}\right) = 0.5401$$

$$\Rightarrow \phi\left(\frac{270-\mu}{\sigma}\right) = \phi\ (0.1)$$

$$\Rightarrow \left(\frac{270-\mu}{\sigma}\right) = 0.1 \quad .....(4)$$

Dividing (3) by (4), we get

$$\frac{\mu-124}{270-\mu} = 20.85$$

$$\Rightarrow \mu - 124 = 5629.50 - 20.85\ \mu$$

$$\Rightarrow \mu = 5753.50/21.85$$

$$= 263.32$$

Substituting this value of μ in (3), we get

$$\frac{263.32 - 124}{\sigma} = 2.085$$

$$\Rightarrow \sigma = 73$$

Thus the coefficient of variation of sales

$$= \sigma/\mu \times 100$$

$$= \frac{73}{263.32} \times 100$$

$$= 25.38$$

Example 17.32: x and y are independent normal variables with mean 100 and 80 respectively and standard deviation as 4 and 3 respectively. What is the distribution of (x + y)?

Solution: We know that if $x \sim N(\mu_1, \sigma_1^2)$ and $y \sim N(\mu_2, \sigma_2^2)$ and they are independent, then $z = x + y$ follows normal with mean $(\mu_1 + \mu_2)$ and

$$SD = \sqrt{\sigma_1^2 + \sigma_2^2} \text{ respectively.}$$

Thus the distribution of (x + y) is normal with mean (100 + 80) or 180

and SD $\sqrt{4^2+3^2} = 5$

Standard Normal Distribution:

If a continuous random variable z follows standard normal distribution, to be denoted by $z \sim N(0, 1)$, then the probability density function of z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{for } -\infty < z < \infty \;\ldots\ldots (17.35)$$

Some important properties of z are listed below :

(i) z has mean, median and mode all equal to zero.

(ii) The standard deviation of z is 1. Also the approximate values of mean deviation and quartile deviation are 0.8 and 0.675 respectively.

(iii) The standard normal distribution is symmetrical about $z = 0$.

(iv) The two points of inflexion of the probability curve of the standard normal distribution are –1 and 1.

(v) The two tails of the standard normal curve never touch the horizontal axis.

(vi) The upper and lower p per cent points of the standard normal variable z are given by

$$P(Z > z_p) = p \;\ldots\ldots (17.36)$$

$$\text{And } P(Z < z_{1-p}) = p$$

$$\text{i.e. } P(Z < -z_p) = p \text{ respectively } \ldots (17.37)$$

(since for a standard normal distribution $z_{1-p} = -z_p$)

Selecting P = 0.005, 0.025, 0.01 and 0.05 respectively,

We have $\mathbf{z_{0.005} = 2.58}$

$\mathbf{z_{0.025} = 1.96}$

$\mathbf{z_{0.01} = 2.33}$

$\mathbf{z_{0.05} = 1.645}$ ………… (17.38)

These are shown in fig 17.3.

(vii) If $\overline{x}$ denotes the arithmetic mean of a random sample of size n drawn from a normal population then,

$$Z = \frac{\sqrt{n}\,(\overline{x} - \mu)}{\sigma} \sim N\,(0, 1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (17.39)$$

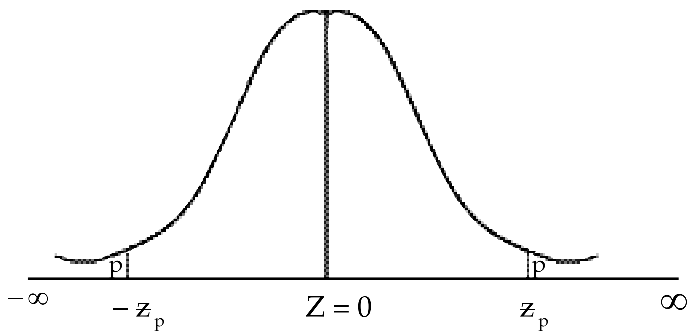


Fig. 17.3

Showing upper and lower p % points of the standard normal variable.

SUMMARY

- A probability distribution also possesses all the characteristics of an observed distribution. We define population mean (μ), population median $(\tilde{\mu})$, population mode (μ_0), population standard deviation (σ) etc. exactly same way we have done earlier. These characteristics are known as population parameters.
- Probability distribution or a Continuous probability distribution depending on the random variable under study.
- Two important discrete probability distributions are (a) Binomial Distribution and (b) Poisson distribution.
- Normal Distribution is a important continuous probability distribution
- A discrete random variable x is defined to follow binomial distribution with parameters n and p, to be denoted by x ~ B (n, p), if the probability mass function of x is given by

  $f(x) = p(X = x) = {}^{n}c_{x}\, p^{x}\, q^{n-x}$ for x = 0, 1, 2, …., n

  = 0, otherwise

- Additive property of binomial distribution.

  If X and Y are two independent variables such that

  $X \sim \beta\ (n_1, P)$

  and $Y \sim \beta\ (n_2, P)$

  Then $(X+Y) \sim \beta\ (n_1 + n_2,\ P)$

- **Definition of Poisson Distribution**

  A random variable X is defined to follow Poisson distribution with parameter λ, to be denoted by X ~ P (m) if the probability mass function of x is given by

  $$f(x) = P(X = x) = \frac{e^{-m}.m^x}{x!} \text{ for } x = 0, 1, 2, \ldots \infty$$

  $$= 0 \quad \text{otherwise}$$

  (i) Since $e^{-m} = 1/e^m > 0$, whatever may be the value of m, m > 0, it follows that $f(x) \geq 0$ for every x.

  Also it can be established that $\sum_x f(x) = 1$ i.e. $f(0) + f(1) + f(2) + \ldots\ldots = 1$

  (ii) Poisson distribution is known as a uniparametric distribution as it is characterised by only one parameter m.

  (iii) The mean of Poisson distribution is given by m i.e $\mu = m$.

  (iv) The variance of Poisson distribution is given by $\sigma^2 = m$

  (v) Like binomial distribution, Poisson distribution could be also unimodal or bimodal depending upon the value of the parameter m.

  (vi) Poisson approximation to Binomial distribution

  (vii) Additive property of Poisson distribution

- A continuous random variable x is defined to follow normal distribution with parameters μ and σ^2, to be denoted by

  $X \sim N(\mu, \sigma^2)$

  If the probability density function of the random variable x is given by

  $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}.e^{-(\bar{x}-u)^2/2\sigma^2}$$

  for $-\infty < x < \infty$

  where μ and σ are constants, and $\sigma > 0$

- **Properties of Normal Distribution**

1. Since π = 22/7 , $e^{-\theta} = 1 / e^{\theta} > 0$, whatever θ may be,

   it follows that f (x) $\geq$ 0 for every x.

   It can be shown that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

2. The mean of the normal distribution is given by μ. Further, since the distribution is symmetrical about x = μ, it follows that the mean, median and mode of a normal distribution coincide, all being equal to μ.

3. The standard deviation of the normal distribution is given by σ.

   Mean deviation of normal distribution is

   $\sigma\sqrt{2} \cong \mathbf{0.8\sigma}$

   The first and third quartiles are given by

   $Q_1 = \mu - 0.675\ \sigma$

   and $Q_3 = \mu + 0.675\ \sigma$

   so that, quartile deviation = 0.675 σ

4. The normal distribution is symmetrical about x = μ. As such, its skewness is zero i.e. the normal curve is neither inclined move towards the right (negatively skewed) nor towards the left (positively skewed).

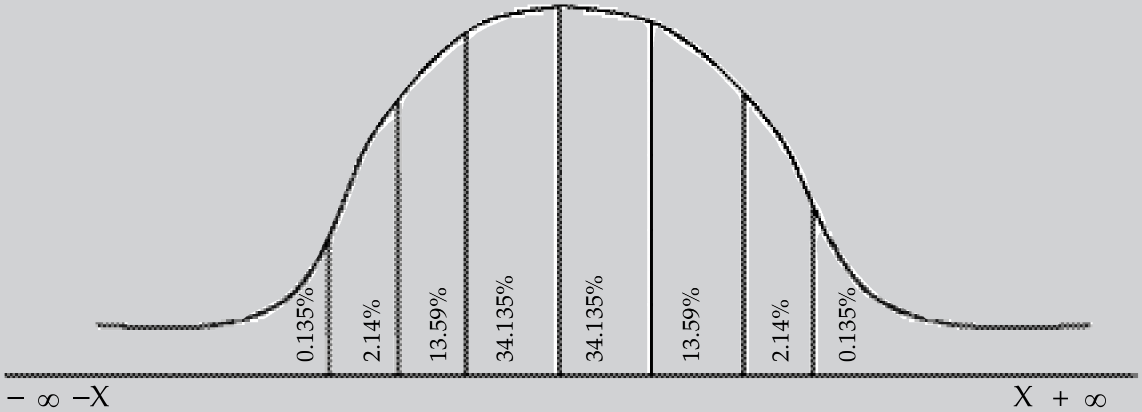
5. The normal curve y = f (x) has two points of inflexion to be given by x = $\mu - \sigma$ and x = $\mu + \sigma$ i.e. at these two points, the normal curve changes its curvature from concave to convex and from convex to concave.

6. If $x \sim N(\mu, \sigma^2)$ then $z = x - \mu/\sigma \sim N(0, 1)$, z is known as standardised normal variate or normal deviate.

   We also have $P(z \leq k) = \phi(k)$

7. Area under the normal curve is shown in the following figure :

$\mu - 3\sigma$	$\mu - 2\sigma$	$\mu - \sigma$	$x = \mu$	$\mu + \sigma$	$\mu + 2\sigma$	$\mu + 3\sigma$
(z = –3)	(z = –2)	(z = –1)	(z = 0)	(z = 1)	(z = 2)	(z = 3)



$P(\mu - \sigma < x < \mu + \sigma) = 0.6828$

$\Rightarrow P(-1 < z < 1) = 0.6828$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9546$

$\Rightarrow P(-2 < z < 2) = 0.9546$

and $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

$\Rightarrow P(-3 < z < 3) = 0.9973$.

8. We note that 99.73 per cent of the values of a normal variable lies between $(\mu - 3\sigma)$ and $(\mu + 3\sigma)$. Thus the probability that a value of x lies outside that limit is as low as 0.0027.

9. If x and y are independent normal variables with means and standard deviations as μ_1 and μ_2 and σ_1, and σ_2 respectively, then $z = x + y$ also follows normal distribution with mean $(\mu_1 + \mu_2)$ and SD = $\sqrt{\sigma_1^2+\sigma_2^2}$ respectively.

- **Standard Normal Distribution**

If a continuous random variable z follows standard normal distribution, to be denoted by $z \sim N(0, 1)$, then the probability density function of z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{for } -\infty < z < \infty$$

Some important properties of z are listed below :

(i) z has mean, median and mode all equal to zero.

(ii) The standard deviation of z is 1. Also the approximate values of mean deviation and quartile deviation are 0.8 and 0.675 respectively.

(iii) The standard normal distribution is symmetrical about $z = 0$.

(iv) The two points of inflexion of the probability curve of the standard normal distribution are –1 and 1.

(v) The two tails of the standard normal curve never touch the horizontal axis.

(vi) The upper and lower p per cent points of the standard normal variable z are given by

$P(Z > z_p) = p$

And $P(Z < z_{1-p}) = p$

i.e. $P(Z < -z_p) = p$ respectively

(since for a standard normal distribution $z_{1-p} = -z_p$)

Selecting P = 0.005, 0.025, 0.01 and 0.05 respectively,

We have $z_{0.005} = 2.58$

$z_{0.025} = 1.96$

$z_{0.01} = 2.33$

$z_{0.05} = 1.645$

These are shown in fig 13.3.

(vii) If $\bar{x}$ denotes the arithmetic mean of a random sample of size n drawn from a normal population then,

$$Z = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0, 1)$$

EXERCISE

Set : A

Write down the correct answers. Each question carries 1 mark.

1. A theoretical probability distribution.
   (a) does not exist.
   (b) exists only in theory.
   (c) exists in real life.
   (d) both (b) and (c).
2. Probability distribution may be
   (a) discrete.
   (b) continuous.
   (c) infinite.
   (d) (a) or (b).
3. An important discrete probability distribution is
   (a) Poisson distribution.
   (b) Normal distribution.
   (c) Cauchy distribution.
   (d) Log normal distribution.
4. An important continuous probability distribution
   (a) Binomial distribution.
   (b) Poisson distribution.
   (c) Geometric distribution.
   (d) Normal distribution.

5. Parameter is a characteristic of
   (a) population. (b) sample. (c) probability distribution. (d) both (a) and (b).
6. An example of a parameter is
   (a) sample mean. (b) population mean.
   (c) binomial distribution. (d) sample size.
7. A trial is an attempt to
   (a) make something possible. (b) make something impossible.
   (c) prosecute an offender in a court of law.
   (d) produce an outcome which is neither certain nor impossible.
8. The important characteristic(s) of Bernoulli trials
   (a) each trial is associated with just two possible outcomes.
   (b) trials are independent. (c) trials are infinite.
   (d) both (a) and (b).
9. The probability mass function of binomial distribution is given by
   (a) $f(x) = p^x q^{n-x}$. (b) $f(x) = {}^nc_x p^x q^{n-x}$.
   (c) $f(x) = {}^nc_x q^x p^{n-x}$. (d) $f(x) = {}^nc_x p^{n-x} q^x$.
10. If x is a binomial variable with parameters n and p, then x can assume
    (a) any value between 0 and n.
    (b) any value between 0 and n, both inclusive.
    (c) any whole number between 0 and n, both inclusive.
    (d) any number between 0 and infinity.
11. A binomial distribution is
    (a) never symmetrical. (b) never positively skewed.
    (c) never negatively skewed. (d) symmetrical when p = 0.5.
12. The mean of a binomial distribution with parameter n and p is
    (a) n (1– p). (b) np (1 – p). (c) np. (d) $\sqrt{np(1-p)}$.
13. The variance of a binomial distribution with parameters n and p is
    (a) $np^2 (1 - p)$. (b) $\sqrt{np(1-p)}$. (c) nq (1 – q). (d) $n^2p^2 (1- p)^2$.
14. An example of a bi-parametric discrete probability distribution is
    (a) binomial distribution. (b) poisson distribution.
    (c) normal distribution. (d) both (a) and (b).

15. For a binomial distribution, mean and mode
    (a) are never equal. (b) are always equal.
    (c) are equal when q = 0.50. (d) do not always exist.
16. The mean of binomial distribution is
    (a) always more than its variance. (b) always equal to its variance.
    (c) always less than its variance. (d) always equal to its standard deviation.
17. For a binomial distribution, there may be
    (a) one mode. (b) two modes. (c) (a). (d) (a) or (b).
18. The maximum value of the variance of a binomial distribution with parameters n and p is
    (a) n/2. (b) n/4. (c) np (1 – p). (d) 2n.
19. The method usually applied for fitting a binomial distribution is known as
    (a) method of least square. (b) method of moments.
    (c) method of probability distribution. (d) method of deviations.
20. Which one is not a condition of Poisson model?
    (a) the probability of having success in a small time interval is constant.
    (b) the probability of having success more than one in a small time interval is very small.
    (c) the probability of having success in a small interval is independent of time and also of earlier success.
    (d) the probability of having success in a small time interval (t, t + dt) is kt for a positive constant k.
21. Which one is uniparametric distribution?
    (a) Binomial. (b) Poisson. (c) Normal. (d) Hyper geometric.
22. For a Poisson distribution,
    (a) mean and standard deviation are equal. (b) mean and variance are equal.
    (c) standard deviation and variance are equal. (d) both (a) and (b).
23. Poisson distribution may be
    (a) unimodal. (b) bimodal. (c) Multi-modal. (d) (a) or (b).
24. Poisson distribution is
    (a) always symmetric. (b) always positively skewed.
    (c) always negatively skewed. (d) symmetric only when m = 2.
25. A binomial distribution with parameters n and p can be approximated by a Poisson distribution with parameter m = np is

(a) $n \to \infty$.
(b) $p \to 0$.
(c) $n \to \infty$ and $p \to 0$.
(d) $n \to \infty$ and $p \to 0$ so that np remains finite..

26. For Poisson fitting to an observed frequency distribution,
(a) we equate the Poisson parameter to the mean of the frequency distribution.
(b) we equate the Poisson parameter to the median of the distribution.
(c) we equate the Poisson parameter to the mode of the distribution.
(d) none of these.

27. The most important continuous probability distribution is known as
(a) Binomial distribution.
(b) Normal distribution.
(c) Chi-square distribution.
(d) Sampling distribution.

28. The probability density function of a normal variable x is given by

(a) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for $-\propto < x < \propto$.

(b) $f(x) = f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{-(x-\mu)^2}{2\sigma^2}}$ for $0 < x < \propto$.

(c) $f(x) = \frac{1}{\sqrt{2\pi\sigma}} . e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for $-\propto < x < \propto$.

(d) none of these.

29. The total area of the normal curve is
(a) one.
(b) 50 per cent.
(c) 0.50.
(d) any value between 0 and 1.

30. The normal curve is
(a) Bell-shaped.
(b) U- shaped.
(c) J-shaped.
(d) Inverted J-shaped.

31. The normal curve is
(a) positively skewed.
(b) negatively skewed.
(c) symmetrical.
(d) all these.

32. Area of the normal curve
(a) between $-\propto$ to μ is 0.50.
(b) between μ to $\propto$ is 0.50.
(c) between $-\propto$ to $\propto$ is 0.50.
(d) both (a) and (b).

33. The cumulative distribution function of a random variable X is given by

(a) $F(x) = P(X \leq x)$.
(b) $F(X) = P(X \leq x)$.
(c) $F(x) = P(X \geq x)$.
(d) $F(x) = P(X = x)$.

34. The mean and mode of a normal distribution

(a) may be equal.
(b) may be different.
(c) are always equal.
(d) (a) or (b).

35. The mean deviation about median of a standard normal variate is

(a) 0.675 σ.
(b) 0.675.
(c) 0.80 σ.
(d) 0.80.

36. The quartile deviation of a normal distribution with mean 10 and SD 4 is

(a) 0.675.
(b) 67.50.
(c) 2.70.
(d) 3.20.

37. For a standard normal distribution, the points of inflexion are given by

(a) $\mu - \sigma$ and $\mu + \sigma$.
(b) $-\sigma$ and σ.
(c) -1 and 1.
(d) 0 and 1.

38. The symbol ϕ (a) indicates the area of the standard normal curve between

(a) 0 to a.
(b) a to ∞.
(c) $-\propto$ to a.
(d) $-\propto$ to $\propto$.

39. The interval $(\mu - 3\sigma, \mu + 3\sigma)$ covers

(a) 95% area of a normal distribution.
(b) 96% area of a normal distribution.
(c) 99% area of a normal distribution.
(d) all but 0.27% area of a normal distribution.

40. Number of misprints per page of a thick book follows

(a) Normal distribution.
(b) Poisson distribution.
(c) Binomial distribution.
(d) Standard normal distribution.

41. The results of ODI matches between India and Pakistan follows

(a) Binomial distribution.
(b) Poisson distribution.
(c) Normal distribution.
(d) (b) or (c).

42. The wage of workers of a factory follow

(a) Binomial distribution.
(b) Poisson distribution.
(c) Normal distribution.
(d) Chi-square distribution.

43. If X and Y are two independent normal random variables, then the distribution of (X+Y) is

(a) normal.
(b) standard normal.
(c) T.
(d) chi-square.

Set B :

Write down the correct answers. Each question carries 2 marks.

1. What is the standard deviation of the number of recoveries among 48 patients when the probability of recovering is 0.75?

   (a) 36. (b) 81. (c) 9. (d) 3.

2. X is a binomial variable with n = 20. What is the mean of X if it is known that x is symmetric?

   (a) 5. (b) 10. (c) 2. (d) 8.

3. If X ~ B (n, p), what would be the greatest value of the variance of x when n = 16?

   (a) 2. (b) 4. (c) 8. (d) $\sqrt{5}$.

4. If x is a binomial variate with parameter 15 and 1/3, what is the value of mode of the distribution?

   (a) 5 and 6. (b) 5. (c) 5.50. (d) 6.

5. What is the number of trials of a binomial distribution having mean and SD as 3 and 1.5 respectively?

   (a) 2. (b) 4. (c) 8. (d) 12.

6. What is the probability of getting 3 heads if 6 unbiased coins are tossed simultaneously?

   (a) 0.50. (b) 0.25. (c) 0.3125. (d) 0.6875.

7. If the overall percentage of success in an exam is 60, what is the probability that out of a group of 4 students, at least one has passed?

   (a) 0.6525. (b) 0.9744. (c) 0.8704. (d) 0.0256.

8. What is the probability of making 3 correct guesses in 5 True – False answer type questions?

   (a) 0.3125. (b) 0.5676. (c) 0.6875. (d) 0.4325

9. If the standard deviation of a Poisson variate X is 2, what is P (1.5 < X < 2.9)?

   (a) 0.231. (b) 0.158. (c) 0.15. (d) 0.144.

10. If the mean of a Poisson variable X is 1, what is P (X = takes the value at least 1)?

    (a) 0.456. (b) 0.821. (c) 0.632. (d) 0.254.

11. If X ~ P (m) and its coefficient of variation is 50, what is the probability that X would assume only non-zero values?

    (a) 0.018. (b) 0.982. (c) 0.989. (d) 0.976.

12. If 1.5 per cent of items produced by a manufacturing units are known to be defective, what is the probability that a sample of 200 items would contain no defective item?

    (a) 0.05. (b) 0.15. (c) 0.20. (d) 0.22.

13. For a Poisson variate X, P (X = 1) = P (X = 2). What is the mean of X?

    (a) 1.00. (b) 1.50. (c) 2.00. (d) 2.50.

14. If 1 per cent of an airline's flights suffer a minor equipment failure in an aircraft, what is the probability that there will be exactly two such failures in the next 100 such flights?

(a) 0.50. (b) 0.184. (c) 0.265. (d) 0.256.

15. If for a Poisson variable X, f(2) = 3 f(4), what is the variance of X?

(a) 2. (b) 4. (c) $\sqrt{2}$. (d) 3.

16. What is the coefficient of variation of x, characterised by the following probability density function: $f(x) = \frac{1}{4\sqrt{2\pi}} e^{-(x-10)^2/32}$ for $-\infty < x < \infty$

(a) 50. (b) 60. (c) 40. (d) 30.

17. What is the first quartile of X having the following probability density function?

$$f(x) = \frac{1}{\sqrt{72\pi}} e^{-(x-10)^2/72} \quad \text{for } -\infty < x < \infty$$

(a) 4. (b) 5. (c) 5.95. (d) 6.75.

18. If the two quartiles of N (μ , σ^2) are 14.6 and 25.4 respectively, what is the standard deviation of the distribution?

(a) 9. (b) 6. (c) 10. (d) 8.

19. If the mean deviation of a normal variable is 16, what is its quartile deviation?

(a) 10.00. (b) 13.50. (c) 15.00. (d) 12.05.

20. If the points of inflexion of a normal curve are 40 and 60 respectively, then its mean deviation is

(a) 40. (b) 45. (c) 50. (d) 60.

21. If the quartile deviation of a normal curve is 4.05, then its mean deviation is

(a) 5.26. (b) 6.24. (c) 4.24. (d) 4.80.

22. If the Ist quartile and mean deviation about median of a normal distribution are 13.25 and 8 respectively, then the mode of the distribution is

(a) 20. (b) 10. (c) 15. (d) 12.

23. If the area of standard normal curve between z = 0 to z = 1 is 0.3413, then the value of ϕ (1) is

(a) 0.5000. (b) 0.8413. (c) –0.5000. (d) 1.

24. If X and Y are 2 independent normal variables with mean as 10 and 12 and SD as 3 and 4, then (X+Y) is normally distributed with

(a) mean = 22 and SD = 7. (b) mean = 22 and SD = 25.

(c) mean = 22 and SD = 5. (d) mean = 22 and SD = 49.

Set : C

Answer the following questions. Each question carries 5 marks.

1. If it is known that the probability of a missile hitting a target is 1/8, what is the probability that out of 10 missiles fired, at least 2 will hit the target?

   (a) 0.4258. (b) 0.3968. (c) 0.5238. (d) 0.3611.

2. X is a binomial variable such that 2 P(X = 2) = P(X = 3) and mean of X is known to be 10/3. What would be the probability that X assumes at most the value 2?

   (a) 16/81. (b) 17/81. (c) 47/243. (d) 46/243.

3. Assuming that one-third of the population is tea drinkers and each of 1000 enumerators takes a sample of 8 individuals to find out whether they are tea drinkers or not, how many enumerators are expected to report that five or more people are tea drinkers?

   (a) 100. (b) 95. (c) 88. (d) 90.

4. If a random variable X follows binomial distribution with mean as 5 and satisfying the condition 10 P (X = 0) = P (X = 1), what is the value of P $(x \geq 1/ x > 0)$?

   (a) 0.67. (b) 0.56. (c) 0.99. (d) 0.82.

5. Out of 128 families with 4 children each, how many are expected to have at least one boy and one girl?

   (a) 100. (b) 105. (c) 108. (d) 112.

6. In 10 independent rollings of a biased die, the probability that an even number will appear 5 times is twice the probability that an even number will appear 4 times. What is the probability that an even number will appear twice when the die is rolled 8 times?

   (a) 0.0304. (b) 0.1243. (c) 0.2315. (d) 0.1926.

7. If a binomial distribution is fitted to the following data:

x:	0	1	2	3	4
f:	16	25	32	17	10

   then the sum of the expected frequencies for x = 2, 3 and 4 would be

   (a) 58. (b) 59. (c) 60. (d) 61.

8. If X follows normal distribution with $\mu = 50$ and $\sigma = 10$, what is the value of P $(x \leq 60 \ / \ x > 50)$?

   (a) 0.8413. (b) 0.6828. (c) 0.1587. (d) 0.7256.

9. X is a Poisson variate satisfying the following condition 9 P (X = 4) + 90 P (X = 6) = P (X = 2). What is the value of P $(X \leq 1)$?

   (a) 0.5655 (b) 0.6559 (c) 0.7358 (d) 0.8201

10. A random variable x follows Poisson distribution and its coefficient of variation is 50. What is the value of P (x > 1 / x > 0)?

(a) 0.1876 (b) 0.2341 (c) 0.9254 (d) 0.8756

11. A renowned hospital usually admits 200 patients every day. One per cent patients, on an average, require special room facilities. On one particular morning, it was found that only one special room is available. What is the probability that more than 3 patients would require special room facilities?

(a) 0.1428 (b) 0.1732 (c) 0.2235 (d) 0.3450

12. A car hire firm has 2 cars which is hired out everyday. The number of demands per day for a car follows Poisson distribution with mean 1.20. What is the proportion of days on which some demand is refused? (Given $e^{1.20} = 3.32$).

(a) 0.25 (b) 0.3012 (c) 0.12 (d) 0.03

13. If a Poisson distribution is fitted to the following data:

Mistake per page	0	1	2	3	4	5
Number of pages	76	74	29	17	3	1

Then the sum of the expected frequencies for x = 0, 1 and 2 is

(a) 150. (b) 184. (c) 165. (d) 148.

14. The number of accidents in a year attributed to taxi drivers in a locality follows Poisson distribution with an average 2. Out of 500 taxi drivers of that area, what is the number of drivers with at least 3 accidents in a year?

(a) 162 (b) 180 (c) 201 (d) 190

15. In a sample of 800 students, the mean weight and standard deviation of weight are found to be 50 kg and 20 kg respectively. On the assumption of normality, what is the number of students weighing between 46 Kg and 62 Kg? Given area of the standard normal curve between $z = 0$ to $z = 0.20 = 0.0793$ and area between $z = 0$ to $z = 0.60 = 0.2257$.

(a) 250 (b) 244 (c) 240 (d) 260

16. The salary of workers of a factory is known to follow normal distribution with an average salary of ₹ 10,000 and standard deviation of salary as ₹ 2,000. If 50 workers receive salary more than ₹ 14,000, then the total no. of workers in the factory is

(a) 2,193 (b) 2,000 (c) 2,200 (d) 2,500

17. For a normal distribution with mean as 500 and SD as 120, what is the value of k so that the interval [500, k] covers 40.32 per cent area of the normal curve? Given ϕ (1.30) = 0.9032.

(a) 740 (b) 750 (c) 656 (d) 800

18. The average weekly food expenditure of a group of families has a normal distribution with mean ₹ 1,800 and standard deviation ₹ 300. What is the probability that out of 5 families belonging to this group, at least one family has weekly food expenditure in excess of ₹ 1,800? Given ϕ (1) = 0.84.

(a) 0.418 (b) 0.582 (c) 0.386 (d) 0.614

19. If the weekly wages of 5000 workers in a factory follows normal distribution with mean and SD as ₹ 700 and ₹ 50 respectively, what is the expected number of workers with wages between ₹ 660 and ₹ 720?

(a) 2,050 (b) 2,200 (c) 2,218 (d) 2,300

20. 50 per cent of a certain product have weight 60 kg or more whereas 10 per cent have weight 55 kg or less. On the assumption of normality, what is the variance of weight?

Given ϕ (1.28) = 0.90.

(a) 15.21 (b) 9.00 (c) 16.00 (d) 22.68

ANSWERS

Set : A

1.	(d)	**2.**	(d)	**3.**	(a)	**4.**	(d)	**5.**	(a)	**6.**	(b)	**7.**	(d)	**8.**	(d)
9.	(b)	**10.**	(c)	**11.**	(d)	**12.**	(c)	**13.**	(c)	**14.**	(a)	**15.**	(c)	**16.**	(a)
17.	(c)	**18.**	(b)	**19.**	(b)	**20.**	(a)	**21.**	(b)	**22.**	(b)	**23.**	(d)	**24.**	(b)
25.	(d)	**26.**	(a)	**27.**	(b)	**28.**	(a)	**29.**	(a)	**30.**	(a)	**31.**	(c)	**32**	(d)
33.	(a)	**34.**	(c)	**35.**	(d)	**36.**	(c)	**37.**	(c)	**38.**	(c)	**39.**	(d)	**40.**	(b)
41.	(a)	**42.**	(c)	**43.**	(a)										

Set : B

1.	(d)	**2.**	(b)	**3.**	(b)	**4.**	(b)	**5.**	(d)	**6.**	(c)	**7.**	(b)	**8.**	(a)
9.	(d)	**10.**	(c)	**11.**	(b)	**12.**	(a)	**13.**	(c)	**14.**	(b)	**15.**	(a)	**16.**	(c)
17.	(c)	**18.**	(d)	**19.**	(b)	**20.**	(a)	**21.**	(d)	**22.**	(a)	**23.**	(b)	**24.**	(c)

Set : C

1.	(d)	**2.**	(b)	**3.**	(c)	**4.**	(c)	**5.**	(d)	**6.**	(a)	**7.**	(d)	**8.**	(b)
9.	(c)	**10.**	(c)	**11.**	(a)	**12.**	(d)	**13.**	(b)	**14.**	(a)	**15.**	(b)	**16.**	(a)
17.	(c)	**18.**	(b)	**19.**	(c)	**20.**	(a)								

ADDITIONAL QUESTION BANK

1. When a coin is tossed 10 times then we use

   (a) Normal Distribution (b) Poisson Distribution
   (c) Binomial Distribution (d) None

2. In Binomial Distribution 'n' means

   (a) Number of trials of the experiment (b) the probability of getting success
   (c) Number of success (d) none

3. Binomial probability Distribution is a

   (a) Continuous (b) discrete
   (c) both (d) none

4. When there are a fixed number of repeated trial of any experiments under identical conditions for which only one of two mutually exclusive outcomes, success or failure can result in each trial then, we use

   (a) Normal Distribution (b) Binomial Distribution

   (c) Poisson Distribution (d) None

5. In Binomial Distribution 'p' denotes Probability of

   (a) Success (b) Failure (c) Both (d) None

6. When p = 0.5 5, the binomial distribution is

   (a) asymmetrical (b) symmetrical (c) Both (d) None

7. When 'p' is larger than 0. 5, the binomial distribution is

   (a) asymmetrical (b) symmetrical (c) Both (d) None

8. Mean of Binomial distribution is

   (a) npq (b) np (c) both (d) none

9. Variance of Binomial distribution is

   (a) npq (b) np (c) both (d) none

10. When p = 0.1 the binomial distribution is skewed to the

    (a) left (b) right (c) both (d) none

11. If in Binomial distribution np = 9 and npq = 2. 25 then q is equal to

    (a) 0.25 (b) 0.75 (c) 1 (d) none

12. In Binomial Distribution

    (a) mean is greater than variance (b) mean is less than variance
    (c) mean is equal to variance (d) none

13. Standard deviation of binomial distribution is

(a) $(npq)^2$ (b) $\sqrt{npq}$

(c) $(np)^2$ (d) $\sqrt{np}$

14. _________ distribution is a limiting case of Binomial distribution

(a) Normal (b) Poisson (c) Both (d) none

15. When the number of trials is large and probability of success is small then we use the distribution

(a) Normal (b) Poisson

(c) Binomial (d) none

16. In Poisson Distribution, probability of success is very close to

(a) 1 (b) – 1 (c) 0 (d) none

17. In Poisson Distribution np is

(a) finite (b) infinite (c) 0 (d) none

18. In ________________ distribution, mean = variance

(a) Normal (b) Binomial (c) Poisson (d) none

19. In Poisson distribution mean is equal to

(a) (λ) (b) np (c) square root mp (d) square root mpq

20. In Binomial distribution standard deviation is equal to

(a) $\sqrt{np}$ (b) $(np)^2$ (c) $\sqrt{npq}$ (d) $(npq)^2$

21. For continuous events _________________ distribution is used.

(a) Normal (b) Poisson (c) Binomial (d) none

22. Probability density function is associated with

(a) discrete random variable (b) continuous random variables

(c) both (d) none

23. Probability density function is always

(a) greater than 0 (b) greater than equal to 0

(c) less than 0 (d) less than equal to 0

24. For continuous random variables probability of the entire space is

(a) 0 (b) –1 (c) 1 (d) none

25. For discrete random variables the probability of the entire space is

(a) 0 (b) 1 (c) –1 (d) none

26. Binomial distribution is symmetrical if

(a) $p > q$ (b) $p < q$ (c) $p = q$ (d) none

27. The Poisson distribution tends to be symmetrical if the mean value is

(a) high (b) low (c) zero (d) none

28. The curve of ____________ distribution has single peak

(a) Poisson (b) Binomial (c) Normal (d) none

29. The curve of _________ distribution is unimodal and bell shaped with the highest point over the mean

(a) Poisson (b) Normal (c) Binomial (d) none

30. Because of the symmetry of Normal distribution the median and the mode have the ______ value as that of the mean

(a) greater (b) smaller (c) same (d) none

31. For a Normal distribution, the total area under the normal curve is

(a) 0 (b) 1 (c) 2 (d) –1

32. In Normal distribution the probability has the maximum value at the

(a) mode (b) mean (c) median (d) none

33. In Normal distribution the probability decreases gradually on either side of the mean but never touches the axis.

(a) True (b) false (c) both (d) none

34. Whatever may be the parameter of __________ distribution, it has same shape.

(a) Normal (b) Binomial (c) Poisson (d) none

35. In Standard Normal distribution

(a) mean=1, S.D=0 (b) mean=1, S.D=1
(c) mean = 0, S.D = 1 (d) mean=0, S.D=0

36. The Number of methods for fitting the normal curve is

(a) 1 (b) 2 (c) 3 (d) 4

37. ____________ distribution is symmetrical around t = 0

(a) Normal (b) Poisson (c) Binomial (d) t

38. As the degree of freedom increases, the ________ distribution approaches the Standard Normal distribution

(a) t (b) Binomial (c) Poisson (d) Normal

39. _________ distribution is asymptotic to the horizontal axis.

(a) Binomial (b) Normal (c) Poisson (d) t

40. ________ distribution has a greater spread than Normal distribution curve

(a) t (b) Binomial (c) Poisson (d) none

41. In Binomial Distribution if n is infinitely large, the probability p of occurrence of event' is close to _______ and q is close to _________

(a) 0, 1 (b) 1, 0 (c) 1, 1 (d) none

42. Poisson distribution approaches a Normal distribution as n

(a) increase infinitely (b) decrease (c) increases moderately (d) none

43. If neither p nor q is very small but n sufficiently large, the Binomial distribution is very closely approximated by _________ distribution

(a) Poisson (b) Normal (c) t (d) none

44. For discrete random variable x, Expected value of x (i.e E(x)) is defined as the sum of products of the different values and the corresponding probabilities.

(a) True (b) false (c) both (d) none

45. For a probability distribution, ____________ is the expected value of x.

(a) median (b) mode (c) mean (d) none

46. _________ is the expected value of $(x - m)^2$, where m is the mean.

(a) median (b) variance (c) standard deviation (d) mode

47. The probability distribution of x is given below :

value of x :	1	0	Total
probability :	p	1–p	1

Mean is equal to

(a) p (b) 1–p (c) 0 (d) 1

48. For n independent trials in Binomial distribution the sum of the powers of p and q is always n , whatever be the no. of success.

(a) True (b) false (c) both (d) none

49. In Binomial distribution parameters are

(a) n and q (b) n and p (c) p and q (d) none

50. In Binomial distribution if n = 4 and p = 1/3 then the value of variance is

(a) 8/3 (b) 8/9 (c) 4/3 (d) none

51. In Binomial distribution if mean = 20, S.D.= 4 then q is equal to

(a) 2/5 (b) 3/8 (c) 1/5 (d) 4/5

52. If in a Binomial distribution mean = 20 , S.D.= 4 then p is equal to

(a) 2/5 (b) 3/5 (c) 1/5 (d) 4/5

53. If is a Binomial distribution mean = 20 , S.D.= 4 then n is equal to

(a) 80 (b) 100 (c) 90 (d) none

54. Poisson distribution is a __________ probability distribution .

(a) discrete (b) continuous (c) both (d) none

55. Number of radio-active atoms decaying in a given interval of time is an example of

(a) Binomial distribution (b) Normal distribution
(c) Poisson distribution (d) None

56. __________ distribution is sometimes known as the "distribution of rare events".

(a) Poisson (b) Normal (c) Binomial (d) none

57. The probability that x assumes a specified value in continuous probability distribution is

(a) 1 (b) 0 (c) –1 (d) none

58. In Normal distribution mean, median and mode are

(a) equal (b) not equal (c) zero (d) none

59. In Normal distribution the quartiles are equidistant from

(a) median (b) mode (c) mean (d) none

60. In Normal distribution as the distance from the __________ increases, the curve comes closer and closer to the horizontal axis.

(a) median (b) mean (c) mode (d) none

61. A discrete random variable x follows uniform distribution and takes only the values 6, 8, 11, 12, 17.

The probability of P(x = 8) is

(a) 1/5 (b) 3/5 (c) 2/8 (d) 3/8

62. A discrete random variable x follows uniform distribution and takes the values 6, 9, 10, 11, 13.

The probability of P(x = 12) is

(a) 1/5 (b) 3/5 (c) 4/5 (d) 0

63. A discrete random variable x follows uniform distribution and takes the values 6, 8, 11, 12, 17

The probability of $P(x \leq 12)$ is

(a) 3/5 (b) 4/5 (c) 1/5 (d) none

64. A discrete random variable x follows uniform distribution and takes the values 6, 8, 10, 12, 18

The probability of $P(x < 12)$ is

(a) 1/5 (b) 4/5 (c) 3/5 (d) none

65. A discrete random variable x follows uniform distribution and takes the values 5, 7, 12, 15, 18

The probability of P(x > 10) is

(a) 3/5 (b) 2/5 (c) 4/5 (d) none

66. The probability density function of a continuous random variable is defined as follows :

f(x) = c when $-1 \le x \le 1 = 0$, otherwise the value of c is

(a) 1 (b) –1 (c) 1/2 (d) 0

67. A continuous random variable x has the probability density fn.f(x) = ½ –ax , $0 \le x \le 4$ When 'a' is a constant. The value of ' a' is

(a) 7/8 (b) 1/8 (c) 3/16 (d) none

68. A continuous random variable x follows uniform distribution with probability density function
f(x) = ½, ($4 \le x \le 6$). Then P($4 \le x \le 5$)

(a) 0.1 (b) 0.5 (c) 0 (d) none

69. An unbiased die is tossed 500 times.The mean of the number of 'Sixes' in these 500 tosses is

(a) 50/6 (b) 500/6 (c) 5/6 (d) none

70. An unbiased die is tossed 500 times. The Standard deviation of the number of 'sixes' in these 500 tossed is

(a) 50/6 (b) 500/6 (c) 5/6 (d) none

71. A random variable x follows Binomial distribution with mean 2 and variance 1.2. then the value of n is

(a) 8 (b) 2 (c) 5 (d) none

72. A random variable x follows Binomial distribution with mean 2 and variance 1.6 then the value of p is

(a) 1/5 (b) 4/5 (c) 3/5 (d) none

73. "The mean of a Binomial distribution is 5 and standard deviation is 3".

(a) True (b) false (c) both (d) none

74. The expected value of a constant k is the constant

(a) k (b) k–1 (c) k+1 (d) none

75. The probability distribution whose frequency function f(x)= 1/n($x = x_1, x_2, \ldots, x_n$) is known as

(a) Binomial distribution (b) Poisson distribution
(c) Uniform distribution (d) Normal distribution

76. Theoretical distribution is a

(a) Random distribution (b) Standard distribution
(c) Probability distribution (d) None

77. Probability function is known as

(a) frequency function (b) continuous function
(c) discrete function (d) none

78. The number of points obtained in a single throw of an unbiased die follows :

(a) Binomial distribution (b) Poisson distribution
(c) Uniform distribution (d) None

79. The Number of points in a single throw of an unbiased die has frequency function

(a) f(x)=1/4 (b) f(x)= 1/5 (c) f(x) = 1/6 (d) none

80. In uniform distribution random variable x assumes n values with

(a) equal probability (b) unequal probability (c) zero (d) none

81. In a discrete random variable x follows uniform distribution and assumes only the values 8 , 9, 11, 15, 18, 20. Then P(x = 9) is

(a) 2/6 (b) 1/7 (c) 1/5 (d) 1/6

82. In a discrete random variable x follows uniform distribution and assumes only the values 8 , 9, 11, 15, 18, 20. Then P(x = 12) is

(a) 1/6 (b) 0 (c) 1/7 (d) none

83. In a discrete random variable x follows uniform distribution and assumes only the values 8, 9, 11, 15, 18, 20. Then $P(x < 15)$ is

(a) 1/2 (b) 2/3 (c) 1 (d) none

84. In a discrete random variable x follows uniform distribution and assumes only the values 8 , 9, 11, 15, 18, 20. Then $P(x \leq 15)$ is

(a) 2/3 (b) 1/3 (c) 1 (d) none

85. In a discrete random variable x follows uniform distribution and assumes only the values 8, 9, 11, 15, 18, 20. Then $P(x > 15)$ is

(a) 2/3 (b) 1/3 (c) 1 (d) none

86. In a discrete random variable x follows uniform distribution and assumes only the values 8, 9, 11, 15, 18, 20. Then $P(|x - 14| < 5)$ is

(a) 1/3 (b) 2/3 (c) 1/2 (d) 1

87. When f(x)= 1/n then mean is

(a) (n–1)/2 (b) (n+1)/2 (c) n/2 (d) none

88. In continuous probability distribution $P(x \leq t)$ means

(a) Area under the probability curve to the left of the vertical line at t .

(b) Area under the probability curve to the right of the vertical line at t .

(c) both (d) none

89. In continuous probability distribution F(x) is called.

(a) frequency distribution function (b) cumulative distribution function
(c) probability density function (d) none

90. The probability density function of a continuous random variable is $y = k(x-1)$, ($1 \leq x \leq 2$) then the value of the constant k is

(a) –1 (b) 1 (c) 2 (d) 0

ANSWERS

1.	(c)	**2.**	(a)	**3.**	(b)	**4.**	(b)	**5.**	(a)
6.	(b)	**7.**	(a)	**8.**	(b)	**9.**	(a)	**10.**	(b)
11.	(a)	**12.**	(a)	**13.**	(b)	**14.**	(b)	**15.**	(b)
16.	(c)	**17.**	(a)	**18.**	(c)	**19.**	(a)	**20.**	(c)
21.	(a)	**22.**	(b)	**23.**	(b)	**24.**	(c)	**25.**	(b)
26.	(c)	**27.**	(a)	**28.**	(c)	**29.**	(b)	**30.**	(c)
31.	(b)	**32.**	(b)	**33.**	(a)	**34.**	(a)	**35.**	(c)
36.	(b)	**37.**	(d)	**38.**	(a)	**39.**	(d)	**40.**	(a)
41.	(a)	**42.**	(a)	**43.**	(b)	**44.**	(a)	**45.**	(c)
46.	(b)	**47.**	(a)	**48.**	(a)	**49.**	(b)	**50.**	(b)
51.	(d)	**52.**	(c)	**53.**	(b)	**54.**	(a)	**55.**	(c)
56.	(a)	**57.**	(b)	**58.**	(a)	**59.**	(c)	**60.**	(b)
61.	(a)	**62.**	(d)	**63.**	(b)	**64.**	(c)	**65.**	(a)
66.	(c)	**67.**	(b)	**68.**	(b)	**69.**	(b)	**70.**	(a)
71.	(c)	**72.**	(a)	**73.**	(b)	**74.**	(a)	**75.**	(c)
76.	(c)	**77.**	(a)	**78.**	(c)	**79.**	(c)	**80.**	(a)
81.	(d)	**82.**	(b)	**83.**	(a)	**84.**	(a)	**85.**	(b)
86.	(c)	**87.**	(b)	**88.**	(a)	**89.**	(b)	**90.**	(c)

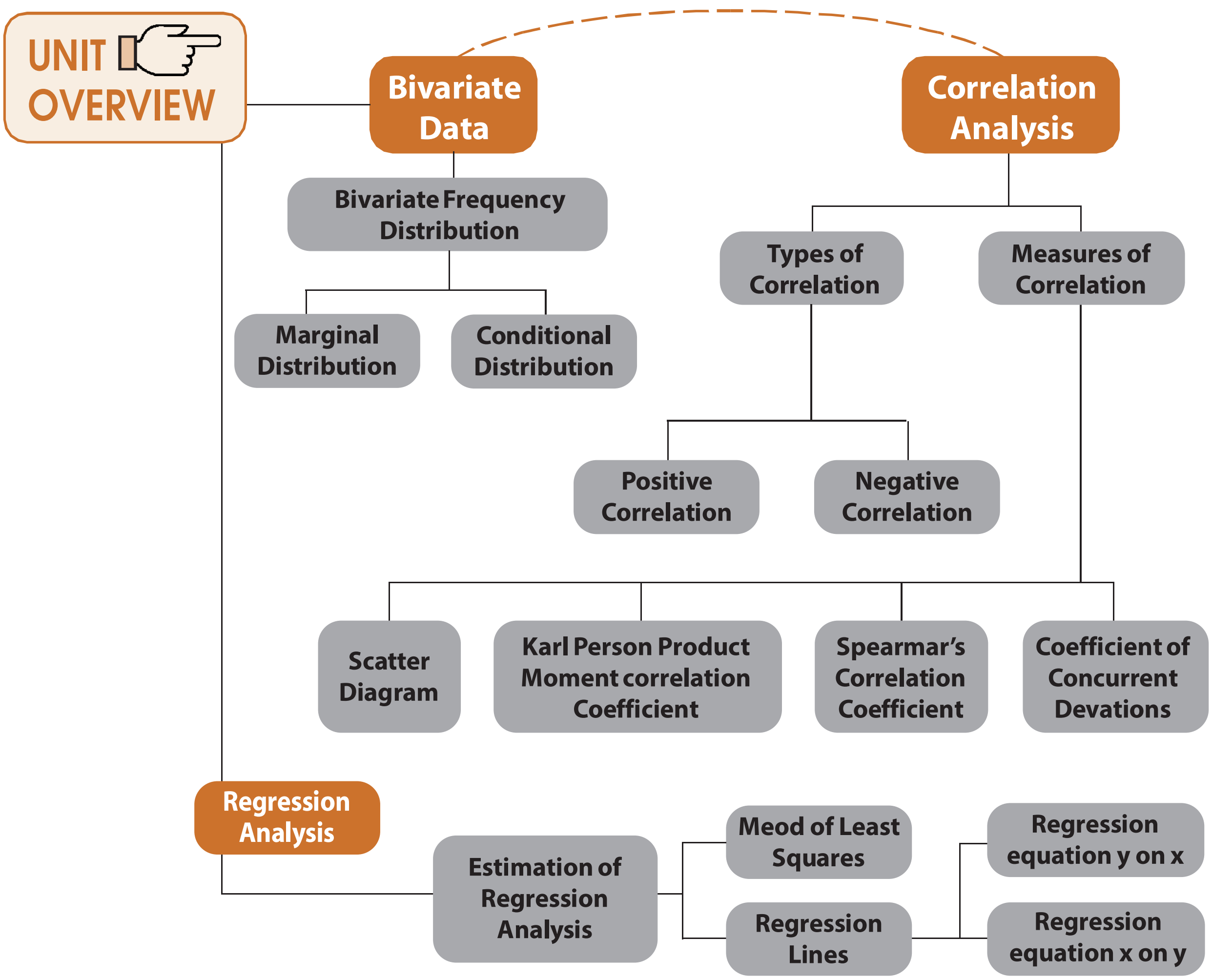
CHAPTER 18

CORRELATION AND REGRESSION

LEARNING OBJECTIVES

After reading this chapter, students will be able to understand:

- The meaning of bivariate data and techniques of preparation of bivariate distribution;
- The concept of correlation between two variables and quantitative measurement of correlation including the interpretation of positive, negative and zero correlation;
- Concept of regression and its application in estimation of a variable from known set of data.



18.1 INTRODUCTION

In the previous chapter, we discussed many a statistical measure relating to Univariate distribution i.e. distribution of one variable like height, weight, mark, profit, wage and so on. However, there are situations that demand study of more than one variable simultaneously. A businessman may be keen to know what amount of investment would yield a desired level of profit or a student may want to know whether performing better in the selection test would enhance his or her chance of doing well in the final examination. With a view to answering this series of questions, we need to study more than one variable at the same time. Correlation Analysis and Regression Analysis are the two analyses that are made from a multivariate distribution i.e. a distribution of more than one variable. In particular when there are two variables, say x and y, we study bivariate distribution. We restrict our discussion to bivariate distribution only.

Correlation analysis, it may be noted, helps us to find an association or the lack of it between the two variables x and y. Thus if x and y stand for profit and investment of a firm or the marks in Statistics and Mathematics for a group of students, then we may be interested to know whether x and y are associated or independent of each other. The extent or amount of correlation between x and y is provided by different measures of Correlation namely Product Moment Correlation Coefficient or Rank Correlation Coefficient or Coefficient of Concurrent Deviations. In Correlation analysis, we must be careful about a cause and effect relation between the variables under consideration because there may be situations where x and y are related due to the influence of a third variable although no causal relationship exists between the two variables.

Regression analysis, on the other hand, is concerned with predicting the value of the dependent variable corresponding to a known value of the independent variable on the assumption of a mathematical relationship between the two variables and also an average relationship between them.

18.2 BIVARIATE DATA

When data are collected on two variables simultaneously, they are known as bivariate data and the corresponding frequency distribution, derived from it, is known as Bivariate Frequency Distribution. If x and y denote marks in Maths and Stats for a group of 30 students, then the corresponding bivariate data would be (x_i, y_i) for i = 1, 2, …. 30 where (x_1, y_1) denotes the marks in Mathematics and Statistics for the student with serial number or Roll Number 1, (x_2, y_2), that for the student with Roll Number 2 and so on and lastly (x_{30}, y_{30}) denotes the pair of marks for the student bearing Roll Number 30.

As in the case of a Univariate Distribution, we need to construct the frequency distribution for bivariate data. Such a distribution takes into account the classification in respect of both the variables simultaneously. Usually, we make horizontal classification in respect of x and vertical classification in respect of the other variable y. Such a distribution is known as Bivariate Frequency Distribution or Joint Frequency Distribution or Two way classification of the two variables x and y.

ILLUSTRATIONS:

Example 18.1: Prepare a Bivariate Frequency table for the following data relating to the marks in Statistics (x) and Mathematics (y):

(15, 13),	(1, 3),	(2, 6),	(8, 3),	(15, 10),	(3, 9),	(13, 19),
(10, 11),	(6, 4),	(18, 14),	(10, 19),	(12, 8),	(11, 14),	(13, 16),
(17, 15),	(18, 18),	(11, 7),	(10, 14),	(14, 16),	(16, 15),	(7, 11),
(5, 1),	(11, 15),	(9, 4),	(10, 15),	(13, 12)	(14, 17),	(10, 11),
(6, 9),	(13, 17),	(16, 15),	(6, 4),	(4, 8),	(8, 11),	(9, 12),
(14, 11),	(16, 15),	(9, 10),	(4, 6),	(5, 7),	(3, 11),	(4, 16),
(5, 8),	(6, 9),	(7, 12),	(15, 6),	(18, 11),	(18, 19),	(17, 16)
(10, 14)						

Take mutually exclusive classification for both the variables, the first class interval being 0-4 for both.

Solution:

From the given data, we find that

Range for x = 19–1 = 18

Range for y = 19–1 = 18

We take the class intervals 0-4, 4-8, 8-12, 12-16, 16-20 for both the variables. Since the first pair of marks is (15, 13) and 15 belongs to the fourth class interval (12-16) for x and 13 belongs to the fourth class interval for y, we put a stroke in the (4, 4)-th cell. We carry on giving tally marks till the list is exhausted.

Table 18.1

Bivariate Frequency Distribution of Marks in Statistics and Mathematics.

		MARKS IN MATHS					
Y / X		0-4	4-8	8-12	12-16	16-20	Total
MARKS IN STATS	0–4	I (1)	I (1)	II (2)			4
	4–8	I (1)	IIII (4)	~~IIII~~ (5)	I (1)	I (1)	12
	8–12	I (1)	II (2)	IIII (4)	~~IIII~~ I (6)	I (1)	14
	12–16		I (1)	III (3)	II (2)	~~IIII~~ (5)	11
	16–20			I (1)	~~IIII~~ (5)	III (3)	9
	Total	3	8	15	14	10	50

We note, from the above table, that some of the cell frequencies (f_{ij}) are zero. Starting from the above Bivariate Frequency Distribution, we can obtain two types of univariate distributions which are known as:

(a) Marginal distribution.

(b) Conditional distribution.

If we consider the distribution of Statistics marks along with the marginal totals presented in the last column of Table 12-1, we get the marginal distribution of marks in Statistics. Similarly, we can obtain one more marginal distribution of Mathematics marks. The following table shows the marginal distribution of marks of Statistics.

Table 18.2

Marginal Distribution of Marks in Statistics

Marks	No. of Students
0-4	4
4-8	12
8-12	14
12-16	11
16-20	9
Total	50

We can find the mean and standard deviation of marks in Statistics from Table 18.2. They would be known as marginal mean and marginal SD of Statistics marks. Similarly, we can obtain the marginal mean and marginal SD of Mathematics marks. Any other statistical measure in respect of x or y can be computed in a similar manner.

If we want to study the distribution of Statistics Marks for a particular group of students, say for those students who got marks between 8 to 12 in Mathematics, we come across another univariate distribution known as conditional distribution.

Table 18.3

Conditional Distribution of Marks in Statistics for Students having Mathematics Marks between 8 to 12

Marks	No. of Students
0-4	2
4-8	5
8-12	4
12-16	3
16-20	1
Total	15

We may obtain the mean and SD from the above table. They would be known as conditional mean and conditional SD of marks of Statistics. The same result holds for marks in Mathematics. In particular, if there are m classifications for x and n classifications for y, then there would be altogether (m + n) conditional distribution.

18.3 CORRELATION ANALYSIS

While studying two variables at the same time, if it is found that the change in one variable is reciprocated by a corresponding change in the other variable either directly or inversely, then the two variables are known to be associated or correlated. Otherwise, the two variables are known to be dissociated or uncorrelated or independent. There are two types of correlation.

(i) Positive correlation

(ii) Negative correlation

If two variables move in the same direction i.e. an increase (or decrease) on the part of one variable introduces an increase (or decrease) on the part of the other variable, then the two variables are known to be positively correlated. As for example, height and weight yield and rainfall, profit and investment etc. are positively correlated.

On the other hand, if the two variables move in the opposite directions i.e. an increase (or a decrease) on the part of one variable results a decrease (or an increase) on the part of the other variable, then the two variables are known to have a negative correlation. The price and demand of an item, the profits of Insurance Company and the number of claims it has to meet etc. are examples of variables having a negative correlation.

The two variables are known to be uncorrelated if the movement on the part of one variable does not produce any movement of the other variable in a particular direction. As for example, Shoe-size and intelligence are uncorrelated.

18.4 MEASURES OF CORRELATION

We consider the following measures of correlation:

(a) Scatter diagram

(b) Karl Pearson's Product moment correlation coefficient

(c) Spearman's rank correlation co-efficient

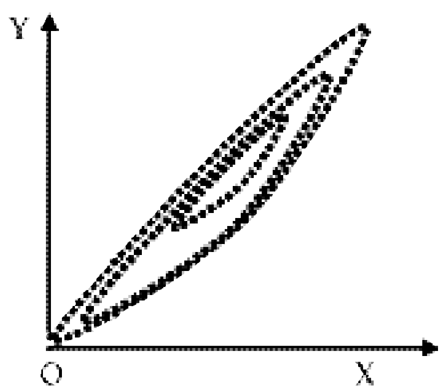
(d) Co-efficient of concurrent deviations

(a) SCATTER DIAGRAM

This is a simple diagrammatic method to establish correlation between a pair of variables. Unlike product moment correlation co-efficient, which can measure correlation only when the variables are having a linear relationship, scatter diagram can be applied for any type of correlation – linear as well as non-linear i.e. curvilinear. Scatter diagram can distinguish between different types of correlation although it fails to measure the extent of relationship between the variables.

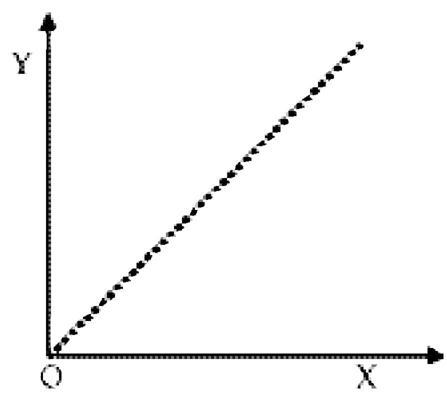
Each data point, which in this case a pair of values (x_i, y_i) is represented by a point in the rectangular axes of cordinates. The totality of all the plotted points forms the scatter diagram. The pattern of the plotted points reveals the nature of correlation. In case of a positive correlation, the plotted points lie from lower left corner to upper right corner, in case of a negative correlation the plotted points concentrate from upper left to lower right and in case of zero correlation, the plotted points would be equally distributed without depicting any particular pattern. The following figures show different types of correlation and the one to one correspondence between scatter diagram and product moment correlation coefficient.

FIGURE 18.1
Showing Positive Correlation



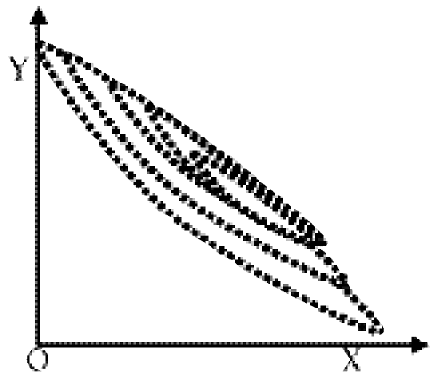
(0 < r <1)

FIGURE 18.2
Showing Perfect Correlation



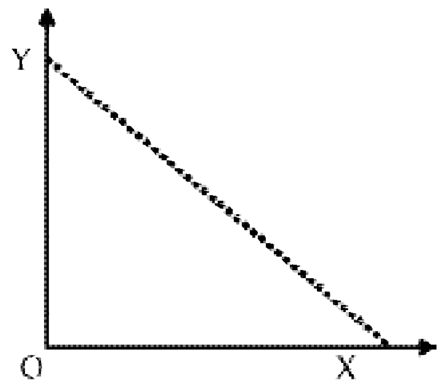
(r = 1)

FIGURE 18.3
Showing Negative Correlation



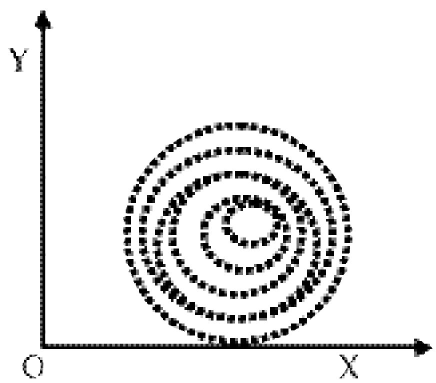
(–1 < r <0)

FIGURE 18.4
Showing Perfect Negative Correlation



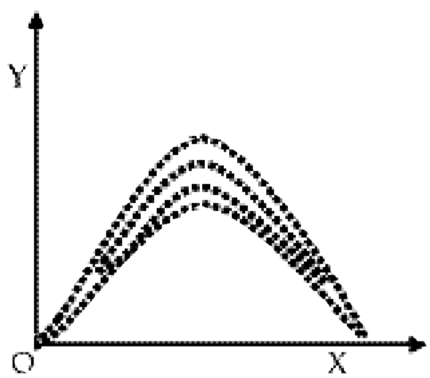
(r = –1)

FIGURE 18.5
Showing No Correlation



(r = 0)

FIGURE 18.6
Showing Curvilinear Correlation



(r = 0)

(b) KARL PEARSON'S PRODUCT MOMENT CORRELATION COEFFICIENT

This is by for the best method for finding correlation between two variables provided the relationship between the two variables is linear. Pearson's correlation coefficient may be defined as the ratio of covariance between the two variables to the product of the standard deviations of the two variables. If the two variables are denoted by x and y and if the corresponding bivariate data are (x_i, y_i) for i = 1, 2, 3, ….., n, then the coefficient of correlation between x and y, due to Karl Pearson, in given by :

$$r = r_{xy} = \frac{Cov(x,y)}{S_x \times S_y} \quad \text{.........(18.1)}$$

where

$$cov(x,y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x}\bar{y} \quad \text{.........(18.2)}$$

$$S_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \quad \text{.........(18.3)}$$

and $$S_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n}} = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2} \quad \text{.........(18.4)}$$

A single formula for computing correlation coefficient is given by

$$r = \frac{n\sum x_i y_i - \sum x_i \times \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \quad \text{.........(18.5)}$$

In case of a bivariate frequency distribution, we have

$$Cov(x,y) = \frac{\sum_{i,j} x_i y_i f_{ij}}{N} - \bar{x} \times \bar{y} \quad \text{.........(18.6)}$$

$$S_x = \sqrt{\frac{\sum_i f_{io} x_i^2}{N} - \bar{x}^2} \quad \text{.........(18.7)}$$

and $$S_y = \sqrt{\frac{\sum_j f_{oj} y_j^2}{N} - \bar{y}^2} \quad \text{.........(18.8)}$$

where x_i = Mid-value of the i^{th} class interval of x.

y_j = Mid-value of the j^{th} class interval of y

f_{io} = Marginal frequency of x

f_{oj} = Marginal frequency of y

f_{ij} = frequency of the $(i, j)^{th}$ cell

$$N = \sum_{i,j} f_{ij} = \sum_i f_{io} = \sum_j f_{oj} = \text{Total frequency} \quad \ldots\ldots\ldots\ldots\ldots (18.9)$$

PROPERTIES OF CORRELATION COEFFICIENT

(i) **The Coefficient of Correlation is a unit-free measure.**

This means that if x denotes height of a group of students expressed in cm and y denotes their weight expressed in kg, then the correlation coefficient between height and weight would be free from any unit.

(ii) **The coefficient of correlation remains invariant under a change of origin and/or scale of the variables under consideration depending on the sign of scale factors.**

This property states that if the original pair of variables x and y is changed to a new pair of variables u and v by effecting a change of origin and scale for both x and y i.e.

$$u = \frac{x-a}{b} \text{ and } v = \frac{y-c}{d}$$

where a and c are the origins of x and y and b and d are the respective scales and then we have

$$r_{xy} = \frac{bd}{|b||d|} r_{uv} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18.10)$$

r_{xy} and r_{uv} being the coefficient of correlation between x and y and u and v respectively, (18.10) established, numerically, the two correlation coefficients remain equal and they would have opposite signs only when b and d, the two scales, differ in sign.

(iii) **The coefficient of correlation always lies between –1 and 1, including both the limiting values i.e.**

$$-1 \le r \le 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18.11)$$

Example 18.2: Compute the correlation coefficient between x and y from the following data n = 10, $\sum xy = 220$, $\sum x^2 = 200$, $\sum y^2 = 262$

$\sum x = 40$ and $\sum y = 50$

Solution:

From the given data, we have by applying (18.5),

$$r = \frac{n\sum xy - \sum x \times \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \times \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{10 \times 220 - 40 \times 50}{\sqrt{10 \times 200 - (40)^2} \times \sqrt{10 \times 262 - (50)^2}}$$

$$= \frac{2200 - 2000}{\sqrt{2000 - 1600} \times \sqrt{2620 - 2500}}$$

$$= \frac{200}{20 \times 10.9545}$$

$$= 0.91$$

Thus there is a good amount of positive correlation between the two variables x and y.

Alternately

As given, $\bar{x} = \frac{\sum x}{n} = \frac{40}{10} = 4$

$\bar{y} = \frac{\sum y}{n} = \frac{50}{10} = 5$

Cov (x, y) $= \frac{\sum xy}{n} - \bar{x}.\bar{y}$

$= \frac{220}{10} - 4.5 = 2$

$S_x = \sqrt{\frac{\sum x2}{n} - (\bar{x})^2}$

$= \sqrt{\frac{200}{10} - 4^2} = 2$

$$S_y = \sqrt{\frac{\Sigma y_i^2}{n} - \bar{y}^2}$$

$$= \sqrt{\frac{262}{10} - 5^2}$$

$$= \sqrt{26.20 - 25} = 1.0954$$

Thus applying formula (18.1), we get

$$r = \frac{\text{cov}(x, y)}{S_x . S_y}$$

$$= \frac{2}{2 \times 1.0954} = 0.91$$

As before, we draw the same conclusion.

Example 18.3: Find product moment correlation coefficient from the following information:

x :	2	3	5	5	6	8
y :	9	8	8	6	5	3

Solution:

In order to find the covariance and the two standard deviation, we prepare the following table:

Table 18.3

Computation of Correlation Coefficient

x_i (1)	y_i (2)	x_iy_i (3)= (1) x (2)	x_i^2 (4)= $(1)^2$	y_i^2 (5)= $(2)^2$
2	9	18	4	81
3	8	24	9	64
5	8	40	25	64
5	6	30	25	36
6	5	30	36	25
8	3	24	64	9
29	39	166	163	279

We have

$$\bar{x} = \frac{29}{6} = 4.8333 \quad \bar{y} = \frac{39}{6} = 6.50$$

$$\text{cov}(x, y) = \frac{\Sigma x_i y_i}{n} - \bar{x}\,\bar{y}$$

$$= 166/6 - 4.8333 \times 6.50 = -3.7498$$

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

$$= \sqrt{\frac{163}{6} - (4.8333)^2}$$

$$= \sqrt{27.1667 - 23.3608} = 1.95$$

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

$$= \sqrt{\frac{279}{6} - (6.50)^2}$$

$$= \sqrt{46.50 - 42.25} = 2.0616$$

Thus the correlation coefficient between x and y in given by

$$r = \frac{\text{cov}(x, y)}{S_x \times s_y}$$

$$= \frac{-3.7498}{1.9509 \times 2.0616}$$

$$= -0.93$$

We find a high degree of negative correlation between x and y. Also, we could have applied formula (18.5) as we have done for the first problem of computing correlation coefficient.

Sometimes, a change of origin reduces the computational labor to a great extent. This we are going to do in the next problem.

Example 18.4: The following data relate to the test scores obtained by eight salesmen in an aptitude test and their daily sales in thousands of rupees:

Salesman :	1	2	3	4	5	6	7	8
scores :	60	55	62	56	62	64	70	54
Sales :	31	28	26	24	30	35	28	24

Solution:

Let the scores and sales be denoted by x and y respectively. We take a, origin of x as the average of the two extreme values i.e. 54 and 70. Hence a = 62 similarly, the origin of y is taken

as $b = \frac{24+35}{2} \cong 30$

Table 18.4

Computation of Correlation Coefficient Between Test Scores and Sales.

Scores (x_i) (1)	Sales in ₹ 1000 (y_i) (2)	$u_i = x_i - 62$ (3)	$v_i = y_i - 30$ (4)	u_iv_i (5)=(3)x(4)	u_i^2 (6)=(3)^2	v_i^2 (7)=(4)^2
60	31	–2	1	–2	4	1
55	28	–7	–2	14	49	4
62	26	0	–4	0	0	16
56	24	–6	–6	36	36	36
62	30	0	0	0	0	0
64	35	2	5	10	4	25
70	28	8	–2	–16	64	4
54	24	–8	–6	48	64	36
Total	—	–13	–14	90	221	122

Since correlation coefficient remains unchanged due to change of origin, we have

$$r = r_{xy} = r_{uv} = \frac{n\sum u_iv_i - \sum u_i \times \sum v_i}{\sqrt{n\sum u_i^2 - \left(\sum u_i\right)^2} \times \sqrt{n\sum v_i^2 - \left(\sum v_i\right)^2}}$$

$$= \frac{8 \times 90 - (-13) \times (-14)}{\sqrt{8 \times 221 - (-13)^2} \times \sqrt{8 \times 122 - (-14)^2}}$$

$$= \frac{538}{\sqrt{1768 - 169} \times \sqrt{976 - 196}}$$

$$= 0.48$$

In some cases, there may be some confusion about selecting the pair of variables for which correlation is wanted. This is explained in the following problem.

Example 18.5: Examine whether there is any correlation between age and blindness on the basis of the following data:

Age in years :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Persons (in thousands) :	90	120	140	100	80	60	40	20
No. of blind Persons :	10	15	18	20	15	12	10	06

Solution:

Let us denote the mid-value of age in years as x and the number of blind persons per lakh as y. Then as before, we compute correlation coefficient between x and y.

Table 18.5

Computation of correlation between age and blindness

Age in years (1)	Mid-value x (2)	No. of Persons ('000) P (3)	No. of blind B (4)	No. of blind per lakh y=B/P × 1 lakh (5)	xy (2)×(5) (6)	x^2 $(2)^2$ (7)	y^2 $(5)^2$ (8)
0-10	5	90	10	11	55	25	121
10-20	15	120	15	12	180	225	144
20-30	25	140	18	13	325	625	169
30-40	35	100	20	20	700	1225	400
40-50	45	80	15	19	855	2025	361
50-60	55	60	12	20	1100	3025	400
60-70	65	40	10	25	1625	4225	625
70-80	75	20	6	30	2250	5625	900
Total	320	—	—	150	7090	17000	3120

The correlation coefficient between age and blindness is given by

$$r = \frac{n\sum xy - \sum x.\sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \times \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{8.7090 - 320.150}{\sqrt{8.17000 - (320)^2} \times \sqrt{8.3120 - (150)^2}}$$

$$= \frac{8720}{183.3030.49.5984}$$

$$= 0.96$$

which exhibits a very high degree of positive correlation between age and blindness.

Example 18.6: Coefficient of correlation between x and y for 20 items is 0.4. The AM's and SD's of x and y are known to be 12 and 15 and 3 and 4 respectively. Later on, it was found that the pair (20, 15) was wrongly taken as (15, 20). Find the correct value of the correlation coefficient.

Solution:

We are given that n = 20 and the original r = 0.4, $\bar{x}$ = 12, $\bar{y}$ = 15, S_x = 3 and S_y = 4

$$r = \frac{\text{cov}(x, y)}{S_x \times S_y} \qquad = 0.4 = \frac{\text{cov}(x, y)}{3 \times 4}$$

$$= \text{Cov}(x, y) = 4.8$$

$$= \frac{\sum xy}{n} - \bar{x}\,\bar{y} = 4.8$$

$$= \frac{\sum xy}{20} - 12 \times 15 = 4.8$$

$$= \sum xy = 3696$$

Hence, corrected $\sum xy$= 3696 – 20 × 15 + 15 × 20 = 3696

Also, $S_x^2 = 9$

$= (\sum x^2 / 20) - 12^2 = 9$

$\sum x^2 = 3060$

Similarly, $S_y^2 = 16$

$$S_y^2 = \frac{\sum y^2}{20} - 15^2 = 16$$

$\sum y^2 = 4820$

Thus corrected $\sum x = n\bar{x}$ – wrong x value + correct x value.

$$= 20 \times 12 - 15 + 20$$

$$= 245$$

Similarly corrected $\sum y = 20 \times 15 - 20 + 15 = 295$

Corrected $\sum x^2 = 3060 - 15^2 + 20^2 = 3235$

Corrected $\sum y^2 = 4820 - 20^2 + 15^2 = 4645$

Thus corrected value of the correlation coefficient by applying formula (18.5)

$$= \frac{20 \times 3696 - 245 \times 295}{\sqrt{20 \times 3235 - (245)^2} \times \sqrt{20 \times 4645 - (295)^2}}$$

$$= \frac{73920 - 72275}{68.3740 \times 76.6480}$$

$$= 0.31$$

Example 18.7: Compute the coefficient of correlation between marks in Statistics and Mathematics for the bivariate frequency distribution shown in Table 18.6

Solution:

For the sake of computational advantage, we effect a change of origin and scale for both the variable x and y.

$$\text{Define } u_i = \frac{x_i - a}{b} = \frac{x_i - 10}{4}$$

$$\text{And } v_j = \frac{y_i - c}{d} = \frac{y_i - 10}{4}$$

Where x_i and y_j denote respectively the mid-values of the x-class interval and y-class interval respectively. The following table shows the necessary calculation on the right top corner of each cell, the product of the cell frequency, corresponding u value and the respective v value has been shown. They add up in a particular row or column to provide the value of $f_{ij}u_iv_j$ for that particular row or column.

Table 18.6

Computation of Correlation Coefficient Between Marks of Mathematics and Statistics

Class Interval	Mid-value	Class Interval / Mid-value →	0-4	4-8	8-12	12-16	16-20				
		Mid-value	2	6	10	14	18				
Class Interval	Mid -value	v_j / u_i	–2	–1	0	1	2	f_{io}	$f_{io}u_i$	$f_{io}u_i^2$	$f_{ij}u_iv_j$
0-4	2	–2	1 (4)	1 (2)	2 (0)			4	–8	16	6
4-8	6	–1	2 (4)	4 (4)	5 (0)	1 (–1)	1 (–2)	13	–13	13	5
8-12	10	0		2 (0)	4 (0)	6 (0)	1 (0)	13	0	0	0
12-16	14	1		1 (–1)	3 (0)	2 (2)	5 (10)	11	11	11	11
16-20	18	2			1 (0)	5 (10)	3 (12)	9	18	36	22
		f_{oj}	3	8	15	14	10	50	5	76	44
		$f_{oj}v_j$	–6	–8	0	14	20	20			
		$f_{oj}v_j^2$	12	8	0	14	40	74			
		$f_{ij}u_iv_j$	8	5	0	11	20	44		CHECK	

A single formula for computing correlation coefficient from bivariate frequency distribution is given by

$$r = \frac{N\sum_{i,j} f_{ij}u_iv_j - \sum f_{io}u_i \times \sum f_{oj}v_j}{\sqrt{N\sum f_{io}u_i^2 - \left(\sum f_{io}u_i\right)^2 \times \sum f_{oj}v_j^2 - \left(\sum f_{oj}v_j\right)^2}} \quad \text{.............................(18.10)}$$

$$= \frac{50 \times 44 - 8 \times 20}{\sqrt{50 \times 76 - 8^2}\sqrt{50 \times 74 - 20^2}}$$

$$= \frac{2040}{61.1228 \times 57.4456}$$

$$= 0.58$$

The value of r shown a good amount of positive correlation between the marks in Statistics and Mathematics on the basis of the given data.

Example 18.8: Given that the correlation coefficient between x and y is 0.8, write down the correlation coefficient between u and v where

(i) 2u + 3x + 4 = 0 and 4v + 16y + 11 = 0

(ii) 2u – 3x + 4 = 0 and 4v + 16y + 11 = 0

(iii) 2u – 3x + 4 = 0 and 4v – 16y + 11 = 0

(iv) 2u + 3x + 4 = 0 and 4v – 16y + 11 = 0

Solution:

Using (18.10), we find that

$$r_{xy} = \frac{bd}{|b||d|} r_{uv}$$

i.e. $r_{xy} = r_{uv}$ if b and d are of same sign and $r_{uv} = -r_{xy}$ when b and d are of opposite signs, b and d being the scales of x and y respectively. In (i), u = (–2) + (-3/2) x and v = (–11/4) + (–4)y.

Since b = –3/2 and d = –4 are of same sign, the correlation coefficient between u and v would be the same as that between x and y i.e. $r_{xy} = 0.8 = r_{uv}$

In (ii), u = (–2) + (3/2)x and v = (–11/4) + (–4)y Hence b = 3/2 and d = –4 are of opposite signs and we have $r_{uv} = -r_{xy} = -0.8$

Proceeding in a similar manner, we have r_{uv} = 0.8 and – 0.8 in (iii) and (iv).

(c) SPEARMAN'S RANK CORRELATION COEFFICIENT

When we need finding correlation between two qualitative characteristics, say, beauty and intelligence, we take recourse to using rank correlation coefficient. Rank correlation can also be applied to find the level of agreement (or disagreement) between two judges so far as assessing a qualitative characteristic is concerned. As compared to product moment correlation coefficient, rank correlation coefficient is easier to compute, it can also be advocated to get a first hand impression about the correlation between a pair of variables.

Spearman's rank correlation coefficient is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad \text{..} (18.11)$$

where r_R denotes rank correlation coefficient and it lies between –1 and 1 inclusive of these two values.

$d_i = x_i - y_i$ represents the difference in ranks for the i-th individual and n denotes the number of individuals.

In case u individuals receive the same rank, we describe it as a tied rank of length u. In case of a tied rank, formula (18.11) is changed to

$$r_R = 1 - \frac{6\left[\sum_i d_i^2 + \sum_j \frac{(tj^3 - t_j)}{12}\right]}{n(n^2-1)} \quad \text{.. (18.12)}$$

In this formula, t_j represents the j^{th} tie length and the summation $\sum_j (t_j^3 - t_j)$ extends over the lengths of all the ties for both the series.

Example 18.9: compute the coefficient of rank correlation between sales and advertisement expressed in thousands of rupees from the following data:

Sales :	90	85	68	75	82	80	95	70
Advertisement :	7	6	2	3	4	5	8	1

Solution:

Let the rank given to sales be denoted by x and rank of advertisement be denoted by y. We note that since the highest sales as given in the data, is 95, it is to be given rank 1, the second highest sales 90 is to be given rank 2 and finally rank 8 goes to the lowest sales, namely 68. We have given rank to the other variable advertisement in a similar manner. Since there are no ties, we apply formula (16.11).

Table 18.7

Computation of Rank correlation between Sales and Advertisement.

Sales (x_i)	Advertisement (y_i)	Rank for Sales (x_i)	Rank for Advertisement (y_i)	$d_i = x_i - y_i$	d_i^2
90	7	2	2	0	0
85	6	3	3	0	0
68	2	8	7	1	1
75	3	6	6	0	0
82	4	4	5	–1	1
80	5	5	4	1	1
95	8	1	1	0	0
70	1	7	8	–1	1
Total	—	—	—	0	4

Since n = 8 and $\sum d_i^2 = 4$, applying formula (18.11), we get.

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 4}{8(8^2-1)}$$

$= 1–0.0476$

$= 0.95$

The high positive value of the rank correlation coefficient indicates that there is a very good amount of agreement between sales and advertisement.

Example 18.10: Compute rank correlation from the following data relating to ranks given by two judges in a contest:

Serial No. of Candidate :	1	2	3	4	5	6	7	8	9	10
Rank by Judge A :	10	5	6	1	2	3	4	7	9	8
Rank by Judge B :	5	6	9	2	8	7	3	4	10	1

Solution:

We directly apply formula (18.11) as ranks are already given.

Table 18.8

Computation of Rank Correlation Coefficient between the ranks given by 2 Judges

Serial No.	Rank by A (x_i)	Rank by B (y_i)	$d_i = x_i - y_i$	d_i^2
1	10	5	5	25
2	5	6	–1	1
3	6	9	–3	9
4	1	2	–1	1
5	2	8	–6	36
6	3	7	–4	16
7	4	3	1	1
8	7	4	3	9
9	8	10	–2	4
10	9	1	8	64
Total	—	—	0	166

The rank correlation coefficient is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 166}{10(10^2-1)}$$

$$= -0.006$$

The very low value (almost 0) indicates that there is hardly any agreement between the ranks given by the two Judges in the contest.

Example 18.11: Compute the coefficient of rank correlation between Eco. marks and stats. Marks as given below:

Eco Marks :	80	56	50	48	50	62	60
Stats Marks :	90	75	75	65	65	50	65

Solution:

This is a case of tied ranks as more than one student share the same mark both for Economics and Statistics. For Eco. the student receiving 80 marks gets rank 1 one getting 62 marks receives rank 2, the student with 60 receives rank 3, student with 56 marks gets rank 4 and since there are two students, each getting 50 marks, each would be receiving a common rank, the average of the next two ranks 5 and 6 i.e. $\frac{5+6}{2}$ i.e. 5.50 and lastly the last rank..

7 goes to the student getting the lowest Eco marks. In a similar manner, we award ranks to the students with stats marks.

Table 18.9

Computation of Rank Correlation Between Eco Marks and Stats Marks with Tied Marks

Eco Mark (x_i)	Stats Mark (y_i)	Rank for Eco (x_i)	Rank for Stats (y_i)	$d_i = x_i - y_i$	d_i^2
80	90	1	1	0	0
56	75	4	2.50	1.50	2.25
50	75	5.50	2.50	3	9
48	65	7	5	2	4
50	65	5.50	5	0.50	0.25
62	50	2	7	–5	25
60	65	3	5	–2	4
Total	—	—	—	0	44.50

For Economics mark there is one tie of length 2 and for stats mark, there are two ties of lengths 2 and 3 respectively.

$$\text{Thus } \frac{\sum\left(t_j^3 - t_j\right)}{12} = \frac{\left(2^3 - 2\right) + \left(2^3 - 2\right) + \left(3^3 - 3\right)}{12} = 3$$

$$\text{Thus } r_R = 1 - \frac{6\left[\sum_i d_i^2 + \sum_j \frac{\left(tj^3 - t_j\right)}{12}\right]}{n\left(n^2 - 1\right)}$$

$$= 1 - \frac{6 \times (44.50 + 3)}{7(7^2 - 1)}$$

$$= 0.15$$

Example 18.12: For a group of 8 students, the sum of squares of differences in ranks for Mathematics and Statistics marks was found to be 50 what is the value of rank correlation coefficient?

Solution:

As given n = 8 and $\sum d_i^2 = 50$. Hence the rank correlation coefficient between marks in Mathematics and Statistics is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)}$$

$$= 1 - \frac{6 \times 50}{8(8^2 - 1)}$$

$$= 0.40$$

Example 18.13: For a number of towns, the coefficient of rank correlation between the people living below the poverty line and increase of population is 0.50. If the sum of squares of the differences in ranks awarded to these factors is 82.50, find the number of towns.

Solution:

As given $r_R = 0.50$, $\sum d_i^2 = 82.50$.

$$\text{Thus } r_R = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)}$$

$$0.50 = 1 - \frac{6 \times 82.50}{n\left(n^2 - 1\right)}$$

$$= n\,(n^2 - 1) = 990$$

$$= n\,(n^2 - 1) = 10(10^2 - 1)$$

$\therefore$ n = 10 as n must be a positive integer.

Example 18.14: While computing rank correlation coefficient between profits and investment for 10 years of a firm, the difference in rank for a year was taken as 7 instead of 5 by mistake and the value of rank correlation coefficient was computed as 0.80. What would be the correct value of rank correlation coefficient after rectifying the mistake?

Solution:

We are given that n = 10,

r_R = 0.80 and the wrong $d_i = 7$ should be replaced by 5.

$$r_R = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)}$$

$$0.80 = 1 - \frac{6\sum d_i^2}{10\left(10^2 - 1\right)}$$

$$\sum d_i^2 = 33$$

Corrected $\sum d_i^2 = 33 - 7^2 + 5^2 = 9$

Hence rectified value of rank correlation coefficient

$$= 1 - \frac{6 \times 9}{10 \times \left(10^2 - 1\right)}$$

$$= 0.95$$

(d) COEFFICIENT OF CONCURRENT DEVIATIONS

A very simple and casual method of finding correlation when we are not serious about the magnitude of the two variables is the application of concurrent deviations. This method involves in attaching a positive sign for a x-value (except the first) if this value is more than the previous value and assigning a negative value if this value is less than the previous value. This is done for the y-series as well. The deviation in the x-value and the corresponding y-value is known to be concurrent if both the deviations have the same sign.

Denoting the number of concurrent deviation by c and total number of deviations as m (which must be one less than the number of pairs of x and y values), the coefficient of concurrent deviation is given by

$$r_C = \pm\sqrt{\pm\frac{(2c-m)}{m}} \quad \text{...(18.13)}$$

If (2c–m) >0, then we take the positive sign both inside and outside the radical sign and if (2c–m) <0, we are to consider the negative sign both inside and outside the radical sign.

Like Pearson's correlation coefficient and Spearman's rank correlation coefficient, the coefficient of concurrent deviations also lies between –1 and 1, both inclusive.

Example 18.15: Find the coefficient of concurrent deviations from the following data.

Year :	1990	1991	1992	1993	1994	1995	1996	1997
Price :	25	28	30	23	35	38	39	42
Demand :	35	34	35	30	29	28	26	23

Solution:

Table 18.10

Computation of Coefficient of Concurrent Deviations.

Year	Price	Sign of deviation from the previous figure (a)	Demand	Sign of deviation from the previous figure (b)	Product of deviation (ab)
1990	25		35		
1991	28	+	34	–	–
1992	30	+	35	+	+
1993	23	–	30	–	+
1994	35	+	29	–	–
1995	38	+	28	–	–
1996	39	+	26	–	–
1997	42	+	23	–	–

In this case, m = number of pairs of deviations = 7

c = No. of positive signs in the product of deviation column = Number of concurrent deviations = 2

Thus $r_C = \pm\sqrt{\pm\frac{(2c-m)}{m}}$

$$= \pm\sqrt{\pm\frac{(4-7)}{m}}$$

$$= \pm\sqrt{\pm\frac{(-3)}{7}}$$

$$= \sqrt{\frac{3}{7}} = -0.65$$

(Since $\frac{2c-m}{m} = \frac{-3}{7}$ we take negative sign both inside and outside of the radical sign)

Thus there is a negative correlation between price and demand.

18.5 REGRESSION ANALYSIS

In regression analysis, we are concerned with the estimation of one variable for a given value of another variable (or for a given set of values of a number of variables) on the basis of an average mathematical relationship between the two variables (or a number of variables). Regression analysis plays a very important role in the field of every human activity. A businessman may be keen to know what would be his estimated profit for a given level of investment on the basis of the past records. Similarly, an outgoing student may like to know her chance of getting a first class in the final University Examination on the basis of her performance in the college selection test.

When there are two variables x and y and if y is influenced by x i.e. if y depends on x, then we get a simple linear regression or simple regression. y is known as dependent variable or regression or explained variable and x is known as independent variable or predictor or explanator. In the previous examples since profit depends on investment or performance in the University Examination is dependent on the performance in the college selection test, profit or performance in the University Examination is the dependent variable and investment or performance in the selection test is the In-dependent variable.

In case of a simple regression model if y depends on x, then the regression line of y on x in given by

$$y = a + bx \text{ .... } (18.14)$$

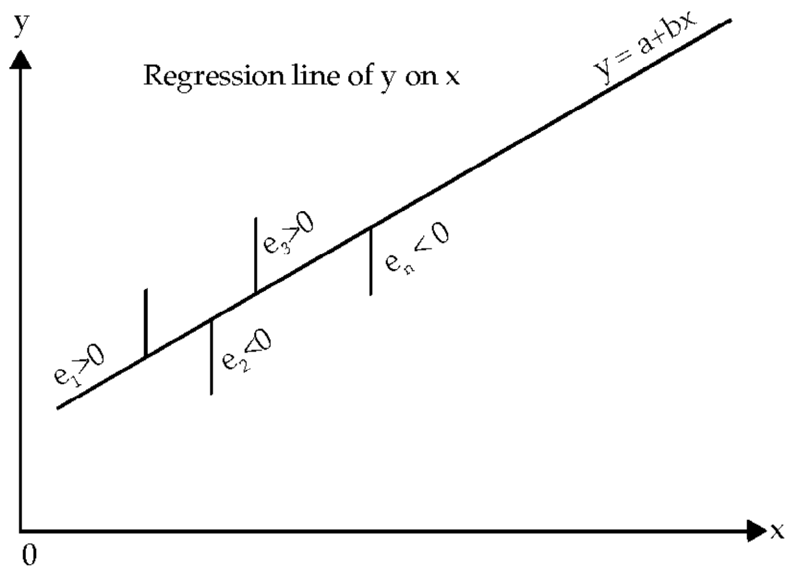
Here a and b are two constants and they are also known as regression parameters. Furthermore, b is also known as the regression coefficient of y on x and is also denoted by b_{yx}. We may define the regression line of y on x as the line of best fit obtained by the method of least squares and used for estimating the value of the dependent variable y for a known value of the independent variable x.

The method of least squares involves in minimizing

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2 \text{ .... (18.15)}$$

where y_i demotes the actual or observed value and $\hat{y}_i = a + b_{xi}$, the estimated value of y_i for a given value of x_i, e_i is the difference between the observed value and the estimated value and e_i is technically known as error or residue. This summation intends over n pairs of observations of (x_i, y_i). The line of regression of y or x and the errors of estimation are shown in the following figure.

FIGURE **18.7**



SHOWING REGRESSION LINE OF y on x

AND ERRORS OF ESTIMATION

Minimisation of (18.15) yields the following equations known as 'Normal Equations'

. $\sum y_i = na + b\sum x_i$ .... (18.16)

$\sum x_i y_i = a\sum x_i + b\sum x_i^2$ .... (18.17)

Solving there two equations for b and a, we have the "least squares" estimates of b and a as

$$b = \frac{Cov(x, y)}{S_x^{\,2}}$$

$$= \frac{r.S_x.S_y}{S_x^2}$$

$$= \frac{r.S_y}{S_x} \quad ....(18.18)$$

After estimating b, estimate of a is given by

$$a = \bar{y} - b\bar{x} \quad ....(18.19)$$

Substituting the estimates of b and a in (18.14), we get

$$\frac{(y-\bar{y})}{S_y} = \frac{r(x-\bar{x})}{S_x} \quad ....(18.20)$$

There may be cases when the variable x depends on y and we may take the regression line of x on y as

$x = a\hat{} + b\hat{}y$

Unlike the minimization of vertical distances in the scatter diagram as shown in figure (18.7) for obtaining the estimates of a and b, in this case we minimize the horizontal distances and get the following normal equation in a^ and b^, the two regression parameters :

$\sum x_i = na\hat{} + b\hat{}\sum y_i$ .... (18.21)

$\sum x_i y_i = a\hat{}\sum y_i + b\hat{}\sum y_i^2$.. (18.22)

or solving these equations, we get

$$b\hat{} = b_{xy} = \frac{\text{cov}(x, y)}{S_y^2} = \frac{r.S_x}{S_y} \quad ....(18.23)$$

and $a\hat{} = \bar{x} - b\hat{}\,\bar{y}$ ....(18.24)

A single formula for estimating b is given by

$$b\hat{} = b_{yx} = \frac{n\sum x_i y_i - \sum x_i . \sum y_i}{n\sum y_i^2 - (\sum y_i)^2} \quad ....(18.25)$$

Similarly, $b\hat{} = b_{yx} = \frac{n\sum x_i y_i - \sum x_i . \sum y_i}{n\sum y_i^2 - (\sum y_i)^2}$ ....(18.26)

The standardized form of the regression equation of x on y, as in (18.20), is given by

$$\frac{x-\bar{x}}{S_x} = r\frac{(y-\bar{y})}{S_y} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18.27)$$

Example 16.15: Find the two regression equations from the following data:

x:	2	4	5	5	8	10
y:	6	7	9	10	12	12

Hence estimate y when x is 13 and estimate also x when y is 15.

Solution:

Table 18.11

Computation of Regression Equations

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
2	6	12	4	36
4	7	28	16	49
5	9	45	25	81
5	10	50	25	100
8	12	96	64	144
10	12	120	100	144
34	56	351	234	554

On the basis of the above table, we have

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{34}{6} = 5.6667$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{56}{6} = 9.3333$$

$$\text{cov}(x, y) = \frac{\Sigma x_i y_i}{n} - \bar{x}\bar{y}$$

$$= \frac{351}{6} - 5.6667 \times 9.3333$$

$$= 58.50 – 52.8890$$

$$= 5.6110$$

$$S_x^2 = \frac{\Sigma x_i^2}{n} - (\bar{x})^2$$

$$= \frac{234}{6} - (5.6667)^2$$

$= 39 - 32.1115$

$= 6.8885$

$$S_y^2 = \frac{\sum y_i^2}{n} - (\bar{y})^2$$

$$= \frac{554}{6} - (9.3333)^2$$

$= 92.3333 - 87.1105$

$= 5.2228$

The regression line of y on x is given by

$y = a + bx$

$$\text{Where } b\hat{} = \frac{\text{cov}(x, y)}{S_x^2}$$

$$= \frac{5.6110}{6.8885}$$

$= 0.8145$

and $a\hat{} = \bar{y} - b\bar{x}$

$= 9.3333 - 0.8145 \times 5.6667$

$= 4.7178$

Thus the estimated regression equation of y on x is

$y = 4.7178 + 0.8145x$

When x = 13, the estimated value of y is given by $\hat{y} = 4.7178 + 0.8145 \times 13 = 15.3063$

The regression line of x on y is given by

$x = a\hat{} + b\hat{}\, y$

$$\text{Where } b\hat{} = \frac{\text{cov}(x, y)}{S_y^2}$$

$$= \frac{5.6110}{5.2228}$$

$= 1.0743$

and $\hat{a} = \bar{x} - \hat{b}\bar{y}$

$= 5.6667 - 1.0743 \times 9.3333$

$= -4.3601$

Thus the estimated regression line of x on y is

$x = -4.3601 + 1.0743y$

When y = 15, the estimate value of x is given by

$\hat{x} = -4.3601 + 1.0743 \times 15$

$= 11.75$

Example 18.16: Marks of 8 students in Mathematics and statistics are given as:

Mathematics:	80	75	76	69	70	85	72	68
Statistics:	85	65	72	68	67	88	80	70

Find the regression lines. When marks of a student in Mathematics are 90, what are his most likely marks in statistics?

Solution:

We denote the marks in Mathematics and Statistics by x and y respectively. We are to find the regression equation of y on x and also of x or y. Lastly, we are to estimate y when x = 90. For computation advantage, we shift origins of both x and y.

Table 18.12

Computation of regression lines

Maths mark (x_i)	Stats mark (y_i)	$u_i = x_i - 74$	$v_i = y_i - 76$	$u_i v_i$	u_i^2	v_i^2
80	85	6	9	54	36	81
75	65	1	–11	–11	1	121
76	72	2	–4	–8	4	16
69	68	–5	–8	40	25	64
70	67	–4	–9	36	16	81
85	88	11	12	132	121	144
72	80	–2	4	–8	4	16
68	70	–6	–6	36	36	36
595	595	3	–13	271	243	559

The regression coefficients b (or b_{yx}) and b′ (or b_{xy}) remain unchanged due to a shift of origin.

Applying (18.25) and (18.26), we get

$$b = b_{yx} = b_{vu} = \frac{n\sum u_i v_i - \sum u_i . \sum v_i}{n\sum u_i^2 - (\sum u_i)^2}$$

$$= \frac{8.(271) - (3).(-13)}{8.(243) - (3)^2}$$

$$= \frac{2168 + 39}{1944 - 9}$$

$$= 1.1406$$

$$\text{and } b\hat{} = b_{xy} = b_{uv} = \frac{n\sum u_i v_i - \sum u_i . \sum v_i}{n\sum v_i^2 - (\sum v_i)^2}$$

$$= \frac{8.(271) - (3).(-13)}{8.(559) - (-13)^2}$$

$$= \frac{2168 + 39}{4472 - 169}$$

$$= 0.5129$$

Also $a\hat{} = \bar{y} - b\hat{}\,\bar{x}$

$$= \frac{(595)}{8} - 1.1406\frac{(595)}{8}$$

$$= 74.375 - 1.1406 \times 74.375$$

$$= -10.4571$$

and $a\hat{} = \bar{x} - b\hat{}\,\bar{y}$

$$= 74.375 - 0.5129 \times 74.375$$

$$= 36.2280$$

The regression line of y on x is

$y = -10.4571 + 1.1406x$

and the regression line of x on y is

$x = 36.2281 + 0.5129y$

For x = 90, the most likely value of y is

$\hat{y}$ = –10.4571 + 1.1406 x 90

= 92.1969

≅ 92

Example 18.17: The following data relate to the mean and SD of the prices of two shares in a stock Exchange:

Share	Mean (in ₹)	SD (in ₹)
Company A	44	5.60
Company B	58	6.30

Coefficient of correlation between the share prices = 0.48

Find the most likely price of share A corresponding to a price of ₹ 60 of share B and also the most likely price of share B for a price of ₹ 50 of share A.

Solution:

Denoting the share prices of Company A and B respectively by x and y, we are given that

$\bar{x}$ = ₹ 44, $\bar{y}$ = ₹ 58

S_x = ₹ 5.60, S_y = ₹ 6.30

and r = 0.48

The regression line of y on x is given by

$$y = a + bx$$

Where $b = r \times \frac{S_y}{S_x}$

$$= 0.48 \times \frac{6.30}{5.60}$$

$$= 0.54$$

$$a = \bar{y} - b\bar{x}$$

= ₹ (58 – 0.54 × 44)

= ₹ 34.24

Thus the regression line of y on x i.e. the regression line of price of share B on that of share A is given by

y = ₹ (34.24 + 0.54x)

When x = ₹ 50, = ₹ (34.24 + 0.54 × 50)

= ₹ 61.24

= The estimated price of share B for a price of ₹ 50 of share A is ₹ 61.24

Again the regression line of x on y is given by

$x = \hat{a} + \hat{b}y$

Where $\hat{b} = r \times \frac{S_x}{S_y}$

$= 0.48 \times \frac{5.60}{6.30}$

$= 0.4267$

$\hat{a} = \bar{x} - \hat{b}\,\bar{y}$

= ₹ (44 – 0.4267 × 58)

= ₹ 19.25

Hence the regression line of x on y i.e. the regression line of price of share A on that of share B in given by

x = ₹ (19.25 + 0.4267y)

When y = ₹ 60, $\hat{x}$ = ₹ (19.25 + 0.4267 × 60)

= ₹ 44.85

Example 18.18: The following data relate the expenditure or advertisement in thousands of rupees and the corresponding sales in lakhs of rupees.

Expenditure on Ad	:	8	10	10	12	15
Sales	:	18	20	22	25	28

Find an appropriate regression equation.

Solution:

Since sales (y) depend on advertisement (x), the appropriate regression equation is of y on x i.e. of sales on advertisement. We have, on the basis of the given data,

$n = 5, \quad \sum x = 8+10+10+12+15 = 55$

$\sum y = 18+20+22+25+28 = 113$

$\sum xy = 8\times18+10\times20+10\times22+12\times25+15\times28 = 1284$

$\sum x^2 = 8^2+10^2+10^2+12^2+15^2 = 633$

$$\therefore b = \frac{n\sum xy - \sum x \times \sum y}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{5 \times 1284 - 55 \times 113}{5 \times 633 - (55)^2}$$

$$= \frac{205}{140}$$

$$= 1.4643$$

$$a = \bar{y} - b\bar{x}$$

$$= \frac{113}{5} - 1.4643 \times \frac{55}{5}$$

$$= 22.60 - 16.1073$$

$$= 6.4927$$

Thus, the regression line of y or x i.e. the regression line of sales on advertisement is given by

$$y = 6.4927 + 1.4643x$$

18.6 PROPERTIES OF REGRESSION LINES

We consider the following important properties of regression lines:

(i) **The regression coefficients remain unchanged due to a shift of origin but change due to a shift of scale.**

This property states that if the original pair of variables is (x, y) and if they are changed to the pair (u, v) where

$$u = \frac{x-a}{p} \text{ and } v = \frac{y-c}{q}$$

$$b_{yx} = \frac{q}{p} \times b_{vu} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18.28)$$

$$\text{and } b_{xy} = \frac{p}{q} \times b_{uv} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18.29)$$

(ii) **The two lines of regression intersect at the point $(\bar{x}, \bar{y})$, where x and y are the variables under consideration.**

According to this property, the point of intersection of the regression line of y on x and the regression line of x on y is $(\bar{x}, \bar{y})$ i.e. the solution of the simultaneous equations in x and y.

(iii) **The coefficient of correlation between two variables x and y in the simple geometric mean of the two regression coefficients. The sign of the correlation coefficient would be the common sign of the two regression coefficients.**

This property says that if the two regression coefficients are denoted by b_{yx} (=b) and b_{xy} (=b′) then the coefficient of correlation is given by

$$r = \pm\sqrt{b_{yx} \times b_{xy}} \quad \text{......................} \; (18.30)$$

If both the regression coefficients are negative, r would be negative and if both are positive, r would assume a positive value.

Example 18.19: If the relationship between two variables x and u is u + 3x = 10 and between two other variables y and v is 2y + 5v = 25, and the regression coefficient of y on x is known as 0.80, what would be the regression coefficient of v on u?

Solution:

$u + 3x = 10$

$$u = \frac{(x - 10/3)}{-1/3}$$

and $2y + 5v = 25$

$$\Rightarrow \qquad v = \frac{(y - 25/2)}{-5/2}$$

From (16.28), we have

$$b_{yx} = \frac{q}{p} \times b_{vu}$$

$$\text{or,} \qquad 0.80 = \frac{-5/2}{-1/3} \times b_{vu}$$

$$\Rightarrow \qquad 0.80 = \frac{15}{2} \times b_{vu}$$

$$\Rightarrow \qquad b_{vu} = \frac{2}{15} \times 0.80 = \frac{8}{75}$$

Example 18.20: For the variables x and y, the regression equations are given as 7x – 3y – 18 = 0 and 4x – y – 11 = 0

(i) Find the arithmetic means of x and y.

(ii) Identify the regression equation of y on x.

(iii) Compute the correlation coefficient between x and y.

(iv) Given the variance of x is 9, find the SD of y.

Solution:

(i) Since the two lines of regression intersect at the point $(\bar{x}, \bar{y})$, replacing x and y by $\bar{x}$ and $\bar{y}$ respectively in the given regression equations, we get

$$7\bar{x} - 3\bar{y} - 18 = 0$$

and $$4\bar{x} - \bar{y} - 11 = 0$$

Solving these two equations, we get $\bar{x} = 3$ and $\bar{y} = 1$

Thus the arithmetic means of x and y are given by 3 and 1 respectively.

(ii) Let us assume that $7x - 3y - 18 = 0$ represents the regression line of y on x and $4x - y - 11 = 0$ represents the regression line of x on y.

Now $7x - 3y - 18 = 0$

$$\Rightarrow \quad y = (-6) + \frac{(7)}{3}x$$

$$\therefore \quad b_{yx} = \frac{7}{3}$$

Again $4x - y - 11 = 0$

$$\Rightarrow \quad x = \frac{(11)}{4} + \frac{(1)}{4}y \qquad \therefore b_{xy} = \frac{1}{4}$$

Thus $r^2 = b_{yx} \times b_{xy}$

$$= \frac{7}{3} \times \frac{1}{4}$$

$$= \frac{7}{12} < 1$$

Since $|r| \leq 1 \Rightarrow r^2 \leq 1$, our assumptions are correct. Thus, $7x - 3y - 18 = 0$ truly represents the regression line of y on x.

(iii) Since $r^2 = \frac{7}{12}$

$\therefore \quad r = \sqrt{\frac{7}{12}}$ (We take the sign of r as positive since both the regression coefficients are positive)

$= 0.7638$

(iv) $b_{yx} = r \times \frac{S_y}{S_x}$

$\Rightarrow \frac{7}{3} = 0.7638 \times \frac{S_y}{3}$ ($\therefore S_x^2 = 9$ as given)

$\Rightarrow S_y = \frac{7}{0.7638}$

$= 9.1647$

18.7 PROBABLE ERROR

The correlation coefficient calculated from the sample of n pairs of value from large population. It is possible to determine the limits of the correlation coefficient of population and which coefficient of correlationof correlation of the population will lie from the knowledge of sample correlation coefficient.

Probable Error is a method of obtaining correlation coefficient of population. It is defined as:

$$P.E = 0.674 \times \frac{1-r^2}{\sqrt{N}}$$

Where r = Correlation coefficient fromn pairs of sample observations

$$PE = \frac{2}{3}\ SE$$

When SE = Standard Error of correlation coefficient

$$S.E = \times \frac{1-r^2}{\sqrt{N}}$$

The limit of the correlation coefficient is given by p = r ± P.E

Where p = Correlation coefficient of the population

The following are the assumption while probable Errors are significant.

(i) If r< PE there is no evidence of correlation

(ii) If the value of 'r 'is more than 6 times of the probable error, then the presence of correlation coefficient is certain

(iii) Since 'r 'lies between -1 and +1 ($-1 \leq r \leq 1$) the probable error is never negative.

Note: The formula PE is valued onlyif

(1) The sample chooses to find 'r'is a sample random sample (2) the population is normal.

Example 18.21:

Compute the Probable Error assuming the correlation coefficient of 0.8 from a sampleof 25 pairs of items.

Solution: r = 0.8 ,n = 25

P.E. = 0.6745 ×

= 0.6745 × 0.07 = 0.0486

Example 18.22:

If r = 0.7 ; and n = 64 find out the probable error of the coefficient of correlationand determine the limits for the population correlation coefficient:

Solution:

r = 0.7 ; n= 64

$$\text{Probable Error (P.E.)} = 0.6745 \times \frac{1 - (0.7)^2}{\sqrt{64}}$$

= (0.6745) × (0.06375)

= 0.043

Limits for the population correlation coefficient

(0.7 ± 0.043)

i.e. (0.743, 0.657)

18.8 REVIEW OF CORRELATION AND REGRESSION ANALYSIS

So far we have discussed the different measures of correlation and also how to fit regression lines applying the method of 'Least Squares'. It is obvious that we take recourse to correlation analysis when we are keen to know whether two variables under study are associated or correlated and if correlated, what is the strength of correlation. The best measure of correlation is provided by Pearson's correlation coefficient. However, one severe limitation of this correlation coefficient, as we have already discussed, is that it is applicable only in case of a linear relationship between the two variables.

If two variables x and y are independent or uncorrelated then obviously the correlation coefficient between x and y is zero. However, the converse of this statement is not necessarily true i.e. if the correlation coefficient, due to Pearson, between two variables comes out to be zero, then we cannot conclude that the two variables are independent. All that we can conclude is that no linear relationship exists between the two variables. This, however, does not rule out the existence of some non linear relationship between the two variables. For example, if we consider the following pairs of values on two variables x and y.

(–2, 4), (–1, 1), (0, 0), (1, 1) and (2, 4), then cov (x, y) = (–2+ 4) + (–1+1) + (0×0) + (1×1) + (2×4) = 0

as $\bar{x} = 0$

Thus $r_{xy} = 0$

This does not mean that x and y are independent. In fact the relationship between x and y is $y = x^2$. Thus it is always wiser to draw a scatter diagram before reaching conclusion about the existence of correlation between a pair of variables.

There are some cases when we may find a correlation between two variables although the two variables are not causally related. This is due to the existence of a third variable which is related to both the variables under consideration. Such a correlation is known as spurious correlation or non-sense correlation. As an example, there could be a positive correlation between production of rice and that of iron in India for the last twenty years due to the effect of a third variable time on both these variables. It is necessary to eliminate the influence of the third variable before computing correlation between the two original variables.

Correlation coefficient measuring a linear relationship between the two variables indicates the amount of variation of one variable accounted for by the other variable. A better measure for this purpose is provided by the square of the correlation coefficient, Known as 'coefficient of determination'. This can be interpreted as the ratio between the explained variance to total variance i.e.

$$r^2 = \frac{\text{Explained variance}}{\text{Total variance}}$$

Thus a value of 0.6 for r indicates that $(0.6)^2 \times 100\%$ or 36 per cent of the variation has been accounted for by the factor under consideration and the remaining 64 per cent variation is due to other factors. The 'coefficient of non-determination' is given by $(1–r^2)$ and can be interpreted as the ratio of unexplained variance to the total variance.

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

Regression analysis, as we have already seen, is concerned with establishing a functional relationship between two variables and using this relationship for making future projection. This can be applied, unlike correlation for any type of relationship linear as well as curvilinear. **The two lines of regression coincide i.e. become identical when r = –1 or 1 or in other words, there is a perfect negative or positive correlation between the two variables under discussion. If r = 0 Regression lines are perpendicular to each other.**

SUMMARY

- The change in one variable is reciprocated by a corresponding change in the other variable either directly or inversely, then the two variables are known to be associated or correlated.

  There are two types of correlation.

  (i) Positive correlation

  (ii) Negative correlation

- We consider the following measures of correlation:

(a) Scatter diagram: This is a simple diagrammatic method to establish correlation between a pair of variables.

(b) Karl Pearson's Product moment correlation coefficient:

$$r = r_{xy} = \frac{Cov(x, y)}{S_x \times S_y}$$

A single formula for computing correlation coefficient is given by

$$r = \frac{n\sum x_i y_i - \sum x_i \times \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum {}_i^2 - \left(\sum y_i\right)^2}}$$

(i) The Coefficient of Correlation is a unit-free measure.

(ii) The coefficient of correlation remains invariant under a change of origin and/or scale of the variables under consideration depending on the sign of scale factors.

(iii) The coefficient of correlation always lies between –1 and 1, including both the limiting values i.e. $-1 \le r \le +1$

(c) Spearman's rank correlation co-efficient: Spearman's rank correlation coefficient is given by

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$, where r_R denotes rank correlation coefficient and it lies between –1 and 1 inclusive of these two values. $d_i = x_i - y_i$ represents the difference in ranks for the i-th individual and n denotes the number of individuals.

In case u individuals receive the same rank, we describe it as a tied rank of length u. In case of a tied rank,

$$r_R = 1 - \frac{6\left[\sum_i d_i + \sum_j \frac{(t_j^3 - t_j)}{12}\right]}{n(n^2-1)}$$

In this formula, t_j represents the j^{th} tie length and the summation extends over the lengths of all the ties for both the series.

(d) Co-efficient of concurrent deviations: The coefficient of concurrent deviation is given by

$$r_C = \pm\sqrt{\pm\frac{(2c-m)}{m}}$$

If (2c–m) >0, then we take the positive sign both inside and outside the radical sign and if (2c–m) <0, we are to consider the negative sign both inside and outside the radical sign.

- In regression analysis, we are concerned with the estimation of one variable for given value of another variable (or for a given set of values of a number of variables) on the basis of an average mathematical relationship between the two variables (or a number of variables).
- In case of a simple regression model if y depends on x, then the regression line of y on x in given by y = a + b, here a and b are two constants and they are also known as regression parameters. Furthermore, b is also known as the regression coefficient of y on x and is also denoted by b_{yx}
- The method of least squares is solving the equations of regression lines

  The normal equations are

  $\Sigma y_i = na + b\Sigma x_i$

  $\Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2$

  Solving the normal equations

  $$b = \frac{cov(x_i y_i)}{S_x^2} = \frac{r.S_x.S_y}{S_x^2}$$

- The regression coefficients remain unchanged due to a shift of origin but change due to a shift of scale.

  This property states that if the original pair of variables is (x, y) and if they are changed to the pair (u, v) where

  $$u = \frac{x-a}{p} \text{ and } v = \frac{y-c}{q}$$

  $$b_{yx} = \frac{p}{q} \times b_{vu} \text{ and } bxy = \frac{q}{p} \times b_{uv}$$

- The two lines of regression intersect at the point , where x and y are the variables under consideration.

  According to this property, the point of intersection of the regression line of y on x and the regression line of x on y is i.e. the solution of the simultaneous equations in x and y.
- The coefficient of correlation between two variables x and y in the simple geometric mean of the two regression coefficients. The sign of the correlation coefficient would be the common sign of the two regression coefficients.

  $$r = \pm\sqrt{b_{yx} \times b_{xy}}$$

- Correlation coefficient measuring a linear relationship between the two variables indicates the amount of variation of one variable accounted for by the other variable. A better measure for this purpose is provided by the square of the correlation coefficient, Known as 'coefficient of determination'. This can be interpreted as the ratio between the explained variance to total variance i.e.

$$r^2 = \frac{\text{Explained variance}}{\text{Total variance}}$$

- The 'coefficient of non-determination' is given by $(1–r^2)$ and can be interpreted as the ratio of unexplained variance to the total variance.
- The two lines of regression coincide i.e. become identical when r = –1 or 1 or in other words, there is a perfect negative or positive correlation between the two variables under discussion. If r = 0 Regression lines are perpendicular to each other.

EXERCISE

Set A

Write the correct answers. Each question carries 1 mark.

1. Bivariate Data are the data collected for
   (a) Two variables
   (b) More than two variables
   (c) Two variables at the same point of time
   (d) Two variables at different points of time.
2. For a bivariate frequency table having (p + q) classification the total number of cells is
   (a) p (b) p + q
   (c) q (d) pq
3. Some of the cell frequencies in a bivariate frequency table may be
   (a) Negative (b) Zero
   (c) a or b (d) Non of these
4. For a p x q bivariate frequency table, the maximum number of marginal distributions is
   (a) p (b) p + q
   (c) 1 (d) 2
5. For a p x q classification of bivariate data, the maximum number of conditional distributions is
   (a) p (b) p + q
   (c) pq (d) p or q
6. Correlation analysis aims at
   (a) Predicting one variable for a given value of the other variable
   (b) Establishing relation between two variables

(c) Measuring the extent of relation between two variables

(d) Both (b) and (c).

7. Regression analysis is concerned with

(a) Establishing a mathematical relationship between two variables

(b) Measuring the extent of association between two variables

(c) Predicting the value of the dependent variable for a given value of the independent variable

(d) Both (a) and (c).

8. What is spurious correlation?

(a) It is a bad relation between two variables.

(b) It is very low correlation between two variables.

(c) It is the correlation between two variables having no causal relation.

(d) It is a negative correlation.

9. Scatter diagram is considered for measuring

(a) Linear relationship between two variables

(b) Curvilinear relationship between two variables

(c) Neither (a) nor (b)

(d) Both (a) and (b).

10. If the plotted points in a scatter diagram lie from upper left to lower right, then the correlation is

(a) Positive (b) Zero

(c) Negative (d) None of these.

11. If the plotted points in a scatter diagram are evenly distributed, then the correlation is

(a) Zero (b) Negative

(c) Positive (d) (a) or (b).

11. If all the plotted points in a scatter diagram lie on a single line, then the correlation is

(a) Perfect positive (b) Perfect negative

(c) Both (a) and (b) (d) Either (a) or (b).

13. The correlation between shoe-size and intelligence is

(a) Zero (b) Positive

(c) Negative (d) None of these.

14. The correlation between the speed of an automobile and the distance travelled by it after applying the brakes is

(a) Negative (b) Zero

(c) Positive (d) None of these.

15. Scatter diagram helps us to
    (a) Find the nature correlation between two variables
    (b) Compute the extent of correlation between two variables
    (c) Obtain the mathematical relationship between two variables
    (d) Both (a) and (c).
16. Pearson's correlation coefficient is used for finding
    (a) Correlation for any type of relation
    (b) Correlation for linear relation only
    (c) Correlation for curvilinear relation only
    (d) Both (b) and (c).
17. Product moment correlation coefficient is considered for
    (a) Finding the nature of correlation
    (b) Finding the amount of correlation
    (c) Both (a) and (b)
    (d) Either (a) and (b).
18. If the value of correlation coefficient is positive, then the points in a scatter diagram tend to cluster
    (a) From lower left corner to upper right corner
    (b) From lower left corner to lower right corner
    (c) From lower right corner to upper left corner
    (d) From lower right corner to upper right corner.
19. When r = 1, all the points in a scatter diagram would lie
    (a) On a straight line directed from lower left to upper right
    (b) On a straight line directed from upper left to lower right
    (c) On a straight line
    (d) Both (a) and (b).
20. Product moment correlation coefficient may be defined as the ratio of
    (a) The product of standard deviations of the two variables to the covariance between them
    (b) The covariance between the variables to the product of the variances of them
    (c) The covariance between the variables to the product of their standard deviations
    (d) Either (b) or (c).
21. The covariance between two variables is
    (a) Strictly positive
    (b) Strictly negative
    (c) Always 0
    (d) Either positive or negative or zero.
22. The coefficient of correlation between two variables

(a) Can have any unit.

(b) Is expressed as the product of units of the two variables

(c) Is a unit free measure

(d) None of these.

23. What are the limits of the correlation coefficient?

(a) No limit (b) –1 and 1

(c) 0 and 1, including the limits (d) –1 and 1, including the limits

24. In case the correlation coefficient between two variables is 1, the relationship between the two variables would be

(a) $y = a + bx$ (b) $y = a + bx, b > 0$

(c) $y = a + bx, b < 0$ (d) $y = a + bx$, both a and b being positive.

25. If the relationship between two variables x and y in given by $2x + 3y + 4 = 0$, then the value of the correlation coefficient between x and y is

(a) 0 (b) 1

(c) –1 (d) negative.

26. For finding correlation between two attributes, we consider

(a) Pearson's correlation coefficient

(b) Scatter diagram

(c) Spearman's rank correlation coefficient

(d) Coefficient of concurrent deviations.

27. For finding the degree of agreement about beauty between two Judges in a Beauty Contest, we use

(a) Scatter diagram (b) Coefficient of rank correlation

(c) Coefficient of correlation (d) Coefficient of concurrent deviation.

28. If there is a perfect disagreement between the marks in Geography and Statistics, then what would be the value of rank correlation coefficient?

(a) Any value (b) Only 1

(c) Only –1 (d) (b) or (c)

29. When we are not concerned with the magnitude of the two variables under discussion, we consider

(a) Rank correlation coefficient (b) Product moment correlation coefficient

(c) Coefficient of concurrent deviation (d) (a) or (b) but not (c).

30. What is the quickest method to find correlation between two variables?

(a) Scatter diagram (b) Method of concurrent deviation

(c) Method of rank correlation (d) Method of product moment correlation

31. What are the limits of the coefficient of concurrent deviations?
    (a) No limit
    (b) Between –1 and 0, including the limiting values
    (c) Between 0 and 1, including the limiting values
    (d) Between –1 and 1, the limiting values inclusive
32. If there are two variables x and y, then the number of regression equations could be
    (a) 1 (b) 2
    (c) Any number (d) 3.
33. Since Blood Pressure of a person depends on age, we need consider
    (a) The regression equation of Blood Pressure on age
    (b) The regression equation of age on Blood Pressure
    (c) Both (a) and (b)
    (d) Either (a) or (b).
34. The method applied for deriving the regression equations is known as
    (a) Least squares (b) Concurrent deviation
    (c) Product moment (d) Normal equation.
35. The difference between the observed value and the estimated value in regression analysis is known as
    (a) Error (b) Residue
    (c) Deviation (d) (a) or (b).
36. The errors in case of regression equations are
    (a) Positive (b) Negative
    (c) Zero (d) All these.
37. The regression line of y on x is derived by
    (a) The minimisation of vertical distances in the scatter diagram
    (b) The minimisation of horizontal distances in the scatter diagram
    (c) Both (a) and (b)
    (d) (a) or (b).
38. The two lines of regression become identical when
    (a) $r = 1$ (b) $r = -1$
    (c) $r = 0$ (d) (a) or (b).
39. What are the limits of the two regression coefficients?
    (a) No limit (b) Must be positive
    (c) One positive and the other negative
    (d) Product of the regression coefficient must be numerically less than unity.

40. The regression coefficients remain unchanged due to a

(a) Shift of origin (b) Shift of scale

(c) Both (a) and (b) (d) (a) or (b).

41. If the coefficient of correlation between two variables is –0 9, then the coefficient of determination is

(a) 0.9 (b) 0.81

(c) 0.1 (d) 0.19.

42. If the coefficient of correlation between two variables is 0.7 then the percentage of variation unaccounted for is

(a) 70% (b) 30%

(c) 51% (d) 49%

Set B

Answer the following questions by writing the correct answers. Each question carries 2 marks.

1. If for two variable x and y, the covariance, variance of x and variance of y are 40, 16 and 256 respectively, what is the value of the correlation coefficient?

(a) 0.01 (b) 0.625

(c) 0.4 (d) 0.5

2. If cov(x, y) = 15, what restrictions should be put for the standard deviations of x and y?

(a) No restriction.

(b) The product of the standard deviations should be more than 15.

(c) The product of the standard deviations should be less than 15.

(d) The sum of the standard deviations should be less than 15.

3. If the covariance between two variables is 20 and the variance of one of the variables is 16, what would be the variance of the other variable?

(a) More than 100 (b) More than 10

(c) Less than 10 (d) More than 1.25

4. If $y = a + bx$, then what is the coefficient of correlation between x and y?

(a) 1 (b) –1

(c) 1 or –1 according as $b > 0$ or $b < 0$ (d) none of these.

5. If $r = 0.6$ then the coefficient of non-determination is

(a) 0.4 (b) –0.6

(c) 0.36 (d) 0.64

6. If $u + 5x = 6$ and $3y – 7v = 20$ and the correlation coefficient between x and y is 0.58 then what would be the correlation coefficient between u and v?

(a) 0.58 (b) –0.58

(c) –0.84 (d) 0.84

7. If the relation between x and u is 3x + 4u + 7 = 0 and the correlation coefficient between x and y is –0.6, then what is the correlation coefficient between u and y?
   (a) –0.6 (b) 0.8
   (c) 0.6 (d) –0.8
8. From the following data

x:	2	3	5	4	7
y:	4	6	7	8	10

Two coefficient of correlation was found to be 0.93. What is the correlation between u and v as given below?

u:	–3	–2	0	–1	2
v:	–4	–2	–1	0	2

   (a) –0.93 (b) 0.93 (c) 0.57 (d) –0.57
9. Referring to the data presented in Q. No. 8, what would be the correlation between u and v?

u:	10	15	25	20	35
v:	–24	–36	–42	–48	–60

   (a) –0.6 (b) 0.6 (c) –0.93 (d) 0.93
10. If the sum of squares of difference of ranks, given by two judges A and B, of 8 students in 21, what is the value of rank correlation coefficient?
   (a) 0.7 (b) 0.65 (c) 0.75 (d) 0.8
11. If the rank correlation coefficient between marks in management and mathematics for a group of student in 0.6 and the sum of squares of the differences in ranks in 66, what is the number of students in the group?
   (a) 10 (b) 9 (c) 8 (d) 11
12. While computing rank correlation coefficient between profit and investment for the last 6 years of a company the difference in rank for a year was taken 3 instead of 4. What is the rectified rank correlation coefficient if it is known that the original value of rank correlation coefficient was 0.4?
   (a) 0.3 (b) 0.2 (c) 0.25 (d) 0.28
13. For 10 pairs of observations, No. of concurrent deviations was found to be 4. What is the value of the coefficient of concurrent deviation?
   (a) $\sqrt{0.2}$ (b) $-\sqrt{0.2}$ (c) 1/3 (d) –1/3
14. The coefficient of concurrent deviation for p pairs of observations was found to be $1/\sqrt{3}$. If the number of concurrent deviations was found to be 6, then the value of p is.
   (a) 10 (b) 9 (c) 8 (d) none of these
15. What is the value of correlation coefficient due to Pearson on the basis of the following data:

x:	–5	–4	–3	–2	–1	0	1	2	3	4	5
y:	27	18	11	6	3	2	3	6	11	18	27

   (a) 1 (b) –1 (c) 0 (d) –0.5

16. Following are the two normal equations obtained for deriving the regression line of y and x:

    $5a + 10b = 40$

    $10a + 25b = 95$

    The regression line of y on x is given by

    (a) $2x + 3y = 5$ (b) $2y + 3x = 5$ (c) $y = 2 + 3x$ (d) $y = 3 + 5x$

17. If the regression line of y on x and of x on y are given by $2x + 3y = -1$ and $5x + 6y = -1$ then the arithmetic means of x and y are given by

    (a) (1, –1) (b) (–1, 1) (c) (–1, –1) (d) (2, 3)

18. Given the regression equations as $3x + y = 13$ and $2x + 5y = 20$, which one is the regression equation of y on x?

    (a) 1st equation (b) 2nd equation (c) both (a) and (b) (d) none of these.

19. Given the following equations: $2x - 3y = 10$ and $3x + 4y = 15$, which one is the regression equation of x on y ?

    (a) 1st equation (b) 2nd equation (c) both the equations (d) none of these

20. If $u = 2x + 5$ and $v = -3y - 6$ and regression coefficient of y on x is 2.4, what is the regression coefficient of v on u?

    (a) 3.6 (b) –3.6 (c) 2.4 (d) –2.4

21. If $4y - 5x = 15$ is the regression line of y on x and the coefficient of correlation between x and y is 0.75, what is the value of the regression coefficient of x on y?

    (a) 0.45 (b) 0.9375 (c) 0.6 (d) none of these

22. If the regression line of y on x and that of x on y are given by $y = -2x + 3$ and $8x = -y + 3$ respectively, what is the coefficient of correlation between x and y?

    (a) 0.5 (b) $-1/\sqrt{2}$ (c) –0.5 (d) none of these

23. If the regression coefficient of y on x, the coefficient of correlation between x and y and variance of y are $-3/4$, $\frac{\sqrt{3}}{2}$ and 4 respectively, what is the variance of x?

    (a) $2/\sqrt{3/2}$ (b) 16/3 (c) 4/3 (d) 4

24. If $y = 3x + 4$ is the regression line of y on x and the arithmetic mean of x is –1, what is the arithmetic mean of y?

    (a) 1 (b) –1 (c) 7 (d) none of these

SET C

Write down the correct answers. Each question carries 5 marks.

1. What is the coefficient of correlation from the following data?

x:	1	2	3	4	5
y:	8	6	7	5	5

   (a) 0.75 (b) –0.75 (c) –0.85 (d) 0.82

2. The coefficient of correlation between x and y where

x:	64	60	67	59	69
y:	57	60	73	62	68

is

(a) 0.655 (b) 0.68 (c) 0.73 (d) 0.758

3. What is the coefficient of correlation between the ages of husbands and wives from the following data?

Age of husband (year):	46	45	42	40	38	35	32	30	27	25
Age of wife (year):	37	35	31	28	30	25	23	19	19	18

(a) 0.58 (b) 0.98 (c) 0.89 (d) 0.92

4. The following results relate to bivariate data on (x, y):
$\sum xy = 414$, $\sum x = 120$, $\sum y = 90$, $\sum x^2 = 600$, $\sum y^2 = 300$, n = 30. Later or, it was known that two pairs of observations (12, 11) and (6, 8) were wrongly taken, the correct pairs of observations being (10, 9) and (8, 10). The corrected value of the correlation coefficient is

(a) 0.752 (b) 0.768 (c) 0.846 (d) 0.953

5. The following table provides the distribution of items according to size groups and also the number of defectives:

Size group:	9-11	11-13	13-15	15-17	17-19
No. of items:	250	350	400	300	150
No. of defective items:	25	70	60	45	20

The correlation coefficient between size and defectives is

(a) 0.25 (b) 0.12 (c) 0.14 (d) 0.07

6. For two variables x and y, it is known that cov (x, y) = 8 =0.4, variance of x is 16 and sum of squares of deviation of y from its mean is 250. The number of observations for this bivariate data is

(a) 7 (b) 8 (c) 9 (d) 10

7. Eight contestants in a musical contest were ranked by two judges A and B in the following manner:

Serial Number of the contestants:	1	2	3	4	5	6	7	8
Rank by Judge A:	7	6	2	4	5	3	1	8
Rank by Judge B:	5	4	6	3	8	2	1	7

The rank correlation coefficient is

(a) 0.65 (b) 0.63 (c) 0.60 (d) 0.57

8. Following are the marks of 10 students in Botany and Zoology:

Serial No.:	1	2	3	4	5	6	7	8	9	10

Marks in

Botany:	58	43	50	19	28	24	77	34	29	75
Marks in Zoology:	62	63	79	56	65	54	70	59	55	69

The coefficient of rank correlation between marks in Botany and Zoology is

(a) 0.65 (b) 0.70 (c) 0.72 (d) 0.75

9. What is the value of Rank correlation coefficient between the following marks in Physics and Chemistry:

Roll No.:	1	2	3	4	5	6
Marks in Physics:	25	30	46	30	55	80
Marks in Chemistry:	30	25	50	40	50	78

(a) 0.782 (b) 0.696 (c) 0.932 (d) 0.857

10. What is the coefficient of concurrent deviations for the following data:

Supply:	68	43	38	78	66	83	38	23	83	63	53
Demand:	65	60	55	61	35	75	45	40	85	80	85

(a) 0.82 (b) 0.85 (c) 0.89 (d) –0.81

11. What is the coefficient of concurrent deviations for the following data:

Year:	1996	1997	1998	1999	2000	2001	2002	2003
Price:	35	38	40	33	45	48	49	52
Demand:	36	35	31	36	30	29	27	24

(a) –0.43 (b) 0.43 (c) 0.5 (d) $\sqrt{2}$

12. The regression equation of y on x for the following data:

x	41	82	62	37	58	96	127	74	123	100
y	28	56	35	17	42	85	105	61	98	73

Is given by

(a) y = 1.2x – 15 (b) y = 1.2x + 15 (c) y = 0.93x – 14.64 (d) y = 1.5x – 10.89

13. The following data relate to the heights of 10 pairs of fathers and sons:

(175, 173), (172, 172), (167, 171), (168, 171), (172, 173), (171, 170), (174, 173), (176, 175) (169, 170), (170, 173)

The regression equation of height of son on that of father is given by

(a) y = 100 + 5x (b) y = 99.708 + 0.405x (c) y = 89.653 + 0.582x (d) y = 88.758 + 0.562x

14. The two regression coefficients for the following data:

x:	38	23	43	33	28
y:	28	23	43	38	8

are

(a) 1.2 and 0.4 (b) 1.6 and 0.8 (c) 1.7 and 0.8 (d) 1.8 and 0.3

15. For y = 25, what is the estimated value of x, from the following data:

X:	11	12	15	16	18	19	21
Y:	21	15	13	12	11	10	9

(a) 15 (b) 13.926 (c) 13.588 (d) 14.986

16. Given the following data:

Variable:	x	y
Mean:	80	98
Variance:	4	9

Coefficient of correlation = 0.6

What is the most likely value of y when x = 90 ?

(a) 90 (b) 103 (c) 104 (d) 107

17. The two lines of regression are given by

$8x + 10y = 25$ and $16x + 5y = 12$ respectively.

If the variance of x is 25, what is the standard deviation of y?

(a) 16 (b) 8 (c) 64 (d) 4

18. Given below the information about the capital employed and profit earned by a company over the last twenty five years:

	Mean	SD
Capital employed (0000 ₹)	62	5
Profit earned (000 ₹)	25	6

Correlation coefficient between capital employed and profit = 0.92. The sum of the Regression coefficients for the above data would be:

(a) 1.871 (b) 2.358 (c) 1.968 (d) 2.346

19. The coefficient of correlation between cost of advertisement and sales of a product on the basis of the following data:

Ad cost (000 ₹):	75	81	85	105	93	113	121	125
Sales (000 000 ₹):	35	45	59	75	43	79	87	95

is

(a) 0.85 (b) 0.89 (c) 0.95 (d) 0.98

ANSWERS

Set A

1.	(c)	**2.**	(d)	**3.**	(b)	**4.**	(d)	**5.**	(b)	**6.**	(d)
7.	(d)	**8.**	(c)	**9.**	(d)	**10.**	(c)	**11.**	(a)	**12.**	(d)
13.	(a)	**14.**	(a)	**15.**	(a)	**16.**	(b)	**17.**	(c)	**18.**	(a)
19.	(a)	**20.**	(c)	**21.**	(d)	**22.**	(c)	**23.**	(d)	**24.**	(b)
25.	(c)	**26.**	(c)	**27.**	(b)	**28.**	(c)	**29.**	(c)	**30.**	(b)
31.	(d)	**32.**	(b)	**33.**	(a)	**34.**	(a)	**35.**	(d)	**36.**	(d)
37.	(a)	**38.**	(d)	**39.**	(d)	**40.**	(a)	**41.**	(b)	**42.**	(c)

Set B

1.	(b)	**2.**	(b)	**3.**	(a)	**4.**	(c)	**5.**	(d)	**6.**	(b)
7.	(c)	**8.**	(b)	**9.**	(c)	**10.**	(c)	**11.**	(a)	**12.**	(b)
13.	(d)	**14.**	(a)	**15.**	(c)	**16.**	(c)	**17.**	(a)	**18.**	(b)
19.	(d)	**20.**	(b)	**21.**	(a)	**22.**	(c)	**23.**	(b)	**24.**	(a)

Set C

1.	(c)	**2.**	(a)	**3.**	(b)	**4.**	(c)	**5.**	(d)	**6.**	(d)
7.	(d)	**8.**	(d)	**9.**	(d)	**10.**	(c)	**11.**	(a)	**12.**	(c)
13.	(b)	**14.**	(a)	**15.**	(c)	**16.**	(d)	**17.**	(b)	**18.**	(a)
19.	(c)										

ADDITIONAL QUESTION BANK

1. _______________ is concerned with the measurement of the "strength of association" between variables.

   (a) correlation (b) regression (c) both (d) none

2. _______________ gives the mathematical relationship of the variables.

   (a) correlation (b) regression (c) both (d) none

3. When high values of one variable are associated with high values of the other & low values of one variable are associated with low values of another, then they are said to be

   (a) positively correlated (b) directly correlated
   (c) both (d) none

4. If high values of one tend to low values of the other, they are said to be

   (a) negatively correlated (b) inversely correlated
   (c) both (d) none

5. Correlation coefficient between two variables is a measure of their linear relationship .

   (a) true (b) false (c) both (d) none

6. Correlation coefficient is dependent of the choice of both origin & the scale of observations.

   (a) True (b) false (c) both (d) none

7. Correlation coefficient is a pure number.

   (a) true (b) false (c) both (d) none

8. Correlation coefficient is _______________ of the units of measurement.

   (a) dependent (b) independent (c) both (d) none

9. The value of correlation coefficient lies between

   (a) –1 and +1 (b) –1 and 0
   (c) 0 and 1 Inclusive of these two values (d) none.

10. Correlation coefficient can be found out by

    (a) Scatter Diagram (b) Rank Method (c) both (d) none.

11. Covariance measures __________ variations of two variables.

    (a) joint (b) single (c) both (d) none

12. In calculating the Karl Pearson's coefficient of correlation it is necessary that the data should be of numerical measurements. The statement is

    (a) valid (b) not valid (c) both (d) none

13. Rank correlation coefficient lies between

    (a) 0 to 1 (b) –1 to +1 inclusive of these value
    (c) –1 to 0 (d) both

14. A coefficient near +1 indicates tendency for the larger values of one variable to be associated with the larger values of the other.

    (a) true (b) false (c) both (d) none

15. In rank correlation coefficient the association need not be linear.

    (a) true (b) false (c) both (d) none

16. In rank correlation coefficient only an increasing/decreasing relationship is required.

    (a) false (b) true (c) both (d) none

17. Great advantage of ____________ is that it can be used to rank attributes which can not be expressed by way of numerical value .

    (a) concurrent correlation (b) regression
    (c) rank correlation (d) none

18. The sum of the difference of rank is

    (a) 1 (b) –1 (c) 0 (d) none.

19. Karl Pearson's coefficient is defined from

    (a) ungrouped data (b) grouped data (c) both (d) none.

20. Correlation methods are used to study the relationship between two time series of data which are recorded annually, monthly, weekly, daily and so on.

    (a) True (b) false (c) both (d) none

21. Age of Applicants for life insurance and the premium of insurance – correlation is

    (a) positive (b) negative (c) zero (d) none

22. "Unemployment index and the purchasing power of the common man"____ Correlation is

    (a) positive (b) negative (c) zero (d) none

23. Production of pig iron and soot content in Durgapur – Correlations are

    (a) positive (b) negative (c) zero (d) none

24. "Demand for goods and their prices under normal times" ____ Correlation is

    (a) positive (b) negative (c) zero (d) none

25. ___________ is a relative measure of association between two or more variables.

    (a) Coefficient of correlation (b) Coefficient of regression
    (c) both (d) none

26. The lines of regression passes through the points, bearing _________ no. of points on both sides

    (a) equal (b) unequal (c) zero (d) none

27. Under Algebraic Method we get ————— linear equations .

    (a) one (b) two (c) three (d) none

28. In linear equations Y = a + bX and X= a + bY 'a' is the

(a) intercept of the line (b) slope
(c) both (d) none

29. In linear equations Y = a + bX and X = a + bY ' b ' is the

(a) intercept of the line (b) slope of the line
(c) both (d) none

30. The regression equations Y = a + bX and X = a + bY are based on the method of

(a) greatest squares (b) least squares (c) both (d) none

31. The line Y = a + bX represents the regression equation of

(a) Y on X (b) X on Y (c) both (d) none

32. The line X = a + bY represents the regression equation of

(a) Y on X (b) X onY (c) both (d) none

33. Two regression lines always intersect at the means.

(a) true (b) false (c) both (d) none

34. r, b_{xy}, b_{yx} all have ______ sign.

(a) different (b) same (c) both (d) none

35. The regression coefficients are zero if r is equal to

(a) 2 (b) –1 (c) 1 (d) 0

36. The regression lines are identical if r is equal to

(a) +1 (b) –1 (c) ± 1 (d) 0

37. The regression lines are perpendicular to each other if r is equal to
(a) 0 (b) +1 (c) –1 (d) ± 1

38. Feature of Least Square regression lines are __________ The sum of the deviations at the Y's or the X's from their regression lines are zero.

(a) true (b) false (c) both (d) none

39. The coefficient of determination is defined by the formula

(a) $r^2 = 1 - \frac{\text{unexplained variance}}{\text{total variance}}$ (b) $r^2 = \frac{\text{explained variance}}{\text{total variance}}$

(c) both (d) none

40. If the line Y = 13 –3X /2 is the regression equation of y on x then byx is

(a) $\frac{2}{3}$ (b) $\frac{-2}{3}$ (c) $\frac{3}{2}$ (d) $\frac{-3}{2}$

41. In the line Y = 19 – 5X/2 is the regresson equation x on y then bxy is,
(a) 19/2 (b) 5/2 (c) –5/2 (d) –2/5

42. The line X = 31/6 — Y/6 is the regression equation of
(a) Y on X (b) X on Y (c) both (d) we can not say

43. In the regression equation x on y, X = 35/8 – 2Y /5, b_{xy} is equal to
(a) –2/5 (b) 35/8 (c) 2/5 (d) 5/2

44. The square of coefficient of correlation 'r' is called the coefficient of
(a) determination (b) regression (c) both (d) none

45. A relationship $r^2 = 1 - \frac{500}{300}$ is not possible
(a) true (b) false (c) both (d) none

46. Whatever may be the value of r, positive or negative, its square will be
(a) negative only (b) positive only (c) zero only (d) none only

47. Simple correlation is called
(a) linear correlation (b) nonlinear correlation
(c) both (d) none

48. A scatter diagram indicates the type of correlation between two variables.
(a) true (b) false (c) both (d) none

49. If the pattern of points (or dots) on the scatter diagram shows a linear path diagonally across the graph paper from the bottom left- hand corner to the top right, correlation will be
(a) negative (b) zero (c) positive (d) none

50. The correlation coefficient being +1 if the slope of the straight line in a scatter diagram is
(a) positive (b) negative (c) zero (d) none

51. The correlation coefficient being –1 if the slope of the straight line in a scatter diagram is
(a) positive (b) negative (c) zero (d) none

52. The more scattered the points are around a straight line in a scattered diagram the ________ is the correlation coefficient.
(a) zero (b) more (c) less (d) none

53. If the values of y are not affected by changes in the values of x, the variables are said to be
(a) correlated (b) uncorrelated (c) both (d) zero

54. If the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable, then correlation is said to be
(a) non linear (b) linear (c) both (d) none

55. Variance may be positive, negative or zero.
(a) true (b) false (c) both (d) none

56. Covariance may be positive, negative or zero.

(a) true (b) false (c) both (d) none

57. Correlation coefficient between x and y = correlation coefficient between u and v

(a) true (b) false (c) both (d) none

58. In case ' The ages of husbands and wives' _______________ correlation is

(a) positive (b) negative (c) zero (d) none

59. In case 'Shoe size and intelligence'

(a) positive correlation (b) negative correlation
(c) no correlation (d) none

60. In case 'Insurance companies' profits and the no of claims they have to pay "___________

(a) positive correlation (b) negative correlation
(c) no correlation (d) none

61. In case 'Years of education and income' ____________

(a) positive correlation (b) negative correlation
c) no correlation (d) none

62. In case 'Amount of rainfall and yield of crop'____________

(a) positive correlation (b) negative correlation
(c) no correlation (d) none

63. For calculation of correlation coefficient, a change of origin is

(a) not possible (b) possible (c) both (d) none

64. The relation r_{xy} = cov (x,y)/ $\sigma_x.\sigma_y$ is

(a) true (b) false (c) both (d) none

65. A small value of r indicates only a _________ linear type of relationship between the variables.

(a) good (b) poor (c) maximum (d) highest

66. Two regression lines coincide when

(a) r = 0 (b) r = 2 (c) r = $\pm$ 1 (d) none

67. Neither y nor x can be estimated by a linear function of the other variable when r is equal to

(a) + 1 (b) – 1 (c) 0 (d) none

68. When r = 0 then cov (x,y) is equal to

(a) + 1 (b) – 1 (c) 0 (d) none

69. When the variables are not independent, the correlation coefficient may be zero

(a) true (b) false (c) both (d) none

70. b_{xy} is called regression coefficient of

(a) x on y (b) y on x (c) both (d) none

71. b_{yx} is called regression coefficient of

(a) x on y (b) y on x (c) both (d) none

72. The slopes of the regression line of y on x is

(a) b_{yx} (b) b_{xy} (c) b_{xx} (d) b_{yy}

73. The slopes of the regression line of x on y is

(a) b_{yx} (b) b_{xy} (c) $1/b_{xy}$ (d) $1/b_{yx}$

74. The angle between the regression lines depends on

(a) correlation coefficient (b) regression coefficient
(c) both (d) none

75. If x and y satisfy the relationship y = –5 + 7x, the value of r is

(a) 0 (b) – 1 (c) + 1 (d) none

76. If b_{yx} and b_{xy} are negative, r is

(a) positive (b) negative (c) zero (d) none

77. Correlation coefficient r lie between the regression coefficients b_{yx} and b_{xy}

(a) true (b) false (c) both (d) none

78. Since the correlation coefficient r cannot be greater than 1 numerically, the product of the regression must

(a) not exceed 1 (b) exceed 1 (c) be zero (d) none

79. The correlation coefficient r is the __________ of the two regression coefficients b_{yx} and b_{xy}

(a) A.M (b) G.M (c) H.M (d) none

80. Which is true?

(a) $b_{yx} = r\frac{\sigma_x}{\sigma_y}$ (b) $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

(c) $b_{yx} = r\frac{\sigma_{xy}}{\sigma_x}$ (d) $b_{yx} = r\frac{\sigma_{yy}}{\sigma_x}$

81. Maximum value of Rank Correlation coefficient is

(a) –1 (b) + 1 (c) 0 (d) none

82. The partial correlation coefficient lies between

(a) –1 and +1 inclusive of these two value (b) 0 and + 1
(c) –1 and (d) none

83. r_{12} is the correlation coefficient between

(a) x_1 and x_2 (b) x_2 and x_1 (c) x_1 and x_3 (d) x_2 and x_3

84. r_{12} is the same as r_{21}

(a) true (b) false (c) both (d) none

85. In case of employed persons 'Age and income' correlation is

(a) positive (b) negative (c) zero (d) none

86. In case 'Speed of an automobile and the distance required to stop the car often applying brakes' – correlation is

(a) positive (b) negative (c) zero (d) none

87. In case 'Sale of woolen garments and day temperature' ____________ correlation is

(a) positive (b) negative (c) zero (d) none

88. In case 'Sale of cold drinks and day temperature' ____________ correlation is

(a) positive (b) negative (c) zero (d) none

89. In case of 'Production and price per unit' – correlation is

(a) positive (b) negative (c) zero (d) none

90. If slopes at two regression lines are equal then r is equal to

(a) 1 (b) ± 1 (c) 0 (d) none

91. Co–variance measures the joint variations of two variables.

(a) true (b) false (c) both (d) none

92. The minimum value of correlation coefficient is

(a) 0 (b) –2 (c) 1 (d) –1

93. The maximum value of correlation coefficient is

(a) 0 (b) 2 (c) 1 (d) –1

94. When r = 0 , the regression coefficients are

(a) 0 (b) 1 (c) –1 (d) none

95. The regression equation of Y on X is, $2x + 3Y + 50 = 0$. The value of b_{YX} is

(a) 2/3 (b) – 2/3 (c) –3/2 (d) none

96. In Method of Concurrent Deviations, only the directions of change (Positive direction / Negative direction) in the variables are taken into account for calculation of

(a) coefficient of S.D (b) coefficient of regression.
(c) coefficient of correlation (d) none

ANSWERS

1.	(a)	2.	(b)	3.	(c)	4.	(c)	5.	(a)
6.	(b)	7.	(a)	8.	(b)	9.	(a)	10.	(b)
11.	(a)	12.	(a)	13.	(b)	14.	(a)	15.	(a)
16.	(b)	17.	(c)	18.	(c)	19.	(b)	20.	(a)
21.	(a)	22.	(b)	23.	(a)	24.	(b)	25.	(a)
26.	(d)	27.	(b)	28.	(a)	29.	(b)	30.	(b)
31.	(a)	32.	(b)	33.	(a)	34.	(b)	35.	(d)
36.	(c)	37.	(a)	38.	(a)	39.	(c)	40.	(d)
41.	(d)	42.	(b)	43.	(a)	44.	(a)	45.	(a)
46.	(b)	47.	(a)	48.	(a)	49.	(c)	50.	(a)
51.	(b)	52.	(c)	53.	(b)	54.	(b)	55.	(b)
56.	(a)	57.	(a)	58.	(a)	59.	(c)	60.	(b)
61.	(a)	62.	(a)	63.	(b)	64.	(a)	65.	(b)
66.	(c)	67.	(c)	68.	(c)	69.	(a)	70.	(a)
71.	(b)	72.	(a)	73.	(b)	74.	(a)	75.	(c)
76.	(b)	77.	(a)	78.	(a)	79.	(b)	80.	(b)
81.	(b)	82.	(a)	83.	(a)	84.	(a)	85.	(a)
86.	(a)	87.	(b)	88.	(a)	89.	(b)	90.	(b)
91.	(a)	92.	(d)	93.	(c)	94.	(a)	95.	(b)
96.	(c)								

CHAPTER 19

UNIT 1: INDEX NUMBERS

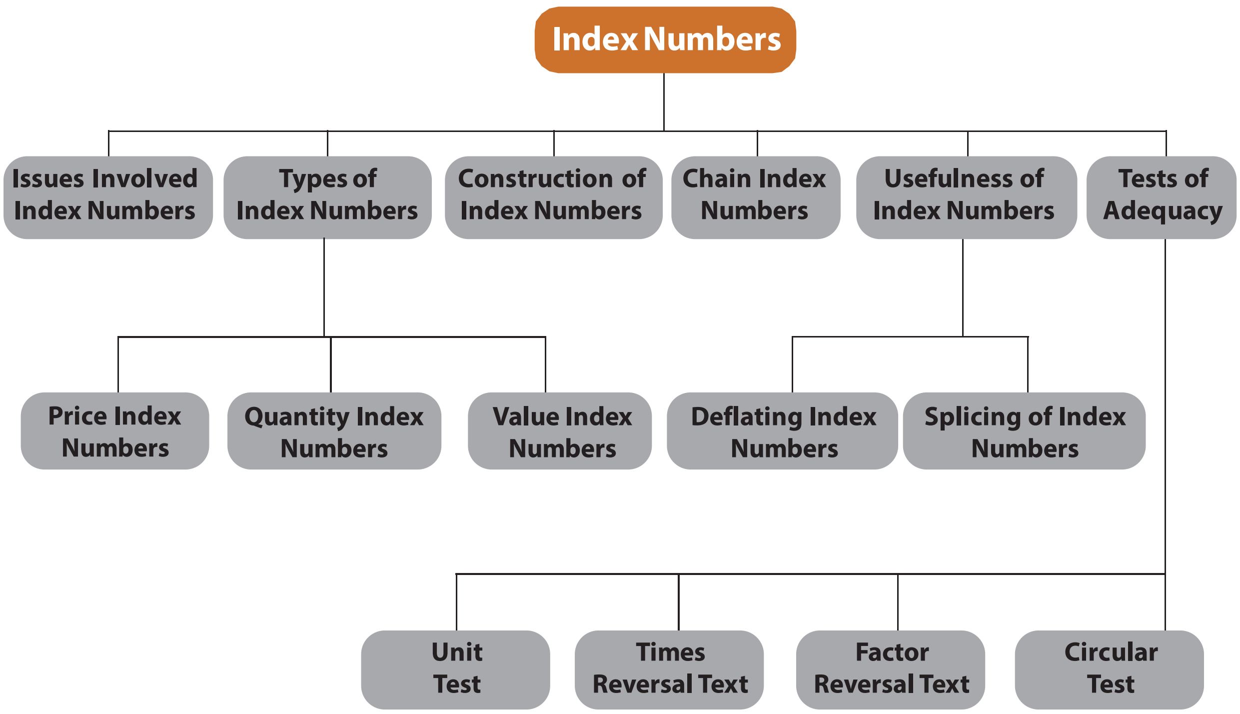
LEARNING OBJECTIVES

Often we encounter news of price rise, GDP growth, production growth, etc. It is important for students of Chartered Accountancy to learn techniques of measuring growth/rise or decline of various economic and business data and how to report them objectively.

After reading the chapter, students will be able to understand:

- Purpose of constructing index number and its important applications in understanding rise or decline of production, prices, etc.
- Different methods of computing index number.

UNIT OVERVIEW



19.1.1 INTRODUCTION

Index numbers are convenient devices for measuring relative changes of differences from time to time or from place to place. Just as the arithmetic mean is used to represent a set of values, an index number is used to represent a set of values over two or more different periods or localities.

The basic device used in all methods of index number construction is to average the relative change in either quantities or prices since relatives are comparable and can be added even though the data from which they were derived cannot themselves be added. For example, if wheat production has gone up to 110% of the previous year's producton and cotton production has gone up to 105%, it is possible to average the two percentages as they have gone up by 107.5%. This assumes that both have equal weight; but if wheat production is twice as important as cotton, percentage should be weighted 2 and 1. The average relatives obtained through this process are called the index numbers.

Definition: An index number is a ratio of two or more time periods are involved, one of which is the base time period. The value at the base time period serves as the standard point of comparison.

Example: NSE, BSE, WPI, CPI etc.

An index time series is a list of index numbers for two or more periods of time, where each index number employs the same base year.

Relatives are derived because absolute numbers measured in some appropriate unit, are often of little importance and meaningless in themselves. If the meaning of a relative figure remains ambiguous, it is necessary to know the absolute as well as the relative number.

Our discussion of index numbers is confined to various types of index numbers, their uses, the mathematical tests and the principles involved in the construction of index numbers.

Index numbers are studied here because some techniques for making forecasts or inferences about the figures are applied in terms of index number. In regression analysis, either the independent or dependent variable or both may be in the form of index numbers. They are less unwieldy than large numbers and are readily understandable.

These are of two broad types: simple and composite. The simple index is computed for one variable whereas the composite is calculated from two or more variables. Most index numbers are composite in nature.

19.1.2 ISSUES INVOLVED

Following are some of the important criteria/problems which have to be faced in the construction of index Numbers.

Selection of data: It is important to understand the purpose for which the index is used. If it is used for purposes of knowing the cost of living, there is no need of including the prices of capital goods which do not directly influence the living.

Index numbers are often constructed from the sample. It is necessary to ensure that it is representative. Random sampling, and if need be, a stratified random sampling can ensure this.

It is also necessary to ensure comparability of data. This can be ensured by consistency in the method of selection of the units for compilation of index numbers.

However, difficulties arise in the selection of commodities because the relative importance of commodities keep on changing with the advancement of the society. More so, if the period is quite long, these changes are quite significant both in the basket of production and the uses made by people.

Base Period: It should be carefully selected because it is a point of reference in comparing various data describing individual behaviour. The period should be normal i.e., one of the relative stability, not affected by extraordinary events like war, famine, etc. It should be relatively recent because we are more concerned with the changes with reference to the present and not with the distant past. There are three variants of the base fixed, chain, and the average.

Selection of Weights: It is necessary to point out that each variable involved in composite index should have a reasonable influence on the index, i.e., due consideration should be given to the relative importance of each variable which relates to the purpose for which the index is to be used. For example, in the computation of cost of living index, sugar cannot be given the same importance as the cereals.

Use of Averages: Since we have to arrive at a single index number summarising a large amount of information, it is easy to realise that average plays an important role in computing index numbers. The geometric mean is better in averaging relatives, but for most of the indices arithmetic mean is used because of its simplicity.

Choice of Variables: Index numbers are constructed with regard to price or quantity or any other measure. We have to decide about the unit. For example, in price index numbers it is necessary to decide whether to have wholesale or the retail prices. The choice would depend on the purpose. Further, it is necessary to decide about the period to which such prices will be related. There may be an average of price for certain time-period or the end of the period. The former is normally preferred.

Selection of Formula: The question of selection of an appropriate formula arises, since different types of indices give different values when applied to the same data. We will see different types of indices to be used for construction succeedingly.

19.1.3 CONSTRUCTION OF INDEX NUMBER

Notations: It is customary to let $P_n(^1)$, $P_n(^2)$, $P_n(^3)$ denote the prices during n^{th} period for the first, second and third commodity. The corresponding price during a base period are denoted by $P_o(^1)$, $P_o(^2)$, $P_o(^3)$, etc. With these notations the price of commodity j during period n can be indicated by $P_n(^j)$. We can use the summation notation by summing over the superscripts j as follows:

$$\sum_{j=1}^{k} P_n(j) \quad \text{or} \quad \sum P_n(j)$$

We can omit the superscript altogether and write as ΣP_n etc.

Relatives: One of the simplest examples of an index number is a price relative, which is the ratio of the price of single commodity in a given period to its price in another period called the base period or the reference period. It can be indicated as follows:

$$\text{Price relative} = \frac{P_n}{P_o}$$

It has to be expressed as a percentage, it is multiplied by 100

$$\text{Price relative} = \frac{P_n}{P_o} \times 100$$

There can be other relatives such as of quantities, volume of consumption, exports, etc. The relatives in that case will be:

$$\text{Quantity relative} = \frac{Q_n}{Q_o}$$

Similarly, there are value relatives:

$$\text{Value relative} = \frac{V_n}{V_o} = \frac{P_n Q_n}{P_o Q_o} = \left(\frac{P_n}{P_o} \times \frac{Q_n}{Q_o}\right)$$

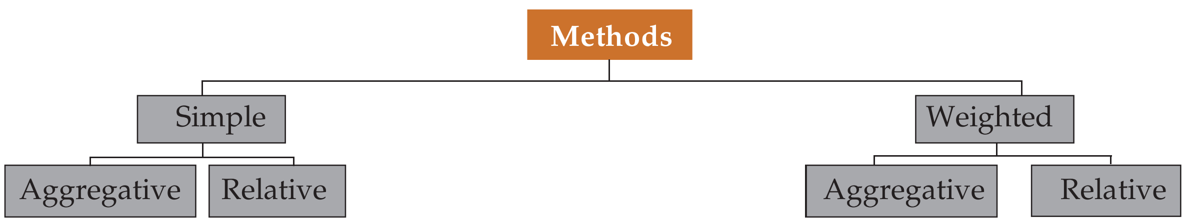
When successive prices or quantities are taken, the relatives are called the link relative,

$$\frac{P_1}{P_o}, \frac{P_2}{P_1}, \frac{P_3}{P_2}, \frac{P_n}{P_{n-1}}$$

When the above relatives are in respect to a fixed base period these are also called the chain relatives with respect to this base or the relatives chained to the fixed base. They are in the form of :

$$\frac{P_1}{P_o}, \frac{P_2}{P_o}, \frac{P_3}{P_o}, \frac{P_n}{P_o}$$

Methods: We can state the broad heads as follows:



19.1.3.1 SIMPLE AGGREGATIVE METHOD

In this method of computing a price index, we express the total of commodity prices in a given year as a percentage of total commodity price in the base year. In symbols, we have

$$\text{Simple aggregative price index} = \frac{\Sigma P_n}{\Sigma P_o} \times 100$$

where ΣP_n is the sum of all commodity prices in the current year and ΣP_o is the sum of all commodity prices in the base year.

ILLUSTRATIONS:

Commodities	1998	1999	2000
Cheese (per 100 gms)	12.00	15.00	15.60
Egg (per piece)	3.00	3.60	3.30
Potato (per kg)	5.00	6.00	5.70
Aggregrate	20.00	24.60	24.60
Index	100	123	123

Simple Aggregative Index for 1999 over 1998 = $\frac{\Sigma P_n}{\Sigma P_o} = \frac{24.60}{20.00} \times 100 = 123$

and for 2000 over 1998 = $\frac{\Sigma P_n}{\Sigma P_o} \times 100 = \frac{24.60}{20.00} \times 100 = 123$

The above method is easy to understand but it has a serious defect. It shows that the first commodity exerts greater influence than the other two because the price of the first commodity is higher than that of the other two. Further, if units are changed then the Index numbers will also change. Students should independently calculate the Index number taking the price of eggs per dozen i.e., ₹ 36, ₹ 43.20, ₹ 39.60 for the three years respectively. This is the major flaw in using absolute quantities and not the relatives. Such price quotations become the concealed weights which have no logical significance.

19.1.3.2 SIMPLE AVERAGE OF RELATIVES

One way to rectify the drawbacks of a simple aggregative index is to construct a simple average of relatives. Under it we invert the actual price for each variable into percentage of the base period. These percentages are called relatives because they are relative to the value for the base period. The index number is the average of all such relatives. One big advantage of price relatives is that they are pure numbers. Price index number computed from relatives will remain the same regardless of the units by which the prices are quoted. This method thus meets criterion of unit test (discussed later). Also quantity index can be constructed for a group of variables that are expressed in divergent units.

ILLUSTRATIONS:

In the proceeding example we will calculate relatives as follows:

Commodities	1998	1999	2000
A	100.0	125.0	130.0
B	100.0	120.0	110.0
C	100.0	120.0	114.0
Aggregate	300.0	365.0	354.0
Index	100.0	127.7	118.0

Inspite of some improvement, the above method has a flaw that it gives equal importance to each of the relatives. This amounts to giving undue weight to a commodity which is used in a small quantity because the relatives which have no regard to the absolute quantity will give weight more than what is due from the quantity used. This defect can be remedied by the introduction of an appropriate weighing system.

19.1.3.3 WEIGHTED METHOD

To meet the weakness of the simple or unweighted methods, we weigh the price of each commodity by a suitable factor often taken as the quantity or the volume of the commodity sold during the base year or some typical year. These indices can be classfied into broad groups:

(i) Weighted Aggregative Index.

(ii) Weighted Average of Relatives.

(i) Weighted Aggregative Index: Under this method we weigh the price of each commodity by a suitable factor often taken as the quantity or value weight sold during the base year or the given year or an average of some years. The choice of one or the other will depend on the importance we want to give to a period besides the quantity used. The indices are usually calculated in percentages. The various alternatives formulae in use are:

(The example has been given after the tests).

(a) Laspeyres' Index: In this Index base year quantities are used as weights:

$$\text{Laspeyres Index} = \frac{\Sigma P_n Q_0}{\Sigma P_0 Q_0} \times 100$$

(b) Paasche's Index: In this Index current year quantities are used as weights:

$$\text{Passche's Index} = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n} \times 100$$

(c) Methods based on some typical Period:

Index $\frac{\Sigma P_n Q_t}{\Sigma P_o Q_t} \times 100$ the subscript t stands for some typical period of years, the quantities of which are used as weight

Note: * Indices are usually calculated as percentages using the given formulae

The Marshall-Edgeworth index uses this method by taking the average of the base year and the current year

$$\text{Marshall-Edgeworth Index} = \frac{\Sigma P_n (Q_o + Q_n)}{\Sigma P_o (Q_o + Q_n)} \times 100$$

(d) Fisher's ideal Price Index: This index is the geometric mean of Laspeyres' and Paasche's.

$$\text{Fisher's Index} = \sqrt{\frac{\Sigma P_n Q_o}{\Sigma P_o Q_o} \times \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n}} \times 100$$

(ii) **Weighted Average of Relative Method:** To overcome the disadvantage of a simple average of relative method, we can use weighted average of relative method. Generally weighted arithmetic mean is used although the weighted geometric mean can also be used. The weighted arithmetic mean of price relatives using base year value weights is represented by

$$\frac{\sum \frac{P_n}{P_o} \times (P_o Q_o)}{\sum P_o Q_o} \times 100 = \frac{\sum P_n Q_o}{\sum P_o Q_o} \times 100$$

Example:

		Price Relatives			*Value Weights*	*Weighted Price Relatives*	
Commodity	Q.	1998	1999	2000	1998	1999	2000
		$\frac{P_n}{P_0}$	$\frac{P_n}{P_0}$	$\frac{P_n}{P_0}$	P_0Q_0	$\frac{P_n}{P_0} P_0Q_0$	$\frac{P_n}{P_0} P_0Q_0$
Butter	0.7239	100	101.1	118.7	72.39	73.19	85.93
Milk	0.2711	100	101.7	126.7	27.11	27.57	34.35
Eggs	0.7703	100	100.9	117.8	77.03	77.72	90.74
Fruits	4.6077	100	96.0	114.7	460.77	442.24	528.50
Vegetables	1.9500	100	84.0	93.6	195.00	163.80	182.52
					832.30	784.62	922.04

Weighted Price Relative

For 1999 : $\frac{784.62}{832.30} \times 100 = 94.3$

For 2000 : $\frac{922.04}{832.30} \times 100 = 110.8$

19.1.3.4 THE CHAIN INDEX NUMBERS

So far we concentrated on a fixed base but it does not suit when conditions change quite fast. In such a case the changing base for example, 1998 for 1999, and 1999 for 2000, and so on, may be more suitable. If, however, it is desired to associate these relatives to a common base the results may be chained. Thus, under this method the relatives of each year are first related to the preceding year called the link relatives and then they are chained together by successive multiplication to form a chain index.

The formula is:

$$\text{Chain Index} = \frac{\text{Link relative of current year} \times \text{Chain Index of the previous year}}{100}$$

Example:

The following are the index numbers by a chain base method:

Year (1)	*Price* (2)	*Link Relatives* (3)	*Chain Indices* (4)
1991	50	100	100
1992	60	$\frac{60}{50} \times 100 = 120.0$	$\frac{120}{100} \times 100 = 120.0$
1993	62	$\frac{62}{60} \times 100 = 103.3$	$\frac{103.3}{100} \times 120 = 124.0$
1994	65	$\frac{65}{62} \times 100 = 104.8$	$\frac{104.8}{100} \times 124 = 129.9$
1995	70	$\frac{70}{65} \times 100 = 107.7$	$\frac{107.7}{100} \times 129.9 = 139.9$
1996	78	$\frac{78}{70} \times 100 = 111.4$	$\frac{111.4}{100} \times 139.9 = 155.8$
1997	82	$\frac{82}{78} \times 100 = 105.1$	$\frac{105.1}{100} \times 155.8 = 163.7$
1998	84	$\frac{84}{82} \times 100 = 102.4$	$\frac{102.4}{100} \times 163.7 = 167.7$
1999	88	$\frac{88}{84} \times 100 = 104.8$	$\frac{104.8}{100} \times 167.7 = 175.7$
2000	90	$\frac{90}{88} \times 100 = 102.3$	$\frac{102.3}{100} \times 175.7 = 179.7$

You will notice that link relatives reveal annual changes with reference to the previous year. But when they are chained, they change over to a fixed base from which they are chained, which in the above example is the year 1991. The chain index is an unnecessary complication unless of course where data for the whole period are not available or where commodity basket or the weights have to be changed. The link relatives of the current year and chain index from a given base will give also a fixed base index with the given base year as shown in the column 4 above.

19.1.3.5 QUANTITY INDEX NUMBERS

To measure and compare prices, we use price index numbers. When we want to measure and compare quantities, we resort to Quantity Index Numbers. Though price indices are widely used to measure the economic strength, Quantity indices are used as indicators of the level of output in economy. To construct Quantity indices, we measure changes in quantities and weight them using prices or values as weights. The various types of Quantity indices are:

1. Simple aggregate of quantities:

   This has the formula $\frac{\Sigma Q_n}{\Sigma Q_o}$

2. The simple average of quantity relatives:

   This can be expressed by the formula $\frac{\Sigma \frac{Q_n}{Q_o}}{N}$

3. Weighted aggregate Quantity indices:

   (i) With base year weight : $\frac{\Sigma Q_n P_o}{\Sigma Q_o P_o}$ (Laspeyre's index)

   (ii) With current year weight : $\frac{\Sigma Q_n P_n}{\Sigma Q_o P_n}$ (Paasche's index)

   (iii) Geometric mean of (i) and (ii) : $\sqrt{\frac{\Sigma Q_n P_o}{\Sigma Q_o P_o} \times \frac{\Sigma Q_n P_n}{\Sigma Q_o P_n}}$ (Fisher's Ideal)

4. Base-year weighted average of quantity relatives. This has the formula $\frac{\Sigma\left(\frac{Q_n}{Q_o} P_o Q_o\right)}{\Sigma P_o Q_o}$

Note : Indices are usually calculated as percentages using the given formulae.

19.1.3.6 VALUE INDICES

Value equals price multiplied by quantity. Thus a value index equals the total sum of the values of a given year divided by the sum of the values of the base year, i.e.,

$$\frac{\Sigma V_n}{\Sigma V_o} = \frac{\Sigma P_n Q_n}{\Sigma P_0 Q_0}$$

19.1.4 USEFULNESS OF INDEX NUMBERS

So far we have studied various types of index numbers. However, they have certain limitations. They are :

1. As the indices are constructed mostly from deliberate samples, chances of errors creeping in cannot be always avoided.
2. Since index numbers are based on some selected items, they simply depict the broad trend and not the real picture.
3. Since many methods are employed for constructing index numbers, the result gives different values and this at times create confusion.

In spite of its limitations, index numbers are useful in the following areas :

1. Framing suitable policies in economics and business. They provide guidelines to make decisions in measuring intelligence quotients, research etc.
2. They reveal trends and tendencies in making important conclusions in cyclical forces, irregular forces, etc.
3. They are important in forecasting future economic activity. They are used in time series analysis to study long-term trend, seasonal variations and cyclical developments.
4. Index numbers are very useful in deflating i.e., they are used to adjust the original data for price changes and thus transform nominal wages into real wages.
5. Cost of living index numbers measure changes in the cost of living over a given period.

19.1.5 DEFLATING TIME SERIES USING INDEX NUMBERS

Sometimes a price index is used to measure the real values in economic time series data expressed in monetary units. For example, GNP initially is calculated in current price so that the effect of price changes over a period of time gets reflected in the data collected. Thereafter, to determine how much the physical goods and services have grown over time, the effect of changes in price over different values of GNP is excluded. The real economic growth in terms of constant prices of the base year therefore is determined by deflating GNP values using price index.

Year	*Wholesale Price Index*	*GNP at Current Prices*	*Real GNP*
1970	113.1	7499	6630
1971	116.3	7935	6823
1972	121.2	8657	7143
1973	127.7	9323	7301

The formula for conversion can be stated as

$$\text{Deflated Value} = \frac{\text{Current Value}}{\text{Price Index of the current year}}$$

$$\text{or Current Value} \times \frac{\text{Base Price }(P_0)}{\text{Current Price }(P_n)}$$

19.1.6 SHIFTING AND SPLICING OF INDEX NUMBERS

These refer to two technical points: (i) how the base period of the index may be shifted, (ii) how two index covering different bases may be combined into single series by splicing.

Shifted Price Index

Year	*Original Price Index*	*Shifted Price Index to base 1990*
1980	100	71.4
1981	104	74.3
1982	106	75.7
1983	107	76.4
1984	110	78.6
1985	112	80.0
1986	115	82.1
1987	117	83.6
1988	125	89.3
1989	131	93.6
1990	140	100.0
1991	147	105.0

The formula used is,

$$\text{Shifted Price Index} = \frac{\text{Original Price Index}}{\text{Price Index of the year on which it has to be shifted}} \times 100$$

Splicing two sets of price index numbers covering different periods of time is usually required when there is a major change in quantity weights. It may also be necessary on account of a new method of calculation or the inclusion of new commodity in the index.

Splicing Two Index Number Series

Year	*Old Price Index [1990 = 100]*	*Revised Price Index [1995 = 100]*	*Spliced Price Index [1995 = 100]*
1990	100.0		87.6
1991	102.3		89.6
1992	105.3		92.2
1993	107.6		94.2
1994	111.9		98.0
1995	114.2	100.0	100.0
1996		102.5	102.5
1997		106.4	106.4
1998		108.3	108.3
1999		111.7	111.7
2000		117.8	117.8

You will notice that the old series upto 1994 has to be converted shifting to the base. 1995 i.e, 114.2 to have a continuous series, even when the two parts have different weights

19.1.7 TEST OF ADEQUACY

There are four tests:

(i) **Unit Test:** This test requires that the formula should be independent of the unit in which or for which prices and quantities are quoted. Except for the simple (unweighted) aggregative index all other formulae satisfy this test.

(ii) **Time Reversal Test:** It is a test to determine whether a given method will work both ways in time, forward and backward. The test provides that the formula for calculating the index number should be such that two ratios, the current on the base and the base on the current should multiply into unity. In other words, the two indices should be reciprocals of each other. Symbolically,

$$P_{01} \times P_{10} = 1$$

where P_{01} is the index for time 1 on 0 and P_{10} is the index for time 0 on 1.

You will notice that Laspeyres' method and Paasche's method do not satisfy this test, but Fisher's Ideal Formula does.

While selecting an appropriate index formula, the Time Reversal Test and the Factor Reversal test are considered necessary in testing the consistency.

Laspeyres:

$$P_{01} = \frac{\Sigma P_1Q_0}{\Sigma P_0Q_0}, \qquad P_{10} = \frac{\Sigma P_0Q_1}{\Sigma P_1Q_1}$$

$$P_{01} \times P_{10} = \frac{\Sigma P_1Q_0}{\Sigma P_0Q_0} \times \frac{\Sigma P_0Q_1}{\Sigma P_1Q_1} \neq 1$$

Paasche's

$$P_{01} = \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1}, \qquad P_{10} = \frac{\Sigma P_0Q_0}{\Sigma P_1Q_0}$$

$$\therefore \quad P_{01} \times P_{10} = \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1} \times \frac{\Sigma P_0Q_0}{\Sigma P_1Q_0} \neq 1$$

Fisher's :

$$P_{01} = \sqrt{\frac{\Sigma P_1Q_0}{\Sigma P_0Q_0} \times \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1}} \qquad P_{10} = \sqrt{\frac{\Sigma P_0Q_1}{\Sigma P_1Q_1} \times \frac{\Sigma P_0Q_0}{\Sigma P_1Q_0}}$$

$$\therefore \quad P_{01} \times P_{10} = \sqrt{\frac{\Sigma P_1Q_0}{\Sigma P_0Q_0} \times \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1} \times \frac{\Sigma P_0Q_1}{\Sigma P_1Q_1} \times \frac{\Sigma P_0Q_0}{\Sigma P_1Q_0}} = 1$$

(iii) Factor Reversal Test: This holds when the product of price index and the quantity index should be equal to the corresponding value index, i.e., $\frac{\Sigma P_1Q_1}{\Sigma P_0Q_0}$

Symbolically: $P_{01} \times Q_{01} = V_{01}$

Fishers'

$$P_{01} = \sqrt{\frac{\Sigma P_1Q_0}{\Sigma P_0Q_0} \times \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1}} \qquad Q_{01} = \sqrt{\frac{\Sigma P_1Q_0}{\Sigma Q_0P_0} \times \frac{\Sigma P_1Q_1}{\Sigma Q_0P_1}}$$

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma P_1Q_0}{\Sigma P_0Q_0} \times \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1} \times \frac{\Sigma Q_1P_0}{\Sigma Q_0P_0} \times \frac{\Sigma Q_1P_1}{\Sigma Q_0P_1}} = \sqrt{\frac{\Sigma P_1Q_1}{\Sigma P_0Q_0} \times \frac{\Sigma P_1Q_1}{\Sigma P_0Q_0}}$$

$$= \frac{\Sigma P_1Q_1}{\Sigma P_0Q_0}$$

Thus Fisher's Index satisfies Factor Reversal test. Because Fisher's Index number satisfies both the tests in (ii) and (iii), it is called an Ideal Index Number.

(iv) Circular Test: It is concerned with the measurement of price changes over a period of years, when it is desirable to shift the base. For example, if the 1970 index with base 1965 is 200 and 1965 index with base 1960 is 150, the index 1970 on base 1960 will be 300. This property therefore enables us to adjust the index values from period to period without referring each time to the original base. The test of this shiftability of base is called the circular test.

This test is not met by Laspeyres, or Paasche's or the Fisher's ideal index. **The simple geometric mean of price relatives and the weighted aggregative with fixed weights meet this test.**

Example: Compute Fisher's Ideal Index from the following data:

	Base Year		*Current Year*	
Commodities	*Price*	*Quantity*	*Price*	*Quantity*
A	4	3	6	2
B	5	4	6	4
C	7	2	9	2
D	2	3	1	5

Show how it satisfies the time and factor reversal tests.

Solution:

Commodities	P_0	Q_0	P_1	Q_1	P_0Q_0	P_1Q_0	P_0Q_1	P_1Q_1
A	4	3	6	2	12	18	8	12
B	5	4	6	4	20	24	20	24
C	7	2	9	2	14	18	14	18
D	2	3	1	5	6	3	10	5
					52	63	52	59

Fisher's Ideal Index: $$P_{01} = \sqrt{\frac{\Sigma P_1Q_0}{\Sigma P_0Q_0} \times \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1}} \times 100 = \sqrt{\frac{63}{52} \times \frac{59}{52}} \times 100$$

$$= \sqrt{1.375} \times 100 = 1.172 \times 100 = 117.3$$

Time Reversal Test:

$$P_{01} \times P_{10} = \sqrt{\frac{63}{52} \times \frac{59}{52} \times \frac{52}{59} \times \frac{52}{63}} = \sqrt{1} = 1$$

∴ Time Reversal Test is satisfied.

Factor Reversal Test:

$$P_{01} \times Q_{01} = \sqrt{\frac{63}{52} \times \frac{59}{52} \times \frac{52}{59} \times \frac{52}{63}} = \sqrt{\frac{59}{52} \times \frac{59}{52}} = \frac{59}{52}$$

Since, $\frac{\Sigma P_1Q_1}{\Sigma P_0Q_0}$ is also equal to $\frac{59}{52}$, the Factor Reversal Test is satisfied.

SUMMARY

- An index number is a ratio or an average of ratios expressed as a percentage, Two or more time periods are involved, one of which is the base time period.
- Issues Involved in index numbers

  (a) Selection of Data
  (b) Base period
  (c) Selection of Weights:
  (d) Use of Averages:
  (e) Choice of Variables
- Construction of Index Number

  Price Index numbers

  (a) Simple aggregative price index = $\frac{\sum P_n}{\sum P_o} \times 100$

  (b) Laspeyres' Index: In this Index base year quantities are used as weights:

  Laspeyres Index = $\frac{\sum P_n Q_n}{\sum P_o Q_o} \times 100$

  (c) Paasche's Index: In this Index current year quantities are used as weights:

  Passche's Index = $\frac{\sum P_n Q_n}{\sum P_o Q_n} \times 100$

  (d) The Marshall-Edgeworth index uses this method by taking the average of the base year and the current year

  Marshall-Edgeworth Index $= \frac{\sum P_n (Q_o + Q_n)}{\sum P_o (Q_o + Q_n)} \times 100$

  (e) Fisher's ideal Price Index: This index is the geometric mean of Laspeyres' and Paasche's.

  Fisher's Index = $\sqrt{\frac{\sum P_n Q_o}{\sum P_o Q_o} \times \frac{\sum P_n Q_n}{\sum P_o Q_n)}} \times 100$

  (g) Weighted Average of Relative Method: $\frac{\sum \frac{P_n}{P_o} \times (P_o Q_o)}{\sum P_o Q_o} \times 100 = \frac{\sum P_n Q_o}{\sum P_o Q_o} \times 100$

  (h) Chain Index = $\frac{\text{Link relative of current year} \times \text{Chain Index of the previous year}}{100}$

Quantity Index Numbers

- Simple aggregate of quantities: $\frac{\sum Q_n}{\sum Q_o}$

- The simple average of quantity relatives: $\frac{\frac{\sum Q_n}{\sum Q_o}}{N}$

- Weighted aggregate quantity indices:

  (i) With base year weight : $\frac{\sum Q_n P_o}{\sum Q_o P_o}$ (Laspeyre's index)

  (ii) With current year weight : $\frac{\sum Q_n P_n}{\sum Q_o P_n}$ (Paasche's index)

  (iii) Geometric mean of (i) and (ii) : $\sqrt{\frac{\sum Q_n P_o}{\sum Q_o P_o} \times \frac{\sum Q_n P_n}{\sum Q_o P_n}}$ (Fisher's Ideal)

- Base-year weighted average of quantity relatives. This has the formula $\frac{\sum\left(\frac{Q_n}{Q_o} P_o Q_o\right)}{\sum P_o Q_o}$

- Value Indices

  $\frac{V_n}{V_o} = \frac{\sum P_n Q_n}{\sum P_o Q_o}$

- Deflated Value = $\frac{\text{Current Value}}{\text{Price Index of the current year}}$

  or Current Value x $\frac{\text{Base Price } (P_0)}{\text{Current Price } (P_n)}$ Base Price (P_0)

- Shifted Price Index = $\frac{\text{Original Price Index}}{\text{Price Index of the year on which it has to be shifted}} \times 100$

- Test of Adequacy

  (1) Unit test
  (2) Time reversal Test
  (3) Factor reversal test
  (4) Circular Test

EXERCISE

Choose the most appropriate option (a) (b) (c) or (d).

1. A series of numerical figures which show the relative position is called
   a) index number b) relative number c) absolute number d) none
2. Index number for the base period is always taken as
   a) 200 b) 50 c) 1 d) 100
3. ________ play a very important part in the construction of index numbers.
   a) weights b) classes c) estimations d) none
4. ________ is particularly suitable for the construction of index numbers.
   a) H.M. b) A.M. c) G.M. d) none
5. Index numbers show ________ changes rather than absolute amounts of change.
   a) relative b) percentage c) both d) none
6. The ________ makes index numbers time-reversible.
   a) A.M. b) G.M. c) H.M. d) none
7. Price relative is equal to
   a) $\dfrac{\text{Price in the given year} \times 100}{\text{Price in the base year}}$ b) $\dfrac{\text{Price in the year base year} \times 100}{\text{Price in the given year}}$
   c) Price in the given year × 100 d) Price in the base year × 100
8. Index number is equal to
   a) sum of price relatives b) average of the price relatives
   c) product of price relative d) none
9. The ________ of group indices given the General Index
   a) H.M. b) G.M. c) A.M. d) none
10. Circular Test is one of the tests of
    a) index numbers b) hypothesis c) both d) none
11. ___________ is an extension of time reversal test
    a) Factor Reversal test b) Circular test
    c) both d) none
12. Weighted G.M. of relative formula satisfy ________test
    a) Time Reversal Test b) Circular test
    c) Factor Reversal Test d) none
13. Factor Reversal test is satisfied by
    a) Fisher's Ideal Index b) Laspeyres Index
    c) Paasches Index d) none

14. Laspeyre's formula does not satisfy

a) Factor Reversal Test b) Time Reversal Test

c) Circular Test d) all the above

15. A ratio or an average of ratios expressed as a percentage is called

a) a relative number b) an absolute number

c) an index number d) none

16. The value at the base time period serves as the standard point of comparison

a) false b) true c) both d) none

17. An index time series is a list of _______ numbers for two or more periods of time

a) index b) absolute c) relative d) none

18. Index numbers are often constructed from the

a) frequency b) class c) sample d) none

19. ________ is a point of reference in comparing various data describing individual behaviour.

a) Sample b) Base period c) Estimation d) none

20. The ratio of price of single commodity in a given period to its price in the preceding year price is called the

(a) base period (b) price ratio (c) relative price (d) none

21. $\frac{\text{Sum of all commodity prices in the current year} \times 100}{\text{Sum of all commodity prices in the base year}}$ is

(a) Relative Price Index (b) Simple Aggregative Price Index

(c) both (d) none

22. Chain index is equal to

(a) $\frac{\text{link relative of current year} \times \text{chain index of the current year}}{100}$

(b) $\frac{\text{link relative of previous year} \times \text{chain index of the current year}}{100}$

(c) $\frac{\text{link relative of current year} \times \text{chain index of the previous year}}{100}$

(d) $\frac{\text{link relative of previous year} \times \text{chain index of the previous year}}{100}$

23. P_{01} is the index for time

(a) 1 on 0 (b) 0 on 1 (c) 1 on 1 (d) 0 on 0

24. P_{10} is the index for time

(a) 1 on 0 (b) 0 on 1 (c) 1 on 1 (d) 0 on 0

25. When the product of price index and the quantity index is equal to the corresponding value index then the test that holds is

(a) Unit Test (b) Time Reversal Test

(c) Factor Reversal Test (d) none holds

26. The formula should be independent of the unit in which or for which price and quantities are quoted in

(a) Unit Test (b) Time Reversal Test

(c) Factor Reversal Test (d) none

27. Laspeyre's method and Paasche's method do not satisfy

(a) Unit Test (b) Time Reversal Test

(c) Factor Reversal Test (d) b & c

28. The purpose determines the type of index number to use

(a) yes (b) no (c) may be (d) may not be

29. The index number is a special type of average

(a) false (b) true (c) both (d) none

30. The choice of suitable base period is at best temporary solution

(a) true (b) false (c) both (d) none

31. Fisher's Ideal Formula for calculating index numbers satisfies the _______ tests

(a) Unit Test (b) Factor Reversal Test

(c) both (d) none

32. Fisher's Ideal Formula dose not satisfy _________ test

(a) Unit Test (b) Circular Test (c) Time Reversal Test (d) none

33. _________________ satisfies circular test

a) G.M. of price relatives or the weighted aggregate with fixed weights

b) A.M. of price relatives or the weighted aggregate with fixed weights

c) H.M. of price relatives or the weighted aggregate with fixed weights

d) none

34. Laspeyre's and Paasche's method _________ time reversal test

(a) satisfy (b) do not satisfy (c) are (d) are not

35. There is no such thing as unweighted index numbers

(a) false (b) true (c) both (d) none

36. Theoretically, G.M. is the best average in the construction of index numbers but in practice, mostly the A.M. is used

(a) false (b) true (c) both (d) none

37. Laspeyre's or Paasche's or the Fisher's ideal index do not satisfy

(a) Time Reversal Test (b) Unit Test
(c) Circular Test (d) none

38. ____________ is concerned with the measurement of price changes over a period of years, when it is desirable to shift the base

(a) Unit Test (b) Circular Test
(c) Time Reversal Test (d) none

39. The test of shifting the base is called

(a) Unit Test (b) Time Reversal Test
(c) Circular Test (d) none

40. The formula for conversion to current value

(a) Deflated value = $\frac{\text{Price Index of the current year}}{\text{previous value}}$

(b) Deflated value = $\frac{\text{Price Index of the current year}}{\text{current value}}$

(c) Deflated value = $\frac{\text{Price Index of the previous year}}{\text{previous value}}$

(d) Deflated value = $\frac{\text{Price Index of the previous year}}{\text{previous value}}$

41. Shifted price Index = $\frac{\text{Original Price} \times 100}{\text{Price Index of the year on which it has to be shifted}}$

(a) True (b) false (c) both (d) none

42. The number of test of Adequacy is

(a) 2 (b) 5 (c) 3 (d) 4

43. We use price index numbers

(a) To measure and compare prices (b) to measure prices
(c) to compare prices (d) none

44. Simple aggregate of quantities is a type of

(a) Quantity control (b) Quantity indices
(c) both (d) none

ANSWERS

Exercise

1. (a)	**2.** (d)	**3.** (a)	**4.** (c)	**5.** (b)	**6.** (b)	**7.** (a)	**8.** (b)
9. (c)	**10.** (a)	**11.** (b)	**12.** (a)	**13.** (a)	**14.** (d)	**15.** (c)	**16.** (b)
17. (a)	**18.** (c)	**19.** (b)	**20.** (c)	**21.** (b)	**22.** (c)	**23.** (a)	**24.** (b)
25. (c)	**26.** (a)	**27.** (d)	**28.** (a)	**29.** (b)	**30.** (a)	**31.** (c)	**32.** (b)
33. (a)	**34.** (b)	**35.** (a)	**36.** (b)	**37.** (c)	**38.** (b)	**39.** (c)	**40.** (b)
41. (a)	**42.** (d)	**43.** (a)	**44.** (b)				

ADDITIONAL QUESTION BANK

1. Each of the following statements is either True or False write your choice of the answer by writing T for True

   (a) Index Numbers are the signs and guideposts along the business highway that indicate to the businessman how he should drive or manage.

   (b) "For Construction index number, the best method on theoretical ground is not the best method from practical point of view".

   (c) Weighting index numbers makes them less representative.

   (d) Fisher's index number is not an ideal index number.

2. Each of the following statements is either True or False. Write your choice of the answer by writing F for false.

   (a) Geometric mean is the most appropriate average to be used for constructing an index number.

   (b) Weighted average of relatives and weighted aggregative methods render the same result.

   (c) "Fisher's Ideal Index Number is a compromise between two well known indices – not a right compromise, economically speaking".

   (d) "Like all statistical tools, index numbers must be used with great caution".

3. The best average for constructing an index numbers is

   (a) Arithmetic Mean (b) Harmonic Mean

   (c) Geometric Mean (d) None of these.

4. The time reversal test is satisfied by

   (a) Fisher's index number. (b) Paasche's index number.

   (c) Laspeyre's index number. (d) None of these.

5. The factor reversal test is satisfied by

(a) Simple aggregative index number. (b) Paasche's index number.

(c) Laspeyre's index number. (d) None of these.

6. The circular test is satisfied by

(a) Fisher's index number. (b) Paasche's index number.

(c) Laspeyre's index number. (d) None of these.

7. Fisher's index number is based on

(a) The Arithmetic mean of Laspeyre's and Paasche's index numbers.

(b) The Median of Laspeyre's and Paasche's index numbers.

(c) the Mode of Laspeyre's and Paasche's index numbers.

(d) None of these.

8. Paasche index is based on

(a) Base year quantities. (b) Current year quantities.

(c) Average of current and base year. (d) None of these.

9. Fisher's ideal index number is

(a) The Median of Laspeyre's and Paasche's index numbers

(b) The Arithmetic Mean of Laspeyre's and Paasche's index numbers

(c) The Geometric Mean of Laspeyre's and Paasche's index numbers

(d) None of these.

10. Price-relative is expressed in term of

(a) $P = \frac{P_n}{P_o}$ (b) $P = \frac{P_o}{P_n}$

(c) $P = \frac{P_n}{P_o} \times 100$ (d) $P = \frac{P_o}{P_n} \times 100$

11. Paasehe's index number is expressed in terms of :

(a) $\frac{\Sigma P_n q_n}{\Sigma P_o q_n}$ (b) $\frac{\Sigma P_o q_o}{\Sigma P_n q_n}$

(c) $\frac{\Sigma P_n q_n}{\Sigma P_o q_n} \times 100$ (d) $\frac{\Sigma P_n q_o}{\Sigma P_o q_o} \times 100$

12. Time reversal Test is satisfied by following index number formula is

(a) Laspeyre's Index number.

(b) Simple Arithmetic Mean of price relative formula

(c) Marshall-Edge worth formula.

(d) None of these.

13. Cost of Living Index number (C. L. I.) is expressed in terms of :

(a) $\frac{\sum P_n q_o}{\sum P_o q_o} \times 100$ (b) $\frac{\sum P_n q_n}{\sum P_o q_o}$

(c) $\frac{\sum P_o q_n}{\sum P_n q_n} \times 100$ (d) None of these.

14. If the ratio between Laspeyre's index number and Paasche's Index number is 28 : 27. Then the missing figure in the following table P is :

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
X	L	10	2	5
Y	L	5	P	2

(a) 7 (b) 4 (c) 3 (d) 9

15. If the prices of all commodities in a place have increased 1.25 times in comparison to the base period, the index number of prices of that place now is

(a) 125 (b) 150 (c) 225 (d) None of these.

16. If the index number of prices at a place in 1994 is 250 with 1984 as base year, then the prices have increased on average by

(a) 250% (b) 150% (c) 350% (d) None of these.

17. If the prices of all commodities in a place have decreased 35% over the base period prices, then the index number of prices of that place is now

(a) 35 (b) 135 (c) 65 (d) None of these.

18. Link relative index number is expressed for period n is

(a) $\frac{P_n}{P_{n+1}}$ (b) $\frac{P_0}{P_{n-1}}$

(c) $\frac{P_n}{P_{n-1}} \times 100$ (d) None of these.

19. Fisher's Ideal Index number is expressed in terms of :

(a) $(P_{on})^F = \sqrt{\text{Laspeyre's Index} \times (\text{Paasche's Index})}$

(b) $(P_{on})^F$ = Laspeyre's Index X Paasche's Index

(c) $(P_{on})^F = \sqrt{\text{Marshall Edge worth Index} \times \text{Paasche's}}$

(d) None of these.

20. Factor Reversal Test According to Fisher is $P_{01} \times Q_{01}$=

(a) $\frac{\sum P_o q_o}{\sum P_n q_n}$ (b) $\frac{\sum P_n q_n}{\sum P_o q_o}$

(c) $\frac{\sum P_o q_n}{\sum P_n q_n}$ (d) None of these.

21. Marshall-edge worth Index formula after interchange of p and q is expressed in terms of :

(a) $\frac{\sum q_n (p_0 + p_n)}{\sum q_0 (p_0 + p_n)}$ (b) $\frac{\sum P_n (q_0 + q_n)}{\sum q P_0 (q_0 + q_n)}$

(c) $\frac{\sum P_0 (q_0 + q_n)}{\sum P_n (P_0 + P_n)}$ (d) None of these.

22. If $\sum P_n q_n = 249$, $\sum P_o q_o = 150$, Paasche's Index Number = 150 and Drobiseh and Bowely's Index number = 145, then the Fisher's Ideal Index Number is

(a) 75 (b) 60 (c) 145.97 (d) None of these.

23. Consumer Price index number for the year 1957 was 313 with 1940 as the base year 96 the Average Monthly wages in 1957 of the workers into factory be ₹ 160/- their real wages is

(a) ₹ 48.40 (b) ₹ 51.12 (c) ₹ 40.30 (d) None of these.

24. If $\sum P_o q_o = 3500$, $\sum P_n q_o = 3850$, then the Cost of living Index (C.L.I.) for 1950 w.r. to base 1960 is

(a) 110 (b) 90 (c) 100 (d) None of these.

25. From the following table by the method of relatives using Arithmetic mean the price Index number is

Commodity	Wheat	Milk	Fish	Sugar
Base Price	5	8	25	6
Current Price	7	10	32	12

(a) 140.35 (b) 148.95 (c) 140.75 (d) None of these.

From the Q.No. 26 to 29 each of the following statements is either True or False with your choice of the answer by writing F for False.

26. (a) Base year quantities are taken as weights in Laspeyre's price Index number.

(b) Fisher's ideal index is equal to the Arithmetic mean of Laspeyre's and Paasche's index numbers.

(c) Laspeyre's index number formula does not satisfy time reversal test.

(d) None of these.

27. (a) Current year quantities are taken as weights in Paasche's price index number.

(b) Edge worth Marshall's index number formula satisfies Time, Reversal Test.

(c) The Arithmetic mean of Laspeyre's and Paasche's index numbers is called Bowely's index numbers.

(d) None of these.

28. (a) Current year prices are taken as weights in Paasche's quantity index number.

(b) Fisher's Ideal Index formula satisfies factor Reversal Test.

(c) The sum of the quantities of the base period and current period is taken as weights in Laspeyre's index number.

(d) None of these.

29. (a) Simple Aggregative and simple Geometric mean of price relatives formula satisfy circular Test.

(b) Base year prices are taken as weights in Laspeyre's quantity index numbers.

(c) Fisher's Ideal Index formula obeys time reversal and factor reversal tests.

(d) None of these.

30. In 1980, the net monthly income of the employee was ₹ 800/- p. m. The consumer price index number was 160 in 1980. It rises to 200 in 1984. If he has to be rightly compensated. The additional D. A. to be paid to the employee is

(a) ₹ 175/- (b) ₹ 185/- (c) ₹ 200/- (d) ₹ 125.

31. The simple Aggregative formula and weighted aggregative formula satisfy is

(a) Factor Reversal Test (b) Circular Test

(c) Unit Test (d) None of these.

32. "Fisher's Ideal Index is the only formula which satisfies"

(a) Time Reversal Test (b) Circular Test

(c) Factor Reversal Test (d) a & c.

33. "Neither Laspeyre's formula nor Paasche's formula obeys" :

(a) Time Reversal and factor Reversal Tests of index numbers.

(b) Unit Test and circular Tests of index number.

(c) Time Reversal and Unit Test of index number.

(d) None of these.

34. Bowley's index number is 150. Fisher's index number is 149.95. Paasche's index number is

(a) 158 (b) 154 (c) 148 (d) 156

35. With the base year 1960 the C. L. I. in 1972 stood at 250. x was getting a monthly Salary of ₹ 500 in 1960 and ₹ 750 in 1972. In 1972 to maintain his standard of living in 1960 x has to receive as extra allowances of

(a) ₹ 600/- (b) ₹ 500/- (c) ₹ 300/- (d) none of these.

36. From the following data base year :-

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	4	3	6	2
B	5	4	6	4
C	7	2	9	2
D	2	3	1	5

Fisher's Ideal Index is

(a) 117.3 (b) 115.43 (c) 118.35 (d) 116.48

37. Which statement is False?

(a) The choice of suitable base period is at best a temporary solution.

(b) The index number is a special type of average.

(c) Those is no such thing as unweighted index numbers.

(d) Theoretically, geometric mean is the best average in the construction of index numbers but in practice, mostly the arithmetic mean is used.

38. Factor Reversal Test is expressed in terms of

(a) $\frac{\Sigma P_1Q_1}{\Sigma P_0Q_0}$ (b) $\frac{\Sigma P_1Q_1}{\Sigma P_0Q_0} \times \frac{\Sigma P_1Q_1}{\Sigma P_0Q_1}$

(c) $\frac{\Sigma P_1Q_1}{\Sigma Q_0P_1}$ (d) $\frac{\Sigma Q_1P_0}{\Sigma Q_0P_0} \times \frac{\Sigma P_1Q_1}{\Sigma Q_0P_1}$

39. Circular Test is satisfied by

(a) Laspeyre's Index number.

(b) Paasche's Index number

(c) The simple geometric mean of price relatives and the weighted aggregative with fixed weights.

(d) None of these.

40. From the following data for the 5 groups combined

Group	Weight	Index Number
Food	35	425
Cloth	15	235
Power & Fuel	20	215
Rent & Rates	8	115
Miscellaneous	22	150

The general Index number is

(a) 270 (b) 269.2 (c) 268.5 (d) 272.5

41. From the following data with 1966 as base year

Commodity	Quantity Units	Values (₹)
A	100	500
B	80	320
C	60	150
D	30	360

The price per unit of commodity A in 1966 is

(a) ₹ 5 (b) ₹ 6 (c) ₹ 4 (d) ₹ 12

42. The index number in whole sale prices is 152 for August 1999 compared to August 1998. During the year there is net increase in prices of whole sale commodities to the extent of

(a) 45% (b) 35% (c) 52% (d) 48%

43. The value Index is expressed in terms of

(a) $\frac{\sum P_1Q_1}{\sum P_0Q_0} \times 100$ (b) $\frac{\sum P_1Q_1}{\sum P_0Q_0}$

(c) $\frac{\sum P_0Q_0}{\sum P_1Q_1} \times 100$ (d) $\frac{\sum P_0Q_1 \times \sum P_1Q_1}{\sum P_0Q_0 \times \sum P_1Q_0}$

44. Purchasing Power of Money is

(a) Reciprocal of price index number. (b) Equal to price index number.

(c) Unequal to price index number. (d) None of these.

45. The price level of a country in a certain year has increased 25% over the base period. The index number is

(a) 25 (b) 125 (c) 225 (d) 2500

46. The index number of prices at a place in 1998 is 355 with 1991 as base. This means
    (a) There has been on the average a 255% increase in prices.
    (b) There has been on the average a 355% increase in price.
    (c) There has been on the average a 250% increase in price.
    (d) None of these.
47. If the price of all commodities in a place have increased 125 times in comparison to the base period prices, then the index number of prices for the place is now
    (a) 100 (b) 125 (c) 225 (d) None of the above.
48. The wholesale price index number or agricultural commodities in a given region at a given date is 280. The percentage increase in prices of agricultural commodities over the base year is :
    (a) 380 (b) 280 (c) 180 (d) 80
49. If now the prices of all the commodities in a place have been decreased by 85% over the base period prices, then the index number of prices for the place is now (index number of prices of base period = 100)
    (a) 100 (b) 135 (c) 65 (d) None of these.
50. From the data given below

Commodity	Price Relative	Weight
A	125	5
B	67	2
C	250	3

Then the suitable index number is
    (a) 150.9 (b) 155.8 (c) 145.8 (d) None of these.

51. Bowley's Index number is expressed in the form of :
    (a) $\frac{\text{Laspeyre's index + Paasche's index}}{2}$ (b) $\frac{\text{Laspeyre's index} \times \text{Paasche's index}}{2}$
    (c) $\frac{\text{Laspeyre's index - Paasche's index}}{2}$ (d) None of these.
52. From the following data

Commodity	Base Price	Current Pricet
Rice	35	42
Wheat	30	35
Pulse	40	38
Fish	107	120

The simple Aggregative Index is

(a) 115.8 (b) 110.8 (c) 112.5 (d) 113.4

53. With regard to Laspeyre's and Paasche's price index numbers, it is maintained that "If the prices of all the goods change in the same ratio, the two indices will be equal for them the weighting system is irrelevant; or if the quantities of all the goods change in the same ratio, they will be equal, for them the two weighting systems are the same relatively". Then the above statements satisfy.

(a) Laspeyre's Price index ≠ Paasche's Price Index.

(b) Laspeyre's Price Index = Paasche's Price Index.

(c) Laspeyre's Price Index may be equal Paasche's Price Index.

(d) None of these.

54. The quantity Index number using Fisher's formula satisfies :

(a) Unit Test (b) Factor Reversal Test.
(c) Circular Test. (d) Time Reversal Test.

55. For constructing consumer price Index is used :

(a) Marshall Edge worth Method. (b) Paasche's Method.
(c) Dorbish and Bowley's Method. (d) Laspeyre's Method.

56. The cost of living Index (C.L.I.) is always :

(a) Weighted index (b) Price Index.
(c) Quantity Index. (d) None of these.

57. The Time Reversal Test is not satisfied to :

(a) Fisher's ideal Index. (b) Marshall Edge worth Method.
(c) Laspeyre's and Paasche Method. (d) None of these.

58. Given below are the data on prices of some consumer goods and the weights attached to the various items Compute price index number for the year 1985 (Base 1984 = 100)

Items	Unit	1984	1985	Weight
Wheat	Kg.	0.50	0.75	2
Milk	Litre	0.60	0.75	5
Egg	Dozen	2.00	2.40	4
Sugar	Kg.	1.80	2.10	8
Shoes	Pair	8.00	10.00	1

Then weighted average of price Relative Index is :

(a) 125.43 (b) 123.3 (c) 124.53 (d) 124.52

59. The Factor Reversal Test is as represented symbolically is :

(a) $P_{01} \times Q_{01} = \frac{\Sigma P_1Q_1}{\Sigma P_0Q_0}$ (b) $I_{01} \times I_{10}$

(c) $\frac{\Sigma P_0Q_0}{\Sigma P_1Q_1}$ (d) $\sqrt{\frac{\Sigma P_1Q_1}{\Sigma P_0Q_0} \times \frac{\Sigma P_0Q_1}{\Sigma Q_{10}P_0}}$

60. If the 1970 index with base 1965 is 200 and 1965 index with base 1960 is 150, the index 1970 on base 1960 will be :

(a) 700 (b) 300 (c) 500 (d) 600

61. Circular Test is not met by :

(a) The simple Geometric mean of price relatives.

(b) The weighted aggregative with fixed weights.

(c) Laspeyre's or Paasche's or the fisher's Ideal index.

(d) None of these.

62. From the following data

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	4	3	6	2
B	5	4	6	4
C	7	2	9	2
D	2	3	1	5

Then the value ratio is:

(a) $\frac{59}{52}$ (b) $\frac{49}{47}$ (c) $\frac{41}{53}$ (d) $\frac{47}{53}$

63. The value index is equal to :

(a) The total sum of the values of a given year multiplied by the sum of the values of the base year.

(b) The total sum of the values of a given year Divided by the sum of the values of the base year.

(c) The total sum of the values of a given year plus by the sum of the values of the base year.

(d) None of these.

64. Time Reversal Test is represented symbolically by :

(a) $P_{01} \times P_{10}$

(b) $P_{01} \times P_{10} = 1$

(c) $P_{01} \times P_{10} \neq 1$

(d) None of these.

65. In 1996 the average price of a commodity was 20% more than in 1995 but 20% less than in 1994; and more over it was 50% more than in 1997 to price relatives using 1995 as base (1995 price relative 100) Reduce the data is :

(a) 150, 100, 120, 80 for (1994–97)

(b) 135, 100, 125, 87 for (1994–97)

(c) 140, 100, 120, 80 for (1994–97)

(d) None of these.

66. From the following data

Commodities	Base Year 1922 Price (₹)	Current Year 1934 Price
A	6	10
B	2	2
C	4	6
D	11	12
E	8	12

The price index number for the year 1934 is :

(a) 140 (b) 145 (c) 147 (d) None of these.

67. From the following data

Commodities	Base Price 1964	Current Price 1968
Rice	36	54
Pulse	30	50
Fish	130	155
Potato	40	35
Oil	110	110

The index number by unweighted methods :

(a) 116.8 (b) 117.25 (c) 115.35 (d) 119.37

68. The Bowley's Price index number is represented in terms of :

(a) A.M. of Laspeyre's and Paasche's Price index number.

(b) G.M. of Laspeyre's and Paasche's Price index number.

(c) A.M. of Laspeyre's and Walsh's price index number.

(d) None of these.

69. Fisher's price index number equal is :

(a) G.M. of Kelly's price index number and Paasche's price index number.

(b) G.M. of Laspeyre's and Paasche's Price index number.

(c) G.M. of Bowley's price index number and Paasche's price index number.

(d) None of these.

70. The price index number using simple G.M. of the n relatives is given by :

(a) $\log I_{on} = 2 - \frac{1}{n}\sum \log \frac{P_n}{P_o}$ (b) $\log I_{on} = 2 + \frac{1}{n}\sum \log \frac{P_n}{P_o}$

(c) $\log I_{on} = \frac{1}{2n}\sum \log \frac{P_n}{P_o}$ (d) None of these.

71. The price of a number of commodities are given below in the current year 1975 and base year 1970.

Commodities	A	B	C	D	E	F
Base Price	45	60	20	50	85	120
Current Price	55	70	30	75	90	130

For 1975 with base 1970 by the Method of price relatives using Geometrical mean, the price index is :

(a) 125.3 (b) 124.3 (c) 128.8 (d) None of these.

72. From the following data

Group	A	B	C	D	E	F
Group Index	120	132	98	115	108	95
Weight	6	3	4	2	1	4

The general Index I is given by :

(a) 111.3 (b) 113.45 (c) 117.25 (d) 114.75

73. The price of a commodity increases from ₹ 5 per unit in 1990 to ₹ 7.50 per unit in 1995 and the quantity consumed decreases from 120 units in 1990 to 90 units in 1995. The price and quantity in 1995 are 150% and 75% respectively of the corresponding price and quantity in 1990. Therefore, the product of the price ratio and quantity ratio is :

(a) 1.8 (b) 1.125 (c) 1.75 (d) None of these.

74. Test whether the index number due to Walsh give by :

$$I = \frac{\sum P_1\sqrt{Q_0Q_1}}{\sum P_0\sqrt{Q_0Q_1}} \times 100$$ Satisfies is :-

(a) Time reversal Test. (b) Factor reversal Test.

(c) Circular Test. (d) None of these.

75. From the following data

Group	Weight	Index Number Base : 1952–53 = 100
Food	50	241
Clothing	2	21
Fuel and Light	3	204
Rent	16	256
Miscellaneous	29	179

The Cost of living index numbers is :

(a) 224.5 (b) 223.91 (c) 225.32 (d) None of these.

76. Consumer price index number goes up from 110 to 200 and the Salary of a worker is also raised from ₹ 325 to ₹ 500. Therefore, in real terms, to maintain his previous standard of living he should get an additional amount of :

(a) ₹ 85 (b) ₹ 90.91 (c) ₹ 98.25 (d) None of these.

77. The prices of a commodity in the year 1975 and 1980 were 25 and 30 respectively taking 1980 as base year the price relative is :

(a) 109.78 (b) 110.25 (c) 113.25 (d) None of these.

78. The average price of certain commodities in 1980 was ₹ 60 and the average price of the same commodities in 1982 was ₹ 120. Therefore, the increase in 1982 on the basis of 1980 was 100%. 80. The decrease in 1980 with 1982 as base is: using 1982, comment on the above statement is :

(a) The price in 1980 decreases by 60% using 1982 as base.

(b) The price in 1980 decreases by 50% using 1982 as base.

(c) The price in 1980 decreases by 90% using 1982 as base.

(d) None of these.

79. Cost of Living Index (C.L.I.) numbers are also used to find real wages by the process of

(a) Deflating of Index number. (b) Splicing of Index number.

(c) Base shifting. (d) None of these.

80. From the following data

Commodities		A	B	C	D
1992 Base	Price	3	5	4	1
	Quantity	18	6	20	14
1993 Current Year	Price	4	5	6	3
	Quantity	15	9	26	15

The Passche price Index number is :

(a) 146.41 (b) 148.25 (c) 144.25 (d) None of these.

81. From the following data

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	7	17	13	25
B	6	23	7	25
C	11	14	13	15
D	4	10	8	8

The Marshall Edge Worth Index number is :

(a) 148.25 (b) 144.19 (c) 147.25 (d) None of these.

82. The circular Test is an extension of

(a) The time reversal Test. (b) The factor reversal Test.

(c) The unit Test. (d) None of these.

83. Circular test, an index constructed for the year 'x' on the base year 'y' and for the year 'y' on the base year 'z' should yield the same result as an index constructed for 'x' on base year 'z' i.e. $I_{01} \times I_{12} \times I_{20}$ equal is :

(a) 3 (b) 2 (c) 1 (d) None of these.

84. In 1976 the average price of a commodity was 20% more than that in 1975 but 20% less than that in 1974 and more over it was 50% more than that in 1977. The price relatives using 1975 as base year (1975 price relative = 100) then the reduce date is :

(a) 8,.75 (b) 150,80 (c) 75,125 (d) None of these.

85. Time Reversal Test is represented by symbolically is :

(a) $P_{01} \times Q_{01} = 1$ (b) $I_{01} \times I_{10} = 1$

(b) $I_{01} \times I_{12} \times I_{23} \times \ldots . I_{(n-1)n} \times I_{n0} = 1$ (d) None of these.

86. The prices of a commodity in the years 1975 and 1980 were 25 and 30 respectively, taking 1975 as base year the price relative is :

(a) 120 (b) 135 (c) 122 (d) None of these.

87. From the following data

Year	1992	1993	1995	1996	1997
Link Index	100	103	105	112	108

(Base 1992 = 100) for the years 1993–97. The construction of chain index is :

(a) 103, 100.94, 107, 118.72 (b) 103, 108.15, 121.3, 130.82

(c) 107, 100.25, 104, 118.72 (d) None of these.

88. During a certain period the cost of living index number goes up from 110 to 200 and the salary of a worker is also raised from ₹ 330 to ₹ 500. The worker does not get really gain. Then the real wages decreased by :

(a) ₹ 45.45 (b) ₹ 43.25 (c) ₹ 100 (d) None of these.

89. Net monthly salary of an employee was ₹ 3000 in 1980. The consumer price index number in 1985 is 250 with 1980 as base year. If the has to be rightly compensated then, 7th dearness allowances to be paid to the employee is :

(a) ₹ 4.800.00 (b) ₹ 4,700.00 (c) ₹ 4,500.0 (d) None of these.

90. Net Monthly income of an employee was ₹ 800 in 1980. The consumer price Index number was 160 in 1980. It is rises to 200 in 1984. If he has to be rightly compensated. The additional dearness allowance to be paid to the employee is :

(a) ₹ 240 (b) ₹ 275 (c) ₹ 250 (d) None of these.

91. When the cost of Tobacco was increased by 50%, a certain hardened smoker, who maintained his formal scale of consumption, said that the rise had increased his cost of living by 5%. Before the change in price, the percentage of his cost of living was due to buying Tobacco is

(a) 15% (b) 8% (c) 10% (d) None of these.

92. If the price index for the year, say 1960 be 110.3 and the price index for the year, say 1950 be 98.4, then the purchasing power of money (Rupees) of 1950 will in 1960 is

(a) ₹ 1.12 (b) ₹ 1.25 (c) ₹ 1.37 (d) None of these.

93. If $\sum P_0Q_0 = 1360$, $\sum PnQ_0 = 1900$, $\sum P_0Q_n = 1344$, $\sum P_nQ_n = 1880$ then the Laspeyre's Index number is

(a) 0.71 (b) 1.39 (c) 1.75 (d) None of these.

94. The consumer price Index for April 1985 was 125. The food price index was 120 and other items index was 135. The percentage of the total weight of the index is

(a) 66.67 (b) 68.28 (c) 90.25 (d) None of these.

95. The total value of retained imports into India in 1960 was ₹ 71.5 million per month. The corresponding total for 1967 was ₹ 87.6 million per month. The index of volume of retained imports in 1967 composed with 1960 (= 100) was 62.0. The price index for retained inputs for 1967 our 1960 as base is

(a) 198.61 (b) 197.61 (c) 198.25 (d) None of these.

96. During the certain period the C.L.I. goes up from 110 to 200 and the Salary of a worker is also raised from 330 to 500, then the real terms is

(a) Loss by ₹ 50 (b) Loss by 75 (c) Loss by ₹ 90 (d) None of these.

97. From the following data

Commodities	Q_0	P_0	Q_1	P_1
A	2	2	6	18
B	5	5	2	2
C	7	7	4	24

Then the fisher's quantity index number is

(a) 87.34 (b) 85.24 (c) 87.25 (d) None of these.

98. From the following data

Commodities	Base year	Current year
A	25	55
B	30	45

Then index numbers from G. M. Method is :

(a) 181.66 (b) 185.25 (c) 181.75 (d) None of these.

99. Using the following data

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
X	4	10	6	15
Y	6	15	4	20
Z	8	5	10	4

the Paasche's formula for index is :

(a) 125.38 (b) 147.25 (c) 129.8 (d) None of these.

100. Group index number is represented by

(a) $\frac{\text{Price Relative for the year}}{\text{Price Relative for the previous year}} \times 100$ (b) $\frac{\sum(\text{Price Relative} \times w)}{\sum w}$

(c) $\frac{\sum(\text{Price Relative} \times w)}{\sum w} \times 100$ (d) None of these.

ANSWERS

1.	(a)	**2.**	(c)	**3.**	(c)	**4.**	(a)	**5.**	(a)
6.	(d)	**7.**	(d)	**8.**	(b)	**9.**	(c)	**10.**	(c)
11.	(c)	**12.**	(c)	**13.**	(a)	**14.**	(b)	**15.**	(c)
16.	(b)	**17.**	(c)	**18.**	(c)	**19.**	(a)	**20.**	(b)
21.	(a)	**22.**	(d)	**23.**	(b)	**24.**	(a)	**25.**	(b)
26.	(b)	**27.**	(d)	**28.**	(c)	**29.**	(d)	**30.**	(c)
31.	(b)	**32.**	(d)	**33.**	(a)	**34.**	(a)	**35.**	(b)
36.	(a)	**37.**	(c)	**38.**	(a)	**39.**	(c)	**40.**	(b)
41.	(a)	**42.**	(c)	**43.**	(a)	**44.**	(a)	**45.**	(b)
46.	(a)	**47.**	(c)	**48.**	(c)	**49.**	(c)	**50.**	(a)
51.	(a)	**52.**	(b)	**53.**	(b)	**54.**	(d)	**55.**	(d)
56.	(a)	**57.**	(c)	**58.**	(b)	**59.**	(a)	**60.**	(b)
61.	(c)	**62.**	(a)	**63.**	(b)	**64.**	(b)	**65.**	(a)
66.	(d)	**67.**	(a)	**68.**	(a)	**69.**	(b)	**70.**	(b)
71.	(b)	**72.**	(a)	**73.**	(b)	**74.**	(a)	**75.**	(d)
76.	(b)	**77.**	(d)	**78.**	(b)	**79.**	(a)	**80.**	(a)
81.	(b)	**82.**	(a)	**83.**	(c)	**84.**	(b)	**85.**	(b)
86.	(a)	**87.**	(b)	**88.**	(c)	**89.**	(c)	**90.**	(a)
91.	(c)	**92.**	(a)	**93.**	(b)	**94.**	(a)	**95.**	(b)
96.	(a)	**97.**	(a)	**98.**	(a)	**99.**	(d)	**100.**	(b)

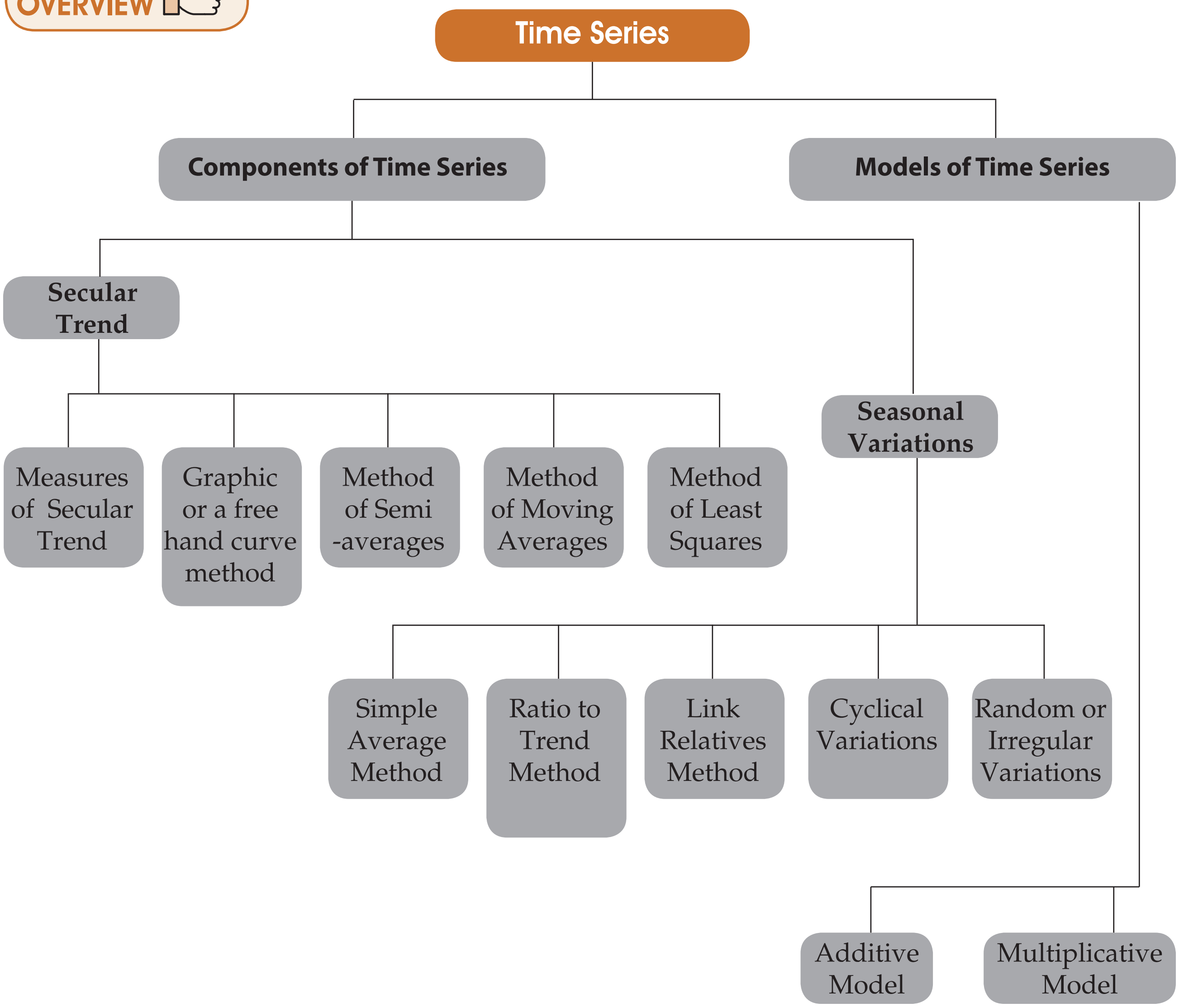
UNIT - II TIME SERIES

LEARNING OBJECTIVES

At the end of this Chapter, you will be able to:

- Understand the components of Time Series
- Calculate the trend using graph, moving averages
- Calculate Seasonal variations for both Additive and Multiplicative models

CHAPTER OVERVIEW



19.2.1 INTRODUCTION

We came across a data which is collected on a variable/s (rainfall, production of industrial product, production of rice, sugar cane, import/export of a country, population, etc.) at different time epochs (hours, days, months, years etc.), such a data is called time series data. Time series is statistical data that are arranged and presented in a chronological order i.e., over a period of time.

Most of the time series relating to Economic, Business and Commerce might show an upward tendency in case of population, production & sales of products, incomes, prices; or downward tendency might be noticed in time series relating to share prices, death, birth rate etc. due to global melt down, or improvement in medical facilities etc.

Definition: According to Spiegel, "A time series is a set of observations taken at specified times, usually at equal intervals."

According to Ya-Lun-Chou, "A time series may be defined as a collection of reading belonging to different time period of same economic variable or composite of variables."

Components of Time Series: There are various forces that affect the values of a phenomenon in a time series; these may be broadly divided into the following four categories, commonly known as the components of a time series.

(1) Long term movement or Secular Trend

(2) Seasonal variations

(3) Cyclical variations

(4) Random or irregular variations

In traditional or classical time series analysis, it is ordinarily assumed that there are:

1. Secular Trend or Simple trend:

Secular trend is the long: Term tendency of the time series to move in an upward or down ward direction. It indicates how on the whole, it has behaved over the entire period under reference. These are result of long-term forces that gradually operate on the time series variable. A general tendency of a variable to increase, decrease or remain constant in long term (though in a small time interval it may increase or decrease) is called trend of a variable. E.g. Population of a country has increasing trend over a years. Due to modern technology, agricultural and industrial production is increasing. Due to modern technology health facilities, death rate is decreasing and life expectancy is increasing. Secular trend is be long-term tendency of the time series to move on upward or downward direction. It indicates how on the whole behaved over the entire period under reference. These are result of long term forces that gradually operate on the time series variable. A few examples of theses long term forces which make a time series to move in any direction over long period of the time are long term changes per capita income, technological improvements of growth of population, Changes in Social norms etc.

Most of the time series relating to Economic, Business and Commerce might show an upward tendency in case of population, production & sales of products, incomes, prices; or downward tendency might be noticed in time series relating to share prices, death, birth rate etc. due to global melt down, or improvement in medical facilities etc. All these indicate trend.

2. Seasonal variations:

Over a span of one year, seasonal variation takes place due to the rhythmic forces which operate in a regular and periodic manner. These forces have the same or almost similar pattern year after year.

It is common knowledge that the value of many variables depends in part on the time of year. For Example, Seasonal variations could be seen and calculated if the data are recorded quarterly, monthly, weekly, daily or hourly basis. So if in a time series data only annual figures are given, there will be no seasonal variations.

The seasonal variations may be due to various seasons or weather conditions for example sale of cold drink would go up in summers & go down in winters. These variations may be also due to man-made conventions & due to habits, customs or traditions. For example, sales might go up during Diwali & Christmas or sales of restaurants & eateries might go down during Navratri's.

The method of seasonal variations are

(i) Simple Average Method

(ii) Ratio to Trend Method

(iii) Ratio to Moving Average Method

(iv) Link Relatives Method

3. Cyclical variations:

Cyclical variations, which are also generally termed as business cycles, are the periodic movements. These variations in a time series are due to ups & downs recurring after a period from Season to Season. Though they are more or less regular, they may not be uniformly periodic. These are oscillatory movements which are present in any business activity and is termed as business cycle. It has got four phases consisting of prosperity (boom), recession, depression and recovery. All these phases together may last from 7 to 9 years may be less or more.

4. Random or irregular variations:

These are irregular variations which occur on account of random external events. These variations either go very deep downward or too high upward to attain peaks abruptly. These fluctuations are a result of unforeseen and unpredictably forces which operate in absolutely random or erratic manner. They do not have any definite pattern and it cannot be predicted in advance. These variations are due to floods, wars, famines, earthquakes, strikes, lockouts, epidemics etc.

19.2.2 MODELS OF TIME SERIES

The following are the two models which are generally used for decomposition of time series into its four components. The objective is to estimate and separate the four types of variations and to bring out the relative impact of each on the overall behaviour of the time series.

(1) Additive model

(2) Multiplicative model

Additive Model: In additive model it is assumed that the four components are independent of one another i.e. the pattern of occurrence and magnitude of movements in any particular component does not affect and are not affected by the other component. Under this assumption the four components are arithmetically additive ie. magnitude of time series is the sum of the separate influences of its four components i.e. $Y_t = T + C + S + I$

Where,

Y_t = Time series

T = Trend variation

C = Cyclical variation

S = Seasonal variation

I = Random or irregular variation

Multiplicative Model: In this model it is assumed that the forces that give rise to four types of variations are interdependent, so that overall pattern of variations in the time series is a combined result of the interaction of all the forces operating on the time series. Accordingly, time series are the product of its four components i.e.

$Y_t = T \times C \times S \times I$

As regards to the choice between the two models, it is generally the multiplication model which is used more frequently. As the forces responsible for one type of variation are also responsible for other type of variations, hence it is multiplication model which is more suited in most business and economic time series data for the purpose of decomposition.

Example 19.2.1: Under the additive model, a monthly sale of ₹ 21,110 explained as follows:

The trend might be ₹ 20,000, the seasonal factor: ₹ 1,500 (The month question is a good one for sales, expected to be ₹ 1,500 over the trend), the cyclical factor: ₹ 800 (A general Business slump is experienced, expected to depress sales by ₹ 800 (per month); and Residual Factor: ₹ 410 (Due to unpredictable random fluctuations).

The model gives:

$Y = T + C + S + R$

$21,110 = 2,000 + 1,500 + (-800) + 410$

The multiplication model might explain the same sale figure in similar way.

Trend = ₹ 20,000, Seasonal Factor: ₹ 1.15 (a good month for sales, expected to be 15 per cent above the trend)

Cyclical Factor: 0.90 (a business slump, expected to cause a 10 per cent reduction in sales) and

Residual Factor: 1.02 (Random fluctuations of + 2 Factor)

$Y = T \times S \times C \times R$

$21114 = 20,000 \times 1.15 \times 0.90 \times 1.02$

19.2.3 MEASUREMENT OF SECULAR TREND

The following are the methods most commonly used for studying & measuring the trend component in a time series -

(1) Graphic or a Freehand Curve Method

(2) Method of Semi Averages

(3) Method of Moving Averages

(4) Method of Least Squares

Graphic or Freehand Curve Method:

The data of a given time series is plotted on a graph and all the points are joined together with a straight line. This curve would be irregular as it includes short run oscillation. These irregularities are smoothened out by drawing a freehand curve or line along with the curve previously drawn.

This curve would eliminate the short run oscillations and would show the long period general tendency of the data. While drawing this curve it should be kept in mind that the curve should be smooth and the number of points above the trend curve should be more or less equal to the number of points below it.

Merits:

(1) It is very simple and easy to understand

(2) It does not require any mathematical calculations

Disadvantages:

(1) This is a subjective concept. Hence different persons may draw freehand lines at different positions and with different slopes.

(2) If the length of period for which the curve is drawn is very small, it might give totally erroneous results.

Example 19.2.2: The following are figures of a Sale for the last nine years. Determine the trend by line by the freehand method.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Sale in lac Units	75	95	115	65	120	100	150	135	175

The trend line drawn by the freehand method can be extended to predicted values. However, since the freehand curve fitting is too subjective, the method should not be used as basis for predictions.

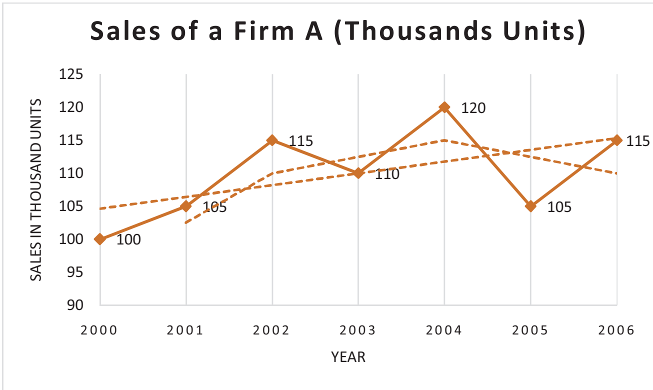
Methods of Semi averages: Under this method the whole time series data is classified into two equal parts and the averages for each half are calculated. If the data is for even number of years, it is easily divided into two. If the data is for odd number of years, then the middle year of the time series is left and the two halves are constituted with the period on each side of the middle year.

The arithmetic mean for a half is taken to be representative of the value corresponding to the midpoint of the time interval of that half. Thus we get two points. These two points are plotted on a graph and then are joined by straight line which is our required trend line.

Example 19.2.3: Fit a trend line to the following data by the method of Semi-averages.

Year	2000	2001	2002	2003	2004	2005	2006
Sale in lac Units	100	105	115	110	120	105	115

Solution: Since the data consist of seven Years, the middle year shall be left out and an average of the first three years and last three shall be obtained. The average of first three years is $\frac{100+105+115}{3}$ or $\frac{320}{3}$ or 106.67 and the average of last three years $\frac{120+105+115}{3}$ or $\frac{340}{3}$ or 133.33.



Moving average method: A moving average is an average (Arithmetic mean) of fixed number of items (known as periods) which moves through a series by dropping the first item of the previously averaged group and adding the next item in each successive average. The value so computed is considered the trend value for the unit of time falling at the centre of the period used in the calculation of the average.

In case the period is odd: If the period of moving average is odd for instance for computing 3 yearly moving average, the value of 1st, 2nd and 3rd years are added up and arithmetic mean is found out and the answer is placed against the 2nd year; then value of 2nd, 3rd and 4th years are added up and arithmetic mean is derived and this average is placed against 3rd year (i.e. the middle of 2nd, 3rd and 4th) and so on.

In case of even number of years: If the period of moving average is even for instance for computing 4 yearly moving average, the value of 1st, 2nd, 3rd and 4th years are added up& arithmetic mean is found out and answer is placed against the middle of 2nd and 3rd year. The second average is placed against middle of 3rd & 4th year. As this would not coincide with a period of a given time series an attempt is made to synchronise them with the original data by taking a two period average of the moving averages and placing them in between the corresponding time periods. This technique is called centring& the corresponding moving averages are called moving average centred.

Example 19.2.4: The wages of certain factory workers are given as below. Using 3 yearly moving average indicate the trend in wages.

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012
Wages	1200	1500	1400	1750	1800	1700	1600	1500	1750

Solution:

Table: Calculation of Trend Values by method of 3 yearly Moving Average

Year	Wages	3 yearly moving totals	3 yearly moving average i.e. trend
2004	1200	–	–
2005	1500	4100	1366.67
2006	1400	4650	1550
2007	1750	4950	1650
2008	1800	5250	1750
2009	1700	5100	1700
2010	1600	4800	1600
2011	1500	4850	1616.67
2012	1750	–	–

Example 19.2.5: Calculate 4 yearly moving average of the following data.

Year	2005	2006	2007	2008	2009	2010	2011	2012
Wages	1150	1250	1320	1400	1300	1320	1500	1700

Solution: (First Method):

Table: Calculation of 4 year Centred Moving Average

Year (1)	Wages (2)	4 yearly moving total (3)	2 year moving total of col. 3 (centred) (4)	4 yearly moving average centred (5) [col. 4/8]
2005	1,150	–	–	–
2006	1,250	–	–	–
		5,120		
2007	1,320		10,390	1,298.75
		5,270		
2008	1,400		10,610	1,326.25
		5,340		
2009	1,300		10,860	1,357.50
		5,520		
2010	1,320		11,340	1,417.50
		5,820		
2011	1,500			
2012	1,700			

Second Method:

Table: Calculation of 4 year Centred Moving Average

Year	Wages	4 yearly moving total (3)	4 yearly moving average (4)	2 year moving total of col. 4 (centered) (5)	4 year centered moving average (col. 5/2)
2005	1,150	–	–	–	–
2006	1,250	–	–	–	– –
		5,120	1,280	–	–
2007	1,320			2,597.75	1,298.75
		5,270	1,317.5	–	–
2008	1,400			2,652.5	1,326.25
		5,340	1,335	–	–
2009	1,300			2,715	1,357.50
		5,520	1,380		
2010	1,320			2,835	1,417.50
		5,820	1,455	–	–
2011	1,500				
2012	1,700				

Example 19.2.6: **Calculate five yearly moving averages for the following data.**

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Value	123	140	110	98	104	133	95	105	150	135

Solution:

Table: Computation of Five Yearly Moving Averages

Year	Value ('000 ₹)	5 yearly moving totals ('000 ₹)	5 yearly moving average ('000 ₹)
2003	123	–	–
2004	140	–	–
2005	110	575	115
2006	98	585	117
2007	104	540	108
2008	133	535	107
2009	95	587	117.4
2010	105	618	123.6
2011	150	–	–
2012	135	–	–

The method of least squares as studied in regression analysis can be used to find the trend line of best fit to a time series data.

The regression trend line (Y) is defined by the following equation – Y = a + bX where Y = predicted value of the dependent variable

a = Y axis intercept or the height of the line above origin (i.e. when X = 0, Y = a)

b = slope of the regression line (it gives the rate of change in Y for a given change in X) (when b is positive the slope is upwards, when b is negative, the slope is downwards) X = independent variable (which is time in this case)

To estimate the constants a and b, the following two equations have to be solved simultaneously –

$\sum Y = na + b\sum X$

$\sum XY = a\sum X + b\sum X^2$

To simplify the calculations, if the mid point of the time series is taken as origin, then the negative values in the first half of the series balance out the positive values in the second half so that $\sum x = 0$. In this case the above two normal equations will be as follows -

$\sum Y = na$

$\sum XY = b\sum X^2$

In such a case the values of a and b can be calculated as under -

Since $\sum Y = na$,

Since $\sum XY = b\sum X^2$

$$b = \frac{\sum XY}{\sum X^2}; a = \frac{\sum Y}{n}$$

Example 19.2.7: Fit a straight line trend to the following data by Least Square Method and estimate the sale for the year 2012.

Year	2005	2006	2007	2008	2009	2010
Sale (in'000s)	70	80	96	100	95	114

Solution:

Table: Calculation of trend line

Year	Sales Y	Deviations from 2007.5	Deviations multiplied by 2 (X)	X^2	XY
2005	70	-2.5	-5	25	-350
2006	80	-1.5	-3	9	-240
2007	96	-.5	-1	1	-96
2008	100	+.5	+1	1	100
2009	95	+1.5	+3	9	285
2010	114	+2.5	+5	25	570
	$\sum Y = 555$			$\sum X^2 = 70$	$\sum XY = 269$

N = 6

Equation of the straight line trend is $Y_o = a + bX$

$$a = \frac{\sum Y}{N} = \frac{555}{6} = 92.5 \; ; \; b = \frac{\sum XY}{\sum X^2} = \frac{269}{70} = 3.843$$

Trend equation is $Y_c = 92.5 + 3.843X$

For 2012, X = 9

$Y_{2012} = 92.5 + 3.843 \times 9 = 92.5 + 34.587$

= 126.59 (in '000 ₹)

Example 19.2.8:

Fit a straight line trend to the following data and estimate the likely profit for the year 2012. Also calculate the trend values.

Year	2003	2004	2005	2006	2007	2008	2009
Profit (in lakhs of ₹)	60	72	75	65	80	85	95

Solution:

Table: Calculation of Trend and Trend Values

Year	Profit Y	Deviation from 2006	X^2	XY	Trend Values ($Y_c = a + bX$)
		X			[Yc= 76 + 4.85X]
2003	60	-3	9	-180	76 + 4.85 (-3) = 61.45
2004	70	-2	4	-144	76 + 4.85 (-2) = 66.30
2005	75	-1	1	-75	76 + 4.85 (-1) = 70.15
2006	65	0	0	0	76 + 4.85 (0) = 76
2007	80	1	1	80	76 + 4.85(1) = 80.85
2008	85	2	4	170	76 + 4.85 (2) = 85.70
2009	95	3	9	285	76 + 4.85 (3) = 90.55
	$\sum y = 532$		$\sum X^2 = 28$	$\sum XY=136$	

N = 7

The equation for straight line trend is $Y_c = a + bX$

Where

$$a = \frac{\sum Y}{N} = \frac{555}{6} = 92.5$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{269}{70} = 3.843$$

The trend equation $Y_c = 92.5 + 3.843.X$

2012, x = 6 (2012 - 2006) $Y_c = 76 + 4.85(6) = 76 + 29.10$

= 105.10

The estimated profit for the year 2012 is ₹ 105.10 lakhs.

Example 19.2.9: Calculate Seasonal Indices for each quarter from the following percentages of whole sale price indices to their moving averages.

Year	Quarter			
	I	II	III	IV
2003	-	-	11.0	11.0
2004	12.5	13.5	15.5	14.5
2005	16.8	15.2	13.1	15.3
2006	11.2	11.0	12.4	13.2
2007	10.5	13.3	-	

Solution:

Year	Quarter			
	I	II	III	IV
Quarterly Total	51.0	53.0	52.0	54.0
Quarterly Average	12.75	13.25	13.0	13.5

Average of the Quarterly Averages = $\frac{52.5}{4} = 13.125$

Year	Quarter			
	I	II	III	IV
Seasonal Indices	$\frac{12.75 \times 100}{13.125} = 97.143$	$\frac{13.25 \times 100}{13.125} = 100.952$	$\frac{13.0 \times 100}{13.125} = 97.143$	$\frac{13.5 \times 100}{13.125} = 102.857$

Seasonal Indices are calculated by converting the respective quarterly averages on the basis that the average of the quarterly average = 100

Example 19.2.10: Calculate 5- year weighted moving averages for the following data, using weights 1, 1, 3, 2 respectively:

Year	1	2	3	4	5	6	7	8	9	10
Codded Sales	40	33	72	81	76	68	91	87	98	97

Solution :

Year I	Sales II	5- Year Weighted Average	IV
1	40	–	–
2	33	–	–
3	72	$\frac{1}{8}(40\times1+33\times1+72\times3+81\times2+76\times1)$	= 65.8
4	81	$\frac{1}{8}(33\times1+72\times1+81\times3+76\times2+68\times1)$	= 71.00
5	76	$\frac{1}{8}(72\times1+81\times1+76\times3+68\times2+91\times1)$	= 76.00
6	68	$\frac{1}{8}(81\times1+76\times1+68\times3+91\times2+87\times1)$	= 78.750
7	91	$\frac{1}{8}(76\times1+68\times1+91\times3+87\times2+90\times1)$	= 86.125
8	87	$\frac{1}{8}(68\times1+91\times1+87\times3+98\times2+97\times1)$	= 89.125
9	98	–	–
10	97	–	–

Example 19.2.11: Assuming no trend, calculate Seasonal variation indices for the following data.

Year	Quarterly Data			
	Q1	Q2	Q3	Q4
2013	3.7	4.1	3.3	3.5
2014	3.7	3.9	3.6	3.6
2015	4.0	4.1	3.3	3.1
2016	3.3	4.4	4.0	4.0

Solution:

	Quarterly Data			
	Q1	Q2	Q3	Q4
	3.7	4.1	3.3	3.5
	3.7	3.9	3.6	3.6
	4.0	4.1	3.3	3.1
	3.3	4.4	4.0	4.0
Quarterly Total	14.7	16.5	14.2	14.2
Quarterly Average	3.675	4.125	3.55	3.55

Average of Quarterly Averages = $\frac{3.675+4.125+3.55+3.55}{4}=\frac{14.9}{4}=3.725$

Example 19.2.12: Calculate the Seasonal Indices from the following ratio to moving averages values expressed in percentage.

Year	Seasons		
	Summer	Rain	Winter
2009	–	101.75	107.14
2010	96.18	92.30	114.00
2011	92.45	95.20	118.18

Solution:

Year	Seasons			
	Summer	Rain	Winter	
2009	–	101.75	107.14	
2010	96.18	92.30	114.00	
2011	92.45.	95.20	118.18	
Total	188.63	289.25	339.832	
Average	94.315	96.417	113.107	303.839
Constructed Seasonal Index	93.127	95.202	111.682	

Correction Factor = $\frac{300}{303.839}=0.9874$

Example 19.2.13: From the following data, calculate the trend values, using Four yearly moving average.

Years	2008	2009	2010	2011	2012	2013	2014	2015	2016
Values	506	620	1036	673	588	696	1116	738	663

Solution:

Year	Values	4 Yearly Moving Totals (a)	2 Period Moving Total of (a)	4 Yearly Moving Averages
2008	506	–	–	–
2009	620	–	–	–
2010	1036	2835	5752	719.0
2011	673	2917	5910	738.8
2012	588	2993	6066	758.3
2013	696	3073	6211	776.4
2014	1116	3138	6311	793.9
2015	738	3213	–	–
2016	663	–	–	–

SUMMARY

1) A time series is set of measurements on a variable taken over some period of time, it has four components.

   (a) Trend (b) Seasonal variations

   (c) Cyclical variations (d) Irregular variations

2) There are two models of time series

   (a) Additive Model (b) Multiplicative Model

3) Trends can be measured in the following measures

   (a) Free hand curve method (b) Semi-averages method

   (c) Moving averages method (d) Least squares method

4) A time series may be determined by eliminating the computed trend values from the given data set. It may done using additive model or multiplicative model.

5) Seasonal variations can be measured in any of the following methods:

   (a) simple averages (b) Ratio to trend method

   (c) Ratio to Moving averages (d) Link relative method

6) Time series data can be deseaonalised by eliminating the effect of seasonal variations from it.

7) Irregular component in a time series is measured as a residue after eliminating all other fluctuations from data.

8) Time Series is useful in forecasting future values.

Unit -II Exercise 1 (a)

1) (a) What is meant by Time Series? Explain its utility.

   (b) Explain clearly the meaning of Time Series Analysis in business.

2) What is a Time Series? What are the different components? Describe briefly each of them.

3) Write Short Notes on:

   (a) Moving average method of measuring Trend

   (b) Seasonal Index

4) Calculate five yearly moving averages for the following data.

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Value (‘000 ₹)	123	140	110	98	104	133	95	105	150	135

5) Write Short notes on:

   (i) Seasonal variations and (ii) Cyclical variations

6) From the following data verify that 5 year weighted moving average with weights 1, 2, 2, 2, 1 respectively is equivalent to the 4 year centred moving average:

(₹ In lakhs)

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sale	5	3	7	6	4	8	9	10	8	9	9

7) Explain the additive and multiplicative models of Time Series.

8) Calculate the Seasonal Indices by the method of Link Relatives for the following data.

Quarter	Quarterly Figures for Five years				
	2003	2004	2005	2006	2007
I	45	48	49	52	60
II	54	56	63	65	70
III	72	63	70	75	83
IV	60	56	65	72	86

9) Calculate the seasonal Indices for each quarter from the following percentages of whole sale prices indices to their moving averages.

Year	Quarter			
	I	II	III	IV
2003	–	–	11.0	80
2004	12.5	13.5	15.5	14.5
2005	16.8	15.2	13.1	15.3
2006	11.2	11.0	12.4	13.2
2007	10.5	13.3	–	–

10) Assuming no trend in the series, Calculate seasonal indices for the following data.

Year	Quarter			
	I	II	III	IV
2004	78	66	84	80
2005	76	74	82	78
2006	72	68	80	70
2007	74	70	84	74
2008	76	74	86	82

11) The annual production in commodity is given as follows:

Year	2000	2001	2002	2003	2004	2005	2006
Production: (in tonnes)	70	80	90	95	102	110	115

(a) Fit a straight line trend by the method of least squares.

(b) Convert the annual trend equation into monthly trend equation.

Unit -II Exercise 1(B)

Choose the most appropriate option (a) or (b) or (c) or (d).

1) An orderly set of data arranged in accordance with their time of occurrence is called:
   (a) Arithmetic series (b) Harmonic series
   (c) Geometric series (d) Time series
2) A time series consists of:
   (a) Short-term variations (b) Long-term variations
   (c) Irregular variations (d) All of the above
3) The graph of time series is called:
   (a) Histogram (b) Straight line
   (c) Histogram (d) Ogive
4) Secular trend can be measured by:
   (a) Two methods (b) Three methods
   (c) Four methods (d) Five methods
5) The secular trend is measured by the method of semi-averages when:
   (a) Time series based on yearly values
   (b) Trend is linear
   (c) Time series consists of even number of values
   (d) None of them
6) Increase in the number of patients in the hospital due to heat stroke is:
   (a) Secular trend (b) Irregular variation
   (c) Seasonal variation (d) Cyclical variation
7) The systematic components of time series which follow regular pattern of variations are called:
   (a) Signal (b) Noise
   (c) Additive model (d) Multiplicative model
8) The unsystematic sequence which follows irregular pattern of variations is called:
   (a) Noise (b) Signal
   (c) Linear (d) Non-linear
9) In time series seasonal variations can occur within a period of:
   (a) Four years (b) Three years
   (c) One year (d) Nine years
10) Wheat crops badly damaged on account of rains is:
   (a) Cyclical movement (b) Random movement

(c) Secular trend (d) Seasonal movement

11) The method of moving average is used to find the:
 (a) Secular trend (b) Seasonal variation
 (c) Cyclical variation (d) Irregular variation

12) Most frequency used mathematical model of a time series is:
 (a) Additive model (b) Mixed model
 (c) Multiplicative model (d) Regression

13) A time series consists of:
 (a) No mathematical model (b) One mathematical model
 (c) Two mathematical models (d) Three mathematical models

14) In semi-averages method, we decide the data into:
 (a) Two parts (b) Two equal parts
 (c) Three parts (d) Difficult to tell

15) Moving average method is used for measurement of trend when:
 (a) Trend is linear (b) Trend is non-linear
 (c) Trend is curvi linear (d) None of them

16) When the trend is of exponential type, the moving averages are to be computed by using:
 (a) Arithmetic mean (b) Geometric mean
 (c) Harmonic mean (d) Weighted mean

17) The long term trend of a time series graph appears to be:
 (a) Straight-line (b) Upward
 (c) Downward (d) Parabolic curve or third degree curve

18) Indicate which of the following an example of seasonal variations is:
 (a) Death rate decreased due to advance in science
 (b) The sale of air condition increases during summer
 (c) Recovery in business
 (d) Sudden causes by wars

19) The most commonly used mathematical method for measuring the trend is:
 (a) Moving average method (b) Semi average method
 (c) Method of least squares (d) None of them

20) A trend is the better fitted trend for which the sum of squares of residuals is:
 (a) Maximum (b) Minimum
 (c) Positive (d) Negative

21) Decomposition of time series is called:

(a) Historigram (b) Analysis of time series

(c) Histogram (d) Detrending

22) The fire in a factory is an example of:

(a) Secular trend (b) Seasonal movements

(c) Cyclical variations (d) Irregular variations

23) Increased demand of admission in the subject of computer in Uttar Pradesh is:

(a) Secular trend (b) Cyclical trend

(c) Seasonal trend (d) Irregular trend

24) Damages due to floods, droughts, strikes fires and political disturbances are:

(a) Trend (b) Seasonal

(c) Cyclical (d) Irregular

25) The general pattern of increase or decrease in economics or social phenomena is shown by:

(a) Seasonal trend (b) Cyclical trend

(c) Secular trend (d) Irregular trend

26) In moving average method, we cannot find the trend values of some:

(a) Middle periods (b) End periods

(c) Starting periods (d) Between extreme periods

27) Moving-averages:

(a) Give the trend in a straight line (b) Measure the seasonal variations

(c) Smooth-out the time series (d) None of them

28) The rise and fall of a time series over periods longer than one year is called:

(a) Secular trend (b) Seasonal variation

(c) Cyclical variation (d) Irregular variations

29) A time series has:

(a) Two Components (b) Three Components

(c) Four Components (d) Five Components

30) The multiplicative time series model is:

(a) $Y = T + S + C + I$ (b) $Y = TSCI$

(c) $Y= a + bx$ (d) $y = a + bx + C x 2$

31) The additive model of Time Series

(a) $Y = T + S + C + I$ (b) $Y = TSCI$

(c) $Y= a + bx$ (d) $y = a + bx + C x 2$

32) A pattern that is repeated throughout a time series and has a recurrence period of at most one year is called:

(a) Cyclical variation (b) Irregular variation

(c) Seasonal variation (d) Long term variation

33) If an annual time series consisting of even number of years is coded, then each coded interval is equal to:

(a) Half year (b) One year

(c) Both (a) and (b) (d) Two years

34) In semi averages method, if the number of values is odd then we drop:

(a) First value (b) Last value

(c) Middle value (d) Middle two values

35) The trend values in freehand curve method are obtained by:

(a) Equation of straight line (b) Graph

(c) Second degree parabola (d) All of the above

ANSWERS

Unit-II

Exercise 1(a)

4) 115, 117, 108, 107, 117.4, 123.6

6) 5.125, 5.625, 6.500, 7.250, 5.475, 5.700

8) 82.86, 98.45, 114.60, 104.08

11) a) y = 94.6 + 7.39 x (b) Monthly trend equation is y = 7.88 + 0.05 x

Exercise 1(b)

1. (d)	**2.** (d)	**3.** (c)	**4.** (c)	**5.** (b)	**6.** (c)	**7.** (a)	**8.** (a)	**9.** (c)	**10.** (b)
11. (a)	**12.** (c)	**13.** (d)	**14.** (b)	**15.** (a)	**16.** (d)	**17.** (b)	**18.** (c)	**19.** (b)	**20.** (b)
21. (d)	**22.** (a)	**23.** (d)	**24.** (c)	**25.** (d)	**26.** (c)	**27.** (c)	**28.** (c)	**29.** (c)	**30.** (b)
31. (b)	**32.** (c)	**33.** (c)	**34.** (c)	**35.** (b)					

TABLE 1(a)

Compound Interest

Annual Compounding

No. of Periods n	10% per Annum $i = 0.10$	$(1 + i)^n$ 14% per Annum $i = 0.14$	18% per Annum $i = 0.18$
1	1.1	1.14	1.18
2	1.21	1.2996	1.3924
3	1.331	1.48154	1.64303
4	1.4641	1.68896	1.93878
5	1.61051	1.92541	2.28776
6	1.77156	2.19497	2.69955
7	1.94872	2.50227	3.18547
8	2.14359	2.85258	3.75886
9	2.35795	3.25194	4.43546
10	2.59374	3.70722	5.23384
11	2.85312	4.22622	6.17593
12	3.13843	4.8179	7.28759
13	3.45227	5.4924	8.59936
14	3.7975	6.26,133	10.1472
15	4.17725	7.13792	11.9738
16	4.59497	8.13723	14.129
17	5.05447	9.27644	16.6723
18	5.55992	10.5751	19.6733
19	6.11591	12.0557	23.2144
20	6.7275	12.7435	27.393

TABLE 1(b)

Present Value of Re. 1

Annual Compounding

No. of Periods n	10% per Annum	$(1 + i)^{-n}$ 14% per Annum	18% per Annum
1	.909091	.877193	.847458
2	.826446	.769468	.718184
3	.751315	.674972	.608631
4	.683014	.592081	.515789
5	.620921	.519369	.437109
6	.564474	.455587	.370432
7	.513158	.399638	.313925
8	.466507	.35056	.266038
9	.424098	.307508	.225456
10	.385543	.269744	.191064
11	.350494	.236618	.161919
12	.318631	.20756	.137219
13	.289664	.18207	.116288
14	.263331	.15971	.0985489
15	.239392	.140097	.083516
16	.217629	.122892	.0707763
17	.197845	.1078	.0599799
18	.179859	.0945614	.0508304
19	.163508	.0829486	.0430766
20	.148644	0.72762	.0365056

TABLE 2(a)

Present Value of an Annuity

Annual Compounding

No. of Periods n	10% per Annum		14% per Annum		18% per Annum	
	$P(n, i)$	$1/P(n, i)$	$P(n, i)$	$1/P(n, i)$	$P(n, i)$	$1/P(n, i)$
1	.909091	1.1	.877192	1.14	.847458	1.18
2	1.73554	.576191	1.64666	.60729	1.56564	.638716
3	2.48685	.402115	2.32163	.430732	2.17427	.459924
4	3.16987	.315471	2.91371	.343205	2.69006	.371739
5	3.79079	.263798	3.43308	.291284	3.12717	.319778
6	4.35526	.229607	3.88867	.257158	3.4976	.28591
7	4.86842	.205406	4.2883	.233193	3.81153	.262362
8	5.33493	.187444	4.63886	.21557	4.07757	.245244
9	5.75902	.173641	4.94637	.202169	4.30302	.232395
10	6.14457	.162745	5.21611	.191714	4.49409	.222515
11	6.49506	.153963	5.45273	.183394	4.65601	.214776
12	6.81369	.146763	5.66029	.176669	4.79323	.208628
13	7.10336	.140779	5.84236	.171164	4.90951	.203686
14	7.36669	.135746	6.00207	.166609	5.00806	.199678
15	7.60608	.131474	6.14217	.162809	5.09158	.196403
16	7.82371	.127817	6.26506	.159615	5.16236	.19371
17	8.02155	.124664	6.37286	.156915	5.22233	.191485
18	8.20141	.12193	6.46742	.154621	5.27316	.189639
19	8.36492	.119547	6.55037	.152663	5.31624	.188103
20	8.51356	.11746	6.62313	.150986	5.35275	.18682

TABLE 2(b)

Amount of an Annuity

Annual Computing

No. of Periods n	10% per Annum		14% per Annum		18% per Annum	
	$A(n, i)$	$1/A(n, i)$	$A(n, i)$	$1/A(n, i)$	$A(n, i)$	$1/A(n, i)$
1	1,000000	.999999994	1.00000001	999999993	1	.999999996
2	2.100000	.476190473	2.14000001	.467289717	2.18000001	.458715595
3	3.310000	.302114802	3.43960003	.290731478	3.57240001	.27992386
4	4,641000	.215470802	4.92114404	.203204782	5.21543202	.19173867
5	6.105100	.16379748	6.61010421	.151283545	7.15420979	.139777841
6	7.71561006	.129607379	8.53551881	.117157495	9.44196755	.105910129
7	9,48717108	.105405499	10.7304915	.0931923765	12.1415217	.082361999
8	11.4358882	.0874440168	13.2327603	0.755700232	15.3269956	.0652443586
9	13.579477	0.736405385	16.0853467	.0621683833	19.0858549	.0523948237
10	15.9374248	.0627453949	19.3372953	0.517135403	23.5213088	.0425146411
11	18.5311672	0.539631415	23.0445166	.043394271	28.7551443	.034776386
12	21.384284	.0467633146	27.270749	.0366693265	34.9310704	.0286278087
13	24.5227124	.0407785234	32.0886539	.0311636631	42.218663	.0236862072
14	27.9749837	.0357462229	37.5810655	.0266091445	50.8180224	0.196780582
15	31.772482	.0314737765	43.8424147	.0228089627	60.9652664	.0164027824
16	35.9497303	.0278166204	50.9803528	.0196153998	72.9390144	.0137100838
17	40.5447033	.0246641341	59.1176022	.0169154357	87.0680371	.011485271
18	45.5991737	.021930222	68.3940666	.0146211514	103.740284	.00963945696
19	51.1590911	.019546868	78.969236	.0126631591	123.413535	.00810283897
20	57.274999	.0174596250	91.0249291	.0109860014	146.627971	.00681998115

TABLE 3

Future Value and Present Value

i = rate of interest per period, n = number of periods

n	$i = \frac{1}{4}\%$ $(1+i)^n$	$i = \frac{1}{4}\%$ $(1+i)^{-n}$	$i = \frac{1}{2}\%$ $(1+i)^n$	$i = \frac{1}{2}\%$ $(1+i)^{-n}$	$i = \frac{3}{4}\%$ $(1+i)^n$	$i = \frac{3}{4}\%$ $(1+i)^{-n}$
1	1.0025 0000	0.9975 0623	1.0050 0000	0.9950 2488	1.0075 0000	0.9925 5583
2	1.0050 0625	0.9950 1869	1.0100 2500	0.9900 7450	1.0150 5625	0.9851 6708
3	1.0075 1877	0.9925 3734	1.0150 7513	0.9851 4876	1.0226 6917	0.9778 3333
4	1.0100 3756	0.9900 6219	1.0201 5050	0.9802 4752	1.0303 3919	0.9705 5417
5	1.0125 6266	0.9875 9321	1.0252 5125	0.9753 7067	1.0380 6673	0.9633 2920
6	1.0150 9406	0.9851 3038	1.0303 7751	0.9705 1808	1.0458 5224	0.9561 5802
7	1.0176 3180	0.9826 7370	1.0355 2940	0.9656 8963	1.0536 9613	0.9490 4022
8	1.0201 7588	0.9802 2314	1.0407 0704	0.9608 8520	1.0615 9885	0.9419 7540
9	1.0227 2632	0.9777 7869	1.0459 1058	0.9561 0468	1.0695 6084	0.9349 6318
10	1.0252 8313	0.9753 4034	1.0511 4013	0.9513 4794	1.0775 8255	0.9280 0315
11	1.0278 4634	0.9729 0807	1.0563 9583	0.9466 1487	1.0856 6441	0.9210 9494
12	1.0304 1596	0.9704 8187	1.0616 7781	0.9419 0534	1.0938 0690	0.9142 3815
13	1.0329 9200	0.9680 6171	1.0669 8620	0.9372 1924	1.1020 1045	0.9074 3241
14	1.0355 7448	0.9656 4759	1.0723 2113	0.9325 5646	1.1102 7553	0.9006 7733
15	1.0381 6341	0.9632 3949	1.0776 8274	0.9279 1688	1.1186 0259	0.8939 7254
16	1.0407 5882	0.9608 3740	1.0830 7115	0.9233 0037	1.1269 9211	0.8873 1766
17	1.0433 6072	0.9584 4130	1.0884 8651	0.9187 0684	1.1354 4455	0.8307 1231
18	1.0459 6912	0.9560 5117	1.0939 2894	0.9141 3616	1.1439 6039	0.8741 5614
19	1.0485 8404	0.9536 6700	1.0993 9858	0.9095 8822	1.1525 4009	0.8676 4878
20	1.0512 0550	0.9512 8878	1.1048 9558	0.9050 6290	1.1611 8414	0.8611 8985
21	1.0538 3352	0.9489 1649	1.1104 2006	0.9005 6010	1.1698 9302	0.8547 7901
22	1.0564 6810	0.9465 5011	1.1159 7216	0.8960 7971	1.1786 6722	0.8484 1589
23	1.0591 0927	0.9441 8964	1.1215 5202	0.8916 2160	1.1875 0723	0.8421 0014
24	1.0617 5704	0.9418 3505	1.1271 5978	0.8871 8567	1.1964 1353	0.8358 3140
25	1.0644 1144	0.9394 8634	1.1327 9558	0.8827 7181	1.2053 8663	0.8296 0933
26	1.0670 7247	0.9371 4348	1.1384 5955	0.8783 7991	1.2144 2703	0.8234 3358
27	1.0697 4015	0.9348 0646	1.1441 5185	0.8740 0986	1.2235 3523	0.8173 0380
28	1.0724 1450	0.9324 7527	1.1498 7261	0.8696 6155	1.2327 1175	0.8112 1966
29	1.0750 9553	0.9301 4990	1.1556 2197	0.8653 3488	1.2419 5709	0.8051 8080
30	1.0777 8327	0.9278 3032	1.1614 0008	0.8610 2973	1.2512 7176	0.7991 8690
31	1.0804 7773	0.9255 1653	1.1672 0708	0.8567 4600	1.2606 5630	0.7932 3762
32	1.0831 7892	0.9232 0851	1.1730 4312	0.8524 8358	1.2701 1122	0.7873 3262
33	1.0858 8687	0.9209 0624	1.1789 0833	0.8482 4237	1.2796 3706	0.7814 7158
34	1.0886 0159	0.9186 0972	1.1848 0288	0.8440 2226	1.2892 3434	0.7756 5418
35	1.0913 2309	0.9163 1892	1.1907 2689	0.8398 2314	1.2989 0359	0.7698 8008
36	1.0940 5140	0.9140 3384	1.1966 8052	0.8356 4492	1.3086 4537	0.7641 4896
37	1.0967 8653	0.9117 5445	1.2026 6393	0.8314 8748	1.3184 6021	0.7584 6051
38	1.0995 2850	0.9094 8075	1.2086 7725	0.8273 5073	1.3283 4866	0.7528 1440
39	1.1022 7732	0.9072 1272	1.2147 2063	0.8232 3455	1.3383 1128	0.7472 1032
40	1.1050 3301	0.9049 5034	1.2207 9424	0.8191 3886	1.3483 4861	0.7416 4796
41	1.1077 9559	0.9026 9361	1.2268 9821	0.8150 6354	1.3584 6123	0.7361 2701
42	1.1105 6508	0.9004 4250	1.2330 3270	0.8110 0850	1.3686 4969	0.7306 4716
43	1.1133 4149	0.8981 9701	1.2391 9786	0.8069 7363	1.3789 1456	0.7252 0809
44	1.1161 2485	0.8959 5712	1.2453 9385	0.8029 5884	1.3892 5642	0.7198 0952
45	1.1189 1516	0.8937 2281	1.2516 2082	0.7989 6402	1.3996 7584	0.7144 5114
46	1.1217 1245	0.8914 9407	1.2578 7892	0.7949 8907	1.4101 7341	0.7091 3264
47	1.1245 1673	0.8892 7090	1.2641 6832	0.7910 3390	1.4207 4971	0.7038 5374
48	1.1273 2802	0.8870 5326	1.2704 8916	0.7870 9841	1.4314 0533	0.6986 1414
49	1.1301 4634	0.8848 4116	1.2768 4161	0.7831 8250	1.4421 4087	0.6934 1353
50	1.1329 7171	0.8826 3457	1.2832 2581	0.7792 8607	1.4529 5693	0.6882 5165

	$i = 1\%$		$i = 1\frac{1}{4}\%$		$i = 1\frac{1}{2}\%$	
n	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$
1	1.0100 0000	0.9900 9901	1.0125 0000	0.9876 5432	1.0150 0000	0.9852 2167
2	1.0201 0000	0.9802 9605	1.0251 5625	0.9754 6106	1.0302 2500	0.9706 6175
3	1.0303 0100	0.9705 9015	1.0379 7070	0.9634 1833	1.0456 7838	0.9563 1699
4	1.0406 0401	0.9609 8034	1.0509 4534	0.9515 2428	1.0613 6355	0.9421 8423
5	1.0510 1005	0.9514 6569	1.0640 8215	0.9397 7706	1.0772 8400	0.9282 6033
6	1.0615 2015	0.9420 4524	1.0773 8318	0.9281 7488	1.0934 4326	0.9145 4219
7	1.0721 3535	0.9327 1805	1.0908 5047	0.9167 1593	1.1098 4491	0.9010 2679
8	1.0828 5671	0.9234 8322	1.1044 8610	0.9053 9845	1.1264 9259	0.8877 1112
9	1.0936 8527	0.9143 3982	1.1182 9218	0.8942 2069	1.1433 8998	0.8745 9224
10	1.1046 2213	0.9052 8695	1.1322 7083	0.8831 8093	1.1605 4083	0.8616 6723
11	1,1156 6835	0.8963 2372	1,1464 2422	0.8722 7746	1.1779 4894	0.8489 3323
12	1.1268 2503	0.8874 4923	1.1607 5452	0.8615 0860	1.1956 1817	0.8363 8742
13	1.1380 9328	0.8786 6260	1.1752 6395	0.8508 7269	1.2135 5244	0.8240 2702
14	1.1494 7421	0.8699 6297	1.1899 5475	0.8403 6809	1.2317 5573	0.8118 4928
15	1.1609 6896	0.8613 4947	1.2048 2918	0.8299 9318	1.2502 3207	0.7998 5150
16	1.1725 7864	0.8528 2126	1.2198 8955	0.8197 4635	1.2689 8555	0.7880 3104
17	1.1843 0443	0.8443 7749	1.2351 3817	0.8096 2602	1.2880 2033	0.7763 8526
18	1.1961 4748	0.8360 1731	1.2505 7739	0.7996 3064	1.3073 4064	0.7649 1159
19	1.2081 0895	0.8277 3992	1.2662 0961	0.7897 5866	1.3269 5075	0.7536 0747
20	1.2201 9004	0.8195 4447	1.2820 3723	0.7800 0855	1.3468 5501	0.7424 7042
21	1.2323 9194	0.8114 3017	1.2980 6270	0.7703 7881	1.3670 5783	0.7314 9795
22	1.2447 1586	0.8033 9621	1.3142 8848	0.7608 6796	1.3875 6370	0.7206 8763
23	1.2571 6302	0.7954 4179	1.3307 1709	0.7514 7453	1.4083 7715	0.7100 3708
24	1.2697 3465	0.7875 6613	1.3473 5105	0.7421 9707	1.4295 0281	0.6995 4392
25	1.2824 3200	0.7797 6844	1.3641 9294	0.7330 3414	1.4509 4535	0.6892 0583
26	1.2952 5631	0.7720 4796	1.3812 4535	0.7239 8434	1.4727 0953	0.6790 2052
27	1.3082 0888	0.7644 0392	1.3985 1092	0.7150 4626	1.4948 0018	0.6689 8574
28	1.3212 9097	0.7568 3557	1.4159 9230	0.7062 1853	1.5172 2218	0.6590 9925
29	1.3345 0388	0.7493 4215	1.4336 9221	0.6974 9978	1.5399 8051	0.6493 5887
30	1.3478 4892	0.7419 2292	1.4516 1336	0.6888 8867	1.5630 8022	0.6397 6243
31	1.3613 2740	0.7345 7715	1.4697 5853	0.6803 8387	1.5865 2642	0.6303 0781
32	1.3749 4068	0.7273 0411	1.4881 3051	0.6719 8407	1.6103 2432	0.6209 9292
33	1.3886 9009	0.7201 0307	1.5067 3214	0.6636 8797	1.6344 7918	0.6118 1568
34	1.4025 7699	0.7129 7334	1.5255 6629	0.6554 9429	1.6589 9637	0.6027 7407
35	1.4166 0276	0.7059 1420	1.5446 3587	0.6474 0177	1.6838 8132	0.5938 6608
36	1.4307 6878	0.6989 2495	1.5639 4382	0.6394 0916	1.7091 3954	0.5850 8974
37	1.4450 7647	0.6920 0490	1.5834 9312	0.6315 1522	1.7347 7663	0.5764 4309
38	1.4595 2724	0.6851 5337	1.6032 8678	0.6237 1873	1.7607 9828	0.5679 2423
39	1.4741 2251	0.6783 6967	16233 2787	06160 1850	1.7872 1025	0.5595 3126
40	1.4888 6373	0.6716 5314	1.6436 1946	0.6084 1334	1.8140 1841	0.5512 6232
41	1.5037 5237	0.6650 0311	1.6641 6471	0.6009 0206	1.8412 2868	0.5431 1559
42	1.5187 8989	0.6584 1892	1.6849 6677	0.5934 8352	1.8688 4712	0.5350 8925
43	1.5339 7779	0.6518 9992	1.7060 2885	0.5861 5656	1.8968 7982	0.5271 8153
44	1.5493 1757	0.6454 4546	1.7273 5421	0.5789 2006	1.9253 3302	0.5193 9067
45	1.5648 1075	0.6390 5492	1.7489 4614	0.5717 7290	1.9542 1301	0.5117 1494
46	1.5804 5885	0.6327 2764	1.7708 0797	0.5647 1397	1.9835 2621	0.5041 5265
47	1.5962 6344	0.6264 6301	1.7929 4306	0.5577 4219	2.0132 7910	0.4967 0212
48	1.6122 2608	0.6202 6041	1.8153 5485	0.5508 5649	2.0434 7829	0.4893 6170
49	1.6283 4834	0.6141 1921	1.8380 4679	0.5440 5579	2.0741 3046	0.4821 2975
50	1.6446 3182	0.6080 3882	1.8610 2237	0.5373 3905	2.1052 4242	0.4750 0468

	$i = 1\frac{3}{4}\%$		$i = 2\%$		$i = 2\frac{1}{4}\%$	
n	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$
1	1.0175 0000	0.9828 0098	1.0200 0000	0.9803 9216	1.0225 0000	0.9779 9511
2	1.0353 0625	0.9658 9777	1.0404 0000	0.9611 6878	1.0455 0625	0.9564 7444
3	1.0534 2411	0.9492 8528	1.0612 0800	0.9423 2233	1.0690 3014	0.9354 2732
4	1.0718 5903	0.9329 5851	1.0824 3216	0.9238 4543	1.0930 8332	0.9148 4335
5	1.0906 1656	0.9169 1254	1.1040 8080	0.9057 3081	1.1176 7769	0.8947 1232
6	1.1097 0235	0.9011 4254	1.1261 6242	0.8879 7138	1.1428 2544	0.8750 2427
7	1.1291 2215	0.8856 4378	1.1486 8567	0.8705 6018	1.1685 3901	0.8557 6946
8	1.1488 8178	0.8704 1157	1.1716 5938	0.8534 9037	1.1948 3114	0.8369 3835
9	1.1689 8721	0.8554 4135	1.1950 9257	0.8367 5527	1.2217 1484	0.8185 2161
10	1.1894 4449	0.8407 2860	1.2189 9442	0.8203 4830	1.2492 0343	0.8005 1013
11	1.2102 5977	0.8262 6889	1.2433 7431	0.8042 6304	1.2773 1050	0.7828 9499
12	1.2314 3931	0.8120 5788	1.2682 4179	0.7884 9318	1.3060 4999	0.7656 6748
13	1.2529 8950	0.7980 9128	1.2936 0663	0.7730 3253	1.3354 3611	0.7488 1905
14	1.2749 1682	0.7843 6490	1.3194 7876	0.7578 7502	1.3654 8343	0.7323 4137
15	1.2972 2786	0.7708 7459	1.3458 6834	0.7430 1473	1.3962 0680	0.7162 2628
16	1.3199 2935	0.7576 1631	1.3727 8571	0.7284 4581	1.4276 2146	0.7004 6580
17	1.3430 2811	0.7445 8605	1.4002 4142	0.7141 6256	1.4597 4294	0.6850 5212
18	1.3665 3111	0.7317 7990	1.4282 4625	0.7001 5937	1.4925 8716	0.6699 7763
19	1.3904 4540	0.7191 9401	1.4568 1117	0.6864 3076	1.5261 7037	0.6552 3484
20	1.4147 7820	0.7068 2458	1.4859 4740	0.6729 7133	1.5605 0920	0.6408 1647
21	1.4395 3681	0.6946 6789	1.5156 6634	0.6597 7582	1.5956 2066	0.6267 1538
22	1.4647 2871	0.6827 2028	1.5459 7967	0.6468 3904	1.6315 2212	0.6129 2457
23	1.4903 6146	0.6709 7817	1.5768 9926	0.6341 5592	1.6682 3137	0.5994 3724
24	1.5164 4279	0.6594 3800	1.6084 3725	0.6217 2149	1.7057 6658	0.5862 4668
25	1.5429 8054	0.6480 9632	1.6406 0599	0.6095 3087	1.7441 4632	0.5733 4639
26	1.5699 8269	0.6369 4970	1.6734 1811	0.5975 7928	1.7833 8962	0.5607 2997
27	1.5974 5739	0.6259 9479	1.7068 8648	0.5858 6204	1.8235 1588	0.5483 9117
28	1.6254 1290	0.6152 2829	1.7410 2421	0.5743 7455	1.8645 4499	0.5363 2388
29	1.6538 5762	0.6046 4697	1.7758 4469	0.5631 1231	1.9064 9725	0.5245 2213
30	1.6828 0013	0.5942 4764	1.8113 6158	0.5520 7089	1.9493 9344	0.5129 8008
31	1.7122 4913	0.5840 2716	1.8475 8882	0.5412 4597	1.9932 5479	0.5016 9201
32	1.7422 1349	0.5739 8247	1.8845 4059	0.5306 3330	2.0381 0303	0.4906 5233
33	1.7727 0223	0.5641 1053	1.9222 3140	0.5205 2873	2.0839 6034	0.4798 5558
34	1,8037 2452	0.5544 0839	1.9606 7603	0.5100 2817	2.1308 4945	0.4692 9641
35	1.8352 8970	0.5448 7311	1.9998 8955	0.5000 2761	2.1787 9356	0.4589 6960
36	1.8674 0727	0.5355 0183	2.0398 8734	0.4902 2315	2.2278 1642	0.4488 7002
37	1.9000 8689	0.5262 9172	2.0806 8509	0.4806 1093	2.2779 4229	0.4389 9268
38	1.9333 3841	0.5172 4002	2.1222 9879	0.4711 8719	2.3291 9599	0.4293 3270
39	1.9671 7184	0.5083 4400	2.1647 4477	0.4619 4822	2.3816 0290	0.4198 8528
40	2.0015 9734	0.4996 0098	2.2080 3966	0.4528 9042	2.4351 8897	0.4106 4575
41	2.0366 2530	0.4910 0834	2.2522 0046	0.4440 1021	2.4899 8072	0.4016 0954
42	2.0722 6624	0.4825 6348	2.2972 4447	0.4353 0413	2.5460 0528	0.3927 7216
43	2.1085 3090	0.4742 6386	2.3431 8936	0.4267 6875	2.6032 9040	0.3841 2925
44	2.1454 3019	0.4661 0699	2.3900 5314	0.4184 0074	2.6618 6444	0.3756 7653
45	2.1829 7522	0.4580 9040	2.4378 5421	0.4101 9680	2.7217 5639	0.3674 0981
46	2.2211 7728	0.4502 1170	2.4866 1129	0.4021 5373	2.7829 9590	0.3593 2500
47	2.2600 4789	0.4424 6850	2.5363 4352	0.3942 6836	2.8456 1331	0.3514 1809
48	2.2995 9872	0.4348 5848	2.5870 7039	0.3865 3761	2.9096 3961	0.3436 8518
49	2.3398 4170	0.4273 7934	2.6388 1179	0.3789 5844	2.9751 0650	0.3361 2242
50	2.3807 8893	0.4200 2883	2.6915 8803	0.3715 2788	3.0420 4640	0.3287 2608

	$i = 2\frac{1}{2}\%$		$i = 3\%$		$i = 3\frac{1}{2}\%$	
n	$(1 + i)^n$	$(1 + i)^{-n}$	$(1 + i)^n$	$(1 + i)^{-n}$	$(1 + i)^n$	$(1 + i)^{-n}$
1	1.0250 0000	0.9756 0976	1.0300 0000	0.9708 7379	1.0350 0000	0.9661 8357
2	1.0506 2500	0.9518 1440	1.0609 0000	0.9425 9591	1.0712 2500	0.9335 1070
3	1.0768 9063	0.9285 9941	1.0927 2700	0.9151 4166	1.1087 1788	0.9019 4271
4	1.1038 1289	0.9059 5064	1.1255 0881	0.8884 8705	1.1475 2300	0.8714 4223
5	1.1314 0821	0.8838 5429	1.1592 7407	0.8626 0878	1.1876 8631	0.8419 7317
6	1.1596 9342	0.8622 9687	1.1940 5230	0.8374 8426	1.2292 5533	0.8135 0064
7	1.1886 8575	0.8412 6524	1.2298 7387	0.8130 9151	1.2722 7926	0.7859 9096
8	1.2184 0290	0.8207 4657	1.2667 7008	0.7894 0923	1.3168 0904	0.7594 1156
9	1.2488 6297	0.8007 2836	1.3047 7318	0.7664 1673	1.3628 9735	0.7337 3097
10	1.2800 8454	0.7811 9840	1.3439 1638	0.7440 9391	1.4105 9876	0.7089 1881
11	1.3120 8666	0.7621 4478	1.3842 3387	0.7224 2128	1.4599 6972	0.6849 4571
12	1.3448 8882	0.7435 5589	1.4257 6089	0.7013 7988	1.5110 6866	0.6617 8330
13	1.3785 1104	0.7254 2038	1.4685 3371	0.6809 5134	1.5639 5606	0.6394 0415
14	1.4129 7382	0.7077 2720	1.5125 8972	0.6611 1781	1.6186 9452	0.6177 8179
15	1.4482 9817	0.6904 6556	1.5579 6742	0.6418 6195	1.6753 4883	0.5968 9062
16	1.4845 0562	0.6736 2493	1.6047 0644	0.6231 6694	1.7339 8604	0.5767 0591
17	1.5216 1826	0.6571 9506	1.6528 4763	0.6050 1645	1.7946 7555	0.5572 0378
18	1.5596 5872	0.6411 6591	1.7024 3306	0.5873 9461	1.8574 8920	0.5383 6114
19	1.5986 5019	0.6255 2772	1.7535 0605	0.5702 8603	1.9225 0132	0.5201 5569
20	1.6386 1644	0.6102 7094	1.8061 1123	0.5536 7575	1.9897 8886	0.5025 6588
21	1.6795 8185	0.5953 8629	1.8602 9457	0.5375 4928	2.0594 3147	0.4855 7090
22	1,7215 7140	0.5808 6467	1,9161 0341	0.5218 9250	2,1315 1158	0.4691 5063
23	1.7646 1068	0.5666 9724	1.9735 8651	0.5066 9175	2.2061 1448	0.4532 8563
24	1.8087 2595	0.5528 7535	2.0327 9411	0.4919 3374	2.2833 2849	0.4379 5713
25	1.8539 4410	0.5393 9059	2.0937 7793	0.4776 0557	2.3632 4498	0.4231 4699
26	1.9002 9270	0.5262 3472	2.1565 9127	0.4636 9473	2.4459 5856	0.4088 3767
27	1.9478 0002	0.5133 9973	2.2212 8901	0.4501 8906	2.5315 6711	0.3950 1224
28	1.9964 9502	0.5008 7778	2.2879 2768	0.4370 7675	2.6201 7196	0.3816 5434
29	2.0464 0739	0.4886 6125	2.3565 6551	0.4243 4636	2.7118 7798	0.3687 4815
30	2.0975 6758	0.4767 4269	2.4272 6247	0.4119 8676	2.8067 9370	0.3562 7841
31	2.1500 0677	0.4651 1481	2.5000 8035	0.3999 8715	2.9050 3148	0.3442 3035
32	2.2037 5694	0.4537 7055	2.5750 8276	0.3883 3703	3.0067 0759	0.3325 8971
33	2.2588 5086	0.4427 0298	2.6523 3524	0.3770 2625	3.1119 4235	0.3213 4271
34	2.3153 2213	0.4319 0534	2.7319 0530	0.3660 4490	3.2208 6033	0.3104 7605
35	2.3732 0519	0.4213 7107	2.8138 6245	0.3553 8340	3.3335 9045	0.2999 7686
36	2.4325 3532	0.4110 9372	2.8982 7833	0.3450 3243	3.4502 6611	0.2898 3272
37	2.4933 4870	0.4010 6705	2.9852 2668	0.3349 8294	3.5710 2543	0.2800 3161
38	2.5556 8242	0.3912 8492	3.0747 8348	0.3252 2615	3.6960 1132	0.2705 6194
39	2.6195 7448	0.3817 4139	3.1670 2698	0.3157 5355	3.8253 7171	0.2614 1250
40	2.6850 6384	0.3724 3062	3.2620 3779	0.3065 5684	3.9592 5972	0.2525 7247
41	2.7521 9043	0.3633 4695	3.3598 9893	0.2976 2800	4.0978 3381	0.2440 3137
42	2.8209 9520	0.3544 8483	3.4606 9589	0.2889 5922	4.2412 5799	0.2357 7910
43	2.8915 2008	0.3458 3886	3.5645 1677	0.2805 4294	4.3897 0202	0.2278 0590
44	2.9638 0808	03374 0376	3.6714 5227	0.2723 7178	4.5433 4160	0.2201 0231
45	3.0379 0328	0.3291 7440	3.7815 9584	0.2644 3862	4.7023 5855	0.2126 5924
46	3.1138 5086	0.3211 4576	3.8950 4372	0.2567 3653	4.8669 4110	0.2054 6787
47	3.1916 9713	0.3133 1294	4.0118 9503	0.2492 5876	5.0372 8404	0.1985 1968
48	3.2714 8956	0.3056 7116	4.1322 5188	0.2419 9880	5.2135 8898	0.1918 0645
49	3.3532 7680	0.2982 1576	4.2562 1944	0.2349 5029	5.3960 6459	0.1853 2024
50	3.4371 0872	0.2909 4221	4.3839 0602	0.2281 0708	5.5849 2686	0.1790 5337

	i = 4%		i = 4½%		i = 5%	
n	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$
1	1.0400 0000	0.9615 3846	1.0450 0000	0.9569 3780	1.0500 0000	0.9523 8095
2	1.0816 0000	0.9245 5621	1.0920 2500	0.9157 2995	1.1025 0000	0.9070 2948
3	1.1248 6400	0.8889 9636	1.1411 6613	0.8762 9660	1.1576 2500	0.8638 3760
4	1.1698 5856	0.8548 0419	1.1925 1860	0.8382 6134	1.2155 0625	0.8227 0247
5	1.2166 5290	0.8219 2711	1.2461 8194	0.8024 5105	1.2762 8156	0.7835 2617
6	1.2653 1902	0.7903 1453	1.3022 6012	0.7678 9574	1.3400 9564	0.7462 1540
7	1.3159 3178	0.7599 1781	1.3608 6183	0.7348 2846	1.4071 0042	0.7106 8133
8	1.3685 6905	0.7306 9021	1.4221 0061	0.7031 8513	1.4774 5544	0.6768 3936
9	1.4233 1181	0.7025 8674	1.4860 9514	0.6729 0443	1.5513 2822	0.6446 0892
10	1.4802 4428	0.6755 6417	1.5529 6942	0.6439 2768	1.6288 9463	0.6139 1325
11	1.5394 5406	0.6495 8093	1.6228 5305	0.6161 9874	1.7103 3936	0.5846 7929
12	1.6010 3222	0.6245 9705	1.6958 8143	0.5896 6386	1.7958 5633	0.5568 3742
13	1.6650 7351	0.6005 7409	1.7721 9610	0.5642 7164	1.8856 4914	0.5303 2135
14	1.7316 7645	0.5774 7508	1.8519 4492	0.5399 7286	1.9799 3160	0.5050 6795
15	1.8009 4351	0.5552 6450	1.9352 8244	0.5167 2044	2.0789 2818	0.4810 1710
16	1.8729 8125	0.5339 0818	2.0223 7015	0.4944 6932	2.1828 7459	0.4581 1152
17	1.9479 0050	0.5133 7325	2.1133 7681	0.4731 7639	2.2920 1832	0.4362 9669
18	2.0258 1652	0.4936 2812	2.2084 7877	0.4528 0037	2.4066 1923	0.4155 2065
19	2.1068 4918	0.4746 4242	2.3078 6031	0.4333 0179	2.5269 5020	0.3957 3396
20	2.1911 2314	0.4563 8695	2.4117 1402	0.4146 4286	2.6532 9771	0.3768 8948
21	2.2787 6807	0.4388 3360	2.5202 4116	0.3967 8743	2.7859 6259	0.3589 4236
22	2.3699 1879	0.4219 5539	2.6336 5201	0.3797 0089	2.9252 6072	0.3418 4987
23	2.4647 1554	0.4057 2633	2.7521 6635	0.3633 5013	3.0715 2376	0.3255 7131
24	2.5633 0416	0.3901 2147	2.8760 1383	0.3477 0347	3.2250 9994	0.3100 6791
25	2.6658 3633	0.3751 1680	3.0054 3446	0.3327 3060	3.3863 5494	0.2953 0277
26	2.7724 6978	0.3606 8923	3.1406 7901	0.3184 0248	3.5556 7269	0.2812 4073
27	2.8833 6858	0.3468 1657	3.2820 0956	0.3046 9137	3.7334 5632	0.2678 4832
28	2.9987 0332	0.3334 7747	3.4296 9999	0.2915 7069	3.9201 2914	0.2550 9364
29	3.1186 5145	0.3206 5141	3.5840 3649	0.2790 1502	4.1161 3560	0.2429 4632
30	3.2433 9751	0.3083 1867	3.7453 1813	0.2670 0002	4.3219 4238	0.2313 7745
31	3.3731 3341	0.2964 6026	3.9138 5745	0.2555 0241	4.5380 3949	0.2203 5947
32	3.5080 5875	0.2850 5794	4.0899 8104	0.2444 9991	4.7649 4147	0.2098 6617
33	3.6483 8110	0.2740 9417	4.2740 3018	0.2339 7121	5.0031 8854	0.1998 7254
34	3.7943 1634	0.2635 5209	4.4663 6154	0.2238 9589	5.2533 4797	0.1903 5480
35	3.9460 8899	0.2534 1547	4.6673 4781	0.2142 5444	5.5160 1537	0.1812 9029
36	4.1039 3255	0.2436 6872	4.8773 7846	0.2050 2817	5.7918 1614	0.1726 5741
37	4.2680 8986	0.2342 9685	5.0968 6049	0.1961 9921	6.0814 0694	0.1644 3563
38	4.4388 1345	0.2252 8543	5.3262 1921	0.1877 5044	6.3854 7729	0.1566 0536
39	4.6163 6599	0.2166 2061	5.5658 9908	0.1796 6549	6.7047 5115	0.1491 4797
40	4.8010 2063	0.2082 8904	5.8163 6454	0.1719 2870	7.0399 8871	0.1420 4568
41	4.9930 6145	0.2002 7793	6.0781 0094	0.1645 2507	7.3919 8815	0.1352 8160
42	5.1927 8391	0.1925 7493	6.3516 1548	0.1574 4026	7.7615 8756	0.1288 3962
43	5.4004 9527	0.1851 6820	6.6374 3818	0.1506 6054	8.1496 6693	0.1227 0440
44	5.6165 1508	0.1780 4635	6.9361 2290	0.1441 7276	8.5571 5028	0.1168 6133
45	5.8411 7568	0.1711 9841	7.2482 4843	0.1379 6437	8.9850 0779	0.1112 9651
46	6.0748 2271	0.1646 1386	7.5744 1961	0.1320 2332	9.4342 5818	0.1059 9668
47	6.3178 1562	0.1582 8256	7.9152 6849	0.1263 3810	9.9059 7109	0.1009 4921
48	6.5705 2824	0.1521 9476	8.2714 5557	0.1208 9771	10.4012 6965	0.0961 4211
49	6.8333 4937	0.1463 4112	8.6436 7107	0.1156 9158	10.9213 3313	0.0915 6391
50	7.1066 8335	0.1407 1262	9.0326 3627	0.1107 0965	11.4673 9979	0.0872 0373

	i = 6%		i = 7%		i = 8%	
n	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$	$(1+i)^n$	$(1+i)^{-n}$
1	1.0600 0000	0.9433 9623	1.0700 0000	0.9345 7944	1.0800 0000	0.9259 2593
2	1.1236 0000	0.8899 9644	1.1449 0000	0.8734 3873	1.1664 0000	0.8573 3882
3	1.1910 1600	0.8396 1928	1.2250 4300	0.8162 9788	1.2597 1200	0.7938 3224
4	1.2624 7696	0.7920 9366	1.3107 9601	0.7628 9521	1.3604 8896	0.7350 2985
5	1.3382 2558	0.7472 5817	1.4025 5173	0.7129 8618	1.4693 2808	0.6805 8320
6	1.4185 1911	0.7049 6054	1.5007 3035	0.6663 4222	1.5868 7432	0.6301 6963
7	1.5036 3026	0.6650 5711	1.6057 8148	0.6227 4974	1.7138 2427	0.5834 9040
8	1.5938 4807	0.6274 1237	1.7181 8618	0.5820 0910	1.8509 3021	0.5402 6888
9	1.6894 7896	0.5918 9846	1.8384 5921	0.5439 3374	1.9990 0463	0.5002 4897
10	1.7908 4770	0.5583 9478	1.9671 5136	0.5083 4929	2.1589 2500	0.4631 9349
11	1.8982 9856	0.5267 8753	2.1048 5195	0.4750 9280	2.3316 3900	0.4288 8286
12	2.0121 9647	0.4969 6936	2.2521 9159	0.4440 1196	2.5181 7012	0.3971 1376
13	2.1329 2826	0.4688 3902	2.4098 4500	0.4149 6445	2.7196 2373	0.3676 9792
14	2.2609 0396	04423 0096	2.5785 3415	0.3878 1724	2.9371 9362	0.3404 6104
15	2.3965 5819	0.4172 6506	2.7590 3154	0.3624 4602	3.1721 6911	0.3152 4170
16	2.5403 5168	0.3936 4628	2.9521 6375	0.3387 3460	3.4259 4264	0.2918 9047
17	2.6927 7279	0.3713 6442	3.1588 1521	0.3165 7439	3.7000 1805	0.2702 6895
18	2.8543 3915	0.3503 4379	3.3799 3228	0.2958 6392	3.9960 1950	0.2502 4903
19	3.0255 9950	0.3305 1301	3.6165 2754	0.2765 0833	4.3157 0106	0.2317 1206
20	3.2071 3547	0.3118 0473	3.8696 8446	0.2584 1900	4.6609 5714	0.2145 4821
21	3.3995 6360	0.2941 5540	4.1405 6237	0.2415 1309	5.0338 3372	0.1986 5575
22	3.6035 3742	0.2775 0510	4.4304 0174	0.2257 1317	5.4365 4041	0.1839 4051
23	3.8197 4966	0.2617 9726	4.7405 2986	0.2109 4688	5.8714 6365	0.1703 1528
24	4.0489 3464	0.2469 7855	5.0723 6695	0.1971 4662	6.3411 8074	0.1576 9934
25	4.2918 7072	0.2329 9863	5.4274 3264	0.1842 4918	6.8484 7520	0.1460 1790
26	4.5493 8296	0.2198 1003	5.8073 5292	0.1721 9549	7.3963 5321	0.1352 0176
27	4.8223 4594	0.2073 6795	6.2138 6763	0.1609 3037	7.9880 6147	0.1251 8682
28	5.1116 8670	0.1956 3014	6.6488 3836	0.1504 0221	8.6271 0639	0.1159 1372
29	5.4183 8790	0.1845 5674	7.1142 5705	0.1405 6282	9.3172 7490	0.1073 2752
30	5.7434 9117	0.1741 1013	7.6122 5504	0.1313 6712	10.0626 5689	0.0993 7733
31	6.0881 0064	0.1642 5484	8.1451 1290	0.1227 7301	10.8676 6944	0.0920 1605
32	6.4533 8668	0.1549 5740	8.7152 7080	0.1147 4113	11.7370 8300	0.0852 0005
33	6.8405 8988	0.1461 8622	9.3253 3975	0.1072 3470	12.6760 4964	0.0788 8893
34	7.2510 2528	0.1379 1153	9.9781 1354	0.1002 1934	13.6901 3361	0.0730 4531
35	7.6860 8679	0.1301 0522	10.6765 8148	0.0936 6294	14.7853 4429	0.0676 3454
36	8.1472 5200	0.1227 4077	11.4239 4219	0.0875 3546	15.9681 7184	0.0626 2458
37	8.6360 8712	0.1157 9318	12.2236 1814	0.0818 0884	17.2456 2558	0.0579 8572
38	9.1542 5235	0.1092 3885	13.0792 7141	0.0764 5686	18.6252 7563	0.0536 9048
39	9.7035 0749	0.1030 5552	13.9948 2041	0.0714 5501	20.1152 9768	0.0497 1341
40	10.2857 1794	0.0972 2219	14.9744 5784	0.0667 8038	21.7245 2150	0.0460 3093
41	10.9028 6101	0.0917 1905	16.0226 6989	0.0624 1157	23.4624 8322	0.0426 2123
42	11.5570 3267	0.0865 2740	17.1442 5678	0.0583 2857	25.3394 8187	0.0394 6411
43	12.2504 5463	0.0816 2962	18.3443 5475	0.0545 1268	27.3666 4042	0.0365 4084
44	12.9854 8191	0.0770 0908	19.6284 5959	0.0509 4643	29.5559 7166	0.0338 3411
45	13.7646 1083	0.0726 5007	21.0024 5176	0.0476 1349	31.9204 4939	0.0313 2788
46	14.5904 8748	0.0685 3781	22.4726 2338	0.0444 9859	34.4740 8534	0.0290 0730
47	15.4659 1673	0.0646 5831	24.0457 0702	0.0415 8747	37.2320 1217	0.0268 5861
48	16.3938 7173	0.0609 9840	25.7289 0651	0.0388 6679	40.2105 7314	0.0248 6908
49	17.3775 0403	0.0575 4566	27.5299 2997	0.0363 2410	43.4274 1899	0.0230 2693
50	18.4201 5427	0.0542 8836	29.4570 2506	0.0339 4776	46.9016 1251	0.0213 2123

TABLE 4
Log-Tables

LOGARITHAMS

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5566	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6345	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

LOG-TABLES

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7769	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9843	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

TABLE 5
ANTILOGARITHMS

LOG-TABLES

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
.00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
.01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
.02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
.03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
.04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
.05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
.06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
.07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
.08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
.09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
.10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
.11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
.12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
.13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
.14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
.15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
.16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
.17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
.18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
.19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
.20	1585	1289	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
.21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
.22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
.23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
.24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
.25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
.26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
.27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
.28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
.29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
.30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
.31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
.32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
.33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
.34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
.35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
.36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
.37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
.38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
.39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
.40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
.41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
.42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
.43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
.44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
.45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
.46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
.47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
.48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
.49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

LOG-TABLES

	0	1	2	3	4	5	6	7	8	9	Mean Difference								
											1	2	3	4	5	6	7	8	9
.50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
.51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
.52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
.53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
.54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
.55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
.56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
.57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
.58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
.59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
.60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
.61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
.62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
.63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
.64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
.65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
.66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
.67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
.68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
.69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
.70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
.71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
.72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
.73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
.74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
.75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
.76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
.77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
.78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
.79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
.80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
.81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
.82	6607	6622	6637	6653	6668	6683	6699	6715	6730	6745	2	3	5	6	8	9	11	12	14
.83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
.84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
.85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
.86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
.87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
.88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
.89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	12	14	16
.90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
.91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
.92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
.93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
.94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
.95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
.96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
.97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
.98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
.99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

Table 6

Areas under the Standard Normal Probability Distribution between the Mean and Positive Values of z

Example:

To find the area under the curve between the mean and a point 2.24 standard deviations to the right of the mean, look up the value opposite 2.2 and under 0.04 in the table; 0.4875 of the area under the curve lies between the mean and a z Value of 2.24

x	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	0.0753
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	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	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	0.2995	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	04956	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

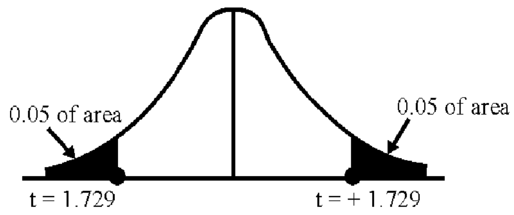


Table 7

Areas in Both Tails Combined for Student's t Distribution

Example:

To find the value of t that corresponds to an area of 0.10 in both tails of the distribution combined, when there are 19 degress of freedom, look under the 0.10 column, and proceed down to the 19 degrees of freedom row; the appropriate t value there is 1.729

Degree of Freedom	Area in Both Tails Combined 0.10	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
Normal Distribution	1.645	1.960	2.326	2.576