

Representation of data

The first and foremost duty of a statistician is to collect raw data from the investigation field. The data stay scattered in the field. Nothing can be realized about the nature of data and cannot be taken any decision. If the area of the primary data is large, statistical method is difficult to use. If it is possible, it consumes time and cost of money. So it needs to condensation. To know the characteristics of data, generally they have to make condensation by statistical method. The collective data should be representative of all the data. The condensation process of data collection is as follows:

- (i) Classification
- (ii) Tabulation
- (iii) Frequency distribution

Classification:

- Classification is the grouping of related facts into different classes.
- Facts in one class differ from those of another with respect to some characteristics is called a basis of classification.

Representation of data

The first and foremost duty of a statistician is to collect raw data from the investigation field. The data stay scattered in the field. Nothing can be realized about the nature of data and cannot be taken any decision. If the area of the primary data is large, statistical method is difficult to use. If it is possible, it consumes time and cost of money. So it needs to condensation. To know the characteristics of data, generally they have to make condensation by statistical method.

The collective data should be representative of all the data. The condensation process of data collection is as follows:

- (i) Classification
- (ii) Tabulation
- (iii) Frequency distribution

Classification:

- Classification is the grouping of related facts into different classes.
- Facts in one class differ from those of another with respect to some characteristics is called a basis of classification.

types of classification:

- (1) Geographical i.e Area wise e.g cities, districts etc.
- (2) Chronological i.e on the basis of time
- (3) Qualitative i.e According to some attributes
- (4) Quantitative i.e in terms of magnitude

Geographical classification:

- Data are classified on the basis of geographical location.
- production of rice for various districts, this would be called geographical classification.

<u>District's Name</u>	<u>Production</u>
Dhaka	240
Rajshahi	200

chronological classification:

- Observed over a period of time.

Ex Sales figure of a company are

<u>Year</u>	<u>Sales Taka (Lakhs)</u>
1981	580
1982	612
1990	1000

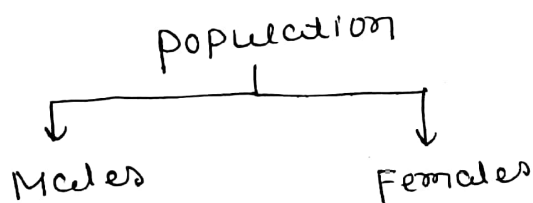
Qualitative classification:

- Data are classified on the basis of some attribute or quality
- Sex, colour of hair, literacy, religion etc
- Study cannot be measured
- One can only find out whether it is present or absent in the units of the population under study.

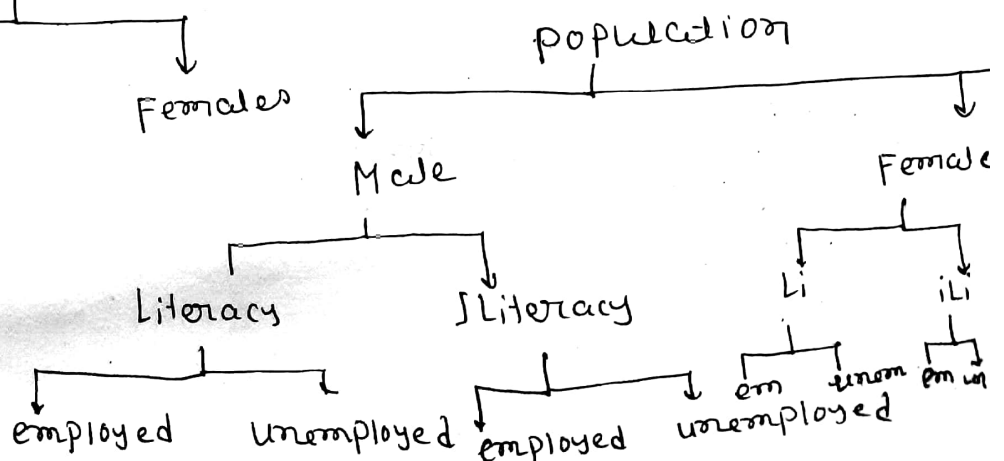
Two types:

Simple classification

- Only one attribute is studied two classes are formed
- population under study may be divided into two categories



Manifold classification
→ Instead of forming two classes, we further divide the data on the basis of some attributes so as to form several classes.



Quantitative classification:

Quantitative classification refers to the classification of data according to some characteristics that can be measured. Such as height, weight, income, sales etc.

<u>Monthly wages</u>	<u>Number of cookers</u>
2000 - 2500	50
2500 - 3000	100
3000 - 3500	200

Tabulation:

- Tabulation is the process of condensation.
- It is applied in summarizing data and presenting them in meaningful fashion in the statistical table.
- Table is a systematic arrangement of statistical data in column and rows.

Parts of table

- Table Number
- Title of the table
- Caption
- Stub
- Body of the table
- Footnote
- Source note

Title
 Sub-title
 Explanatory notes

Sub Head	Caption		column Head	
	column Head			
	Sub column Head	Sub-column Head	Sub column Head	Sub column Head
Sub Entries	B	O	D	Y

Footnote —

Source-note —

Frequency :

the number of item falling within the limits of a class interval is known as frequency of that class and called class frequency. Frequency is in general, the number of occurrences of the items.

For example, 15 Observations within the range 10-14. Then the frequency is 15 of the class 10-14.

Frequency distribution:

If we divide the observation into classes with certain characteristic of any population, showing the corresponding class frequencies with some other required facts, in a tabulated form is called frequency distribution.

The following are the two examples of discrete and continuous frequency distribution.

<u>Number of children</u>	<u>Number of Families</u>	<u>Age (year)</u>	<u>Number of employee</u>
0	10	20-25	10
1	10	25-30	15
2	200	30-35	17
3	250	35-40	40
4	120	40-45	15

Formulation of frequency distribution:

- (i) Range
- (ii) Number of class
- (iii) class intervals
- (iv) class limits
- (v) classifying data
- (vi) class frequency and tally marks

Problem: The statistics marks of 100 students are given below. prepare a frequency distribution taking a suitable class interval from this raw data

53, 32, 59, 73, 56, 97, 90, 82, 33, 41, 50, 42, 55, 62, 63,
39, 81, 71, 77, 78, 22, 48, 46, 44, 33, 61, 66, 62,
63, 64, 53, 60, 50, 72, 67, 98, 83, 85, 68, 69,
45, 22, 21, 27, 37, 67, 65, 64, 88, 63, 47, 58,
59, 60, 71, 72, 73, 80, 75, 64, 61, 35, 49, 46,
73, 52, 66, 92, 64, 84, 96, 47, 57, 69, 64,
74, 65, 70, 76, 48, 72, 93, 41, 27, 95, 2,
61, 67, 66, 76, 23, 68, 84, 60, 87, 35, 68,
67, 32, 61, 69

Solving The highest value of the observation is 98. Lowest value is 21. $\text{Range} = 98 - 21 = 77$. If we divide range by 5, the quotient is found to be 15.40 and divided by 15, quotient will be 3.08. We can take a convenient value 10 within the range 3.08 to 15.40. We take the class interval

$$\begin{aligned}\text{Number of class} &= \frac{\text{Range}}{\text{class interval}} \\ &= \frac{77}{10} = 7.7 \approx 8\end{aligned}$$

OR

By the Sturges formula,
we get the number of classes

$$\begin{aligned}k &= 1 + 3.322 \log_{10} N \\ &= 1 + 3.322 \log_{10} 100 \\ &= 7.644 \approx 8\end{aligned}$$

$$\begin{aligned}\text{class interval} &= \frac{\text{Range}}{\text{Number of class}} \\ &= \frac{77}{7.644} \approx 10.07 \approx 10\end{aligned}$$

Frequency distribution with inclusive method:

Class Limit	Tally	Frequency
20-29		6
30-39		8
40-49		12
50-59		12
60-69		32
70-79		15
80-89		9
90-99		6

Frequency distribution using exclusive method:

Frequency distribution with exclusive method:

<u>Class interval</u>	<u>Tally</u>	<u>Frequency</u>
20-30		6
30-40		8
40-50		12
50-60		12
60-70		32
70-80		15
80-90		9
90-100		6

Ex The statistics marks of 40 students are given below. prepare a frequency distribution table taking a suitable class interval from this raw data

53, 32, 59, 73, 56, 97, 90, 82, 33, 41, 50, 42, 55, 62, 63, 39, 81, 71, 77, 78, 22, 48, 46, 44, 33, 61, 66, 62, 63, 64, 53, 60, 50, 72, 67, 98, 83, 85, 68, 69

Soln: the highest value of the observation is 98 and the lowest value is 22.

$$\text{Range} = 98 - 22 = 76$$

If we divide range by 5, the quotient is 15.2 and divided by 15, quotient will be 5.06. We take a convenient value 10 within the range 5.06 to 15.2. The class interval = 10

$$\begin{aligned}\text{Number of class} &= \frac{\text{Range}}{\text{Class interval}} \\ &= \frac{76}{10} = 7.6 \approx 8\end{aligned}$$

By Sturges formula, we get the number of classes

$$k = 1 + 3.322 \log_{10} 40 = 6.322 \approx 7$$

$$\text{class interval} = \frac{\text{Range}}{\text{Number of class}}$$

$$= \frac{76}{6.322} = 12.02 \approx 12$$

Frequency distribution using inclusive method:

Class limit	Tally	Frequency
20-31	I	1
32-43		6
44-55		8
56-67		11
68-79		7
80-91		5
92-103		2

By exclusive method:

<u>Class limit</u>	<u>Tally</u>	<u>Frequency</u>
20-32	I	1
32-44		6
44-56		8
56-68		11
68-80		7
80-92		5
92-104		2

H.W.

- ① The monthly earnings of various employ in a company in thousands are given below. Make a frequency distribution table from this raw data using exclusive and inclusive method:

70, 74, 82, 90, 110, 132, 142, 133,

116, 176, 123, 145, 102, 190, 186,

167, 175, 162, 164, 187

- ② Construct a frequency distribution table using these raw data

128	100	180	150	200	90	340	105	85	270
200	65	230	150	150	120	130	80	230	200
110	126	170	132	140	112	90	340	170	190

100

2e Cumulative frequency:

The cumulative frequency is given in the following table;

The cumulative frequency of a class is the sum of the frequencies of that class and all previous classes. The cumulative frequency of the last class is equal to the sample size.

<u>Class Limit</u>	<u>Frequency</u>	<u>Cumulative frequency</u>
20-32	1	1
32-44	6	7
44-56	8	15
56-68	11	26
68-80	7	33
80-92	5	38
92-104	2	40

Class boundary;

If the grouped frequency distribution is not continuous, we first convert it into continuous distribution as follows:

<u>Class Limit</u>	<u>Class boundaries</u>	<u>Frequency</u>
5-9	4.5-9.5	5
10-14	9.5-14.5	10
15-19	14.5-19.5	17
20-24	19.5-24.5	22
25-29	24.5-29.5	30
30-34	29.5-34.5	33
35-39	34.5-39.5	39

28.5/14
28
8