Organisation of Data

Raw Data, Classification of data: Meaning objective and Methods

Objective

After going through this lesson, you shall be able to understand the following concepts.

- Classification of Data- Meaning
- Objectives of Classification
- Features of Classification
- Methods of Classification- Chronological Classification, Geographical Classification, Qualitative Classification and Quantitative Classification

Introduction

In the previous lesson, we studied how the data is collected through different methods. After the collection of data, the next step in any study or research is to classify or organise the data in a systematic and proper manner.

You all must have seen a grocery store. It offers a large number of items for sale. However, you must have noticed that the different items in the store are properly organised and classified in different categories.

What would happen if all these items were placed in a huge heap? It would be very difficult for the seller to find the item which the customer demands from the heap. Similarly, the buyer would find it difficult to choose the item he wants.

Thus, classification of items in the store eases the process of sale and purchase. Similarly, the data collected by a researcher is in raw form. In other words, it is in an unorganised and unclassified form.

From such a data no meaningful conclusions can be drawn. For example, suppose data for the income of 20 people in a locality is collected. The data collected is as follows.

8190	7598	7389	32683
10310	4326	12368	18266
14666	22046	13000	9500
12937	43598	25650	25550

15000	21967	17350	11948

This data is nothing but an unsorted or an unclassified set of numbers. Extracting any information from such a data can be quite difficult. For instance, if we need to locate the highest or the lowest income in the group then, we have to go through the entire set of numbers.

This can be tedious, especially in the cases where the data is quite large. The task can be simplified if the data is arranged in the ascending or descending order. In other words, organisation or classification facilitates analysis of the data. Thus, it can be said that the collected data needs to be classified or organised before being used for further analysis.

Classification of Data- Meaning

Classification of data refers to the arrangement of data according to its characteristics. Under such an arrangement of data, similar or homogenous data are grouped together. For example, students in a college can be classified on the basis of their age, the cities to which they belong, their course subjects etc.

Objectives of Classification

Some of the major objectives of classification of data are as follows.

(i) ***Arrangement of data in the simplest form***: The main objective of classification of data is to present the raw, unorganised data in a simple and systematic form so as to make it understandable. The classification of data arranges the data in the simplest form. With such simple arrangement, it becomes easier to derive conclusions from the data.

(ii) ***To ease comparisons***: Classification makes comparison among different characteristics of the data easy.

(iii) ***To draw inferences of relationship***: Classified data helps in drawing out cause and effect relationship between two or more variables.

(iv) ***Facilitates further statistical analysis***: For any further statistical analysis, the raw needs to be properly classified. An unclassified data cannot be used for analysis and drawing out conclusions.

(v) ***Drawing out similarities and dissimilarities***: By grouping together similar or homogenous data, similarities or dissimilarities between different variables can be easily drawn.

(vi) *Makes data more attractive*: A classified data is far more attractive and effective than raw data.

Features of Classification

A proper classified data must possess certain essential characteristics. The following are the important characteristics of classification of data.

(i) *Clear and unambiguous*: The classification of raw data should be absolutely clear and unambiguous. There should not be any confusion in the placement of items in different groups or categories.

(ii) *Complete*: The classification should be complete in all respect. This means that, every item in the data should get placed in some or the other group. In addition, caution must be taken that there should be no room for placing an item in more than one group.

(iii) *Suit the objective*: The data must be classified in a manner so as to suit the objective and the purpose of the study.

(iv) *Flexibility*: The classification must be flexible so as to facilitate change according to the need of the study.

(v) *Stable*: The classification should be stable and a particular investigation should be based on a particular classification.

Methods of Classification

Classification of data can be done in the following four manner.

(i) Chronological Classification

(ii) Geographical Classification

(iii) Qualitative Classification

(iv) Quantitative Classification

(i) *Chronological Classification*: The classification of data on the basis of their occurrence at different time period such as year, month, week, etc. is called chronological classification. The data can be arranged in ascending or descending order of time whether in years, months, weeks
etc.

Example-

Sales of Firm A

Year	Sales (Rs Cr.)
2001	100
2002	110
2003	120
2004	125
2005	127

Here, the data for sales is classified for different years. For instance, in the year 2001, sales was worth Rs 100 crore. Similarly, for the year 2002, the sales was worth Rs 110 crore

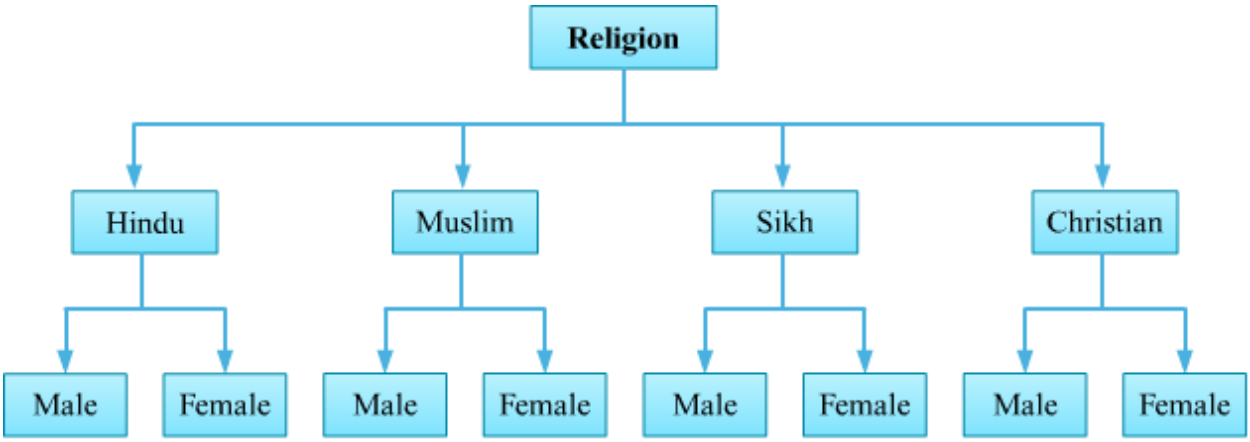
(ii) *Geographical Classification*: The classification of data on the basis of geographical locations such as, states, countries, cities, districts, etc. is called geographical classification.

Example-

Place	Number of Students
Delhi	70
Mumbai	990
Chennai	50
Kolkata	45
Punjab	30
Haryana	20

Here, the students are classified according to the cities to which they belong. For instance, there are 70 students who belong to Delhi. Similarly, 990 students belong to Mumbai and so on.

(iii) *Qualitative Classification*: When the data is classified according to some characteristics or qualities, it is called as qualitative classification of data. For example, the workers in a factory can be classified on the basis of religion as follows.



(iv) *Quantitative Classification*: The classification of data on the basis of certain numerical values into different classes or groups is known as quantitative classification.

Example-

Sales Group (Tonnes)	Number of Firms
0 – 10	2
10 – 20	3
21 – 30	7
31 – 40	4
41 – 50	2

Here, the sales is classified as per the number of firms. Thus, there are three years for which the sales is in the bracket 10 to 20 tonnes. Similarly, there are 7 firms for which the sales is between 21 to 30 tonnes.

Variables and their Classification, Types of Statistical Series

Objective

After going through this lesson, you shall be able to understand the following concepts.

- Variables- Meaning and Types
- Statistical Series- Meaning and Types

Variables- Meaning and Types

A variable refers to any measurable characteristic or phenomena whose value changes overtime.

For example, age, height, weight, prices etc. are variables. Consider the age of a person. It increases with time. Similarly, prices of commodities change. Variables can be classified into two broad categories as follows.

- Discrete Variables
- Continuous Variables

i. *Discrete Variables*: Discrete variables refer to those variables that ***can take only whole numbers as their value***. Such variables cannot take fraction values. In this sense, they are discontinuous. In other words, they jump from one whole value to another.

For example, number of persons in a family, number of students in a class, etc. are discrete variables. They take only whole values. For instance, the number of persons can be 1, 2, 3, 4 and so on. It cannot take any fraction value such as 1.5, 2.5, 1.75

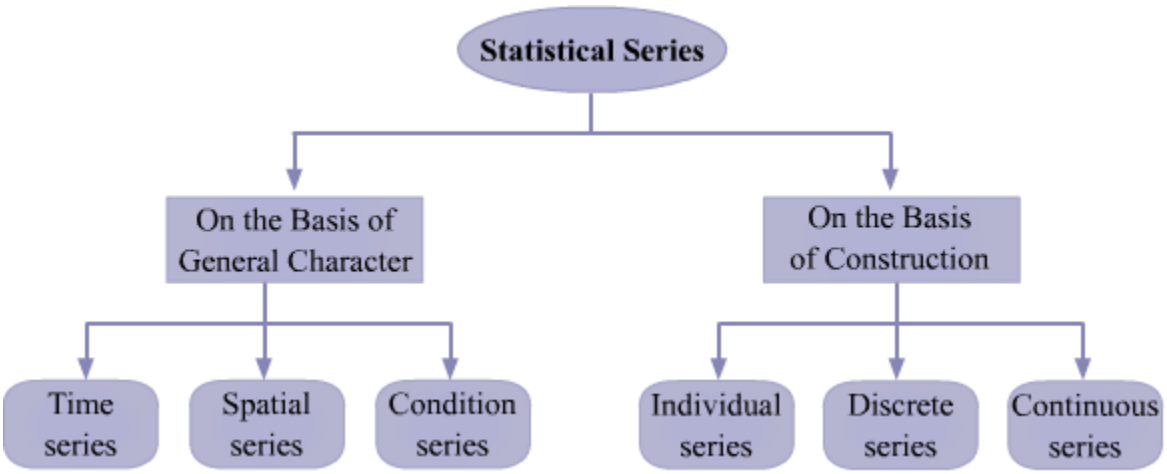
ii. *Continuous Variables*: Continuous variables refer to those variables that can take any value within a reasonable limit. In other words, such variables ***can take both whole number values as well as fraction values***.

For example, age, height, weight, etc. are continuous variables. Consider, weight of a person. It can take values as 2 kg, 2.5 kg, 2.35 kg and so on. As against discrete variables, the values of continuous variables do not increase in jumps rather they increase continuously.

Types of Statistical Series

When the data is presented in some specific order and sequence it is called as a statistical series. In other words, statistical series is the arrangement of a data in a certain systematic order. It can be broadly classified into the following two categories.

- On the Basis of General Characteristic
- On the Basis of Construction



Statistical Series on the Basis of General Characteristics

On the basis of general characteristics, a statistical series can be divided into three broad categories.

- Time series
- Spatial series
- Condition series

(i) *Time Series*: Time series refer to those series where ***time forms the basis of information***. Thus, in such series the variable is measured against different periods of time be it years, months, weeks or days. The following table illustrates a time series. It presents the calorie consumption of a family on different days of a week. Here, days of a week are taken as the basis of reference.

Days	Consumption (in calories)
Sunday	200
Monday	150
Tuesday	180
Wednesday	190
Thursday	160
Friday	210
Saturday	220

According to the series, the family consumed 200 calories on Sunday, 150 calories on Monday and so on.

(ii) *Spatial series*: Spatial series refers to those series where ***geographical area forms the basis of information***. For example, the following table presents the number of children in a school on the basis of their locality.

Locality	Number of Children
A	50
B	40
C	70
D	90
E	60

According to the series, 50 children in the school belong to locality A, 40 children belong to locality B and so on.

(iii) *Condition series*: Under condition series, the data is arranged ***with reference to a physical characteristic*** such as height, weight, etc. The following is a condition series where weight is taken as the reference. It classifies people according to their weight.

Weight (Kgs)	Number of Persons
40-45	25
45-50	17
50-55	22
55-60	28
60-65	30

As per the series, there are 25 persons whose weight lie between 40 Kgs to 45 Kgs. Similarly, there are 17 persons who have weight in the bracket of 45 Kgs to 50 kgs.

Statistical Series on the Basis of Construction

On the basis of construction, statistical series can be divided into three broad categories.

- Individual Series
- Discrete Series
- Continuous Series

(i) *Individual series*: Under this type of series, all items of the data set are expressed individually. For example, a record of marks of students in a class is an individual

series. They are presented simply in serial number order. In addition, items in an individual series can be arranged in either ascending or descending order.

The following is a simple arrangement of marks according to the roll number.

02	10	04	09	15	12	06	12	19	13
14	09	08	17	11	20	11	15	18	13

If, the above data is arranged in *ascending order*, it takes the following form.

02	04	06	08	09	09	10	11	11	12
12	13	13	14	15	15	17	18	19	20

Similarly, if the data is arranged in *descending order* it as follows.

20	19	18	17	15	15	14	13	13	12
12	11	11	10	09	09	08	06	04	02

(ii) *Discrete Series*: Discrete series refer to those series in which the data is presented in the form of exact measurement. Thus, in such series every item takes a definite value. The following is an example of discrete series which presents the marks scored by students of a class.

Marks	Number of Students
5	12
10	17
15	9
20	5
25	11
30	8

According to the series, twelve students score 5 marks each. Seventeen students score 10 marks each and so on. The value corresponding to each item is called its ***frequency***. Here, 12, 17, 9, etc. are frequencies. They show the number of times an item occurs in the series. Thus, 5 occurs twelve times in the data set, 10 occurs seventeen times and so on.

(iii) *Continuous Series*: Continuous series refer to those series in which items assume a range of values and are placed within those ranges (or class intervals). This series is generally used when the data is too large.

The following table presents the marks of students in a class in the form of continuous series.

Marks	Number of Students
0-5	12
5-10	15
10-15	19
15-20	22
20-25	27
25-30	32

According to the series, there are twelve students whose score lie between 0-5. Similarly, there are fifteen students whose score lies between 5-10 marks and so on.

Some Important Terms Related to Continuous Series

The following are some of the important terms related to a continuous series.

(i) *Class*: Class refers to the range of values that incorporates a set of items. In the above example, the classes are (0-5), (5-10) and so on.

(ii) *Class limit*: Class limit refers to the extreme values of a particular class. For example, in class (5-10), 5 is the lower limit and 10 is the upper limit.

(iii) *Magnitude of class interval*: The difference between the upper and lower limit is called the magnitude of that class interval**.**

Algebraically, $i = l_2 - l_1$, where, i = magnitude of class interval, l_2 = upper limit of class interval and l_1 = lower limit of class interval.

(iv) *Mid-Value*: Mid-value of a class is the average value of the upper limit and the lower limit of that class.

Algebraically,

$$\text{Mid-value} = \frac{\text{Upper Limit} + \text{Lower Limit}}{2}$$

For example, mid-value of class interval (10-15):

$$\text{Mid-value} = \frac{10+15}{2} = 12.5$$

Frequency Distribution and its Types

Objective

After going through this lesson, you shall be able to understand the following concepts.

- Construction of Discrete Frequency Distribution
- Construction of Continuous Frequency Distribution
- Ways of Presenting a Continuous Frequency Distribution
- Cumulative Frequency Distribution
- Relative Frequency Distribution
- Bivariate Frequency Distribution

Introduction

In the previous lesson, we studied about the basic characteristics of the different types of statistical series on the basis of construction, namely individual series, discrete series and continuous series. In this lesson, we will try to understand how these series are constructed.

Since the construction of individual series (ascending order and descending order) was dealt with in the previous lesson, we will now try to understand the construction of discrete and continuous series. ***The discrete and continuous series are together known as the frequency distribution series.*** In each of these series, each item corresponds to a particular value that shows how many times that item occurs in the data set. This value is known as its frequency.

Discrete Frequency Distribution

We know that discrete series refer to those series in which data is presented in the form of exact measurement. Now, let us understand the construction of a discrete series (or discrete frequency distribution).

Steps to Construct Discrete Frequency Distribution

The steps involved in the construction of a discrete frequency distribution from a raw data are:

Step 1: Prepare a table comprising of three columns. Represent the variables in the first column; tally bars in the second column and frequencies in the third column.
Step 2: Arrange all items of the first column in a systematic order (i.e. either in ascending order or in descending order).

Step 3: Read each item in the raw data. Put a bar across that value in the tally bars column every time that value occurs in the data. It should be noted that tally bars are

grouped in five. The first four times are represented by simple vertical lines. But if the item occurs for the fifth time, a cross is marked across the first four lines. Now, if the item occurs for the sixth time, a vertical line is again drawn and so on.

Step 4: Corresponding to each item, count the groups of five's in the column of tally bars. This gives the frequency of each item.

Step 5: Record the frequencies in the last column.

Step 6: Resulting table is the discrete frequency distribution table.

Example: The following data gives the income of 24 families in a locality. Present the data in the form of discrete frequency distribution.

Income (in '000)

25	32	9	21	12	16	5	16
12	16	5	14	25	16	25	9
19	11	7	15	16	32	16	9

Solution:

Discrete Frequency Distribution		
Income	**Tally Bars**	**Frequency**
5		2
7		1
9		3
11		1
12		2
14		1
15		1
16		6
19		1
21		1
25		3
32		2
	Total	24

Continuous Frequency Distribution

As we know, continuous series or, continuous frequency distribution refer to those series in which items assume a range of values and are placed within a certain ranges (or class intervals).

In other words, the items from the raw data are grouped together according to the range in which they lie.

Steps to Construct Continuous Frequency Distribution

The following are the steps involved in the construction of a continuous frequency distribution.

***Step* 1**: Determine the range of the data i.e. the difference between the highest and lowest value of the data.

***Step* 2**: Based on the range of the data, determine the number of classes to be taken for the frequency distribution. For example, if the range of the data is 50 and we wish to keep the width of the class as 10, the number of classes would be 5 (i.e. 50 ÷ 10).

***Step* 3**: Determine the class limits. The class limits can be determined using the following two methods.

- Inclusive Method*
- Exclusive Method*

* These two methods are discussed in detail in the next topic.

Step 4: Corresponding to each class, count the number of items that belong to that range. This gives the frequency corresponding to each class.

Step 5: Resulting table is the continuous frequency distribution table.

Inclusive Method: In this method, both the upper and lower limit of a class interval are included in that particular class interval. In such a case, the upper limit of one class does not overlap the lower limit of the next class. Example-

Income
0-4
5-9
10-14
15-19

20-24
25-29
29-34

According to the table, those values of income that lie between zero and four thousand (including both zero and four thousand) are included in the first class. Similarly, all the incomes lying between five thousand and nine thousand are included in the second class interval and so on.

Exclusive Method: In this method, the upper limit of one class becomes the lower limit of the following (next) class. However, it must be noted that in this method, the value of the upper limit of a class is not included in the concerned class interval, rather it is included in the next class.

Example-

The following table presents the income of 20 families in a locality in a continuous series using the exclusive method.

Income (In thousands)	**Number of Families**
0-5	1
5-10	5
10-15	10
15-20	3
20-25	1

In the above example, the family having an income between zero to five thousand will be included in the first class interval. However, an income of five thousand will be included in the second class interval and not in the first class interval. Similarly, the family having income between five thousand to ten thousand are grouped in the class interval (10-15). However, an income of ten thousand is included in the next class interval i.e. 15 to 20.

Note: In exclusive method, one can exclude either the upper limit or the lower limit of the class interval.

Example- Consider the following distribution.

Marks	**Number of Students**
0-5	1
5-10	5

10-15	7
15-20	6
20-5	1

If the lower limit is excluded, we exclude 0, 5, 10, 15 and 20 from the respective classes. These values are included in their previous classes.

For example, 5 is included in the class interval (0-5) and not in the class interval (5-10). Similarly, 10 is included in the class interval (5-10) and not in (10-15).

As against this, if the upper limit is excluded then, we exclude 5, 10, 15, 20, 25 and 30 from the upper limits. In this case, a value of 20 is included in the class interval 20-25 and not in the class interval 15-20.

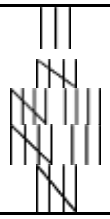
Note: By default, we assume that in the preparation of continuous frequency distributions, the upper limit of the classes is excluded.

Example- Following is the data for marks of 30 students in a class

40	38	37	20	17	35	30	29	20	15
11	10	07	08	15	20	22	49	32	48
30	22	27	09	39	45	47	23	39	24

Prepare a frequency distribution for the data using the class interval 0-10, 10-20, 20-30, 30-40 and 40-50.

Solution

Class Interval (Marks)	**Tally Bars**	**Frequency (Number of Students)**
0-10		3
10-20		5
20-30		9
30-40		8
40-50		5
		Total = 30

Note: In statistical analysis, we generally use the exclusive continuous series rather than the inclusive continuous series. However, if the data given in the question is of the inclusive form, we can covert it in the exclusive form.

Method of Converting Inclusive Series into Exclusive Series

For converting an inclusive series into an exclusive series, adjustments are made in the class interval. For this, the following formula is used.

$$\text{Value of Adjustment} = \frac{\text{Value of the Lower Limit of One Class} - \text{Value of Upper Limit of the Preceeding Class}}{2}$$

The value of adjustment as calculated is then added to the upper limit of each class and subtracted from the lower limit of each class.

Example- Convert the following inclusive series into exclusive series.

Class Interval	Frequency
11-20	9
21-30	12
31-40	7
41-50	8

Solution:

$$\text{Here, the value of adjustment of each class} = \frac{21-20}{2} = 0.5$$

Therefore, we add 0.5 to the upper limit and subtract 0.5 from the lower limit of each class.

Class Interval	Frequency
10.5-20.5	9
20.5-30.5	12
30.5-40.5	7
40.5-50.5	8

Ways of Presenting a Continuous Frequency Distribution

A continuous frequency distribution can be presented in different ways. The following are the four ways of presenting a continuous frequency distribution.

- By Presenting Only the Lower Limits
- By Presenting Only the Upper Limits
- With Open Ended Class Interval

- With Mid-Points

(i) *By presenting only the lower limits*

Consider the given frequency distribution.

Income	New C.I.
0-	0-5
5-	5-10
10-	10-15
15-	15-20
20-	20-25
25-	25-30

Here, only the lower limits of the classes are given. However, as we know that in an exclusive series, the upper limit of one class becomes the lower limit of the next class, the class intervals become 0-5, 5-10, 10-15 and so on.

(ii) *By presenting only the upper limits*

In the following illustration, only the upper limits are given.

Income	New C.I.
-5	0-5
-10	5-10
-15	10-15
-20	15-20
-25	20-25
-30	25-30

Here, to determine the lower limit we again follow the fact that in an exclusive series, the upper limit of one class becomes the lower limit of the next class. So, the class intervals become 0-5, 5-10, 10-15 and so on.

(iii) *With open ended class interval*

An open ended continuous frequency distribution is the one where lower limit of the first class and the upper limit of last class are not explicitly mentioned.

In such cases, the first and the last class interval is determined on the basis of the

remaining class intervals. The following is an open-ended frequency distribution.

Income	New C.I.
Below 5	0-5
5-10	5-10
10-15	10-15
15-20	15-20
20-25	20-25
25 and above	25-30

In this case, the size of each class is five. Thus, maintaining the uniformity, we take the first class as 0-5 and the last class as 25-30.

(iv) *With mid-points*

Sometimes, instead of the class intervals the mid points of various classes are given.

In such cases, we can determine the class intervals using the following adjustment formula.

$$\text{Value of Adjustment} = \frac{\text{Mid-Point of One Class} - \text{Mid-Point of Preceeding Class}}{2}$$

The value obtained is then added to the mid point to obtain the upper limit and subtracted from the mid-point to obtain the lower limit.

***Example*-** For the following distribution, determine the class intervals.

Mid-point	Frequency
5	1
15	5
25	3
35	4
45	2

Solution

Here, the class interval is calculated by the following value of adjustment.

$$\text{Value of adjustment} = \frac{15-5}{2} = 5$$

Thus, we add and subtract 5 to each mid-point to obtain the class interval.

For instance:

The lower limit of first class = 5 – 5 = 0
Upper limit of first class = 5 + 5 = 10.

Thus, the first class interval is 0-10. Similarly, we can calculate the remaining class intervals.

Class Interval	Frequency
0-10	1
10-20	5
20-30	4
30-40	3
40-50	2

Sometimes in the analysis we are required to calculate the mid-points of the class intervals. The mid-point of the class interval is calculated using the following formula.

$$\text{Mid-Point} = \frac{\text{Upper Limit} + \text{Lower Limit}}{2}$$

Example: For the class intervals given below, calculate the mid-points.

Income	Mid-Point
0-5	2.5
5-10	7.5
10-15	12.5
15-20	17.5
20-25	22.5
25-30	27.5

Solution:

$\text{Mid-Point}_{(\text{FirstClass})} = \frac{5-0}{2} = 2.5$ and so on.

Cumulative Frequency Distribution

In cumulative frequency distribution, the frequencies are presented in the cumulative form. In other words, they are presented as aggregate of frequencies of various classes. Cumulative frequency distribution can be of the following two types.

- Less Than Type
- More Than Type

Cumulative Frequencies of the Less Than Type

Less than cumulative frequencies are based on the upper limits of each class. For the calculation of the cumulative frequency of a particular class, we add the frequencies of all the preceding classes.

Example: Convert the following frequency distribution in less than type.

Marks	Frequency
0-5	6
5-10	9
10-15	12
15-20	18
20-25	25

Solution

Marks	Frequency (*f*)	Cumulative Frequency (*c.f*)
Less than 5	6	6

Less than 10	9	15 (= 6 + 9)
Less than 15	12	27 (= 6 + 9 + 12)
Less than 20	18	45 (= 6 + 9 + 12 + 18)
Less than 25	25	70 (= 6 + 9 + 12 + 18 + 25)

The cumulative frequency for less than 10 is calculated by adding all the frequencies prceeding 9 (i.e. 6 + 9). Similarly, cumulative frequency for less than 15 is calculated by adding all the frequencies prceeding 12 (i.e. 6 + 9+12) and so on.

Cumulative Frequencies of the More Than Type

More than cumulative frequencies are based on the lower limit of each class. Here, the cumulative frequencies are calculated as follows.

***Step* 1**: Calculate the total of frequencies.

***Step* 2**: Subtract the value of first frequency from the total of the frequencies.

***Step* 3**: For the cumulative frequency of the next class, subtract the frequency of the preceding class from the cumulative frequency of the preceding class. The cumulative frequencies of the subsequent classes are calculated in a similar manner.

Example: Consider again the illustration given above. Let us convert the frequency distribution in more than type.

Solution

Marks	**Frequency (*f*)**	**Cumulative Frequency (*c.f*)**
More than 0	6	70
More than 5	9	64 (= 70-6)

More than 10	12	55 (= 64-9)
More than 15	18	43 (= 55-12)
More than 20	25	25 (= 43-18)

Relative Frequency Distribution

Relative frequency distribution refers to the distributions having frequencies is relative terms. Relative frequencies are obtained when actual frequencies are expressed as percent of total frequencies.

For example, the income of different individual are given as follows.

Individual	Income (in Rs thousand)	Individual	Income (in Rs Thousand)
1	17	11	13
2	12	12	15
3	15	13	19
4	18	14	12
5	13	15	11
6	19	16	17
7	22	17	17
8	11	18	22
9	15	19	18
10	17	20	18

The relative frequency distribution for the given data set is as follows. The frequencies are calculated in a similar manner as discussed above. Each of the frequencies is then expressed as a percent of the total of frequencies.

Income (in thousand)	Tally Bars	Frequency (*f*)	Relative Frequency (%)
10-12	𝍸	4	20
12-14	||	2	10
14-16	𝍸	3	15
16-18	𝍸 ||	7	35
18-20	||	2	10
20-22	||	2	10
		20	100

Frequency Distribution with Unequal Classes

Sometimes, the given classes are of unequal size. In such cases, to make the distribution uniform certain adjustments are made in the frequencies. A general formula for the adjustment of the frequency is as follows.

$$\text{Adjusted Frequency} = \frac{\text{Required Class Interval} \times \text{Frequency}}{\text{Actual Class Interval}}$$

Example- Consider the following frequency distribution. Here, the class intervals are of unequal size. The size for the class (0 - 10) is 10, for the class (20 -40), the size is 20 and for the class (40 -70) is 30.

Class Interval	Frequency
0-10	2
10-20	5
20-40	7
40-70	9
70-80	10

Here, the class intervals are of unequal size. the size for the class (0-10) is 10, for the class (20-40), the size is 20 and for the class (40-70) is 30. So we adjust the

frequencies to make the distribution uniform. The adjustment is carried out in the following manner.

Class Interval	Frequency	Adjusted Frequency
0-10	2	–
10-20	5	–
20-40	8	$\frac{10\times8}{20}=4$
40-70	9	$\frac{10\times9}{30}=3$
70-80	10	–

Loss of Information

Classification and grouping of raw data into classes makes it more concise and understandable. But simultaneously there exists loss of information. The calculation involved in the classified data are generally on the mid-point.

The items in such series cannot be exactly measured and consequently, an individual observation loses its importance during statistical calculations. Also, the statistical calculations are based on the values of the class marks, ignoring the exact observation of the data resulting in the problem of loss of information.

Bivariate Frequency Distribution

Bivariate frequency distributions refer to those distributions where the value of two variables are presented simultaneously. Herein, the two variables are measured for the same set of items and are thereby, presented through cross classification. Such type of distributions are also known as '*two-way frequency distribution*'. For example, consider the the following table that presents the marks of students of two classes, A and B.

Class A	Class B	Class A	Class B
12	25	16	24
16	29	19	37
14	32	12	43
11	37	13	40
19	26	16	35
17	39	19	38
13	33	14	2

Here, the two variables are marks of students of class A and the marks of students of class B. This data can be presented in the form of a bivariate distribution as follows.

Marks (Class A)	Marks (Class B)					Total
	20-25	25-30	30-35	35-40	40-45	
10-12				| (1)		1
12-14		| (1)	| (1)		|| (2)	4
14-16		| (1)	| (1)			2
16-18	| (1)	| (1)		|| (2)		4
18-20		| (1)		|| (2)		3
Total	1	4	2	5	2	14

Here, marks of the students of class A are presented in the first column and the marks of the students of class B are presented in the first row. Now, to fill the values in the table, we find the combination of marks for class A and class B as per the different class intervals.

For instance, we first find a combination such that the marks for class A lies in the interval (10-12) and the marks for class B lies in the class interval (20-25). Next, we find a combination for the class interval (10-12) for class A and class interval (25-30) for class B. In the similar manner, we find the combination for class interval (10-12) for class A with the remaining class intervals of class B.

Next, we move to the second row to find combination for class interval (12-14) for class A with the different class intervals of class B. That is, combination of class interval (12-14) with the class intervals (20-25), (25-30), (30-35), (35-40) and (40-45) of class B. In the similar manner, we find the combination for different class intervals of class A and class B.

Following this procedure, consider first the combination (10-12) and (20-25) for class A and class B, respectively. In the given data we can see that there is no such combination of values. In a similar manner, consider the combination (10-12) and (35-40). Here, we can see that the combination (13, 33), i.e. 13 marks for class A and 33 marks for class B corresponds to the required combination. So, the frequency for this combination is 1. In a similar manner we can find the frequencies for the remaining combination of marks.

Difference between Univariate frequency Distribution and Bivariate frequency distribution

Univariate Frequency Distribution	Bivariate Frequency Distribution
The word '*Uni*' means one. A series of statistical data showing the frequency of only one variable is called Univariate Frequency Distribution. In other words, the frequency distribution of single variable is called Univariate Frequency Distribution. For example- income of people, marks scored by students, etc.	The word '*Bi*' means two. A series of statistical data showing the frequency of two variables simultaneously is called Bivariate Frequency Distribution. In other words, the frequency distribution of two variables is called Bivariate Frequency Distribution. For example- sales and advertisement expenditure, weight and height of individuals, etc.