

Chapter 2: Collection of data

2.1. Principal of collection

Statistical data are collected from different sources and different methods are adopted to collect adequate and reliable data. These data are collected to conduct some inquiries or to analyse some problems. Various aspects relating to statistical inquiries and collection of data are the subject matter of this chapter. The business organizations keep the records of their routine transactions, development plans and research reports. Similarly, the Government departments maintain the records regarding the exports, imports, taxes, Government expenditure, births and deaths, price index numbers etc,... Those records are the main sources of data.

2.2. Data Presentation

Usually the data are collected in tables, which summarize the given information such as:

Example 1.

In January 1995, a firm employed 90 staff whom 79 were men. During the year, 17 staff left and 13 of these were men. The total recruitment during the year was 13 out of whom 3 were women.

The number of men

During 1996 wastage declined by 3 amongst men compared with 1995 and no women left. 16 more men but 2 fewer women were recruited than in the previous year. The total number employed on 1st January 1996 amounted to 93. Represent those data in a table. For that, arrange the above information in concise tabular form showing all relevant totals and sub-totals.

	1995			1996		
Employees at January	Men	Women	Total	Men	Women	Total
	79	11	90	76	10	86
Recruited in the year	10	3	13	16	1	17
Left in the year	13	4	17	10	0	10
Total	76	10	86	82	11	93

Example 2.

The following report was prepared by an Examinations Officer on the performance of X District in a National Examination.

Out of 3,500 male candidates below 20 years of age, 500 passed and 3,000 failed. Out of the 1,100 male candidates, 20 years old and over 200 passed and 900 failed. As regards the female candidates, out of 500 below 20 years of age, 100 passed and 400 failed. Of the 340 females, 20 years old and over, 80 passed and 260 failed. Present the above information in a tabular form

Solution: National Examination; **X District**

		Passed	Failed	Total
< 20 Years	Men	500	3000	3500
	Women	100	400	500
>= 20 Years	Men	200	900	1100
	Women	80	260	340
Total		880	4560	5440

These tables show how to collect the data and summarize them in a table like shown above, and we can easily read all information through the given table.

2.2. Types of data

There are different types of statistical data, which are ungrouped, grouped, discrete and continuous data with frequencies. When the data are few, they can be presented as ungrouped data, in that case the data are called simple and their variation Marge is less than 15.

Example 3: Representation of results in table of frequency:

18	17	18	17	14	12	13	8	9	13	15	16
7	15	14	16	20	16	13	13	15	17	16	10
16	18	19	14	16	16	13	14	9	17	15	12
17	17	16	13	14	7	13	16	14	15	18	16
13	19	17	16	14	18	12	20	16	14	13	18
20	13	12	15	12	15	18	17	18	20	11	15
17	14	15	18	14	9	11	14	18	16	13	18
12	15	14	16	18	13	16	16	15	17	19	15

The Marge of variation or variation Marge (VM) is found by using the following formula: $VM = (X_H - X_B) + 1 = (20 - 7) + 1 = 14 < 15$, in the above example. Then, the data can be presented in form of simple data as follows:

X	Tally	F	X^2	fX^2	Observations
7	II	2	49	98	Complete these table yourselves, which will help you to calculate some parameters.
8	I	1	64	64	
9	III	3	81	243	
10	I	1	100	100	
11	II	2	121	242	
12	IIII I	6	144	864	
13	IIII IIII II	12	169	2028	
14	IIII IIII II	12	196	2352	
15	IIII IIII II	12	225	2700	
16	IIII IIII IIII I	16	256	4096	
17	IIII IIII	10	289	2890	
18	IIII IIII II	12	324	3888	
19	III	3	361	1083	
20	IIII	4	400	1600	
		96			

As the number of variables are 14 and less than 15, our variables are considered as simple data, therefore as ungrouped data.

Let us use another activity of data and calculate the Marge of variation, which helps us to see if our data can be ungrouped or grouped into class intervals.

Example 4: Consider the table below of data collected on the recruitment test in a firm.

26	36	44	42	38	48	42	44	46
30	38	50	48	46	48	50	50	50
32	38	50	40	40	44	42	46	42
34	36	48	38	40	50	44	50	48
34	40	46	50	50	46	40	44	48
48	50	54	52	52	54	48	50	52
46	54	52	54	56	52	52	52	52
52	60	60	54	56	58	56	54	56
58	62	56	62	58	60	54	48	60
64	50	64	66	62	64	62	66	70
48	58	44	46	42	48	50	56	60

76

Construct the frequency table. Solution to the table 3.

In some cases, the values of variables are large and are not possible to represent the table with details; therefore, the results are grouped in class intervals by using the following method:

Look for the extremes values of variable (the largest and smallest) and calculate the number of possible values of variable, otherwise, calculate the Marge of variation between two extreme values or numbers. $(X_H - X_B) + 1 = (76 - 26) + 1 = 51$

Why plus 1 at the difference $(X_H - X_B)$? Suppose a distribution of results whose $X_H = 12$ and $X_B = 8$, the values of variable are: 8, 9, 10, 11, 12 and are five (5) but the difference $X_H - X_B = 12 - 8 = 4$, so it is necessary to add a value (+1) at the difference $(X_H - X_B)$ to get the number of possible values of variable.

The number of classes are chosen (normally between 10 and 20). In practical: 15 divide the Marge of variation; number lies between 10 and 20, numbers of minimum and maximum of classes, and the quotient determine the length or width of class. $51 : 15 = 3.4$ one of the two odd numbers that approach that quotient is chosen to determine the elements in each class interval. Here 3 or 5 are

both odd numbers which approach the obtained quotient. We take 5 in way to use a few number of class intervals, which help us in calculation.

The classes are built: The large number in the table is placed at the middle of the higher class, then $X_H=76$, the high class will be now: 74 – 75 – 76 – 77 – 78 or 74 – 78

The tally is made and the numbers, which fall in each class, are found. In addition, the mid- points or mid-values are determine, the central point or mid - point of a class is the average mean of its limits, the two extreme values.

$$X_1 = \frac{24 + 28}{2} = 26 \quad ; \quad X_2 = \frac{29 + 33}{2} = 31, \dots$$

The length or width is found by using the difference of the Consecutive mid-point of two classes that are such that: $c = 36 - 31 = 31 - 26 = 5$.

Class-Intervals	Tally	Frequencies(f_i)
24 – 28	I	1
29 – 33	I I	2
34 – 38	IIII III	8
39 – 43	IIII IIII	10
44 – 48	IIII IIII IIII IIII III	23
49 – 53	IIII IIII IIII IIII II	22
54 – 58	IIII IIII IIII III	18
59 – 63	IIII IIII	9
64 – 68	IIII	5
69 – 73	I	1
74 - 78	I	1
		$\sum f = N = 100$

Feedback

To answer this question, you should be able to complete your reading with other materials. And if you are not able to do it, then consult your tutor and discuss with your group classmates.