# STATISTICS

# **BASIC CONCEPT OF STATISTICS**

## **INTRODUCTION**

The branch of science known as Statistics has been used in India from ancient times. Statistics deals with collection of numerical facts i.e., data, their classification & tabulation and their interpretation. In statistics we shall try to study, in detail about collection, classification and tabulation of such data.

## (a) Importance of Data :

Expressing facts with the helps of data is of great importance in our day-today life. For example, instead of saying that India has a large population it is more appropriate to say that the population of India, based on the census of 2000 is more than one billion.

## (b) Collection of Data :

On the basis of methods of collection, data can be divided into two categories :

### (i) Primary data :

Data which are collected for the first time by the statistical investigator or with help of his workers is called primary data. As example if an investigator wants to study the condition of the workers working in a factory then fro this he collects some data like their monthly income, expenditure, number or brother, sisters, etc.

## (ii) Secondary data :

These are the data already collected by a person or a society and these may be in published or unpublished form. These data should be carefully used. These are generally obtained from the following two sources.

- (A) Published sources
- (B) unpublished sources
- (c) Classification of Data :

When the data is complied in the same form and order in which it is collected, it is known as Raw Data, It is also Crude Data. For example, the marks obtained by 20 students of class X in English out of 10 marks are as follows :

7, 4,9,5,8,9,6.7,9,2,0 3,7,6,2,1,9,8,3,8,

## (i) Geographical basis :

Here, the data is classified on the basis of place or region. For example the production of food grains of different state is shown in the following table :

S.No.	State	Production (in Tons)
1	Andhdra Pradesh	9690
2	Bihar	8074
3	Haryana	10065
4	Pubjab	17065
5	Uttar Pradesh	28095

## (ii) Chronological classification :

If data's classification is based on hour, day, week and month or year, then it is called chronological classification, For example, the population of India in different year is shown in following table :

S.No	Year	Production (in Crores)
1	1951	46.1
2	1961	53.9
3	1971	61.8
4	1981	68.5
5	1991	88.4
6	2001	100.01

## (iii) Qualitative basis :

When the data are classified into different groups on the basis of their descriptive qualities and properties, such a classification is known as descriptive or qualitative classification. Since the attributes can not be measured directly, they are counted on the basis of presence or absence of qualities. For example intelligence, literacy, unemployment, honesty etc. The following table shows classification on the basis of sex and employment.

Table Population (in lacs)

Gender →	Male	Female
Position of Employment $\downarrow$		
Employed	16.2	13.7
Unemployed	26.4	24.8
Total	42.6	38.5

## (iv) Quantitative basis :

if facts are such that they can be measured physically e.g. marks obtained height, weight, age, income, expenditure etc. Such facts are known as variable values. If such facts are kept into classes then it is called classification according to quantitative or class intervals.

Marks obtained	10-20	20-30	30-40	40-50
No. of students	7	9	15	6

## DEFINATION

## (i) Variate :

The numerical quantify whose value varies in objective is called a variate, generally a variate is represented by x. There are two types of variate.

## (A) Discrete variate :

its magnitude is fixed. For example, the number of teacher in different branches of a institute are 30, 35, 40 etc.

## (B) Continuous variate :

is magnitude is not fixed. It is expressed in groups like 10 - 20, 20 - 30, .... etc.

## (ii) Rage :

The difference of the maximum and the minimum values of the variable x is called range.

## (iii) Class frequency:

In each class the number of times a data is repeated in known as its class frequency.

# (iv) Class Interval

 $= \frac{\text{Range}}{\text{Numberfclasse}}$ 

It is generally denoted by h or i.

# (v) Class limits :

The lowest and the highest value of the class are known as lower and upper limited restively of that class.

# (vi) Class mark:

The average of the lower and the upper limits of a class is called the mid value or the class mark of that class. It is generally denoted by x.

If x be the mid value and h be the class interval, then the class limits are  $\left(x - \frac{h}{2}x + \frac{4}{2}\right)$ .

- **Ex.1** The mid values of a distribution are 54, 64, 74, 84 and 94. Find the class interval and class limits.
- Sol. The class interval is the difference of two consecutive class marks, therefore class interval (h) = 64-54 = 10.

Here the mid values are given and the class interval is 10.

So class limits are

For 1<sup>st</sup> class 
$$54 - \frac{10}{2}t054 + \frac{10}{2}$$
 or  $49$  to 59  
For 2<sup>nd</sup> class  $64 - \frac{10}{2}t064 + \frac{10}{2}$  or  $59$  to 69  
For 3<sup>rd</sup> class  $74 - \frac{10}{2}t074 + \frac{10}{2}$  or  $69$  to 79  
For 4<sup>th</sup> class  $84 - \frac{10}{2}t084 + \frac{10}{2}$  or  $79$  to 89  
For 5<sup>th</sup> class  $94 - \frac{10}{2}t094 + \frac{10}{2}$  or  $89$  to 99

Therefore class limits are 49 - 59, 59 - 69, 79 - 89, and 89 - 99.

## **FREQUENCY DISTRIBUTION**

The marks scored by 30 students of IX class, of a school in the first test of Mathematics our of 50 marks are as follows :

6	32	10	17	22	28	0	48	6	22
32	6	36	26	48	10	32	48	28	22
22	22	28	26	17	36	10	22	28	0

The number of times a mark is repeated is called its frequency. It is denoted by f.

Marks obtained	Taly mark	Frequency	Marks obtained	Tally mark	Frequency
0	II	2	26	II	2
6	III	3	28	IIII	4
10	III	3	32	III	3
17	II	2	36	II	2
22	IIII I	6	48	III	3

Above type of frequency distribution is called ungrouped frequency distribution. Although this representation of data is shorter than representation of raw data, but from the angle of comparison and analysis it is quite bit. So to reduce the frequency distribution, it can be classified into groups is following ways and it is called grouped frequency distribution.

Class	Frequency
0-10	8
11-20	2
21-30	12
31-40	5
41-50	3

# (a) Kinds of Frequency Distribution :

Statistical methods like comparison, decision taken etc. depends of frequency distribution. Frequency distribution are of three types.

# (i) Individual frequency distribution :

Here each item or original price of unit is written separately. In n this category, frequency of each variable is one.

**Ex.2** Total marks obtained by 10 students in a class.

S.No.	1	2	3	4	5	6	7	8	9	10
Marks obtained	46	18	79	12	97	80	5	27	67	54

# (ii) Discrete frequency distribution :

When number of terms is large and variable are discrete, i.e., variate can accept some particular values only under finite limits and is repeated then its called discrete frequency distribution. For example the wages of employees and their numbers is shown in following table.

Monthly wages	No. Of employees
4000	10
6000	8
8000	5
11000	7
20000	2
25000	1

The above table shows ungrouped frequency distribution the same facts can be written in grouped frequency as follows :

Monthly wages	No. of employees
0-10.000	23
11 000 20 000	0
11,000-20,000	9
21,000-30,000	1

## NOTE:

If variable is repeated in individual distribution then it can be converted into discrete frequency distribution.

# (iii) Continuous frequency distribution :

When number of terms is large and variate is continuous. i.e., variate can accept all values under finite limits and they are repeated then it is called continuous frequency distribution. For example age of students in a school is shown in the following table :

Age (in year)	Class	No. of students
Less than 5 year	0-5	72
Between 5 and 10 y ear	5-10	103
Between 10 and 15 year	10-15	50
Between 15 and 20 year	15-20	25

## NOTE:

Continuous frequency distribution is generally represented in form of grouped frequency distribution and variate is continuous in i, so 0 - 5, 6 - 10, 11 - 15, 16 - 20 types of classes can not be made here. If such classes are made in the table then students of age 5 to 6 year or 10 to 11 year or 15 to 16 years can not be classified. if such type of classes are given then they should be made continuous by following methods. Half of the difference between classes should be added to the upper limit of lower class and subtracted from lower limit o upper class. Thus the classes 0 - 5.5, 5.5 - 10.5, 10.5 - 15.5, 15.5 - 19.5 are obtained which are continuous.

## CLASSES CAN BE MADE MAINLY BY TWO METHODS:

## (i) Exclusive series :

In this method upper limit of the previous class and lower limit of the next class is same. In this method the term of upper limit in a class is not considered in the same class, it is considered in the next class.

## (ii) Inclusive series :

In this method value of upper and lower limit are both contained in same class. In this method the upper limit of class and lower limit of other class are not same. Some time the

I	ł	В		
Class	Frequency	Or Class	Frequency	
0-9	4	0-9.99	4	
10-19	7	10-19.99	7	
20-29	6	20-29.99	6	
30-39	3	30-39.99	3	
40-49	3	40-49.99	3	

value is not a whole number, it is a fraction or in decimals and lies in between the two intervals then in such situation the class interval can be constructed as follows

# **CUMULIVE FREQUENCY:**

## (i) Discrete frequency distribution :

From the table of discrete frequency distribution, it can be identified that number of employees whose monthly income is 4000 or how many employees of monthly income 1100 are there. But if we want to know how many employees whose monthly income is upto 11000, then we should add 10 + 8 + 57 i.e., number of employees whose monthly income is upto 11000 is 30. Here we add all previous frequency and get cumulative frequency. If will be more clear from the following table

Class	Frequency (f)	Cumulative frequency (cf)	Explanation
4000	10	10	10 = 1 0
6000	8	18	10 + 8
8000	5	23	18 + 5
11000	7	30	23 + 7
20000	2	32	30 + 2
25000	1	33	32 + 1

# (ii) Continuous frequency distribution :

In the previous page we obtained cumulative frequency for discrete series. Similarly cumulative frequency table can be made from continuous frequency distribution also. For example, for table :

Monthly income	No. of employee	Cumulative	Explanation
Variate (x)	Frequency (f)	Frequency (cf)	
0 – 5	72	72	72 = 72
5 - 10	103	175	72 + 103 = 175
10 – 15	50	225	175 + 50 = 225
15 - 20	25	250	225 + 25 = 250

Above table can also be written as follows :

Clas	Cumulative Frequency
Less than 5	72
Less than 10	175
Less than 15	225
Less than 20	250

From this table the number of students of age less than the upper limit of a class, i.e., number of student whose age is less than 5, 10, 15, 20 year can determined by merely seeing the table but if we need the number students whose age is more than zero, more than 5, more than 10 or more than 15, then table should be constructed as follows :

Class	Frequency	Age Cumulative frequency	Explanation
0 – 5	72	0 and more 50	250 = 250
5 - 10	103	5 and more 78	250 - 72 = 178
10 - 15	50	10 and more 75	178 - 103 = 75
15 - 20	25	15 and more 25	75 - 50 = 25