CHAPTER - 01
Statistical Data and Descriptive Statistics

Introduction:

Statistics is very old discipline. In the olden days it was used in the administrative departments of the states and the scope was limited. Earlier it was used by governments to keep record of birth, death, population etc., for administrative purpose. **John Graunt** was the first person to make a systematic study of birth and death statistics and the calculation of expectation of life at different age in the 17th century which led to the idea of Life Insurance. In 18th century mathematics was introduced for the purpose of collection, classification, and presentation of data. Today, there is no person can live without the application of statistics.

Origin and Growth of Statistics:

The word **'Statistics'** and **'Statistical'** are all derived from the Latin word 'Status', Italian word 'Stato', German word 'Statistik', or the French word 'Statistique' each of which means a Political State.

Meaning of Statistics:

"Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analyzing data as well as deriving valid conclusions and making reasonable decisions based on the analysis".

DEFINITIONS OF STATISTICS:

(STATISTICS DEFINED IN PLURAL SENSE)

According to A.L. Bowley "Statistics are numerical facts in any department of enquiry placed in relation to each other".

(STATISTICS DEFINED IN SINGULAR SENSE)

According to Wallis & Roberts "Statistics is a branch of science, which provides tools and techniques for decision making in the face of uncertainty".

Characteristics OR Features of Statistics:

Following are the important features of Statistics................

1. **Aggregate of Facts:**

    In statistics single or isolated figures are not considered to be statistics because such figures are unrelated and cannot be compared.

2. **Systematically Arranged:**

    A proper plan should be prepared before collecting the statistical data. Data collected in a haphazard manner would lead to false conclusions. Therefore, data should be collected in a systematic manner.

3. **Statistics are Estimated or Enumerated:**

    According to this feature of statistics, data can be enumerated or estimated. If the numerical statements are precise & accurate, then they can be enumerated.

4. **Statistics are numerically expressed:**

    All statistics are expressed in numerical figures i.e. expressed in numbers and related to quantitative information only. Qualitative aspects do not come under statistics.

5. **Statistical data Collected for a Pre-Determined Purpose:**

    The data should be collected for a pre-determined purpose. The figures are collected with some goal or objective in mind. Without any objective the data collected will be useless.

6. **Placed in relation to Each Other:**

    Numerical information must be related and comparable. Data collected should be uniform in nature. For example, statistics related to number of children born, exports of a country, coal production etc..,

Objectives of Statistics:

There are some of the important Objectives as follows.......

a) To throw light on the complex mass data.
b) To take action on the basis of analyzed data.
c) To draw conclusion on the basis of analyzed data.
d) To estimate & forecast the trends in future from the past data.
e) To ascertain the reliability of statement.
f) To provide the basis for the development of knowledge.

Functions of Statistics:

There are many functions of statistics. Let us consider the important functions are as follows.............

1) Present facts in simple form:

Statistics presents facts and figures in a definite form. That makes the statement logical and convincing than mere description.

2) Reduces the complexity of data:

Statistics simplifies the complex data. The raw data are complex. We make them simple and intelligible by using different statistical measures.

3) Helps in comparison:

Classification and tabulation are the two methods that are used to condense the data. They help us to compare data collected from different sources.

4) Forecasting:

The future is uncertain. Statistics helps in forecasting the trend and tendencies. Statistical techniques are used for predicting the future values of a variable.

5) Testing hypothesis:

Formulating and testing of hypothesis is an important function of statistics. This helps in developing new theories. So statistics examines the truth and helps in innovating new ideas.

6) Helps in formulation of policies:

Statistics helps in formulating plans and policies in different fields. Statistical analysis of data forms the beginning of policy formulations. Hence, statistics is essential for planners, economists, scientists, and administrators to prepare different plans and programs.

7) Derives valid inferences:

Statistical methods mainly aim at deriving inferences from an enquiry. Statistical techniques are often used by scholars, planners, and scientists to evaluate different projects.

Limitations of Statistics:

Statistics are very essential in the context of any economy but, it also have some limitations is given below: ....

1) **Statistics does not study qualitative facts:**

Statistics is applicable to study only numerical aspects of enquiry, which are expressed in terms of quantitative measurements. For qualitative aspects like honesty, poverty, beauty, intelligence etc., any statistical analysis cannot be directly applied. For example, the intelligence of a group of students can be studied on the basis of their marks in a particular examination.

2) **Statistics does not study individual items:**

Statistics does not give any specific importance to the individual items; it deals with an aggregate of objects. When they are taken items individually do not constitute any statistical data and do not serve any purpose for statistical enquiry.

3) **Statistical laws are not exact:**

Statistical laws are not exact like mathematical and physical sciences. Statistical laws are only approximations. Statistical conclusions are not universally true only on an average.

4) **Statistics may be misused:**

Statistics must be used only by experts. Statistical methods are most dangerous tools in the hands of inexpert. The use of statistical tools by the inexperienced and untraced persons might lead to wrong conclusions. Statistics can be easily misused by quoting wrong figures of data.

5) **Statistics is only one of the methods of studying a problem:**

Statistical methods does not provide complete solution to the problems. Because, problems are to be studied by taking the background of the countries culture, philosophy or religion into consideration. Thus the statistical study should be supplemented by other evidences.

6) **Statistical results may be misleading:**

Without any reference, statistical results may provide doubtful conclusions. For example: on the basis of increasing number of prisoners in the prison, it may be conclude that crime is increasing.

Collection of Data:

Statistics are set of numerical data. They relate to a particular characteristic in an enquiry. Collection of facts and figures is the first step in all statistical enquiries or investigation.

The person who collects the data or information is called as the **'investigator' or 'enumerator'.** One who provides the information is called as **'informant' or 'respondent'.**

Meaning of Collection of Data:

The process of collecting, counting and systematic recording of the results of statistical enquiry is called **"Collection of Data".**

Meaning of Data:

Data are the measurements or observations that are collected as a source of information. For example, the number of people in India, the value of sales of a particular product etc.

Types of Data:

Any statistical data can be classified under two categories depending upon the sources utilized. They are,

1. Primary data   **2.** Secondary data

Primary data:

Primary data is one, which is collected by the investigator himself for the first time for the purpose of a specific enquiry or study.

Methods for Collecting of Primary data:

The primary data can be collected by the following methods: ...
a) Direct personal interview
b) Indirect Oral interview
c) Information from correspondents
d) Mailed Questionnaire method
e) Schedules sent through enumerators
f) Indirect interview or Telephone interview

a) <u>Direct Personal Interview:</u>

The investigator personally meets respondents and ask questions to collect necessary information. It is the best suitable method for intensive study of the limited field.

b) <u>Indirect Oral Interview:</u>

In this method, the investigator contacts witnesses or neighbors or friends or some other third parties who are capable for supplying necessary information. This method is suitable whenever direct sources do not exist.

c) Information from Correspondents:

The investigator appoints local or correspondents in different places to collect the information. The main advantage of this method is that it is Cheap and appropriate for extensive investigations. But it may not ensure accurate results because the correspondents are likely to be negligent, prejudiced and biased.

d) Mailed Questionnaire Method:

Under this method a list of questions is prepared and it is sent to all the respondents by post. The list of questions is technically called questionnaire.

Meaning of Questionnaire:

It is a list of questions sent to the respondents by post along with a covering letter to collect certain information from a group. Respondents have to fill in the blank questionnaire and should return the same with in specified time.

e) Schedules sent through Enumerators:

In this method enumerators or interviewers take the schedules, meet the respondents and fill up their replies in their own handwriting.

The difference is made between the schedule and a questionnaire.   A schedule is filled by the interviewers in a face-to-face situation with the respondent. A questionnaire is filled by the respondent which he receives and returns by post.

f) Indirect interview or Telephone Interview:

In this method of collecting information consists of contacting on telephone or mobile phones itself.

Secondary Data:

Secondary data is the data which have been already collected and analyzed by someone or their own use and later the same data are used by others.

Sources of Secondary Data:

The sources of secondary data can broadly classified under two heads:

1. Published Sources and

2. Unpublished Sources.

Published Sources:

The various sources of published data are reports and official publications of international organizations (IMF, World Bank, UNO), central and state governments, local bodies. Financial and economic journals, Annual reports, research agencies, research scholars, Govt., agencies, websites in internet, etc.

Unpublished Sources:

All statistical material is not always published. There various sources of unpublished data such as records maintained by various Government and private offices, studies made by research institutions, scholars, etc.

Classification of Data

Introduction:

After collection of data, the data may be in the form of schedules and questionnaires. But, it is not possible to understand easily. So, it requires grouping of uniform data on certain basis.

Meaning of Classification:

The process of arranging data into homogenous groups or classes according to some common characteristics present in the data is called classification.

(OR)

The process of dividing the data into different groups which are homogeneous in nature.

For Example: The letters in the post office are classified according to areas, places, streets etc..,

Frequency:

The frequency is the number of times a particular value for a variable (data item) has been observed to occur. It is denoted by **'f '.**

Frequency Distribution:

The way of tabulating a group of data of a variable and their respective frequencies side by side is called 'Frequency distribution'.

Frequency Table:

A frequency table is a simple way to display the number of occurrences of a particular value or characteristic.

Problems under Frequency distribution table:

Problem No: 01

Jaya is asked to collect data for size of shoes in her class. Prepare a frequency distribution table using tally marks.

5	4	7	5	6	7	6	5	6	6	5	4	5
6	8	7	4	6	5	6	4	6	5	7	6	7
5	7	6	4	8	7							

Problem No: 02

Form a frequency distribution from the following data by inclusive method taking five as the width of the class.

42	62	46	54	41	37	54	44	32	45	47	50	58
49	51	42	46	37	42	39	54	39	51	58	47	64
43	48	49	48	49	61	41	40	58	49	59	57	57
34	56	38	45	52	46	40	63	41	51	41		

Problem No: 03

Form a frequency distribution from the following data by inclusive method taking four as the width of the class.

40	62	48	54	42	37	55	45	38	46	47	50	58
49	51	42	46	38	42	38	54	39	51	58	47	66
43	48	49	48	49	62	45	40	58	48	59	57	58
34												

Problem No: 04

Form a frequency distribution from the following data by exclusive method taking five as the width of the class.

42	62	46	54	41	37	54	44	32	45	47	50	58
49	51	42	46	37	42	39	54	39	51	58	47	64
43	48	49	48	49	61	41	40	58	49	59	57	57
34	56	38	45	52	46	40	63	41	51	41		

Problem No: 05

Form a frequency distribution from the following data by exclusive method taking four as the width of the class.

10	17	15	22	11	16	19	24	29	18
25	26	32	14	17	20	23	27	30	12
15	18	24	36	18	15	21	28	33	38
34	13	10	16	20	22	29	19	23	31

Meaning and Definition of Average:
A single value which can be considered as typical or representative of a set of observations and around which the observations can be considered as Centered is called as an Average.

According to **Clark** defines as, "Average is an attempt to find one single figure to describe the whole of figure".

Types or Kinds of Averages:
  A. Mathematical Averages
    1. Arithmetic mean
        a. Simple arithmetic mean
        b. Weighted mean
    2. Geometric mean
    3. Harmonic mean

  B. Positional Averages
    1. Median
    2. Mode

Meaning and Definition of Arithmetic Mean:
Arithmetic mean refers to the sum of the values of a group of items divided by number of items. It is also called as Simple arithmetic mean. It is denoted by X (X Bar).

According to **Horace Sacrist** defines as, "Arithmetic mean is the amount secured by dividing the sum of values of the items in a series by their number".

Merits of Arithmetic Mean:
Following are merits of arithmetic mean: ....
    1) It is easy to calculate.
    2) It is simple and easy to understand.
    3) It calculated on the basis of all the items of the series.
    4) It can be easily used for comparison.

Demerits of Arithmetic mean:
Following are de-merits of arithmetic mean: ....
    1) It cannot be computed for qualitative aspects like intelligence, honesty etc.,
    2) It is very much affected by extreme values.
    3) Sometimes the averages may not be actual item in the series.
    4) Value of mean cannot be graphically ascertained.

Problems under Mean
Calculation of Arithmetic mean under individual Series
Problems under Direct Method

Formula:

$$\overline{X} = \frac{\sum X}{N}$$

Problem No: 06
Calculate the arithmetic mean:

Wages: 200, 220, 190, 160, 140, 400, 500.

Problem No: 07
Calculate mean from the following:

Roll No.	1	2	3	4	5	6	7	8
Age	21	18	19	21	20	17	22	23

Problems under Indirect Method

$$\overline{X} = A + \frac{\sum d}{N}$$

Problem No: 08
From following information calculate Arithmetic mean:
Marks: 75, 68, 80, 92, 56.

Problem No: 09
From following information calculate Arithmetic mean:
Variables: 82, 45, 68, 89, 74, 62, 58, and 94.

Problems under Step Deviation Method

Formula $\overline{X} = A + \frac{\sum d'}{N} \times C$

Problem No: 10
The following table gives the marks of 10 Students
Marks: 30, 70, 40, 20, 60, 40, 30, 80, 50, and 90.

Problem No: 11
The following table gives the marks of 10 Students
Marks: 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50.

Calculation of Arithmetic mean under Discrete Series

Problems under Direct and Indirect Method

Formula under direct Method $\quad \overline{X} = \dfrac{\Sigma fx}{N}$

Formula under Indirect Method $\quad \overline{X} = A + \dfrac{\Sigma fd}{N}$

Problem No: 12

From the following information calculate Arithmetic mean under direct and indirect method:

Marks	64	63	62	61	60	59
No. of Students	8	18	12	9	7	6

Problems under Step Deviation Method

Formula under Step deviation Method $\quad \overline{X} = A + \dfrac{\Sigma fd'}{N} \times C$

Problem No: 13

From the following information calculate Arithmetic mean:

Marks	64	66	68	70	72
No. of Students	4	12	4	8	2

Problem No: 14

In a class the number of books issued to the students from the library were as follows:

No. of Books issued	0	1	2	3	4	5	6	7
No. of Students	7	12	10	7	6	4	3	1

Find the average number of books issued to a student of the class.

Calculation of Arithmetic mean under Continuous Series

Problems under Direct Method

Formula under direct Method $\quad \overline{X} = \dfrac{\Sigma fm}{N}$

Problem No: 15
Calculate the arithmetic mean

Height (in cm's)	0 - 10	10 - 20	20-30	30-40	40-50	50-60	60-70
No. of Students	2	4	6	8	10	12	14

Problems under Deviation or Indirect Method

Formula under Indirect Method $\quad \overline{X} = A + \frac{\Sigma fd}{N}$

Problem No: 16
From the following calculate average height of the students

Height (in cm's)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of Students	5	12	30	45	50	37	21

Problems under Step Deviation Method
Formula under Step deviation Method $\overline{X} = A + \frac{\Sigma fd'}{N} \times C$

Problem No: 17
From the following information calculate Arithmetic Mean

Weights (in Kgs)	No. of Persons
0 – 10	5
10 – 20	10
20 – 30	25
30 – 40	30
40 – 50	20
50 – 60	10

Problem No: 18
From the following information find out Arithmetic Mean

Marks below	10	20	30	40	50	60	70	80
No. of Students	15	35	60	84	96	127	198	250

Problem No: 19
Calculate the average marks from the following information.

Marks	Below 5	5-10	10-15	15-20	20-25	25-30	30-35
No. of Students	2	4	5	3	2	4	5

Problem No: 20
Calculate the Average weekly wages from the following data.

Weights (in Kgs)	No. of Persons
100 & above	50
200 & above	46
300 & above	36
400 & above	18
500 & above	6
600 & above	3
700 & above	1

Meaning of Weighted Mean:

Weighted mean refers to the average whose component items are being multiplied by weights and the aggregate of the multiplied results are being divided by the total of weights.

Problems for Calculating Weighted Mean

$$\overline{X}w = \frac{\sum Wx}{\sum w}$$

Problem No: 21
Rain fall for a month as recorded in an observatory is as follows:

Days	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Rain fall in mm	87	63	74	56	62	85	73

If there are 5 Mondays, Tuesdays, and Wednesdays, 4 Sundays, Thursdays, Fridays and Saturdays, during that month, calculate the daily average rain fall during the month.

Problem No: 22
A candidate obtains the following marks in PUC Exam in 2022.

Subjects	Marks
English	60
Kannada	65
Physics	80
Chemistry	70
Mathematics	85
Biology	80

Find out weighted mean if weights are 1, 1, 3, 2, 4, and 1 respectively.

Meaning of Median:

The median is the middle value of the distribution. It divides the distribution into two equal parts. All the values with one part will greater than the median and all the values with other part will be less than median.

Definition:

According to **Connor** defines as, "The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other all values lesser than the median".

Merits or Advantages of Median:

Following are the merits of Median as follows: ....

a) It is easy to calculate and simple to understand.
b) It is rigidly defined.
c) It can be found by mere inspection.
d) It can be obtained graphically.

Demerits of Median:

Following are the demerits of Median as follows: ....

a) A slight change in the series may bring drastic change in median value.
b) It is not capable of future algebraic treatment.

Problems under Median:
Calculation of Median under Individual Series:

Formula: **Med = Size of [$\frac{N+1}{2}$]th item**

Where, N = Number of items

Problem No: 23

The following figures represent the number of books issued at the counter of a library on different days. 96, 180, 98, 75, 270, 80, 102, 100, 94, 75, and 200. Calculate the Median.

Problem No: 24

The marks of 10 students are given as under:

Roll No	1	2	3	4	5	6	7	8	9	10
Marks	40	38	60	62	65	35	20	80	86	90

Calculate Median.

Calculation of Median under Discrete Series:

Formula: **Med = Size of [$\frac{N+1}{2}$]th item**

Problem no: 25
Locate the median in the following distribution.

Size   : 8   10   12   14   16   18   20
Frequency: 3   7   12   28   10   9   6

Problem No: 26
Locate the median in the following distribution.

Size   : 8   12   16   16   18   20   24
Frequency: 3   7   14   26   10   9   8

Calculation of Median Under Continuous Series:

Formula: 1. **Med = Size of [$\frac{N}{2}$]th item**

$$2.\ \textbf{Med = L}_1 + \frac{\frac{N}{2} - cf}{f} \times C$$

N = Total Frequency
L_1 = lower limit of the median class
cf = cumulative frequency of the class preceding the median class
 f = simple frequency of the median class
C = width of the median class ($L_2 - L_1$ of the median class)

Problem No: 27
The following table gives the frequency distribution of 325 workers of a factory, according to their average monthly income in a certain year.

Income Group	Number of workers
Below 100	1
100-150	20
150-200	42
200-250	55
250-300	62
300-350	45
350-400	30
400-450	25
450-500	15
500-550	18
550-600	10
600 and above	2
Total	**325**

Calculate Median income.

Problem No: 28

Calculate Median from the following data:

Marks	No. of students
1-5	6
6-10	9
11-15	15
16-20	12
21-25	8

Problem No: 29

Following are the daily wages of workers in a cement industry. Find the Median.

Wages (in Rs.)	Number of workers
Less than 100	5
Less than 200	12
Less than 300	20
Less than 400	32
Less than 500	40
Less than 600	45
Less than 700	52
Less than 800	60
Less than 900	68
Less than 1000	75

Problem No: 30

Compute the Median from the following information:

Wages (in Rs.)	Number of workers
More than 10	100
More than 20	98
More than 30	85
More than 40	70
More than 50	45
More than 60	30
More than 70	15
More than 80	8
More than 90	5

Problem No: 31

Calculate the Median from the following data:

X	F
Less than 5	7
Less than 10	20
5-15	38
15 and above	55
20-25	20
25 and above	5
30 and above	1

Compute Median from the following data:

Mid-Value	:	5	15	25	35	45	55	65	75
Frequency	:	7	10	15	17	8	4	6	7

Meaning of Mode:

It is a positional average. It shows the centre of concentration of the frequency around a given value. Where the purpose is to know the point of the highest concentration it is preferred.

In other words, the observation which occurs most frequently is known as Mode. It is denoted by **Z.**

Definition of Mode:

According to **A M Tuttle** defines as, "Mode is the value which has the greatest frequency or density in its immediate neighborhood".

Merits of Mode:

Following are the important Merits of Mode: .....
   a) It is easy to calculate and simple to understand.
   b) It can be calculated for open-end classes.
   c) It is an actual value of an important part of the series.
   d) The value of mode can also be determined graphically.
   e) In individual or discrete distribution it can be located by mere inspection.
   f) It is not affected by sampling fluctuations.

De-merits of Mode:

Following are the important demerits of Mode: .....
   a) It is not based on all observations.
   b) It is not capable of further mathematical treatment.
   c) In some cases Mode is ill-defined.
   d) It is unsuitable in cases where relative important of items has to be considered.

Steps for Preparation of Grouping and Analysis Table:

Grouping table is prepared with **6** columns:
Column I : The actual frequency (simple frequency) is taken.
Column II : Two frequencies are combined as one group from the beginning.
Column III: Two frequencies are combined as one group leaving the first frequency.

Column IV: Three frequencies are combined as one group from the beginning.
Column V : Three frequencies are combined as one group leaving the first frequency.
Column VI: Three frequencies are combined as one group leaving the first two frequencies.
We have to mark the highest totals in all six columns.

Analysis Table:
An analysis table is prepared out of grouping table. In this table, the columns are recorded in rows and variables/class intervals are arranged into columns. The highest totals marked in six columns of grouping table, are marked against variables/class intervals with the help of tally bars. Find the totals of all 6 columns. The variable or class intervals with highest tally bars is considered Mode or modal class.

Problems under Mode
Problems under Individual Series:
Under individual series the Mode can be located by inspection method. The mode can be obtained by simply identifying the item that occurs most frequently.
Mode = The value which is repeated for more number of times.

Problem No: 33
Calculate mode from the following data:
Marks: 20, 24, 15, 20, 10, 20, 28, 20, 15, 20, 12, 20

Problem No: 34
Calculate mode from the following:
Age (in years): 7, 10, 15, 12, 7, 14, 24, 10, 7, 20, 10

Calculation of Mode under Discrete Series:
Under discrete series Mode can be calculated by inspection method if the highest frequency is well defined. Whereas, the highest frequency is repeated, the mode cannot be calculated by inspection method.

Mode = The value which has the highest frequency.

Problem No: 35
Calculate Mode:

Height :	2	3	5	6	7	8	4
Frequency:	12	23	45	50	33	15	10

Problem No: 36
Determine the Mode for the following Group:

Variable	28	30	32	34	36	38	40	42	44
Frequency	10	12	16	14	10	8	17	5	4

Problem No: 37
Find Mode by using grouping and analysis tables from the data given below:

Marks	No. of Students
20	40
30	50
40	70
50	90
60	85
70	91
80	89
90	65

Problems under Continuous Series:

In continuous series, by the inspection method we have to identify the highest frequency. The corresponding class interval is called as modal class. Then we have to apply the formula to get exact value of Mode.

Formula: $Z = L_1 + \dfrac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$

L_1= Lower limit of the model class
f_1= Frequency of model class
f_0= Frequency of the class preceding the modal class
f_2= Frequency of the class succeeding the modal class
C= Width of the modal class ($L_2 - L_1$)

Problem No: 38
Find the Mode of the following series:

Wages (Rs.)	No. of Persons
Below 250	10
250 – 500	30
500 – 750	40
750 – 1000	25
1000 – 1250	20
Above 1250	15

Problem No: 39

Form a Continuous frequency table from the following data, using class intervals such as 40-50, 50-60 etc., and determine the Model value.

90	78	86	51	96	104	51	78	50	72
76	49	77	90	84	76	42	72	69	70
68	69	104	80	79	54	79	73	58	91
65	60	77	78	67	50	84	76	110	53
74	40	60	42	82	41	61	75	115	81

Problem No: 40
Calculate Mode for the following:

Class Intervals	Frequency
0-50	5
50-100	14
100-150	40
150-200	91
200-250	150
250-300	87
300-350	60
350-400	38
400 and Above	15

Problem No: 41

Calculate Mode for the following frequency distribution.

C I	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
No. of students	29	195	241	117	52	10	6	3	2

Problem No: 42

Calculate mode by using grouping and analysis table:

Expenditure	0-10	10-20	20-30	30-40	40-50
No. of Families	14	23	27	21	15

Problem No: 43

Calculate Mean, Median and Mode from the following:

Roll No	:	1	2	3	4	5	6	7	8
Age	:	21	18	19	21	20	17	22	23

Problem No: 44

Compute the Mean, Median and Mode from the following data:

MV	115	125	135	145	155	165	175	185	195
F	6	25	48	72	116	60	38	22	3

Problem No: 45
　　Calculate the value of Mean, Median and Mode from the following data:

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	14	23	35	20	8	5

Problem No: 46
　　Calculate the value of Mean, Median and Mode from the following data:

Profits (in lakhs)	4-7	8-11	12-15	16-19	20-23	24-27	28-31	32-35	36-39
Frequency	6	10	18	30	15	12	10	6	2

Problem No: 47
　　From the following data, draw a less than ogive curve and locate the value of Median graphically.

Marks	No. of students
Less than 20	5
Less than 30	13
Less than 40	24
Less than 50	39
Less than 60	52
Less than 70	60

Empirical Relationship Between Averages
　　In a symmetrical distribution the three simple averages mean = median = mode. For a moderately asymmetrical distribution, the relationship between them are brought by Prof. Karl Pearson as **Mode = 3Median – 2 mean.**

Problem No: 48
If mean = 38.2, Median = 41.6, find Z?

Problem No: 49
Find Median if A.M. = 12 and Z = 13.

Problem No: 50
　　For a distribution the value of Mode is 28 and Median 30. What will be the value of Arithmetic?

CHAPTER – 02

MEASURES OF VARIATION AND SKEWNESS

Introduction:

The measure of central tendency serves to locate the center of the distribution, but they do not reveal how the items are spread out on either side of the center. This characteristic of a frequency distribution is commonly referred to as dispersion. In a series all the items are not equal. There is difference or variation among the values. The degree of variation is evaluated by various measures of dispersion. Small dispersion indicates high uniformity of the items, while large dispersion indicates less uniformity.

Meaning of Dispersion:

Dispersion or Variation is the measurement of the scatter of the size of the items of a series about the average.

Definition of Dispersion:

According to **A.L. Bowley** defines as, **"**Dispersion is a measure of the items**".**

Objectives of measure of Dispersion:

Following are the important objectives of Dispersion:
   a) To use other statistical methods
   b) To compare variability
   c) To test reliability of average
   d) To establish trends in time series
   e) To control the undesired variability.

Features or Characteristics of a good measure of Dispersion:

An ideal measure of dispersion is expressed to possess the following:
   1) It should be simple to understand
   2) It should be easy to compare
   3) It should have sampling stability
   4) It should be unduly affected by extreme items
   5) It should be amenable to further algebraic treatment
   6) It should be based on each and every item of the distribution.

Types of measures of Dispersion:
There are two kinds of measures of dispersion: ....
1. Absolute measure of dispersion
      a) The Range
      b) Quartile deviation
      c) Mean deviation
      d) Standard deviation

2. Relative measure of dispersion
      a) Co-efficient of range
      b) Co-efficient of quartile deviation
      c) Co-efficient of mean deviation
      d) Co-efficient of variation

Absolute measures of dispersion:
Absolute measure of dispersion are expressed in same units in which original data is presented but these measures cannot be used to compare the variation between the two series.

Relative measures of dispersion:
Relative measures are not expressed in units but it is a pure number. It is the ratios of absolute dispersion to an appropriate average such as co-efficient of standard deviation or co-efficient of mean deviation.

Meaning of Range:
The range is the most obvious measure of dispersion and it is the difference between the lowest and highest values in a dataset.
Formula   **Range = L – S**
L = Largest value   **S =** Smallest value

Meaning of Co-efficient of Range:
It is relative measure of dispersion and is based on the value of range. It is also called Range co-efficient of dispersion.
$$\text{Range} = \frac{L - S}{L + S}$$

Calculation of Range and Co-efficient of Range under individual Series:
Problem No: 01
Calculate Range and its co-efficient from the following data:
      53, 46, 18, 16, 75, 84, 28

Problem No: 02

Calculate Range and its co-efficient from the following data:

  10, 5, 14, 25, 3, 8, 6

Calculation of Range and its Co-efficient of Range under Discrete Series:

Problem No: 03

Calculate Range and its co-efficient from the following data:

Variables : 5, 10, 4, 20, 15, 22
Frequency: 4, 8, 12, 15, 10, 5

Problem No: 04

Calculate Range and its co-efficient from the following data:

Height   : 2,   3,   5,   6,   7,   8, 4
Frequency: 12, 23, 45, 50, 33, 15, 10

Calculation of Range and its Co-efficient of Range under Continuous Series:

Problem No: 05

Calculate Range and its co-efficient from the following data:

X : 40-50   50-60   60-70   70-80   80-90
F :   8   24   46   22   10

Problem No: 06

Calculate Range and its co-efficient from the following data:

Wages (Rs.)   : 110-140 140-160 160-180 180-200 200-220
No. of workers :   10   15   20   30   15

Quartile Deviation:

   It is based on the lower quartile (Q_1) and the upper quartile (Q_3). The differences $Q_3 Q_1$ is called the inter quartile range. The difference between (Q_3 and Q_1) $Q_3 - Q_1$ divided by 2 is called as semi-inter-quartile range or the Quartile deviation.

   Formula:

   Quartile Deviation = $\dfrac{Q_3 - Q_1}{2}$

   Co-efficient of Quartile Deviation = $\dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

Q_1 = Size of $\left[\dfrac{1(N+1)}{4}\right]$ th item

Q_3 = Size of $\left[\dfrac{3(N+1)}{4}\right]$ th item

Steps to be followed:

1. Arrange the in Ascending or Descending order.
2. Apply the formula, Q_1 = Size of $\frac{[1(N+1)]}{4}$ th item
3. Value found through above formula is Q_1
4. If the value is in fraction (like 2.75^{th} item), the value of Q_1 will be 2nd item + 0.75 (3rd item – 2nd item)
5. Apply the formula, Q_3 = Size of $\frac{[3(N+1)]}{4}$ th item
6. Value found thorough above formula is Q_3
7. If the value is in fraction (8.25^{th} item), the value of Q_3 will be 8th item + 0.25 (9th item – 8th item)
8. Apply the Quartile deviation formula.

Calculation of Quartile Deviation Under Individual Series:
Problem No: 07
Compute the quartile deviation from the following data:
37, 78, 86, 91, 93, 94, 35, 42, 44, 53, 55, 57

Problem No: 08
Compute the quartile deviation from the following data:
Marks: 15, 30, 37, 43, 50, 54, 61, 65, 72, 80, 90, 100

Calculation of Quartile Deviation under discrete Series:
Problem No: 09
Calculate QD and its co-efficient:
Values: 10, 20, 30, 40, 50, 60
F   : 28, 36, 24, 32, 40, 16

Problem No: 10
Calculate QD and its co-efficient:

Size	2	4	6	8	10	12	14	16	18	20
F	2	9	11	14	20	24	20	16	5	2

Calculation of Quartile Deviation under continuous Series:
Formula:

1. Q_1 = **Size of** $\frac{[1(N)]}{4}$ th item and $Q_1 = L_1 + \frac{\frac{N}{4} - cf}{f} \times C$

2. Q_3 = **Size of** $\frac{[3(N)]}{4}$ th item and $Q_3 = L_1 + \frac{\frac{3N}{4} - cf}{f} \times C$

Problem No: 11
Calculate quartile deviation:

Value	F
15-25	32
25-35	38
35-45	45
45-55	98
55-65	122
65-75	80
75-85	50
85-95	25

Problem No: 12
Calculate quartile deviation:

Weekly wages (Rs.)	No. of workers
0-20	20
20-40	30
40-60	50
60-80	60
80-100	80
100-120	40
120-140	20

Mean Deviation and co-efficient of Mean deviation:

The mean deviation is the first measure of dispersion that actually uses each data value in its computation. It is the mean of the distances between each value and the mean.

The difference (X – average) is called deviation and when we ignore the negative sign, this deviation is written as |D| and is read as Modular D.

Co-efficient of Mean Deviation:

A relative measure of dispersion based on the mean deviation is called as co-efficient of mean deviation.

Formula: Mean Deviation = $\dfrac{\sum|D|}{N}$

Co-efficient of M D = $\dfrac{\text{MD from Mean}}{\text{Mean}}$

Calculation of Mean Deviation under individual Series:
Problem No: 13
Calculate the mean deviation from the following data:
Variables: 68, 49, 32, 21, 54, 38, 59, 66, 41

Problem No: 14
Calculate mean deviation (from mean) and it's co-efficient.
Marks: 59, 32, 67, 43, 22, 17, 64, 55, 47, 80, 25

Problem No: 15
Calculate mean deviation (from median) and it's co-efficient.
Wages (Rs.): 400, 200, 150, 440, 460, 700, 120

Problem No: 16
Calculate mean deviation (from mode) and its co-efficient.
Marks: 58, 35, 62, 25, 48, 15, 62, 50, 49, 85, 27, 13

Calculation of Mean Deviation under Discrete Series and continuous Series:
Formula: Mean Deviation = $\dfrac{\sum f|D|}{N}$

Problem No: 17
Compute the mean deviation and co-efficient from mean.

Marks:	10	20	30	40	50	60
No. of students:	18	22	35	20	13	12

Problem No: 18
Find out the co-efficient of mean deviation from the following with median as base.

X:	4	6	8	10	12	14	16
F:	10	13	22	24	15	12	9

Problem No: 19
Calculate the mean deviation from the following data with mode as base.

X:	10	20	30	40	50	60
F:	4	7	15	8	7	2

Problem No: 20
Calculate average deviation and its co-efficient (from mean & median)

Marks:	0	5	10	15	20	25	30	35
Students:	8	16	22	35	15	5	3	1

Problem No: 21
Calculate mean deviation and co-efficient of mean deviation from mean.

CI :	0-10	10-20	20-30	30-40	40-50
F :	5	10	15	10	5

Problem No: 22

Compute mean deviation from median:

Wages (in Rs.)	No. of workers
10-20	8
20-30	10
30-40	15
40-50	25
50-60	20
60-70	18
70-80	9
80-90	5

Problem No: 23

Calculate mean deviation from mean of the following:

Marks less than	No. of students
10	4
20	10
30	20
40	40
50	48
60	55
70	60

Problem No: 24

Calculate the mean deviation from Median from the following data:

Size	F
More than 10	3
More than 20	11
More than 30	26
More than 40	46
More than 50	81
More than 60	93
More than 70	100

Problem No: 25

Calculate mean deviation from median:

C I	F
0-10	5
10-20	8
20-30	12
30-40	15
40-50	20
50-60	14
60-70	12
70-80	6

Meaning and Definition of Standard Deviation:

It is defined as the positive square-root of the arithmetic mean of square of the deviations of the given observations from their arithmetic mean.

It is denoted by the Greek letter **σ** (sigma).

Merits of Standard Deviation:

Following are the merits of standard deviation: …….

1) Standard deviation is rigidly defined and its value is always definite.
2) It is possible to calculate the combined standard deviation of two or more groups.
3) It is based on all the observation of the data.
4) It is less affected by the fluctuations of sampling than most other measures of dispersions.

De-merits of Standard Deviation:

Following are the demerits of standard deviation: …….

1) It is not easy to calculate and understand.
2) As it is an absolute measure of variability, it cannot be used for the purpose of comparison.

Problems for Calculating Standard Deviation under Individual Series.

(a) Deviations are taken from actual mean:

Problem No: 26

From the following information find the most consistent batsman and better run getter: ……….

Batsman A:	5	7	16	27	39	53	56	61	80	101	105
Batsman B:	0	4	16	21	41	43	57	78	83	90	95

Problem No: 27

The following are the runs scored by two batsman A and B in ten innings.

Batsman A:	101	0	27	82	36	45	13	7	65	14
Batsman B:	97	40	12	13	96	8	8	85	56	15

Who is the better run getter? Who is the more consistent batsman?

Problem No: 28
Calculate the Standard Deviation.

Values	73	16	10	80	90	40	15	10	60	46

Problem No: 29
Calculate the Standard Deviation from the following data.
14, 22, 9, 15, 20, 17, 12, 11

(b) Deviations taken from assumed mean:

Problem No: 30
Calculate Standard deviation for the following data:
Marks: 5, 10, 20, 25, 40, 42, 45, 48, 70, 80

Problem No: 31
     The table below gives the marks obtained by 10 students in statistics: Calculate Standard deviation.

Roll Number:	1	2	3	4	5	6	7	8	9	10
Marks :	43	48	65	57	31	60	37	48	78	59

Calculation of Standard Deviation under Discrete Series

Problem No: 32
Find the S.D & C.V from the following information:

Age [in years] :	10	11	12	13	14	15	16	17
No. of Students:								
Boys:	1	4	7	12	15	15	5	3
Girls:	1	7	13	10	12	12	2	1

Problem No: 33
Find the Standard Deviation from the following series:

X:	20	22	25	31	35	40	42	45
F:	5	12	15	20	25	14	10	6

Problem No: 34
Goals scored by two teams in football session were as follows:

No. of Goals Scored:	0	1	2	3	4	5
Team A :	15	10	7	5	3	2
Team B :	20	10	5	4	2	1

Which Team is more variable?

Problem No: 35

Calculate Standard deviation from the following data

Marks	:	10	20	30	40	50	60
No. of Students:		8	12	20	10	7	3

Calculation of Standard Deviation under Continuous Series:

Problem No: 36

Calculate Standard Deviation

CI	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F	25	30	60	80	50	15	10

Problem No: 37

Calculate S.D for the following data.

Midpoint	1	2	3	4	5	6	7	8	9
Frequency	6	45	103	147	205	160	80	40	4

Problem No: 38

A dealer received samples of tires from two suppliers X and Y. Had the samples tested for length of life with the following results?

Length of life in months	Supplier X	Supplier Y
4-8	10	2
8-12	16	42
12-16	30	12
16-20	4	4

Which supplier's tyres are more uniform regarding their length of life?

Problem No: 39

Calculate Standard deviation from the following:

C I	15	25	35	45	55	65	75	85
F	18	22	30	50	45	30	20	15

Problem No: 40

Calculate S.D. for the following data.

Age under: (in years)	10	20	30	40	50	60	70	80
No. of Persons	10	40	85	125	160	200	220	280

Problem No: 41

From the following data state which of the two series is more variable.

Variable	Frequency A	Frequency B
10-20	10	18
20-30	18	22
30-40	32	40
40-50	40	32
50-60	22	18
60-70	18	10

Problem No: 42

Two brands of tyres are tested for their life, the following results were obtained. ....

Life (in months)	20-25	25-30	30-35	35-40	40-45
Brand X	1	22	64	10	3
Brand Y	3	21	74	1	1

If consistency is the criterion which brand of tyres would you prefer?

Problem No: 43

Find out by calculating co-efficient of variation, which of the following two series has more variation?

Age in years	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Population (000's)CityA:	18	16	15	12	10	5	2
CityB:	10	12	24	32	29	11	3

Problem No: 44

Following information is related to the marks scored in a test of Girls section and Boys section. From this you required to calculate:
(A) Which section scores more marks in an average?
(B) Which section has more consistency in scoring?

Marks:	0-10	10-20	20-30	30-40	40-50
Boys section:	9	13	20	10	8
Girls section:	5	12	24	12	7

Problem No: 45

Your given below daily wages paid to workers in two factories A & B.
[A] Find which factory pays higher wages?
[B] In which factory wages are more variable?

Daily wages:	50-60	60-70	70-80	80-90	90-100
No. of workers:					
Factory A:	15	30	45	20	10
Factory B:	20	35	50	10	5

Calculation of C.V (Co-efficient Variance)

Problem No: 46

If σ = 9.87, Mean= 50 find C.V.

Problem No: 47

If σ = 100.23, Mean= 129.25 find C.V.

Problem No: 48

Find Mean if C.V=40% and variance =144

Problem No: 49

Find C.V if Variance = 144, Mean=56.

Problem No: 50

(a) Distribution "A" has the same standard deviation of "B". Comment on their variability. (b) Coefficient of variation of series is 40% and its variance is 144. What is its mean?

	A	B
Mean	200	175
S.D	30	30

Problem No: 51

The **A M** of runs scored by 3 batsman A, B, and C in a series of 10 innings are 50, 48 and 12 respectively. The **S.D** of their runs are 15, 12, and 2 respectively. Find who are more consistent batsman?

Problem No: 52

Co-efficient of variation of two series is 30% and 50% and their variance is 100 and 144 respectively. What are their arithmetic means?

Problem No: 53

What is C.V given that mean = 56, variance = 144, of 60 items.

Problem No: 54

In a distribution variance = 121 and coefficient of variance =23.16%. Find the value of mean of the distribution.

Problem No: 55

Coefficient of variation of a series is 58% and its standard deviation is 21.2. Find mean.

Calculation of Karl Pearson's coefficient of skewness

(a) Calculation of Karl Pearson's Coefficient of skewness under Individual Series:

Problem No: 56
Compute Karl – Pearson's coefficient of skewness for the following data.

      25, 15, 23, 40, 27, 25, 23, 25, 20.

Under Discrete Series:

Problem No: 57
Find the coefficient of skewness form the data given below.

Size:	3	4	5	6	7	8	9	10
Frequency:	7	10	14	35	102	136	43	8

Problem No: 58
    Find Mean, Mode, Standard deviation and Karl Pearson's coefficient of skewness.

No. of Children per family	0	1	2	3	4	5	6
No. of families	7	10	16	25	18	11	8

Under Continuous Series

Problem No: 59
Calculate Karl Pearson's coefficient of skewness from the following series, where mode is ill defined.

Marks	No. of Students
Above 0	150
Above 10	140
Above 20	100
Above 30	80
Above 40	80
Above 50	70
Above 60	30
Above 70	14
Above 80	0

Problem No: 60
Calculate Karl Pearson's co-efficient of skewness of the following data:

Wages(Rs)	40-50	50-60	60-70	70-80	80-90	90-100	100-110
No. of Workers	8	20	36	24	16	14	2

Problem No: 61
Find Karl Pearson's co-efficient of skewness for the following data:

Marks	: 0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of Students:	5	12	16	25	10	8	4

Problem No: 62
Calculate Karl Pearson's co-efficient of skewness from the following frequency distribution of profits earned by 100 partnership firms during the year 2013.

Profits (in Rs. 000's)	No. of Partnership firms
More than 10	100
More than 20	97
More than 30	90
More than 40	70
More than 50	40
More than 60	25
More than 70	15
More than 80	08
More than 90	03

Problem No: 63
Find Karl Pearson's coefficient of skewness for the given distribution:

X:	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
F:	2	5	7	13	21	16	8	3

Problem No: 64
Find out Karl Pearson's co-efficient of skewness from the following.

Wages	No. of Persons
270 – 280	12
280 – 290	18
290 – 300	35
300 – 310	42
310 – 320	50
320 – 330	45
330 – 340	20
340 – 350	08

Problem No: 65

Calculate Pearson's coefficient of skewness from the following data.

Life time (in hours)	No. of tubes
300 – 400	25
400 – 500	56
500 – 600	60
600 – 700	75
700 – 800	48
800 – 900	30
900 – 1000	15

Problem No: 66

If Med=17.4, SKp= 0.35, Z= 15.3, Find the CV?

Problem No: 67

If CV = 5%, Skp= 0.5, S.D. = 2

Find the value of mean and mode.

Problem No: 68

Given CV = 30%, Mean = 25, Mode=16, Find Skp.

Problem No: 69

Given Mean = 20, Median= 18.5, and CV= 30% find Karl Pearson's coefficient of skewness.

Problem No: 70

If Σfx=350, N=10, Median=38, and Variance=49. find Skp.

CHAPTER – 03
PROBABILITY DISTRIBUTIONS

Introduction:

Theory of probability was found in the 17th century by the two French Mathematicians called **Blaise Pascal** and **Pierre de Fermat.** It was made applicable in games and gambling. It is the branch of mathematics concerned with probability, which deals with the analysis of random phenomena.

A probability is a numerical statement about the likelihood that an event will occur. Probability is expressed in numbers between **0** and **1.** Probability = 0 means the never happens; Probability = 1 means it always happens.

Meaning of Probability:

Probability is defined as the chance or the likelihood that an event will happen.　　　　　　　　**OR**

Probability is the study of the chance that a particular event or series of events will occur.

Approaches of Probability:

There are four important approaches of probability: ………

1. **Classical approach:**
   This method is also called as theoretical or prior approach. The probabilities of events are known well in advance. All the outcomes are equally likely or equal probable. They are derived logically prior to the conduct of any experiment.
   For example: Find the probability of getting a tail in tossing of a coin. **Solution:** Sample space S = {H, T} ….n = 2
   　　　　　Event 'A'　　　= {T}
   　　　　　Therefore **P** (T) 　= 1/2
2. **Statistical approach:**
   This method is also called as empirical or relative approach. The probabilities of events are not known well in advance.
   They are based on observed data or experiment.
3. **Subjective approach:**
   It is also called as persona-list approach. The probabilities are based on individual judgments or opinions or guess about the occurrence of an event.
   This approach has been extensively studied, especially with regards to gambling and securities markets. Under this approach, the probabilities are assigned to each event by personal opinion.

4. Axiomatic approach:

It is also called as statement or objective approach. Generally a few recognized statements are traditionally used in studying probabilities.

Importance of Probability:

Following are importance of probability: ....

⇨ The theory of probability was developed with the study of games of chance such as playing with dice and cards.
⇨ Along with games we also face uncertainty in different phases of life like business, projects, economy and in many other day to day activities of life.
⇨ They study of probability is essential to the advancement of sciences.
⇨ The probability rules and probability models help the researchers to take decisions on the subject under research.

Meaning of Experiment:

Experiment is an operation which results in some well-defined outcomes is called an experiment.

⇨ Random Experiment:

It is an experiment whose outcome cannot be predicted with certainty is called as random experiment.

⇨ Probability Experiment:

It is an action through which specific results are obtained. For example tossing a coin, throwing an unbiased die and drawing of a card from a well shuffled pack of 52 playing cards.

Meaning of an Event:

An event is a collection or any combination of outcomes to which a probability is assigned.

Types of Events:

Following are different types of events in probability: ....

1. Certain Event:

An event which has the probability value of 1 (i.e. 100%), then it is a **certain event.** In this event we surely get the output in every sample experiment is performed.

2. Uncertain Event:

An event which has the probability values between **1** and **0,** then it is an uncertain event. In this event we probably get some output in every sample experiment is performed.

3. **Impossible Event:**
   An event which has the probability of **0** (0%), then it is an impossible event. The sample space does not have an event and 0. In this experiment, a particular event never occurs in every sample experiment is performed. **For example**, rolling a dice and obtaining a number **7.**

4. **Elementary / Simple Event:**
   Every one of the elements, which forms the sample space is called an elementary event. It is the event of a single outcome.

5. **Complex Event:**
   The union of simple events is called as complex event. It can be decomposed into simple events. They further sub-divided into smaller events.

6. **Compound / Mixed Event:**
   An event which is having more than one sample point is called as compound event. It is a subset of the sample space but not a simple event. **For example,** if a dice is thrown, an **even number** would be a compound event, which consists of 2, 4, 6 elementary events.

7. **Equally Likely Events:**
   Two or more events which have an equal probability of occurrence are called as equally likely events. For example, in tossing a coin, events **Head** & **Tail** are said to be equally likely events, which have the same chance of probability.

8. **Not Equally Likely Events:**
   Two or more events which have an unequal probability of occurrence are said to be **not equally likely events.** In throwing dice, events {2, 4, 6} and {3, 5} are said to be not likely events, which have different chance of occurrence i.e. 1/2 and 1/3 probabilities respectively.

9. **Independent Event:**
   Each event not affected by other events is called an independent event. It is an unconditional event. The two events **A** and **B** are independent events, if the probability of the succeeding event is not affected by the outcome of the proceeding event.
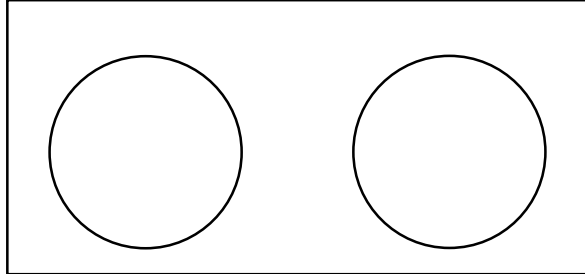
10. **Dependent Event:**
    Each event affected by other events is called as dependent event. It is a conditional event. The two events – **A** and **B** are dependent if the probability of the succeeding event is affected by the outcome of the preceding event.
    There are **3** blues and **5** red balls in a box. If we pick one ball from the box and the probability picking blue is **3/8** and that of red ball is **5/8.**

Suppose, the first outcome is a **red** ball. Without replacing the first ball picked, the probability of second pick of red ball is **4/7,** and the probability of blue ball is **3/7.**

11. Mutually Exclusive Events:

Two or more events are called **mutually exclusive** when they do not occur at the same time simultaneously. It means they do not have any element in common. It is either one or the other, but not both.
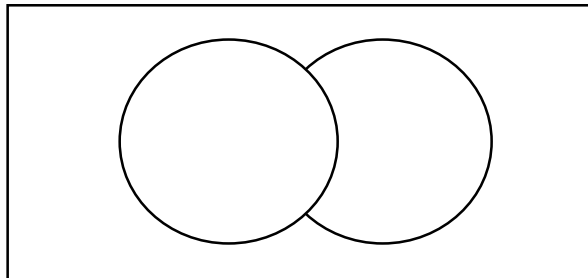

Mutually Exclusive Events (A U B)

If an event **A** occur, it excludes and event **B.** For example, in a deck of 52 cards, **Aces** and **Kings** are mutually exclusive (they cannot occur both) (i.e. = 4 Aces + 4 Kings)

12. Not Mutually Exclusive Events:

Two or more events are called **Not mutually Exclusive** when they are similar in character and occurring simultaneously. It means the occurrence of one does not exclude the occurrence of other.

Non-Mutually Exclusive Events (A ∩ B)

If an event **A** occurs, it does not exclude occurrence of an event **B.** For example, in a deck of 52 cards, **Hearts and Kings** are not mutually exclusive (they can occur both). Because we can have a King of Hearts.

(i.e. 16 cards = 13 Hearts + 4 Kings – 1 King Heart)

13. Exhaustive Events:

A set of events is called exhaustive events. It is the union of all the experiments in the complete space. In rolling a six-sided dice, the outcomes 1, 2,3,4,5, and 6 are collectively exhaustive.

14. Not exhaustive Events:

The events are not exhaustive when all the possible elementary events under the experiment are not covered by the event considered together.

Addition Theory:

The probability of happening an event can easily understood by the addition rules. A theorem known as **"Addition Theorem".**

It is the process of determination of the probability that either event A or event B occurs or both occur.

General rule:

If **A** and **B** be two events, the probability of their union is,

$$P (A∪B) = P (A) + P (B) – P (A) ∩ B)$$

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

Where,

P   = Probability function.

P (A)   = Probability of occurrence of event A.

P (B)   = Probability of occurrence of event B.

P (A∪B)   = Probability of occurrence of event A or event B.

P (A∩B)   = Probability of occurrence of both event A and event B.

(**Note:** "U" symbol indicates **"or"** and "∩" indicates **"and")**

The addition theorem of probability is studied under two headings:

1) Addition theorem of mutual exclusive events:

If A and B are mutually exclusive events, then

P (A or B) = 0, so the rule can be simplified

P (A or B) = P (A) + P (B)

2) Addition theorem of Not mutually exclusive events:
If A and B are not mutually exclusive events then,
P (A or B) = P (A) + P (B) = 1 so the rule can be simplified

P (A or B) = P (A) + P (B) – P (A and B)

Problem NO: 01

What is the probability of getting a **Queen** or a **King** from selecting a card from a deck of cards?

Problem NO: 02

What is the probability of getting a **Queen** or a **Heart** from selecting a card from a deck of cards?

Problem NO: 03

A coin is tossed successfully three times. Find the probability of getting exactly one head or two heads.

Problem NO: 04

If the probabilities of scoring a century by two batsmen **Kapil** and **Rahul** are **1/2** and **1/3** respectively then, what is the probability of the century to be scored?

Problem NO: 05

A single 6 sided dice is rolled. What is the probability of rolling a 3 or a 5?

Problem NO: 06

A glass jar contains **4 red, 5 green, 3 blue** and **6 yellow** marbles. If a single marble is chosen at random from the jar, what is the probability that it is blue or yellow?

Problem NO: 07

In a statistic class of 70 students, 49 are Hindus, 14 are Muslims and 7 are Christians. If a student is chosen at random from the class, what is the probability of choosing a Muslim or a Christian?

Problem NO: 08

In a group of 24 cricket players. 14 are batsmen, 10 bowlers. 6 batsmen and 2 bowlers are from Karnataka. If a player is chosen at random, what is the probability of choosing a batsmen or a player from Karnataka?

Problem NO: 09

A single 6 sided dice is rolled. What is the probability of rolling a number greater than 3 or an even number?

Problem NO: 10

A bag contains 30 balls number from 1 to 30. One ball is drawn at random. Find the probability that the number of ball is a multiple of 5 or 6?

Problem NO: 11

A single 6 sided dice rolled. What is the probability of rolling a number greater than 3 or odd number?

Problem NO: 12

One card is drawn from a standard pack of 52 cards. What is the probability that it is
   a) A spade
   b) A king
   c) The ace club

Problem NO: 13

A single card is choosing at random from a standard pack of 52 playing cards. What is the probability of choosing_?
   a) A king or a club
   b) A king or queen

Problem NO: 14

The unbiased dice are drawn find the probability that,
   a) Both the dice show the same digit
   b) The first dice show 5

Problem NO: 15

The probability of an Indian having a car is 26%. The probability of an Indian having a house is 40%. The probability of an Indian owning a car and house is 18%. What is the probability that the Indians owns a car or a house?

Problem NO: 16

A single six sided dice is rolled. What is the probability of rolling a number less than 5 or an even number?

Problem NO: 17

A coin is tossed successfully two times. Find the probability of getting exactly one tail or exactly two tails.

Problem NO: 18

A card is drawn from a deck of 52 playing cards. What is the probability that it is?
   a) A red card
   b) A block card
   c) A red or block card

CHAPTER – 04
Correlation and Regression Analysis

Introduction:

Statistical methods of measures of central tendency, dispersion and skewness are helpful for the purpose of comparison and analysis of distributions involving only one variable i.e. *univariate* distributions. Sometimes we may, however, come across certain series where each item of the series may assume the values of two or more variables. So numerical data comes in pairs. The distribution in which each unit of series assumes two values is called **bivariate distribution.** In a bivariate distribution, we are interested to find out whether there is any relationship between two variables.

Meaning of Correlation:

Correlation is a statistical technique which describes the relationship between two or more variables.

It describes and indicates the size and direction of a mutual relationship between two or more variables. However, it does not mean that the change in one variable is the **cause** and **effect** of change in the values of the other variable.

For example, in the two variables "**hours worked** and **income earned**" there is a relationship between the two if the increase in hours worked is associated with an increase in income earned. If we consider the two variables **"price** and **purchasing power",** as the price of goods increases the ability of a person to buy the goods decreases with his constant income. If the students are missing the classes, they may get lower grades in the examination as there might be relationship between the **classes missed** and **grades secured.**

Definition of Correlation:

According to **A.M. Tuttle** defines **"Correlation is an analysis of the covariation between two or more variables".**

Meaning of Correlation Analysis:

Correlation analysis involves various methods and techniques used for studying and measuring the extent of relationship between the two variables.
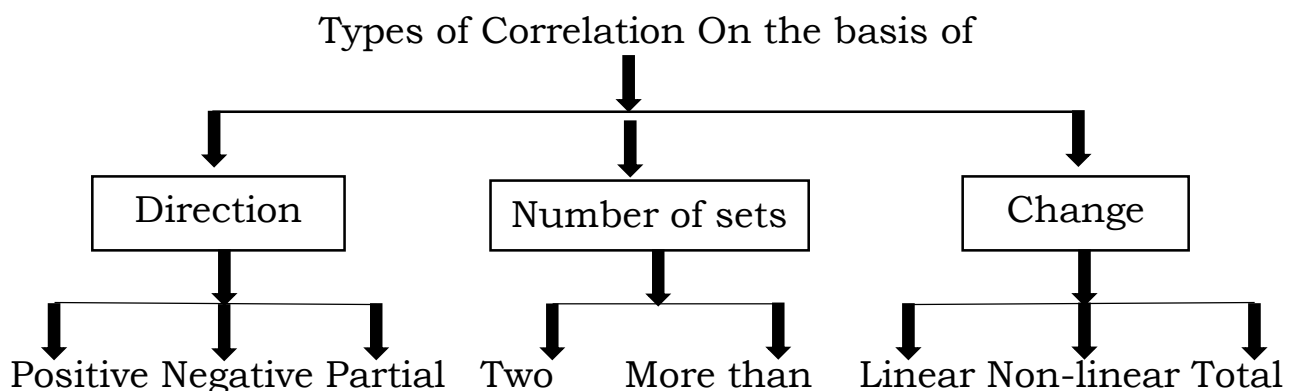
Uses or Merits or Advantages of Correlation:

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. The study of correlation is of immense use in practical life.

Following are the uses of correlation: ....

✓ Correlation analysis is used in advanced portfolio management.

✓ It helps to take important decisions when two or more variables lies in the same decision making process.

✓ We can estimate the value of one variable given the value of another by studying the relationship between the two variables.

✓ Correlation analysis contributes of economic behaviour of one variable depending upon the other.

✓ Correlation can be used as the step before conducting a statistical experiment.

Types of Correlation:

Correlation can be classified in several ways. The important ways of classification are as follows: ....

Types of Correlation On the basis of

Direction	Number of sets	Change

Positive Negative Partial   Two   More than   Linear Non-linear Total

1. Positive Correlation:

If both the variables move in the same direction, is called as positive correlation i.e. if one variable increases, the other variable also increases or if one variable is decreases, the other variable also decreases. **For example:** Positive correlation would be the correlation between **Price** and **Supply.** Typically, higher the price of a product, the more will be the supply of a product.

2. Negative Correlation:

If both the variables move in the opposite direction, is called as Negative correlation, i.e. if one variable increases, the other variable decreases on an average.

For example: a negative correlation between **Price** and **Demand**. Typically, higher the price of a product, the lower will be the demand for a product.

3. Partial Correlation:

In a partial correlation, more than two variables are recognized but only two variables are considered keeping the other variable constant. It means one variable goes on changing and the other remains constant.

4. Simple Correlation:

In this correlation, only two variables are involved in a study and both the variables are changing in any direction. It determines the strength and the direction of the relationship between two variables.

5. Multiple Correlation:

In a multiple correlation, three or more variables are studied simultaneously.

6. Linear Correlation:

If change in one variable is accompanied by change in another variable in a constant ratio is called as linear correlation.

The graph of the variables having a linear relationship will form a straight line.

7. Non-linear Correlation:

If the amount of change in one variable does not follow a constant ratio with the change in another variable.

The graph of the variables have a curvy relationship.

8. Total Correlation:

It based on all the relevant variables, which is normally not feasible. It is one of several generalizations of the mutual information. It is also called as the multivariate constraint.

Methods of Correlation:

We can measure the correlation mathematically how close the two variables are related. It is in terms of **Correlation Coefficient,** which is denoted by "**r**" as a real number.

There are two methods of calculation of the correlation coefficient:

A. Karl Pearson's Coefficient method

B. Spearman's Rank Correlation.

A. Karl Pearson's Coefficient Correlation Method

It is a measure of the degree and strength of the linear relationship between x and y variables. It is the covariance of the two variables divided by the product of their standard deviations.

Assumptions of Pearson's Correlation Coefficient:

✓ There is linear relationship between two variables.
✓ Cause and effect relationship.
✓ Materiality.

Merits of Karl Pearson's correlation:

Following are the merits of Pearson's correlation: …………

✓ It helps to find direction of correlation.
✓ It is most widely used method.
✓ It helps in comparison.
✓ It is useful in further algebraic treatment.
✓ It is based on all observations or variables.

Demerits of Karl Pearson's correlation:

✓ It assumes linear relationship regardless of the assumption is correct or not.
✓ It is Time consuming.
✓ It is affected by extreme values
✓ It requires careful interpretation.

Coefficient of Determination:

On the basis of the formula given by Karl Pearson, we can state approximately the degree of correlation as under

Degree of Correlation	Positive		Negative	
Correlation lies between +1 and -1 Approximation	Range		Range	
	From +1	To 0	From 0	To -1
1. Perfect	+1		-1	
2. Very High	+1.00	+0.90	-0.90	-1.00
3. High	+0.90	+0.75	-0.75	-0.90
4. Moderate	+0.75	+0.60	-0.60	-0.75
5. Low	+0.60	+0.30	-0.30	-0.60
6. Very low	+0.30	+0.00	-0.00	-0.30
7. No correlation	0		0	

Probable Error:
Meaning:

The probable error is a statistical measure which provides for two limits within which all the answers, obtained from different sample pairs of the population, will fall.

Definition:

According to Secrist, "_The probable error of the correlation coefficient is an amount, which if added to and substracted from the average correlation coefficient, produces amounts within which the chances are even that a coefficient of correlation from a series selected at random will fall_**".**

It is obtained by the following formula:

$$PE = 0.6745 \; \frac{1 - r^2}{\sqrt{N}}$$

Uses of Probable Error:

Following are the some of the uses of Probable error: …………

- It is used to determine the limits within which the population correlation coefficient may be expected to lie.
- It may be used to test if an observed value of sample correlation coefficient is significant of any correlation in population.

Conditions for the use of Probable Error:

The measure of probable error can be properly used only when the following three conditions exist: ………………

- ✓ The data must have been drawn from a normal population.
- ✓ The conditions of random sampling should prevail in selecting the samples from the population.
- ✓ The coefficient of correlation must have been computed from the sample.

Coefficient of Correlation:

It is the numerical measure of the amount of correlation existing between the two variables X and Y – the subject and the relative respectively.

Formulas for calculating Karl Pearson's Co-efficient of Correlation:

Formula under Direct Method:

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2} \ \sqrt{\sum dy^2}}$$

Formula under Indirect Method:

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

Steps to be followed while Calculating Karl Pearson's Co-efficient of Correlation:

Step 1: Find out arithmetic mean of both the variables i.e., x and y variables $\left(\overline{X} = \dfrac{\sum x}{N}\right)$ and $\left(\overline{Y} = \dfrac{\sum y}{N}\right)$

Step 2: Find out the deviations of X variable i.e., (dx= X-$\overline{X}$)

Step 3: Square the deviations of X variables i.e., (dx^2)

Step 4: Find out the deviations of Y variable i.e., (dy= Y-$\overline{Y}$)

Step 5: Square the deviations of Y variables i.e., (dy^2)

Step 6: Multiply the deviations of x and y variables i.e., (dxdy)

Step 7: Apply the Formula

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2} \ \sqrt{\sum dy^2}}$$

OR

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

Problems for Calculating Karl Pearson's Correlation Coefficient:

Problem No: 01

From the following table calculate the coefficient of correlation by Karl Pearson's method. Arithmetic means of X and Y variables are 6 and 8 respectively.

X: 6   2   10   -   8

Y: 9   11   -   8   7

Problem No: 02

Find the coefficient of correlation between the following two variables. Comment on the result through the probable error.

X: 6   8   12   15   18   20   24   28   31

Y: 10   12   15   15   18   25   22   26   28

Problem No: 03

Calculate the coefficient of correlation form the following data and calculate its probable error.

Marks in

Statistics (X): 30   60   30   66   72   24   18   12   42   06

Accounts (Y): 06   36   12   48   30   06   24   36   30   12

Problem No: 04

Calculate the coefficient of correlation between income and weight from the following data. Comment on the result.

Income (Rs.): 100 200 300 400 500 600

Weight (Kg's): 120 130 140 150 160 170

Problem No: 05

Calculate Karl Pearson's coefficient of correlation between percentages of pass and fail from the following data.

No. of students :   800 600 900 700 500 400
No. of Passed :   480 300 450 560 450 300

Problem No: 06

Calculate Karl Pearson's coefficient of correlation between price and demand from the following data.

Price in Rs: 21 22 23 24 25 26 27 28 29
Demand in units: 20 19 19 17 17 16 16 15 14

Problem No: 07

Find Karl Pearson's correlation co-efficient between age and playing habits of the following students.

Age   :   15, 16, 17, 18, 19, 20

No. of students :   250, 200, 150, 120, 100, 80

Regular players: 200, 150, 90, 48, 30, 12

Problem No: 08

Calculate Karl Pearson's co-efficient of correlation with the following data of 6 cities between density of population and death rates. Interpret the correlation by finding probable error.

Cities   :   A   B   C   D   E   F

Area [in Sq.km]   :   150 180 100 60   120 80

Population [in 000's] :   30   90   40   42   72   44

No. of deaths   :   300 1440 560 840 1224 312

Problem No: 09

The following table gives the distribution of the total population and those are totally or partially blind among them. Find out if there is any relation between age and blindness.

Age in years	No. of persons (000's)	Blind
0-10	100	55
10-20	60	40
20-30	40	40
30-40	36	40
40-50	24	36
50-60	11	22
60-70	6	18
70-80	3	15

Problem No: 10

Find Pearson's coefficient of correlation between average profits and average advertisement expenditure per shop and interpret result.

No. of Shops   :   12   18   25   20   10

Total profits (Rs): 7,200   5400   10,000   3000   1800

Total Advert.

Expenditure (Rs): 1,200   3,600   7,500   1,000   600

Problem No: 11
     Compute Karl Pearson's coefficient of correlation between per capita National income and per capita Consumer expenditure from the data given below:

Year	N.I. per capita	Per capita Con. Exp.
1999	249	237
2000	251	238
2001	248	236
2002	252	240
2003	258	245
2004	269	255
2005	271	254
2006	272	252
2007	280	258
2008	275	251

Also calculate probable error.

Problem No: 12
     Calculate the Pearson's coefficient of correlation from the following data using 44 and 26 (as assumed mean) respectively as the origin of X and Y.

X: 43, 44, 46, 40, 44, 42, 45, 42, 40, 42, 57, 48

Y: 29, 31, 19, 18, 18, 27, 27, 29, 41, 30, 26, 10

Problem No: 13
     Compute Karl Pearson's coefficient of correlation from the following data. What conclusion do you draw from the result?

Supply (Quintals)   :30, 29, 29, 25, 24, 24, 24, 21, 18, 15
Price (Rs.)   :11, 12, 13, 14, 15, 16, 15, 17, 18, 20

Problem No: 14
     Calculate Karl Pearson's coefficient of correlation between X and Y series.

	X	Y
Number of pairs	15	15
Variance	10	12

Summation of products of deviations of X & Y series from the respective means is 135.

Problem No: 15

Calculate Karl Pearson's coefficient of correlation between X and Y series.

	X	Y
Number of pairs	15	15
Mean	25	18

Sum of the product of squared deviation from mean is 136 & 138
Summation of products of deviations of X & Y series from the respective means is 122.

Problem No: 16

Calculate the correlation coefficient between the variables X and Y from the following figures:

$\Sigma x = 118$ $\Sigma x^2 = 556$ $\Sigma xy = 368$
$\Sigma y = 93$ $\Sigma y^2 = 309$ $n = 30$

Problem No: 17

Calculate the coefficient of correlation between the X and Y series from the following data:

		X	Y
Arithmetic mean	:	74.5	125.5
Assumed mean	:	69.0	112.0
Standard deviation	:	13.07	15.85

Sum of the products of
deviations of x and y from: +2176
their assumed means
No. of pairs of observations: 8

Problem No: 18

Calculate the correlation coefficient between the variables X and Y from the following figures:

$\Sigma x = 125$ $\Sigma x^2 = 650$ $\Sigma xy = 508$
$\Sigma y = 100$ $\Sigma y^2 = 460$ $n = 25$

Problem No: 19

Calculate probable error for the distribution if r = 0.6 and N = 64.

Problem No: 20

Coefficient of correlation between two variables X and Y is 0.8. Their covariance is 20. The variance of X is 16. Find the standard deviation of Y series.

Spearman's Rank Correlation:

Introduction:

A nonparametric rank statistic proposed by spearman in 1904 as measure of the strength of the associations between two variables.

The spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables.

Meaning and Definition of Rank Correlation:

Spearman's Rank Correlation is a technique used to test the direction and strength of the relationship between two variables. This method is based on the ranks of the items rather than on their actual values and makes no assumptions about the population from which actual observations are drawn. The advantage of this method over the others in that it can be used even when the actual values if item are unknown.

The following Formula will be used: ……………………….
Where,

$$r_s = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

r_s = rank correlation coefficient
ΣD^2 = Sum of squares of differences between the pairs of ranks.
N = Number of pairs of observations.

Features of Spearman's Rank Correlation:

1. It is based on ranks rather than actual observations.
2. It is nonparametric because no assumptions are made about the form of population from which sample observations are drawn.
3. The sum of the differences of rank correlation two variables shall be zero.
4. It can be interpreted like Karl Pearson's coefficient of correlation.
5. It lies between -1 and +1.

Advantages of Spearman's Rank Correlation:

1. It is easy to understand and simple to calculate.
2. It is suitable for qualitative data.
3. This method is also suitable for abnormal data.
4. It is only method for ranks.
5. This method is applicable even for actual data.

Limitations of Spearman's Rank Correlation:

✓ Unsuitable data
✓ Tedious calculations
✓ Approximation

Computation of Rank Correlation Coefficient:

Rank correlation can be calculated in different situations. They are: ....

✓ **When Ranks are given**
✓ **When the ranks are not given**

When Ranks are given following steps to be followed:

1. Find the difference between the ranks of two corresponding variables and denote it by D. **(D = R₁ – R₂)**

2. Square the differences and find the total. **(ΣD²)**

3. Apply the formula $r_s = 1 - \dfrac{6\Sigma D^2}{N^3 - N}$

Problem No: 01

Ranks by Judge A: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Ranks by Judge B: 8, 7, 6, 4, 3, 2, 5, 1, 9, 10

Problem No: 02

The ranks of the 15 students in two subjects A and B are given below: The two numbers within the brackets denoting the ranks of the same student in A and B respectively.

(1, 10) (2, 7) (3, 2) (4, 6) (5, 4) (6, 8) (7, 3) (8, 1) (9, 11) (10, 9) (11, 15) (12, 5) (13, 14) (14, 12) (15, 13) find rank correlation coefficient.

Problem No: 03

Ranks by Judge A: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Ranks by Judge B: 3, 1, 5, 2, 7, 4, 9, 6, 10, 8

Problem No: 04

Ten competitors in a summer fall design show are ranked by three judges. Using rank correlation find out which pair of judges have the nearest approach to the common taste in fashion design:

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

Problem No: 05

Ten competitors in a beauty contest are ranked by three judges. Using rank correlation find out which pair of judges have the nearest approach to the common taste in beauty: ....

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

Problem No: 06

Six competitors in a boxing contest are ranked by three judges. Using rank correlation find out which pair of judges have the nearest approach to the common interest in boxing: ....

Judge A	Judge B	Judge C
1	3	6
6	5	4
5	2	3
4	1	2
3	4	1
2	6	5

When Ranks are not given following steps to be followed:

1. Assign ranks to the actual numbers through either by ascending order or descending order. That is taking highest value as first rank or the lowest value as first rank. But, same method should be followed for both the variables.

2. Find the difference between the ranks of two corresponding variables and denote it by D. **(D = R₁ – R₂)**

3. Square the differences and find the total. **(ΣD²)**

4. Apply the formula $r_s = 1 - \dfrac{6\Sigma D^2}{N^3-N}$

Problem No: 07

Ranks by Judge A: 40, 50, 30, 80, 60, 90, 70, 20

Ranks by Judge B: 80, 30, 40, 50, 20, 10, 90, 60

Problem No: 08

Find the rank correlation coefficient between marks in accountancy and statistics.

Marks in Accountancy	Marks in statistics
48	62
60	78
72	65
62	70
56	38
40	84
39	60
52	32
30	31

Problem No: 09

Find the rank correlation coefficient from the following data:

X	Y
75	120
88	134
95	150
70	115
60	110
80	140
81	142
50	100

Problem No: 10

In a marketing survey the price of Tea and Coffee in a city based on quality was found as shown below. Could you find any relation between tea and coffee?

Price of Tea	Price of Coffee
88	120
90	134
95	150
70	115
60	110
75	140
50	100

When Ranks are Equal OR Tied Ranks OR Repeated Ranks:

There may be two or more items having equal values in the distribution. In such a case the same rank is to be given to equal values. The ranking is said to be tied. In such situations an average rank is to be given to each individual item. For Example: if the value repeated twice at the 4th rank, the common rank is to be assigned to each item is $\frac{4 + 5}{2}$ = 4.5 which is the average of 3rd and 4th item.

At same time if any value repeated three times at 6th place, the common is assigned to each item will be $\frac{6 + 7 + 8}{3}$ = 7 which is the average of 6th, 7th and 8th item.

If the ranks are repeated or tied, it is required to add $\frac{1}{12}$ (m³ – m) as correction factor to the value of ΣD^2 in the formula. Where '**m**' stands for the number of times a value is repeated and whose ranks are common.

If one repeated two times,

The formula will be $\mathbf{r_s = 1 - 6\left[\dfrac{\Sigma D^2 + \frac{1}{12}(m^3 - m)}{N^3 - N}\right]}$

If one value repeated 2 times and another value repeated 3 times:

The formula will be $\mathbf{r_s = 1 - 6\left[\dfrac{\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)}{N^3 - N}\right]}$

Problem No: 11

Following are the marks obtained by 6 students in two class tests. Calculate rank correlation coefficient between the marks of two tests.

Students	Test 1	Test 2
A	70	65
B	68	65
C	67	80
D	55	60
E	60	68
F	60	58

Problem No: 12

Calculate rank correlation coefficient from the following data:

X	Y
70	13
80	15
65	12
78	14
68	13
65	11
82	16
65	10

Problem No: 13

Calculate rank correlation coefficient from the following data:
X: 53, 98, 95, 81, 70, 81, 65, 81
Y: 25, 47, 82, 76, 53, 61, 75, 70

Problem No: 14

The competitors in a beauty contest were ranked by 3 judges as follows:

Judge A	Judge B	Judge C
3	5.5	9
5	5.5	9
4	1	7
10	8.5	5
8	4	2
8	10	2
1	2	2
6	7	5
8	8.5	5
2	3	9

Which pair of judges has the nearest approach to common taste in beauty?

Problem No: 15

From the marks obtained by 8 students in accountancy and statistics, compute coefficient of correlation by rank difference method.

Marks in Accountancy	Marks in statistics
60	10
15	40
20	30
28	50
12	30
40	20
80	60
20	30

Problem No: 16

10 students obtained the following marks in statistics and accounts.

Students	Statistics	Accounts
A	115	75
B	109	73
C	112	85
D	87	70
E	98	76
F	120	82
G	98	65
H	100	73
I	98	68
J	118	80

Problem No: 17

Calculate the coefficient of rank correlation from the following:

Marks in Mathematics	Marks in statistics
60	10
15	50
20	30
60	50
12	30

Problem No: 18

If N = 8, and $\sum D^2 = 4$ find r_s?

Problem No: 19

Given N = 6, and $r_s = 0.57142$ find $\sum D^2$?

Problem No: 20

If the coefficient of rank correlation = 0.75, $\sum D^2 = 30$ and find N?

Regression Analysis

Meaning of Regression:

The word "Regression" means returning back to averages or previous state or earlier stage in some way.

Regression is a technique used to predicting the value of dependent variable on the basis of independent variable.

Regression lines:

Regression line is the line which gives the best estimate of one variable from the value of any other given variable.

Regression Analysis:

Regression Analysis is a statistical technique for estimating the relationships among the variables. Regression analysis involves identifying the relationship between a dependent variable and one or more independent variables.

Advantages or Benefits or Uses of Regression Analysis:

Following are the important uses or benefits of regression: ……

1) Regression analysis is used to predict a dependent variable from a number of independent variables.
2) Regression analysis is used by management accountant for both planning and control.
3) It estimates the value of dependent variable from the value of independent variable.
4) It helps obtain a measure of the error involved in using the regression line as a basis for estimations.
5) It helps to calculate co-efficient of correlation (r) with the help of regression co-efficient.
6) Regression analysis widely used in statistical analysis of demand curves and supply curves.
7) It is widely used in production function, consumption function and cost function etc.
8) It is useful in indicating the degree of association that exists between the two variables.

Regression co-efficient:

"It is a statistical technique used to explain or predict the behaviour of a dependent variable".

Regression Equations:

There are two important regression equations i.e. Y on X and X on Y.

Linear regression equation of Y on X is

$\qquad$ Y= a+bx .... Equation -1

Linear regression equation of X on Y is

$\qquad$ X=a+by .... Equation – 2

1) Regression equation of Y on X is

$$(Y - \overline{Y}) = \frac{\Sigma xy}{\Sigma x^2} (X - \overline{X})$$

2) Regression equation of X on Y is

$$(X - \overline{X}) = \frac{\Sigma xy}{\Sigma y^2} (Y - \overline{Y})$$

1. Regression Co-efficient of X on Y:

The regression co-efficient of X on Y is represented by the symbol bxy or b_1. It measures the change in X corresponding to a unit change in Y.

The regression co-efficient of X on Y is given by $bxy = r \dfrac{\sigma X}{\sigma Y}$

When the deviations are taken from actual means of X and Y series: ....

$$bxy = \frac{\Sigma xy}{\Sigma y^2}$$

2. Regression Co-efficient of Y on X:

The regression co-efficient of Y on X is represented by the symbol byx or b_2. It measures the change in Y corresponding to a unit change in X.

The regression co-efficient of Y on X is given by $byx = r \dfrac{\sigma y}{\sigma x}$

When the deviations are taken from actual means of Y and X series: ....

$$byx = \frac{\Sigma xy}{\Sigma x^2}$$

Calculation of co-efficient of correlation from regression co-efficient: ....

$$r = \sqrt{bxy \times byx}$$

Assumptions of Regression Analysis:

Regression analysis is based on some assumptions as under: ......

1) Existence of actual linear relationship.
2) The regression analysis is used to estimate the values within the range of which it is valid.
3) The relationship between the dependent variables remains the same till the regression equation is calculated.
4) The dependent variable takes any random value but the values of the independent variables are fixed.
5) In regression, we have only one dependent variable in our estimating equation.

Properties of Regression Co-efficient:

Following are the important properties of regression co-efficient:

a) Both regression co-efficient must have the same sign (positive or negative).
b) The co-efficient of correlation is geometric mean of two regression co-efficient. $r = \sqrt{bxy \times byx}$
c) The correlation co-efficient and regression co-efficient must have the same sign.
d) If one regression co-efficient is greater than one the other must be less than one.
e) Arithmetic mean of bxy and byx is equal to or greater than co-efficient of correlation symbolically $\frac{bxy + byx}{2} \geq r$
f) Regression co-efficient are independent of origin but not of scale.

Dis-advantages of Regression Analysis:

Following are the important dis-advantages of regression analysis: ....

✓ Outputs of regression can lie outside of the range (0,1).

✓ It has limitations in the shapes that linear models can assume over long ranges.

✓ The extrapolation properties will be possibly poor.

✓ It often gives optimal estimates of the unknown parameters.

Differences between Correlation and Regression:

Following are the important differences between Correlation and regression as under: ....

Correlation	Regression
1. It precedes regression.	**1.** It succeeds correlation.
2. It tests the closeness between the two variable.	**2.** It establishes the future dependent variables.
3. It measures the degree of co-variation.	**3.** It studies the nature of co-variation.
4. It does not study cause and effect.	**4.** It studies cause and effect.
5. Both the variables are interdependent.	**5.** Time is independent and other one is dependent.
6. It studies a two-way average relationship.	**6.** It studies one-way average relationship.
7. It establishes just a relationship.	**7.** It establishes functional relationship.

Problems on Regression Analysis

When the deviations are taken from Actual Mean:

Problem No: 01

Construct two regression equations.

X: 6, 2, 10, 4, 8   **Y:** 9, 11, 5, 8, 7

Problem No: 02

Construct two regression equations.

X: 8, 4, 10, 6, 5   **Y:** 3, 12, 5, 9, 8

Regression Equations through Regression Co-efficient:

When the deviations are taken from Actual Mean:

Problem No: 03

Obtain the two regression equations and determine value of X = 73.

X: 57, 58, 59, 59, 60, 61, 62, 64

Y: 67, 68, 65, 68, 72, 72, 69, 71

Problem No: 04

From the data given below find: ....
a) The two regression equations.

b) The most likely marks in statistics when marks in income tax is 50

c) The coefficient of correlation with b_{xy} and b_{yx}

Marks in income tax   : 35, 37, 28, 42, 48, 54, 65, 40, 35, 50

Marks in statistics   : 60, 55, 75, 70, 80, 35, 65, 35, 50, 60

Problem No: 05

The following information gives relative value of two variables.

X: 42, 44, 58, 55, 89, 98, 66

Y: 56, 49, 53, 58, 65, 76, 58

Find out:
a) The two regression equations.
b) Most probable value of X and Y = 60.
c) The coefficient of correlation X and Y

Problem No: 06

The following data related to the age of husbands and wives:

Age of Husbands	Age of wives
25	20
28	26
30	29
35	25
36	18
38	26
39	35
42	35
55	46

Obtain the two regression equations and determine the most likely age of husband for the age of wife 25 years.

Problem No: 07

Formulate the two regression equations for the following data: ....

Marks in economics : 80, 45, 55, 56, 58, 60, 65, 69
Marks in statistics : 82, 56, 50, 48, 60, 62, 64, 66

Problem No: 08

Calculate a) two regression equations
b) Estimate the value of X where Y = 20
c) Determine the value of correlation coefficient through regression coefficients
X: 10, 12, 13, 17, 18
Y: 5, 6, 7, 9, 13

Problem No: 09

Find the two regression equations for the following data and predict the value of X when Y is 9.

X: 3   6   5   4   7   2   8   1

Y: 3   2   3   5   3   6   6   4

Problem No: 10

From the following data: ....

a) Calculate two regression equations

b) Estimate the value of X where Y = 40

c) Determine the value of correlation coefficient through regression coefficients

X: 20, 24, 26, 34, 36

Y: 10, 12, 14, 18, 26

Problem No: 11

Construct two regression equations for the following data and estimate the value of X when Y=70 and Y when X=650.

X: 100, 200, 300, 400, 500, 600, 700

Y: 30, 50, 60, 80, 100, 110, 130

Problem No: 12

From the following data relating to years of service in a factory of seven persons in a specialized field and their monthly income.

Year of service	:	11	7	9	5	8	6	10
Income (in '000' Rs.)	:	7	5	3	2	6	4	8

Find two regression equations and also estimate the income of a person of 12 years' service.

Problem No: 13

From the following data of the age of husband and age of wife, form two regression equations and calculate the wife's age when the husband's age is 25 years

Husband's age : 31, 32, 35, 29, 33

Wife's age : 23, 25, 29, 21, 26

Problem No: 14

Find the two regression equations by taking deviations from actual means of X and Y series.

X: 2 4 6 8 10

Y: 5 7 9 8 11

Problem No: 15

Find the two regression equations from the following and estimate the yield for 10 inches rainfall.

Rain fall (in inches):	2	4	6	8	10
Yield (tonnes) :	5	7	9	8	11

Problem No: 16

Find likely sales when advertisement expenditure is Rs. 30 crores

	Ad. Expenditure (crores)	Sales (crores)
Mean	40	240
Variance	25	625

Correlation coefficient = 0.8

Problem No: 17

Find likely sales when advertisement expenditure is Rs. 25 crores

	Ad. Expenditure (crores)	Sales (crores)
Mean	20	120
S.D	5	25

Correlation coefficient = 0.8

Problem No: 18   Given:

	Economics	Statistics
Mean marks	55	80
S.D	9	12

Correlation coefficient = +0.8

Find marks in Economics if marks in statistics = 70.

Problem No: 19

The following data based on 450 students, are given for marks in statistics and economics at a certain examinations.

	Statistics	Economics
Mean marks	40	48
S.D	12	16

Sum of the two products of deviations of marks from their respective means = 42,075.

Problem No: 20   you're given the following data:

	X	Y
Mean marks	47	96
S.D	64	81

Correlation coefficient = 0.36. Compute the regression line of X on Y and calculate X when Y = 88.

Problem No: 21

If Bxy = 0.8, Byx = 0.45 find r?

Problem No: 22

Find r when the two regression coefficients are (-0.6) and (-1.4)?

Problem No: 23

If r=0.62, σx = 10.5, σy = 8.5, find bxy?

Problem No: 24

Is it possible for two regression equations to be as follows?

Y= (-1.5) X+7, X= 0.6Y+9? Give reason.

Problem No: 25

Find two regression coefficients when r=0.9, σx = 10, σy = 15.

Problem No: 26

If bxy = 0.85, byx = 0.89 find r?

Problem No: 27

Find r when the two regression coefficients are (-0.8) and (-1.6)?

Problem No: 28

If r=0.62, σx = 15, σy = 12, find bxy?

Problem No: 29

Is it possible for two regression equations to be as follows?

Y= 1.5 X+7, X= 0.6Y+9? Give reason.

Problem No: 30

   Find two regression equations from the following data, when the deviations were obtained from mean:

   n=20, x=4, y=2, Σx^2=1680, Σy^2=320, Σxy=480.

CHAPTER - 05
Index Numbers

Introduction:
Index numbers are designed to measure the magnitude of economic changes over a period of time. Because they work in a similar way to percentage they make such changes easier to compare.

Meaning of Index Numbers:
An index number is a statistical device for comparing the general level of magnitude of a group of related variables in two or more situation.

Definition of Index Numbers:
According to Bowley defines "Index numbers are used to measure the changes in some quantity which we cannot observe directly".

Characteristics of Index Numbers:
Following are the important features of index numbers: ......
   a) Index numbers are specified averages.
   b) Index numbers are expressed in percentages.
   c) Index numbers are for comparison.
   d) Index numbers measure the effect of changes over a period of time.
   e) Index numbers measure the net changes in a group of related variables.
   f) Index numbers measure the changes which are not capable of direct investment.

Uses or Benefits of Index Numbers:
Following are the some important uses or advantages: ......
   a) They measure the relative change.
   b) They are better comparison.
   c) They are economic barometers.
   d) They compare the standard of living.
   e) They compare the wage structure.
   f) They are the pulse of the economy.
   g) They are special type of averages.
   h) They provide guidelines in policy making.

Index Numbers are called Economic Barometers
Index numbers are called as economic barometers. Because, they are used to check the economic growth of a nation. They can also influence the economic activities of the country.

Types of Classification of Index Numbers

Index numbers may be classified in terms of the variables that they are intended to measure. They are as follows: ....

1) Price Index Numbers:

Price index numbers measure the relative changes in prices of commodities between two periods.

2) Quantity Index Numbers:

Quantity index numbers measure the physical volume of production, employment etc. The most common type of the quantity index is that of quantity produced.

3) Volume Index Numbers:

Value index numbers compare the total value of a certain period with total value in the base period. Here total value is equal to the price of commodity multiplied by the quantity consumed.

For any index numbers, two time periods are needed for comparison. These are called the Base period and Current period.

Meaning of Base Year:

A Base year is the year used for comparison for the level of a particular economic index. It will be indicated as 0.

Meaning of Current Year:

The year for which comparisons are required is called as Current Year. It will be indicated as 1.

The various notations used are as given Below:

P_1= Price of Current year
P_0= Price of Base year
Q_1= Quantity of Current year
Q_0= Quantity of Base year

Methods of constructing Index Numbers:

A. Simple Index Numbers
B. Weighted Index Numbers

Meaning of Simple Index Numbers:

Under this method all the commodities are given equal importance while calculating index numbers.

They classified into two types: ....
1. Simple aggregative method
2. Simple average of price relatives method

Meaning of Simple aggregative method:

This is the simplest of all the methods of constructing index numbers and consists in expressing the total price in the current year as a percentage of the aggregate of prices in the base year.

Formula: $P_{01} = \dfrac{\Sigma P_1}{\Sigma P_0} \times 100$

P_{01} = Index number
ΣP_1 = Total prices of the current year
ΣP_0 = Total prices of the base year

Problems under Simple Aggregative Method:
Problem No: 01

Calculate index numbers from the following data by using simple aggregative method:

Items	Price per unit (in Rupees)	
	2010	2011
P	180	195
Q	150	160
R	190	220
S	130	175

Problem No: 02

Calculate index numbers from the following data by using simple aggregative method:

Items	Price per unit (in Rupees)	
	2010	2011
P	280	185
Q	160	260
R	240	420
S	80	275

Meaning of Simple average of Price Relatives Method:

Under this method the actual price of each variable will be converted into percentage of the base period. These percentages are called relatives because they are relative to the value of the base period.

Formula: $P_{01} = \dfrac{\Sigma P}{N}$

N = Number of items P = Price relatives [$\dfrac{P_1}{P_0} \times 100$ for each item]

Problems under Simple Average of price relatives Method:
Problem No: 03
Calculate index numbers from the following data by using simple average of price relatives' method:

Items	Price per unit (in Rupees)	
	Prices in 2010	Prices in 2011
P	180	195
Q	150	160
R	190	220
S	130	175

Problem No: 04
Calculate index numbers from the following data by using simple average of price relatives' method:

Items	Price per unit (in Rupees)	
	2010	2011
P	280	185
Q	160	260
R	240	420
S	80	275

Fisher's Ideal Index Numbers:
Fisher's price index number is the geometric mean of the laspeyres and paasche indices.

Formula: Fisher's Ideal index numbers

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

It is known as ideal index number because
a) It is based on geometric
b) It is based on the current year as well as the base year
c) It confirms certain tests of consistency
d) It is free from bias

Tests of Consistency of index numbers:
1) Time Reversal Test:
Time reversal test is a test to determine whether a given method will work both ways in time, forward and backward.

Time reversal test: $P_{01} \times P_{10} = 1$

Formula for Testing Time Reversal Test

$$P_{01} \times P_{10} = \sqrt{\frac{\Sigma P1Q0}{\Sigma P0Q0} \times \frac{\Sigma P1Q1}{\Sigma P0Q1} \times \frac{\Sigma P0Q1}{\Sigma P1Q1} \times \frac{\Sigma P0Q0}{\Sigma P1Q0}}$$

2) Factor Reversal Test:

It holds that the product of a price index and the quality index should be equal to the corresponding value index.

Factor Reversal Test: $P_{01} \times Q_{01} = \frac{\Sigma P1Q0}{\Sigma P0Q0}$

Formula for Testing Factor Reversal Test

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma P1Q0}{\Sigma P0Q0} \times \frac{\Sigma P1Q1}{\Sigma P0Q1} \times \frac{\Sigma Q1P0}{\Sigma Q0P0} \times \frac{\Sigma Q1P1}{\Sigma Q0P1}}$$

Problem No: 05
Construct Fisher's Ideal Index number from the following data:

Items	Base Year		Current year	
	Quantity	Price/Kg	Quantity	Price/Kg
Wheat	20	12	20	14
Rice	22	16	24	18
Gram	20	32	18	36
Pulses	8	29	12	29
Ghee	1	62	2	70
Sugar	5	14	4	16

Problem No: 06
Compute Fisher's Ideal Index number from the following data:

Items	Base Year		Current year	
	Value	Quantity	Value	Quantity
A	300	150	480	4
B	50	10	90	6
C	48	12	50	5
D	120	60	100	2
E	60	20	105	3.5

Problem No: 07

Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Commodities	Price		Quantity	
	Base Year	Current year	Base Year	Current year
A	12	20	55	120
B	10	12	100	80
C	14	15	60	80
D	16	18	30	70
E	18	20	40	40
F	20	15	70	60
G	20	16	90	100
H	16	18	80	80

Problem No: 08

Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Items	2015		2016	
	Price	Quantity	Price	Quantity
Wheat	20	8	30	10
Ghee	50	10	40	8
Fire wood	40	5	50	12
Sugar	60	20	60	16
Cloth	10	6	40	10

Problem No: 09

Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Items	Base Year		Current year	
	Value	Quantity	Price	Value
A	120	12	12	180
B	105	15	5	100
C	120	24	9	180
D	80	5	14	70

Problem No: 10

Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

	Base Year 2007		Current year 2008	
Items	Price	Quantity	Price	Quantity
M	20	O8	30	10
N	50	10	40	08
O	40	05	50	12
P	60	20	60	16
Q	10	06	40	10

Weighted Average of Price Relative Method:

When the specific weights are given for each commodity, the weighted index number is calculated.

Problems under Weighted Average Price Relative Method:

Formula: $P_{01} = \dfrac{\Sigma PW}{\Sigma W}$

Where, P = Price relatives $\left[\dfrac{P_1 \times 100}{P_0}\right]$

W = Weights assigned to price relatives.

Problem No: 11

Calculate index numbers from the following data by using weighted average of price relatives' method: ….

	Price per unit (in Rupees)		Weights
Items	Current year	Base year	
P	5	4	60
Q	3	2	50
R	2	1	30

Problem No: 12

Calculate index numbers from the following data by using weighted average of price relatives' method: ….

	Price per unit (in Rupees)		Weights
Items	Current year	Base year	
A	16	10	12
B	25	20	10
C	8	5	20
D	14	7	1

Consumer Price Index Numbers:

A consumer price index measures changes in the price level of consumer goods and services purchased by households.

Consumer price index is also called as cost of living index.

Uses of Consumer Price Index:

The consumer price indices are of great significance and is given below: ....

a) It is very useful in wage negotiations, wage contracts and dearness allowance adjustment in many countries.

b) At govt. level, it is used for wage policy, rent control, taxation and general economic policies.

c) Change in the purchasing power of money and real income can be measured.

d) It is used for analyzing market price for particular kinds of goods and services.

Methods of Constructing Consumer Price Index:

There are two methods of constructing consumer price index. They are

A. Aggregate Expenditure Method

B. Family Budget Method or Method of Weighted Relative Method.

1) Aggregate Expenditure Method:

When this method is applied the quantities of commodities consumed by the particular group in the base year are estimated which constitute the weight.

Formula:

Consumer Price Index Number = $P_{01} = \dfrac{\Sigma P1Q0}{\Sigma P0Q0} \times 100$

2) Family Budget Method:

When this method is applied the family budgets of a large number of people for whom the index meant are carefully studied and the aggregate expenditure of an average family on various items is estimated.

Formula: $P_{01} = \dfrac{\Sigma PW}{\Sigma W}$

Where, P = Price relatives $\left[\dfrac{P_1 \times 100}{P_0}\right]$

W = Weights assigned to price relatives.

Problem No: 13

Construct the cost of living index number for 2011 on the basis of 2010 from the following data using aggregate expenditure method.

Items	Quantity consumed in 2010	Price (Rs.) 2010	Price (Rs.) 2011
A	100	8	12
B	25	6	7
C	10	5	8
D	20	15	18

Problem No: 14

Calculate consumer price index by using family budget method for the year 2010 with 2009 as base year for the following data.

Items	Weights	Price (Rs.) 2009	Price (Rs.) 2010
Food	35	150	140
Rent	20	75	90
Clothing	10	25	30
Fuel	15	50	60
Miscellaneous	20	60	80

Problem No: 15

Construct consumer price index number for the following data by using

1) Aggregate expenditure method 2) Family Budget method.

Items	Units	Weights	Price (Rs.) 1990	Price (Rs.) 2000
A	Quintals	5 quintals	1200	1800
B	Kg	25 Kgs	20	25
C	Litre	20 Litres	20	30
D	Meter	25 Meter	50	45
E	Unit	20 units	100	200

Business Statistics
2 MARKS QUESTIONS FOR EXAMINATION

1) Define Statistics in Singular sense?

According to Wallis & Roberts "Statistics is a branch of science, which provides tools and techniques for decision making in the face of uncertainty".

2) What is Statistics?

"Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analyzing data as well as deriving valid conclusions and making reasonable decisions based on the analysis".

3) Define Statistics in Plural sense?

According to A.L. Bowley "Statistics are numerical facts in any department of enquiry placed in relation to each other".

4) State any four objectives of Statistics?

There are some of the important Objectives as follows.......
a) To throw light on the complex mass data.
b) To take action on the basis of analyzed data.
c) To draw conclusion on the basis of analyzed data.
d) To ascertain the reliability of statement.
e) To provide the basis for the development of knowledge.

5) State any four functions of Statistics?

1) Present the facts in simple form
2) Reduces complexity of data
3) Helps in comparison
4) Helps in formulation policies

6) State any four limitations of Statistics?

✓ Statistics is not suitable for calculating qualitative aspects.
✓ Statistical laws are not exact.
✓ Statistics may be misused
✓ Statistics does not study individual items.

7) What is Primary data?

Primary data is one, which is collected by the investigator himself for the first time for the purpose of a specific enquiry or study.

8) What is Secondary data?

Secondary data is the data which have been already collected and analyzed by someone or their own use and later the same data are used by others.

9) State any four methods for collecting primary data?
Following are the methods for collecting primary data:
  ✓ Direct personal interview
  ✓ Indirect Oral interview
  ✓ Information from correspondents
  ✓ Mailed Questionnaire method

10) Define an Average / Simple Average?
According to Clark defines as "Average is an attempt to find one single figure to describe the whole of figure**".**

11) Mention the types of Averages.
Following are the types of averages: ....
  a) Mathematical averages
  b) Positional averages
  c) Commercial averages

12) What is Arithmetic Mean / Simple Arithmetic Mean?
Arithmetic mean is the amount secured by dividing the sum of values of the items in a series by their number.

13) What is Weighted Mean?
Weighted mean refers to the average whose component items are being multiplied by weights and the aggregate of the multiplied results are being divided by the total of weights.

14) Define Median?
According to Connor defines "The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other all values lesser than the median".

15) Define Mode?
According to A.M. Tuttle "Mode is the value which has the greatest frequency or density in its immediate neighborhood".

16) How do you calculate 'Mode' when it is ill defined?
If Mode is ill defined then we go through the grouping and analysis table method. OR otherwise we can use following formula to locate the mode. Z = 3 median – 2 mean

17) Define Variation or Dispersion?
According to **A.L. Bowley** defines as, "Dispersion is a measure of variation of the items".

18) State ant two objectives of measure of Dispersion.
Following are the objectives of measure of dispersion: .....
   a) To use other statistical methods
   b) To compare variability
   c) To test the reliability of average
   d) To establish trends in time series

19) Mention the types of Dispersion OR mention any four methods of studying variation.
Types of Dispersion are as follows: ....
   a) The Range
   b) Quartile deviation
   c) Mean deviation
   d) Standard deviation

20) How range is calculated?
To calculate the range, we need to find the largest value of a variable and subtract the smallest value of given variables.
Range = L – S

21) What is meant by Mean Deviation?
The mean deviation is defined as the mean of the absolute deviations of observations from suitable average which may be the arithmetic mean, median or the mode.

22) Define Standard Deviation?
According to **Karl Pearson** defines as, "It is the positive square-root of the arithmetic mean of the square of the deviation of the given observations from their arithmetic mean".

23) State any four Merits of Standard deviation?
Following are the merits of Standard deviation: ....
   a) It is rigidly defined and its value is always definite.
   b) It is possible to calculate the combined standard deviation of two or more groups.
   c) It is based on all the observation of the data.
   d) It is less affected by the fluctuations of sampling.

24) Define Co-efficient of Variation (CV)?
According to **Karl Pearson** defines as, "Co-efficient of variation is the percentage variation in mean, standard deviation being considered the total variation in the mean".

25) What is meant be Skewness?

Skewness is a measure of asymmetry of a distribution. Skewness can come in the form of negative or positive depending upon data points are skewed to left (negative) or to the right (positive) of the data average.

26) Give the meaning of probability?

Probability is the study of the chance that a particular event or series of events will occur.

27) What is an experiment?

Experiment is an operation which results in some well-defined outcomes is called an experiment.

28) What is an event?

An event is a collection or any combination of outcomes to which a probability is assigned.

29) What is an Outcome?

An outcome is a possible result of an experiment or trial. Each possible outcome of a particular experiment is unique, and different outcomes are mutually exclusive.

30) What are equally likely events?

Two or more events which have an equal probability of occurrence are called as equally likely events. For example, in tossing a coin, events **Head** & **Tail** are said to be equally likely events, which have the same chance of probability.

31) What do you mean by not equally likely events?

Two or more events which have an unequal probability of occurrence are said to be **not equally likely events.** In throwing dice, events {2, 4, 6} and {3, 5} are said to be not likely events.

32) Give the meaning of mutually exclusive event?

Two or more events are called **mutually exclusive** when they do not occur at the same time simultaneously. It means they do not have any element in common. It is either one or the other, but not both.

33) Give the meant by not mutually exclusive event?

Two or more events are called **Not Mutually Exclusive** when they are similar in character and occurring simultaneously. It means the occurrence of one does not exclude the occurrence of other.

34) What is an addition theory?

The probability of happening an event can easily have understood by the addition rules. A theorem known as **"Addition Theorem".**

35) Define correlation?

According to A.M. Tuttle defines "Correlation is an analysis of the covariation between two or more variables".

36) What is meant by correlation analysis?

Correlation analysis involves various methods and techniques used for studying and measuring the extent of relationship between the two variables.

37) State any two uses or merits of correlation?

Following are the uses of correlation: ....
- ✓ Correlation analysis is used in advanced portfolio management.
- ✓ It helps to take important decisions when two or more variables lies in the same decision making process.

38) What is meant by Positive Correlation?

If both the variables move in the same direction, is called as positive correlation i.e. if one variable increases, the other variable also increases or if one variable is decreases, the other variable also decreases.

39) What is Negative Correlation?

If both the variables move in the opposite direction, is called as Negative correlation, i.e. if one variable increases, the other variable decreases on an average.

40) What is Perfect Correlation?

The correlation co-efficient must be exactly 1 whether it can be positive or negative i.e. -1 or +1 is called as Perfect Correlation.

41) What is Linear Correlation?

If change in one variable is accompanied by change in another variable in a constant ratio is called as linear correlation.

42) State any two merits of Karl Pearson's correlation?

Following are the merits of Pearson's correlation: …………
- ✓ It helps to find direction of correlation.
- ✓ It is most widely used method.
- ✓ It helps in comparison.
- ✓ It is useful in further algebraic treatment.

43) State any two limitations of Karl Pearson's correlation?

Following are the merits of Pearson's correlation: ....
- ✓ It is Time consuming.
- ✓ It is affected by extreme values
- ✓ It requires careful interpretation.

44) Give the meaning of Probable Error?

The probable error is a statistical measure which provides for two limits within which all the answers, obtained from different sample pairs of the population, will fall.

$$PE = \mathbf{0.6745} \ \frac{\mathbf{1- r^2}}{\mathbf{\sqrt{N}}}$$

45) What do you mean by Rank correlation?

Spearman's Rank Correlation is a technique used to test the direction and strength of the relationship between two variables.

46) State any two features of Rank Correlation?

Following are the Features of Spearman's Rank Correlation:
1. It is based on ranks rather than actual observations.
2. The sum of the differences of rank correlation two variables shall be zero.
3. It can be interpreted like Karl Pearson's coefficient of correlation.
4. It lies between -1 and +1.

47) State any two merits of Rank Correlation?

Following are the Advantages of Spearman's Rank Correlation:
1. It is easy to understand and simple to calculate.
2. It is suitable for qualitative data.
3. This method is also suitable for abnormal data.
4. It is only method for ranks.

48) Define the term regression?

Regression means returning back to averages or previous stage or earlier stage. It is a technique for predicting, the value of a dependent variable on the basis of independent variable.

49) Give the meaning of regression line?

Regression line is the line in which gives the best estimate of one variable from the value of any other given variable. It gives the relationship between the two variables in mathematical form.

50) What is meant by regression analysis?

Regression analysis is the statistical approach to forecasting change in dependent variable on the basis of change in one or more independent variable.

51) State any two assumptions of regression analysis?

Following are the important assumptions of regression analysis:
- ✓ Existence of actual linear relationship.
- ✓ It is used to estimate the values within the range for which it is valid.
- ✓ The relationship between the dependent and independent variable remains the same till the regression is calculated.
- ✓ The dependent variable takes any random value but the value of the independent variable are fixed.

52) What is meant by Index Numbers?

An index number is a statistical device for comparing the general level of magnitude of a group of related variables in two or more situation.

53) State different kinds of Index Numbers.

Following are the kinds / types of index numbers: ....
- a) Price index numbers
- b) Quantity index numbers
- c) Value index numbers

54) State any two features of Index numbers.

Following are features of Index numbers: ....
- a) Index numbers are specified averages
- b) Index numbers are expressed in percentage
- c) Index numbers are for comparison
- d) Index numbers measure the effect of changes over a period of time.

55) State any two Uses / Benefits of Index numbers.

Following are the Uses / Benefits of Index numbers: ......
- a) They measure the relative change.
- b) They are better comparison.
- c) They are good guides.
- d) They are economic barometers.

56) What is Base Year?

A Base year is the year used for comparison for the level of a particular economic index. It will be indicated as **0**.

57) What is Current Year?

The year for which comparisons are required is called as Current Year. It will be indicated as **1**.

58) Why index numbers are called "Economic barometers"?

Index numbers are called as economic barometers. Because, they are used to check the economic growth of a nation. They can also influence the economic activities of the country.

59) What is Consumer Price Index number?

A consumer price index measures changes in the price level of consumer goods and services purchased by households.

Consumer price index is also called as cost of living index.

60) What is Factor Reversal Test?

It holds that the product of a price index and the quality index should be equal to the corresponding value index.

Factor Reversal Test: $P_{01} \times Q_{01} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0}$

61) What is Time Reversal Test?

Time reversal test is a test to determine whether a given method will work both ways in time, forward and backward.

Time reversal test: $P_{01} \times P_{10} = 1$

62) Why fisher's index is Ideal?

Fisher's index number is called as Ideal. Because, it corrects the positive price bias in the Laspeyres price index and the negative price bias in the Paasche price index.

Time: 2 ½ Hours   **Max. Marks: 60**

SECTION - A

1. Answer **any Five** sub-questions. Each question carries **2** Marks

(5x2=10)

a) Define Statistics in Plural sense?

b) Define dispersion?

c) What is meant by Skewness?

d) Give the meaning of Probability?

e) What is meant by Positive correlation?

f) If, r = 0.6 and N = 64 of a distribution, find the probable error.

g) What is meant by Index Numbers?

SECTION – B

Answer **any Three** of the following. Each carries **Four** Marks **(3x4=12)**

2. Calculate simple mean and weighted mean for the following data:

Subject	Marks	Weights
Kannada	60	1
English	75	2
Business statistics	63	1
Corporate accounting	59	3
Cost accounting	53	3

3. Find the co-efficient of Mean deviation from the following data taking median as base.

Daily wages (Rs.)	No. of workers
Less than 60	8
" 70	24
" 80	56
" 90	96
" 100	136
" 110	178
" 120	191
" 130	200

4. A single card is drawn at random from a standard pack of 52 playing cards. What is the probability that it is: ....

   a) A King or a club

   b) A King or Queen

   c) Black or Spade

5. The correlation co-efficient between the variables X and Y is r = 0.60. if σx = 1.50, σy = 2, $\overline{X}$=10, $\overline{Y}$ = 20. Calculate two regression equations.

6. Construct the consumer price index number by using aggregate expenditure method from the following data:

Commodities	Quantity consumed in 2010	Price (Rs.) 2010	Price (Rs.) 2014
A	12	50	60
B	8	40	45
C	4	70	80
D	9	70	90
E	5	20	40
F	2	200	200

SECTION – C

Answer **any Three** of the following. Each carries **10** Marks **(3x10=30)**

7. Calculate the value of Mean, Median and Mode from the following data:

Profits (in lakhs)	4-7	8-11	12-15	16-19	20-23	24-27	28-31	32-35	36-39
Frequency	6	10	18	30	15	12	10	6	2

8. From the prices of shares of X and Y company given below, state which company's share prices are more variable.

Share prices of X Co.,	55	54	52	53	56	58	52	50	51	59
Share prices of Y Co.,	108	107	105	105	106	107	104	103	104	101

9. From the following data of the marks obtained by 10 students in Accounts and Statistics. Calculate Pearson's correlation.

Roll No :	1	2	3	4	5	6	7	8	9	10
Marks in A/c:	20	25	60	45	80	28	55	65	30	75
Marks in Statistics :	25	50	55	56	60	70	72	78	80	63

10. From the data given below find: ….
  a) The two regression equations.
  b) The most likely marks in statistics when marks in income tax is 50
  c) The coefficient of correlation with bxy and byx

  Marks in income tax   : 35, 37, 28, 42, 48, 54, 65, 40, 35, 50

  Marks in statistics   : 60, 55, 75, 70, 80, 35, 65, 35, 50, 60

11. Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Commodities	2010		2014	
	Price (Rs.)	Quantity	Price (Rs.)	Quantity
A	25	110	32	150
B	82	95	94	120
C	125	30	143	30
D	92	185	89	230
E	29	200	31	235
F	110	20	100	22

SECTION – D
Answer **any 1** of the following questions. Each carries **8** Marks **(1x8=8)**

12. A) Draw less than and more than ogive curves for the following frequency distribution and obtain median graphically. Verify the result.

C.I.	0-20	20-40	40-60	60-80	80-100	100-120	120-140	140-160
F	5	12	18	25	15	12	8	5

  B) Find Karl Pearson's co-efficient of skewness for the following data:

Marks	: 0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of Students:	5	12	16	25	10	8	4

III Semester B. Com. Examination, March/April – 2023
(NEP) (Semester Scheme) Model Paper - 02
COMMERCE
Paper – 3.2: Business Statistics

Time: 2 ½ Hours   Max. Marks: 60

SECTION - A

1. Answer **any Five** sub-questions. Each question carries **2** Marks

(5x2=10)

   a) State any four functions of statistics.
   b) Define Standard deviation?
   c) State any two objectives of measure of dispersion.
   d) Give the meaning of an Event?
   e) Define correlation?
   f) What is meant by Regression?
   g) What is meant by Time Reversal Test?

SECTION – B

Answer **any Three** of the following. Each carries **Four** Marks **(3x4=12)**

2. Calculate Median from the following data:

Marks	: 50	40	30	20	10
Frequency	: 10	40	20	12	16

3. From the following data compute Quartile Deviation (QD) and its co-efficient.

Marks	: 10-20	20-30	30-40	40-50	50-60
No. students	: 15	18	20	9	6

4. In a group of 24 cricket players. 14 are batsmen, 10 bowlers. 6 batsmen and 2 bowlers are from Karnataka. If a player is chosen at random, what is the probability of choosing a batsmen or a player from Karnataka?

5. From the following data calculate the rank correlation co-efficient.
Ranks by Judge A: 40, 50, 30, 80, 60, 90, 70, 20
Ranks by Judge B: 80, 30, 40, 50, 20, 10, 90, 60

6. Construct the Consumer Price Index from the following data:

Commodities	P_0	P_1	W
A	2	4	2
B	4	6	4
C	6	6	3
D	2	3	1
E	1	1	1

SECTION – C

Answer **any Three** of the following. Each carries **10** Marks **(3x10=30)**

7. Calculate the value of Mean, Median and Mode from the following data:

X	Less than 10	20	30	40	50	60	70	80
F	4	16	40	76	96	112	120	125

8. You are given below the daily wages paid to workers in two factories X and Y. Find:
   a) Which factory pays more wages?
   b) In which factory are more variable?

Daily wages (Rs.)	50-60	60-70	70-80	80-90	90-100
No. of workers: Factory X	30	60	90	40	20
Factory Y	40	70	100	20	10

9. Find Pearson's coefficient of correlation between average profits and average advertisement expenditure per shop and interpret result.

No. of Shops   :   12   18   25   20   10
Total profits (Rs): 7,200   5400   10,000   3000   1800
Total Advert.
Expenditure (Rs): 1,200   3,600   7,500   1,000   600

10. From the data given below: ....
Construct two regression equations for the following data and estimate the value of X when Y=70 and Y when X=650.

   X: 100, 200, 300, 400, 500, 600, 700

   Y: 30, 50, 60, 80, 100, 110, 130

11. Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Commodities	2015		2016	
	Price (Rs.)	Quantity	Price (Rs.)	Quantity
M	20	8	30	10
N	50	10	40	8
O	40	5	50	12
P	60	20	60	16
Q	10	6	40	10

Answer **any 1** of the following questions. Each carries **8** Marks **(1x8=8)**

12. A) Calculate the Pearson's coefficient of correlation from the following data using 44 and 26 (as assumed mean) respectively as the origin of X and Y.

        X: 43, 44, 46, 40, 44, 42, 45, 42, 40, 42, 57, 48

        Y: 29, 31, 19, 18, 18, 27, 27, 29, 41, 30, 26, 10

OR

B) Find Karl Pearson's coefficient of skewness for the given distribution:

X:	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
F:	2	5	7	13	21	16	8	3

III Semester B. Com. Examination, March/April – 2023
(NEP) (Semester Scheme) Model Paper - 03
COMMERCE
Paper – 3.2: Business Statistics

Time: 2 ½ Hours   Max. Marks: 60

SECTION - A

1. Answer **any Five** sub-questions. Each question carries **2** Marks

(5x2=10)

   a) State any four objectives of statistics.
   b) Define Co-efficient of variation?
   c) List any four methods of studying variation.
   d) Give the meaning of an Experiment?
   e) Give the meaning of Perfect correlation?
   f) State any two assumptions of regression analysis.
   g) Mention any two methods of measuring consumer price index.

SECTION – B

Answer **any Three** of the following. Each carries **Four** Marks **(3x4=12)**

2. Calculate Median from the following data:

Income in Rs.	: More than 0	10	20	30	40	50
No. of workers	: 100		90	75	50	25	10

3. From the following data compute Quartile Deviation:

X :	15	24	33	42	51
F :	3	5	15	11	5

4. A bag contains 30 balls number from 1 to 30. One ball is drawn at random. Find the probability that the number of ball is a multiple of 5 or 6?

5. Calculate the Rank correlation co-efficient from the following data:

X :	70	80	65	78	62	65	82	65
Y :	13	15	12	14	13	11	16	10

6. Calculate consumer price index by using family budget method for the year 2010 with 2009 as base year for the following data.

Items	Weights	Price (Rs.)	
		2009	**2010**
Food	35	150	140
Rent	20	75	90
Clothing	10	25	30
Fuel	15	50	60
Miscellaneous	20	60	80

SECTION – C

Answer **any Three** of the following. Each one carries **10** Marks **(3x10=30)**

7. Calculate the Mean, Median and Mode from the following data:

X :	0-10	10-20	20-30	30-50	50-70	70-100
F :	5	10	17	40	62	60

8. Following are the marks obtained by two students Suraj and Dheeraj in ten tests of 100 marks each:

Tests	1	2	3	4	5	6	7	8	9	10
Suraj	40	80	76	48	52	72	68	56	60	56
Dheeraj	48	75	54	60	63	69	72	51	72	60

Find who is the better scorer and if consistency is the criteria for awarding prize who should get the prize?

9. Calculate Karl Pearson's co-efficient of correlation with the following data of 6 cities between density of population and death rates. Interpret the correlation by finding probable error.

Cities	:	A	B	C	D	E	F
Area [in Sq.km] :		150	180	100	60	120	80
Population [in 000's]:		30	90	40	42	72	44
No. of deaths :		300	1440	560	840	1224	312

10. The following table gives the results of capital employed and profit earned by a firm in 10 successive years.

	Mean	Standard Deviation
Capital employed (Rs. in 000'):	55	28.7
Profit earned (Rs. in 000') :	13	8.5
Co-efficient of correlation	0.96	

a) Obtain two regression equations.

b) Estimate the amount of profit to be earned if capital employed is Rs. 50,000.

c) Estimate the amount of capital to be employed if profit earned is Rs. 20,000.

11. Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Commodities	2015		2016	
	Price (Rs.)	Quantity	Price (Rs.)	Quantity
M	8	80	10	110
N	10	90	12	108
O	16	256	20	340
P	20	420	24	456
Q	25	550	32	704

SECTION – D

Answer **any One** of the following questions. Each carries **8** Marks **(1x8=8)**

12. A) Draw an ogives (less than and more than), calculate median from the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	3	8	12	20	24	12	7

OR

B) Find Karl Pearson's coefficient of skewness for the given distribution:

X:	40-50	50-60	60-70	70-80	80-90	90-100	100-110
F:	8	20	36	24	16	14	2

III Semester B. Com. Examination, March/April – 2023
(NEP) (Semester Scheme) Model Paper – 04
COMMERCE
Paper – 3.2: Business Statistics

Time: 2 ½ Hours   **Max. Marks: 60**

SECTION - A

1. Answer **any Five** sub-questions. Each question carries **2** Marks

(5x2=10)

    a) Mention any four types of statistical averages.
    b) What is meant by Mean deviation?
    c) If variance = 36, $\sum X$ = 150 and N = 10, find CV.
    d) Give the meaning of mutually exclusive event?
    e) Define rank correlation?
    f) What is meant by Regression analysis?
    g) What do you mean by current year?

SECTION – B

Answer **any Three** of the following. Each carries **Four** Marks   **(3x4=12)**

2. Calculate Median and mode from the following data:

Marks	10-20	20-30	30-40	40-50	50-60
Frequency	10	40	20	12	16

3. From the following data compute Mean Deviation from mean (MD) and its co-efficient.

Marks	10-20	20-30	30-40	40-50	50-60
No. students	15	18	20	9	6

4. Calculate the probability of picking a card out of a pack:
    a) A diamond or a spade
    b) A king or a Queen

5. Obtain two regression equations from the following data, where the deviations were obtained from mean.
    $\overline{X}$ = 4, $\overline{Y}$ = 2, $\sum x^2$=1680, $\sum y^2$=320, n=20, $\sum xy$=480.

6. Construct the Consumer Price Index number by using family budget method:

Items	Units	Price in 2015	Price in 2018
A	100	40	60
B	80	30	45
C	30	70	100
D	10	30	50
E	8	20	40
F	2	900	1200

SECTION – C

Answer **any Three** of the following. Each one carries **10** Marks **(3x10=30)**

7. Calculate the value of Mean, Median and Mode from the following data:

X	More than 10	20	30	40	50	60	70	80
F	5	16	50	80	98	118	125	140

8. Life of two models of refrigerators obtained in a survey in given below. Find which model has greater uniformity of life:

Life (No. of years)	0-2	2-4	4-6	6-8	8-10	10-12
Model X	6	15	13	8	5	3
Model Y	3	16	12	18	10	1

9. Calculate Karl Pearson's correlation coefficient between net income and expenditure from the data given below and also calculate the probable error.

Income : 150 148 190 220 245 262 237
Expenses : 140 132 168 214 237 260 222

10. From the following data relating to years of service in a factory of seven persons in a specialized field and their monthly income.

Year of service : 11 7 9 5 8 6 10
Income (in '000' Rs.) : 7 5 3 2 6 4 8

Find two regression equations and also estimate the income of a person of 12 years' service.

11. Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Commodities	Price		Quantity	
	Price (Rs.)	Quantity	Price (Rs.)	Quantity
A	12	20	55	120
B	10	12	100	80
C	14	15	60	80
D	16	18	30	70
E	18	20	40	40
F	20	15	70	60
G	20	16	90	100
H	15	18	80	80

SECTION – D

Answer **any One** of the following questions. Each carries **8** Marks **(1x8=8)**

12. A) Convert the following distribution into more than cumulative frequency curve and less than frequency curve and find median graphically.

Wages (Rs.)	0-20	20-40	40-60	60-80	80-100
No. of workers	82	112	150	95	48

OR

B) Calculate Karl Pearson's coefficient of correlation between price and demand from the following data.

Price in Rs   :   21 22 23 24 25 26 27 28 29
Demand in units: 20 19 19 17 17 16 16 15 14

III Semester B. Com. Examination, March/April – 2023
(NEP) (Semester Scheme) Model Paper – 05
COMMERCE
Paper – 3.2: Business Statistics

Time: 2 ½ Hours   Max. Marks: 60

SECTION - A

1. Answer **any Five** sub-questions. Each question carries **2** Marks
(5x2=10)

a) How do you calculate 'Mode' when it is ill-defined?
b) State any two objectives of measures of dispersion.
c) If Q_1 is 20 Q_3 is 45, what is its co-efficient of quartile deviation?
d) Give the meaning of not mutually exclusive event?
e) What do you mean by Probable error?
f) What is meant by Regression line?
g) Why index numbers are called "Economic barometers"?

SECTION – B

Answer **any Three** of the following. Each carries **Four** Marks   **(3x4=12)**

2. Calculate mean wages of worker from the following data:

Wages :	48-56	56-64	64-72	72-80	80-88	88-96
No. of Workers :	8	3	11	14	5	7

3. From the following data compute standard Deviation.

Marks :	10-20	20-30	30-40	40-50	50-60
No. students:	15	18	20	9	6

4. The unbiased dice are drawn find the probability that,
a) Both the dice show the same digit
b) The first dice show 5

5. Calculate rank correlation coefficient from the following data:
X: 53, 98, 95, 81, 70, 81, 65, 81
Y: 25, 47, 82, 76, 53, 61, 75, 70

6. Construct the Consumer Price Index number by aggregate expenditure method:

Items	Units	Price in 2015	Price in 2018
A	100	40	60
B	80	30	45
C	30	70	100
D	10	30	50
E	8	20	40
F	2	900	1200

SECTION – C

Answer **any Three** of the following. Each one carries **10** Marks **(3x10=30)**

7. From the following data of the wages of 122 workers, determine the model wages with help of grouping table and analysis table.

Wages	No. of workers
100-110	4
110-120	6
120-130	20
130-140	32
140-150	33
150-160	17
160-170	8
170-180	2

8. The scores of two batsman 'A' and 'B' in ten cricket matches is given below:

A: 32 28 47 63 71 39 10 60 96 14

B: 19 31 48 53 67 90 10 62 40 80

Find whether batsman A or B is more consistent in scoring.

9. Find Karl Pearson's correlation co-efficient between age and playing habits of the following students.

Age   :   15, 16, 17, 18, 19, 20

No. of students :   250, 200, 150, 120, 100, 80

Regular players :   200, 150, 90, 48, 30, 12

10. The following data related to the age of husbands and wives:

Age of Husbands	Age of wives
25	20
28	26
30	29
35	25
36	18
38	26
39	35
42	35
55	46

Obtain the two regression equations and determine the most likely age of husband for the age of wife 25 years.

11. Construct Fisher's Ideal Index number from the following data and how it satisfies TRT and FRT.

Commodities	Price		Quantity	
	Price (Rs.)	Quantity	Price (Rs.)	Quantity
A	12	20	55	120
B	10	12	100	80
C	14	15	60	80
D	16	18	30	70
E	18	20	40	40
F	20	15	70	60

SECTION – D

Answer **any One** of the following questions. Each carries **8** Marks **(1x8=8)**

12. A) Convert the following distribution into more than cumulative frequency curve and less than frequency curve and find median graphically.

Wages (Rs.)	0-10	10-20	20-30	30-40	40-50
No. of workers	80	96	120	90	45

OR

B) Find Karl Pearson's coefficient of skewness for the given distribution:

X:	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F:	10	12	24	32	28	11	3

III Semester B. Com. Examination, March/April – 2023

(NEP)

COMMERCE

Paper – 3.2 (DSC): Business Statistics

Time: 2 ½ Hours   Max. Marks: 60

SECTION - A

1. Answer **any five** sub-questions. Each sub-question carries **2** Marks

(5x2=10)

- **a)** Give the meaning of simple average.
- **b)** How range is calculated?
- **c)** If Q_1 =60.125, Q_3 = 62.6875 and Median = 61.345. Find Bowley's co-efficient of skewness.
- **d)** Give the meaning of outcome?
- **e)** Why Fisher's idex number is called as "Ideal"?
- **f)** If N = 8 and $\sum D^2$ = 40, find rank correlation co-efficient.
- **g)** If b_{xy} = 0.8 and b_{yx} = 0.45, find r.

SECTION – B

Answer **any Three** of the following. Each carries **Four** Marks   **(3x4=12)**

2. Form a frequency distribution from the following data by exclusive method taking four as the width of the class.

10	17	15	22	11	16	19	24	29	18
25	26	32	14	17	20	23	27	30	12
15	18	24	36	18	15	21	28	33	38
34	13	10	16	20	22	29	19	23	31

3. The following data is available for 10 innings of T20 cricket of two teams:

	Runs scored by RCB	Runs scored by CSK
Mean	210	186
Variance	14400	15625

Correlation co-efficient between runs scored by RCB and runs scored by CSK = 0.36. Compute two regression equations. Also estimate the runs scored by RCB when CSK scored 250 and runs scored by CSK when RCB scored 200 runs.

4. Calculate mean deviation (from mean) and it's co-efficient.

Marks: 59   32   67   43   22   17   64   55   47   80   25

5. From a pack of 52 playing cards, a card is drawn at random. What is the probability that it is either queen or ace?

6. Construct the Consumer Price Index number for the following data by family budget method.

Items	Units	Units consumed in 2015	Price in 2015	Price in 2022
A	Quintals	5 quintals	1200	2800
B	Kg	25 kgs	20	35
C	Litre	20 litres	120	150
D	Meter	25 meters	50	65
E	Unit	20 units	100	200

SECTION – C

Answer **any Three** of the following. Each one carries **10** Marks **(3x10=30)**

7. Compute Mean, Median and Mode from the following data:

Wages	No. of workers
100-200	2
200-300	10
300-400	16
400-500	32
500-600	10
600-700	8
700-800	2

8. Find Mean, Mode, S.D. and Karl Pearson's co-efficient of skewness for the following data:

Kids per family : 0 1 2 3 4 5 6
No. of families : 7 10 16 25 18 11 8

9. Compute Fisher's Ideal Index number.

Items	2021		2023	
	Value	Quantity	Value	Price
A	300	150	480	4
B	50	10	90	6
C	48	12	50	5
D	120	60	100	2
E	60	20	105	3.5

10. In a class of Sri Subhash Chandra Bose College a preparatory exam is conducted for 10 students in Mathematics and Statistics subjects. Calculate the Karl Pearson's co-efficient of correlation. (Take 40 in mathematics and 50 in statistics as assumed mean).

Marks in Mathematics	45	70	65	30	90	40	50	75	85	60
Marks in Statistics	35	90	70	40	95	40	60	80	80	50

11. Weights (In Kgs) of Brothers and Sisters of a particular locality is given below: ....

Weights of Brothers	Weights of Sisters
40	20
48	24
52	28
68	36
72	52

a) Find the two regression equations.

b) Estimate Brother Weight when sister Weight = 80 kgs.

c) Estimate the value of co-efficient of correlation through regression co-efficient.

SECTION – D

Answer **any One** of the following questions. Each carries **8** Marks **(1x8=8)**

12. a) Find Mode graphically for the following frequency distribution:

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	5	8	12	7	5	3

OR

b) Two judges vijayprakash and Rajeshkrishnan participated and assigned marks in a singing competition as under for 9 participants

Marks assigned by Vijayprakash:

Singer Roll No.	1	2	3	4	5	6	7	8	9
Marks	48	60	72	62	56	40	39	52	30

Marks assigned by RajeshKrishnan:

Singer Roll No.	1	2	3	4	5	6	7	8	9
Marks	62	78	65	70	38	84	60	32	31

You are required to indicate the degree of agreement about the taste in singing between two judges according to spearmen's rank correlation co-efficient.