Chapter: 3

FREQUENCY DISTRIBUTION

INTRODUCTION

The data collected with some definite purpose in any field of enquiry is large. As such it will not be possible to visualize the main features or characteristics of the data easily. It is difficult to take decisions by observing large data. If large data is simplified or condensed and presented in the form of tables and diagrams it is possible to take decisions about the subject of experiment. This process is known as data presentation.

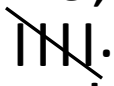
Thus to make the main features of the data clearer and more meaningful we group the data into ordered groups (called intervals). These grouping result in a table called the frequency table because it shows groups and frequency of scores within each group. By constructing a frequency table for the data and observing it, we can know many features of it. The behavior of the variable , the underlining characteristics of the data will be clearly known from the frequency distribution of the data.

While preparing frequency distributions the following points should be taken care of:

1. **Classes should be non-overlapping.**
2. **There should be no gaps between the classes.**
3. **As far as possible the classes should be of the same size.**
4. **Classes should be clear. There should not be any ambiguity.**
5. **The number of classes should be neither very large nor very small (i,.e.,5 to 16)**
6. **Every score of the data should find a place in one class or the other distinctively.**

PREPARATION OF A FREQUENCY DISTRIBUTION

1. **The difference between the highest and lowest scores of data is called the range. First the range of the data is found . Class interval is taken depending on the nature of the data.**
2. **The range is divided into a number of equal classes. The number of classes should be neither too large nor too small. We commonly use from 5 to 16.**
3. **After we determine the class interval and the number of classes the next step is to set up the intervals. Classes are written with the lowest score of the data of any number smaller than that.**

4. The next step is to find the number of scores in each interval. To do this we take each score from the data and place a tally mark opposite the class in which the score lies. For convenience of counting and for the sake of elegance we record tally marks in bundles of five , the fifth one crossing diagonally the other four like  **. After the tally marks are drawn corresponding to each score of the data, the tally marks against each class are counted and the number is written against that class in the next column of the table. This number of scores in the data as the total should be equal to the number of scores in the raw data.**

Example:1 Let us consider the following data.
The percentage of marks scored by 30 students of the 10^{th} class in mathematics is as follows:

2	**50**	**0**	**4**	**6**	**12**	**6**	**15**	**16**	**17**
20	**25**	**23**	**27**	**28**	**29**	**30**	**34**	**40**	**43**
40	**50**	**55**	**54**	**56**	**60**	**64**	**69**	**70**	**12**

Construct a frequency table for the above data.

Sol: The frequency table of marks of 30 students given in the example is as follows:

Frequency Table/ Distribution

Class	Tally marks	Frequency
0-9	𝍸	5
10-19	𝍸	5
20-29	𝍸 |	6
30-39	||	2
40-49	|||	3
50-59	𝍸	5
60-69	|||	3
70-79	|	1
	Total	30

In this table each class is represented by two numbers. The first number is called its lower limit and the second number is called its upper limit. The lower limit of the class 0-9 is 0 and its upper limit is 9. Similarly, the lower limit of class 50-59 is 50 and its upper limit 9s 59.

The difference between the lower limits or the upper limits of two consecutive classes is called the class interval or size of the class.

Example:2 Construct a frequency table for the following data taking 4 as the size of the class interval.

15	17	18	19	20	21	22	23	24	25
24	23	21	20	25	27	28	29	29	29
30	30	29	30	29	30	31	33	35	37
27	39	41	43	45	44	45	30	39	40

Sol: Range = 45 – 15; Size of the class interval =4

Number of classes = $\frac{Range}{Class\ interval} = \frac{30}{4}$ = 7.5 ~ 8

Note: If the number of classes is in the form of a mixed fraction the next higher integer is taken as the number of classes.

Frequency Distribution

Class	Tally marks	Frequency
15-18	|||	3
19-22	𝍸 |	6
23-26	𝍸 |	6
27-30	𝍸 𝍸 ||	12
31-34	||	2
35-38	|||	3
39-42	||||	4
43-46	||||	4
	Total	40

Sometimes we may find it difficult to decide the class to which a particular item of the data belongs to, for example, we cannot decide to which of the classes 15-18, 19-22 , the children whose ages are between 18 and 19 years belong to. This kind of data is called *continuous data*.

A variable which takes only specific value is called a *discrete variable*. The data pertaining to discrete variable is called *discrete data*.

When a frequency table is to be prepared for a continuous data, care is to be taken that every item of the data pertains to some class or the other. This is possible if the upper limit of a class is equal to the lower limit of the next class. For example , in the above table the average of the upper limit of the first class and the lower limit of the second class is $\frac{18+19}{2}$=18.5 . Similarly , the average of 22 and 23 is 22.5.

Now the classes can be written as 14.5-18.5, 18.5-22.5, 22.5-26.5, 26.5-30.5, 30.5-34.5, 34.5-38.5, 38.5-42.5, 42.5-46.5.

The present limit of each class are called *lower* and *upper boundaries*. In this case , the upper boundary of a class becomes the lower boundary of the next class. These boundaries are called *true class limits*.

When the classes are arranged with true class limits an item of the data equal to these boundaries is put in the next class. This is followed as a convention. For example , an item of the data with value 30.5 is considered to belong to the class 30.5-34.5, but not to the class 26.5-30.5.

Example:3 The maximum day temperature in May 1974 in Bombay is as follows:

32.5	30.5	33.8	31.0	28.6	33.9	33.3	32.4
30.4	32.6	34.7	34.9	31.6	35.2	35.3	35.5
36.4	36.6	37.0	34.3	32.5	31.4	34.4	35.6
35.3	37.5	36.9	37.0	36.3	36.9	36.9	36.7

Construct a frequency table for the data.

Sol: Range = 37.5 – 28.6 = 8.9; Size of class interval =2; Number of classes = 5

Frequency Table

Class	Tally marks	Frequency
28.6-30.6	|||	3
30.6-32.6	𝍸 ||	7
32.6-34.6	||||	4
34.6-36.6	𝍸 𝍸 |	11
36.6-38.6	𝍸 ||	7
	Total	32

Classes like 0-9, 10-19 , 20-29,… are called *inclusive classes* and classes like 0-10, 10-20, 20-30,.. are called *exclusive classes*. In the inclusive classes the upper limit of one class is included in that class itself whereas in the exclusive classes the upper limit of one class is the lower limit of the next class. In the exclusive classes the limits of the classes are true limits of the classes. In the exclusive method the difference between the upper and lower limits of a class represents the class interval or size of the class.

CUMULATIVE FREQUENCY DISTRIBUTION

Frequency distribution helps us to find out the number of items in each class and the nature of frequencies but it does not help to tell the number of items from the beginning upto and including a particular class. Sometimes we may need to find the total frequency of item greater or less than a particular item of the data. In such cases cumulative frequency distribution helps us in finding the number of items from the beginning upto and including any particular class. Cumulative frequency distributions are of two kinds.

1. **Less than cumulative frequency distribution.**
2. **Greater than cumulative frequency distribution.**

1. Less than Cumulative Frequency Distribution

The number of items from the beginning in the upper boundary of a particular class of a frequency distribution is called the less than cumulative frequency.

or

Less than cumulative frequency for any value of the class is obtained on adding successively the frequencies of all the previous classes including the frequency of a class against which totals are written provided the classes are arranged in ascending order of magnitude.

Example:4 Convert the following series into less than cumulative frequency distribution.

Class	Frequency
50-59	30
60-69	25
70-79	20
80-89	11
90-99	8
total	94

Sol: The above series in the less than cumulative frequency distribution can be written as follows:

Class	Frequency	Upper boundary of the class	Less than cumulative frequency
49.5-59.5	30	Less than 59.5	30
59.5-69.5	25	Less than 69.5	30+25 =55
69.5-79.5	20	Less than 79.5	20+30+25=75
79.5-89.5	11	Less than 89.5	11+20+30+25=86
89.5-99.5	8	Less than 99.5	8+11+20+30+25=94

2. Greater than (OR) More than Cumulative Frequency Distribution

The number of items from the end to the lower boundary of a particular class of frequency distribution is called greater than cumulative frequency.

Or

The more than cumulative frequency is obtained similarly by finding the cumulative totals of frequencies starting from the highest value of the class to the lowest class.

Example:5 Convert the series given in example:4 into more than cumulative frequency distribution.

Sol. The given data can be converted into more than the cumulative frequency distribution as follows.

Class	Frequency	Upper boundary of the class	Less than cumulative frequency
49.5-59.5	30	More than 49.5	8+11+20+25+30=94
59.5-69.5	25	More than 59.5	11+20+25+30=86
69.5-79.5	20	More than 69.5	20+25+30=75
79.5-89.5	11	More than 79.5	25+30 =55
89.5-99.5	8	More than 89.5	30 = 30

END