

Lesson-6

Collection and classification of Data

Summary

Collection and classification of data provides the basis to understand the methods of data collection and also how to arrange them for further statistical analysis. In this chapter we will discuss about methods of collecting primary data, secondary data, arranging raw data for the purpose of utilizing various statistical tools with a view to find out conclusion of statistical enquiry. Here, we will also learn about various terms like – arrays, frequency arrays, class interval, tally marks, frequency, mid-point, etc. Besides these, we will be aware about types and method of constructing different types of statistical series.

Sources of collection of Data

a) Primary Data and

b) Secondary Data

Primary Data- Primary data refers to the first hand data gathered by the investigator himself/herself.

Methods for collecting primary data-

- Direct Personal Interview
- Indirect Oral Investigation
- Questionnaire Method

Secondary Sources of Data- Secondary data means data collected by someone else earlier. It has two sources of data- a) Published and b) Unpublished.

Sources of Secondary Data-

Published Sources of data

- Published Sources like journals or reports.
- Data published by Central Statistical Organization (CSO)
- National Sample Survey Organization, under the Ministry of Finance publishes statistical data on National Income, Saving, Capital Formation etc
- Reserve Bank of India publishes data on currency and finance also Statistical Table related to Banking Sector.
- Labour Bureau publishes Indian Labour Statistics, Indian Labour Year Book, Indian Labour Journal.
- Population Census of India under the Ministry of Home Affairs publishes

- Per capita income, literacy rate etc.
- Papers and Magazines are also effective published sources which provide sufficient statistical information and data .

Unpublished Sources of Data

- Secondary data are also available from unpublished sources because some organization do not published statistical data.

Organising and Condensing data

- It is a peocess of arranging and grouping the data in some meaningful arrangement for further study of statistical analysis. Data can be arranged in two ways- a) Arrays and b) Frequency distribution.

Arrays

- An orderly arrangement of raw data in individual series is called arrays. It can be of two types- a) Simple array snd b) Frequency arrays
- **Simple array** - A simple arrays is an arrangement of data in ascending or descending order.
- **Frequency Arrays** - It is a series in the formed on the basis of frequency with which each item is repeated in series.

The main steps in constructing frequency array are -

- Prepare a table with columns-first for value of items, second for tally marks and third for corresponding frequency.
- Put the item in first column in ascending order in such a way that one item is recorded once only.
- Prepare a tally sheet in second column making one bar for each item. Make block of five tally bars and every fifth bar is shown by crossing the previous four bars like e.g. |||| .

Frequency Distribution

A frequency distribution classifies the data into group. It is necessary to learn the following important concept for preparing frequency distribution table.

Class - It is a group of magnitude shaving two ends called class. For example- 20-25, 25-30 etc or 20-24, 25-29etc.

Class Limit - Every class has two boundaries or limit called lower limit (L_1) upper limit (L_2)

Class Interval - The difference between two limits of a class is called class interval.

Class Interval = $L_2 - L_1$ or $30 - 20 = 10$.

Cumulative Frequency Class

Frequency – Total no of items falling in a class that is having value within L_1 and L_2 is known as class interval.

Mid- point – It is obtained by dividing sum of lower limit and upper limit by 2.

$$\text{Mid- point} = \frac{L_2 - L_1}{2}$$

Construction of Frequency Distribution

It can be constructed in many ways -

- Exclusive Series
- Inclusive Series
- Open end classes
- Cumulative frequency

Exclusive Series – In this type the upper class limit is generally excluded while making tally sheet. The upper limit of a class interval becomes lower limit of next class interval and then frequency of the value of that upper limit is counted.

Class	Tally Sheet	Frequency
20-25	/	1
25-30	///	3
30-35	////	5
35-40	//// //	7
40-45	//// ////	10
45-50	//// ///	8
50-55	////	4
55-60	//	2
		Total=40

Inclusive Series – In this type lower limit is increased by one over the upper limit of previous class. Both the limit of class interval are counted to mark tally sheet and frequency.

Class	Tally Sheet	Frequency
20-24	/	1
25-39	///	3
30-34	////	5
35-49	//// //	7
40-44	//// ////	10
45-49	//// ///	8
50-54	////	4
55-59	//	2
		Total=40

Open End Classes – Open – end frequency distribution is one which has at least one of its end open. It is observed that either lower limit of first class or upper limit of last class or both are not given in such series.

Class	Tally Sheet	Frequency
20-24	/	1
25-39	///	3
30-34	////	5
35-49	//// //	7
40-44	//// ////	10
45-49	//// ///	8
50-54	////	4
55-59	//	2
		Total=40

Cumulative Frequency – A cumulative frequency distribution is formed by taking successive totals of a given frequencies.

Cumulative frequency distribution is of two types – a) Less - than and b) More – than.

Less – than Cumulative Frequency Distribution

Class	F.	C. F.
Less than 25	1	1
Less than 30	3	4(3+1)
Less than 35	5	9 (4+5)
Less than 40	7	16 (9+7)
Less than 45	10	26 (16+10)
Less than 50	8	34 (26+8)
Less than 55	4	38 (34+4)
Less than 60	2	40 (38+2)

F. – Frequency

C. F. – Cumulative Frequency

More – than Cumulative Frequency Distribution

Class	F.	C. F.
More than 20	1	40
More than 25	3	39(40-1)
More than 30	5	36(39-3)
More than 35	7	31(36-5)
More than 40	10	24(31-7)
More than 45	8	14(24-10)
More than 50	4	6(14-8)
More than 55	2	2(6-4)

F. – Frequency

C. F. – Cumulative Frequency

Evaluate Yourself

Q.” Primary data provide more reliable information than secondary sources of data”. Why?

Q. Distinguish between exclusive method and inclusive method of data collection.

Q. Prepare less-than and more-than frequency distribution table of below given data

Expenditure (in Rs.)	No. Of Family
0-10	14
10-20	23
20-30	27
30-40	21
40-50	15

Q. Give a brief description of sources of secondary data collection.