CHAPTER 1:
Data Condensation and Presentation of Data

1. Definition, importance, scope and limitations of statistics.
2. Data Condensation: Types of data (Primary and secondary), Attributes and variables, discrete and Continuous variables.
3. Graphical Representation: Histogram, Ogive Curves, Steam and leaf chart. [Note: Theory paper will contain only procedures. Problems to be included in practical]
4. Numerical problems related to real life situations.

1.1 Definition, importance, scope and limitations of statistics.

Definition of Statistics:

Statistics is a branch of mathematical science which deals with:

- ❑ Collection of data
- ❑ Classification or tabulation of data
- ❑ Analysis of data
- ❑ Interpretation of results.

1.2 Data Condensation: Types of data (Primary and secondary), Attributes and variables, discrete and Continuous variables.

1) Data: It is the information in the form of ‘Numbers’ or ‘Categories’.
   Ex. i) Salaries of employees of a company
   ii) Blood groups of blood donors.
2) Statistical data: The observations or units in the population or in the sample is called as Statistical data.
3) Raw Data: It is the statistical data before any further processing or data before any mathematical treatments.
   Ex. Marks of students.
   45,58,25,29,75,35,48,68,15,85,45,78,59,65,75,95,65,42,8,25,67
   .

❑ Attributes and variables:

➢ **Attribute**:

A qualitative characteristics which can not be measured by numbers but can be described in categories called as Attributes.

Ex: Gender, Blood groups, States of country, Names of vehicles etc.

Above characteristics can be characterized by their qualities or categories, hence these are called as Attributes.

➢ **Variable**:

A quantitative characteristics which can be measured by numbers is called as Variable.

Ex: Age, Height , Weight, Blood pressure, Temperature, Marks, Heart beat rate, Mileage of car, Speed, Distance, Salary, Income, Bills, Price, Rent, etc. Data File

Type of Variables:

- There are two types of variables:

1. Discrete variable
2. Continuous variable

Discrete variable:

A variable which can take isolated values and it has finite possible values is called as Discrete variable.

Ex. i) Number of students present in class

ii) Number of workers in a factory

iii) Number of children in a family

Continuous Variable:

It is a variable which can take infinite possible values within certain interval or range.

Ex.

i. *Height / Weight of a person*
ii. *Temperature of a certain place*
iii. *Speed of a vehicle.*

Classification of Data:

- After collection of data, next step is classification of data.
- Classification is a grouping of observations in different classes on the basis of suitable criteria.

Objective of Classification:

1) To condense the mass of data in such a way that similarities and dissimilarities can be easily understood.
2) To facilitate comparison.
3) To enable the statistical treatment of collected data.
4) To give more importance for necessary information and dropping unnecessary information.

Types of Classification of data:

1) Geographical classification
2) Chronological classification
3) Qualitative classification
4) Quantitative classification.

1) Geographical classification:

Classification according to locational or geographical differences are called as geographical classification. Data

Ex. Villages, Cities, States, Countries, Sea area, Mountain Area, Rivers.

2) Chronological classification:

The classification based on time period that is: Hourly, Daily, Weekly, Monthly, yearly etc. are called as chronological classification or also called as time series. Data

Ex. Yearly sales of Vehicles/Mobiles etc.

3) Qualitative Classification:

The classification on different types of categories of Attribute is called as Qualitative classification.

There are two types of Qualitative classification.

i)Simple Classification: The classification which has only two possible categories is called as Simple classification. Data

Ex. Yes/No, True/False, Domestic/International, Head/Tail.

ii)Manifold Classification: The classification which has more than two possible categories is called as Manifold classification.

Data

Ex. Blood groups, Nationality, Religion

4) Quantitative Classification:

The classification of data according to some characteristics that can be measured numerically.

For Ex.

i) height, weight, income, age, sales, etc

ii) The employees of an institute may be classified according to their pay scales as follows:

Table-13.3: Quantitative Classification of 840 Employees According to their Pay Scales

Scale of Pay	Number of Employees
9300 - 34800	467
15600 - 39100	215
37400 - 67000	158
Total	840

1.3 Classification of variables:

1) Discrete series of observations:

Each possible value of the variable forms a class is said to form a discrete series of observations.

For Ex. i) Number of children in a family

Number of children	Tally marks	Number of families
0	|||||	5
1	||||| ||	7
2	||||| ||||| ||	12
3	|||||	5
4	||||| |	6
5	|||	3
6	|||	3

Continued...

2) Class:

The Variables that are discrete or continuous are classified by dividing the total range into suitable number of intervals and each interval represents a class.

Range = 100-0=100

And classes = 10

Then class = 100/10 = 10

3) Class limits:

The two numbers designating the class are called class limits.

Lower value of the class is called lower limit and upper value is called upper class limit.

4. Class width:

It is the difference between two successive lower limits or two successive upper limits or difference between upper and lower limit of the class.

If we have class intervals in which Upper limits is not equal to lower limit of its succeeding class interval, then class width h is given by:

Class Width = h = Upper limit of Second class- Upper limit of First class

Or Class Width = h = Lower limit of Second class- Lower limit of First class

Frequency Distribution

- The **class width** is the distance between lower (or upper) limits of consecutive classes.
  - Example: 6 – 1 = 5

Class	Frequency
1 – 5	5
6 – 10	8
11 – 15	6
16 – 20	8
21 – 25	5
26 – 30	4

If we have class intervals in which Upper limits is equal to lower limit of its succeeding class interval, then class width h is given by:

Class Width = h = Upper class limit – Lower class limit

(Note: Class width is also called as Class size)

Class Interval	Class Limits		Class Size
	Lower Limit	Upper Limit	
0 - 10	0	10	10 - 0 = 10
10 - 20			
20 - 30			
30 - 40			

If class width is not known, we use following formula to get class width:

$$h = \frac{L-S}{k}$$

Where,

L = Largest observation in the data

S= Smallest observation in the data

k = Number of classes

And k can be determined by Sturges' rule and it is given by formula:

$$k = 1 + 3.322 * Log_{10}N$$

And N = Total observations in the data set.

- Another approach is: k = Number of classes = $\sqrt{N}$

5. Class Mark or Mid value:

It is the mid-point of class interval:

That is:

$$Mid\ Value = \frac{Lower\ Class\ Limit + Upper\ Class\ Limit}{2}$$

Age (years) of Best actor when Oscar was won	Mid points	Class Width
20-29	$\frac{20+29}{2} = 24.5$	10
30-39	$\frac{30+39}{2} = 34.5$	10
40-49	$\frac{40+49}{2} = 44.5$	10
50-59	$\frac{50+59}{2} = 54.5$	10
60-69	$\frac{60+69}{2} = 64.5$	10
70-79	$\frac{70+79}{2} = 74.5$	10

6. Open End Classes:

Classes having only one limit are known as open end classes.

Ex.

Marks	*Number of Students*
Less than 10	4
10 – 20	8
20 – 30	5
30 – 40	3
40 and above	10

7. Class Boundaries:

If we have class intervals in which Upper limit of a class is not equal to lower limit of its succeeding class interval, then such classes can be extended to form a continuous frequency distribution. Such a extended limits are called class boundaries.

Step 1) Find Gap = Lower limit of second class – Upper limit of first class

Step 2) Find Gap/2

Step 3) Subtract Gap/2 from each lower limit and add Gap/2 in each upper limit

Thus

Lower Class Boundary = Lower Class Limit – Gap/2

Upper Class Boundary = Upper Class Limit + Gap/2

- Gap = 49 – 48 = 1
- Gap/2 = ½ = 0.5
- Subtract 0.5 each lower limit
- Add 0.5 in each upper limit
- 44-48 44-0.5 = 43.5 and 48+0.5 = 48.5
- 43.5 to 48.5

Class Limit	Class boundary
44 - 48	43.50 - 48.50
49 - 53	48.50 - 53.50
54 - 58	53.50 - 58.50
59 - 63	58.50 - 63.50
64 - 68	63.50 - 68.50
69 - 73	68.50 - 73.50

8. Class frequency:

The number of observations corresponding to each class gives that class frequency.

Number of children	Tally marks	Number of families
0	𝍸	5
1	𝍸 ||	7
2	𝍸 𝍸 ||	12
3	𝍸	5
4	𝍸 |	6
5	|||	3
6	|||	3

Weight in kg (Class Interval)	Tally marks	No. of Students (Frequency)
44-48	III	3
49-53	IIII	4
54-58	𝍸	5
59-63	𝍸 II	7
64-68	𝍸 IIII	9
69-73	𝍸 III	8
Total	–	36

9. Inclusive type of Classification:

The classification in which the classes are defined or formed such way that both the limits are included in the same class.

Weight in kg (Class Interval)	Tally marks	No. of Students (Frequency)
44-48	III	3
49-53	IIII	4
54-58	𝍸	5
59-63	𝍸 II	7
64-68	𝍸 IIII	9
69-73	𝍸 III	8
Total	–	36

In this classification, lower and upper limit 44 and 48 are included in the class 44-48, lower and upper limit 49 and 53 are included in the class 49-53 and so on.

10. Exclusive type of Classification:

The classification in which the classes are defined or formed by excluding upper limit in the same class and we include this upper limit value in next class interval.

Class interval	Frequency
0–10	3
10–20	8
20–30	9
30–40	15
40–50	5
	Total = 40

Class interval	Tally marks	Frequency
60-75	II	2
75-90	IIII	4
90-105	𝍸 I	6
105-120	II	2
120-135	𝍸 I	6
135-150	IIII	4

In this classification, upper limit 10 is excluded from the class 0-10 and we include this value 10 in next class 10-20, similarly exclude 20 from 10-20 and include 20 in next class 20-30 and so on.

11. Frequency Distribution:

The number of observations belonging to a class is called as frequency of that class and various classes together with their frequencies is called a frequency distribution.

There are two types:

i) Ungrouped or Discrete frequency distribution:

In this classification, we form classes according values of the variable and then we count the frequency of each value.

Number of children	Tally marks	Number of families
0	𝍸	5
1	𝍸 ||	7
2	𝍸 𝍸 ||	12
3	𝍸	5
4	𝍸 |	6
5	|||	3
6	|||	3

- Discrete data: possible values are countable

Example: An advertiser asks 200 customers how many days per week they read the daily newspaper.

Number of days read	Frequency
0	44
1	24
2	18
3	16
4	20
5	22
6	26
7	30
Total	200

ii. Grouped or Continuous Frequency Distribution:

When the number of observations are very large, so that we can not form a class for each possible value, so we divide total range of observations into suitable number of intervals and each interval is called as class. Next we count number of observations that fall in each class and write against respective class.

Thus table obtained of class intervals with frequencies called as a Grouped or Continuous Frequency Distribution.

Class interval	Tally marks	Frequency
60-75	II	2
75-90	IIII	4
90-105	𝍸I	6
105-120	II	2
120-135	𝍸I	6
135-150	IIII	4

12. Relative frequency:

Relative frequency of a class is defined as the ratio of the class frequency to total frequency.

That is: $Relative\ Frequency = \frac{Frequency\ of\ that\ class}{Total\ Frequency}$

The classes together with their relative frequencies is called relative frequency distribution.

Relative frequency table		
Weight (in Kg.)	No. of persons (f_i)	Relative frequency
60 - 62	5	$\frac{5}{100} \times 5\%$ or 0.05
63 - 65	18	$\frac{18}{100} \times 18\%$ or 0.18
66 - 68	42	$\frac{42}{100} \times 42\%$ or 0.42
69 - 71	27	$\frac{27}{100} \times 27\%$ or 0.27
72 - 74	8	$\frac{8}{100} \times 8\%$ or 0.08
Total	$\Sigma f_i = 100$	

class interval	frequency Fi	Relative Frequency R Fi	
0 - 20	4	=4/15	0.27
20 - 40	1	=1/15	0.07
40 - 60	4	=4/15	0.27
60 - 80	1	=1/15	0.07
80 - 100	5	=5/15	0.33
sum	15		1

13. Frequency Density:

Frequency density of a class is the ratio of the class frequency to class width.

That is: $Frequency\ Denisty = \frac{Class\ Frequency}{Class\ Width}$

It is used when class widths are not uniform, that is class widths are different for each class interval.

HEIGHT (CM)	FREQUENCY	CLASS WIDTH	FREQUENCY DENSITY
$65 < h \leq 75$	2	10	2/10 = 0.2
$75 < h \leq 80$	7	5	7/5 = 1.4
$80 < h \leq 90$	21	10	21/10 = 2.1
$90 < h \leq 105$	15	15	15/15 = 1
$105 < h \leq 110$	12	5	12/5 = 2.4

Frequency Densities

- To calculate the frequency densities, we use a table as shown here:

Mass (Kg)	Class width	Frequency	Frequency Density
6 – 8	3	4	$\frac{4}{3} = 1\frac{1}{3}$
9 – 11	3	6	$\frac{6}{3} = 2$
12 – 17	6	10	$\frac{10}{6} = 1\frac{2}{3}$
18 – 20	3	3	$\frac{3}{3} = 1$
21 – 29	9	12	$\frac{12}{9} = 1\frac{1}{3}$

Cumulative Frequency Distribution:

❖ Cumulative Frequency:

When frequencies are added they are called as cumulative frequencies.

There are two types of cumulative frequencies:

i. Less than cumulative frequency

ii. More than cumulative frequency

i. Less than cumulative frequency

When frequencies are added from top to bottom and obtained cumulative frequencies which give the number of observations less than given values are called as less than cumulative frequencies.

Less than cumulative frequency of a given class is the number of observations having their values less than the upper boundary of that class.

Classes	Frequency	Upper Boundary		Cumulative Frequency calculations	Less than Cumulative Frequency
0--10	3	10	Less than 10	3	3
10--20	8	20	Less than 20	=3+8 = 11	11
20--30	11	30	Less than 30	=11+11 = 22	22
30--40	14	40	Less than 40	=22+14 = 36	36
40--50	20	50	Less than 50	=36+20 = 56	56
50--60	17	60	Less than 60	=56+17 = 73	73
60--70	13	70	Less than 70	=73+13 = 86	86
70--80	13	80	Less than 80	=86+13 = 99	99
80--90	12	90	Less than 90	=99+12 = 111	111
90--100	7	100	Less than 100	=111+7 = 118	118

ii. More than cumulative frequency

When frequencies are added from bottom to top and obtained cumulative frequencies which give the number of observations more than given values are called as more than cumulative frequencies.

More than cumulative frequency of a given class is the number of observations having their values more than the lower boundary of that class.

Classes	Frequency	Lower Boundary		Cumulative Frequency calculations	More than Cumulative Frequency
0--10	3	0	More than 0	=115+3 = 118	118
10--20	8	10	More than 10	=107+8 = 115	115
20--30	11	20	More than 20	=96+11 = 107	107
30--40	14	30	More than 30	=82+14 = 96	96
40--50	20	40	More than 40	=62+20 = 82	82
50--60	17	50	More than 50	=45+17 = 62	62
60--70	13	60	More than 60	=32+13 = 45	45
70--80	13	70	More than 70	=19+13 = 32	32
80--90	12	80	More than 80	=7+12 = 19	19
90--100	7	90	More than 90	7	7

Graphical Representation of Frequency Distribution:

i) Histogram

ii) Stem and Leaf Chart

iii) Ogive or Cumulative frequency curve

Histogram:

- In this graph grouped data is represented by series of adjacent rectangles.
- The base of each rectangle is the class interval of that class
- If classes are not continuous or classes are discontinuous, then they are to be converted into continuous classes by converting the limits in class boundaries.
- The height of each rectangle is proportional to:
  - The Frequency of the corresponding class if class widths are uniform
  - The frequency densities of the corresponding class if class widths are not uniform

Intervals	Frequency
0 - 5	2
5 - 10	2
10 - 15	5
15 - 20	5
20 - 25	3
25 - 30	3