J.K. SHAH®
CLASSES
a Veranda Enterprise

Foundation → Intermediate → Final CA ↗

CA FOUNDATION

FAST TRACK

STATISTIC

Head Office

Shraddha, 4th Floor, Old Nagardas Road,
Near Chinai College, Andheri (E), Mumbai - 400 069.

022 - 2683 66 66

/officialjksc jkshahclasses.com

J.K. SHAH® CLASSES
a Veranda Enterprise

INDEX

1A STATISTICAL DESCRIPTION OF DATA
(Introduction to Statistics)

Introduction:

The word "STATISTICS" has its origin from the following:

• Latin	-	STATUS
• German	-	STATISTIK
• French	-	STATISTIQUE
• Italian	-	STATISTA

Statistics in India

- Kautilya recorded birth and death in Arthashastra during Chandragupta Maurya's regime.
- Abul Fazal, during Akbar's regime, recorded agriculture in the book Ain-i-Akbari.

"STATISTICS" DEFINED

IN SINGULAR SENSE

It is defined as the scientific method of collecting, presenting, analyzing the data and drawing inference from the same.

IN PLURAL SENSE

By Statistics, we mean aggregate of facts which are known as "DATA" (Singular Datum).

Features of Statistics:

a) Statistics deals with masses and not individuals.

b) Statistics deals with quantitative data . Qualitative data are also to be expressed in quantitative terms.

c) It is aggregate of facts (plural sense).

d) It refers to scientific methods of analyzing data.(Singular Sense)

e) It is science as well as an art.

f) Data are affected by multiplicity of causes.

g) Data should be collected in a systematic manner and for a pre-determined purpose.

h) Data should be comparable.

i) All Statistics are Numerical Statements but all Numerical Statements are not statistics

APPLICATION OF STATISTICS

Statistics is used in

a) Mathematics

b) Economics

c) Accountancy

d) Auditing

e) Business and industry

f) Social Science

g) Medical Sciences & Biology

h) Different Statistical techniques used in Business, Economics and Industry.

i) Management.

LIMITATIONS OF STATISTICS

i. Statistics does not study qualitative phenomenon directly.

ii. Statistics does not study individuals.

iii. Statistical laws are not exact.

iv. Statistical data are liable to be misused.

v. Statistics results are true on the average sense only. They are not exact

FEW TERMS COMMONLY USED IN STATISTICS.

i. **Data :** It is a collection of observations, expressed in numerical figures, obtained by measuring or counting.

ii. **Population :** It is used to denote the totality of the set of objects under considering.

iii. **Sample :** A sample is a selected no. of individuals each of which is a member of the population. It is examined with a view to assessing the characteristics of the population.

iv. **Characteristic :** A quality possessed by an individual person, object or item of a population is called a characteristic e.g. Height, age, nationality, etc.

v. **Variable & Attribute :** Measurable characteristics which are expressed numerically in terms of some units are called as variables or variates e.g. age, height, income, etc. Non-measurable characteristics is a qualitative characteristic which is called as attribute e.g. sex, marital status, employment status, etc.

vi. **Continuous & Discrete Variable :** A variable which can assume for its value any real quantity within a specified interval is a continuous variable e.g height, weight etc and the variables which can assume only whole numbers are discrete variables eg :-. number of members in the family, no of accidents etc.

CLASSWORK SECTION

Related MCQ's:

1. Which of the following statement is true?
   a) Statistics is derived from the French word "Statistik".
   b) Statistics is derived from the Italian word "Statista".
   c) Statistics is derived from the Latin word "Statistique".
   d) None of these

2. Statistics is considered with:
   a) Qualitative information
   b) Quantitative information
   c) Both a) and b)
   d) Either a) or b)

3. Which of the following would you regard as discrete variable:
   a) height
   b) weight
   c) number of persons in a family
   d) wages paid to workers

4. An attribute is:
   a) A measurable characteristics
   b) A quantitative characteristics
   c) A qualitative characteristic
   d) All of the above

5. Annual income of a person is:
   a) An attribute
   b) A continuous variable
   c) A discrete variable
   d) Either b) or c)

❖ A STATISTICAL ENQUIRY PASSES THROUGH THE FOLLOWING PHASES :

1. COLLECTION OF DATA

2. SCRUTINY OF DATA

3. CLASSIFICATION OF DATA

4. PRESENTATION OF DATA

LLECTION OF DATA (DATUM IN SINGULAR)

Data : Data are aggregate of facts i.e. Quantitative information about characteristic under study.

Types of Data

Primary Data

These data are collected for a specific purpose directly from the field of enquiry. These are original in nature

Secondary Data

1. Secondary Data are numerical information which have been previously collected as primary data by some agency for a specific purpose but are now complied from that source for use in a different connection. Sources of Secondary Data.

   i. Publications of Central and State Governments, of Foreign Governments, and international bodies like ILO, UNO, UNESCO, WHO, etc.

   ii. Publications of various Chambers of Commerce, Trade Associations, Co-operative Societies, etc.

Methods of Collecting Primary Data

- Direct Observation Method
- Mailed Questionnaire Method
- Interview Method
  - Direct Personal Interview
  - Indirect Interview
  - Telephonic Interview

(1) DIRECT OBSERVATION METHOD:

It is the best method of data collection, but time consuming, laborious and covers only a small area.

(2) MAILED QUESTIONNAIRE METHOD:

Under this method, data are collected by means of framing a well drafted and properly sequenced questionnaire covering all the important aspects of the problem under study and sending them to the respondents. (Although a wide area can be covered but non-response is maximum under this method).

(3) INTERVIEW METHOD:

a. Direct Personal Interview Method:

Under this method, the investigator collects information directly from the respondents. In case of natural calamities like earthquake, cyclone or epidemic the data can be collected much more quickly and accurately.

b. Indirect Interview Method:

It is used when the respondents can't be reached directly and the data is collected from the persons associated with the problems. E.g. in case of accidents this method is used.

Note : The above two methods are more accurate but not suitable for large area.

c. Telephonic Interview Method:

It is quick, less expensive and covers largest area. Under this method, the researcher himself gathers information by contacting the interviewee over the phone. It is less consistent compared to the other two methods. Amount of non –response is maximum under this method.

Related MCQ's:

6. A statistical survey may either be _________ purpose or _________ purpose survey.
   a) general, specific
   b) general, without
   c) all, individual
   d) none of the above

7. Data originally collected for an investigation are known as:
    a) primary data
    b) secondary data
    c) both primary and secondary data
    d) none of the above

8. Primary data are:
    a) always more reliable compared to secondary data
    b) less reliable compared to secondary data
    c) depends upon the care with which data have been collected
    d) depends upon the agency collecting the data

9. In case of a rail accident, the appropriate method of data collection is by :
    a) Direct interview
    b) Personal interview
    c) Indirect interview
    d) All of the above

2. SCRUTINY OF DATA

It means checking the data for accuracy & consistency. Intelligence, patience & experience is used by scrutinizing the data.

3. CLASSIFICATION OF DATA

Definitions : When the items / individuals are classified, according to some common non-measurable characteristics processed by them, they are said to form a statistical class, and when they are classified according to some common measureable characteristics processed by them, they are said to form a statistical group.

Types of Classifications

Geographical (or) Spatial i.e. Areawise	Chronological (or) Temporal or Time Series i.e. on the basis of time	Qualitative (or) Ordinal	Quantitative(or) Cardinal

10. The primary rules that should be observed in classification:

I. As far as possible, the class should be of equal width.

II. The classes should be exhaustive.

III. The classes should be un-ambiguously defined.

a) Only I and II

b) Only II and III

c) Only I and III

d) All I, II and III

4. Presentation of Data

Presentation of Data

Textual | Tabular | Graphical

Textual

Textual Presentation : It is in written form. It is simple but dull, monotonous & comparison is not possible

Tabular

Tabular Presentation : Presentation of data with the help of a statistical table having rows & columns.

Advantages of Tabulation are as follows:

1. Complicated data can be represented.
2. It is a must for diagrammatic representation.
3. Statistical analysis is not possible without tabulation.
4. It facilitates comparison between rows & columns.

3. DIFFERENT PARTS OF A TABLE (4 Parts)

TABLE

STUB | CAPTION | BODY | BOXHEAD

1 **Stub :** Stubs are the headings or designations for the horizontal rows.

2. **Captions :** Captions are the headings or designations for vertical columns.

3. **Body :** The arrangement of the data according to the descriptions given in the captions (columns) and stubs(rows) forms the body of the table. It contains the numerical information which is to be presented to the readers and forms the most important part of the table.

4. **Box-head:** The entire upper part of the table is known as box-head.

Other Parts :

5. **Title :** Every Table must be given a suitable title, which usually appears at the top of the table (below the table number or next to the table number). A title is meant to describe in brief and concise form the contents of the table and should be self-explanatory.

6. **Table Number :**

7. **Head Note :**

8. **Foot Note :**

9. **Source Note**

FORMAT OF A BLANK TABLE

Title

[Head Note or Prefatory Note (if any)]

<table>
<tr><td rowspan="3">Stub
Heading
↓</td><td colspan="5">Captions</td><td rowspan="3">Total</td></tr>
<tr><td colspan="2">Sub-Heads</td><td colspan="3">Sub-Heads</td></tr>
<tr><td>Column
Head</td><td>Column
Head</td><td>Column
Head</td><td>Column
Head</td><td>Column
Head</td></tr>
<tr><td></td><td colspan="5">Body</td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Foot Note :

Source Note :

Types of Tabulatio

Types of Tabulation

Simple | Complex

Simple Tabulation : In this type the number or measurement of the items are placed below the headings showing the characteristics.

Complex Tabulation : In this type each numerical figure in the table is the value of the measurement having the characteristics shown both by the column and the row headings.

Related MCQ's:

11. When the accuracy in presentation is more important than the method of presentation it is done through:

a) Textual b) Diagrammatic

c) Tabular d) Either b) or c)

12. The unit of measurement in tabulation is shown in

a) box head b) body c) caption d) stub.

13. For tabulation, 'caption' is :

a) the lower part of the table.

b) the main part of the table.

c) the upper part of the table.

d) the upper part of a table that describes the column and sub-column.

14. 'Stub' of a table is the

a) right part of the table describing the columns.

b) left part of the table describing the columns.

c) right part of the table describing the rows

d) left part of the table describing the rows.

15. A table has ______ parts.

a) Two b) Three c) Four d) Five

Diagrammatic Representation of Data

1. Diagrammatic Representation are mainly done by charts (or graphs) and figures.

2. A chart or graph is inferior to a table or numbers as a method of presenting data, since one can get only approximate idea from it, but its advantage is that it emphasizes certain facts and relations more than numbers do.

Advantages :

1. It is more attractive and informative to an ordinary person.

2. A complex problem can sometimes be clarified easily by a diagram.

3. It reveals the hidden facts which are not apparent from the tabular presentation.

4. Two or more sets of values can be compared very easily from a diagram.

5. It shows the relation of the parts to the whole.

Types of Diagrams

Without Frequency	With Frequency (Frequency Curves)
1. Line Chart or Line Graph or Line Diagram or Historigram Chart (one dimensional)	1. Histogram or Area Diagram (Two dimensional)
2. Bar Diagram or Bar Chart (one dimensional)	2. Frequency Polygon (Two dimensional)
3. Pie Chart (Two dimensional)	3. Frequency Curve (Two dimensional)
	4. Cumulative Frequency Polygon or Ogive (Two dimensional)

Each of the Diagram is described below:

Line Diagram :

It is used for time related data (Time series).
When there is wide range of fluctuations, logarithmic or ratio charts are used.

Multiple Line Chart :

It is used for representing 2 or more related series expressed in same units.

Multiple Axis Chart :

Multiple Axis Chart is used for representing two or more related series expressed in different units.

Semi-Logarithmic Graph or Ratio Chart :

Semi-Logarithmic Graph or Ratio Chart is a line diagram drawn on a special type of graph paper which shows the natural scale in the horizontal direction and the logarithmic or ratio scale in the vertical direction. The semi-log graph is used where ratios of change are more important than absolute amounts of change.

Bar Diagram

1. **Vertical Bar Chart (or Colum Chart) :**

   This is generally used to represent a time series data or a data which is classified by the values of the variable. (Measurable characteristics).

2. **Horizontal Bar Chart :**

   This is used to represent data classified by attributes **or data varying over space.** (i.e. non-measurable characteristics).

3. **Grouped or Multiple or Compound Bar Chart):**

   These are used to compare related series.

4. **Component /Sub divided Bar Chart:**

   These are used for representing the data divided into different components

5. **Percentage Bars :**

   Percentage Bars are particularly useful in statistical work which requires the portrayal of relative changes.

6. **Deviation Bars**

   Deviation Bars are popularly used for representing net quantities – excess or deficit i.e. net profit, net loss, net exports or imports, etc. Such bars can have both positive and negative values. Positive values are shown above the base line and negative values below it.

7. Broken Bars

In certain series there may be wide variations in values – some value may be very small and others very large. In order to gain space for the smaller bars of the series, larger bars may be broken.

PIE CHART / PIE DIAGRAM / CIRCLED DIAGRAM

This is a very useful diagram to represent data which are divided into a number of categories. The diagram consists of a circle divided into a number of sectors whose areas are proportional to the values they represent. Again the areas of the sectors are proportional to their angles at the centre. Therefore, ultimately the angles of the different sectors are proportional to the values of different components. The total value is represented by the full circle. Comparison among the various components or between a part and the whole of data can be made easily by this diagram.

Example :

Draw a pie chart to represent the following data on the proposed outlay during a Five-year Plan of a Government :

Items	₹ (in crores)
Agriculture	12,000
Industry & Minerals	9,000
Irrigation & Power	6,000
Education	8,000
Communication	5,000

Calculations for the angles of the pie chart

Items	Outlay (in crores ₹)	Angles (in egrees)
Agriculture	12,000	108
Industry & Minerals	9,000	81
Irrigation & Power	6,000	54
Education	8,000	72
Communication	5,000	45
Total	40,000	360

Working Note : 40,000 is represented by 360^0

1,000 is represented by $\frac{360}{40} = 9^0$

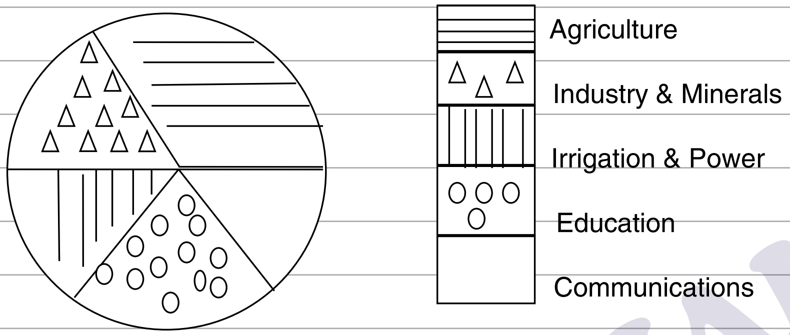
12,000 is represented by 12 x 9 = 108^0

9,000 is represented by 9 x 9 = 81^0

6,000 is represented by 6 x 9 = 54^0

8,000 is represented by 8 x 9 = 72^0

And 5,000 is represented by 5 x 9 = 45^0



DIAGRAMMATIC/GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTION

1. Histogram or Area Diagram

i) It consists of a set of adjoining vertical rectangles whose widths represent the class intervals and the heights represent the corresponding frequencies (for equal class width) and frequency densities (for unequal class width). Boundaries are plotted along the horizontal axis and the frequencies (or frequency densities) are plotted along the vertical axis

ii) The area of each rectangle is proportional to the frequency of the corresponding class.

iii) Mode is calculated graphically from Histogram.

iv) It helps us to get an idea about the frequency curve and frequency polygon.

v) Comparison among the frequencies can be made for different class intervals.

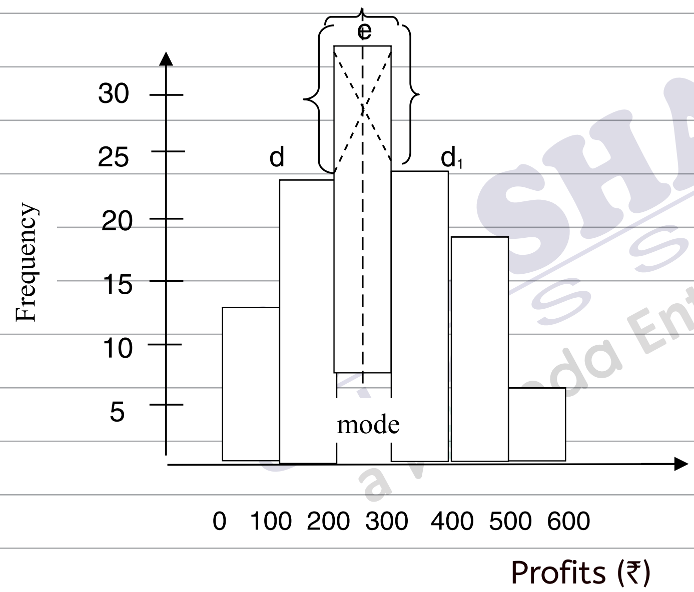
Example

The monthly profits in rupees of 100 shops are distributed as follows:

Profits per Shop	0-100	100-200	200-300	300-400	400-500	500-600
No. of Shops	12	18	27	20	17	6

Draw the histogram to the data and hence find the modal value.

In the histogram, the top right corner of the highest rectangle is joined by a straight line to the top right corner of the preceding rectangle. Similarly, top left corner of the highest rectangle is joined to the top left corner of the following rectangle. From the point of intersection of these two lines a perpendicular is drawn on the horizontal axis. The foot of the perpendicular indicates the Mode. This is read from the horizontal scale and the modal value is found to be 256 (in ₹) approximately.



2. Frequency Polygon and Frequency Curve

i) In this method, the frequency of each class is plotted against the mid-value of the corresponding class. The points thus obtained are joined successively by straight lines. The polygon is then completed by joining two end-points to the mid-values of two empty classes assumed in either side of the frequency distribution.

ii) Frequency polygon can be obtained from the histogram by joining the successive mid-points of the top of the rectangles which constitute the histogram and the polygon is completed in the same manner as before.

iii) If in a frequency distribution the widths of the classes are reduced, then the number of classes will increase. As a result the vertices of a frequency polygon will come very close to each other. In that case, if we join the points by smooth free hand line instead of straight lines, a smooth curve is obtained which is known as a Frequency Curve.

iv) Frequency Curve is a limiting curve case of frequency polygon.

3. Cumulative Frequency Polygon / Ogive Curve

1. It is a graphical representation of cumulative frequency distribution.
2. Median and all other partition values are calculated from ogives.
3. There are two types of ogives (i) Less Than Ogive (ii) More Than Ogive.
4. IN LESS THAN OGIVE LESS THAN CUMULATIVE FREQUENCIES ARE USED. AND IN CASE OF MORE THAN OGIVE, MORE THAN CUMULATIVE FREQUENCIES ARE USED AND THE OGIVE CURVE LOOKS LIKE ELONGATED "S". THESE ARE ALSO KNOWN AS "S" CURVE.

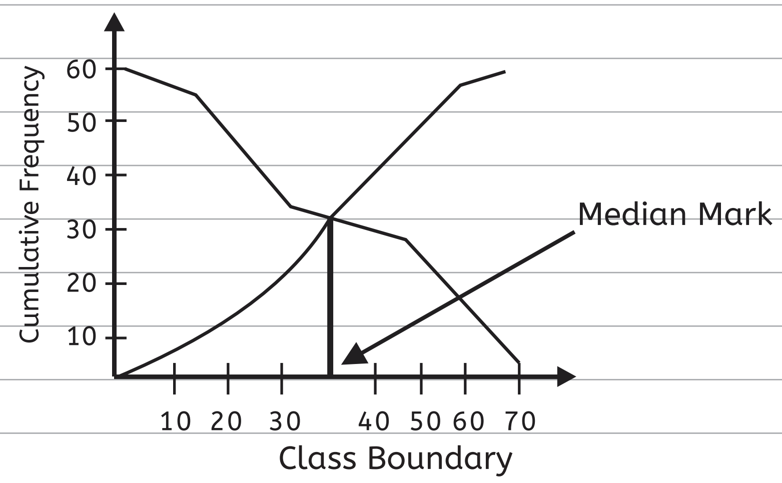
Example

Draw the cumulative frequency diagram (both more-than and less-than ogive) of the following frequency distribution and locate graphically the Median:

Marks-Group	0-10	10-20	20-30	30-40	40-50	50-60	60-70	Total
No. of Students	4	8	11	15	12	6	3	59

Calculation for Cumulative Frequencies

Class Boundary	Cumulative Frequency	
	Less than	More than
0	0	59
10	4	55
20	12	47
30	23	36
40	38	21
50	50	9
60	56	3
70	59	0



Less than and More than ogive of a frequency distribution

From the graph the median is found to be 34.5.

4. Other Frequency Curves

1. Bell Shaped (Symmetrical Curve):

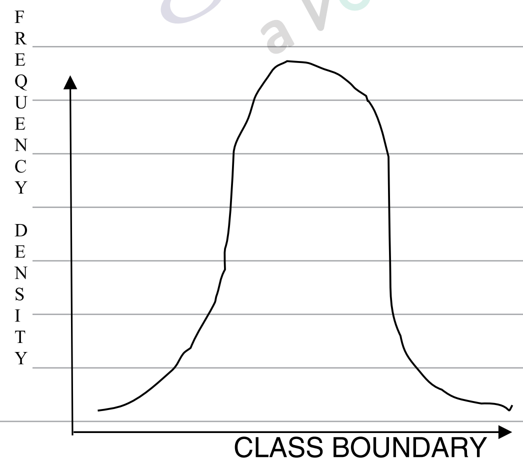
The most commonly used frequency curve use for the distribution of height, weight, profit, etc.

i. It is the limiting form of histogram and frequency polygon

ii. The area under the curve is taken to be unity.

iii. It enables us to understand symmetry of the distribution.

Diagram

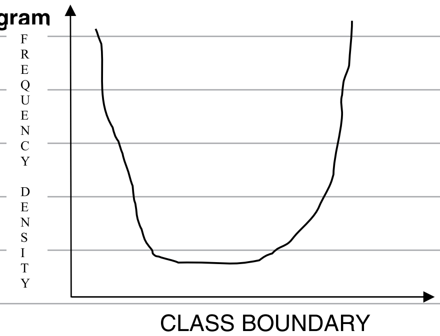


Bell Shaped Curve

2. U Shaped Curve

In this curve, the frequency is minimum at the central part, and slowly but steadily it reaches to two extremities. The distribution of people travelling on streets will be exhibited through this kind of curves.

Diagram

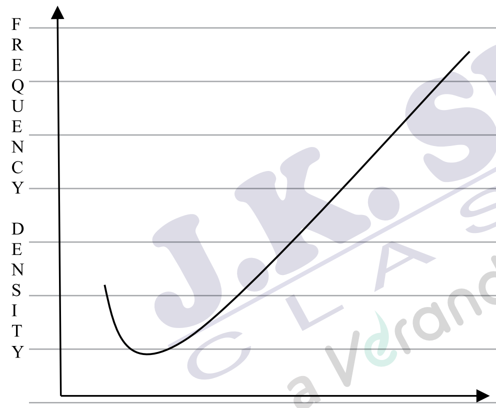


U- Shaped Curve

3. **J Shaped Curve:**

The J Shaped Curve starts with the minimum frequency and then gradually reaches its maximum frequency at the other extremity. The distribution of commuters in a particular time interval will be exhibited through this kind of curves.

Diagram



CLASS BOUNDARY

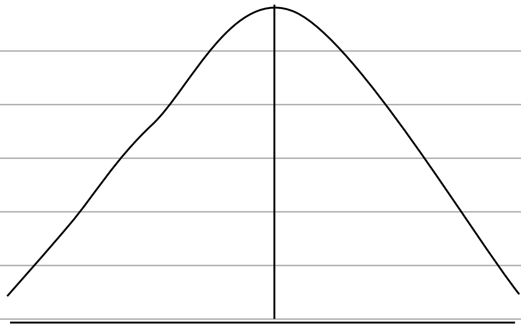
J- Shaped Curve

4. **Asymmetrical Curves**

(A) In case of symmetrical curves or bell shaped curves the

(i) Mean (M) = Median (Me) = Mode (Mo)

ii) Skewness = 0



Zero Skewness

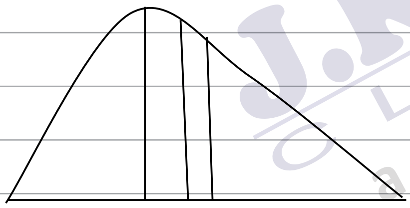
(B) In case of Asymmetrical curves Mean, Median & Mode are unequal and accordingly skewness ≠ 0

Asymmetrical Curves

Positively Skewed

(Mean > Median > Mode)

(i) Frequency curve as a longer tail to the right

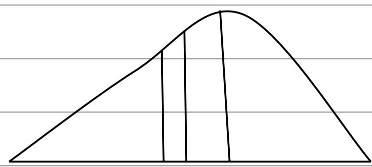


Positive Skewness

Negatively Skewed

(Mean < Median < Mode)

(i) Frequency curve as a longer tail to the Left



Negative Skewness

Related MCQ's:

16. The most common form of diagrammatic representation of a grouped frequency distribution is :

a) ogive
b) histogram
c) frequency polygon
d) none of the above

17. Frequency density is used in the construction of

a) histogram
b) frequency polygon
c) ogive
d) none of the above

18. When the width of all classes is same, frequency polygon has not the same area as the Histogram :

a) true b) false
c) both a) and b) above d) none of the above

19. The breadth of the rectangle is equal to the length of the class-interval in

a) ogive b) histogram
c) both a) and b) above d) none of these.

20. From which graphical representation, we can calculate partition values?

a) Lorenz Curve b) Ogive Curve
c) Histogram d) None of these

21. Arrange the dimensions of Bar Diagram, Cube Diagram, Pie Diagram in sequence.

a) 1, 3, 2 b) 2, 1, 3 c) 2, 3, 1 d) 3, 2, 1

FREQUENCY DISTRIBUTION

1. There are two types of frequency distribution

i. For discrete variable it is known as simple or ungrouped or discrete frequency distribution.

ii. For continuous variable it is known as continuous or grouped frequency distribution.

2. SOME IMPORTANT TERMS

i) Frequency : (Tally Mark)

Frequency of a value of variable is the number of times it occurs in a given series of observations. A Tally Mark (/) is put against the value when it occurs in the raw data. Having occurred four times, the fifth occurrence is represented by putting a Cross Tally Mark (\ ) on the first four tally marks.

ii) Range : Range of a given data is the difference between the largest measure and the smallest measure in a given set of observations.

iii) Class Interval (or class) : A large number of observations having wide range, is usually classified into number of groups. Each of these groups is known as a class.

iv) **Class frequency, Total Frequency :** The number of observations which is class contains, is known as its class frequency. The total number of observations in the frequency distribution is known as 'Total Frequency'.

v) **Class Limit :** The two ends of a class interval are known as class limits of that class. The smaller of the two ends is called LOWER Class Limits and the greater is called Upper Class Limit. These classification are called non-overlapping or mutually inclusive classification.

vi) **Class Boundaries :** When we consider a continuous variable, the observation are recorded nearest to a certain unit. For example, let us consider the distribution of weight of a group of persons. If we measure the weight nearest to the pound, then a class interval like (100-109) will include all the observations between 99.5 lb to 109.5 lb. Similarly, all the observations between 109.5 lb to 119.5 lb will be included in the class interval (110- 119). For the class interval (100-109), 99.5 is the lower class-boundary and 109.5 is the upper class boundary. For the class (110-119), the lower and upper class boundary respectively 109.5 and 119.5. These classifications are called overlapping or mutually exclusive classification.

Class boundaries can be calculated from the class limits by the following rule:

Lower Class boundary = Lower Class limit - $\frac{1}{2}$d;

Upper Class boundary = Upper Class limit + $\frac{1}{2}$d;

where, d is the common difference between the upper limit of a class and the lower limit of the next class. d/2 is called the Correction Factor

vii) **Mid-value (or class mark or mid point or class point) :**

Mid-value is the mid-Point of the class interval and is given by Class Mark= $\frac{UCL+LCL}{2}=\frac{UCB+LCB}{2}$

viii) **Width or Size :** This is the length of a class and is obtained by the difference between the upper and lower class boundaries of that class.

Class width / size = Difference between 2 successive LCL's / UCL's

= Difference between 2 successive LCB's / UCB's

= Difference between 2 successive mid values if all the class are of the same width.

= Difference between UCB and LCB

Note : Class width ≠ UCL-LCL

ix) **Frequency Density :** This is defined as the frequency per unit width of the class.

$$\text{Frequency Density} = \frac{\text{Class frequency}}{\text{Class width}}$$

It measures the concentration of the frequency of different classes.

x) **Relative Frequency :** This is the ratio of the class frequency to the total frequency,

i.e. $\text{Relative frequency} = \dfrac{\text{Class frequency}}{\text{Total Frequency}}$

- Relative Frequency of any class lies between 0 and 1

xi) **Percentage Frequency :**

$\dfrac{\text{Class frequency}}{\text{Total Frequency}} \times 100$ = or Relative frequency x 100

CUMULATIVE FREQUENCY DISTRIBUTION

1. There is another type of frequency distribution known as Cumulative Frequency Distribution where the frequencies are cumulated.
2. This distribution is prepared from the grouped frequency distribution by taking the end values (ie. class boundaries and not class limits)
3. Number of observation less than or equal to the class boundaries are called "Less-Than" Type Cumulative Frequency Distribution.
4. Number of observation greater than or equal to class boundaries are called " More-Than" Type Cumulative Frequency Distribution.
5. It can be made both for discrete series i.e. ungrouped data as well as for grouped data.

Example 2 :

From the following frequency distribution construct the cumulative frequency distribution:

Weights of 60 students in a class

Weights of 60 students in a class

Weight (kg)	Frequency
30-34	3
35-39	5
40-44	12
45-49	18
50-54	14
55-59	6
60-64	2
Total	60

Cumulative Frequency Distribution of weights of 60 students

Class Boundaries (Weight in kg)	Cumulative Frequency	
	Less Than	More Than
29.5	0	60
34.5	3	57
39.5	8	52
44.5	20	40
49.5	38	22
54.5	52	8
59.5	58	2
64.5	60	0

Otherwise

Cumulative Frequency Distribution of weights of 60 students

Class Boundaries (Weight in kg)	Cumulative Frequency	
	Less Than	More Than
30-34	3	60
35-39	8	57
40-44	20	52
45-49	38	40
50-54	52	22
55-59	58	8
60-64	60	2

Here the less than cumulative frequency of the second class is 8. This implies that there are 8 students whose weights are less than 39.5 kg (the upper boundary of that class). The more than cumulative frequency of the second class is 57, i.e. there are 57 students whose weights are more than 34.5 kg(the lower boundary of that class).

Note : **By Cumulative Frequency we usually mean less than type.**

Example 3 :

(a)

Marks	CF (Less than)	C.I	Frequency
Less than 20	5	10-20	5
Less than 30	18	20-30	13
Less than 40	30	30-40	12
Less than 50	35	40-50	5
		----	-----
			N= 35 = $\sum f$

(b)

Marks	C.I	CF (more than)	Frequency
More than 20	20-30	35	17
More than 30	30-40	18	8
More than 40	40-50	10	7
More than 50	50-60	3	3
		----	----
		CF	35

Related MCQ's:

22. For determining the class frequency it is necessary that these classes are:

a) Mutually exclusive b) Not mutually exclusive

c) Independent d) None of these

23. Mutually exclusive classification usually meant for

a) an attribute b) a continuous variable

c) a discrete variable d) any of the above

24. The lower class boundary is :
    a) an upper limit to Lower Class Limit
    b) a Lower limit to Lower Class Limit
    c) both a) and b) above
    d) none of the above

25. Relative frequency for a particular class
    a) lies between 0 and 1.
    b) lies between – 1 and 0.
    c) lies between 0 and 1, both inclusive.
    d) lies between – 1 to 1.

26. The lower extreme point of a class is called :
    a) lower class limit.
    b) lower class boundary
    c) both a) and b) above
    d) none of the above

27. Frequency Density corresponding to a class interval is the ratio of:
    a) Class Frequency to the Total Frequency
    b) Class Frequency to the Class Length
    c) Class Length to the Class Frequency
    d) Class Frequency to the Cumulative Frequency

Theory Answers

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

Numerical Problems

In 1995, out of the 2,000 students in a college; 1,400 were for graduation and the rest of Post-Graduation (PG). Out of 1,400 Graduate students 100 were girls, in all there were 600 girls in the college. In 2000, number of graduate students increased to 1,700 out of which 250 were girls, but the number of PG students fall to 500 of which only 50 were boys. In 2005, out of 800 girls 650 were for graduation, whereas the total number of graduates was 2,200. The number of boys and girls in PG classes were equal.

28. When the class intervals are 10 – 19, 20 – 29, 30 – 39, ... Upper class boundaries (UCB) and the Upper class limits (UCL) of the 2^{nd} class interval are:

a) 29, 29 b) 20, 29 c) 29.5, 29.5 d) 29.5, 29

1B SAMPLING THEORY

1. Population or Universe

Population in statistics means the whole of the information which comes under the purview of statistical investigation. It is the totality of all the observations of a statistical experiment or enquiry.

A population may be finite or infinite according as the number of observations or items in it are finite or infinite. The population of weights of students of class XII in a government school is an example of a finite population. The population of pressure at different points in the atmosphere is an example of an infinite population.

Types of Population:

a) Finite Population: When the items in the population are fixed and limited.
   Example : No. of students in the class

b) Infinite Population: If a population consist of infinite no. of items its an infinite population. If a sample is known to have been drawn from a continuous probability distribution, then the population is infinite. Example : Population of all real numbers lying between 5 and 20.

c) Real Population: A Population consisting of the items which are all present physically is termed as real population.

d) Hypothetical Population: The Population consists of the results of the repeated trails is named as hypothetical population The tossing of a coin repeatedly results into a hypothetical population of heads and tails.

2. Sample

A part of the population selected for study is called a sample. In other words, the selection of a group of individuals or items from a population in such a way that this group represents the population, is called a sample.

1. Sampling is a process whereby we judge the characteristics or draw inference about the totality or Universe (known as population) on the basis of judging the characteristics of a selected portion taken from that totality (known as sample).

2. Sample: Sample is the part of population selected on some basis it is a finite subset of the population.
3. Sample Units : Units forming the samples are called Sample Units.
4. Sample Frame : A complete list of sampling units is called Sample Frame
5. Sample Faction : $\frac{n}{N}$ is called Sampling Fraction where n = Sample Size and N = Population Size.
6. Complete enumeration or census : In case of enumeration, information is collected for each and every unit. The aggregate of all the units under consideration is called the 'population' or the 'universe'. The results are more accurate and reliable but it involves lot of time, money and man power

3. Parameter and Statistic

There are various statistical measures in statistics such as mean, median, mode, standard deviation, coefficient of variation etc. These statistical measures can be computed both from population (or universe) data and sample data.

Parameter : Any statistical measure computed from population data is known as parameter.

Statistics : Any statistical measure computed from sample data is known as statistic. Thus a parameter is a statistical measure which relates to the population and is based on population data, whereas a statistic is a statistical measure which relates to the sample and is based on sample data. Thus a population mean, population median, population variance, population coefficient of variation etc., are all parameters. Statistic computed from a Sample such as sample mean, sample variance etc.

Notations

Statistical Measure	Population	Sample
Mean	μ	x
Standard deviation	σ	s
Proportion	P	p
Size	N	n

Related MCQ's:

1. The aggregate or totality of statistical data forming a subject of investigation is known as :

a) Sample b) Population
c) Both a) and b) above d) None of the above

2. If a sample is known to have been drawn from a continuous probability distribution then the population is .
   a) Large b) Finite
   c) Infinite d) Nothing can be said about the population

3. The possibility of reaching valid conclusions concerning a population by means of a population by means of a properly chosen sample is based on which of the following laws?
   a) Law of Inertia b) Law of Large Number
   c) Law of Statistical Regularity d) All of the above

4. When the population is infinite we should use the:
   a) Sample Method b) Census Method
   c) Either Sample or Census Method d) None of the above

5. A border patrol checkpoint which stops every passenger van is utilizing:
   a) simple random sampling. b) systematic sampling
   c) systematic sampling. d) complete enumeration

6. A population consisting of all real numbers is an example of:
   a) an infinite population b) a finite population
   c) an imaginary d) none of the above

4. Basic principle of Sample Survey

a) **Law of Statistical Regularity :** It states that a reasonably larger number of items selected at random from a large group of items, will on the average, represent the characteristics of the group.

b) **Law of Inertia of Large Numbers :** This law states that other things same, as the sample size increases, the results tend to be more reliable and accurate.

c) **Principle of Optimization :** The principle of optimization ensures that an optimum level of efficiency at a minimum cost or the maximum efficiency at the given level of cost can be achieved with the selection of an appropriate sampling design.

d) **Principle of Validity :** The principle of validity states that a sampling design is valid only if it is possible to obtain valid estimates and valid tests about population parameters. Only a probability sampling ensures this validity.

Related MCQ's:

7. Law of Statistical Regularity states that:
   a) A sample of reasonably small size when selected at random, is almost not sure to represent the characteristics of the population
   b) A sample of reasonably large size when selected, is almost not sure to represent the characteristics of the population.
   c) A sample of reasonably large size when selected at random, is almost sure to represent the characteristics of the population, on an average
   d) None of the above

8. Law of Inertia states that:
   a) Sample of high size show a high degree of stability.
   b) Sample of low size shows a high degree of stability.
   c) Results obtained from sample of high size are expected to be very far.
   d) None of the above.

9. Sampling error increases with an increase in the size of the sample.
   a) The above statement is true.
   b) The above statement is not true.
   c) Sampling error do not depends upon the sample size
   d) None of the above

5. Sampling and Non sampling Errors

i) **Sampling Errors:** Sampling Errors have their origin in sampling and arise due to the fact that only a part of the population (i.e. sample) has been used to estimate population parameters and draw inference about them. As such the sampling errors are totally absent in a census enumeration.

Sampling errors can never be completely eliminated but can be minimize by choosing a proper sample of adequate size.

ii) **Non Sampling Errors or Bias:** As distinct from sampling errors, the non-sampling errors primarily arise at the stages of observation, approximation and processing of the data and are thus **present in both the complete enumeration and the sample survey.** These error usually arise due to faulty planning, defective schedule of questionnaire from non-response from the respondents.

iii) Sampling error is totally absent in "Complete Enumeration" or "Census"
But, Non-Sampling errors are present in both "Complete Enumeration" and "Sample survey"

- Parameter is a statistical measure on population. Statistic is a statistical measure on sample.

Related MCQ's:

10. Bias is also known as:

a) Sampling Error b) Non-Sampling Error
c) Error d) None of the above

11. Sampling error are:

a) Particularly detectfull
b) Can be corrected
c) Arise because the information collected relates only to a part of the population.
d) All of the above.

12. _Can occur in census.

a) Standard Error b) Sampling Error
c) Bias d) None of the above

13. "Sampling errors are present both in census as well as a sample survey." -State whether the given statement is correct or not.

a) Correct b) Incorrect
c) Nothing cannot be said d) None of the above

6. Sampling Distribution of a Statistic

From a population of size N, number of samples of size n can be drawn. These samples will give different values of a statistic. E.g. if different samples of size n are drawn from a population, different values of sample mean are obtained. The various values of a statistic thus obtained, can be arranged in the form of a frequency distribution known as Sampling Distribution. Thus we can have sampling distribution of sample mean $\bar{x}$, sampling distribution of sample proportion p etc.

Errors in Sampling

Any statistical measure say, mean of the sample, may not be equal to the corresponding statistical measure (mean) of the population from which the sample has been drawn. Thus there can be discrepancies in the statistical measure of population, i.e., parameter and the statistical measures of sample drawn from the same population i.e., statistic. These discrepancies are known as Errors in Sampling.

Standard Error of a Statistic

Standard error is used to measure the variability of the values of a statistic computed from the samples of the same size drawn from the population, whereas standard deviation is used to measure the variability of the observations of the population itself.

The standard deviation of the sample statistics is called standard error of that statistic. E.g. if different samples of the same size n are drawn from a population, we get different values of sample mean $\bar{x}$. The S.D. of $\bar{x}$. is called standard error of $\bar{x}$ . . It is obvious that the standard error of $\bar{x}$. will depend upon the size of the sample and the variability of the population.

i) Standard error of sample mean SE ($\bar{x}$) = $\frac{\sigma}{\sqrt{n}}$ or $\frac{s}{\sqrt{n}}$

σ=Population S.D

and s=Sample S.D

ii) Standard error of proportion SE (p) = $\sqrt{\frac{P\ (1\ -\ P)}{n}}$ or $\sqrt{\frac{p(1-p)}{n}}$

Where P=Population proportion P=Sample proportion

If i) Population size is Finite and the **Sampling Fraction** $\frac{n}{N}$ ≥.05

And ii) **Samples are drawn Without Replacement(SRSWOR)**

Then , each of the above formula for Standard Error will be multiplied by the factor

$\sqrt{\frac{N-n}{N-1}}$ (Finite Population correction or Finite Population Multiplier)FPC

- Formula for standard Error when i) n<30(small sample)

  ii) Population S.D σ is unknown in such a case SE ($\bar{x}$)= $\frac{s}{\sqrt{n-1}}$

The following table will provide us a better understanding of the situations while calculating SE ($\bar{x}$)

Sample Size	Parameter	Formula
Large (n ≥ 30)	SD is known	$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Large (n ≥ 30)	SD is unknown	$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$
Small (n < 30)	SD is known	$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Small (n < 30)	SD is unknown	$SE_{\bar{x}} = \frac{s}{\sqrt{n-1}}$
Rule of multiplying FPC will remain unaltered in a cases		

Summary

Concept of Sampling Distribution of Statistic and Standard Error:

- ⇨ Samples can be drawn with or without replacement
- ⇨ Probability distribution of a statistic is called sampling of statistic. Example: sampling distribution of ($\bar{x}$)., sampling distribution of (p)
- ⇨ Standard deviation of the sampling distribution of the sampling is called Standard Error of statistic
- ⇨ As sample size increases standard error decreases proportionately.
- ⇨ Precision of the sample is reciprocal to standard Errors..
- ⇨ Standard Error measures sampling fluctuations. i.e fluctuations in the value of statistics due to sampling

Related MCQ's:

14. Values of a particular statistic with their relative frequencies will constitute the of the concerned statistic.

    a) Probability Distribution
    b) Sampling Distribution
    c) Theoretical Distribution
    d) None of these

15. The population standard deviation describes the variation among elements of the universe, whereas, the standard error measures the:
    a) variability in a statistic due to universe
    b) variabillity in a statistic due to sampling
    c) variablity in a parameter due to universe
    d) variablity in a statistic due to parameter

16. Standard error can be described as:
    a) The error committed in sample survey
    b) The error committed in estimating a parameter
    c) Standard deviation of a statistic
    d) The error committed in sampling.

17. The reciprocal of the standard error is:
    a) Precision of the sample
    b) Error of the sample
    c) Error of the Universe
    d) None of the above

18. Precision of random sample:
    a) increases directly with increase in sample size
    b) increases with the increase in sample size
    c) increases proportionately with sample size
    d) none of these.

19. Sampling Fluctuations may be described as :
    a) the variation in the values of a statistic.
    b) the variation in the values of a sample.
    c) the differences in the values of a parameter.
    d) the variation in the values of observations.

7. Types of Sampling

A sample can be selected from a population in various ways. Different situations call for different methods of sampling. There are three methods of Sampling:

1. Random Sampling or Probability Sampling Method.
2. Non-Random Sampling or Non-Probability Sampling Method.
3. Mixed Sampling.

1. Random Sampling or Probability Sampling

Random Sampling: Random or Probability sampling is the scientific technique of drawing samples from (he population according to some laws of chance in which each unit in the universe or population has some definite pre-assigned probability of being selected in the sample. It is of two types.

(a) Simple Random Sampling (SRS):

It is the method of selection of a sample in such a way that each and every member of population or universe has an equal chance or probability of being included in the sample. Random sampling can be carried out in two ways.

1. **Lottery Method:** It is the simplest, most common and important method of obtaining a random sample. Under this method, all the members of the population or universe are serially numbered on small slips of a paper. They are put in a drum and thoroughly mixed by vibrating the drum. After mixing, the numbered slips are drawn out of the drum one by one according to the size of the sample. The numbers of slips so drawn constitute a random sample.
2. **Random Number Method:** In this method, sampling is conducted on the basis of random numbers which are available from the random number tables. The various random number tables available are:
   a. Trippet's Random Number Series;
   b. Fisher's and Yales Random Number Series;
   c. Kendall and Badington Random Number Series;
   d. Rand Corporation Random Number Series;

One major disadvantage of random sampling is that all the members of the population must be known and be serially numbered. It will entail a lot of difficulties in case the population is of large size and will be impossible in case the population is of infinite size.

(b) Restricted Random Sampling:

It is of three types

- Stratified Sampling
- Systematic Sampling
- Multi-stage Sampling

Stratified Sampling: In stratified random sampling, the population is divided into strata (groups) before the sample is drawn. Strata are so designed that they do not overlap. An elementary unit from each stratum is drawn at random and the units so drawn constitute a sample. Stratified sampling is suitable in those cases where the population is hetrogeneous but there is homogeneity within each of the groups or strata.

Advantages

(i) It is a representative sample of the hetrogeneous population.

(ii) It lessens the possibility of bias of one sidedness.

Disadvantages

(i) It may be difficult to divide population into homogeneous groups.

(ii) There may be over lapping of different strata of the population which will provide an unrepresentative Sample.

Systematic Sampling: In this method every elementary unit of the population is arranged in order and the sample units are distributed at equal and regular intervals. In other words, a sample of suitable size is obtained (from the orderly arranged population) by taking every unit say tenth unit of the population. One of the first units in this ordered arrangement is chosen at random and the sample is computed by selecting every tenth unit (say) from the rest of the lot. If the first unit selected is 4, then the other units constituting the sample will be 14, 24, 34, 44, and so on.

Advantages: It is most suitable where the population units are serially numbered or serially arranged.

Disadvantages: It may not provide a desirable result due to large variation in the items selected.

Multi-stage Sampling: In this sampling method, sample of elementary units is selected in stages. Firstly a sample of cluster is selected and from among them a sample of elementary units is selected. It is suitable in those cases where population size is very big and it contains a large number of units.

2. Non-Random Sampling or Non-Probability Sampling Method

A sample of elementary units that is being selected on the basis of personal judgment is called a non-probability sampling. It is of four types.

- Purposive Sampling;
- Quota Sampling;
- Convenience Sampling;
- Sequential Sampling.

Purposive Sampling: Purposive sampling is the method of sampling by which a sample is drawn from a population based entirely on the personal judgement of the investigator. It is also known as Judgement Sampling or Deliberate Sampling. A randomness finds no place in it and so the sample drawn under this method cannot be subjected to mathematical concepts used in computing sampling error.

Quota Sampling: In quota sampling method, quotas are fixed according to the basic parameters of the population determined earlier and each field investigator is assigned with quotas of number of elementary units to be interviewed.

Convenience Sampling: In convenience sampling, a sample is obtained by selecting convenient population elements from the population.

Sequential Sampling: In sequential sampling a number of sample lots are drawn one after another from the population depending on the results of the earlier samples draw from the same population. Sequential sampling is very useful in Statistical Quality Control. If the first sample is acceptable, then no further sample is drawn. On the other hand if the initial lot is completely unacceptable, it is rejected straightway. But if the initial lot is of doubtful and marginal character falling in the area lying between the acceptance and rejection limits, a second sample is drawn and if need be a third sample of bigger size may be drawn in order to arrive at a decision on the final acceptance or rejection of the lot. Such sampling can be based on any of the random or non-random method of selection.

Advantages of Random (OR Probability) Sampling

1. Random sampling is objective and unbiased. As a 'result, it is defensible before the superiors or even before the court of law. 8
2. The size of sample depends on demonstrable statistical method and therefore, it has a justification for the expenditure involved.
3. Statistical measures, i.e. parameters based on the population can be estimated and evaluated by sample statistic in terms of certain degree of precision required.
4. It provides a more accurate method of drawing conclusions about the characteristics of the population as parameters.
5. It is used to draw the statistical inferences.
6. The samples may be combined and evaluated, even though accomplished by different individuals.
7. The results obtained can be assessed in terms of probability, and the sample is accepted or rejected on a consideration of the extent to which it can be considered representative.

3. Mixed Sampling

Cluster Sampling: Cluster Sampling involves arranging elementary items in a population into hetrogeneous subgroups that are representative of the overall population. One such group constitutes a sample for study.

Related MCQ's:

20. Simple random sampling is
    (a) A probabilistic sampling
    (b) A non- probabilistic sampling
    (c) A mixed sampling
    (d) Both (b) and (c).

21. Which sampling provides separate estimates for population means for different segments and also an over all estimate?
    (a) Multistage sampling
    (b) Stratified sampling
    (c) Simple random sampling
    (d) Systematic sampling

8. SAMPLING WITH REPLACEMENT (SRSWR)

While selecting the units for a sample, when a unit of sample selected is replaced before the next unit is selected then it is called sampling with replacement.

In this case the total number of samples that can be drawn = $(N)^n$

For E.g.: Let Population = {a, b, c}

N = 3, let n = 2

No. of samples = (N)n = $(3)^2$ = 9

No. of samples = {(a, b) (a, c) (b, c) (b, a) (c, a) (c, b) (a, a) (b, b) (c, c)}

9. SAMPLING WITHOUT REPLACEMENT (SRSWOR)

While selecting the units for a sample, when a unit of sample is selected but not replaced before the next unit is selected then it is called Sampling Without Replacement.

In this case the total number of samples that can be drawn =

For E.g.: Let population = {a, b, c}

N = 3, let n = 2

No. of samples = ${}^NC_n = {}^3C_2 = {}^3C_1 = 3$

No. of samples = {(a, b), (a, c), (b, c)}

Related MCQ's:

22. In simple random sampling with replacement, the total number of possible sample with distinct permutation of member is:
(N = Size of Population, n = Sample size)
a) $N \times n$ b) N^n c) N d) n

23. In simple random sampling without replacement, the total number of possible sample with distinct permutation of member is:
(N = Size of Population, n = Sample size)
a) N^n b) P(N, n) c) C(N,n) d) None of the above

Theory Answers

1	b	7	c	13	b	19	a
2	c	8	a	14	b	20	a
3	d	9	b	15	b	21	b
4	a	10	b	16	c	22	b
5	d	11	d	17	a	23	c
6	a	12	c	18	c		

Note : Students shall workout in the class for prof

2A MEASURES OF CENTRAL TENDENCY (Averages of First Order)

INTRODUCTION:

- Central tendency is defined as the tendency of the data to concentrate towards the central or middle most region of the distribution.

- In other words, Central Tendency indicates average.

- Any average is a representative value of the entire distribution value

- Average discovers uniformity in variability.

- The tendency of the variables to accumulate at the center of the distribution (data) is known as measures of central tendency.

- Measures are popularly also known as averages.

Average
- Mathematical Avg.
  - A. M
  - G. M
  - H. M
- Positional Avg.
  - Median
  - Mode

The criteria for Ideal Measures of Central Tendency

1. It should be simple to understand. (Mean, Median & Mode are easy to compute)

2. It should be based on all the observations. (AM,GM,HM are based on all the observations)

3. It should be rigidly defined (except Mode).

4. It should not be affected by extreme values (Median & Mode are not affected by extreme values.

5. It should have sampling stability or it should not be affected by sampling fluctuations. (A.M, G.M, H.M. not affected).

6. It should be capable of further algebraic treatment. (AM,GM,HM)

ARITHMETIC MEAN

- It is the best measure of central tendency and most commonly used measure
- The only drawback of this measure is that it gets highly affected by presence of extreme values in the distribution.
- Calculation of AM

1. For Simple series: A.M. $= \bar{x} = \dfrac{\sum x}{n}$

2. For simple frequency distribution :

Let $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ be a series, occuring with frequency $f_1, f_2, f_3, \ldots\ldots, f_n$ respectively, then

$$\text{A.M.} = \bar{x} = \frac{\sum fx}{N} = \frac{\sum fx}{\sum f} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n};\ \text{N} = \text{Total Frequency}$$

3. For Grouped Frequency Distribution:

a) Direct Method

$$\text{A.M.} = \bar{x} = \frac{\sum fx}{N} = \frac{\sum fx}{\sum f} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n}$$

Where, x = mid - values or class marks

b) Method of Assumed Mean using Step Deviation (By changing of origin and scale)

$$AM = \bar{x} = A + \left(\frac{\sum fd}{\sum f}\right) \cdot i \qquad \bullet\ d = \frac{x - a}{i}$$

Where,

X = mid-values or original values if it is a discreet series

a = Assumed Mean i.e., a value arbitrarily chosen from mid-values or any other values

I = class width or any arbitrary value

PROPERTIES

1. If all values of the variable are constant, then AM is constant.

2. $\bar{x} = \frac{\sum x}{n}$; Thus, Sum of the observations = (no. of observations) x (average).

3. Sum of deviations of values from their arithmetic mean is always zero.

4. When the values of x are equi-distant, then AM = $\frac{\text{First value} + \text{Last value}}{2}$

5. If the frequencies of variable increases or decreases by the same proportion, the value of AM will remain unaltered.

6. Weighted AM of first "n" natural numbers, when the values are equal to their corresponding weights, will be given by $\bar{x} = \frac{2n+1}{3}$

7. Sum of squares of deviation is minimum when the deviation is taken from AM.

8. AM is dependent on the change of origin and scale.

   If Y = a ± bx,

   then, $\bar{Y} = a \pm b\bar{x}$

9. Formula for calculating Combined Mean is given by: $\bar{x}_c = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$

   Where,

   $\bar{x}_1$ = mean of the first group

   $\bar{x}_2$ = mean of the second group

   n_1 = number of samples in the first group

   n_2 = number of samples in the second group

GEOMETRIC MEAN (GM)

1. Let $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ be a simple series, then G.M. = $\sqrt[n]{x_1 . x_2 . x_3 \ldots\ldots. x_n}$ (n^{th} root of the product)

2. Let $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ be a series, occuring with frequency $f_1, f_2, f_3, \ldots.., f_n$ respectively, then

   G.M. = $\sqrt[N]{x_1^{f_1} . x_2^{f_2} . x_3^{f_3} \ldots\ldots x_n^{f_n}}$

3. $(G.M)^n$ = Product of the observation

4. It is capable of further algebraic treatment.

5. It is less affected by sampling fluctuations compare to mode and median.

6. It is less affected by extreme values compare to AM.

7. GM cannot be calculated if any variable assumes value 0 or negative value.

8. GM is particularly useful in cases where we have to find out average rates or ratios of quantities which are changing at a cumulative rate, i.e., the change is related to the immediate preceding data. For example, average rate of depreciation by WDV method or average rate of growth of population.

9. GM is extensively used in the construction of index numbers.

10. GM is the most difficult average to calculate and understand because it involves the knowledge of logarithms.

11. Logarithm of GM of "n" observations is equal to the AM of the logarithm of these "n" observations.

12. GM is based on all observations

13. If all the observations assumed by a variable constant, say K, then the GM of the observations is also K

14. GM of the product of two variables is the product of their GM's i.e.,

if z = xy,

then GM of z = (GM of x) . (GM of y)

15. GM of the ratio of two variables is the ratio of GM's of two variables i.e.,

if z = x/y

then GM of $z = \dfrac{\text{GM of x}}{\text{GM of y}}$

16. Combined GM: $G_{12} = [G_1^{n_1} \; G_2^{n_2}]^{\frac{1}{n_1+n_2}}$ $\therefore \log G_{12} = \dfrac{n_1 \log G_1 + n_2 \log G_2}{n_1+n_2}$

HARMONIC MEAN (HM)

1. Let $x_1, x_2, x_3, ..., x_n$ be a simple series, then H.M. $= \dfrac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+.......+\frac{1}{x_n}}$

2. Let $x_1, x_2, x_3, ..., x_n$ be a series, occuring with frequency $f_1, f_2, f_3, ..., f_n$ respectively, then

$$\text{H.M.} = \frac{N}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\frac{f_3}{x_3}+.......+\frac{f_n}{x_n}}$$

3. HM cannot be calculated if any variable assumes value 0, as inverse of 0 is undefined.

4. HM has a very restricted use, and they are usually used for calculating average speed, average rates of quantities, etc.

5. It is based on all the values.

6. It is capable of further algebraic treatment.

7. It is less affected by extreme values and sampling fluctuations compare to AM and GM.

8. If y = ax then

HM(y) = a HM (x) | GM(y) = a GM (x)

9. If all the observations are constant, HM is constant

10. Combined H.M: $H_{12} = \dfrac{n_1+n_2}{\frac{n_1}{H_1}+\frac{n_2}{H_2}}$

RULE FOR USING AM AND HM

When the average to be calculated is of the form a/b, where a and b are different quantities then

i. Use HM when 'a' is constant

ii. Use AM when 'b' is constant

For eg,

Avg. speed = ?	Distance = same (given)	we know that Speed $= \frac{\text{Distance}}{\text{Time}}$
Use H. M		
Avg. speed = ?	Time = same (given)	
Use A. M		

RELATION BETWEEN AM, GM & HM

1. If the values are equal,

   AM = GM = HM.

2. If the values are distinct,

   AM > GM > HM.

3. $G^2 = A.M \times H.M.$

   $G = \sqrt{A.M. \times H.M.}$

MEDIAN:

1. Median is defined as the positional average and is regarded as the second best average after arithmetic mean.

2. Median is suitable when there is a wide range of variation in data or distribution pattern is to be studied at a varying level.

3. Median is suitable for qualitative data.

4. Median is suitable for distributions with open ends.

5. Median can be located graphically using Cumulative Frequency Polygon or Ogives.

6. The absolute sum of deviations is minimum when the deviations are taken from Median, and this property of Median is known as "Minimal Property".

7. Median is dependent on change of Origin & Scale.
If $Y = a \pm bx$
Then, $Me(Y) = a \pm bMe(x)$

Calculation

For Simple Series

Median = value corresponding to (n + 1)/2th term in the distribution

Note 1: Arrange the data in the ascending or descending order

Note 2: If the value of (n+1)/2th term is a fraction then the average of the values within which it is lying is the median.

Note 3: If n is odd median = simply the middle most value and if n is even median = average of 2 mid values

For Simple Frequency Distribution:

Median = value corresponding to the (N+1)/2th Term in the 'less than' type Cumulative Frequency column where,

N = Total Frequency

For Grouped Frequency Distribution:

$$\text{Median} = l_1 + \left(\frac{\frac{N}{2} - F}{f_m} \right) . i$$

l_1 = Lower boundary of the median class i.e., the class where Cumulative Frequency N/2 falls

N = Total frequency

F = Cumulative frequency of the pre-median class.

f_m = Frequency of the median class

i = Width of the median class

MODE

1. Mode is that value of the distribution which occurs with highest frequency.

2. Mode is a crude method of finding out average and it provides only a Bird's Eye view of the distribution.

3. It is the most unstable average and the quickest method of finding out the average where we need to find out the most common value of the distribution

4. It is not affected by extreme values but it is more affected by sampling fluctuations compare to AM, GM, HM.

5. In case when distribution is Multimodal, mode is ill-defined

6. Mode is dependent on the change of origin and scale

7. If $y = a \pm bx$ then, $Mo(y) = a \pm b\, Mo(x)$

8. Mode can be located graphically using Histogram or Area Diagram or Frequency Diagram.

9. Mode does not take into account all of the observations.

10. When the classes are of unequal width, we consider frequency densities instead of class frequency to locate mode,

    where frequency density = $\dfrac{\text{Class Frequency}}{\text{Width of the Class}}$

Calculation of Mode for Simple Series:

1. For simple series, there is no mode as all values occur with frequency = 1, i.e., same frequency.

2. For simple frequency distribution Mode can be calculated by mere inspection. The variable occurring with the highest frequency is the mode of the distribution. A distribution can be uni-modal or bi-modal, but not multi-modal.

- o If only one value of variable occurs with the highest frequency, then there is only one mode.
- o If two values of variable occurs with the same highest frequency, then there are two modes.
- o If all values of variable occurs with same frequency, then there is no mode.
- o If more than two values of variable occurs with same highest frequency, then also there is no mode.

Calculation of Mode for Grouped Frequency Distribution:

$$\text{Mode} = l_1 + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right).i$$

L_1	=	Lower boundary of the modal class i.e., the class with highest frequency.
f_m	=	Frequency of the modal class
f_1	=	Frequency of the pre-modal class
f_2	=	Frequency of the post-modal class
i	=	Class width

CONCEPT OF SYMMETRICAL & ASYMMETRICAL DISTRIBUTION:

1. When in a distribution all the measures of central tendencies are equal, the distribution is said to be symmetrical.

2. For symmetrical distribution; Mean = Median = Mode.

3. Any deviation from this symmetry makes the distribution asymmetrical or skewed.

4. For moderately skewed distribution: Mean – Mode = 3(Mean – Median)

OTHER PARTITION VALUES (FRACTILES)

Partition values divides distribution in equal parts.

- **QUARTILES**
  - o There are 3 quartiles (Q_1, Q_2, Q_3), which divides the distribution in 4 equal parts representing 25%, 50% and 75% of the data respectively.

- o Q_2 is nothing but the median of the data.

- o For symmetrical data, Q_2 is simple average of the extreme quartiles Q_1 (lower quartile) and Q_3 (upper quartile).

- **DECILES**
  - o There are 9 deciles (D_1, D_2, ..., D_9), which divides the distribution in 10 equal parts representing 10%, 20% .... 90% of the data respectively.
  - o D_5 is nothing but the median of the data.

- **PERCENTILES**
  - o There are 99 percentiles (P_1, P_2, ..., P_{99}), which divides the distribution in 100 equal parts representing 1%, 2% ..... 99% of the data respectively.
  - o P_{50} is nothing but the median of the data

- **NOTE**
  - o All partition values are dependent on the change of Origin and Scale.
  - o All partition values can be calculated graphically through Cumulative Frequency Polygon or ogives.

Calculation of Partition Values

Type of Series	Quartiles	Deciles	Percentiles
Simple Series	$Q_i = i\left(\frac{n+1}{4}\right)$ $i = 1,2,3$	$D_i = i\left(\frac{n+1}{10}\right)$ $i = 1,2,3,.....,9$	$P_i = i\left(\frac{n+1}{100}\right)$ $i = 1,2,3,.....,99$
Simple Frequency Dist	Q_i = *value* correspo--nding to CF; $i\left(\frac{N+1}{4}\right)$	D_i = *value* correspo--nding to CF; $i\left(\frac{N+1}{10}\right)$	P_i = *value* correspo--nding to CF; $i\left(\frac{N+1}{100}\right)$
Group Frequency Dist	$Q_i = l_1 + \left(\frac{\frac{iN}{4} - f}{f_q}\right) i$	$D_i = l_1 + \left(\frac{\frac{iN}{10} - f}{f_d}\right) i$	$P_i = l_1 + \left(\frac{\frac{iN}{100} - f}{f_p}\right) i$

CLASSWORK SECTION

AIRHTEMATIC MEAN

1. The arithmetic mean of 8, 1, 6 with weights 3, 2, 5 respectively is:
   a) 5 b) 5.6 c) 6 d) 4.6

2. The average weight of students in a class of 35 students is 40 kg. If the weight of the teacher be included, the average rises by (1/2) kg; the weight of the teacher is :
   a) 40.5 kg b) 50 kg c) 41 kg d) 58 kg

GEOMETRIC MEAN

3. The interest paid on the same sum yielding 3%, 4%, and 5% compound interest for 3 consecutive year respectively. What is the average yield percent on the total sum invested.
   a) 3.83% b) 4.83% c) 2.83% d) 3.99%

HARMONIC MEAN

4. What is the HM of 1,1/2, 1/3,.... ....1/n?
   a) n b) 2n c) $\frac{2}{(n+1)}$ d) $\frac{n(n+1)}{2}$

MEDIAN

5. Calculate median for the following data :

No. of students	6	4	16	7	8	2
Marks	20	9	25	50	40	80

   a) 20 b) 25 c) 35 d) 28

PARTITION VALUE

6. The third decile for the numbers 15, 10, 20, 25, 18, 11, 9, 12 is
   a) 13 b) 10.70 c) 11 d) 11.50

COMBINED PROPERTIES OF AM, MEDIAN AND MODE

7. If the Mean and Mode of a certain set of numbers be 60.4 and 50.2 respectively, find approximately the value of the Median.

a) 55 b) 56 c) 57 d) 58

MISCELLANEOUS SUM

8. The mean and mode for the following frequency distribution

Class interval :	350-369	370-389	390-409	410-429	430-449	450-469
Frequency:	15	27	31	19	13	6

are

a) 400 and 390 b) 400.58 and 390

c) 400.58 and 394.50 d) 400 and 394.

9. For the following incomplete distribution of marks of 100 pupils, median mark is known to be 32.

Marks:	0–10	10–20	20–30	30–40	40–50	50–60
No. of Students:	10	-	25	30	-	10

What is the mean mark?

a) 32 b) 31 c) 31.30 d) 31.50

THEORETICAL ASPECTS

10. Measures of central tendency for a given set of observations measures

a) The scatterness of the observations

b) The central location of the observations

c) Both (a) and (b)

d) None of these.

11. While computing the AM from a grouped frequency distribution, we assume that

a) The classes are of equal length

b) The classes have equal frequency

c) All the values of a class are equal to the mid-value of that class

d) None of these.

12. Which of the following statements is wrong?
    a) Mean is rigidly defined
    b) Mean is not affected due to extreme values.
    c) Mean has some mathematical properties
    d) All these

13. For open-end classification, which of the following is the best measure of central tendency?
    a) AM b) GM c) Median d) Mode

14. The presence of extreme observations does not affect
    a) AM b) Median c) Mode d) (b) and (c) both

15. In case of an even number of observations which of the following is median?
    a) Any of the two middle-most value
    b) The simple average of these two middle values
    c) The weighted average of these two middle values
    d) Any of these

16. Which one of the following is not uniquely defined?
    a) Mean b) Median c) Mode d) All of these measures

17. Weighted averages are considered when
    a) The data are not classified
    b) The data are put in the form of grouped frequency distribution
    c) All the observations are not of equal importance
    d) Both (a) and (c).

18. Which of the following results hold for a set of distinct positive observations?
    a) AM ≥ GM ≥ HM b) HM ≥ GM ≥ AM
    c) AM > GM > HM d) GM > AM > HM

19. Which of the following measure(s) possesses (possess) mathematical properties?
    a) AM b) GM c) HM d) All of these

20. Which of the following measure(s) satisfies (satisfy) a linear relationship between two variables?

a) Mean b) Median c) Mode d) All of these

21. The sum of the squares of deviations of a set of observations has the smallest value, when the deviations are taken from their

a) A.M b) H.M c) G.M d) none

22. For 899, 999, 391, 384, 590, 480, 485, 760, 111, 240

Rank of median is

a) 2.75 b) 5.5 c) 8.25 d) none

Theory Answers

ANSWERS - SUMS				ANSWERS - THEORITICAL ASPECTS			
Q. No.	Ans	Q. No.	Ans	Q. No.	Ans	Q. No.	Ans
1	b	7	c	13	c	19	d
2	d	8	c	14	d	20	d
3	d	9	c	15	b	21	a
4	c	10	b	16	c	22	b
5	b	11	c	17	c		
6	b	12	c	18	c		

2B MEASURES OF DISPERSION (Average of Second Order)

THEORY

Introduction:

- Dispersion is defined as deviation or scattering of values from their central values i.e, average (Mean, Median or Mode but preferably Mean or Median)
- Dispersion discovers variability in uniformity.
- In other words, dispersion measures the degree or extent to which the values of a variable deviate from its average
- Dispersion indicates the degree of heterogeneity among observation and as heterogeneity increases dispersion increases
- If all values are equal then any measure of dispersion is always zero
- All measures of dispersion are positive
- All measures of dispersions are independent of the change of origin but dependent on the change of scale
- All pre requisites of a good measure of central tendency are equally applicable for good measure of dispersion
- TWO DISTRIBUTIONS MAY HAVE;
  - i. Same central tendency and same dispersion
  - ii. Different central tendency but same dispersion
  - iii. Same central tendency but different dispersion
  - iv. Different central tendency and different dispersion

Types of Measures of Dispersion

There are two types of measures of dispersion,

Absolute Measure	Relative Measure
a. These measures of dispersion will have the same units as those of the variables	a. These are usually expressed as ratios or percentages and hence unit free
b. Absolute measures are related to the distribution itself.	b. Relative measures are used i) to compare variability between two or more series. ii) To check the relative accuracy of the data

MEASURES OF DISPERSION (AVERAGE OF SECOND ORDER)

A good measure of dispersion should obey conditions similar to those for a satisfactory average and are as follows :

i. It should be rigidly defined.

ii. It should be based on all observations.

iii. It should be readily comprehensible.

iv. It should be fairly easily calculated.

v. It should affected as little as possible by fluctuations of sampling;

vi. It should readily lend itself to algebraic treatment and

vii. It should be east affected by the presence by extreme values

Measure of Dispersion

- **Absolute**
  - Range
  - Quartile Deviation or Semi Inter Quartile Range
  - Mean Deviation Or Mean Absolute Deviation
  - Standard Deviation
- **Relative**
  - Coefficient of Range
  - Coefficient of Quartile Deviation
  - Coefficient of Mean Deviation
  - Coefficient of Variation

RANGE

- It is the quickest measure, of finding out Dispersion
- It does not depend on all observations
- It's a crude method of finding out dispersion and most unreliable
- Range is unaffected by the presence of frequency
- Range is independent of the change of origin but dependent on change of scale
- If $y=a\pm bx$

  $R(y)=|b| \times R(x)$

Calculation Of Range:

- For simple series and simple Frequency Distribution :

  Range = Highest Value – Lowest Value (H – L).

- For grouped frequency distribution:
  - o Range = Upper boundary of last class – Lower boundary of 1st class
  - o Range = Upper Limit of last class – Lower limit of 1st class + 1

- Co-efficient of Range (Relative Range) = $\frac{H - L}{H + L} \times 100$

Quartile Deviation or Semi-inter quartile Range:

- QD is defined as the half of the range between the quartiles

- It is based on the upper and the lower Quartile and covers 50% of the observations.

- It does not depend on all observations

- For distributions with the Open Ends Q.D is the best and only measure of dispersion.

- QD is independent of the change of Origin but dependent on the change of Scale.

- If $y=a\pm bx$

  $QD(y)=|b| \times QD(x)$

- Quartile Deviation (QD) = $\frac{Q_3 - Q_1}{2}$, Where Q3 is the upper quartile and Q1 is the lower quartile.

- Co-efficient of QD(Relative Measure) = $\frac{QD}{Median} \times 100 = \frac{\frac{Q_3 - Q_1}{2}}{Q_2} \times 100 = \frac{Q_3 - Q_1}{2Q_2} \times 100$

- For symmetrical distribution; $Q_2 = \frac{Q_1 + Q_3}{2}$, i.e., median is the average of two extreme quartiles.

  Thus coefficient of QD for symmetrical distribution = $\frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} \times 100 = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

Mean Deviation / Mean Absolute Deviation

- It is based on all observations and hence it provides much better dispersion than Range and Quartile Deviation

- Mean deviation of a set of values of a variable is defined as the AM of the Absolute Deviation taken about Mean, Median or Mode.(Preferably AM or Median)

- Absolute Deviation implies Deviation without any regard to sign

- If nothing is specified Mean Deviation will imply Deviation about AM only.

- Since sum of Deviations is least when Deviations are taken about Median hence MD about Median will have the least value.

- MD is the independent of the change of origin but dependent on the change of scale

- If $y=a\pm bx$
  $MD(y)=|b| \times MD(x)$

- **Formula to calculate Mean Deviation:**

Simple Series	Simple / Grouped Frequency Distribution
$MD = \frac{\sum\lvert x-\bar{x}\rvert}{n}$	$MD = \frac{\sum f\lvert x-\bar{x}\rvert}{\sum f}$
$MD = \frac{\sum\lvert x-M\rvert}{n}$	$MD = \frac{\sum f\lvert x-M\rvert}{\sum f}$

Where n = number of observation
$\sum f=N$ = Total frequency
$\bar{x}$ =A.M
M = Median
X=Either actual values of the variables or mid values if it a group frequency distributions

o Coefficient of MD(Relative Measure) = $\frac{MD}{Mean/Median} \times 100$

Standard Deviation

- It is the best measure and the most commonly used Measure of Dispersion.

- It takes into consideration the magnitude of all the observations and gives the minimum value of dispersion possible.

- SD has all the pre-requisites of a good measure of dispersion, except the fact that it gets unduly affected by the presence of extreme values,

- It is also known as Root Mean Square Deviation about mean.

- It is denoted by σ
- SD^2 = Variance= σ^2
- If all observations are equal variance =SD=0
- SD is the independent of the change of origin but dependent on the change of scale
- If y=a±bx

  SD(y)=|b| ×SD(x)

  V(y)=b^2×v(x)

Definition of SD:

- SD of a set of values of a variable is defined as the positive Square Root of the AM of the Square of Deviations of the values from their AM
- Thus, SD is also known as Root - Mean – Square - Deviations (RMSD)

Calculation of SD

Simple Series(Without Frequency)	Simple /Grouped Frequency Distribution
i) $\sigma = \sqrt{\dfrac{\sum (x-\bar{x})^2}{n}}$	i) $\sigma = \sqrt{\dfrac{\sum f(x-\bar{x})^2}{\sum f}}$
ii) $\sigma = \sqrt{\dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2}$	ii) $\sigma = \sqrt{\dfrac{\sum fx^2}{\sum f} - \left(\dfrac{\sum fx}{\sum f}\right)^2}$
iii) $\sigma_x = \sqrt{\dfrac{\sum d^2}{n} - \left(\dfrac{\sum d}{n}\right)^2} \times i$	iii) $\sigma_x = \sqrt{\dfrac{\sum fd^2}{\sum f} - \left(\dfrac{\sum fd}{\sum f}\right)^2} \times i$

- Where, $d = \dfrac{x-A}{i}$, ,

x= mid-values if it is a grouped frequency distribution or original values if it is a discrete series

A = Assumed Mean i.e., a value arbitrarily chosen from mid-values or any other value.

i = class width or any arbitrary value

Note1 : Use form i) when you find that $\bar{x}$ is whole number

Note2 : Use form ii) when the value of the variable x are small

Note3 : Use Form iii) when you find that the values of x are large $\bar{x}$ is not a whole number(usually to be used for grouped frequency distribution)

USEFUL RESULTS:

- SD of two numbers is the half of their absolute difference(Range), i.e., if numbers are a and b, then SD = $\left|\frac{a-b}{2}\right|$
- Variance of first "n" natural numbers (1, 2, 3, ..., n) is $\frac{n^2-1}{12}$
- Sum of the squares of observations $\sum x^2 = n(\sigma^2 + \bar{x}^2)$

Formula for combined or composite or pooled S.D. of two groups

	Group I	Group II
Numbers	n_1	n_2
Mean	$\bar{x}_1$	$\bar{x}_2$
Standard Deviation	σ_1	σ_2

- Step 1 – Find Combined Mean: $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$
- Step 2 – Find Deviations : $d_1 = \bar{x}_1 - \bar{x}$ $\quad d_2 = \bar{x}_2 - \bar{x}$
- Step 3 – Use Formula: $\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}$
- Coefficient of Variation (C.V)(Relative Measure) = $\frac{\text{SD}}{\text{Mean}}$ x $100 = \frac{\sigma}{\bar{x}} \times 100$
- C.V is the best relative measure of dispersion
- C.V is used to compare variability or consistency between 2 or more series
- More C.V implies more variability indicating thereby less stability or consistency and vice versa.
- Regarding choice of an item always choose that item which has less C.V, because the item with lower C.V is more stable.

CLASSWORK SECTION

RANGE

1. If R_x and R_y denote ranges of x and y respectively where x and y are related by 3x+2y+10=0,
   what would be the relation between x and y?
   a) $R_x = R_y$ b) $2 R_x = 3 R_y$ c) $3 R_x = 2 R_y$ d) $R_x = 2 R_y$

2. If the range of x is 2, what would be the range of −3x +50 ?
   a) 2 b) 6 c) −6 d) 44

QUARTILE DEVIATION

3. If x and y are related as 3x+4y = 20 and the quartile deviation of x is 12, then the quartile deviation of y is
   a) 16 b) 14 c) 10 d) 9.

MEAN DEVIATION

4. What is the value of mean deviation about mean for the following numbers?
   5, 8, 6, 3, 4.
   a) 5.20 b) 7.20 c) 1.44 d) 2.23

5. If the relation between x and y is 5y−3x = 10 and the mean deviation about mean for x is 12, then the mean deviation of y about mean is
   a) 7.20 b) 6.80 c) 20 d) 18.80.

6. If two variables x and y are related by 2x + 3y −7 =0 and the mean and mean deviation about mean of x are 1 and 0.3 respectively, then the coefficient of mean deviation of y about its mean is
   a) −5 b) 12 c) 50 d) 4.

7. What is the mean deviation about median for the following data?

X	3	5	7	9	11	13	15
F	2	8	9	16	14	7	4

a) 2.50 b) 2.46 c) 2.43 d) 2.37

STANDARD DEVIATION

8. What is the coefficient of variation of the following numbers?
53, 52, 61, 60, 64.

a) 8.09 b) 18.08 c) 20.23 d) 20.45

9. If the SD of x is 3, what is the variance of (5−2x)?

a) 36 b) 6 c) 1 d) 9

10. If x and y are related by y = 2x+ 5 and the SD and AM of x are known to be 5 and 10 respectively, then the coefficient of variation of y is

a) 25 b) 30 c) 40 d) 20

11. What is the coefficient of variation for the following distribution of wages?

Daily Wages (₹):	30 − 40	40 − 50	50 − 60	60 − 70	70 − 80	80 − 90
No. of workers	17	28	21	15	13	6

a) ₹ 14.73 b) 14.73 c) 26.93 d) 20.82

COMBINED STANDARD DEVIATION

12. If two samples of sizes 30 and 20 have means as 55 and 60 and variances as 16 and 25 respectively, then what would be the SD of the combined sample of size 50?

a) 5.00 b) 5.06 c) 5.23 d) 5.35

CORRECTION IN STANDARD DEVIATION

13. The mean and SD of a sample of 100 observations were calculated as 40 and 5.1 respectively by a CA student who took one of the observations as 50 instead of 40 by mistake. The correct value of SD would be

a) 4.90 b) 5.00 c) 4.88 d) 4.85.

THEORETICAL ASPECTS

14. When it comes to comparing two or more distributions we consider
    a) Absolute measures of dispersion
    b) Relative measures of dispersion
    c) Both (a) and (b)
    d) Either (a) or (b).

15. Which one is an absolute measure of dispersion?
    a) Range
    b) Mean Deviation
    c) Standard Deviation
    d) All these measures

16. Which measures of dispersions is not affected by the presence of extreme observations?
    a) Range
    b) Mean deviation
    c) Standard deviation
    d) Quartile deviation

17. Which measure of dispersion is based on all the observations?
    a) Mean deviation
    b) Standard deviation
    c) Quartile deviation
    d) (a) and (b) but not (c)

18. The appropriate measure of dispersion for open-end classification is
    a) Standard deviation
    b) Mean deviation
    c) Quartile deviation
    d) All these measures.

19. A shift of origin has no impact on
    a) Range
    b) Mean deviation
    c) Standard deviation
    d) All these and quartile deviation.

20. If all the observations are increased by 10, then
    a) SD would be increased by 10
    b) Mean deviation would be increased by 10
    c) Quartile deviation would be increased by 10
    d) All these three remain unchanged.

21. If all the observations are multiplied by 2, then

a) New SD would be also multiplied by 2

b) New SD would be half of the previous SD

c) New SD would be increased by 2

d) New SD would be decreased by 2.

ANSWERS - SUMS				ANSWERS - THEORITICAL ASPECTS			
Q. No.	Ans	Q. No.	Ans	Q. No.	Ans	Q. No.	Ans
1	c	8	a	14	b	20	d
2	b	9	a	15	d	21	a
3	d	10	c	16	d		
4	c	11	b	17	d		
5	a	12	b	18	c		
6	b	13	b	19	d		
7	d						

3A CORRELATION ANALYSIS

- **Correlation is the degree of association between two or more variables**
- In other words, correlation measures the degree or extent to which two variables move in sympathy.
- This association or lack of association is measured by means of a coefficient called correlation coefficient.
- It is a pure number without any unit and the value of which lies between -1 and +1
  a. When correlation coefficient is +1, perfect positive Correlation
  b. When correlation coefficient is -1, perfect negative Correlation
  c. When correlation coefficient is 0, no correlation

  In the given context we are concerned with,
  i. Correlation between two variables i.e., x and y (Bivariate Correlation).
  ii. Correlation implies Linear correlation only.
- Correlation coefficient is independent of change in Origin and Scale.

Note:

Concept of Spurious or Nonsense correlation:

Sometimes it is found that there is no casual relation between two variables but due to presence of a third variable a correlation can be observed between the two. This variable which is responsible for the correlation other two variable is called "Lurking variable".

Methods of calculating correlation coefficient:

1. Karl-Pearson's Coefficient of Correlation or Product-Moment Correlation Coefficient or Correlation Coefficient by Covariance Method (r)

i. $$r = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

Where,

Cov(x, y) = Covariance between x and y

$$Cov(x, y) = \frac{1}{n}\sum(x - \bar{x})(y - \bar{y}) = \frac{\sum xy}{n} - \left(\frac{\sum x}{n}\right)\left(\frac{\sum y}{n}\right)$$

ii. Thus, $$r = \frac{\frac{\sum xy}{n} - \left(\frac{\sum x}{n}\right)\left(\frac{\sum y}{n}\right)}{\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}\sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}}$$

iii. When deviations are taken from actual means say $\bar{x}$ and $\bar{y}$ such that u=x-$\bar{x}$ and v=y-$\bar{y}$ in such a case r will be given by, $$r = \frac{\sum uv}{\sqrt{\sum u^2 \cdot \sum v^2}}$$

iv. When deviations are taken from assumed means say 'a' from X and 'b' from Y such that u=X-a and v=Y-b in such a case 'r' is given by,

$$r = \frac{\frac{\sum uv}{n} - \frac{\sum u}{n} \cdot \frac{\sum v}{n}}{\sqrt{\frac{\sum u^2}{n} \left(\frac{\sum u}{n}\right)^2}\sqrt{\frac{\sum v^2}{n} - \left(\frac{\sum v}{n}\right)^2}}$$

Note 1: Use (i) when you find that cov (x, y), σ_x and σ_y are provided

Note 2: Use (ii) when you find that the values of x and y are small

Note 3: Use (iii) when you find that $\bar{x}$ and $\bar{y}$ are whole numbers

Note 4: Use (iv) when you find that $\bar{x}$ and $\bar{y}$ are not whole numbers or the values of x and y are large or the problems specifically directs that the deviations are to be taken from assumed mean only.

2. Spearman's Rank Correlation Coefficient:

- Rank correlations is used for Qualitative data like beauty, intelligence etc.
- It is used for measuring correlation between two attributes.
- It is denoted by 'R' $(-1 \leq R \leq +1)$

Formula for rank correlation,

Case 1: without tie-when all the variables have different ranks

$$R = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

Where,

n = Total number of individuals

D = Rank difference= $R_x - R_y$

Case 2: Tied Ranks

i. In such cases two or more variables have the same score and accordingly average ranks are assigned to the variables which are involved in the tie.

ii. The Formula in such a case

$$R = 1 - \frac{6\left[\sum D^2 + \sum \frac{t^3 - t}{12}\right]}{n(n^2 - 1)}$$

Where,

t = number of variables are involved in tie.

n = total number of variables

D = $R_x - R_y$ =Rank difference

3. Concurrent Deviation Method or Coefficient of Concurrent Deviation [r]:

- **It is the simplest and quickest method of calculating correlation**
- **It is used to know the direction changes between two variables**
- **It is suitable only when the variable includes short term fluctuations**
- **It lies between -1 and +1**

- Let (x_1,y_1), (x_2,y_2), …. …., (x_{n+1},y_{n+1}) be a set of (n+1) pairs of values of x and y. Let C_x and C_y denote the direction changes in the values of x and y i.e., C_x and C_y will have positive signs if there is an increase in the values of x and y w.r.t its immediate preceding value and will have negative signs in case of decrease.

If C denotes the number of concurrent deviations i.e., total number of positive signs in the $C_x . C_y$ column then the coefficient of concurrent deviation is given by,

$$r = \pm\sqrt{\pm\left(\frac{2C-n}{n}\right)}$$

Where,

n = pairs of deviations compared

c = number of concurrent deviations

i. If $\frac{2C-n}{n}$ is positive, positive sign is to be assigned both inside and outside the square root.

ii. If $\frac{2C-n}{n}$ is negative, negative sign is to be assigned both inside and outside the square root.

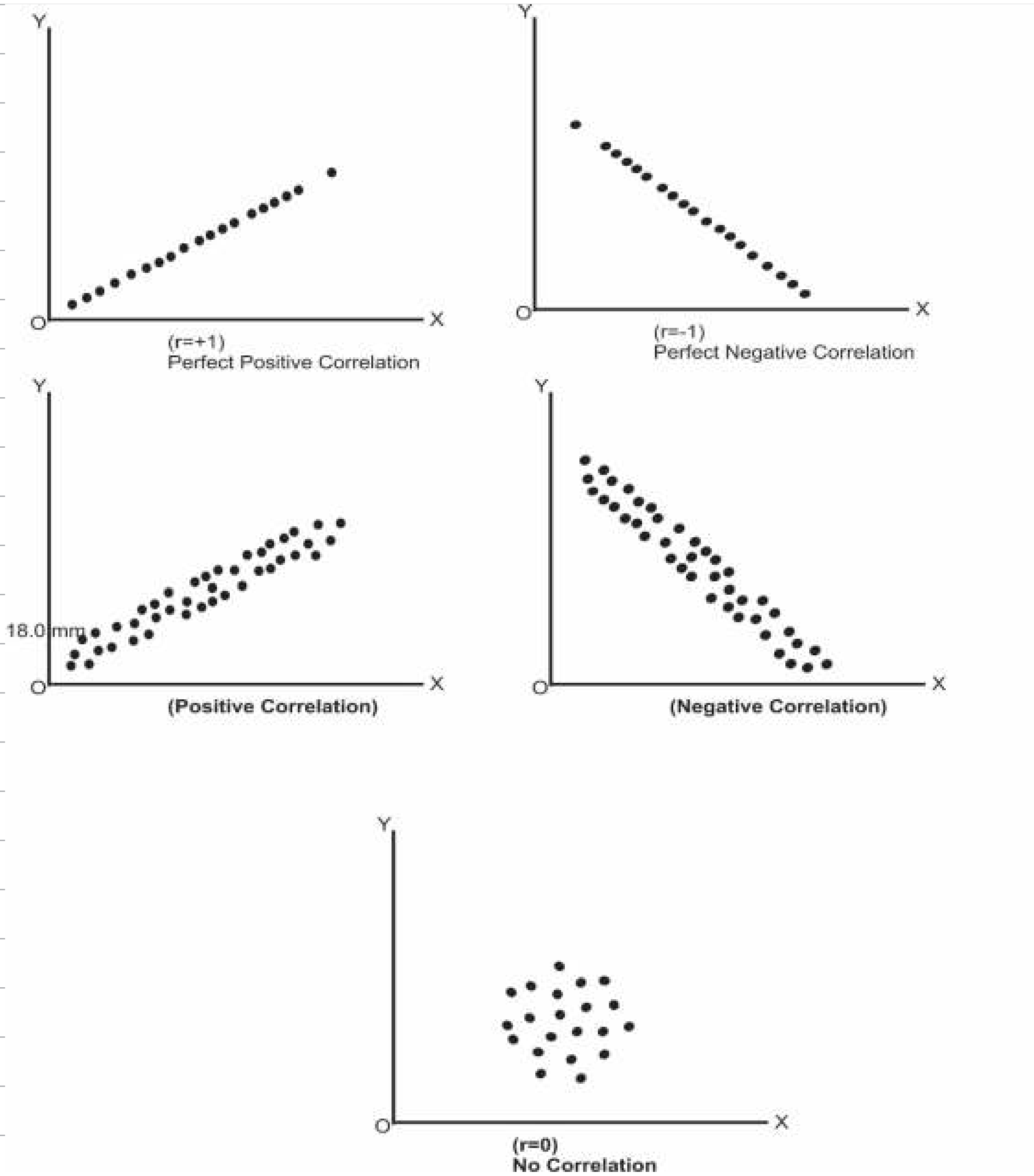
iii. When C = 0, r = -1

iv. When C = n, r = 1

v. When C = $\frac{n}{2}$, r = 0

4. Diagramatic representation of correlation through scatter diagram or scatter plot:

- It the simplest way to represent bivariate data
- It gives a vague idea about the nature of correlation between two variables
- It helps us to distinguish between different types of correlation but fails to measure the extent of relationship between the variables
- Through scatter diagram we can get an idea about the nature of correlation; positive, negative, zero or curvilinear



Properties of Correlation of Coefficient 'r':

- Coefficient correlation is symmetric i.e., $r_{xy} = r_{yx}$
- If y = a+ bx then,

  i. r = +1 when b>0 and

  ii. r = -1 when b<0

- correlation coefficient is independent of the change of origin and scale.

  If $u = \frac{x-a}{c}$ and $v = \frac{y-b}{d}$ then,

a. $r_{uv} = r_{xy}$, if c and d are of the same sign

b. $r_{uv} = -r_{xy}$, if c and d are of the opposite sign

Miscellaneous Properties:

- Coefficient of determination = r^2

$$r^2 = \frac{Explained Variance}{Total Variance} = 1 - \frac{Unexplained Variance}{Total Variance}$$

- Coefficient of Non -Determination : $1 - r^2 = \frac{Unexplained Variance}{Total Variance}$

- Coefficient of alienation = square root of coefficient of non-determination = $\sqrt{1 - r^2}$

- Percentage of explained variation = $r^2 \times 100$

- Percentage of unexplained variation = $(1-r^2)\times 100$

- Standard error of r (S.E of r) = $\frac{1-r^2}{\sqrt{n}}$

- Probable error of r [P.E (r)] = 0.6745 x SE(r)

- Probable error and standard error both are used for determining the reliability of correlation coefficient. For this purpose the following rule is followed,

  1) **If r < P.E. $\longrightarrow$ there is no significant correlation in population.**
  2) **If r > 6 P.E. $\longrightarrow$ there is significant correlation in population and we can rely on the value of r**
  3) **Otherwise, in the intermediate interval $\longrightarrow$ there is no clear idea about the correlation in the population and hence no inference can be drawn about the population correlation coefficient (ρ).**

  Using probable Error (P.E.), we can find the probable limits for population correlation coefficient (ρ) as follows

  Probable limits = r $\pm$ P.E

  = (r-P.E) to (r+ P.E)

- Let x and y be two correlated variables, then: $V(x \pm y) = V(x) + V(y) \pm 2Cov(x, y)$

- Let x and y are two uncorrelated variables, then Cov(x,y) = 0 and hence,
$V(x \pm y) = V(x) + V(y)$

BIVARIATE DATA

- When a set of data is collected for two variables simultaneously it is called a **Bivariate Data**

- When a frequency distribution is formed with these bivariate data it is known as **Bivariate Frequency Distribution** or **Joint Frequency Distribution** or **Two Way Distribution**

- The tabular representation of this frequency distribution is known as **Two Way Frequency Table**

- Following is a bivariate table for the data relating to marks in maths and statistics

	Marks in Mathematics					
Marks in Stats	0-4	4-8	8-12	12-16	16-20	Total
0-4	1	1	2	0	0	4
4-8	1	4	5	1	1	12
8-12	1	2	4	6	1	14
12-16	0	1	3	2	5	11
16-20	0	0	1	5	3	9
Total	3	8	15	14	10	50

Observations:

- **A bivariate frequency distribution having m rows and n columns has m x n cells**
- **Some of the cell frequencies may be zero**

From a bivariate distribution we can have the following two types of Uni-variate distributions

i. **Two Marginal Distributions**

ii. **m+n Conditional Distributions**

From the above table the two marginal distributions are as follows,

Marginal Distribution of Marks in Mathematics

Marks	No of students
0-4	3
4-8	8
8-12	15
12-16	14
16-20	10
Total	50

Similarly, we can have Marginal Distribution for marks in statistics

From the above table, an example a Conditional distribution of marks in Statistics when the mathematics marks lie between <u>8-12</u>

Marks	No of students
0-4	2
4-8	5
8-12	4
12-16	3
16-20	1
Total	15

Bivariate Relationship

Between two variables x and y there can exist any of the following three relationship

a. **Direct or Positive –** with change in one variable x, the other variable y will also change in the same direction. Eg: Price and quantity supplied: amount of rainfall and crop yield

b. **Indirect or Inverse or Negative –** With change in one variable, the other variable will change in the opposite direction. Eg: Price and quantity demanded.

c. **No relation –** With change in one variable x, if another variable y doesn't show any specific trend (increasing or decreasing), then we say there exist no relation between x and y.

CLASSWORK SECTION

Product Moment Method/ Covariance Method

1. The Cov (x, y) =15, what restrictions should be put for the standard deviations of x and y?
   a) No restriction
   b) The product of the standard deviations should be more than 15
   c) The product of the standard deviations should be less than 15
   d) The sum of the standard deviations should be less than 15

2. Find the coefficient of correlation from the following data:

X:	1	2	3	4	5
Y:	6	8	11	8	12

   a) + 0.775 b) − 0.775 c) + 0.895 d) + 0.956

3. Calculate correlation coefficient from the following data: n = 12, $\sum x = 120$, $\sum y = 130$, $\sum (x-8)^2 = 150$, $\sum (y-10)^2 = 200$, $\sum (x-8)(y-10) = 50$.
   a) 0.215 b) − 0.215 c) − 0.317 d) None of the above

4. Find the number of pairs of observation from the following data: r = 0.25, $\sum(x-\bar{x})(y-\bar{y}) = 60$, $\sum(x-\bar{x})^2 = 90$, $SD_y = 4$.
   a) 30 b) 40 c) 20 d) 10

Rank Correlation Coefficient "R"

5. The coefficient of rank correlation between the marks in Statistics and Mathematics obtained by a certain group of students is 2/3 and the sum of the squares of the differences in ranks is 55. How many students are there in the group?
   a) 10 b) 9 c) 12 d) more than 15

6. From the following data calculate the value of coefficient of Rank correlation:

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

   a) 0.93 b) − 0.85 c) 0.85 d) 0.63

Concurrent Deviation Method

7. What is the coefficient of concurrent deviations for the following data:

Supply:	68	43	38	78	66	83	38	23	83	53	48
Demand:	65	60	55	61	35	75	45	40	85	80	85

a) 0.82 b) 0.85 c) 0.89 d) − 0.81

8. The coefficient of concurrent deviation for p pairs of observation was found to be $1/\sqrt{3}$. If the number of concurrent deviations was found to be 6, then the value of p

a) 10 b) 9 c) 8 d) None of these

Change of Origin and Change of Scale

9. If u + 5x = 6 and 3y + 7v = 20 and the correlation coefficient between x and y is 0.58 then where would be the correlation coefficient between u and v?

a) 0.58 b) -0.58 c) -0.84 d) 0.84

Theoretical Aspects

10. Correlation co-efficient is ______ of the units of measurement

a) Dependent b) Independent
c) Both d) None

11. In Case of “insurance companies” profit and the number of claims They have pay there is _______ correlation.

a) Positively b) Negative
c) No of correlation d) None of these

12. Which of the following regarding value of “r” is TRUE?

a) “r” is a pure number
b) “r” lies between −1 and +1 both inclusive
c) Neither (a) nor (b)
d) Both a) and b) are true

13. For which of the following statements the correlation will be negative?
   a) Production and price per unit
   b) Sale of woolen garments and day temperature
   c) Neither (a) nor (b)
   d) Both a) and b) above

14. Karl Pearson's correlation coefficient may be defined as:
   a) The ratio of covariance between the two variables to the product of the standard deviations of the two variables.
   b) The ratio of covariance between the two variables to the product of the variance of the two variables.
   c) The ratio of product of standard deviations of the two variables to the covariance between the two variables.
   d) None of the above.

Rank Correlation:

15. Rank of beauty contest by two judges are in reverse orders the find the value of spearmen's rank correlation co-efficient
   a) -1   b) 0   c) 1   d) 0.75

16. Sum of the difference in ranks is always ______
   a) 1   b) 2   c) -1   d) 0

Properties:

17. In case the correlation coefficient between two variables is 1, which of the following would be the relationship between the two variables?
   a) $y = p + qx, q > 0$ b) $y = p + qx, q < 0$
   c) $y = p + qx, p > 0, q < 0$ d) Both a) and b) above

18. If the relationship between two variables x and y is given by $22x + 33y + 84 = 0$, then the value of correlation coefficient between x and y will be:
   a) 1.00   b) 0
   c) – 1.00   d) Between 0 and 1.00

19. The co-efficient of correlation between x and y is 0.6. if x and y both are multiplied by -1, then the co-efficient of correlation will be:

a) 0.6 b) - 0.6 c) $\frac{1}{0.6}$ d) $1-(0.6)^2$

20. Which of the following regarding value of "r" is TRUE?

a) It is not affected by change in scale.
b) It is not affected by change of origin.
c) It is both affected by change in scale and origin.
d) Both a) and b) above are true.

Application of r:

21. A relationship $r^2 = 1 - \frac{500}{300}$ is not possible

a) True b) False c) Both d) None

Scatter Diagram:

22. When the correlation coefficient r=+1, all the points in a scatter diagram would be

a) On a straight line directed from upper left to lower right
b) On a straight line directed from lower left to upper right
c) On a straight line
d) Both (a) and (b)

Bivariate Data:

23. From the Bivariate Frequency Distribution, we can obtain which of the following Univariate distribution?

a) Marginal distribution
b) Conditional distribution
c) Both a) and b) above
d) Neither a) nor b) above

3B REGRESSION ANALYSIS

Introduction

- Regression is the average linear relationship between two or more variables.

- The word regression implies "estimation or prediction". In other words through regression equations we can quantify the relationship between two variables and we can predict the average value of one variable corresponding to a specific value of the other.

- It establishes a functional relationship between two variables.

- Regression equation enables us to find the nature and the extent of relationship between two variables. Correlation can measure only the degree of association between the two variables whereas regression quantifies such relationship.

- The two variables are dependent and independent variable. Thus, we try to estimate the average value of dependent variable, for a specified value of independent variable using regression analysis.

- If there are two variables, then the independent variable is called the "Regressor" or "Explaining Variable" and the dependent variable is called the "Regressed" or "Explained Variable".

- Regression analysis is an absolute measure showing a change in the value of y or x for a corresponding unit change in the value of x or y whereas correlation coefficient is a relative measure of linear relationship between x and y.

- This average linear relationship between two variables is expressed by means of two straight line equation known as regression lines or regression equations.

- If there are two variables x and y we can have the following two types of regression lines,

  i. Regression equation of y on x (y dependent, x independent)

  ii. Regression equation of x on y (x dependent, y independent)

REGRESSION LINES

Regression equation of y on x:	Regression equation of x on y:
• $(Y-\bar{Y}) = b_{yx}(X-\bar{X})$	• $(X-\bar{X}) = b_{xy}(Y-\bar{Y})$
• b_{yx} stands for regression coefficient of y on x	• b_{xy} stands for regression coefficient of x on y
• Here y depends on x	• Here x depends on y
• Here y is a dependent/explained and x is an independent variable	• Here x is a dependent/explained and x is an independent variable
• This equation will be of the form y = a + bx	• This equation will be of the form x = a + by
• This equation is used to estimate the value of y given the value of x	• This equation is used to estimate the value of x given the value of y
• The slope of this equation is byx	• The slope of this equation is bxy
• The regression line of y on x is the straight line on the scatter diagram for which the sum of squares of vertical distances of the points is minimum.	• The regression line of x on y is derived by the minimization of horizontal distance in the scatter diagram using method of least square.
• The principle which is applied for deriving the two lines of regression is known as "Method of Least Squares".	• The principle which is applied for deriving the two lines of regression is known as "Method of Least Squares".

CALCULATION OF REGRESSION COEFFICIENTS

Regression coefficient of y on x (b_{yx}):	Regression coefficient of x on y (b_{xy}):
1. Using co-variance: $\frac{Cov(x,y)}{\sigma_x^2}$	1. Using co-variance: $\frac{Cov(x,y)}{\sigma_y^2}$
2. Without any deviations (Directly from x and Y values) $= \dfrac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$	2. Without any deviations (Directly from x and Y values) $= \dfrac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}$
3. When deviations are taken from actual mean i.e., $\bar{x}$ and $\bar{y}$ such that $u = x - \bar{x},\ v = y - \bar{y}$ $= \dfrac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \dfrac{\sum uv}{\sum u^2}$	3. When deviations are taken from actual mean i.e., $\bar{x}$ and $\bar{y}$ such that $u = x - \bar{x},\ v = y - \bar{y}$ $= \dfrac{\sum (x-\bar{x})(y-\bar{y})}{\sum (y-\bar{y})^2} = \dfrac{\sum uv}{\sum v^2}$
4. When deviations are taken from assumed mean say A & B for x and y, u=x-A, v = y-B $b_{yx} = \dfrac{\frac{\sum uv}{n} - \frac{\sum u}{n} \cdot \frac{\sum v}{n}}{\frac{\sum u^2}{n} - \left(\frac{\sum u}{n}\right)^2}$	4. When deviations are taken from assumed mean say A & B for x and y, u=x-A, v = y-B $b_{xy} = \dfrac{\frac{\sum uv}{n} - \frac{\sum u}{n} \cdot \frac{\sum v}{n}}{\frac{\sum v^2}{n} - \left(\frac{\sum v}{n}\right)^2}$
5. Using 'r' $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$ $\sigma_x = S.D(x), \sigma_y = S.D(y)$ and r = Correlation co-efficient between x and y	5. Using 'r' $b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$ $\sigma_x = S.D(x), \sigma_y = S.D(y)$ and r = Correlation co-efficient between x and y

PROPERTIES OF REGRESSION COEFFICIENTS

1. **b_{yx} = slope of the regression line of y on x which measures the change in variable y for a unit change in variable x.**

2. b_{xy} = slope of the regression line of x on y which measures the change in variable x for a unit change in variable y.

3. Correlation coefficient is symmetric i.e., $r_{yx} = r_{xy}$ but regression coefficients are not symmetric $b_{yx} \neq b_{xy}$.

4. When r = 0, both the regression coefficients are 0.

5. Both the regression coefficients will have same sign.

6. Correlation coefficient is the geometric mean between regression coefficients i.e.,

$$r = \pm\sqrt{b_{yx} \cdot b_{xy}}$$

7. Sign analogy of b_{yx}, b_{xy} and r

b_{yx}	b_{xy}	r
+	+	+
–	–	–

Note:

When b_{yx} and b_{xy} are of opposite signs, data are inconsistent, r is imaginary.

8. Regression coefficients are independent of the Change of Origin but they are dependent on Change of Scale. If $u = \frac{x-a}{c}$ and $v = \frac{y-b}{d}$ then

i. $b_{yx} = b_{vu} \cdot \frac{d}{c}$

ii. $b_{xy} = b_{uv} \cdot \frac{c}{d}$

9. There is no specific range within which two regression coefficients will lie but their values should be such that the square root of the product of two regression coefficients must lie between -1 and +1 (both inclusive). Thus, if one of the regression coefficient, is greater than unity then the other must be less than unity.

Properties of regression lines:

- Two regression lines always intersect at their mean or average values $(\bar{x}, \bar{y})$. In other words if we solve two regression equations we get the average values of x and y.

- When r = 0, then

i. $b_{yx} = b_{xy} = 0$

ii. The two regression lines thus reduces to; y = $\bar{y}$ and x = $\bar{x}$

iii. Nothing can be predicted from the two regression lines since, the variables become independent.

iv. The angle between the two regression lines becomes 90° i.e., they are perpendicular to each other.

- When r = ± 1, then
  i. The two regression lines become identical i.e., they coincide.

  ii. $b_{yx} = \frac{1}{b_{xy}}$

  iii. Perfect linear co-relationship is observed and the angle between the two regression lines becomes 0°.

  iv. For a particular value of x we shall obtain a specific value of y.

- As the angle between two regression lines numerically decreases from 90° to 0°, the correlation increases from 0 to 1 and the two regression lines comes closer to each other.

- Angle between two regression lines; if A is the angle between two regression lines then

$$\tan A = \pm \frac{1-r^2}{r}\left(\frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

Miscellaneous Properties:

- In regression analysis, the difference between the Observed value and the Estimated value is known as Residue or Error.

- Proportion of Total Variance explained by regression analysis is r^2.

- Proportion of Total Unexplained Variance is $(1 - r^2)$.
- Standard error of estimate of $x(S_{xy})$ is given by $S_{xy} = \sqrt{\frac{\sum (x - \bar{x}_c)^2}{N}}$ or $\sigma_x \sqrt{1 - r^2}$

- Standard error of estimate of $y(S_{yx})$ is given by $S_{yx} = \sqrt{\frac{\sum (y - \bar{y}_c)^2}{N}}$ or $\sigma_y \sqrt{1 - r^2}$

- When $r^2 = 1$, then;

  i. $\frac{\text{Explained variance}}{\text{Total variance}} = 1$

  ii. Explained variance = Total Variance

  iii. The whole of the total variance is explained by regression.

  iv. The unexplained variation is zero

  v. All the points on the scatter diagram will lie on the regression line

  vi. There is a perfect linear dependence between the variables

  vii. The two regression lines coincide

  viii. For a given value of one variable, we have a fixed value of the other variable

CLASSWORK SECTION

1. Given the following data:

Variable:	x	y
Mean:	80	98
Variance:	4	9

Coefficient of correlation = 0.6

What is the most likely value of y when x = 90 ?

a) 90 b) 103 c) 104 d) 107

2. If 4y − 5x = 15 is the regression line of y on x and the coefficient of correlation between x and y is 0.75, what is the value of the regression coefficient of x on y?

a) 0.45 b) 0.9375 c) 0.6 d) none of these

3. Regression equation of Y on X is 8X − 10Y + 66 = 0 and SD(x) = 3, find the value of Cov (x, y).

a) 11.25 b) 7.2 c) 2.4 d) None of the above

Properties of Regression Coefficients

4. If bxy = - 1.2 and byx = - 0.3, then the coefficient of correlation between x and y is:

a) − 0.698 b) − 0.36 c) − 0.51 d) − 0.6

5. Given b_{xy} = 0.756, b_{yx} = 0.659, then the value of coefficient of non-determination is given by:

a) 0.402 b) 0.502 c) 0.602 d) 0.702

Change of Origin and Change of Scale

6. If u = 2x + 5, v = -3y + 1, and the regression coefficient of y on x is − 1.2, the regression coefficient of v on u is:

a) 1.8 b) − 1.8 c) 3.26 d) 0.8

Identification Problems

7. Two random variables have the regression lines 3x+2y=26 and 6x+y=31. The coefficient of correlation between x and y is :

a) -0.25 b) 0.5 c) -0.5 d) 0.25

8. The two lines of regression are given by

8x + 10y = 25 and 16x + 5y = 12 respectively.

If the variance of x is 25, what is the standard deviation of y?

a) 16 b) 8 c) 64 d) 4

Theoretical Aspects

9. The word regression is used to denote __________ of the average value of one variable for a specified value of the other variable.

a) Estimation b) Prediction
c) Either a) or b) above d) None of the above

10. Regression methods are meant to determine:

a) The nature of relationship between the variables.
b) The functional relationship between the two variables.
c) Both a) and b) above
d) Neither a) nor b) above.

11. The dependent variable in the regression analysis is one:

a) Which influences the value of the independent variable.
b) Whose value is to be predicted.
c) Which can choose its value independently.
d) None of the above.

12. The line of regression is:

a) The line which gives the best estimate to the value of one variable for any specified value of the other variable.
b) The line which gives the best estimate to the value of all variables for any arbitrary value of a constant variable.
c) The line showing the nature of relationship between two or more variables.
d) None of the above.

13. Since Yield of a crop depends upon amount of rainfall, we need to consider:
    a) The regression equation of yield on rainfall
    b) The regression equation of rainfall on yield
    c) Any one of a) or b) above can be considered
    d) Neither of a) or b) can be considered

Properties:

14. If r = +1, the two lines of regression become:
    a) Perpendicular to each other.
    b) Identical
    c) Parallel to each other.
    d) Either a) or c) above.

15. Correlation coefficient is the ________ of the two regression coefficients.
    a) Harmonic Mean
    b) Geometric Mean
    c) Arithmetic Mean
    d) Both b) and c) above

16. The sign analogy of correlation coefficient and two regression coefficients is:
    a) −, +, +   b) −, −, −   c) +, +, +   d) Both b) and c) above

17. When r = 0, the regression lines are:
    a) Parallel to each other
    b) Perpendicular to each other
    c) Coincides
    d) Either a) or b) above

18. Which of the following(s) is/are TRUE regarding regression coefficient?
    a) If $b_{xy} > 0$, then $r < 0$
    b) If $b_{xy} < 0$, then $r > 0$
    c) If the variable X and Y are independent, the regression coefficient is zero.
    d) The range of regression coefficient is −1 to +1.

19. Which of the following statement/s is/are FALSE regarding the regression coefficient?

    a) If one of the regression coefficient is greater than unity the other one is less than unity.
    b) The product of two regression coefficient is equal to the square of the correlation coefficient between the two variables.
    c) The regression coefficient lies between – infinity to + infinity.
    d) None of the above is FALSE.

20. Regression Coefficient of y on x=0.8. Regression coefficient of x on y =0.2 coefficient of correlation = -0.4. Given data is:

    a) Accurate   b) Inaccurate   c) True   d) None

21. If the regression coefficient of y on x is 4/3, then the regression coefficient of x on y is:

    a) More than 1   b) Less than 1
    c) Less than zero   d) None of the above

4 INDEX NUMBERS

Basic Concepts

- Index Numbers are special kind of averages, expressed in ratio, calculated as percentage and used as numbers.

- Index number is a number which is used as a tool for comparing prices and quantities of a particular commodity or a group of commodities in a particular time period with respect to other time period or periods.

- Index numbers indicate relative change in price or quantity or value expressed in percentage.

- Index numbers are always unit free.

- The year in which the comparison is made is called the "Current Year" and the year with respect to which the comparison is made is the "Base Year".

- Suppose Price Index in 2011 is 800 based on 1980 prices, then
  - o 1980 means base year with help of which comparison is done.
  - o If nothing is mentioned, base prices are always taken as 100.
  - o 2011 is the current year or present year.
  - o 800 is the index number or price index number.

- **Index numbers are of three types:**
  - o **Price Index –** When the comparison is made in respect of prices it is called price index numbers.

  - o **Quantity Index –** When the comparison is made in respect of quantities it is called Quantity of Volume Index Numbers.

  - o **Value Index –** When comparison is made in respect of values (Value = Price x Quantity), it is called Value Index Number.

- Terminology (Unless otherwise mentioned we shall be using the following notations)
  - I_{01} means Index Number for year "1" based on year "0"(Current with respect to base)
  - I_{10} means Index Number for year "0" based on year "1"(base with respect to current)
  - P_1 = Prices prevailing in current year (year 1)
  - P_0 = Prices prevailing in base year (year 0)
  - Q_1 = Quantity in current year
  - Q_0 = Quantity in base year
  - P_0Q_0 = Price x Quantity of Base Year (Value of the base year)
  - P_1Q_1 = Price x Quantity of Current Year (Value of Current Year)
  - V_{01} = Value Index of current year with respect to base year
  - V_{10} = Value Index of base year with respect to current year

- Concept of price Relative (PR) :

  Price relative is defined as the ratio of Current Year's price to the Base Year's price expressed as percentage Symbolically,

$$I = PR = \frac{P_1}{P_0} \times 100$$

Construction of Price Index Numbers

Method of Aggregates

Case: 1	Case: 2
Simple Aggregate of prices $P_{01} = \frac{\sum P_1}{\sum P_0} \text{x}100$	Weighted Aggregate of prices $P_{01} = \frac{\sum P_1 w}{\sum P_0 w} \text{x}100$

CALCULATION OF WEIGHTED AGGREGATE OF PRICES UNDER DIFFERENT TYPE OF WEIGHTS	
If w = Q_0 Laspeyre's Index $L_{01} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \text{x}100$	If w = Q_1 Paasche's Index $P_{01} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \text{x}100$
Fisher's Index GM of L and P $F_{01} = \sqrt{L \text{ x } P}$	Bowley's Index AM of L and P $B_{01} = \frac{L+P}{2}$

If w = $Q_0 + Q_1$ Marshall-Edgeworth Index $P_{01} = \frac{\sum P_1(Q_0+Q_1)}{\sum P_0(Q_0+Q_1)} \times 100 = \frac{\sum P_1Q_0 + \sum P_1Q_1}{\sum P_0Q_0 + \sum P_0Q_1} \times 100$	

Relative Method

First calculate Price Relative (PR) of each commodity. Price Relative (PR) is defined as the ratio of the current year's price to the base year's price, expressed as percentage and is given by $PR = \frac{P_1}{P_0} \times 100$

Case: 1	Case: 2
Simple AM of Price Relative	Weighted AM of Price Relative
$P_{01} = \frac{\sum PR}{n}$	$P_{01} = \frac{\sum PR.w}{\sum w}$
n=number of Commodities	$\sum$w= Total Weight

Note :

- GM is the best average in the construction of index numbers but practically we use AM, because G.M is difficult to compute.
- Marshall- Edgeworth's Index number is an approximation to Fisher's index number.
- Methods of Relatives are also known as Arithmetic Mean Method.
- When a series of Index Numbers for different years are expressed in a tabular form to compare the changes in different years, then this tabular representation of numbers is known as "Index Time Series".

Construction of Quantity Index Numbers

All the formula will remain same as in price index numbers, just interchange p and q, i.e., p to q and q to p. For example; if Laspeyer's Price Index is $\frac{\sum P_1Q_0}{\sum P_0Q_0} \times 100$, then Laspeyer's Quantity Index we can get by interchanging P to Q and Q to P, and hence it will be $\frac{\sum Q_1P_0}{\sum Q_0P_0} \times 100$.

Construction of Value Index Number

$$V_{01} = \frac{\sum P_1Q_1}{\sum P_0Q_0} x100$$

Cost of Living Index (CLI)

- **CLI is also known as Wholesale Price Index, Consumer Price Index or General Index.**
- **CLI is defined as the weighted AM of index numbers of few groups of basic necessities. Generally for calculating CLI; food, clothing, house rent, fuel & lightning and miscellaneous groups are taken into consideration.**
- $CLI = \frac{\sum Iw}{\sum w}$ **, where I = Individual Group Index and w = Group weight.**
- **Application of Cost of Living Index**
  - o **It helps to calculate the purchasing power of money and real income of the consumer.**
  - o **Increase in CLI implies increase in price index causing thereby an inflation i.e. reduction in the purchasing power.**
  - o **Purchasing Power of ₹ 1 =** $\frac{100}{\text{Cost of Living Index}} \text{ x } 1$
  - o **Real Income =** $\frac{\text{Money or Nominal Income}}{\text{Cost of Living Index}} \text{ x } 100$
- **Concept of Equivalent Salary – Calculation of Dearness Allowances(D.A)**

Suppose a person was getting a money income of ₹ X_1 in Year 1 (Y_1) when the CLI was I_1 and in Year 2 (Y_2) the CLI is I_2. If the person wants to maintain his former standard of living as in Y_1, then Real Income (RI) of Y_1 should be equal to RI of Y_2.

Thus Money Income required in Y_2 = $\frac{\text{CLI of } Y_2}{\text{CLI of } Y_1} \text{ x Salary (Money Income) of } Y_1 \text{ (Rs. } X_1)$

Let the money income in Y_2 is X_2. If X_2 is less than or equal to X_1, then no allowances are required to be given. But if X_2 is greater than X_1, then amount of Dearness Allowances = ₹ $(X_2 - X_1)$

Base Shifting in Index Numbers

- Base Shifting is a process whereby a new series of Index Numbers with a new base year is formed from a given series of Index Numbers with another base year.

- Index Number for any year (with base year shifted) is given by:

$$\frac{\text{Old Index Number for the year}}{\text{Old Index Number for the New Base Year}} \times 100$$

Tests of Adequacy of Index Number

- **Unit Test** – An Index Number is a good index number if it is unit free. All index numbers will satisfy this test except Simple Aggregate of Prices.

- **Time Reversal Test (TRT)** – According to this test $I_{01} \times I_{10} = 1$ (ignore 100). This test is satisfied by:
  - o Simple Aggregate of Prices
  - o Weighted GM of Price Relative
  - o Marshall Edgeworth Index
  - o Fisher's Ideal Index

- **Factor Reversal Test (FRT)** – According to this test Price Index x Quantity Index = Value Index. Only Fisher's Ideal Index satisfies this test.

- **Circular Test** – Circular Test is an extension of Time Reversal Test. According to this test $I_{01} \times I_{12} \times I_{23} \times .... \times I_{(n-1), n} \times I_{n,0} = 1$. This test is satisfied by:
  - o Simple Aggregate of Prices (ie. Weighted Aggregate of Prices with Fixed Weights)
  - o Simple GM of Price Relatives

Fixed Base Method – Chain Base Method

- Under Fixed Base Method (FBM), all the index numbers are calculated with respect to a fixed base period.

- Under Chain Base Method (CBM), all the index numbers are calculated with respect to the price of immediate preceding period.

- Under CBM, the index number for the first year will always be 100.
- For the first year, Chain Base Index = Fixed Base Index.

- FBI for any year = $\dfrac{\text{Chain Base Index for the year x Fixed Base Index for the preceding year}}{100}$

- **Chain Index Numbers**
  - o Chain Index Numbers are calculated from Link Index Numbers or Link Relatives.
  - o Chain Index for any year = $\dfrac{\text{Link Index for the year x Chain Index for the preceding year}}{100}$

o Link Relative = $\dfrac{\text{Price Relative of the Current Year}}{\text{Price Relative of the preceding Year}} \times 100$

Note: Always start with one year preceding to the given years from which you are to calculate the chain index numbers. In that year (i.e. the preceding year) take both the link relative and the chain index to be 100.

Splicing of Index Numbers

- Splicing is a process whereby two or more discontinued series of index numbers with different base years are merged to form a new continuous series of index numbers with a new base year.

- The factor which is multiplied for such conversion is called **"Conversion Multiplier"**.

- Let there are two series Y_1 and Y_2. When the series Y_1 is merged into the series Y_2, it is known as "Forward Splicing" and when series Y_2 is merged into series Y_1, it is known as "Backward Splicing".

Stock Market Index:

It represents the entire stock market. It shows the changes taking place in the stock market. Movement of index is also an indication of average returns received by the investors. With the help of an index, it is easy for an investor to compare performance as it can be used as a benchmark, for e.g. a simple comparison of the stock and the index can be undertaken to find out the feasibility of holding a particular stock.

Each stock exchange has an index. For instance, in India, it is Sensex of BSE and Nifty of NSE. On the other hand, in outside India, popular indexes are Dow Jones, NASDAQ, FTSE etc.

(a) Bombay Stock Exchange Limited: It is the oldest stock exchange in Asia and was established as "The Native Share & Stock Brokers Association" in 1875. The

Securities Contract (Regulation) Act, 1956 gives permanent recognition to Bombay Stock Exchange in 1956. BSE became the first stock exchange in India to obtain such permission from the Government under the Act. One of the Index as BSE Sensex which is basket of 30 constituent stocks. The base year of BSE SENSEX is 1978-79 and the base value is 100 which has grown over the years and quoted at about 592 times of base index as on date. As the oldest Index in the country, it provides the time series data over a fairly long period of time (from 1979 onward).

(b) National Stock Exchange: NSE was incorporated in 1992. It was recognized as a stock exchange by SEBI in April 1993 and commenced operations in 1994.NIFTY50 is a diversified 50 stocks Index of 13 sectors of the economy. The base period of NIFTY 50 Index is 3 November 1995 and base value is 1000 which has grown over years and quoted at 177 times as on date.

Computation of Index

Following steps are involved in calculation of index on a particular date:

- Calculate market capitalization of each individual company comprising the index.
- Calculate the total market capitalization by adding the individual market capitalization of all companies in the index.
- Computing index of next day requires the index value and the total market capitalization of the previous day and is computed as follows:

$$\text{IndexValue} = \text{Index on Previous Day} \times \frac{\text{Total market capitalisation for current day}}{\text{Total market capitalisation for previous day}}$$

- It should also be noted that Indices may also be calculated using the price weighted method. Here, the share price of the constituent companies forms the weight. However, almost all equity indices worldwide are calculated using the market capitalization weighted method.

- It is very important to note that constituents' companies does not remain the same. Hence , it may be possible the stocks of the company constituting index at the time of index inspection , may not be aprt of index as on date and new companies stock may have replaced them.

CPI- Consumer Price Index/ Cost of living Index or Retail Price Index is the Index which measures the effect of change in prices of basket of goods and services on the purchasing power of specific class of consumer during any current period w.r.t to some base period.

WPI- Whole Sale Price Index - The WPI measures the relative changes in prices of commodities traded in wholesale market.

CLASSWORK SECTION

SIMPLE / UNWEIGHTED INDEX NUMBER :

1. From the following table by the method of relatives using Arithmetic mean the price Index number is

Commodity	Wheat	Milk	Fish	Sugar
Base Price	5	8	25	6
Current Price	7	10	32	12

a) 140.35 b) 148.25 c) 140.75 d) None of these.

2. From the following data

Commodities	Base year	Current year
A	25	55
B	30	45

Then index numbers from G. M. Method is :

a) 181.66 b) 185.25 c) 181.75 d) None of these.

WEIGHTED INDEX NUMBER :

3. From the following data for the 5 groups combined

Group	Weight	Index Number
Food	35	425
Cloth	15	235
Power & Fuel	20	215
Rent & Rates	8	115
Miscellaneous	22	150

The general Index number is

a) 270 b) 269.2 c) 268.5 d) 272.5

4. In calculating a certain cost of living index number the following weights were used. Food 15, Clothing 3, Rent 4, Fuel & Light 2, Miscellaneous 1. Calculate the index for the data when the average percentages rise in prices of items in the various groups over the base period were 32, 54, 47, 78 & 58 respectively.

a) 139.76 b) 141.99 c) 141.76 d) 139.87

BASE SHIFTING

5. Shift the base period of the following series of index numbers from 1978 to 1985:

Year	1982	1983	1984	1985	1986	1987	1988
Index No. [Base 1978 =100]	120	125	132	140	150	158	175

a) 85.71, 89.29, 100, 94.29, 107.14, 112.86, 125
b) 85.71, 89.29, 94.29, 100, 107.14, 112.86, 125
c) 85.71, 89.29, 101.98, 94.29, 107.14, 112.86, 125
d) 85, 89, 94, 100, 107, 112, 125

CHAIN BASED AND FIXED BASED INDEX

6. From the following data

Year	1992	1993	1994	1995	1996
Link Index	100	103	105	112	108

(Base 1992 = 100) for the years 1993–96. The construction of chain index is:

a) 103, 100.94, 107, 118.72 b) 103, 108.15, 121.13, 130.82
c) 107, 100.25, 104, 118.72 d) None of these.

DEARNESS ALLOWANCES/ EXTRA ALLOWANCES

7. Net Monthly income of an employee was ₹ 800 in 1980. The consumer price Index number was 160 in 1980. It is rises to 200 in 1984. If he has to be rightly compensated. The additional dearness allowance to be paid to the employee is :

a) ₹ 240 b) ₹ 275 c) ₹ 250 d) 200

MISCELLANEOUS SUMS

8. The price of a commodity increases from ₹ 5 per unit in 1990 to ₹ 7.50 per unit in 1995 and the quantity consumed decreases from 120 units in 1990 to 90 units in 1995. The price and quantity in 1995 are 150% and 75% respectively of the corresponding price and quantity in 1990. Therefore, the product of the price ratio and quantity ratio is:

a) 1.8 b) 1.125 c) 1.75 d) None of these.

THEORETICAL ASPECTS

9. ___________ play a very important part in the construction of index numbers.

a) weights b) classes c) estimations d) none

10. The __________ makes index numbers time-reversible.

a) A.M. b) G.M. c) H.M. d) none

11. The __________ of group indices given the General Index

a) H.M. b) G.M. c) A.M. d) none

12. Factor Reversal test is satisfied by

a) Fisher's Ideal Index b) Laspeyres Index
c) Paasches Index d) none

13. Laspeyre's formula does not satisfy

a) Factor Reversal Test b) Time Reversal Test
c) Circular Test d) all the above

14. $\dfrac{\text{Sum of all commodity prices in the current year} \times 100}{\text{Sum of all commodity prices in the base year}}$ is

(a) Relative Price Index (b) Simple Aggregative Price Index
(c) both (d) none

15. When the product of price index and the quantity index is equal to the corresponding value index then the test that holds is

(a) Unit Test (b) Time Reversal Test
(c) Factor Reversal Test (d) none holds

16. Fisher's Ideal Formula for calculating index numbers satisfies the ________ tests
    a) Unit Test   b) Factor Reversal Test
    c) both   d) none

17. If the index number of prices at a place in 1994 is 250 with 1984 as base year, then the prices have increased on average by
    a) 250%   b) 150%   c) 350%   d) None of these.

18. Theoretically, G.M. is the best average in the construction of index numbers but in practice, mostly the A.M. is used
    a) false   b) true   c) both   d) none

19. Time Reversal Test is represented by symbolically is :
    a) $P_{01} \times Q_{01} = 1$ b) $I_{01} \times I_{10} = 1$
    b) $I_{01} \times I_{12} \times I_{23} \times \ldots . I_{(n-1)n} \times I_{n0} = 1$ d) None of these.

5A PROBABILITY
Theory of Chance

Probability

- **Subjective**
  - It is influenced by personal belief, bias, attitude, etc and this is used in decision making management.
- **Objective**

Definitions

a) **Experiment or Random Experiment :** When an operation or series of operations are conducted under identical conditions it is called as experiment.

b) **Sample Space :** A set of all possible outcomes of a random experiment is called a sample space (S or U). Sample space may be finite or infinite.

c) **Event:** The outcome of an experiment is called an event.

d) **Elementary and Compound (or Composite) Events:** An event is said to be elementary, if it cannot be de-composed into simpler events. A composite event is an aggregate of several elementary events.

e) **Mutually Exclusive Events :** Events are said to be mutually exclusive when the occurrence of any one event excludes the occurrence of other or otherwise e.g. if a coin is tossed occurrence of head and tail are mutually exclusive events because of head will automatically exclude the occurrence of tail or vice versa.

f) **Equally likely events:** Events are said to be equality likely when they are equi-probable i.e. the event should occur with same chance of occurrence (None can be preferred over the other).

g) **Exhaustive events:** The events are said to be exhaustive when they include all possible outcomes. Events will necessarily occur.

h) **Independent Events:** Events are said to be independent of each other if happening or non happening of any one of them is not affected by and does not affect the happening of any one of others.

APPROACHES TO PROBABILITY		
Classical or Mathematical or a Priori	Empirical or Posteriori or Statistical	Axiomatic

1. Classical Definition of Probability

If a random experiment has "n" possible outcomes, which are mutually exclusive, exhaustive and equally likely and "m" of these are favourable to any event A, then the probability of the event A is defined as the ratio m/n, i.e.,

$$P(A) = \frac{m}{n} = \frac{\text{Favourable Outcomes}}{\text{Total Outcomes}}$$

Note1:

a) Probability as defined above will always lie between 0 and 1, both inclusive i.e., $0 \le P(A) \le 1$ **and** $P(A) \ge 0$.

b) If **P(A) = 0**, it means that event is impossible.

c) **P(A) = 1** signifies that event is certain or sure event.

Note2:

Complementary Probability

Let P(A) be the Probability of occurrence of event A.

Then $P(\bar{A}) / P(A^C) / P(A') = 1 - P(A)$ = Probability of non-occurrence of event A.

Note3:

a) $P(A) + P(A^C) = 1$, which implies that A and A^c are collectively exhaustive.

b) $P(A \cap A^C) = 0$, which implies that A and A^c are mutually exclusive.

Limitations Of Classical Probability

a. It fails if the no. of outcomes of an experiment. is very large $n \to$ infinite (∞).

b. It fails if the outcomes are not equally likely.

c. The definition holds if the possible events are known well in advance.

2. Empirical or posteriori or Statistical definition

If a random exp. is repeated large no. of times say n under identical conditions & let event A occurs m times then

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

3. Axiomatic definition

It is totally dependent on set theory

(i) $P(A) \geq 0$ for all $A \subseteq S$

(ii) $P(S) = 1$

(iii) If A & B are mutually exclusive events $P(A \cap B) = 0$

$P(A \cup B) = P(A) + P(B)$.

Total Number of Outcomes

To find the total number of outcomes, when an experiment is conducted "n" times in succession or with "n" objects only once.

Total outcomes = [No of outcomes in one experiment]^n

Where "n" = either number of objects or number of times the experiment gets repeated.

Examples:

a) 2 coins are tossed. Total outcomes = $2^2 = 4$

b) A coin is tossed five times. Total outcomes = $2^5 = 32$

c) 2 dice are rolled together. Total outcomes = $6^2 = 36$

Concepts of 'At least', 'At most' and 'At least one'

- **At least**

Let x = 0, 1, 2, 3, ..., n

Then, x is at least k, implies $x \geq k$, which implies that x = k, (k+1), (k+2), ... n

- **At most**

  x is at most k implies $x \le k$, which means x = 0, 1, 2, ..., k

- **At least One**

  x is at least one implies that $x \ge 1$, i.e., x = 1, 2, 3, ..., n

  Hence, P(at least 1) = 1 − P(none) = 1 − P(0)

Facts about Card

- A well shuffled deck of 52 cards are bi-colored -26 red and 26 black
- There are 4 suits or categories

  Clubs -13 Spades -13

  Hearts -13 Diamonds-13

- In each category , there is 1 king , 1 Queen

  1Jack or knave and 1 Ace(Ace implies 1)

  Therefore,

  King =4
  Queen =4
  Jack =4
  Ace =4

  King, Queen and Jack together are called Face cards.
  King Queen Jack and Ace are together called Honour cards.
  Total face Cards=4+4+4=12
  Honour Cards =4+4+4+4(K,Q,J,A)

Rolling of Dice

- If a die is rolled outcomes are 1, 2, 3, 4, 5, 6
- It two unbiased dice are rolled, outcomes = 6^2 = 36.

Sample Space

1,1	2,1	3,1	4,1	5,1	6,1
1,2	2,2	3,2	4,2	5,2	6,2
1,3	2,3	3,3,	4,3	5,3	6,3
1,4	2,4	3,4	4,4	5,4	6,4
1,5	2,5	3,5	4,5	5,5	6,5
1,6	2,6	3,6	4,6	5,6	6,6

Observations:

A. Sum of faces on two Dice and the no. of ways of getting sum

Sum	2	3	4	5	6	7	8	9	10	11	12
No. of ways	1	2	3	4	5	6	5	4	3	2	1

B. Distribution of sample space

Face	F = S	F > S	F < S	F Face on the First die
Cases	6	15	15	S Face on the Second die

No. of Children in a family

It treated same as in case of tossing of a coin.

For instance, if there are 3 children in a family, then outcomes = 2^3 = 8

(BBB) (BBG) (BGB) (BGG) (GBB) (GBG) (GGB) (GGG)

- **Leap Year**

  A leap year contains 52 weeks and 2 extra days. These two extra days can be either of the following out-comes:

  (M, T) (T, W) (W, Th) (Th, F) (F, Sat) (Sat, Sun) (Sun, M)

- Simple drawing of Balls from Bag – Using Combination Techniques

  A Bag contains m Red Balls and n Black Balls. Then if r balls are drawn, then it can be done in $^{m+n}C_r$ ways.

  Similarly use combination techniques to choose the required number of objects from the total objects given.

THEORM OF TOTAL PROBABILITY (Rule of Addition)

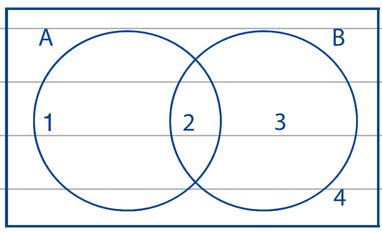
Statement: if A and B are two events, not mutually exclusive, then the probability of occurrence of at least any of the two events, A and B will be given by;

P (A $\cup$ B) or P (A+B) = P (A) + P (B) - P (A $\cap$ B) or P (AB)

Note 1: Union ($\cup$) implies "OR" $\Rightarrow$ Addition (+)

Note 2: Intersection ($\cap$) implies "AND" $\Rightarrow$ Multiplication (×)

Partitioning of events



1. A and B $\Rightarrow$ $(A \cap B)$ or AB
2. A and not B $\Rightarrow$ A but not B $\Rightarrow$ $A \cap B^c$ $\Rightarrow A - (A \cap B)$
3. B but not A $\Rightarrow$ B and not A $\Rightarrow$ $B \cap A^c$ $\Rightarrow B - (A \cap B)$
4. Neither A nor B $\Rightarrow$ A "not" and B "not" $\Rightarrow A^c \cap B^c$
5. $A^c = (3) + (4)$
6. $B^c = (1) + (4)$
7. $A^c \cup B^c = (1) + (3) + (4) = [2]^c = (A \cap B)^c$
8. $A^c \cap B^c = [4] = [1 + 2 + 3]^c = (A \cup B)^c$

Proof of P $(A \cup B)$:

$$
\begin{aligned}
P(A \cup B) &= P(1) + P(2) + P(3) \\
&= P(A \cap B^C) + P(A \cap B) + P(A^C \cap B) \\
&= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\
&= P(A) + P(B) - P(A \cap B)
\end{aligned}
$$

Hence proved

Note 1:

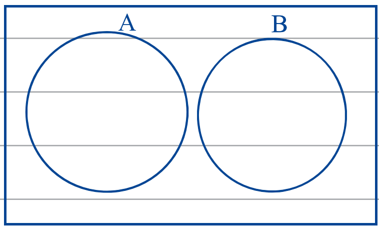
For 3 events, A, B and C, not mutually exclusive, +

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

Note 2:

When A and B are mutually exclusive, the two sets are disjoint and accordingly

$P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$



$(A \cap B) = \varnothing$

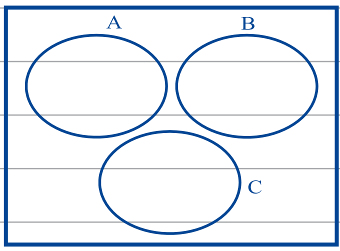
$P(A \cap B) = 0$

Note 3:

When 3 events A, B and C are mutually exclusive then

$P(A \cap B) = P(B \cap C) = P(A \cap C) = P(A \cap B \cap C) = 0$ and accordingly

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$



Note 4:

When 3 events A, B and C are mutually exclusive and collectively exhaustive then,

$P(A \cup B \cup C) = P(A) + P(B) + P(C) = 1$

Note 5:

Working Rules:

i. $P(A \cap B^c) = P(A) - P(A \cap B)$

ii. $P(A^c \cap B) = P(B) - P(A \cap B)$

iii. $P(A^c \cup B^c) = P(A \cap B)^c = 1 - P(A \cap B)$

iv. $P(A^c \cap B^c) = P(A \cup B)^c = 1 - P(A \cup B)$

v. $P(A^c \cup B) = P(A^c) + P(B) - P(A^c \cap B)$

vi. $P(A \cup B^c) = P(A) + P(B^c) - P(A \cap B^c)$

CONCEPT OF 'ODDS IN FAVOR' AND 'ODDS AGAINST'

- Odds in favor of an event is defined as 'the ratio of the favorable to the unfavorable cases and is denoted by u : v

  Where,

  U = favorable cases and

  V = unfavorable cases

  $\therefore\ P(A) = \frac{u}{u+v}$ and $P(A^C) = \frac{v}{u+v}$

- Odds against an event A is defined as 'the ratio of the unfavorable to the favorable cases and is given by v : u

  Where,

  U = favorable cases

  V = Unfavorable cases

  $\therefore\ P(A) = \frac{u}{u+v}$ and $P(A^C) = \frac{v}{u+v}$

THEOREM OF COMPOUND PROBABILITY (RULE OF MULTIPLICATION)

Statement:

If A and B are two events, not mutually independent, then the probability of joint or simultaneous occurrence of the two events A and B would be given by the product of the probability of event A and the conditional probability of event B assuming that, A has already occurred,

Symbolically, the fact is expressed as, $P(A \cap B) = P(A) \times P\left(\frac{B}{A}\right)$

Similarly product of the probability B and the conditional probability of event A assuming that, B has already occurred, is given by

$P(A \cap B) = P(B) \times P\left(\frac{A}{B}\right)$

Note 1:

- $P(A \cap B) = P(A) \times P\left(\frac{B}{A}\right)$

  $\therefore P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)}$

  Where, $P(A) \neq 0$ i.e, P(A) should not be an impossible event

- $P(A \cap B) = P(B) \times P\left(\frac{A}{B}\right)$

  $\therefore P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$

  Where, $P(B) \neq 0$ i.e, P(B) should not be an impossible event

- $P\left(\frac{A}{B^C}\right) = \frac{P(A \cap B^C)}{P(B^C)}$

  $P\left(\frac{A}{B^C}\right) = \frac{P(A) - P(A \cap B)}{1 - P(B)}$

  Where,
  $P(B) \neq 1$ i.e, P(B) is not a sure event

- $P\left(\frac{A^C}{B^C}\right) = \frac{P(A^C \cap B^C)}{P(B^C)} = \frac{P(A \cup B)^C}{1 - P(B)} = \frac{1 - P(A \cup B)}{1 - P(B)}$

- $P\left(\frac{A^C}{B}\right) = \frac{P(A^C \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)}$

- $P\left(\frac{B^C}{A^C}\right) = \frac{P(B^C \cap A^C)}{P(A^C)} = \frac{P(A \cup B)^C}{1 - P(A)} = \frac{1 - P(A \cup B)}{1 - P(A)}$

Note 2:

When the events A and B are independent, in such a case $P(A \cap B) = P(A) \times P(B)$

Note 3:

a. When the events A and B are independent, then,

$$P\left(\frac{B}{A}\right) = P(B)$$

Proof :

$$P\left(\frac{B}{A}\right) = P(B)$$

$$\Rightarrow P\left(\frac{A \cap B}{P(A)}\right) = P(B)$$

$$\Rightarrow P(A \cap B) = P(A) \times P(B)$$

Hence, proved

b. When the events A and B are independent, then,

$$P\left(\frac{A}{B}\right) = P(A)$$

Proof:

$$P\left(\frac{A}{B}\right) = P(A)$$

$$\Rightarrow \frac{P(A \cap B)}{P(B)} = P(A)$$

$$\Rightarrow P(A \cap B) = P(A) \times P(B)$$

Hence, proved

Note 4:

For three events, A, B and C which are not independent,

$$P(A \cap B \cap C) = P(A) \times P\left(\frac{B}{A}\right) \times P\left(\frac{C}{A \cap B}\right)$$

Note 5:

When 3 events, A and B and C are independent,

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$$

Note 6:

Two events A and B are,

i. Mutually exclusive, if $P(A \cap B) = 0$

ii. Independent, if $P(A \cap B) = P(A) \times P(B)$

iii. Equally likely, if $P(A) = P(B)$

iv. Exhaustive, if $P(A \cup B) = 1$

v. Mutually exclusive and exhaustive e, if $P(A \cup B) = P(A) + P(B) = 1$

Note 7:

Two events with non-zero probability cannot be simultaneously mutually exclusive and independent.

Note 8:

If two events A and B are independent, then

i. A^c and B^c are independent $\Rightarrow P(A^C \cap B^{C)} = P(A^C) \times P(B^C)$

ii. A and B^c are independent $\Rightarrow P(A \cap B^{C)} = P(A) \times P(B^C)$

iii. A^c and B are independent $\Rightarrow P(A^C \cap B) = P(A^C) \times P(B)$

Note 9:

If $A_1, A_2, A_3 ....A_n$ are n events, then the number of conditions to be satisfied for proving their mutual independence are $2^n - (n+1)$

CLASSWORK SECTION

Children in a Family

In a family of three children there is at least one girl. Find the probability that;

1. There are at least two girls.
   a) 4/7 b) 2/7 c) 2/8 d) 1/7

2. There is exactly 1 boy.
   a) 1/8 b) 2/7 c) 3/7 d) 1/7

Drawing of Balls from Bag

From a bag containing 7 white and 5 red balls, 4 balls are drawn at random. What is the chance that;

3. All are red.
   a) 5/495 b) 1/495 c) 3/495 d) None of these

4. Three white and one red.
   a) 165/495 b) 185/495 c) 175/495 d) 195/495

Addition Theorem

A number is selected at random from a set of first 120 natural numbers. What is the probability that it is divisible by:

5. 5 or 6
   a) 1/3 b) 1/4 c) 2/12 d) None of the above

Formula

If $P(A) = 1/4$, $P(B) = 2/5$, $P(A \cup B) = 1/2$. Find:

6. $P(A \cap B^c)$
   a) 3/20 b) 1/10 c) ¼ d) ½

7. $P(A^c \cap B^c)$

a) 3/20 b) 1/10 c) ¼ d) ½

8. $P(A^c/B^c)$

a) 4/10 b) 5/10 c) 6/10 d) None of the above

Independent Events

9. If for two independent events A and B, P(A U B) = 2/3 and P(A) = 2/5, what is P(B)?

a) 4/15 b) 4/9 c) 5/9 d) 7/15

A problem in Statistics is given to three students A, B and C whose respective chances of solving are 1/3, 1/4, 1/5. Find the probability that:

10. It is solved by at least 2 of them.

a) 2/6 b) 1/6 c) 5/6 d) None of these

Odds in Favour / Odds Against

11. The odds that a book will be favorably reviewed by three independent critics are 5 to 2, 4 to 3, and 3 to 4 respectively. What is the probability that majority of the critics reviewed the book favorably?

a) 225 / 343 b) 209 / 343 c) 391 / 400 d) 420 / 840

Bags and Balls – Important Cases

Case: 2 – Two bags are given, a bag is chosen at random, then ball(s) is/are drawn

A bag contains 5 red and 3 black balls and another bag contains 4 red and 5 black balls. A bag is selected at random and a ball is selected. Find the chance that:

12. It is red.

a) 77 / 177 b) 87 / 144 c) 97 / 854 d) 77 / 144

Case: 3 – Two bags are given, 1 ball is chosen from Bag 1 and transferred to Bag 2. Now a ball is drawn from Bag 2

There are two bags. The first contains 2 red and 1 white ball, whereas the 2nd bag contains 1 red and 3 white balls. One ball is taken out at random from the 1st bag and put into second bag. Then a ball is chosen at random from the second bag. What is the probability that;

13. The last ball is red.
 a) ½ b) 1/3 c) ¼ d) 1/5

Miscellaneous Cases

14. For a group of students, 30 %, 40% and 50% failed in Physics , Chemistry and at least one of the two subjects respectively. If an examinee is selected at random, what is the probability that he passed in Physics if it is known that he failed in Chemistry?
 a) 1/2 b) 1/3 c) 1/4 d) 1/6

15. Four digits 1, 2, 4 and 6 are selected at random to form a four digit number. What is the probability that the number so formed, would be divisible by 4?
 a) 1/2 b) 1/5 c) 1/4 d) 1/3

Theoretical Aspects

16. An experiment is known to be random if the results of the experiment
 a) Can not be predicted
 b) Can be predicted
 c) Can be split into further experiments
 d) Can be selected at random.

17. Which of the following pairs of events are mutually exclusive?
 a) A : The student reads in a school. B : He studies Philosophy.
 b) A : Raju was born in India. B : He is a fine Engineer.
 c) A : Ruma is 16 years old. B : She is a good singer.
 d) A : Peter is under 15 years of age. B : Peter is a voter of Kolkata.

18. If $P(A \cap B) = 0$, then the two events A and B are
    a) Mutually exclusive
    b) Exhaustive
    c) Equally likely
    d) Independent.

19. If for two events A and B, $P(AUB) = 1$, then A and B are
    a) Mutually exclusive events
    b) Equally likely events
    c) Exhaustive events
    d) Dependent events.

20. If an unbiased coin is tossed once, then the two events Head and Tail are
    a) Mutually exclusive
    b) Exhaustive
    c) Equally likely
    d) All these (a), (b) and (c).

21. If $P(A/B) = P(A)$, then
    a) A is independent of B
    b) B is independent of A
    c) B is dependent of A
    d) Both (a) and (b).

22. If two events A and B are independent, then
    a) A and the complement of B are independent
    b) B and the complement of A are independent
    c) Complements of A and B are independent
    d) All of these (a), (b) and (c).

23. If two events A and B are mutually exclusive, then
    a) They are always independent
    b) They may be independent
    c) They can not be independent
    d) They can not be equally likely.

24. If a coin is tossed twice, then the events 'occurrence of one head', 'occurrence of 2 heads' and 'occurrence of no head' are
    a) Independent
    b) Equally likely
    c) Not equally likely
    d) Both (a) and (b).

25. $P(B/A)$ is defined only when
    a) A is a sure event
    b) B is a sure event
    c) A is not an impossible event
    d) B is an impossible event.

26. For two events A and B, $P(A \cup B) = P(A) + P(B)$ only when
    a) A and B are equally likely events
    b) A and B are exhaustive events
    c) A and B are mutually independent
    d) A and B are mutually exclusive.

27. For any two events A and B,
    a) $P(A) + P(B) > P(A \cap B)$
    b) $P(A) + P(B) < P(A \cap B)$
    c) $P(A) + P(B) \geq P(A \cap B)$
    d) $P(A) \times P(B) \leq P(A \cap B)$

28. According to the statistical definition of probability, the probability of an event A is the
    a) limiting value of the ratio of the no. of times the event A occurs to the number of times the experiment is repeated
    b) the ratio of the frequency of the occurrences of A to the total frequency
    c) the ratio of the frequency of the occurrences of A to the non-occurrence of A
    d) the ratio of the favourable elementary events to A to the total number of elementary events.

29. If $P(A-B) = P(B-A)$, then the two events A and B satisfy the condition
    a) $P(A) = P(B)$.
    b) $P(A) + P(B) = 1$
    c) $P(A \cap B) = 0$
    d) $P(A \cap B) = 1$

5B RANDOM VARIABLE

Theory of Expectation)

A. RANDOM VARIABLES

Definition of Random Variables or Stochastic Variable

1. A variable whose value is determined by the outcome of a random experiment is called a random variable.

2. In other words, a random variable "x" is a real valued function defined on a sample space "S" of a random experiment such that for each point 'x' on the sample space f(x) = probability of the occurrence of the event represented by x.

3. Random Variables are also known as Chance Variables

   e.g. If we toss 3 coins then S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}.

   If 'X' denotes the number of heads obtained then 'x' assumes the following values with the corresponding probabilities.

X	0	1	2	3
P(x)	1/8	3/8	3/8	1/8

   These values of 'x' {0,1,2,3} are called the values of the random variables which are the outcomes of a random experiment.

4. Random Variables can be divided into the following two categories. They are
   a. Discrete Random Variable
   b. Continuous Random Variable

A. **Discrete Random Variable**

Definition : If a variable can assume only discrete set of values i.e. a finite set of values or countably infinite set of values then it is called a Discrete Random Variable. In other words, discrete random variable can assume only whole numbers. (0, 1, 2, 3..........) e.g. In a roll of a die the random variable x assumes values {1, 2, 3, 4, 5, 6}, these are discrete random variables.

B. **Continuous Random Variables**

Definition : If a random variable can assume an uncountably infinite number of values or all real numbers in a given interval is called Continuous Random Variable. E.g. height or weight of a person is an example of continuous random variable.

5. Concept of Probability Function of a Random Variable

A. For a discrete random variable, the probability function $f(x) = P(X = x_i)$ is called Probability Mass Function (p. m. f.) of a discrete random variable 'x' which satisfies the following two conditions (i) $f(x) \geq 0$ (ii) $\sum f(x) = 1$

B. If 'x' is a continuous random variable the probability function f(x) is called Probability Density Function (p. d. f.) which has the following two properties

(i) $f(x) \geq 0$

(ii) $\int_a^b f(x)dx = 1$ where $a \leq x \leq b$ is the range of 'x'. Since the continuous random variable can assume any real value, therefore Random Variable can be any real number.

C. For a Continuous Random Variable, the probability of occurrence of any specific value is 0 because for a continuous variable, probability are associated only with intervals of numbers.

B. MATHEMATICAL EXPECTATION OR EXPECTED VALUE OR MEAN

Definition of Mathematical Expectation or Expected value or Expectation of Random Variable "X"

Let $x_1, x_2, x_3 \ldots\ldots x_n$ be a set of n values of a variable "X" with the corresponding probabilities of occurrences $p_1, p_2, p_3, \ldots$pn then the mathematical expectation or Expectation or Expected value of random variable "X" is given by

$$E(x) = x_1p_1 + x_2p_2 + \ldots. + x_np_n = \sum_{i=1}^{n} x_i p_i$$

E.g. Calculation of Expectation of 'x' (where 'x' are the random variables generated as a result of throwing an unbiased die)

X	P	XP
1 (x_1)	1/6 (p_1)	1/6 (x_1p_1)
2 (x_2)	1/6 (p_2)	2/6 (x_2p_2)
3 (x_3)	1/6 (p_3)	3/6 (x_3p_3)
4 (x_4)	1/6 (p_4)	4/6 (x_4p_4)
5 (x_5)	1/6 (p_5)	5/6 (x_5p_5)
6 (x_6)	1/6 (p_6)	6/6 (x_6p_6)
		$\sum xp = 21/6 = (x_1p_1 + x_2p_2 + \ldots + x_6p_6)$

Therefore $E(x) = \sum xp = 21/6 = 3.5$ i.e. $\sum_{i=1}^{6} x_i p_i$

Properties of Mathematical Expectation

1. $E(x) = \bar{x}$ = mean of random variable 'x'.

2. E(x) can assume any real number since 'x' can assume any real value.

3. If all the value of the random variable 'x' are equal then E(x) will be equal to constant. i.e. E(c) = c

4. $E(x \pm y) = E(x) \pm E(y)$

5. $E(xy) = E(x)\,E(y)$ provided x and y are independent

6. $E(cx) = c.\,E(x)$ e.g. $E(5x) = 5\,E(x)$

7. $E(a \pm bx) = a \pm bE(x)$

e.g. given that E(x) = 5 find E(2 -3x)?

$E(2 - 3x) = 2 - 3\,E(x)$

$= 2 - 3\,(5) = -13$

8. $E(ax \pm by) = aE(x) \pm bE(y)$

e.g. given that $E(x) = 3$ and $E(y) = 4$ find $E(7x + 9y)$?

$E(7x + 9y) = 7E(x) + 9E(y) = 7(3) + 9(4) = 21 + 36 = 57$

9. $E(x - \bar{x}) = 0$

Proof

$E(x - \bar{x}) = E(x) - E(\bar{x}) = \bar{x} - \bar{x}$ (since $E(x) = \bar{x}$ = mean is constant) $= 0$

10. Variance and Standard Deviation of a Random Variable X

A. **Definition :** Variance of a random variable X is defined as the Arithmetic Mean of the Square of Deviations taken about Arithmetic Mean i.e.

B. Symbolically $\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{n}$

$\Rightarrow \sigma^2$ = A. M of $(x - \bar{x})^2$

$\Rightarrow \sigma^2$ = expectation of $(x - \bar{x})^2$ [Since expectation = A. M]

$\Rightarrow \sigma^2 = E(x - \bar{x})^2$ or $E(x - \mu)^2$ or $E(X - \{E(x)\})^2$

Where $\bar{x} = \mu = E(X)$ = Mean of the random variable X

C. σ^2 or variance of x is also denoted by Var (X) or V(X) and $V(X) = E(x - \bar{x})^2 = E(x^2) - [E(x)]^2$

Proof : $E(x - \bar{x})^2 = E(X^2 - 2\bar{x} + \bar{x}^2) = E(x^2) - \bar{x}\,E(X) + E(\bar{x}^2)$

$= E(x^2) - 2\bar{x}.\bar{x} + \bar{x}^2$ $\quad \left(\because \text{ i. } E(X) = \bar{x} \text{ and} \quad \text{ii. } E(\bar{x}) = \bar{x} = \text{constant} \right)$

$= E(x^2) - 2\bar{x}^2 + \bar{x}^2$

$= E(x^2) - \bar{x}^2 = E(x^2) - [E(x)]^2$ - **(Proved)**

D. Thus $V(X) = E(x^2) - [E(x)]^2 = \Sigma x^2p - (\Sigma xp)^2$

E. Standard Deviation of x i.e. S.D. (X) $= \sqrt{\text{var}(x)} = \sqrt{E(X^2) - [E(X)]^2} = \sqrt{\Sigma x^2p - (\Sigma xp)^2}$

11. Properties of Variance and Standard Deviation

1. When all the value of the variable are equal :

   Variance = 0 and S.D. = 0 i.e. V(C) = 0 where C is any constant.

   e.g. V(2) = 0

2. $Var(aX) = a^2V(X)$

   e.g. Given V(X) = 3

   Calculate V(3X)

   Solution :

   V(3X) = 9V(X) = 9(3) = 27

3. $Var(a \pm bx) = b^2\ V(X)$ $[\because Var(a) = 0]$

   e.g. Given V(X) = 2

   Find : (i) V(3 + 2x), (ii) V(2 − 3x)

   Solution : (i) (3 + 2x) = 4V(X) = 4 (2) = 8, (ii) V(2 − 3x) = 9V(X) = 9 (2) = 18

4. $Var\ (aX \pm bY) = a^2\ V(X) + b^2\ V(Y)$

   e.g. Given V(X) = 4 and V(Y) = 9 Find

   Find : (i) V(7X + 4Y), (ii) V(2X + 3Y)

   Solution:

   (i) V(7X − 4Y)= 49V(X) + 16V(Y)
   = 49 × 4 + 16 × 9
   = 196 + 144 = 340

   (ii) V(2X + 3Y) = 4V(X)+ 9V(Y)
   = 4 × 4 + 9 × 9
   = 16 + 81 = 97

CONCEPT OF UNIFORM DISTRIBUTION (DISCRETE VARIABLE)

1. If a discrete random variable ‘x’ assumes n possible values namely x_1 x_2, ...x_n with equal probabilities, then the probability of its taking any particular value is always constant and is equal to (1/n). The p.m.f (Probability Mass Function) of such distribution is given by **f(x) = 1/n** where x = x_1 x_2, ... x_n. These distributions are known as Uniform Distribution because the probability is uniform for all values of x.

   e.g. Probability Distribution of the no. of points in a throw of a die.

x	1	2	3	4	5	6
p	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

2. Mean of Uniform Distribution is : $\frac{n+1}{2}$ and variance of uniform distribution is $\frac{n^2-1}{12}$

CLASSWORK SECTION

Theoretical Aspect

1. When X is a continuous function, f(X) is called:
   a) Probability mass function
   b) Probability density function
   c) Both a) and b)
   d) None of the above

2. If P(a) = 0, P(b) = 1/3, P(c) = 2/3, then s = a, b, c is a probability space.
   a) True   b) False
   c) Both true and false   d) None of the above

3. For a probability distribution, _________ is expected value of x.
   a) Median   b) Mean
   c) Mode   d) None of the above

Probability Mass Function (P.M.F)

4. Let X be a random variable assuming values -3, 6 and 9 with probabilities 1/6, ½ and 1/3 respectively. Then find the value of E(X), $E(X^2)$ and $E(2X+1)^2$
   a) 5.5, 46.5, 209   b) 6.5, 45.5, 207
   c) 6, 40, 200   d) None of these

5. A player tosses three fair coins. He wins Rs. 12 if three tails occur, Rs. 7 if two tails occur and Rs. 2 if only one tail occurs. If the game is fair, how much should he win or lose in case no tail occurs?
   a) Loss of Rs. 39   b) Income of Rs. 39
   c) Neither Income nor Loss   d) None of the above

6. A man draws 2 balls from a bag containing 3 white and 6 black balls. If he is to receive Rs. 14 for every white ball and Rs. 7 for every black ball; what is his expectation?
   a) 18.67   b) 19.25   c) 20.25   d) 25.19

7. A number is chosen at random from the set 1, 2, 3, ..., 100 and another number is chosen at random from the set 1, 2, 3 ..., 50. What is the expected value of their product?

a) 5151 b) 5151/4 c) 5151/2 d) None of the above

A random variable x has the following probability distribution:

X:	0	1	2	3	4	5	6	7
P(x):	0	2k	3k	k	2k	k^2	$7k^2$	$2k^2+k$

8. What is the value of k?

a) 1 / 2 b) 1 / 8 c) 1 / 9 d) 1 / 10

9. What is the value of P(x < 6)?

a) 0.19 b) 0.80 c) 0.81 d) 0.91

10. What is the value of P(0 < x < 5)?

a) 0.19 b) 0.29 c) 0.80 d) 0.91

A probability mass function for a random variable x is given as:

$$\left.\begin{aligned} f(x) &= k, x = 1,2,3,4,5,6 \\ &= 0, \quad \text{elsewhere} \end{aligned}\right\}$$

11. The expected value of sum of points on n unbiased dice is:

a) $\frac{3n}{2}$ b) $\frac{5n}{2}$

c) $\frac{7n}{2}$ d) $\frac{8n}{3}$

UNIFORM DISTRIBUTION

12. The probability distribution whose frequency function f(x) = 1/n, $x = x_1, x_2, \ldots, x_n$ is known as:

a) Binomial distribution b) Poisson distribution
c) Normal distribution d) Uniform distribution

13. If a discrete random variable x follows uniform distribution and assumes only the values 8, 9, 11, 15, 18, 20. Then find $P(|x - 14| < 5)$.

a) 1 b) ½ c) 2/3 d) 1/3

6 THEORETICAL DISTRIBUTION

THEORETICAL DISTRIBUTION

(Exist in theory as well as real life)

1. Theoretical Distribution is a distribution where the values of a variable are distributed according to some definite mathematical laws.

2. In other words, Theoretical Distributions are mathematical models; where the frequencies/probabilities are calculated by mathematical computation.

3. Theoretical Distribution are also called as Expected Variance Distribution or Frequency Distribution

THEORETICAL DISTRIBUTION

- DISCRETE
  - BINOMIAL Distribution
  - POISSON Distribution
- CONTINUOUS
  - NORMAL OR GAUSSIAN Distribution

A. Binomial Distribution (Bernoulli Distribution)

1. The probabilities of 'x' number of success or the p.m.f (Probability Mass Function) of a Binomial Distribution is given by :

$P(x) = n_{c_x} p^x q^{n-x}$

where, p = probability of success

q = probability of failure $=(1-p)$

'x'= no. of success

And $(n-x)$ = no. of failures

Note 1: Sum of powers of p and q will always add up to 'n' irrespective of no. of success.

Note 2: There are $(n + 1)$ possible value of 'x' i.e. $x = \{0,1,2,3,\ldots,n\}$

2. This distribution is a discrete probability Distribution where the variable 'x' can assume only discrete values i.e. $x = 0, 1, 2, 3, \ldots\ldots n$

3. This distribution is derived from a special type of random experiment known as Bernoulli Experiment or Bernoulli Trials, which has the following characteristics

   (i) Each trial must be associated with two mutually exclusive & exhaustive outcomes – SUCCESS and FAILURE. Usually the probability of success is denoted by 'p' and that of the failure by 'q' where $q = 1 - p$ and therefore $p + q = 1$.

   (ii) The trials must be independent under identical conditions.

   (iii) The number of trial must be finite (countably finite).

   (iv) Probability of success and failure remains unchanged throughout the process.

Note 1 : A 'trial' is an attempt to produce outcomes which is neither sure nor impossible in nature.

Note 2 : The conditions mentioned may also be treated as the conditions for Binomial Distributions.

4. Characteristics or Properties of Binomial Distribution

   (i) It is a bi parametric distribution i.e. it has two parameters n & p where

   n = no. of trials

   p = probability of success.

   (ii) Mean of distribution is np.

   (iii) Variance = npq

   (iv) Mean is greater than variance always i.e. $np > npq$.

   (v) SD = $\sqrt{npq}$

   (vi) Maximum variance is equal to (n/4)

   (vii) Binomial Distribution may be Symmetrical or Asymmetrical (i.e. skewed) where $q > p$; i.e. $P > ½$ its positively skewed and when $q < p$ i.e. $P > ½$ its negatively skewed.

   When $q = p = 0.5$ skewness is equal to zero. In such a case, the distribution is said to be symmetrical.

(viii) Binomial Distribution may be Uni-Modal or Bi-Modal depending on the values of the parameters n & p.

Case I : When (n + 1).p is not an integer the distribution is uni-modal and the greatest integer contained in (n+1) p is the value of the mode.

E.g. n = 6; p = 1/3; find modal value.

Solution : (n + 1)p = (6 + 1) x 0.3

= 7 x 0.3 = 2.1 which is not an integer. Hence the given distribution is unimodal and the value of mode is equal to 2 (Greatest integer integral value in 2.1)

Case II: When (n + 1)p is an integer; the distribution is bi-modal and the modal values are (n+1)p and (n+1)p − 1 respectively.

E.g. n = 7 and p = 0.5; find mode or modes.

Solution : (n +1)p = (7 + 1)p

= 8(0.5)

= 4. Which is an integer.

Hence the two modes are :4 & (4 -1) =3

(ix) Additive Property of Binomial Distribution: If 'x' and 'y' are two independent binomial variates with parameters(n_1,p) and (n_2,p) respectively,then x + y will also follow a binomial distribution with parameters {(n_1 + n_2), p} Symbolically the fact is expressed as follows:

$X \sim B(n_1, p)$

$Y \sim B(n_2, p)$

$X + Y \sim B(n_1 + n_2, p)$

(x) The method applied for fitting a binomial distribution to a given set of data is called "Method of Moments".

5. The distribution is called Binomial as the probabilities can be obtain by deferent terms of the expansion of Binomial series $(q+p)^n$

CLASSWORK SECTION

1. If in a Binomial distribution mean 20; S.D. = 4, then p is equal to:
   a) 1/5 b) 2/5 c) 3/5 d) 4/5

2. Mean =10, SD= $\sqrt{5}$, Mode=
   a) 10 b) 12 c) 9 d) 8

3. X is binomial variable with n = 20, what is the mean of X if it is known that x is symmetric?
   a) 5 b) 10 c) 2 d) 8

4. What is the probability of making 3 correct guesses in 5 True – False answer type questions?
   a) 0.3125 b) 0.5676 c) 0.6875 d) 0.4325

6 coins are tossed. Find the probability of getting

5. The probability that a student is not a swimmer is 4/5, then the probability that out of five students four are swimmers is
   a) $\left(\frac{4}{5}\right)^4$ b) ${}^5C_1\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^4$ c) ${}^5C_4\left(\frac{4}{5}\right)^1\left(\frac{1}{5}\right)^4$ d) None of these

6. At least 3 successes.
   a) 80 / 243 b) 192 / 243 d) 77 / 243 d) None of the above

A man takes a step forward with a probability 0.6 and a step backward with a probability of 0.4. Find the probability that at the end of 11 steps, the man is:

7. If x and y are 2 independent binomial variable with parameters 6 and ½, 4 and ½ respectively, what is P(x + y ≥ 1)?
   a) 1023/1024 b) 1056/1923
   c) 1234/2678 d) None of the above

8. Assuming that one-third of the population is tea drinkers and each of 1000 enumerators takes a sample of 8 individuals to find out whether they are tea drinkers or not, how many enumerators are expected to report that five or more people are tea drinkers?
   a) 100 b) 95 c) 88 d) 90

Calculation of Parameters

9. A binomial random variable x satisfies the relation 9P(x = 4) = P(x =2) when n = 6. Find the value of the parameter 'P'?
   a) 1 / 2 b) 1 / 3 c) 1 / 4 d) 1 / 5

Theoretical Aspect

10. Binomial distribution is a:
    a) Discrete Probability Distribution
    b) Continuous Probability Distribution
    c) Both a) and b) above
    d) Neither a) nor b) above

11. The important characteristic(s) of Bernoulli trials is:
    a) Trials are independent
    b) Each trial is associated with just two possible outcomes.
    c) Trials are infinite
    d) Both a) and b) above

12. The mean of binomial distribution is :
    a) Always more than its variance
    b) Always equal to its standard deviation
    c) Always less than its variance
    d) Always equal to its variance

13. The maximum value of the variance of a Binomial distribution with parameters and pis :
    a) $\frac{n}{p}$ b) $\frac{n}{3}$ c) $\frac{n}{4}$ d) $\frac{n}{2}$

14. For a binomial distribution, there may be

a) one mode b) two mode c) zero mode d) (a) or (b)

15. For n independent trials in Binomial distribution, the sum of the powers of p and q is always n,whatever be the number of successes.

a) True b) False

c) both of a) and b) above d) None of the above

16. For a binomial distribution if variance = mean/2, then the values of n and p will be

a) 1 and 1/2 b) 2 and 1/2 c) 3 and ½ d) Any value and 1/2

Theory Answer Key

10	a	11	d	12	a	13	c	14	d
15	b	16	a						

B. POISSON DISTRIBUTION

1. The probability of ‘x’ no. of success or the p.m.f (Probability Mass Function) of a Poisson Distribution is given by

$$P(x) = e^{-m} \cdot \frac{m^x}{x!} \text{ or } e^{-\lambda} \cdot \frac{\lambda^x}{x!} \quad (\lambda = m)$$

where x = desired no. of success.

$e \cong 2{\cdot}71828$

Note1: (λ = m) Mean = variance = parameter of the distribution

Note2: e^{-m} or $e^{-\lambda}$ is a constant and the value of which can be obtained from the table.

Note3: When the parameter ‘m’ is not provided but n and p are provided we shall use m = np for evaluating the parameter.

2. It is a discrete probability distribution where the variable ‘x’ can assume values ‘x’= 0, 1, 2, 3,……∞.

3. This distribution is a limiting case of Binomial Distribution when
   (i) $n \to \infty$ (i.e. no. of trials become very large)
   (ii) $p \to 0$, (i.e. probability of success is very small)
   (iii) $q \to 1$, (i.e. probability of failure is very high)
   (iv) np is finite and constant which is denoted by ‘m’ i.e. np = m or λ

4. Some examples of Poisson Distribution:
   (i) No. of telephones calls per minute at a switch board
   (ii) The no. of printing mistake per page in a large text.
   (iii) The no. of cars passing a certain point in 1 minute
   (iv) The emission of radio active (alpha) particles.

5. The conditions under which the Poisson Distribution is used or the condition for Poisson Model are as follows:
   (i) The probability of having success in a very small time interval (t, t + dt) is K. dt (where k > 0 and is constant)
   In other words, probability of success in a very small time interval is directly proportional to time internal dt.
   (ii) The probability of having more than one success in this time interval is very low.
   (iii) Statistical independence is assumed i.e. the probability of having success in this time interval is independent of time ‘t’ as well as of the earlier success.

6. Poisson Distribution is also known as "Distribution of Improbable Events" or "Distribution of Rare Events".

7. **Characteristic or Properties of Poisson Distribution.**

(i) Poisson Distribution is uniparametric i.e. it has only one parameter 'm' or 'λ'

(ii) Mean of distribution = m

(iii) Variance = m

(iv) In poisson distribution mean = variance and hence they are always positive

(v) SD = $\sqrt{m}$

(vi) Since 'm' is always positive Poisson Distribution is always positively skewed.

(vii) The distribution can be either unimodal or bimodal depending on values of m.

Case I : When 'm' is not an integer then the distribution is uni-modal and the value of the mode will be highest integral value contained in 'm'.
E.g. m = 5.6 then modal value is 5 (greatest integer contained in 5.6)

Case II: When 'm' is an integer; the distribution is bimodal and the modal values are m, m − 1
E.g . if 'm' = 4 (an integer, hence the distribution is bimodal and the modes are 4 and 4 − 1 i.e. 4 and 3)

(viii) Additive Property of Poisson Distribution: If 'x' and 'y' are two independent Poisson Variates with parameters(m1) and (m_2) respectively then (x + y) will also follow a Poisson Distribution with parameter ($m_1 + m_2$). Symbolically the fact is expressed as follows: $X \sim P(m_1)$, $Y \sim P(m_2)$
$X + Y \sim P(m_1 + m_2)$ provided x and y are independent

CLASSWORK SECTION

1. In a Poisson Distribution P(X = 0) = P(X = 1) = k, the value of "k" is:
   a) 1   b) $\frac{1}{e}$ c) e^2 d) $\frac{1}{\sqrt{e}}$

2. If x is Poisson variety with a parameter 4 find the Mode of the Distribution?
   a) 4,2   b) 4,3   c) 4,4   d) None

Between 4 and 5 PM, the average number of phone calls per minute coming into the switchboard of the company is 3. Find the probability that in one particular minute there will be: (Given e^{-3} = 0.0498)

3. Exactly 2 phone calls
   a) 0.1422   b) 0.2214   b) 0.2251   d) 0.2241

It is found that the number of accidents occurring in a factory follows Poisson distribution with a mean of 2 accidents per week. (Given e^{-2} = 0.1353)

4. A radioactive source emits on the average 2.5 particles per second. Calculate that 2 or more particles will be emitted in an interval of 4 seconds.
   a) $11e^{-10}$ b) $1-10e^{-10}$ c) $1-11e^{-10}$ d) None of the above

5. A renowned hospital usually admits 200 patients every day. One per cent patients, on an average, require special room facilities. On one particulars morning, it was found that only one special room is available. What is the probability that more than 3 patients would require special room faculties?
   a) 0.1428   b) 0.1732   c) 0.2235   d) 0.3450

Binomial Approximation to Poisson Distribution

Experience has shown that, as the average, 2% of the airline's flights suffer a minor equipment failure in an aircraft. Estimate the probability that the number of minor equipment failures in the next 50 flights will be(e^{-1}=.3679)

6. In a company manufacturing toys, it is found that 1 in 500 is defective. Find the probability that there will be at the most two defectives in a sample of 2000 units. [Given e^{-4} = 0.0183]

a) 0.2597 b) 0.3549 c) 0.2549 d) 0.2379

Miscellaneous Problems

7. A car hire firm has 2 cars which is hired out every day. The number of demand per day for a car follows Poisson distribution with mean 1.20. What is the proportion of days on which some demand is refused?
(Given $e^{1.20}$ = 3.32)

a) 0.25 b) 0.3012 c) 0.12 d) 0.03

Theoretical Aspects

8. Which one is uni-parametric distribution?

a) Normal Distribution b) Poisson Distribution
c) Hypergeometric Distribution d) Binomial Distribution

9. ____________ Distribution is a limiting case of Binomial distribution.

a) Normal Distribution b) Poisson Distribution
c) Chi-Square Distribution d) (a) & (b) both

10. Poisson distribution may be

a) Bimodal b) Uni modal
c) Multi Modal d) Either a) or b) above and not c)

11. For a Poisson distribution

a) Standard Deviation and Variance are equal.
b) Mean and Variance are equal.
c) Mean and Standard Deviation are equal.
d) Both a) and b) above

12. In Poisson Distribution, probability of success is very close

a) 1 b) 0.8 c) 0 d) None of the above

13. Poisson distribution is

a) Always negatively skewed

b) Always positively skewed

c) Always symmetric

d) Symmetric only when m = 2

Theoretical Aspect Answer Key

8	B	9	D	10	D
11	B	12	C	13	B

C. NORMAL OR GAUSSIAN DISTRIBUTION

1. It is a continuos probability distribution where the variable 'X' can assume any value between $-\infty$ to $+\infty$.

2. The Probability Density Function of a Normal Distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ (-\infty < x < \infty)$$

where μ = mean

σ = Standard Division

Note 1 : μ and σ are the two parameters of Normal Distribution and hence it is bi-parametric in nature.

Note 2 : $\pi \cong 3.1416$ and $e \cong 2.71828$ which are constant.

3. Replacing $\frac{x-\mu}{\sigma}$ by 'z' we obtain another distribution called Standard Normal Distribution with mean 0 and S.D. 1 and is given by the density function

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \ (-\infty < z < \infty)$$

Note1 : $N(\mu, \sigma^2)$ implies Normal Distribution with μ (mean) and σ^2 (variance)

Note2 : N(0,1) implies Standard Normal Distribution with Mean = 0 and S.D. = 1.

Note3 : 'z' is called Standard Normal Variate or Variable.

4. Caculation of Mean & S.D. of Z

(i) Calculation of Mean of Z

$z = \dfrac{x-\mu}{\sigma}$

$\therefore$ Mean of z = E(z)

$= E\left(\dfrac{x-\mu}{\sigma}\right)$

$= \dfrac{1}{\sigma}[E(x) - E(\mu)]$

$= \dfrac{1}{\sigma}(\mu - \mu) = \dfrac{1}{\sigma} \times 0 = 0$

(ii) Calculation of Variance of z

$Var(z) = E(z^2) - [E(z)]^2$

$= E(z^2) - 0^2$ [$\because E(z) = 0$]

$= E(z^2)$

$= E\left(\dfrac{x-\mu}{\sigma}\right)^2$

$= \dfrac{1}{\sigma^2} E(x-\mu)^2$

$= \dfrac{1}{\sigma^2} \times \sigma^2$

$= 1$

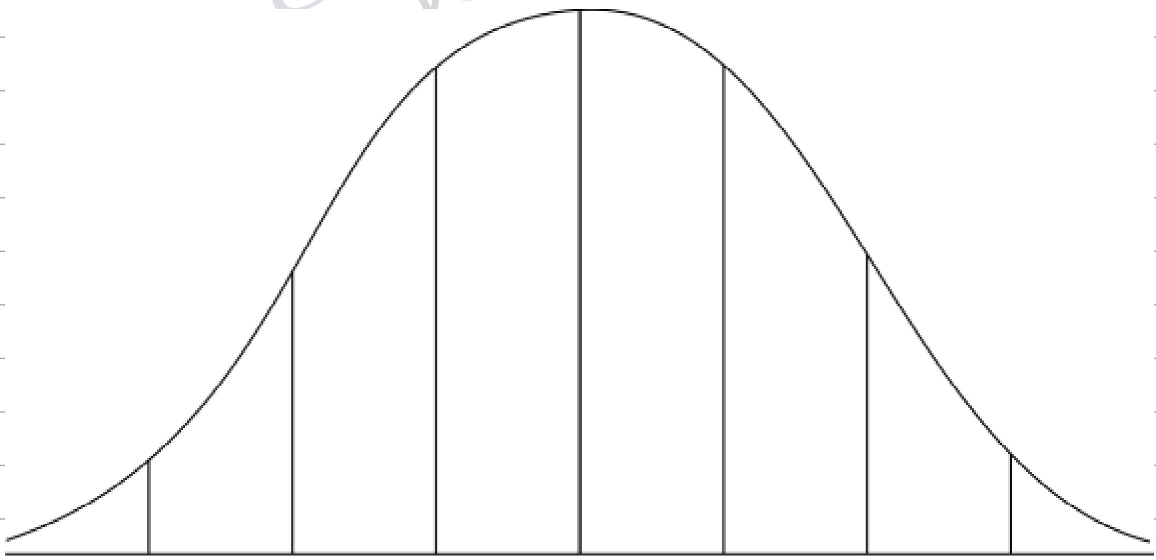
Var (z) = 1

hence S.D. (z) =1

$\boxed{\begin{array}{l}\because E(x-\bar{x})^2 \\ = E(x-\mu)^2 = \sigma^2\end{array}}$

5. The probability of success under Normal Distribution in calculated by evaluating the area under a curve called Normal Frequency curve which in shown in the following diagram

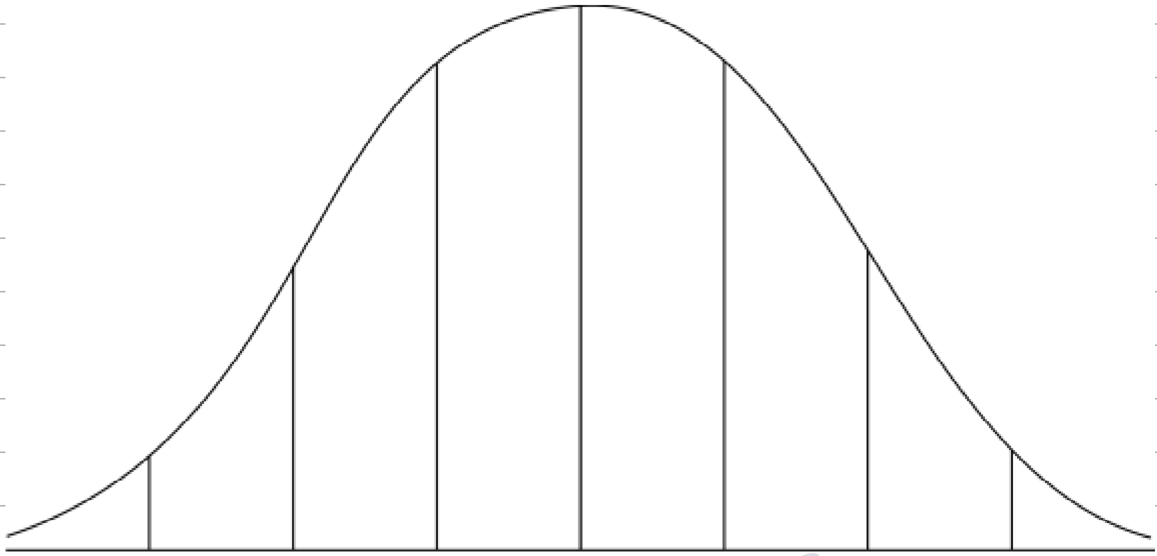
Normal Curve



$x = \mu - 3\sigma$ $x = \mu - 2\sigma$ $x = \mu - \sigma$ $x = \mu$ $x = \mu + \sigma$ $x = \mu + 2\sigma$ $x = \mu + 3\sigma$

Mean = Median = Mode

Standard Normal Curve



Mean = Median = Mode = 0

6. CONVERSION OF X VALUES FROM NORMAL FREQUENCY CURVE TO STANDARD NORMAL CURVE VALUES (Z - VALUES)

1. $x = \mu$ or, $X - \mu = 0$ or, $\frac{x-\mu}{\sigma} = \frac{0}{\sigma}$ $Z = 0$

2. $x = \mu + \sigma$ $\quad$ $x = \mu - \sigma$

   $\frac{x-\mu}{\sigma} = 1$ $\quad$ $\frac{x-\mu}{\sigma} = -1$

   $Z = 1$ $\quad$ $Z = -1$

3. $x = \mu + 2\sigma$ $\quad$ $x = \mu - 2\sigma$

   $\frac{x-\mu}{\sigma} = 2$ $\quad$ $\frac{x-\mu}{\sigma} = -2$

   $Z = 2$ $\quad$ $Z = -2$

4. $x = \mu + 3\sigma$ $\quad$ $x = \mu - 3\sigma$

   $\frac{x-\mu}{\sigma} = 3$ $\quad$ $\frac{x-\mu}{\sigma} = -3$

   $Z = 3$ $\quad$ $Z = -3$

PROPERTIES OF NORMAL CURVE AND NORMAL DISTRIBUTION

1. It is a bell shaped curve symmetrical about the line $x = \mu$ and assymptotic to the horizontal axis (x = axis)

2. The two tails extend upto infinity at both the ends.

3. As the distance from the mean increases, The curve comes closer to the horizontal axis (x = axis)

4. The curve has a single peak at $x = \mu$.

5. The two points of inflection of the normal curve are at $x = \mu - \sigma$ and $x = \mu + \sigma$ respectively where the normal curve changes its curvature.

6. The same points of inflection under standard normal curve are at z = – 1 and z = 1.

7. It is a continous prob. distribution where $-\infty < \chi < \infty$

8. The distribution has two parameters μ and σ . Where μ = mean σ = standard deviation. Hence normal is bi-parametric distribution.

9. The normal curve has a single peak. Hence it is unimodal and mean. Median and mode coincide. at $x = \mu$.

10. The maximum ordinate (i.e. y) lies at $x = \mu$.

11. The distribution being symmetrical,
    i) Mean = Median = Mode
    ii) Skewness = 0

12. The two Quartiles are $Q_1 = \mu - .675\sigma$ (Lower Quartile)
    And $Q_3 = \mu + .675\sigma$ (Upper Quartile)

13. Quartile Deviation (Q. D.)

$$Q.D. = \frac{(Q_3 - Q_1)}{2},\ = \frac{(\mu + .675\sigma) - (\mu - .675\sigma)}{2} = \frac{\mu + .675\sigma - \mu + .675\sigma}{2},\ = \frac{2 \times .675\sigma}{2},\ = .675\sigma = \frac{2}{3}\sigma$$

(Approximately)

14. Mean Deviation (M. D.) = 0.8σ = $\frac{4}{5}$ σ (Approximately)

15. QD : MD : SD = 10 : 12 : 15

16. (i) The total area under the Normal or Standard Normal Curve = 1 ($\because$ Total Probability = 1), Symbolically,

(i) $\int_{-\infty}^{+\infty} f(x)dx = 1$ or (ii) $\int_{-\infty}^{+\infty} f(z)dz = 1$

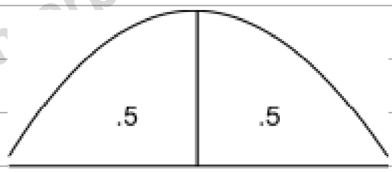
(ii) $f(x) \geq 0$ for all X

17. The curve being Symmetrical,

$x = \mu$ **divides curve into two equal halves**

such that (Area between – ∞ to μ)

= (Area between μ to + ∞) = 0.5

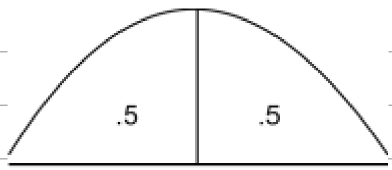


$X = \mu$

18. Similarly, under standard normal curve,

(area between – ∞ to z = 0)

= (area between z = 0 to z = + ∞) = 0.5



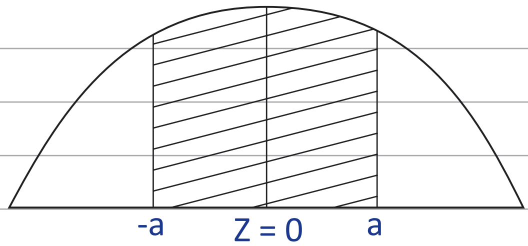
$Z = 0$

19. Symbilically

i) $P(-\infty < X \leq \mu) = P(\mu \leq X < +\infty) = 0.5$

ii) $P(-\infty < Z \leq 0) = P(0 \leq Z < +\infty) = 0.5$

20. The curve being symmetrical area of portions cut off from right and left of $X = \mu$ (or z = 0) are equal.

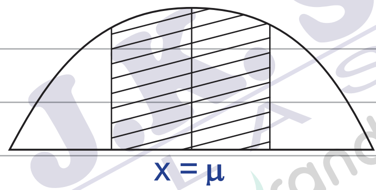


Symbolically, $P(-a \le Z \le 0) = P(0 \le Z \le a)$.

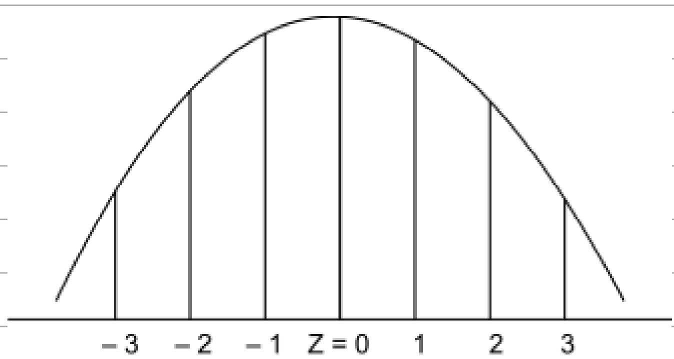
Note : Here "Area" implies "Probability"

21. The probability that a normal variate Z will take a value less than or equal to a particular value (say Z = K) will be denoted by $\Phi(K) = P(Z \le K)$

Note : The probability of success is calculated by evaluating the areas from the standard normal curve, and the areas obtained from normal table.



22. % Distribution of areas under Normal Curve / Standard Normal Curve



C-I

$P(-1 \le Z \le 0) = .3413$,

$P(0 \le Z \le 1) = .3413$.

$\therefore P(-1 \le Z \le 1) = .6826$.

$\therefore$ 68.26% of total area lies between Z = – 1 and Z = + 1 or $X = \mu - \sigma$ and $Z = \mu + \sigma$

C-II

$P(-2 \le Z \le 0) = .4772.$

$P(0 \le Z \le 2) = .4772.$

$\therefore P(-2 \le Z \le 2) = .9544.$

$\therefore$ 95.44% of total area lies between $Z = -2$ and $Z = +2$ or $X = \mu - 2\sigma$ and $X = \mu + 2\sigma$

C-III

$P(-3 \le Z \le 0) = .4987.$

$P(0 \le Z \le 3) = .4987.$

$\therefore P(-3 \le Z \le 3) = .9974.$

$\therefore$ 99.74% of total area lies between $Z = -3$ and $Z = +3$ or $X = \mu - 3\sigma$ and $X = \mu + 3\sigma$

23. **Additive Property of Normal Distribution**

If X & Y are independent normal variates with means μ_1 & μ_2 and standard deviation σ_1 & σ_2 respectively, then Z = X + Y will also follow a Normal Distribution with mean = $(\mu_1 + \mu_2)$ and S.D. = $\sqrt{\sigma_1^2 + \sigma_2^2}$

symbolically, $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $Z = X + Y \sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$

24. In continuous probability Distribution, Probability is to be assigned to intervals and not to individual values and accordingly the Probability that a Random Variable X will take any specific value will be "0" i.e. $P(X = C) = 0$ when Distribution is continuous.

25. **Concept of Cumulative Distribution Function (C. D. F.)**

Cumulative Distribution Function (C. D. F.) is defined as the Probability that a Random Variable X takes a value less than or equal to A specified value x and is denoted by F(X)

$\therefore F(x) = P(X \le x)$

$\because$ F(X) represents Probability; $0 \le F(X) \le 1$

26. $F(X) = P(X \le C)$ will imply the area under the probability curve to the left of vertical line at C.

27. **Uniform Distribution (Continuous)**

A. A continuous Random Variable is said to follow uniform distribution if the probabilities associated with intervals of same width are always equal at all parts and for any range of values.

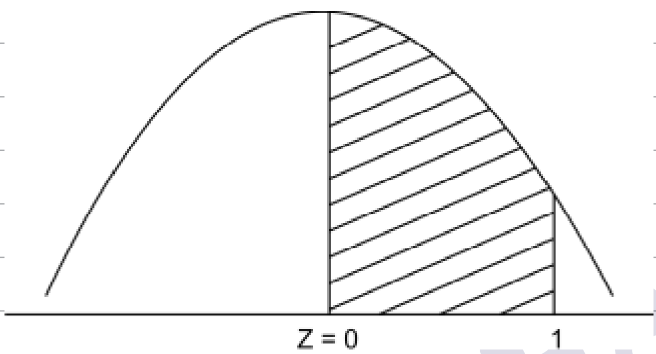
B. P. D. F. of uniform distribution is given by : $f(x) = \frac{1}{b-a}$ $(a \le x \le b)$

C. It is also known as "Rectangular Distribution"

D. Probability that X lies between any two specified values C and D within the range ("b and a") is given by : $P(c \le x \le d) = \frac{d-c}{b-a}$

28. Areas under Standard Normal Curve

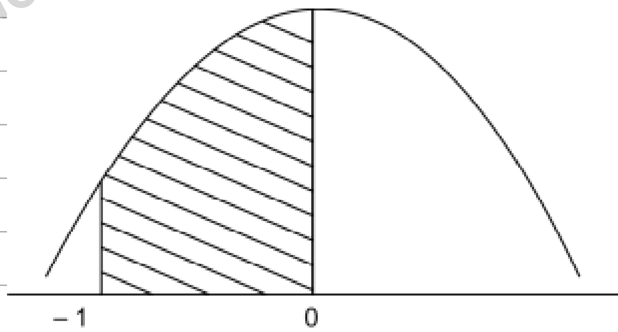
C-I



$P(0 \le Z \le 1) = .3413$, [Z = 0 & Z = any $+^{VE}$ value]

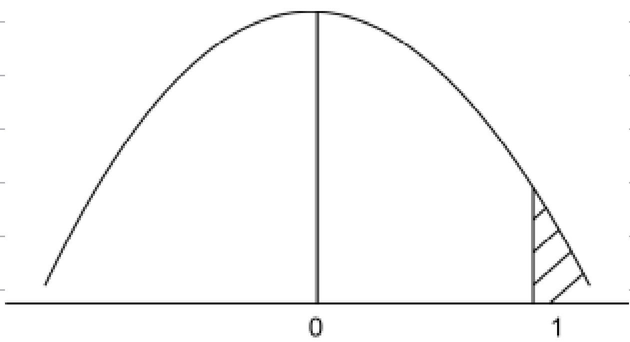
C-II

$P(-1 \le Z \le 0)$
$= P(0 \le Z \le 1)$
$= .3413$
[Z = $-^{VE}$ value to Z = 0] (Due to Symmetry)



C-III

$P(Z \ge 1)$
$= .5 - P(0 \le Z \le 1)$
$= .5 - .3413$
$= .1587$



[Area of the Region Greater than A $+^{VE}$ value]

C-IV

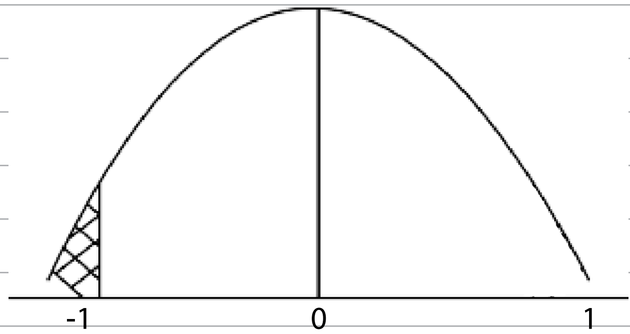
$P(Z \leq -1)$

$= P(Z \geq 1)$ (Due to symmetry)

$= .5 - P(0 \leq$

$= .5 - .3413$

$= .1587$



[Area of the Region Less than A negative value]

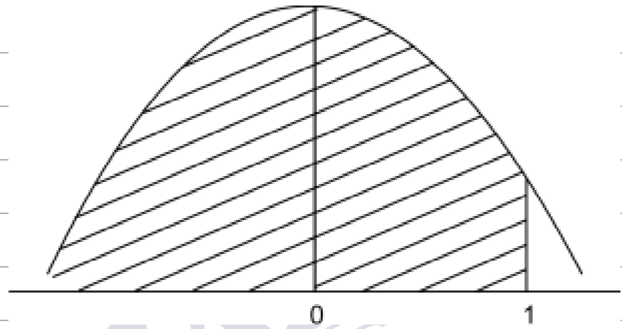
C-V

$P(Z \leq 1)$

$= .5 + P(0 \leq Z \leq 1)$

$= .5 + .3413$

$= .8413$



Area of the Region less than A +VE value]

C-VI

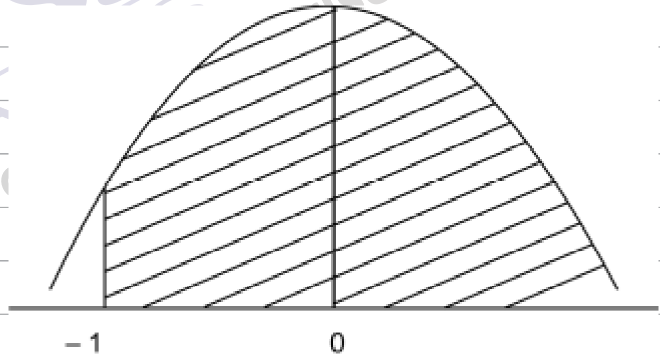
$P(Z \geq -1)$

$= P(Z \leq 1)$ (Due to symmetry)

$= .5 + P(0 \leq Z \leq 1)$

$= .5 + .3413$

$= .8413$



[Area of the Region greater than A –VE value]

C-VII

$P(1 \leq Z \leq 2)$

$= P(0 \leq Z \leq 2)$

$\quad - P(0 \leq Z \leq 1)$

$= .4772 - .3413$

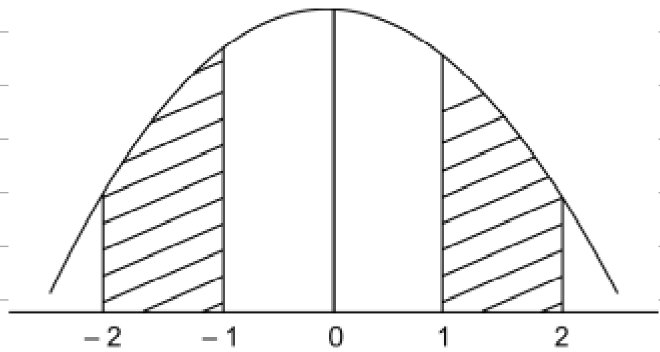
$= .1359$

Or $P(-2 \leq Z \leq -1)$

$= P(1 \leq Z \leq 2)$ (Due to symmetry)

$= P(0 \leq Z \leq 2) - P(0 \leq Z \leq 1)$

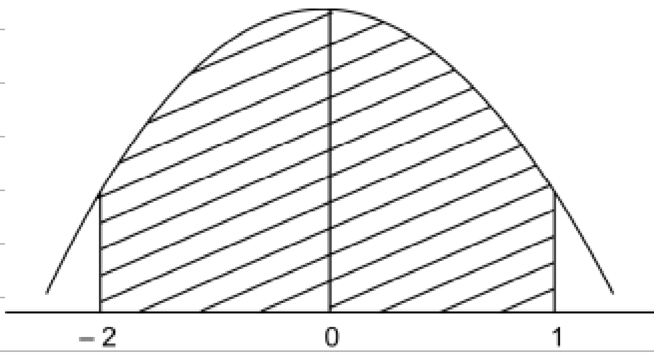
$= .4772 - .3413 = .1359$



(Area between two +VE or Two –VE values of Z]

C-VIII

$P(-2 \le Z \le 1)$
$= P(-2 \le Z \le 0) + (0 \le Z \le 1)$
$= P(0 \le Z \le 2) + (0 \le Z \le 1)$
$= .4772 + .3413$
$= .8185$



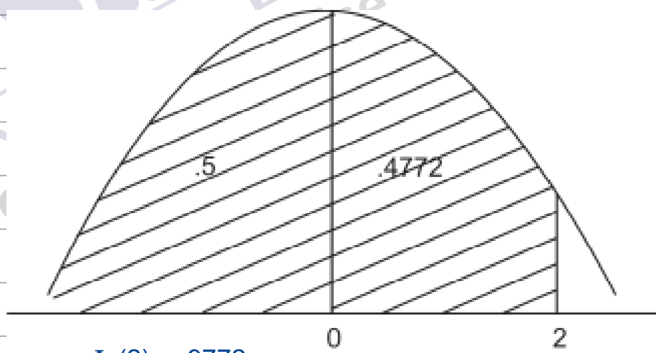
(AREA BETWEEN ONE -^{VE} AND ONE + ^{VE} VALUE)

NOTE:

1) **If the $^{-ve}$ and $^{+ve}$ values happen to be identical .i.e P $(-a \le z \le a)$ in such a case the total area will be = 2P $(0 \le z \le a)$**

2) **When in the problem the magnitude of the given area is greater than ".5" it implies that area from -α to that particular value of 'z' is provided, for evaluating the area from 0 to that particular value of 'z' subtract .5 from it.**

E.G. Φ (2) = .9772
$= P(Z \le 2) = .9772$
$= .5 + P(0 \le Z \le 2) = .9772$
$P(0 \le Z \le 2) = .9772 - .5$
$= .4772$



Φ (2) = .9772

29. Methods of fitting Normal Distribution or a Normal Curve

There Are Two Methods Of Fitting Normal Distribution

1) Ordinate Method
2) Area Method

30. Condition under which "Binomial" and "Possion" approaches "Normal Distribution"

Case I

Normal Distribution as a limiting case of Binomial Distribution when

a) n, the number of trials is infinitely large I.e. n $\rightarrow \alpha$
b) Neither p(or q) is very small, i.e. p and q are fairly near equal

c) In other words, if neither p nor q is very small but n is sufficiently large Binomial Distribution approaches Normal Distribution.

d) In such a case, the Standard Normal Variate is given by $Z = \frac{x-np}{\sqrt{npq}}$

Case II

Poission Distribution tends to Normal Distribution with standardised Variable

$Z = \frac{x-m}{\sqrt{m}}$

Where m = Mean = μ = Variance

$\sqrt{m}$ = S.d = σ as n increases indefinitely (i.e. as $n \rightarrow \alpha$)

CLASSWORK SECTION

1. If the mean deviation of a normal variable is 16, what is its quartile deviation?
   a) 10 b) 15 c) 13.5 d) 12.05

2. If the quartile deviation of a normal curve is 4.05, then its mean deviation is:
   a) 5.26 b) 6.24 c) 4.24 d) 4.80

3. If the two quartiles of normal distribution are 14.6 and 25.4 respectively, what is the standard deviation of the distribution?
   a) 6 b) 8 c) 9 d) 10

4. What is the first quartile of x having the following probability density function?

$$f(x) = \frac{1}{\sqrt{72\prod}} e^{\frac{-(x-10)^2}{72}} \text{ for } -\infty<x<\infty$$

   a) 4 b) 5 c) 5.95 d) 6.75

5. If x and y are 2 independent normal variable with mean 10 and 12 and SD 3 and 4 respectively, then (x + y) is also a normal distribution with mean _____ and SD _____.
   a) 22, 7 b) 22, 25 c) 22, 5 d) 22, 49

Area under Normal / Standard Normal Curve

Find the area under the standard normal curve for the following values of standard normal variate:

6. If the standard normal curve between z = 0 to z =1 is 0.3413, then the value of ϕ (1) is:
   a) 0.5000 b) 0.8413 c) - 0.5000 d) 1

7. For certain normal variate x, the mean is 12 and S.D is 4 find P(X≥20):
   [Area under the normal curve from z=0 to z=2 is 0.4772]
   a) 0.5238 b) 0.0472 c) 0.7272 d) 0.0228

8. If the weekly wages of 5000 workers in a factory follows normal distribution with mean and SD as ₹700 and ₹50 respectively, what is the expected number of workers with wages between ₹ 660 and ₹ 720?

a) 2050 b) 2200 c) 2218 d) 2300

9. 50 per cent of a certain product have weight 60kg or more whereas 10 per cent gave weight 55 kg or less. On assumption of normality, what is the variance of weight? Given ϕ (1.28) = 0.90.

a) 15.21 b) 9.00 c) 16.00 d) 22.68

Theoretical Aspects

10. For a normal distribution, P(X ≥ μ) = ____________

a) 0 b) 1 c) 0.5 d) 0.6826

11. The probability distribution of z is called Standard Normal Distribution and is defined by the probability density function:

a) $f(x)=\frac{1}{\sqrt{2\pi}}e;\ -\infty<x<\infty$

b) $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2};\ -\infty<x<\infty$

c) $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}};\ -\infty<z<\infty$

d) $f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2};\ -\infty<z<\infty$

12. If a random variable is normally distributed with mean μ and standard deviation σ, then $z=\frac{x-\mu}{\sigma}$ is called:

a) Normal Variate b) Standard Normal Variate

c) Chi-square Variate d) Uniform Variate

13. The curve of which of the following distribution is uni-modal and bell shaped with the highest point over the mean

a) Poisson b) Binomial c) Normal d) All of the above

14. In Normal distribution as the distance from the ________ increases, the curve comes closer and closer to the horizontal axis.

a) Standard Deviation b) Mean

c) Both a) and b) above d) Neither a) nor b) above

15. For Standard Normal distribution, which of the following is correct?

a) Mean = 1; S.D. = 1 b) Mean = 1, S.D. = 0

c) Mean = 0, S.D. = 1 d) Mean = 0, S.D. = 0.

16. The mean deviation about median of a Standard Normal Variate is:

a) 0.675σ b) 0.675 c) 0.80σ d) 0.80

17. The interval $(\mu - 3\sigma, \mu + 3\sigma)$ covers

a) 96% area of a normal distribution.

b) 95% area of a normal distribution.

c) 99% area of a normal distribution.

d) All but 0.27% area of a normal distribution

18. The symbol ϕ (a) indicates the area of standard normal curve between

a) 0 to a b) a to ∞ c) $-\infty$ to a d) $-\infty$ to ∞

19. An approximate relation between Quartile deviation (QD) and Standard Deviation (SD) of normal distribution is:

a) 5 QD = 4 SD b) 4 QD = 5 SD

c) 2 QD = 35 SD d) 3 QD = 2 SD

20. The probability that x assumes a specified value in continues probability distribution is ___________.

a) 1 b) 0

c) -1 d) None

Theory Answer Key

10	c	11	c	12	b	13	c	14	b
15	c	16	d	17	d	18	a	19	d
20	b								

7 APPENDIX

Table I Area Under Standard Normal Curve

(Proportion of area under standard normal curve between the ordinates at z = 0 and given values of z)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	..4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974

2.8	.4974	.4975	.4976	.4977	.4977	.4973	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

Table II Values of e^{-m}

m	e^{-m}	m	e^{-m}	m	e^{-m}
0.0	1.0000	1.5	0.2231	3.0	0.0498
0.1	0.9048	1.6	.2019	3.2	.0408
0.2	.8187	1.7	.1827	3.4	.0334
0.3	.7408	1.8	.1653	3.6	.0273
0.4	.6703	1.9	.4497	3.8	.0224
0.5	.6065	2.0	.1353	4.0	.0183
0.6	.5488	2.1	.1225	4.2	.0150
0.7	.4966	2.2	.1108	4.4	.0123
0.8	.4493	2.3	.1003	4.6	.0100
0.9	.4066	2.4	.0907	4.8	.00823
1.0	.3679	2.5	.0821	5.0	.00674
1.1	.3329	2.6	.0743	5.5	.00409
1.2	.3012	2.7	.0672	6.0	.00248
1.3	.2725	2.8	.0608	6.5	.00150
1.4	.2466	2.9	.0550	7.0	.00091

Table III - LOGARITHM

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170						5	9	13	17	21	26	30	34	38
						0212	0253	0294	0334	0374	4	8	12	16	20	24	32	36	36
11	0414	0453	0492	0531	0569						4	8	12	16	20	23	27	31	35
						0607	0645	0682	0719	0755	4	7	11	15	18	22	26	29	33
12	0792	0828	0964	0899	0934						3	7	11	14	18	21	25	28	32
						0969	1004	1038	1072	1106	3	7	10	14	17	20	24	27	31
13	1139	1173	1208	1239	1271						3	6	10	13	16	19	23	26	29
						1303	1335	1367	1399	1430	3	7	10	13	16	19	22	25	29
14	1461	1492	1523								3	6	9	12	15	19	22	25	28
				1553	1584	1614	1644	1673	1703	1732	3	6	9	12	14	17	20	23	26
15	1761	1790	1818								3	6	9	11	14	17	20	23	26
				1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	19	22	25
16	2041	2068	2095	2122	2148						3	6	8	11	14	16	19	22	24
						2175	2201	2227	2253	2279	3	5	8	10	13	16	18	21	23
17	2304	2330	2355	2380	2405						3	5	8	10	13	15	18	20	23
						2430	2455	2480	2504	2529	3	5	8	10	12	15	17	20	22
18	2553	2577	2601	2625	2648						2	5	7	9	12	14	17	19	21
						2672	2695	2718	2742	2765	2	4	7	9	11	14	16	19	21
19	2788	2810	2833	2856	2878						2	4	7	9	11	13	16	18	20
						2900	2923	2945	2967	2989	2	4	6	8	11	13	15	17	19
20	3010	3023	3054	3075	3096	3116	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3909	3927	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	11	13	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	12	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	10	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	11	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11

36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	631	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6235	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6575	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6037	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8

Example:

Log 2 = 0.3010: Log 20 = 1.3010: Log 200 = 2.3010: Log 2,000 = 3.3010 etc.

Log 2 = 0.3010 - 1 – (–) 0.699

Log 0.02 = 0.3010 - 2 – (–) 1.699

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7166	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7306	7314	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7358	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7452	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7848	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7958	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8158	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8428	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8673	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8738	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	2	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	2	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	2	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	2	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	2	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4

88	9445	9450	9450	9455	9460	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9945	9952	0	1	1	2	2	3	3	4	4
99	9958	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4

Table IV - ANTILOGARITHM

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
100	1000	1002	1005	1007	1009	1012	1014	1016	1018	1021	0	0	1	1	1	1	2	2	2
101	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
102	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
103	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
104	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
105	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
106	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
107	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
108	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
109	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
110	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
111	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
112	1381	1321	1324	1327	1330	1334	1337	1340	1342	1348	0	1	1	1	2	2	2	2	3
113	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
114	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
115	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
116	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
117	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
118	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
119	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
120	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
121	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
122	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
123	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
124	1738	1742	1746	1750	1754	1758	1762	1768	1770	1774	0	1	1	2	2	2	3	3	4
125	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
126	1820	1824	1828	1832	1837	1841	1845	1849	1897	1858	0	1	1	2	2	3	3	3	4
127	1862	1866	1871	1875	1879	1884	1888	1892	1941	1901	0	1	1	2	2	3	3	3	4
128	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
129	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
130	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
131	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
132	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
133	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
134	2188	2193	2198	2203	2206	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
135	2239	2244	2249	2254	2259	2265	2270	2275	2280	2256	1	1	2	2	3	3	4	4	5

136	2291	2286	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
137	2344	2350	2355	2359	2366	2271	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
138	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
139	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
140	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
141	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
142	2630	2636	2642	2649	2655	2661	2667	2673	2679	2624	1	1	2	2	3	4	4	5	6
143	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
144	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
145	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
146	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
147	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
148	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
149	3090	3097	3105	3112	3118	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
150	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
151	3236	3243	3251	3258	3268	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
152	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
153	3388	3396	3404	3412	3420	3428	3436	3442	3451	3459	1	2	2	3	4	5	6	6	7
154	3467	3475	3483	3491	3499	3508	3516	3524	3532	2540	1	2	2	3	4	5	6	6	7
155	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	6	7
156	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
157	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
158	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
159	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	4	5	6	7	8
160	3981	3990	3999	4009	4018	4027	4036	4046	4055	4065	1	2	3	4	5	6	6	7	8
161	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
162	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
163	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
164	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
165	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
166	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
167	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
168	4788	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
169	4898	4909	4920	4932	4943	4955	4986	4977	4989	5000	1	2	3	5	6	7	8	9	10
170	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
171	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
172	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
173	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
174	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
175	5632	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
176	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
177	5858	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
178	6028	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
179	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
180	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
181	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
182	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
183	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
184	6918	6934	6950	6965	6982	6598	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
185	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
186	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
187	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16

188	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
189	7762	7780	7796	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
190	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
191	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
192	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
193	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
194	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
195	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
196	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
197	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
198	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
199	9772	9795	9817	9840	9836	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

Example:

If Log x = 0.301. then x = Antilog 0.301 = 2

If Log x = 1.301. then x = (Antilog 0.301) × 10 = 20

If Log x = 2.301. then x = (Antilog 0.301) × 100 = 200

If Log x = (–) 0.699, then we can write Log x = (– 1 + 0.301) : Thus x = Antilog (0.301) / 10 = 0.2

If Log x = (–) 1.699, then we can write Log x = (– 2 + 0.301) : Thus x = Antilog (0.301) / 100 = 0.02